



Universidad Autónoma del Estado de México

Centro Universitario UAEM Valle de Chalco

**PROPUESTA DE UNA MÉTRICA CUALITATIVA PARA
DETERMINAR EL SESGO ÉTICO POST PROCESAMIENTO
EN MACHINE LEARNING EDUCATIVO**

T E S I S

QUE PARA OBTENER EL GRADO DE

MAESTRA EN CIENCIAS DE LA COMPUTACIÓN

P R E S E N T A

ING. LUISA REYNA BEATRIZ AYON MUÑOZ

DIRECTOR

DR. SAMUEL OLMOS PEÑA

CO-DIRECTORA

MTRA. MA. DE LA PAZ SILVA BORJAS

TUTORA

DRA. MAGALLY MARTINEZ REYES

VALLE DE CHALCO SOLIDARIDAD, MÉXICO NOVIEMBRE 2024.



CUVCH

**PROPUESTA DE UNA MÉTRICA CUALITATIVA
PARA DETERMINAR EL SESGO ÉTICO POST
PROCESAMIENTO EN MACHINE LEARNING
EDUCATIVO**

ÍNDICE DE CONTENIDO

RESUMEN	8
I. INTRODUCCIÓN	9
Planteamiento del problema.....	13
Hipótesis.....	14
Objetivos	15
Objetivo general.....	15
Objetivos específicos	15
Delimitación o alcances de la investigación	15
Metodología de investigación	16
Estado del arte	18
Ética e inteligencia artificial aplicada a la educación	18
Estudio acerca del sesgo algorítmico	19
Evaluación de la equidad algorítmica	21
Vacíos localizados sobre el sesgo algorítmico y su evaluación	22
II. MARCO TEÓRICO-CONCEPTUAL.....	25
Tecnoética	25
Inteligencia artificial	26
Aprendizaje automático	28
Ciclo de vida de aprendizaje automático.....	30
Principios FATE.....	32
Sesgo algorítmico.....	34
Equidad algorítmica.....	38
Métricas cualitativas.....	45
Escala de Likert	47
Equidad sustantiva	49
Educación inclusiva y factores de discriminación.....	50
III. CONSTRUCCIÓN DE LA MÉTRICA	53
Identificación de variables sensibles.....	53
Proceso de construcción.....	54
1.- Comprender el Constructo.....	54
2.- Construcción de los Ítems	57
3.- Determinar el espacio de resultados.....	57
4.- Especificar el modelo de medida.....	58
5.- Recopilar comentarios y poner a prueba el cuestionario.....	59
IV. RESULTADOS	61
Fase de pruebas.....	65

Discusión de resultados	106
Conclusiones	107
Trabajo Futuro	108
V. REFERENCIAS	109
VI. ANEXOS.....	117
Anexo A. Primera versión de la escala de Likert.....	117
Anexo B. Interfaz para la escala de Likert.....	118
Anexo C. Prueba de la métrica al modelo 2.....	120
Anexo D. Prueba de la métrica al modelo 3.....	122
Anexo E. Prueba de la métrica al modelo 4.....	124
Anexo F. Prueba de la métrica al modelo 5.....	126
Anexo G. Prueba de la métrica al modelo 6.....	128
Anexo H. Prueba de la métrica al modelo 7.....	130
Anexo I. Prueba de la métrica al modelo 9.....	132
Anexo J. Prueba de la métrica al modelo 10.....	134

ÍNDICE DE ILUSTRACIONES

Ilustración 1. Ciclo de vida de los modelos de aprendizaje automático.....	31
Ilustración 2. Etapas de Aprendizaje automático.....	32
Ilustración 3. Principios FATE.....	33
Ilustración 4. Pasos para la construcción de una escala de Likert.....	49
Ilustración 5. Marco de apego a la equidad sustantiva.....	50
Ilustración 6. Espacio de Resultados propuesto.....	58
Ilustración 7. Interfaz construida.....	66
Ilustración 8. Primera parte del conjunto de datos a prueba.....	70
Ilustración 9. Segunda parte del conjunto de datos a prueba.....	70
Ilustración 10. Tercera parte del conjunto de datos a prueba.....	70
Ilustración 11. Predicciones del modelo 1.....	71
Ilustración 12. Variables sensibles del modelo 1.....	72
Ilustración 13. Primera parte de la contestación de ítems para el modelo 1.....	78
Ilustración 14. Segunda parte de la contestación de ítems para el modelo 1.....	79
Ilustración 15. Tercera parte de la contestación de los ítems para el modelo 1.....	79
Ilustración 16. Cuarta parte de la contestación de los ítems para el modelo 1.....	79
Ilustración 17. Quinta parte de la contestación de los ítems para el modelo 1.....	80
Ilustración 18. Sexta parte de la contestación de los ítems para el modelo 1.....	80
Ilustración 19. Séptima parte de la contestación de los ítems para el modelo 1.....	81
Ilustración 20. Octava parte de la contestación de los ítems para el modelo 1.....	81
Ilustración 21. Novena parte de la contestación de los ítems para el modelo 1.....	81
Ilustración 22. Parte final de la contestación de los ítems para el modelo 1.....	82
Ilustración 23. Nivel de sesgo obtenido en el modelo 1.....	82
Ilustración 24. Primera parte del nuevo conjunto de datos a prueba.....	87
Ilustración 25. Segunda parte del nuevo conjunto de datos a prueba.....	87
Ilustración 26. Tercera parte del nuevo conjunto de datos a prueba.....	88
Ilustración 27. Predicciones del modelo 8.....	88
Ilustración 28. Variables sensibles del modelo 8.....	89
Ilustración 29. Primera parte de la contestación de los ítems para el modelo 8....	102
Ilustración 30. Segunda parte de la contestación de los ítems para el modelo 8..	102
Ilustración 31. Tercera parte de la contestación de los ítems para el modelo 8. ...	102
Ilustración 32. Tercera parte del contestación de los ítems para el modelo 8.....	103
Ilustración 33. Cuarta parte de la contestación de los ítems para el modelo 8.....	103
Ilustración 34. Quinta parte de la contestación de los ítems para el modelo 8.....	103
Ilustración 35. Quinta parte de la contestación de los ítems para el modelo 8.....	104
Ilustración 36. Sexta parte de la contestación de los ítems para el modelo 8.....	104
Ilustración 37. Séptima parte de la contestación de los ítems para el modelo 8. ..	104
Ilustración 38. Parte final de la contestación de los ítems para el modelo 8.....	105
Ilustración 39. Nivel de sesgo obtenido en el modelo 8.....	105
Ilustración 40. Primer borrador de la dimensión A en la métrica.....	117
Ilustración 41. Primer borrador de la dimensión B en la métrica.....	117
Ilustración 42. Primer borrador de la dimensión C en la métrica.....	117
Ilustración 43. Primer borrador de la dimensión D en la métrica.....	117
Ilustración 44. Primera parte de la interfaz creada.....	118

Ilustración 45. Tercera parte de la interfaz creada.....	118
Ilustración 46. Quinta parte de la interfaz creada.	118
Ilustración 47. Segunda parte de la interfaz creada.....	118
Ilustración 48. Cuarta parte de la interfaz creada.	119
Ilustración 49. Sexta parte de la interfaz creada.....	119
Ilustración 50. Octava parte de la interfaz creada.....	119
Ilustración 51. Novena parte de la interfaz creada.....	119
Ilustración 52. Décima parte de la interfaz creada.	119
Ilustración 53. Parte final de la interfaz creada.	119
Ilustración 54. Predicciones del modelo 2.....	120
Ilustración 55. Variables sensibles del modelo 2.	120
Ilustración 56. Nivel de sesgo obtenido en el modelo 2.....	122
Ilustración 57. Predicciones del modelo 3.....	122
Ilustración 58. Variables sensibles del modelo 3.	123
Ilustración 59. Contestación de los ítems para el modelo 3.....	124
Ilustración 60. Predicciones del modelo 4.....	124
Ilustración 61. Variables sensibles del modelo 4.	125
Ilustración 62. Nivel de sesgo obtenido en el modelo 4.....	126
Ilustración 63. Predicciones del modelo 5.....	126
Ilustración 64. Variables sensibles del modelo 5.	127
Ilustración 65. Nivel de sesgo obtenido en el modelo 5.....	128
Ilustración 66. Predicciones del modelo 6.....	128
Ilustración 67. Variables sensibles del modelo 6.	129
Ilustración 68. Nivel de sesgo obtenido en el modelo 6.....	130
Ilustración 69. Predicciones del modelo 7.....	130
Ilustración 70. Variables sensibles del modelo 7.	131
Ilustración 71. Nivel de sesgo obtenido en el modelo 7.....	132
Ilustración 72. Predicciones del modelo 9.....	132
Ilustración 73. Variables sensibles del modelo 9.	133
Ilustración 74. Nivel de sesgo obtenido en el modelo 9.....	134
Ilustración 75. Predicciones del modelo 10.....	134
Ilustración 76. Variables sensibles del modelo 10.	135
Ilustración 77. Nivel de sesgo obtenido en el modelo 10.....	136

ÍNDICE DE GRÁFICAS

Gráfica 1. Predicciones del modelo 1.	72
Gráfica 2. Reprobados del modelo 1 respecto al sexo.	74
Gráfica 3. Reprobados del modelo 1 respecto a la zona donde vive.	75
Gráfica 4. Reprobados del modelo 1 respecto a la escuela donde estudia.	75
Gráfica 5. Reprobados del modelo 1 respecto al trabajo de la madre.	75
Gráfica 6. Reprobados del modelo 1 respecto al trabajo del padre.	76
Gráfica 7. Reprobados del modelo 1 respecto a la escolaridad del padre.	77
Gráfica 8. Reprobados del modelo 1 respecto a la escolaridad de la madre.	77
Gráfica 9. Predicciones del modelo 8.	89
Gráfica 10. Calificación DD obtenida del modelo 8 respecto al género.	93
Gráfica 11. Calificación DD obtenida del modelo 8 respecto si el estudiante tiene empleo.	95
Gráfica 12. Calificación DD obtenida del modelo 8 respecto al salario del estudiante.	95
Gráfica 13. Calificación DD obtenida del modelo 8 respecto al tipo de alojamiento.	95
Gráfica 14. Calificación DD obtenida del modelo 8 respecto al tipo de transporte. .	96
Gráfica 15. Calificación DD obtenida del modelo 8 respecto al tipo de bachillerato.	96
Gráfica 16. Calificación DD obtenida del modelo 8 respecto al trabajo de la madre.	96
Gráfica 17. Calificación DD obtenida del modelo 8 respecto al trabajo del padre. .	97
Gráfica 18. Calificación DD obtenida del modelo 8 respecto a la escolaridad de la madre.	97
Gráfica 19. Calificación DD obtenida del modelo 8 respecto a la escolaridad del padre.	98
Gráfica 20. Calificación DD obtenida del modelo 8 respecto a la nacionalidad.	99
Gráfica 21. Calificación DD obtenida del modelo 8 respecto a la raza.	99
Gráfica 22. Calificación DD obtenida del modelo 8 respecto a los antecedentes migratorios.	100
Gráfica 23. Predicciones del modelo 2.	121
Gráfica 24. Predicciones del modelo 3.	123
Gráfica 25. Predicciones del modelo 4.	125
Gráfica 26. Predicciones del modelo 5.	127
Gráfica 27. Predicciones del modelo 6.	129
Gráfica 28. Predicciones del modelo 7.	131
Gráfica 29. Predicciones del modelo 9.	133
Gráfica 30. Predicciones del modelo 10.	135

ÍNDICE DE TABLAS

Tabla 1. Clasificación de sesgo.	37
Tabla 2. Nociones de equidad.	43
Tabla 3. Métricas de equidad grupal.	43
Tabla 4. Niveles de sesgo y su rango de puntuación.	59
Tabla 5. Retroalimentación de expertos sobre la escala	60
Tabla 6. Análisis de variables sensibles en el modelo 1.	71
Tabla 7. Análisis del modelo 1 respecto al sexo.	73
Tabla 8. Análisis del modelo 1 respecto a la zona donde vive.	74
Tabla 9. Análisis del modelo 1 respecto a la escuela donde estudia.	74
Tabla 10. Análisis del modelo 1 respecto al trabajo de la madre y el padre.	74
Tabla 11. Análisis del modelo 1 respecto a la escolaridad de los padres.	76
Tabla 12. Reprobados del modelo 1 respecto a las variables sensibles.	78
Tabla 13. Análisis de variables sensibles en el modelo 8.	88
Tabla 14. Consideración de otras variables en el modelo 8.	91
Tabla 15. Análisis del modelo 8 respecto al género.	92
Tabla 16. Análisis del modelo 8 respecto a si el estudiante tiene empleo.	93
Tabla 17. Análisis del modelo 8 respecto al salario del estudiante.	93
Tabla 18. Análisis del modelo 8 respecto al tipo de alojamiento.	93
Tabla 19. Análisis del modelo 8 respecto al tipo de transporte.	94
Tabla 20. Análisis del modelo 8 respecto al tipo de bachillerato.	94
Tabla 21. Análisis del modelo 8 respecto al empleo de los padres.	94
Tabla 22. Análisis del modelo 8 respecto a la escolaridad de los padres.	97
Tabla 23. Análisis del modelo 8 respecto a la nacionalidad.	98
Tabla 24. Análisis del modelo 8 respecto a la raza.	98
Tabla 25. Análisis del modelo 8 respecto a los antecedentes migratorios.	98
Tabla 26. Calificaciones DD del modelo 8 respecto a las variables sensibles de entrada.	100
Tabla 27. Calificaciones DD del modelo 8 respecto al resto de variables sensibles.	101
Tabla 28. Reprobados del modelo 2 respecto a las variables sensibles	121
Tabla 29. Reprobados del modelo 3 respecto a las variables sensibles.	123
Tabla 30. Reprobados del modelo 4 respecto a las variables sensibles	125
Tabla 31. Reprobados del modelo 5 respecto a las variables sensibles	127
Tabla 32. Reprobados del modelo 6 respecto a las variables sensibles	129
Tabla 33. Calificaciones CB del modelo 7 respecto a las variables sensibles de entrada.	131
Tabla 34. Calificaciones CB del modelo 7 respecto al resto de variables sensibles.	132
Tabla 35. Calificaciones reprobatorias y DD del modelo 9 respecto a las variables sensibles de entrada.	133
Tabla 36. Calificaciones reprobatorias y DD del modelo 9 respecto al resto de variables sensibles.	134
Tabla 37. Calificaciones reprobatorias del modelo 10 respecto a las variables sensibles de entrada.	135

Tabla 38. Calificaciones reprobatorias del modelo 10 respecto al resto de variables sensibles.	136
---	-----

RESUMEN

Con el acelerado uso de la tecnología y la inteligencia artificial en el área educativa, resulta importante prestar atención a una rama conocida como aprendizaje automático (machine learning en inglés). La relevancia radica en que numerosas investigaciones han identificado la presencia de respuestas sesgadas y, por lo tanto, discriminatorias hacia determinado grupo de estudiantes en implementaciones desarrolladas, debido a que es utilizado en la predicción de patrones, sistemas de recomendación y toma de decisiones automatizada, lo cual impacta directamente y de manera negativa en grupos estudiantiles. El sesgo mencionado, ha sido abordado desde la perspectiva de la detección y mitigación como sesgo algorítmico por medio de métricas cuantitativas. En el presente trabajo de investigación, opta por abordar un enfoque cualitativo para proponer una métrica que permita la detección del sesgo mediante el uso de una escala de Likert en la fase de post procesamiento, siendo esta etapa en la que se obtienen los datos después de procesar entradas con un algoritmo seleccionado, es decir, se presta atención a la respuesta obtenida; de igual forma, se llevan a cabo pruebas con modelos disponibles en repositorios de acceso libre y se observa el cumplimiento de su objetivo.

I. INTRODUCCIÓN

A lo largo del tiempo, el ser humano ha desarrollado e implementado tecnologías para alcanzar metas y cubrir determinadas necesidades en diferentes sectores, las cuales se componen de técnicas para la elaboración de actividades; estas han sido implementadas, replicadas y mejoradas con la finalidad de obtener mayor precisión en los resultados, facilitar su realización y optimizar los recursos involucrados en su ejecución.

Entre los intentos por crear mejoras, destaca la automatización de procesos, donde interviene el uso de la tecnología, como el hardware y el software para realizar actividades y tareas de forma automática en lugar de hacerlo manualmente, o bien, con la mínima intervención humana posible.

Sumado a lo anterior, se observa una tendencia creciente del uso de la inteligencia artificial para tratar de replicar e imitar la inteligencia de las personas lo cual ha llevado a automatizar procesos mediante esta disciplina. Existen varios campos en los cuales se divide la inteligencia artificial, siendo uno de ellos el machine learning o aprendizaje automático, empleado en gran medida para la toma de decisiones en diversos sectores tales como la medicina (Shailaja, Seetharamulu y Jabbar, 2018), farmacología (Lo et al., 2018; Patel, et al., 2020), finanzas (Henrique, Bobreiro y Kimura, 2019), o la educación (Williams et al., 2016; Kashi, Srinivasa y Deshpande, 2016). Para el desarrollo de la presente investigación, se hace referencia a esta rama de la inteligencia artificial con su nombre en español, es decir, aprendizaje automático.

El aprendizaje automático busca imitar el aprendizaje basado en la experiencia de los seres humanos, por lo que no emplea una programación específica de funciones o procedimientos que dicten cada una de las acciones realizadas por el software para obtener una salida, en su lugar, hace uso de un entrenamiento mediante un conjunto de datos que permiten “aprender” a

dar una respuesta adecuada ante casos similares.

El aprendizaje automático en el área educativa tiene variadas aplicaciones, Kučak, Juričić y Đambić (2018) identifican, en una revisión sistemática de la literatura, principalmente cuatro usos de estos algoritmos: calificar a los estudiantes y evaluarlos (Ndukwe, Daniel y Amadi, 2019), predecir el desempeño académico (Hussain et al., 2018) y disminuir la deserción escolar buscando apoyar a aquellos educandos que presenten más probabilidades de abandonar sus estudios. (Cardona et al., 2020; Hussain et al., 2018; Opazo et al., 2021).

Otra línea de esta tecnología apunta a la personalización del aprendizaje basado en un diagnóstico al alumno para ofrecer herramientas adecuadas que favorezcan a mejorar el proceso enseñanza-aprendizaje en un estudiante específicamente, por lo general en ambientes de e-learning (Deena y Raja, 2020; Khanal et al., 2020; Li y Zhang, 2017); de igual forma, aparecen investigaciones relacionadas con la evaluación de la enseñanza o la motivación académica de cada estudiante (Babic, 2018).

Los datos de entrada para cada modelo dependen del objetivo a lograr y los criterios de cada autor, por ejemplo, en su investigación, Akmese, Kör y Erbay (2021) emplean diferentes algoritmos de aprendizaje automático considerando variables sociodemográficas para determinar el rendimiento académico de estudiantes que toman cursos a distancia y su posible éxito escolar; Ndukwe, Daniel y Amadi (2019) requieren que los alumnos contesten algunos cuestionamientos específicos realizados por un chatbot con la finalidad de asignarles una calificación; o bien Hussain et al., (2018) que además de las respuestas a ciertos ejercicios de diseño digital, toman en cuenta las actividades realizadas para cada uno, el tiempo promedio de inactividad, el promedio de pulsación de teclas, entre otros.

A pesar de los beneficios brindados por los algoritmos de Aprendizaje

automático, es importante considerar que los resultados obtenidos pueden presentar deficiencias y ocasionar afectaciones de manera considerable en la toma de decisiones, predicciones y en los actores que dependen de ello; a esto se le ha denominado como sesgo algorítmico y ha sido abordado desde diversas perspectivas (Sun, Nasraoui, y Shafto, 2020), no obstante, para esta investigación se recurre al término “sesgo ético”, abordándolo como aquel sesgo que ocasiona problemas éticos y no únicamente variaciones del resultado esperado.

Si bien, los algoritmos de Aprendizaje automático han adquirido gran relevancia para la toma de decisiones automatizada en modelos enfocados al campo educativo, autores como Baker y Hawn (2021), Yu et al., (2020) y Smith (2020) hacen notable en sus investigaciones la existencia de sesgo en las respuestas obtenidas y remarcan el impacto negativo que esto trae a los estudiantes, quienes en realidad deberían ser los beneficiados en cuanto a su aprendizaje; además, dicho evento genera dudas respecto a la integridad y ética considerada para la obtención del resultado.

Baker y Hawn (2021) expresan su preocupación frente a dicha situación, pues si bien, detectan investigaciones sobre el tema en la educación, la mayoría de ellas se enfocan en el análisis de la discriminación respecto a categorías como raza o etnia, nacionalidad y género, dejando sin atención otras variables como la presencia de discapacidades, por lo cual invitan a la profundización al respecto.

Hagendorff (2020) y Mörch et al., (2021) ponen en evidencia que aún existen omisiones de la relación entre la ética y la inteligencia artificial; ante los esfuerzos por buscar impulsar el estudio y enseñanza de aspectos éticos en el desarrollo de estas tecnologías (Bogina et al., 2021; Skirpan et al., 2018) incluyendo el Aprendizaje automático (Saltz et al., 2019), se puede apreciar la necesidad de contribuir en mayor medida al área, puesto que en su aplicación para la toma de decisiones, el resultado sesgado impacta de manera negativa

a la población dependiente (Chen et al., 2021; Li y Zhang, 2017; Miller, 2019; Zook et al., 2017).

Dicho lo anterior, es imperativo que las respuestas obtenidas de los modelos de Aprendizaje automático en las cuales puede basarse la determinación de una decisión reflejen ecuanimidad. Ante esta situación, los desarrolladores implementan métricas cuantitativas de equidad, con las que se llega a determinar si hay errores y por ende el modelo no es equitativo; estas son aplicadas en diversas fases del procesamiento en aprendizaje automático (Caton y Haas, 2020; Cantero, 2021) es decir antes del procesamiento (Celis, Keswani y Vishnoi, 2020; Kamiran y Calders, 2011), durante (Zhang, Lemoine y Mitchell, 2018) y después del mismo (Lohia, 2021).

El presente trabajo de investigación se centra en la fase de post procesamiento con la finalidad de analizar las respuestas obtenidas en los modelos de Aprendizaje automático aplicados al área educativa y, mediante una métrica cualitativa propuesta, determinar el nivel de presencia de sesgo ético detectado.

En este capítulo se muestra un breve contexto respecto al Aprendizaje automático dentro del área educativa, así como el planteamiento del problema identificado, el cual, da paso a la creación de la pregunta de investigación que guía el trabajo. Posteriormente, se presenta la hipótesis propuesta, los objetivos a alcanzar y una breve descripción de la metodología de investigación utilizada para el desarrollo de esta tesis. Finalmente, se muestra el estado del arte construido a partir de la revisión documental.

En el resto del documento, se localizan los capítulos enfocados al marco teórico-conceptual, donde se sitúan algunos términos, conceptos, principios y teorías relevantes que dan sustento al desarrollo de la investigación; la construcción de la métrica propuesta y la discusión de resultados, donde se abordan las conclusiones obtenidas y trabajo a futuro.

Por último, se presenta el apartado de referencias consultadas y los anexos que complementan el documento.

Planteamiento del problema

La responsabilidad depositada en los algoritmos de Aprendizaje automático para la toma de decisiones va en aumento, manifestándose en gran cantidad de áreas donde el ser humano se desarrolla. Una predicción o decisión errónea por parte de esta tecnología, podría traer graves consecuencias como la pérdida de la libertad, del empleo, etc. Ejemplo de ello fue el sistema COMPAS, cuyo objetivo era ayudar al sistema judicial de Estados Unidos prediciendo la posible reincidencia delictiva de personas arrestadas; el problema radicaba en la notable inclinación del sistema a calificar individuos de raza afroamericana como sujetos con un riesgo mayor de cometer delitos nuevamente, aunque la realidad reflejaba una situación distinta. (Angwin, et al., 2016).

Ante sucesos similares (Garcia, 2016; Lee, 2018), la preocupación por disminuir el sesgo ético que aparece en los resultados de Aprendizaje automático se ha ido extendiendo. Actualmente existe una variedad de métricas cuantitativas para determinar si hay equidad en los algoritmos durante las etapas de preprocesamiento, in procesamiento y post procesamiento (Caton y Haas, 2020; Ashokan y Haas, 2021). Una vez que se aplican las métricas correspondientes, se busca eliminar el sesgo presente; así también, ha sido posible desarrollar algunas librerías y herramientas que contribuyen a su detección y mitigación, utilizadas en lenguajes de programación como Python (Cantero, 2021).

Estas evaluaciones han sido llevadas a campos distintos para contribuir con soluciones al problema del sesgo ético, sin embargo, en la educación aún es necesario profundizar su estudio (Baker y Hawn, 2021), pues la integración de aprendizaje automático en este campo es cada vez más visible y la

aparición de resultados sesgados también se ha observado (Smith, 2020).

La mayoría de las métricas encontradas en la literatura son cuantitativas, y surge el interés por proponer un enfoque cualitativo que defina una medición del sesgo ético en aplicaciones educativas, puesto que autores como Fazelpour y Danks (2021) y Sahlgren (2021) mencionan la importancia de integrar elementos apegados al contexto social donde implementa el modelo, así como tratarlo y considerarlo desde una perspectiva sociotécnica. Aunado a ello, la formalización de aspectos sociales como la equidad, también puede contener sesgo.

Si se analizan sólo los datos para entrenamiento del sistema, la métrica estaría principalmente enfocada a la información entrante, y, por otra parte, si el sesgo se evalúa durante el procesamiento, podría existir un conflicto con aquellos algoritmos cuyo funcionamiento es difícil de comprender. En consecuencia, la pregunta que surge de lo antes mencionado y guía el presente trabajo de investigación es la siguiente:

¿Cómo determinar el nivel de presencia del sesgo ético durante la fase “post procesamiento” de algoritmos de Aprendizaje automático educativo con un enfoque diferente a métricas ya diseñadas?

Hipótesis

Establecer una escala de Likert con base en variables sensibles relacionadas a la discriminación dentro del área educativa y en características del contexto de implementación, permite determinar el nivel de presencia del sesgo ético durante el post procesamiento en Aprendizaje automático aplicado a la educación.

Objetivos

Objetivo general

Desarrollar una métrica cualitativa que determine la existencia del sesgo ético y el nivel que representa su presencia tras el procesamiento de datos reflejado en los resultados obtenidos por algoritmos de Aprendizaje automático aplicados al ámbito educativo.

Objetivos específicos

- Identificar variables sensibles relacionadas al sesgo ético en algoritmos de Aprendizaje automático aplicados en el área educativa.
- Localizar características del contexto de implementación involucradas con la presencia de sesgo en Aprendizaje automático aplicado a la educación.
- Especificar el concepto de equidad que se busca satisfacer con las respuestas proporcionadas por algoritmos de Aprendizaje automático enfocados a la educación.
- Construir ítems que permitan la detección de sesgo conforme el concepto de equidad y las variables sensibles identificadas en algoritmos de Aprendizaje automático educativo.
- Determinar los niveles de respuesta en forma de variables cualitativas ordinales para los ítems propuestos.
- Implementar el modelo de medición adecuado para interpretar y evaluar los niveles de respuesta construidos.

Delimitación o alcances de la investigación

Este trabajo de investigación se limita a elaborar la propuesta de una métrica cualitativa que se encargará de evaluar aquellos modelos de Aprendizaje automático aplicados exclusivamente al área educativa, ya que la selección de variables se enfoca a los problemas de discriminación y características detectadas por diversos autores en dicho campo.

Por otra parte, la métrica construida para esta investigación representa una herramienta complementaria para la detección de sesgo y no pretende sustituir las métricas ya existentes, sino coadyuvar a la elaboración de modelos que no remarquen diferencias entre grupos sociales favoreciendo a la discriminación y perjuicio hacia los estudiantes.

Esto conlleva a mencionar que, si bien, dentro de la literatura se localizan métricas y herramientas relacionadas con la identificación y mitigación del sesgo, la presente propuesta se encarga de abordar únicamente la medición cualitativa del sesgo ético post procesamiento.

Diversos autores, conciben cada métrica como una noción de equidad, por lo cual es importante aclarar que la métrica planteada en este documento pretende ser un instrumento de detección haciendo uso de una noción específica conocida como equidad sustantiva.

Finalmente, la realización de pruebas para determinar el correcto funcionamiento de la métrica va enfocada a evaluar modelos de Aprendizaje automático aplicados al área educativa que se encuentren disponibles en repositorios de acceso abierto junto con el conjunto de datos empleado para su entrenamiento.

Metodología de investigación

El seguimiento de una metodología para llevar a cabo la investigación, permite estructurarla adecuadamente y explicar cómo se desarrolló para obtener los resultados finales y llegar a conclusiones. El presente trabajo involucra seis etapas descritas brevemente a continuación.

En primer lugar, se llevó a cabo la revisión estructurada de la literatura en diversas bases de datos como Springer, Mendeley y Google académico para localizar la información referente al tema de investigación; en los documentos recuperados se consideraron incluir artículos, actas de

conferencias y algunos capítulos de libro relevantes que contribuyeron a la elaboración del trabajo. La búsqueda de información involucró el uso de palabras claves y operadores booleanos, así como otros criterios de inclusión y exclusión que facilitan la selección de documentos.

Posteriormente y a partir de la revisión, se identificaron y sintetizaron puntos clave como las perspectivas con las cuales ha sido analizada la ética en la inteligencia artificial aplicada a la educación, el sesgo algorítmico y su evaluación; las preguntas formuladas que han guiado las investigaciones tomadas en cuenta; los vacíos detectados en la literatura y aquellas áreas de debate existentes. Estos elementos permitieron la construcción del estado del arte, tomando en cuenta el periodo de 2019 a 2023 para la selección de los documentos analizados.

De igual forma, se localizaron conceptos, principios y teorías relevantes para brindar el sustento adecuado a la presente propuesta, considerando lo relacionado a la tecnoética y aterrizando en el área de la inteligencia artificial y Aprendizaje automático en el campo educativo; la educación inclusiva, factores de discriminación detectados por diversos autores, el método empleado para construir la métrica cualitativa, etc. En este apartado, se incluyeron documentos con diferentes años de publicación y algunos ya analizados para la elaboración del marco teórico conceptual.

A partir de la información detectada, se identificaron aquellas características presentadas por los estudiantes que han sido asociadas con la discriminación dentro del área educativa y con resultados discriminatorios en modelos de Aprendizaje automático aplicados a la educación, así como del contexto donde se implementan.

Después de ello, se tomaron como base las características localizadas considerándolas variables sensibles, para construir una escala de Likert bajo la propuesta de Nemoto y Beglar (2014) consistente en 5 puntos: comprensión

del constructo, desarrollo de los ítems, determinación del espacio de resultados, especificación del modelo de medida, así como la recopilación de los comentarios y puesta a prueba del cuestionario.

Finalmente, se procedió a buscar modelos de Aprendizaje automático aplicados a la educación en repositorios como GitHub, los cuales fueran de acceso libre para la implementación de la propuesta y determinación del sesgo ético en los modelos seleccionados. Estas últimas fases se reflejan en el capítulo 3.

Estado del arte

De acuerdo con el análisis y síntesis de la literatura publicada entre 2019 y 2023, se localizaron aspectos relevantes sobre el estado actual de la temática.

Ética e inteligencia artificial aplicada a la educación

Los últimos años se ha visto el crecimiento considerable en la implementación de Inteligencia artificial en diferentes áreas; como antecedente en el área educativa, se observa un aumento en el interés sobre el paradigma basado en datos (Hilbert et al., 2021). Además, es visible el intento por documentar el desarrollo, implementación y evaluación de sistemas de Inteligencia artificial, incluyendo modelos de aprendizaje automático. Sus usos en la educación se aprecian en tareas como las evaluaciones STEM, calificación de ensayos (Ramesh y Sanampudi, 2022; Zhai, Shi y Nehm, 2021), evaluaciones científicas (Zhai, Krajcik y Pellegrino, 2021), y otras tareas automatizadas (Chaudhry y Kazim, 2022; Xu y Ouyang, 2022; Selwyn et al., 2023).

En este rubro, se localiza el término Inteligencia artificial Aplicada a la Educación o AIED, sobre la cual varios autores han tenido la intención de estudiar su relación con la ética y estado actual, abordando generalidades y desafíos (Dieterle, 2022; Akgun y Greenhow, 2022); su responsabilidad para evitar los daños ocasionados por los algoritmos (Hagendorff, 2020; Suresh y Guttag, 2021); una nueva perspectiva para observar la ética como logro

sociotécnico identificado en las relaciones de poder existentes entre estudiantes, profesores y empresas (Henry y Oliver, 2022); la consideración de principios éticos en el desarrollo de las tecnologías (Chaudhry y Kazim, 2022; Nguyen et al., 2023); el nivel de conocimiento y educación que tienen los involucrados acerca de procesos de diseño, despliegue, uso de las tecnologías y principios de equidad, rendición de cuentas, transparencia y ética (FATE) (Bogina, 2021); o bien principios éticos ya existentes que guían la creación de nuevas propuestas (Nguyen et al., 2023).

Dentro de esa línea de investigación, se aprecian algunos trabajos sobre rendición de cuentas (Sahlgren, 2021; Infante, 2021), principio que requiere hacer uso de la equidad algorítmica para demostrar la ausencia de discriminación y con ello la confiabilidad de los sistemas analizados. Así mismo, se encuentra una inclinación hacia la identificación de problemas éticos en los resultados de modelos de Aprendizaje automático ya diseñados para el área educativa, los cuales amplifican la discriminación existente en los sistemas educativos de la nación donde se realiza el estudio o se implementa la tecnología a evaluar (Jeong et al., 2022; Smith, 2020).

Si bien la ética en la tecnología y el sesgo algorítmico son abordados por separado en los trabajos seleccionados, es notoria la relación entre ambos términos, puesto que la discriminación resultante en el proceso de toma de decisiones automatizada involucra la identificación de problemas éticos.

Estudio acerca del sesgo algorítmico

Sobre el sesgo algorítmico en Aprendizaje automático, se han tomado diferentes direcciones: algunos trabajos van orientados a una descripción teórica del término, donde se incluyen su definición, detección, fuentes, desafíos, taxonomía, etc. (Pagano et al., 2023; Mehrabi et al., 2021), de forma general o apegadas a un contexto educativo (Kizilcec y Lee, 2021; Hilbert et al., 2021); así como la noción respecto a su relación con el sesgo del sistema social donde se despliega y utiliza (Hernandez, 2023).

Otros estudios se enfocan en identificar mecanismos de evaluación orientados a métricas de equidad, considerando la presencia de ésta característica como base para reconocer si existe sesgo (Caton y Haas, 2020; Pessach y Shmueli, 2020); implementar dichas métricas en sistemas construidos que solucionan problemáticas del área educativa (Rzepka et al., 2022; Le Quy, Fridge y Ntoutsy, 2023; Palacios et al., 2021); y relacionar los mecanismos de evaluación con procesos de auditoría y evaluación de riesgos éticos (Hasan et al., 2022).

También se observa el análisis sobre el manejo de variables sensibles que pueden o no condicionar los resultados erróneos por parte de los algoritmos en cuestión y son un elemento crucial en la aplicación de las métricas (Baker y Hown, 2021; Rajendran, Chamundeswari y Sinha, 2022; Yu, Lee y Kizilcec, 2021); así como la explicación o uso de herramientas para la mitigación de sesgo (Zhang, Xing y Li, 2023; Deho et al., 2022).

El sesgo algorítmico de forma técnica es definido como la ausencia de equidad (Fazelpour y Danks, 2021), sin embargo, llega a entenderse de manera general como la discriminación hacia determinado grupo que representa un sector de la población, (Pagano et al., 2023) guardando relación con categorías protegidas por la ley (Rzepka et al., 2022; Le Quy, Fridge y Ntoutsy, 2023; Yu, Lee y Kizilcec, 2021; Baker y Hown, 2021). Se encuentran varios tipos de sesgo que pueden manifestarse en los modelos de Aprendizaje automático (Suresh y Gutttag, 2021; Fazelpour y Danks, 2021):

- Moral: Muestra la dependencia de los resultados a categorías protegidas.
- Estadístico: Observable cuando hay diferencias entre los datos de entrenamiento y las predicciones finales.
- Histórico: Referente al sesgo ya existente dentro del contexto social.

- De representación: Relacionado a un muestreo erróneo de los datos de entrenamiento.
- De medición: Surge en algoritmos de Aprendizaje automático supervisado al definir etiquetas.
- De agregación: Originado por el uso indiscriminado de un mismo modelo para problemáticas distintas.
- De aprendizaje: Proviene de una amplificación en las desigualdades por parte del algoritmo.
- De evaluación: Involucra un manejo inadecuado de los datos que proveen el conjunto de entrenamiento.
- De despliegue o de implementación: Ocurre por un mal uso del sistema.

Evaluación de la equidad algorítmica

Respecto a la evaluación de equidad, se estiman al menos 21 definiciones para este concepto, siendo tratadas como métricas dentro de alguna implementación. Las más abordadas son:

- Paridad demográfica. Es una métrica grupal en la cual se identifica si existe la misma probabilidad de obtener un resultado favorable en todos los grupos.
- Probabilidades igualadas. Consiste en determinar si la probabilidad de obtener resultados positivos o negativos es la misma.
- Igualdad de oportunidades. Está centrada en la tasa de verdaderos positivos, observando si es igual para todos los grupos.

Las tasas de falsos positivos y negativos son ampliamente utilizadas y pueden considerarse pieza clave para determinar la equidad en un modelo de aprendizaje automático, puesto que forman parte de los cálculos en distintas métricas (Pagano et al., 2023) o facilitan la identificación de sesgo al deducir si existe un “beneficio de la duda”, es decir, falsos positivos que reflejan la preferencia hacia determinado grupo (Jeong et al., 2022).

Autores han contribuido a la realización de propuestas que funcionan como métricas de equidad cuya naturaleza es cuantitativa, algunas de ellas son: ABROCA, PreCOF, herramientas exploratorias gráficas o puntuaciones de equidad (Gardner, Brooks y Baker, 2019; Goethals, Martens y Calders, 2023; Mukhopadhyay, 2022; Agarwal, Agarwal y Agarwal, 2023) y la medición conjunta de equidad y precisión (Lässig, Oppold y Herschel, 2022).

A pesar de estar presentes diversos planteamientos con el fin de analizar la temática con otras perspectivas, es notoria la orientación hacia una implementación práctica y diseño de métricas cuantitativas para la equidad orientadas al conjunto de datos que participa en el entrenamiento del algoritmo o bien, sin situarlas explícitamente en una etapa del ciclo de vida del aprendizaje automático (Mehrabi et al., 2021; Bogina et al., 2021) para determinar la presencia o ausencia de sesgo.

Estas métricas han sido dirigidas de forma general a modelos de Aprendizaje automático sin importar el área del conocimiento donde son empleados. No obstante, se aprecia el comienzo de una apertura hacia visiones de evaluación acorde al contexto social involucrado (Sahlgren, 2021; Hasan et al., 2022; Madaio, et al., 2021), el diseño especulativo (Gaskins, 2022), la incorporación de métodos de ciencias sociales ante la naturaleza del problema (Grote, 2024); la consideración de la interacción entre usuarios, docentes, las respuestas obtenidas y el código (Hasan et al., 2022); el fortalecimiento de las relaciones entre los actores involucrados (Henry y Oliver, 2022) y la participación de profesionales en filosofía.

Vacíos localizados sobre el sesgo algorítmico y su evaluación

La indagación sobre el sesgo en algoritmos de Aprendizaje automático aún se encuentra en desarrollo, pues existen algunas vertientes que no han sido del todo exploradas. Un claro ejemplo es el nulo desarrollo de métricas para el sesgo, puesto que sólo se realiza una identificación de este a través de métricas de equidad.

No hay un consenso respecto a la más adecuada para cualquier sistema (Sahlgren, 2021) y varios autores reconocen el “teorema de la imposibilidad”, asegurando que no pueden satisfacerse todas las métricas de equidad en forma simultánea al realizar una evaluación (Sahlgren, 2021; Deho et al., 2022). Pocos trabajos examinan una perspectiva sociotécnica y apegada al contexto, ya que predomina una visión estadística (Hasan et al., 2022; Henry y Oliver, 2022; Sahlgren, 2021).

Por otra parte, dentro del reconocimiento de variables sensibles requiere ampliarse la exploración para integrar elementos apegados a la educación en general y no sólo aquellos establecidos por las legislaciones del país donde se implementa el modelo, ya que dentro de la literatura se reconoce la tendencia a involucrar en la mayoría de los casos el marco legal de Estados Unidos y la Unión Europea al ser las naciones donde se originó el estudio (Akgun y Greenhow, 2022; Jeong et al., 2022; Catania, Guerrini y Accinelli, 2023), lo cual excluye el contexto de otras naciones.

Algunos trabajos han explicado una controversia sobre quiénes deberían analizar y definir la ética en la tecnología, área en la cual se incluye la equidad algorítmica (Sahlgren, 2021), argumentando la necesidad de considerar argumentos y elementos filosóficos para su definición (Fazelpour y Danks, 2021; Giovanola y Tiribelli, 2022) y abogando por la realización de investigaciones éticas. Por ende, aún existe la discusión sobre cómo se define realmente el sesgo no permitido en los sistemas (Sahlgren, 2021) y si éste proviene de una falla en el modelo de Aprendizaje automático o es un reflejo de los sesgos en la sociedad (Mayfield et al., 2019; Hernández, 2023).

Por encima de las diferencias entre los objetivos de cada estudio, es posible localizar la necesidad de adentrarse aún más en el rubro de los principios éticos, la equidad y el sesgo algorítmico debido al aumento en su participación para colaborar con procesos de toma de decisiones educativas

y el gran impacto que genera cada determinación en la vida de los estudiantes (Gaskins, 2022; Dieterle, Dede y Walker, 2022). De igual forma, se busca impulsar la creación de soluciones ante problemas de discriminación (Baker y Hown, 2021; Bogina et al., 2021; Hagendorff, 2020) para evitar continuar y fomentar las injusticias dentro del contexto educativo.

Bajo los resultados obtenidos en la revisión de la literatura, es posible reconocer la intención de centrarse en el último principio FATE de la inteligencia artificial, es decir, la ética. De igual forma, la investigación actual, denota la importancia de abordar el sesgo en Aprendizaje automático para contribuir en el campo educativo; la formación de enfoques diferentes a la formalización para determinar el sesgo aún se encuentra en desarrollo, no obstante, ya existen propuestas como se describió anteriormente, por lo que la presente investigación busca participar en el avance de ello y aportar en las Ciencias Computacionales.

II. MARCO TEÓRICO-CONCEPTUAL

A lo largo del segundo capítulo se muestra la descripción de conceptos y teorías relevantes para el trabajo de investigación. Comenzando por los aspectos generales como la tecnoética e inteligencia artificial, se identifican elementos más específicos de cada uno, tal como el Aprendizaje automático en el cual se enfoca esta investigación, los principios FATE, el sesgo y equidad algorítmica. También se integra todo lo relacionado a métricas cualitativas y escalas de medición, aterrizando en la escala de Likert y el método seleccionado para su construcción. Finalmente, se retoma la noción de equidad sustantiva a la cual se apegan la métrica propuesta, educación inclusiva y aquellos factores de discriminación identificados en el área educativa.

Tecnoética

Ante la continua evolución y transformación que se vive actualmente, es necesario considerar las repercusiones de cada implementación tecnológica dentro del contexto social (Vivas, 2018). La tecnoética es una rama de estudio que abarca más de una disciplina, pues involucra el conocimiento y aplicación de la ética y la moral presente en la tecnología que interactúa con un grupo específico de la sociedad (o de manera general) (Luppicini 2009).

Este campo se encuentra relacionado con el área de la comunicación, lógicamente con los estudios de tecnología para el desarrollo de alguna innovación y con información relevante sobre las ciencias sociales. Por otra parte, Gearhart (2004) menciona su posible injerencia para evaluar políticas y leyes existentes encargadas de proteger a las personas afectadas por problemáticas que pueden desembocar tras el desarrollo y uso de una tecnología.

Echeverría (2010) marca la década de 1970 como el periodo donde se concibió el término en cuestión, siendo la aportación más relevante aquella

emitida por Mario Bunge en 1977, quien mencionó la importancia de establecer una ética que permitiera a los tecnólogos considerar su responsabilidad moral en el ámbito donde se desenvuelven. Su idea implicaba construir reglas y teorías que abarcaran principios éticos y morales dirigidos a guiar el quehacer científico y tecnológico.

Omand y Phythian (2022) explican que el tecnólogo debe acercarse a la parte ética del proceso, y no cerrarse a relacionar la aparición de problemas por un mal uso del usuario final, sino tener en cuenta las posibles consecuencias negativas ocasionadas por alguna tecnología desde el momento del desarrollo y la implementación.

Existen varias amenazas reconocidas por la tecnoética. Lee (2020) muestra algunas de ellas relacionadas a la Inteligencia artificial y subdisciplinas de esta, destacando entre ellas el Aprendizaje automático. Dicho autor, enuncia como peligros directos la presencia de sesgo en los algoritmos y falta de equidad en los resultados, atribuyendo la responsabilidad de ambas problemáticas a las personas encargadas de diseñar y entrenar el modelo.

Así también, considera que el desafío por contrarrestar las amenazas es cada vez más grande y debe ser considerado dentro de la investigación, puesto que hay vacíos al respecto y el impacto sobre la sociedad tras recibir respuestas sesgadas es significativo, ya que se utilizan en procesos de toma de decisiones que afectan la vida de los involucrados.

Inteligencia artificial

El término fue utilizado por primera vez en los años 50's por John McCarthy, quien enfocó el concepto a la realización de acciones o comportamientos hechos por una máquina que podrían ser vistos como "inteligentes" si un ser humano los realizara (Kaplan, 2016). Si bien, Wang (2019) expone la falta de una definición universalmente aceptada, la mayoría de los autores se apegan

a describirla como aquella disciplina en la cual se intenta replicar, reproducir o imitar el intelecto humano mediante el uso de máquinas y computadoras (Garnham, 2017). Rouhiainen (2018) menciona como característica destacable la capacidad de los dispositivos para el manejo de datos en grandes cantidades, así como el trabajo continuo.

La inteligencia artificial se encuentra dividida en campos diferentes, los cuales han sido abordados y desarrollados para diversas aplicaciones. Algunas de ellas son mencionadas por Köchling y Wehner (2020), Hashimoto et al. (2018), Rouhiainen (2018) y Tandon et al. (2019):

- Procesamiento de Lenguaje Natural.
- Reconocimiento de voz e imágenes.
- Aprendizaje automático o Aprendizaje automático.
- Aprendizaje profundo o Deep Learning.
- Visión por computadora.

A lo largo del tiempo, ha sido aplicada a otras áreas del conocimiento como la educación, donde se busca mejorar los procesos de enseñanza-aprendizaje y la toma de decisiones. Gracias a la implementación y desarrollo de tecnologías destinadas al área mencionada, se ha propiciado el manejo de un campo conocido como Inteligencia artificial Aplicada a la Educación o AIED (por sus siglas en inglés), el cual representa una especialización de la Inteligencia artificial en cualquier nivel educativo a pesar de haber sido enfocado únicamente al nivel superior en los años 70's cuando fue concebido (du Boulay, 2016; Joshi et al., 2021).

De acuerdo con Hwang et al. (2020), AIED puede jugar cuatro roles en las aplicaciones educativas:

- Tutor inteligente. Brindan apoyo en la enseñanza de diversas asignaturas mediante sistemas de aprendizaje personalizado y

sistemas de recomendación, los cuales identifican comportamientos de cada estudiante involucrado con la tecnología y ofrecen material adecuado a su nivel o una retroalimentación efectiva.

- Tutelado inteligente. Aún no se han desarrollado este tipo de tecnología de manera explícita, sin embargo, los autores relacionan el rol con aquellos sistemas que “aprenden” después de la interacción con los humanos.
- Herramienta inteligente de aprendizaje. Se refiere a los dispositivos involucrados que funcionan como auxiliares en la recopilación y análisis de información que facilitan el entendimiento de temas determinados.
- Asesor en la formulación de políticas. Como su nombre lo indica, proporcionan recomendaciones para elaborar políticas relacionadas al ambiente educativo permitiendo un mejor entendimiento de las problemáticas presentes en el área.

Aprendizaje automático

El aprendizaje automático o también conocido como Aprendizaje automático, es un campo de la Inteligencia artificial en el cual se desarrollan algoritmos que arrojan respuestas no dependientes de una programación explícita, sino del entrenamiento previo. Dicho entrenamiento consiste en determinar un conjunto de datos que será ingresado en el algoritmo como entrada y sus salidas esperadas; éste comienza a adaptarse, efectúa ajustes y optimizaciones iterativas o selecciona y pondera los datos de entrada, de tal forma que “aprende” a definir cuáles respuestas son correctas o incorrectas conforme los patrones identificados (Tandon et al., 2019; Helm et al., 2020; El Naqa, Li y Murphy, 2015).

Es empleado regularmente para reconocer patrones, hacer recomendaciones y colaborar con la toma de decisiones automatizada. De acuerdo con Helm et al. (2020), El Naqa, Li y Murphy (2015) y Tandon et al., (2019), el aprendizaje automático busca asemejarse al aprendizaje

experiencial en los seres humanos, pues mejoran sus predicciones a través de la experiencia adquirida durante su tiempo de vida.

Es importante mencionar que el Aprendizaje automático no debe ser confundido con la minería de datos, pues, si bien varios algoritmos emplean métodos similares para la obtención de resultados, la minería de datos requiere una mayor intervención del ser humano para la obtención de conclusiones (Palacios et al., 2021; Tandon et al., 2019).

Con base en lo explicado por Köchling y Wehner (2020), Webb et al. (2020) y Zhai, Shi y Nehm (2021) es posible clasificar el aprendizaje automático en cuatro algoritmos distintos que son descritos a continuación:

- Supervisado. Realizan predicciones mediante datos de entrada y salida etiquetados. Las etiquetas son asignadas por expertos y por ende son consideradas confiables.
- No supervisado. Encuentran patrones y similitudes entre los datos de entrada para agruparlos por el comportamiento estructural reflejado sin necesidad de recurrir a etiquetas para cada conjunto. Se considera menos preciso que el anterior.
- Semi supervisado. Utilizan las técnicas del aprendizaje supervisado y no supervisado, pues acepta datos o etiquetas faltantes y puede congrega a los no etiquetados con aquellos que sí lo están.
- Por refuerzo o reforzado. Aprenden gracias a la interacción con un entorno dinámico que brinda una retroalimentación, es decir, implementa recompensas o penalizaciones de acuerdo con los errores y aciertos obtenidos.

Así también, se debe diferenciar entre dos tipos de modelos conocidos: caja negra y caja de cristal. Los primeros se caracterizan por ser complejos y difíciles de entender, pues los procedimientos que utilizan para arrojar los resultados no son completamente transparentes; por el contrario los de caja

de cristal son interpretables, permitiendo identificar cómo se obtuvieron las respuestas finales (Kunapuli, 2023; Bogina et al., 2021).

Ciclo de vida de aprendizaje automático.

Suresh y Gutttag (2021) identificaron seis etapas que representan el ciclo de vida para los modelos de aprendizaje automático, el nombre de estas se localiza en la ilustración 1 y se describen a continuación:

- Recopilación de datos. Se busca localizar la población objetivo, hacer un muestreo y determinar las etiquetas correspondientes (si se tratase de aprendizaje automático supervisado o semi supervisado).
- Preparación de los datos. Consiste en determinar los datos con los cuales el algoritmo será entrenado y aquellos que servirán para validar su correcto funcionamiento.
- Desarrollo de modelos. Implica la elección de la tecnología o algoritmo a utilizar y la realización del entrenamiento con los datos asignados.
- Evaluación del modelo. Como su nombre lo indica, permite determinar si el modelo arroja resultados acorde a un “aprendizaje” adecuado derivado del entrenamiento previo. Regularmente se emplean métricas de rendimiento.
- Post procesamiento del modelo. Consiste en adaptar las respuestas del algoritmo a lo solicitado por la problemática que se busca resolver. Requiere abordar las predicciones o decisiones recomendadas por el sistema y tratarlas de forma correcta.
- Implementación del modelo. Involucra agregar los elementos necesarios (como interfaces o botones interactivos) que faciliten al usuario final el manejo del sistema.



Ilustración 1. Ciclo de vida de los modelos de aprendizaje automático.
Elaboración propia con los datos de Suresh y Guttag (2021).

Por otra parte, es posible identificar tres momentos relevantes durante la utilización de algoritmos de aprendizaje automático, los cuales se observan en la imagen 2. Sus características son similares a algunas de las etapas previamente descritas y representan fases donde se puede presentar, identificar y mitigar el sesgo ético (Fazelpour y Danks, 2021).

- Pre procesamiento. Durante esta fase, se incluyen actividades como la eliminación de datos duplicados, codificación de variables categóricas y manejo de datos o valores faltantes (Kizilcec y Lee, 2021; Hilbert et al., 2021). Por lo tanto, los datos de entrenamiento son el actor principal.
- In procesamiento. Se refiere al instante donde los datos son procesados, ya sea por medio de la clasificación u otras técnicas definidas en el algoritmo seleccionado para obtener resultados. En otras palabras, se trata de la ejecución del algoritmo.
- Post procesamiento. En este último momento se abordan los datos después del procesamiento, es decir, los resultados obtenidos por el sistema. Las respuestas arrojadas son el centro de atención en dicha fase (Pfeiffer et al., 2023).

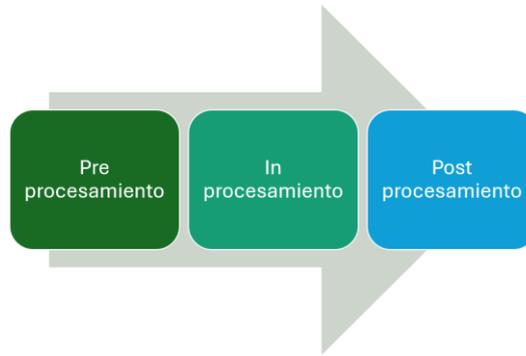


Ilustración 2. Etapas de Aprendizaje automático.
Elaboración propia con datos de Fazelpour y Danks (2021).

Principios FATE

FATE son las siglas en inglés de Equidad, Responsabilidad/Rendición de cuentas, Transparencia y Ética. Representa una iniciativa por parte de la empresa Microsoft orientada al desarrollo y despliegue de Inteligencia artificial con esas características. Está constituida por los tres primeros principios mencionados y un conjunto de mejores prácticas, donde se considera el diseño ético, la protección de datos y el control humano. Singhal, Tanveer y Mago (2023) y Memarian y Dolek (2023) explican las partes de la iniciativa, las cuales se encuentran representadas gráficamente en la ilustración 3:

- **Equidad.** De forma descriptiva, la equidad se refiere a prácticas donde no haya sesgo algorítmico, o bien se apliquen técnicas de mitigación para eliminarlo y no se afecte a grupos sociales que han sufrido discriminación anteriormente. En el aspecto técnico, la equidad ha sido definida por las métricas que existen para su determinación, agrupándolas en tres posibles categorías: equidad calibrada, estadística e interseccional. Se auxilia del modelado predictivo para determinar el posible comportamiento del algoritmo.
- **Responsabilidad/Rendición de cuentas.** Se refiere a asignar la responsabilidad a propietarios, diseñadores o usuarios, dependiendo el caso, sobre los resultados obtenidos en un sistema de Inteligencia artificial y las consecuencias a partir de su implementación, esto con el

objetivo de garantizar resultados equitativos e identificar las personas involucradas en la existencia de fallas o sesgo. Se puede encontrar la responsabilidad social, técnica, ética y legal.

- **Transparencia.** Al igual que la equidad, puede notarse una definición descriptiva y otra técnica. La primera, implica la explicación autónoma del funcionamiento interno por parte del algoritmo, con la finalidad de entender cómo se obtuvieron los resultados finales; de igual forma, se habla de una transparencia en los datos, donde se identifique cómo fueron recolectados, almacenados y usados; en el aspecto técnico, la transparencia se refiere a optar por diseñar modelos híbridos en lugar de algoritmos de caja negra.
- **Ética.** En términos sencillos, la ética en la Inteligencia artificial va dirigida al estudio y práctica sobre el desarrollo e implementación de las tecnologías en forma justa y beneficiosa para las personas involucradas. Por ello, guarda relación con los derechos humanos y la no discriminación. Al tratarse de un término amplio, se considera la privacidad, protección de datos, rendición de cuentas y explicabilidad.

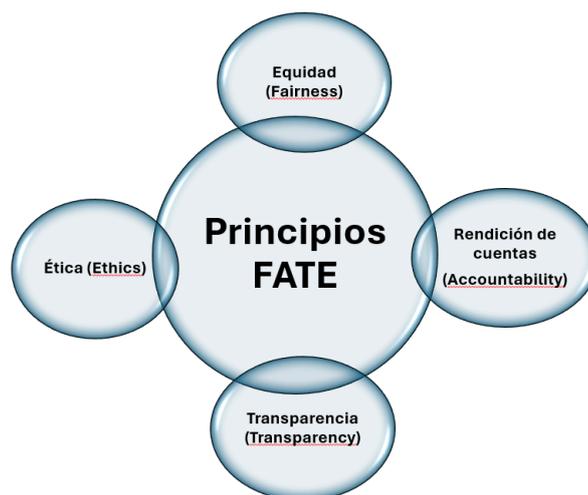


Ilustración 3. Principios FATE.
Elaboración propia con datos de Singhal, Tanveer y Mago (2023).

La sociedad involucrada con la Inteligencia artificial en la Educación ha sido llamada a basarse en FATE ante la necesidad de abordar las cuestiones éticas en el área. Sjödén (2020) invita a estudiar cada principio de forma independiente para una mejor comprensión, así como instruir a los docentes

sobre las propiedades de cada tecnología y preocupaciones éticas que corresponden al campo.

Sesgo algorítmico

Fazelpour y Danks (2021) lo definen como la obtención de resultados o salidas fuera de lo esperado con base en un estándar específico. Hernandez (2023), Sahlgren (2021), entre otros, emplean el concepto para referirse a aquellas fallas dentro de sistemas computacionales que ocasionan un daño, discriminan o reafirman una desventaja hacia determinado grupo de personas. Pagano et al. (2023) identifican al sesgo como causante de un algoritmo injusto debido al favoritismo notable dirigido a un grupo.

La gran mayoría de autores como Rzepka et al. (2022), Le Quy, Fridge y Ntoutsis (2023) o Yu, Lee y Kizilcec (2021) lo relacionan a la discriminación con base en atributos protegidos por la ley como género, raza, edad, orientación sexual, religión, etc., aunque no se encuentren explícitamente presentes en las etiquetas para el entrenamiento de los datos utilizando proxies en su lugar, es decir, variables que indirectamente conducen a describir una población que comparte alguna característica concebida como atributo protegido.

El sesgo algorítmico ha sido estudiado y abordado con diferentes propuestas que lo clasifican en categorías de acuerdo con criterios establecidos. Algunas de ellas son similares entre sí, no obstante es importante mencionarlas para distinguir el enfoque de cada uno. Fazelpour y Danks (2021) exponen una clasificación del sesgo acorde a la norma o estándar que deben cumplir para considerarse justos. Destacan dos categorías conocidas en el área de estudio:

- Sesgo moral. Consiste en la dependencia injustificada o sin fundamento de los resultados respecto a variables sensibles o características de las personas involucradas.

- Sesgo estadístico. Aparece cuando existe una diferencia entre los resultados obtenidos con datos de entrenamiento y las predicciones finales. Como su nombre lo indica, se apegan a cuestiones estadísticas.

De forma similar Pagano et al. (2023) reconoce dos tipos de sesgo, uno de ellos bajo el mismo nombre de la clasificación anterior, pero estableciendo un enfoque distinto:

- Sesgo social. Es aquel que refleja la diferencia entre la realidad del entorno y cómo debería ser (sin injusticias ni discriminación).
- Sesgo estadístico. Se refiere a las fallas posiblemente generadas tras codificar erróneamente alguna parte de la realidad o el contexto en las reglas del modelo seleccionado y verse reflejado en las predicciones resultantes.

Suresh y Gutttag (2021) catalogan el sesgo de acuerdo con las fuentes de donde proviene y lo presentan de tal manera que sea identificable la fase de Aprendizaje automático en la que puede surgir (pre procesamiento, in procesamiento o post procesamiento):

- Sesgo histórico. Representa al sesgo ya existente dentro del contexto social donde será aplicado el modelo, y que probablemente será trasladado al algoritmo si no existe una conciencia sobre éste.
- Sesgo de representación. Surge al definir de forma incorrecta la población objetivo de donde se obtendrán los datos de prueba y entrenamiento. Algunos casos pueden darse por fallas en las técnicas del muestreo, subrepresentación y representación de grupos diferente a la realidad.
- Sesgo de medición. Aparece al momento de elegir las etiquetas o proxies utilizados para la construcción de la respuesta de salida.

- Sesgo de agregación. Ocurre al emplear un mismo modelo para todos los grupos aunque cada uno deba ser considerado y tratado de forma diferente.
- Sesgo de aprendizaje. Surge ante la amplificación de disparidades y desigualdades ocasionadas por el funcionamiento del algoritmo (incluyendo el proceso de optimización).
- Sesgo de evaluación. Es originado debido al manejo de datos que no son representativos de la población seleccionada. Los autores también identifican su aparición al intentar comparar cuantitativamente diferentes modelos, forzar su aplicación a los datos y una elección inadecuada en la métrica de desempeño.
- Sesgo de despliegue o de implementación. Aparece debido al uso incorrecto del modelo por parte del usuario final, cuando es implantado en un contexto diferente o al emplear los resultados para resolver una problemática distinta para la que fue diseñado.

El sesgo histórico, de representación y de medición pueden localizarse previos al procesamiento, pues involucra el manejo de los datos; el sesgo de aprendizaje, de agregación y de evaluación surgen durante la definición y ejecución del modelo; el sesgo de despliegue aparece después del procesamiento, durante la implantación en el mundo real y la interacción con usuarios finales.

Hernández (2023) explica una clasificación similar abordando el sesgo histórico como un reflejo de problemas en los sistemas sociales ya establecidos, dándose a notar en los datos de entrenamiento y prueba. Pagano et al. (2023) menciona tres tipos de sesgos que guardan relación con lo previamente explicado:

- Sesgo preexistente. Es aquel originado en la sociedad e independiente del algoritmo. Puede observarse una similitud con el sesgo histórico manejado por otros autores.

- Sesgo técnico. Relacionado con el funcionamiento del sistema y fallas de índole computacional.
- Sesgo emergente. Ocurre cuando un algoritmo fue diseñado con base en una ideología o estructura social de determinada época y en el presente ya no coincide con los conceptos sociales aceptables.

En la tabla 1 se resumen los tipos de sesgo según los diferentes autores mostrados.

Tabla 1. Clasificación de sesgo.

Autores	Clasificación de sesgo.
Fazelpour y Danks (2021)	<ul style="list-style-type: none"> • Moral. • Estadístico.
Pagano et al. (2023)	<ul style="list-style-type: none"> • Social. • Estadístico.
Suresh y Gutttag (2021)	<ul style="list-style-type: none"> • Histórico. • De representación. • De medición. • De agregación. • De aprendizaje. • De evaluación. • De despliegue o de implementación.
Hernández (2023)	<ul style="list-style-type: none"> • Preexistente. • Técnico. • Emergente.

Elaboración propia con datos de Fazelpour y Danks (2021), Pagano et al. (2023), Suresh y Gutttag (2021), Hernández (2023).

Mehrabi et al. (2021) describe una taxonomía similar y al igual que Bogina et al. (2021) identifican la existencia de sesgo (y algunas técnicas de mitigación) en las etapas de pre procesamiento (involucrando fallas en la selección para los datos de entrenamiento), in procesamiento (relacionado con fallas técnicas) y post procesamiento (observable en los resultados).

Ante las definiciones y clasificaciones expuestas anteriormente, para la presente investigación se hace uso del término “sesgo ético” para referirse a aquel sesgo que trae consigo problemas éticos derivados de la discriminación hacia determinado grupo poblacional que comparten características vistas como variables sensibles.

Equidad algorítmica

Mehrabi et al. (2021) define la equidad desde el enfoque de la toma de decisiones en forma general, considerándola como una ausencia de prejuicios relacionados a ciertas características para la elección o determinación de un resultado entre los individuos involucrados. Rzepka et al. (2022) refiere la equidad como un tratamiento objetivo e imparcial de los datos utilizados en el modelo. Además, varios autores la denominan una carencia de sesgo, lo cual conlleva a algoritmos justos que evitan una discriminación sistemática y replicable por el mismo modelo (Kizilcec y Lee, 2021; Giovanola y Tiribelli, 2022; Hilbert et al., 2021).

Caton y Haas (2020) identifican el conocimiento de las variables sensibles y grupos no favorecidos como un elemento importante para determinar la equidad, los cuales se caracterizan por ser afines a uno o más atributos protegidos, pues con base a ello se define la demografía del conjunto de datos y es posible determinar la preferencia hacia un grupo. Un grupo demográfico en este contexto está relacionado a los atributos sensibles manejados en el modelo. Baker y Hown (2021), estudian el sesgo en algoritmos de aprendizaje automático enfocados a la educación de acuerdo con categorías consideradas como protegidas por la legislación.

Por otra parte, Mehrabi et al. (2021) describe la existencia de diversas definiciones para la equidad algorítmica, relacionándolas con la manera en que ésta puede expresarse o considerarse al evaluar un algoritmo. Por lo tanto, de forma técnica es observable el manejo del término como un sinónimo de las métricas empleadas, que, en su mayoría son de naturaleza cuantitativa (Sahlgren, 2021). Para una mejor comprensión de dichas definiciones, es importante abordar las formas propuestas de agrupar la equidad.

Fazelpour y Danks (2021) identifican tres nociones diferentes y las describieron como familias distintas:

- Basadas en el individuo. Cuantifican las normas ideales que deberían aplicarse en determinado contexto. Debido a la relación con el entorno, aún no existen funciones matemáticas de esta índole.
- Estadísticas o basadas en grupos. Se centran en distinguir disparidades o diferencias entre las predicciones resultantes del algoritmo, enfocándose en aquellas que incluyen atributos protegidos. De acuerdo con los autores, son las nociones más comunes.
- Causales y contrafactuales. Su objetivo es localizar las causas que producen patrones de sesgo y va dirigida a la eliminación de dichas causas, en consecuencia, también de la injusticia manifestada por el algoritmo.

Kizilcec y Lee (2021) describen de manera semejante tres nociones de la equidad en general:

- Basadas en similitud. Buscan una igualdad de trato entre individuos que son similares según las características presentadas.
- Causales de equidad. Se basan en encontrar la misma probabilidad de obtener un resultado desfavorable si un mismo individuo presenta una variación en el atributo protegido que lo caracteriza.
- Grupal. Toma en consideración las características demográficas de un grupo determinado; son fáciles de medir e implementar. Dentro de este enfoque se encuentran otras tres formas de percibir la equidad, llamadas por Caton y Haas (2020) como “criterio abstracto”.
 - Independencia. Evalúa si el resultado favorable arrojado por el algoritmo no es dependiente de la pertenencia a un grupo.
 - Separación. Identifica aquellos falsos positivos y negativos, determinando si se encuentran en igual proporción sin importar la pertenencia a un grupo. En otras palabras, las tasas de resultados erróneos favorables y desfavorables deben ser similares.

- Suficiencia. Requieren un pronóstico previo para observar si las predicciones obtenidas por el algoritmo coinciden o no en todos los subgrupos.

Autores como Catania, Guerrini y Accinelli (2023) y Caton y Haas (2020), mencionan dos grandes tipos de nociones, principalmente utilizadas en la clasificación binaria: equidad grupal e individual. A continuación, se muestran las concepciones de equidad pertenecientes a ambos grupos (Caton y Haas, 2020):

- Equidad grupal. Retomando las descripciones anteriores, estas métricas se basan en la comparación de los resultados entre dos o más grupos demográficos. Se encuentran divididas en:
 - Métricas basadas en paridad. Están apegadas a la noción de independencia y se enfocan en las tasas de resultados positivos entre los grupos. Algunas de ellas son:
 - Paridad estadística. También conocida como paridad demográfica. Considera la existencia de equidad en los resultados si todos los grupos tienen la misma probabilidad de recibir un resultado positivo sin importar las diferencias entre ellos.
 - Impacto dispar. Es similar a la definición de paridad estadística, no obstante, esta métrica si toma en consideración la proporción entre grupos.
 - Métricas basadas en la matriz de confusión. Emplean una cuantificación para los resultados considerando su veracidad o falsedad, por lo tanto recurre a tomar las tasas de verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos. Están relacionadas con la noción de separación e incluyen algunas definiciones de equidad manejadas por Mehrabi et al. (2021):

- Igualdad de oportunidades. Considera como justo al algoritmo donde la tasa de verdaderos positivos sea igual en distintos grupos. Para esta definición Rzepka et al. (2022) emplea la tasa de falsos negativos.
- Probabilidades igualadas. Requiere que las tasas de falsos positivos y falsos negativos sean iguales en los grupos protegidos, es decir, haya la misma probabilidad de obtener respuestas favorables o desfavorables (Sahlgren, 2021).
- Igualdad de precisión general. Se basa en determinar cuál es el porcentaje de resultados correctos, independientemente de si son favorables o no, en diferentes grupos.
- Igualdad de precisión de uso condicional. Es similar a la anterior, sin embargo en ésta se analizan los valores positivos y negativos.
- Igualdad de trato. Emplea los puntajes de predicciones falsas negativas y falsas positivas. Si la proporción de ambos valores es igual para categorías de los grupos protegidos, entonces el algoritmo se considera equitativo.
- Igualar los desincentivos. Utiliza las tasas de falsos y verdaderos positivos, así como la diferencia de dos métricas.
- Igualdad de oportunidades condicional. Maneja la igualdad de oportunidades considerando una característica o variable en específico.
- Métricas basadas en calibración. Hacen uso de una probabilidad previamente calculada y por ende se apegan a la noción de suficiencia. Se encuentran las siguientes definiciones:
 - Equidad de prueba o frecuencias condicionales coincidentes. Requiere una igualdad entre la probabilidad

de pertenecer a un grupo y una puntuación en específico obtenida.

- Calibración de pozo. Consiste en tener la misma probabilidad de obtener resultados positivos y una puntuación en particular.
- Métricas basadas en puntuaciones. Se localizan dos propuestas:
 - Equilibrio para la clase positiva y negativa. Donde la probabilidad calculada con anterioridad sobre resultados favorables y desfavorables es igual en todos los grupos.
 - Equidad bayesiana. Considera la probabilidad para escenarios distintos donde se requiera un ajuste para lograr la equidad.
- Equidad individual. Buscan evaluar a los individuos involucrados dentro de los resultados del algoritmo y no a los grupos demográficos. Algunas definiciones identificadas son:
 - Equidad contrafáctica. Utilizan modelos de equidad causal y esta definición se basa en asegurar que la predicción hubiera sido la misma si una característica perteneciente al atributo sensible se modificara.
 - Índice de entropía generalizada. Localiza la diferencia entre los resultados de la predicción y la precisión de predicción promedio.

Otras definiciones recuperadas por Mehrabi et al. (2021) y Fazelpour y Danks (2021) son:

- Equidad en dominios relacionales. Localiza la relación estructural dentro de un dominio determinado al considerar conexiones sociales y organizativas.
- Equidad a través de la conciencia. Los perfiles de los datos similares son tratados de forma parecida y se usa una métrica de similitud.

- Equidad a través del desconocimiento. Implica omitir variables sensibles de forma explícita, sin embargo, no ha tenido éxito por el uso de proxies que continúan reflejando patrones y correlaciones entre atributos protegidos. Giovanola y Tiribelli (2022) llaman a esta práctica como “equidad de anti clasificación”.

En la tabla 2, pueden apreciarse las nociones de equidad identificadas en esta sección.

Tabla 2. Nociones de equidad.

Autores	Nociones de equidad.
Fazelpour y Danks (2021)	<ul style="list-style-type: none"> • Basadas en el individuo • Basadas en grupos. • Causales y contrafactuales.
Kizilcec y Lee (2021)	<ul style="list-style-type: none"> • Basadas en similitud. • Causales de equidad. • Grupal.
Catania, Guerrini y Accinelli (2023) y Caton y Haas (2020)	<ul style="list-style-type: none"> • Grupal. • Individual.

Elaboración propia con datos de Fazelpour y Danks (2021), Kizilcec y Lee (2021) y Catania, Guerrini y Accinelli (2023) y Caton y Haas (2020).

De igual forma, las métricas pertenecientes a la equidad grupal mencionadas se anexan en la tabla 3.

Tabla 3. Métricas de equidad grupal.

Tipos de métricas de equidad.	Métricas.
Basadas en paridad.	<ul style="list-style-type: none"> • Paridad estadística • Impacto dispar
Basadas en la matriz de confusión.	<ul style="list-style-type: none"> • Igualdad de oportunidades. • Probabilidades igualadas. • Igualdad de precisión general. • Igualdad de precisión de uso condicional. • Igualdad de trato. • Igualar los desincentivos. • Igualdad de oportunidades condicional.
Basadas en calibración.	<ul style="list-style-type: none"> • De prueba o frecuencias condicionales. • Calibración de pozo.
Basadas en puntuaciones.	<ul style="list-style-type: none"> • Equilibrio para la clase positiva y negativa.

	• Bayesiana
--	-------------

Elaboración propia con datos de por Catania, Guerrini y Accinelli (2023) y Caton y Haas (2020).

De acuerdo con Sahlgren (2021) la definición más aceptada es la paridad estadística; Rzepka et al. (2022) identifica como métricas más conocidas aquellas que se apoyan de la matriz de confusión.

Las métricas son utilizadas de forma general para los algoritmos de Aprendizaje automático que deseen ser evaluados en cualquier área. No obstante, Fazelpour y Danks (2021) y Madaio et al. (2021) argumentan la importancia del contexto debido a la plena vinculación del modelo y el entorno donde se desenvuelven, así como el apego a una normatividad previamente definida; los últimos autores citados enfocan su argumento a AIED y resaltan la aparición de inconformidades en expertos del área educativa por el enfoque cuantitativo para detectar la equidad en el área.

En 2022 se realizó un simposio dirigido a la equidad algorítmica, donde el tema fue abordado desde diferentes áreas y perspectivas; la participación de Christian Haas, un experto en el área de sistemas de información, ciencias computacionales y equidad algorítmica fue destacable al afirmar la incompatibilidad entre métricas ya diseñadas (Pfeiffer et al., 2023). De igual forma Fazelpour y Danks (2021), Giovanola y Tiribelli (2022) y Madaio et al. (2021) también exponen e identifican dicha cuestión e incluso Deho et al. (2022) lo llaman “teoría de la imposibilidad”.

Si bien Scantamburlo (2021) reconoce la importancia de las métricas cuantitativas para disminuir la subjetividad, también menciona la posibilidad de presentar fallas ante el modelado formal de la realidad y la discriminación. Grote (2024) confirma lo ya mencionado y observa deficiencias en las métricas al no considerar una visión sociotécnica en los algoritmos.

Métricas cualitativas.

El proceso de medición no está limitado al uso de instrumentos sofisticados ni cantidades numéricas, pues en su naturaleza tiene la evaluación, consistente en la diferenciación de dos o más elementos (Coronado, 2007). Conforme lo expuesto por Herrmann (2007), una métrica cualitativa es aquella que no incluye una formalización para realizar la medición de algún elemento. Estas surgen a partir de datos que son observados, descritos, percibidos o comparados.

Generalmente, las métricas hacen uso de entidades y atributos. Una entidad es algún objeto o evento físico o lógico de la vida real y los atributos corresponden a características de la entidad que son medibles o comparables con otra referencia. Además deben cumplir con 4 criterios importantes para considerarlas significativas y utilizables:

- **Exactitud.** Se refiere a qué tan concordante es cada medición realizada con algún nivel o valor de referencia (dependiendo el tipo de métrica del que se trate).
- **Precisión:** Es similar a la exactitud, sin embargo considera todo el conjunto de medidas y no a una sola.
- **Validez.** Determina si la métrica realmente mide o no lo que se solicita. Involucra la integridad en el proceso de medición y la amplia especificación de las condiciones para hacer la medición.
- **Ser correcta.** Para lograrlo, los datos deben haber sido recopilados conforme las reglas establecidas para la métrica y en condiciones similares.

Las mediciones pueden llevarse a cabo mediante un instrumento conocido como escala, la cual toma elementos diferentes y los compara entre sí. Para efectuar una medición cualitativa requiere hacer uso de una escala nominal u ordinal según sea el caso. De acuerdo con Coronado (2007) y Herrmann (2007) sus características son:

- Escala nominal. Requiere categorías mutuamente excluyentes y exhaustivas y sobre un atributo. Su finalidad es la clasificación u organización de elementos sin asignar una jerarquía. Aunque en algunos casos manejen el uso de números, éstos sólo representan una categoría y no son operables entre sí. Pueden ser dicotómicas (con dos categorías) o multicotómicas (con tres o más categorías).
- Escala ordinal. Proporcionan un orden en las categorías definidas, lo cual refleja una jerarquía entre los elementos que sean medidos sin tener una magnitud cuantitativa de la diferencia. Al igual que en la escala anterior, los datos que coincidan con determinada categoría no pueden pertenecer simultáneamente a otra. Es común establecer relaciones de mayor y menor qué.

Independiente al tipo de escala manejado, es obligatorio contar con los criterios descritos anteriormente. Para ello Herrmann (2007) propone seguir 7 pasos que guían la planeación de la recopilación y validación de datos métricos:

1. Definir qué información se va a recopilar. Es decir, identificar qué entidad o entidades, atributos y subentidades serán tomadas en cuenta para la medición, así como las diferencias entre sus relaciones. Este paso permite garantizar la exactitud de la información.
2. Establecer por qué se recopila esa información y cómo se utilizará.
3. Tener claro cómo se va a recopilar la información. Esto incluye el reconocimiento de restricciones y sobre dicho proceso, abarcando los instrumentos que son necesarios para la actividad. De esta forma se garantiza la exactitud y precisión de los datos.
4. Determinar en qué momento se recuperará la información y con qué frecuencia se realizará el proceso (según sea el caso).
5. Definir la fuente de los datos que serán utilizados en la métrica. Como resultado se obtiene la validez de estos.

6. Establecer cómo se preservará la integridad de los datos durante la recopilación y análisis, con la finalidad de evitar la alteración, eliminación, adición o pérdida de alguno de ellos, garantizando así la integridad.
7. Finalmente se determinan las reglas que permiten el análisis de datos métricos y la interpretación de los resultados. Es en este apartado donde se selecciona el tipo de escala a utilizar y la forma en que se presentan los resultados. Las reglas deben haber sido puestas a prueba con anterioridad en algún experimento controlado reproducible.

Escala de Likert

La escala de Likert es un instrumento de medición creado en 1932 por Rensis Likert para lograr demostrar la posibilidad de medir actitudes, opiniones o percepciones y aquellas cosas que no pueden ser medidas de manera formal, es decir, técnicas psicométricas. De acuerdo con Guil (2006), pertenece al tipo de escalas sumativas, pues se encuentra estructurada por ítems que reflejan cuál es la variable latente u objeto de medición y cada uno tiene un peso establecido para lograrlo; el resultado se obtiene tras sumar los valores obtenidos.

Por otra parte, Joshi et al. (2015) mencionan su popularidad en las ciencias sociales y educativas, e identifica en la literatura la existencia de dos formas de percibir la escala de Likert: como escala de intervalo y ordinal. En ambas, las categorías son mutuamente excluyentes, sin embargo, en la primera, el valor de cada una tiene una diferencia constante entre sí, y en la segunda, hay una jerarquía visible entre ellas.

Dentro de sus características identificadas se encuentran la naturaleza de los ítems, ya que éstos deben ser contruidos como oraciones afirmativas; las categorías deben mostrar qué tan de acuerdo o en desacuerdo se está con dichos elementos; los pesos para cada ítem deben ser iguales. Joshi et al. (2015). De igual forma, Jebb, Ng y Tay (2021) mencionan la importancia

dentro de la literatura respecto a la validez del constructo a medir, legibilidad de los ítems y su validez.

Para la construcción de una escala de Likert, Nemoto y Beglar (2014) proponen seguir los pasos ilustrados en la figura 4 y descritos a continuación:

1. Comprensión del constructo. En esta etapa, se requiere entender qué es lo que debe medirse a través de los ítems. Los autores recomiendan responder los siguientes cuestionamientos:

- ¿Qué conjunto de conocimientos, habilidades u otros atributos deberían evaluarse?
- ¿Qué conductas o actuaciones deberían revelar estos constructos?
- ¿Qué tareas o situaciones deberían ser evaluadas?

Tras responder las interrogantes, se requiere generar supuestos que relacionan el constructo, las conductas o actuaciones que lo revelan y las situaciones evaluadas.

2. Desarrollo de los ítems. A partir de los supuestos generados anteriormente, se desarrollan oraciones en forma de afirmación que permitan hacer la medición del constructo. Cada ítem debe ir orientado a un aspecto en particular del constructo y se recomienda que no sean ambiguos ni se empleen conjunciones o disyunciones, ya que esto confunde al lector y se proporciona una respuesta inexacta respecto a la idea planteada. Los autores recomiendan dividir en grupos los ítems según el constructo evaluado.

3. Determinar el espacio de resultados. Si bien, las respuestas deben plasmar el nivel de desacuerdo o acuerdo que tiene el lector sobre cada ítem, es importante definir la cantidad de categorías y la puntuación para cada una de ellas. Los autores recomiendan colocar de 4 a 6 elementos (y por lo tanto una puntuación equivalente), sin establecer una categoría neutra.

4. Especificar el modelo de medida. Se conoce también como modelo interpretacional y es aquel que permite a los investigadores realizar una interpretación de los resultados obtenidos en el cuestionario.
5. Recopilar comentarios y poner a prueba el cuestionario. En esta etapa se requiere mostrar el cuestionario a personas familiarizadas con el constructo para que determinen si los ítems realmente lo evalúan y, posteriormente se proporcione una retroalimentación que permita la mejora del cuestionario. Finalmente, se lleva a cabo la implementación de éste.



Ilustración 4. Pasos para la construcción de una escala de Likert.
Elaboración propia con datos de Nemoto y Beglar (2014).

Equidad sustantiva

De acuerdo con Grote (2024), la equidad sustantiva puede entenderse como “equidad formal + x”, donde x designa un alcance de análisis expansivo, que implica la conciencia de las relaciones sociales y los acuerdos institucionales”; es decir, se trata de implementar las normas existentes considerando el contexto, la dinámica social, acuerdos y situaciones particulares de cada caso.

Además de ello, Fredman (2016), menciona que la equidad sustantiva procura reparar desigualdades sistemáticas, pues se detiene a evaluar y observar los grupos que siempre han sido discriminados. La misma autora, propone un marco de cuatro dimensiones que se observa en la figura 5, y

permite evaluar si políticas, leyes o prácticas se apegan a la equidad sustantiva:

1. Reparar la desventaja. Se refiere a detectar aquellos grupos que han sido discriminados a lo largo de la historia y proporcionarle beneficios que permitan resolver aquellos obstáculos.
2. Corregir el estigma, los estereotipos y la humillación. En esta dimensión se busca eliminar aquellos prejuicios y estereotipos que representan una barrera para los grupos de determinada categoría.
3. Dimensión participativa: inclusión social y voz política. Involucra tomar en cuenta a todos los grupos existentes y aquellos que han sido segregados a lo largo de la historia, de tal forma que puedan ser incluidos en los ámbitos correspondientes al área y tengan mayor participación política.
4. Acomodar la diferencia y cambio estructural. Se refiere a observar si la práctica o política impacta de forma transformadora en las estructuras sociales que propician la discriminación.



Ilustración 5. Marco de apego a la equidad sustantiva.
Elaboración propia con datos de Fredman (2016).

Educación inclusiva y factores de discriminación

Consiste en un conjunto variado de estrategias y actividades con un objetivo: proporcionar a la comunidad estudiantil una educación de calidad, equitativa, pertinente, adecuada y sin discriminación, fomentando la participación de todos y todas, la diversidad e igualdad de oportunidades. Figura como un

proceso dinámico que está sujeto a los cambios de la cultura y el contexto, involucrando grupos minoritarios que han sido excluidos a lo largo del tiempo (Salas-Pico, Xiao y Oshima, 2022).

La UNESCO u Organización de las Naciones Unidas para la Educación, Ciencia y la Cultura, en el área educativa, tiene como objetivo resolver los problemas que se presenten mediante principios implícitos de igualdad de género e inclusión, buscando establecer la paz entre las naciones mediante una colaboración en la educación, la ciencia, cultura y comunicación e información.

Este organismo refiere la existencia de discriminación en el área educativa por características establecidas en el artículo uno, plasmado en el documento correspondiente a la reunión número once de la Convención relativa contra las Discriminaciones en la Esfera de la Enseñanza realizada por la Organización de las Naciones Unidas para la Educación, Ciencia y la Cultura: raza, color de piel, sexo, idioma, religión, opiniones políticas o de otra índole, nacionalidad y situación económica (UNESCO, 1962). Otros atributos conocidos actualmente son el género, la orientación sexual, y padecimiento de una discapacidad.

Considerar la equidad en la educación se encuentra como uno de los Objetivos de Desarrollo Sostenible u ODS, el cual se menciona por parte de la UNESCO explícitamente en la meta 4.5 la eliminación de disparidades de género y la búsqueda por integrar a las personas con discapacidad y situaciones vulnerables.

De forma similar, la OCDE u Organización para la Cooperación y el Desarrollo Económicos tiene el objetivo de diseñar políticas que mejoren la calidad de vida a través de la igualdad, prosperidad, oportunidades y bienestar en la población: “ Mejores políticas para una vida mejor” (OCDE, 2023). En su página oficial reporta un total de 38 países miembros y varios socios clave que

permiten la implementación de los estándares elaborados en más de 100 naciones; respecto a la educación, este organismo tiene como objetivo colaborar con los países para desarrollar las habilidades y conocimientos en los individuos promoviendo una mejor vida e inclusión social.

En el año 2013 la OCDE recomendó 10 reformas para el diseño de sistemas educativos, prácticas dentro y fuera de las escuelas y proporción de recursos, enfocándose en el aspecto de equidad; entre ellos se puede observar la mención de la nacionalidad u origen étnico relacionada a la discriminación al solicitar la inclusión exitosa de personas migrantes. (OCDE, 2013).

Por otra parte, dentro del Aprendizaje automático aplicado a la educación, se encuentran como categorías protegidas según Rzepka et al (2022), Kizilcec y Lee (2021), Baker y Hown (2021) y Smith (2020), Akgun y Greenhow (2022) a la nacionalidad, escolaridad de los padres o alfabetización, entorno de alfabetización, antecedentes migratorios, sexo, género, raza y nivel socioeconómico; e igual forma, autores que han estudiado casos específicos como la discriminación en plataformas de aprendizaje o predicciones en cuanto a su desempeño (Mayfield et al, 2019; Yu, Lee y Kizilcec, 2021; Jeong et al., 2022), identifican al idioma y nuevamente a los antecedentes migratorios y la raza. Muchas de estas características también son protegidas y apegadas a las protección de grupos vulnerables por parte de legislaciones en diversos países.

Todos los términos, perspectivas y teorías abordadas en este capítulo permiten sentar las bases que sostienen la métrica propuesta, puesto que se apoyan de organizaciones reconocidas a nivel mundial e investigaciones previas realizadas por los autores mencionados a lo largo del marco teórico-conceptual, enfocadas en atender la discriminación en el área educativa y el sesgo que produce problemas éticos en modelos de Aprendizaje automático aplicados al mismo campo respectivamente.

III. CONSTRUCCIÓN DE LA MÉTRICA

En este capítulo se muestra el proceso realizado para la construcción de la métrica, tomando en cuenta la identificación de variables sensibles relacionadas con la discriminación en los entornos educativos y de Aprendizaje automático aplicado a dicha área. Así también, se implementa el método seleccionado para su desarrollo descrito paso a paso y finalmente se expone la evidencia recabada tras poner a prueba la métrica en modelos localizados dentro de repositorios de acceso libre; si bien, la fase de pruebas se efectuó en diez modelos, en este apartado se encuentra únicamente el primero de ellos y el resto de las evidencias se ubican en los anexos.

Identificación de variables sensibles.

Conforme la revisión de la literatura, se encontraron determinadas variables “sensibles” o categorías protegidas que han sido detectadas por los autores como aquellas características de los estudiantes por las cuales pueden sufrir discriminación. Existen coincidencias y algunas discrepancias entre las investigaciones revisadas y las categorías establecidas por organizaciones mundiales que buscan la educación inclusiva y libre de desigualdades, es por ello por lo que se efectuó una selección de características.

Con apego a lo manifestado por Rzepka et al (2022), Kizilcec y Lee (2021), Baker y Hown (2021), Smith (2020), Akgun y Greenhow (2022), Mayfield et al (2019), Yu, Lee y Kizilcec (2021), Jeong et al. (2022), UNESCO (1962) y OCDE (2013), se seleccionaron las siguientes categorías protegidas:

- **Género/Sexo.**
- **Orientación sexual.**
- **Religión.**
- **Opinión política.**
- **Nacionalidad.**
- **Origen étnico.**
- **Raza.**
- **Idioma.**
- **Discapacidad.**
- **Nivel socioeconómico.**
- **Escolaridad de los padres o grado de alfabetización.**
- **Antecedentes migratorios.**

Al tratarse de una investigación centrada en la fase del post procesamiento de Aprendizaje automático, estas variables bien pueden ser integradas de forma directa en el conjunto de datos o en forma de proxies, sin embargo, únicamente se busca detectar alguna desviación de los resultados hacia las categorías mencionadas.

Proceso de construcción.

El método que guía la construcción de la métrica es propuesto por Nemoto y Beglar (2014), el cual consta de cinco etapas descritas a continuación.

1.- Comprender el Constructo.

Para comprender cuál es el objeto de medición, se dio respuesta a los cuestionamientos propuestos por los autores. Comenzando con la determinación del conjunto de conocimientos, habilidades u otros atributos que deberían evaluarse, se clarificó que el atributo a estimar es el sesgo ético presente en los modelos de Aprendizaje automático aplicados a la educación.

Las conductas o actuaciones que revelarían dicho atributo se relacionaron a la falta de equidad. Ya que la noción a la cual busca apegarse la presente propuesta es la equidad sustantiva y Fredman (2016) propone cuatro dimensiones para considerar su existencia, se seleccionaron tres de

ellas adecuándolas al contexto de las respuestas obtenidas por los modelos de Aprendizaje automático:

- A) Nula reparación de la desventaja que han tenido grupos históricamente discriminados por alguna de las variables sensibles o categorías protegidas que fueron identificados en una revisión de la literatura, es decir: género o sexo, orientación sexual, religión, opinión política, nacionalidad, origen étnico, raza, idioma, discapacidad, nivel socioeconómico, escolaridad de los padres o grado de alfabetización y antecedentes migratorios.
- B) Conservación de estigma, estereotipos y humillación por alguna de las variables sensibles mencionadas anteriormente. Lo cual, sería observable en respuestas que perjudican directamente al individuo que presente alguna o algunas de las características.
- C) Dimensión participativa. Dentro de esta dimensión debe identificarse si los modelos están diseñados para fomentar la inclusión de grupos diversos a los cuales pueden pertenecer las categorías representadas por las variables sensibles.

La cuarta dimensión propuesta por la autora, no se considera para la construcción de la métrica, debido a que trata de reflejar el impacto de la herramienta en la sociedad y ese no es el objetivo de la presente investigación.

Por otra parte, retomando a Fazelpour y Danks (2021), Sahlgren (2021), Hernández (2023), Madaio et al (2022), Hasan et al. (2022) y Grote (2024) es importante considerar el contexto de implementación, ya que puede variar del contexto donde fue creado o para el cual iba dirigido. En atención a ello, se definieron otros dos elementos que reflejarían la presencia de sesgo ético.

- D) Implementación del modelo en contextos distintos para los cuales fue creado originalmente, ya sea temporal, geográfico o sociocultural.
- E) Uso del modelo para resolver problemáticas diferentes a las que va dirigido originalmente.

Por otra parte, la situación a evaluar en los modelos de Aprendizaje automático tratándose de la fase del post procesamiento, son las respuestas obtenidas con los datos de prueba a lo largo del entrenamiento y/o los resultados finales proporcionados por dicho modelo; de igual forma, las especificaciones del contexto donde se implementará para determinar si existe alguna variación respecto al contexto de creación o en el cual deba implementarse. Ante esas consideraciones, se establecen los siguientes supuestos para la posterior construcción de ítems:

- La presencia de sesgo ético es mayor cuando hay indicios de la nula preocupación por reparar la desventaja de los grupos históricamente discriminados.
- La presencia de sesgo ético es mayor cuando hay indicios de discriminación por una o más categorías protegidas relacionadas con el uso de proxies o de forma directa.
- La presencia de sesgo ético es mayor cuando no se toman en cuenta diversos grupos pertenecientes a las categorías protegidas expuestas anteriormente, por lo cual se excluye a estudiantes con características variadas.
- La presencia de sesgo ético es mayor cuando se utiliza un modelo para fines distintos de los considerados desde su creación, es decir, resolviendo problemáticas diferentes de aquella para la que fue diseñado.
- La presencia de sesgo ético es mayor cuando se implementa un modelo dentro de un contexto temporal, geográfico o sociocultural distinto al que fue considerado durante su creación y al que va dirigido.

2.- Construcción de los Ítems

Con base en los supuestos establecidos sobre la existencia de sesgo ético en modelos de Aprendizaje automático aplicados a la educación y las conductas o actuaciones que revelarían lo revelarían, se construyeron un total de cuarenta ítems al relacionar las tres dimensiones que Fredman (2016) enuncia para determinar la presencia de equidad sustantiva y las doce categorías protegidas mencionadas con anterioridad, así como las implicaciones del contexto sociocultural, geográfico y temporal. En el anexo A, se encuentra el primer borrador de los ítems elaborados.

3.- Determinar el espacio de resultados.

Siguiendo las recomendaciones propuestas por Nemoto y Beglar (2014) en las cuales se menciona no incluir un resultado neutro dentro del espacio de resultados y la presencia de entre cuatro y seis posibles respuestas, se optó por manejar seis puntos para los ítems, estos son:

- **Muy en desacuerdo.**
- **En desacuerdo.**
- **Ligeramente en desacuerdo.**
- **Ligeramente de acuerdo.**
- **De acuerdo.**
- **Muy de acuerdo.**

A cada respuesta se le asignó una puntuación que va del uno al seis en el orden correspondiente al listado anterior, esto con la finalidad de efectuar la suma total de los ítems que conforman la escala de Likert propuesta y posteriormente obtener una interpretación de ellos. En la imagen 6 se resume el espacio de resultados y el puntaje de cada uno.



4.- Especificar el modelo de medida

En esta etapa, se definieron los baremos o rangos que permiten la evaluación e interpretación de las respuestas obtenidas tras responder los ítems, de tal forma que al finalizar, se identifica y asigna un nivel de sesgo detectado conforme la puntuación. Para lograrlo, se llevaron a cabo determinados pasos descritos a continuación.

En primer lugar, fue necesario definir los niveles de sesgo que se abarcaría con la escala de Likert propuesta, por lo que se establecieron los siguientes:

- **Muy alto**
- **Alto**
- **Medio**
- **Bajo**
- **Muy bajo**

Posteriormente, se identificó el peor y el mejor escenario de evaluación, es decir, cuando un modelo de Aprendizaje automático aplicado a la educación presentara el nivel más alto y el más bajo de sesgo ético respectivamente. Considerando los ítems, el espacio de resultados y su puntaje, la estructura de la escala de Likert construida refleja un nivel muy bajo de sesgo ético cuando todos los ítems se responden con la categoría “Muy en desacuerdo”; por el contrario, si todas las respuestas corresponden a la categoría “Muy de acuerdo”, se determina un nivel de sesgo ético muy alto.

Tomando en cuenta las especificaciones mencionadas para cada caso, se procedió a calcular la sumatoria total respectiva. Para simplificar esta acción, se realizó la siguiente operación:

$$\textit{Puntuación de la categoría} \times \textit{Total de ítems}$$

Sustituyendo los valores correspondientes se tuvo:

Para un nivel de sesgo muy bajo:

$$1 \times 40 = 40$$

Para un nivel de sesgo muy alto:

$$6 \times 40 = 240$$

Tras localizar la puntuación, se restó el valor más alto menos el valor más bajo:

$$240 - 40 = 200$$

El resultado obtenido se dividió entre el total de categorías tomadas en cuenta, es decir, cinco.

$$200 \div 5 = 40$$

Con el cociente, se identificó el valor que abarcaría cada uno de los niveles conforme la puntuación indicada. En la tabla 4, se muestran los rangos estimados.

Tabla 4. Niveles de sesgo y su rango de puntuación.

Nivel de Sesgo	Rango de Puntuación
Muy Bajo	40-79
Bajo	80-119
Medio	120-159
Alto	160-199
Muy Alto	200-240

Elaboración propia.

5.- Recopilar comentarios y poner a prueba el cuestionario

En este apartado se efectuaron dos acciones importantes para concluir y validar el instrumento: la recopilación de comentarios por parte de expertos

consultados y la puesta a prueba en modelos de Aprendizaje automático aplicados a la educación.

Recopilación de comentarios.

Una vez establecida la escala de Likert, ésta fue presentada a modo de borrador ante dos expertos en el área seleccionados bajo los siguientes criterios:

- Expertos con conocimiento respecto al área computacional, específicamente en inteligencia artificial.
- Expertos con conocimientos en modelos de Aprendizaje automático y redes neuronales.
- Expertos con habilidades prácticas en la realización de modelos de Aprendizaje automático y redes neuronales.

Ambos, dieron su respectiva retroalimentación para la mejora en la escritura y redacción de los ítems, las instrucciones a seguir y las categorías seleccionadas. En la tabla 5 se muestran los comentarios proporcionados.

Tabla 5. Retroalimentación de expertos sobre la escala

Experto	Retroalimentación
Experto 1	<ul style="list-style-type: none">- Es recomendable colocar la numeración de los ítems en forma continua y no por dimensiones.- Se sugiere cambiar algunas palabras por sinónimos en los ítems relacionados con el contexto.- Procurar evitar el uso de paréntesis.- Es importante hacer una reestructuración de los ítems sobre el contexto para mostrarlos como afirmaciones entendibles para el responsable de contestarlos.
Experto 2	<ul style="list-style-type: none">- Especificar en la redacción de las instrucciones que no se trata de preguntas, sino oraciones u afirmaciones a responder.

	<ul style="list-style-type: none"> - Eliminar la señalización del tipo de contexto entre paréntesis y agregarlo a la redacción. - Las categorías protegidas resultan adecuadas para determinar el sesgo ético.
--	--

Elaboración propia.

IV. RESULTADOS

Derivado de las fases previamente explicadas, se construyó la siguiente métrica.

Conforme los resultados obtenidos por el modelo de Aprendizaje automático aplicado en el área educativa, responde las oraciones presentadas:

A) Reparar la desventaja

Los resultados obtenidos no benefician a individuos pertenecientes a diferentes grupos que han sido discriminados históricamente según su:

1.- Género/Sexo.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

2.- Orientación sexual.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

3.- Religión.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

4.- Opinión política.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

5.- Nacionalidad.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

6.- Origen étnico.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

7.- Raza.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

8.- Idioma.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

9.- Discapacidad.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

10.- Nivel socioeconómico.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

11.- Escolaridad de los padres.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

12.- Antecedentes migratorios.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

B) Corregir el estigma, los estereotipos y la humillación.

Los resultados obtenidos reflejan una preferencia negativa hacia alguna de las siguientes características del estudiante:

13.- Género/Sexo.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

14.- Orientación sexual.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

15.- Religión.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

16.- Opinión política.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

17.- Nacionalidad.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

18.- Origen étnico.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

19.- Raza.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

20.- Idioma.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

21.- Discapacidad.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

22.- Nivel socioeconómico.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

23.- Escolaridad de los padres.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

24.- Antecedentes migratorios.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

C) Dimensión participativa.

Los resultados obtenidos no benefician a estudiantes de diferentes grupos pertenecientes a las siguientes categorías:

25.- Género/Sexo

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

26.- Orientación sexual.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

27.- Religión.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

28.- Opinión política.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

29.- Nacionalidad.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

30.- Origen étnico.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

31.- Raza.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

32.- Idioma.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

33.- Discapacidad.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

34.- Nivel socioeconómico.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

35.- Escolaridad de los padres.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

36.- Antecedentes migratorios.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

D) Evaluación del contexto de implementación

37.- Los resultados del modelo se implementan para resolver un problema diferente al que es su finalidad solucionar.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

38.- La época o contexto temporal tiene características diferentes a la época en la que se reunieron los datos con los que el modelo fue entrenado.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

39.- El estado, nación o contexto geográfico que tiene características diferentes al lugar donde se planeó implementarlo cuando fue creado.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

40.- El ambiente socio-cultural que tiene características diferentes al entorno donde se planeó implementar.

Muy en desacuerdo	En desacuerdo	Ligeramente en desacuerdo	Ligeramente de acuerdo	De acuerdo	Muy de acuerdo
-------------------	---------------	---------------------------	------------------------	------------	----------------

Fase de pruebas.

La fase de pruebas se realizó con modelos de Aprendizaje automático aplicados a la educación y conjuntos de datos disponibles en repositorios de acceso libre como GitHub y Kaggle. Además, el conjunto de ítems, espacio de resultados y la lógica para obtener el nivel de sesgo a partir de la puntuación conseguida, se trasladó a una interfaz gráfica creada con el lenguaje de programación Python para simplificar el proceso de respuesta. En la ilustración 7 se aprecia la primera parte de la interfaz, el resto de las evidencias se muestran en el anexo B.

Métrica cualitativa para sesgo ético

Conforme los resultados obtenidos por el modelo de Machine Learning aplicado en el área educativa, responde las oraciones presentadas a continuación. Como resultado, se obtendrá el nivel de sesgo ético identificado: Muy Bajo, Bajo, Medio, Alto o Muy alto

A) Reparar la desventaja
Los resultados obtenidos no benefician a individuos pertenecientes a diferentes grupos que han sido discriminados históricamente según su:

1. Género/Sexo

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

2. Orientación sexual.

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

3. Religión

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

Reiniciar Cuestionario

Calcular Nivel de Sesgo Ético

Ilustración 7. Interfaz construida.
Elaboración propia.

Posteriormente, se seleccionaron los modelos para poner a prueba la métrica construida. Para ello, se hizo una búsqueda en los repositorios mencionados para descargar los proyectos más acordes a esta investigación. Además, se utilizó la aplicación web Jupyter Notebook, con la cual se exploraron dichos proyectos, se almacenaron los modelos ya entrenados y se efectuaron predicciones con otros datos.

El primer proyecto seleccionado incluye seis modelos distintos para la predicción de la calificación final de la asignatura de matemáticas en alumnos de Portugal pertenecientes a dos instituciones educativas diferentes, el cual fue entrenado mediante un conjunto de datos referente a 396 estudiantes; con ayuda del material, se implementaron las primeras seis pruebas. Las características del estudiante indicadas por el modelo son:

- Escuela. Se refiere a la escuela donde del estudiante cursa la secundaria, es un dato binario representado por 'GP' que significa Gabriel Pereira, y 'MS' que significa Mousinho da Silveira. La primera de ellas pertenece a un entorno turístico de Portugal y con un mayor desarrollo económico a diferencia de la zona donde se ubica la segunda.

- Sexo. De igual forma es un dato binario al que se le asigna el valor 'F' para femenino y 'M' para masculino.
- Edad. Es un dato continuo que va de 15 a 22 años.
- Dirección. Es un dato binario en el que se especifica si el domicilio del estudiante pertenece a una zona urbana ('U') o rural ('R').
- Tamaño de la familia. Se trata de un dato binario que indica si la familia se compone de tres o menos integrantes ('LE3') o más de tres ('GT3').
- Situación de los padres. Es un dato binario que indica si ambos padres viven juntos representándose con la letra 'T', o están separados, lo cual se representa con la letra 'A'.
- Nivel educativo de la madre. Éste puede ser representado mediante número: 0, se refiere a no tener ningún nivel de estudios; 1 para educación primaria (hasta cuarto grado), 2 referente a educación primaria de quinto a noveno grado, 3 para a educación secundaria y 4 referente a educación superior.
- Nivel educativo del padre. Se representa de la misma forma que el punto anterior.
- Trabajo de la madre. Es un dato nominal que puede adquirir los siguientes valores: 'teacher' para la enseñanza, 'health' para lo relacionado al cuidado de la salud, 'services' para aquellos trabajos cercanos al área civil como cargos administrativos o policíacos, 'at_home' para el hogar, y 'other' para otra ocupación que no abarca ninguna de las opciones anteriores.
- Trabajo del padre. Involucra los mismos valores del punto previo.
- Razón. Se refiere al motivo por el cual se eligió la escuela correspondiente. Es un dato de tipo nominal que adquiere los siguientes valores: 'home' por la cercanía al domicilio del estudiante, 'reputation' por la reputación de la escuela, 'course' por las clases impartidas y 'other' por alguna otra opción no enlistada.
- Tutor. Es un dato nominal que se refiere a la persona que es responsable de la tutela del estudiante. Puede adquirir los siguientes

valores: 'mother' para la madre, 'father' para el padre y 'other' para otra persona.

- Tiempo de traslado a la escuela. Este dato de tipo numérico abarca cuatro distintos valores: 1 si es menor a 15 minutos, 2 si es de 15 a 30 minutos, 3 si va de 30 minutos a 1 hora, y 4 si es mayor a 1 hora.
- Tiempo semanal dedicado al estudio. De igual forma, este dato es de tipo numérico con cuatro diferentes opciones: 1 si es menos de 2 horas, 2 si es de 2 a 5 horas, 3 si es de 5 a 10 horas o 4 si es mayor a 10 horas.
- Número de clases reprobadas. Es un dato de tipo numérico que va de 0 a 3 según el número de asignaturas; si son más de 3, se toma como valor el número 4.
- Cursos extra clase. Es un dato de tipo binario que puede únicamente los valores si o no.
- Apoyo educativo familiar. También se trata de un dato binario y toma los valores si o no.
- Clases extra pagadas de la materia. Puede tomar solamente los valores si o no.
- Actividades extracurriculares. Indica si el estudiante realiza actividades extracurriculares. Es un dato binario que toma los valores si o no.
- Guardería. Indica si el estudiante asistió a guardería durante su niñez. Al igual que los anteriores, es un dato binario que puede ser si o no.
- Internet. Especifica si el estudiante tiene acceso a internet en casa. Puede tomar los valores si o no.
- Educación superior. Refleja si el estudiante en cuestión desea realizar estudios a nivel superior. Las posibles respuestas son si o no.
- Relación. Se refiere al estado de un estudiante respecto a las relaciones románticas. Es un dato binario que puede ser si o no.
- Calidad de las relaciones familiares. Es un dato numérico que va de 1 (si es muy mala) a 5 (si es excelente).
- Tiempo libre después de la escuela. También se trata de un dato numérico que va de 1 (muy poco) a 5 (valor muy alto).

- Salidas con amigos. Es un dato numérico que va de 1 (muy poco) a 5 (valor muy alto) y se refiere a una valoración respecto a las salidas con amistades.
- Consumo de alcohol en jornada estudiantil. De igual forma es un dato numérico que va de 1 (muy poco) a 5 (muy alto).
- Consumo de alcohol en fin de semana. Es un dato numérico que va de 1 (muy poco) a 5 (muy alto).
- Estado de salud. Este es un dato numérico que indica el estado de salud del estudiante, va de 1 (muy malo) a 5 (muy bueno).
- Faltas. Indica el número de faltas escolares, este dato numérico va de 0 a 93.
- G1. Corresponde a la calificación de la primera evaluación respecto a la materia. Es un dato numérico que va de 0 a 20.
- G2. Es la calificación de la segunda evaluación de la materia. También se trata de un dato numérico que va de 0 a 20.

En el conjunto de datos de entrenamiento, se localiza el dato G3, el cual se refiere a la calificación de la tercera evaluación de la materia. De igual forma se trata de un dato numérico que va de 0 a 20. Es utilizada para que el modelo “aprenda” a predecir adecuadamente conforme los casos empleados.

Como ya se mencionó, la variable objetivo es de tipo continuo y se refiere a la calificación obtenida por el alumno, la cual, dentro del contexto educativo portugués va de 0 a 20, siendo el 10 la calificación mínima aprobatoria. La definición de seis modelos se debe al uso de seis algoritmos diferentes. El código está construido para recibir los datos desde un archivo .csv con cadenas de texto o números según corresponde y posteriormente mapear las categorías a números enteros mediante un diccionario de datos para que se procesen correctamente. Fue necesario agregar código con la finalidad de almacenar todos los modelos ya entrenados.

El primer modelo utilizó el algoritmo de regresión lineal; para llevar a cabo las pruebas de la métrica, se consideró elaborar un nuevo archivo .csv con los datos asignados a cada columna. En este caso, se trató de diez alumnos de nacionalidad portuguesa dentro del mismo contexto educativo, geográfico y socio-cultural. En las imágenes 8, 9 y 10 se observan los datos contenidos en el archivo.

	A	B	C	D	E	F	G	H	I	J	K	L
1	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	reason	guardian
2	GP	F	15 U	GT3	T		2	1	services	other	reputation	father
3	GP	M	15 U	LE3	T		4	4	health	services	course	father
4	GP	M	15 U	GT3	T		4	3	teacher	other	course	mother
5	GP	M	15 U	GT3	A		2	2	other	other	home	other
6	GP	F	16 U	GT3	T		4	4	health	other	home	mother
7	MS	F	17 R	GT3	T		1	2	other	other	course	mother
8	MS	F	18 R	LE3	T		4	4	other	other	reputation	mother
9	MS	F	18 R	GT3	T		1	1	other	other	home	mother
10	MS	F	20 U	GT3	T		4	2	health	other	course	other
11	MS	F	18 R	LE3	T		4	4	teacher	services	course	mother

Ilustración 8. Primera parte del conjunto de datos a prueba.
Elaboración propia.

	traveltime	studytime	failures	schoolsup	famsup	paid	activities	nursery	higher	internet	romantic	famrel	t
2	3	3	0 no	yes	no	yes	yes	yes	yes	yes	no		5
3	1	1	0 no	yes	yes	yes	yes	yes	yes	yes	no		4
4	2	2	0 no	yes	yes	no	no	yes	yes	yes	no		5
5	1	3	0 no	yes	no	no	no	yes	yes	yes	yes		4
6	1	1	0 no	yes	no	no	no	yes	yes	yes	no		4
7	1	1	0 no	no	no	yes	yes	yes	yes	yes	no		3
8	2	3	0 no	no	no	no	no	yes	yes	yes	no		5
9	4	3	0 no	no	no	no	no	yes	yes	yes	no		4
10	2	3	2 no	yes	yes	no	no	no	yes	yes	yes		5
11	1	2	0 no	no	no	yes	yes	yes	yes	yes	no		5

Ilustración 9. Segunda parte del conjunto de datos a prueba.
Elaboración propia.

	famrel	freetime	goout	Dalc	Walc	health	absences	G1	G2	G3
2	5	2	2	1	1	4	4	10	12	12
3	4	3	3	1	3	5	2	14	14	14
4	5	4	3	1	2	3	2	10	10	11
5	4	5	2	1	1	3	0	14	16	16
6	4	4	4	1	2	2	4	14	14	14
7	3	5	5	1	3	1	14	6	5	5
8	5	4	4	1	1	1	0	19	18	19
9	4	3	2	1	2	4	2	8	8	10
10	5	4	3	1	1	3	4	15	14	15
11	5	4	3	3	4	2	4	8	9	10

Ilustración 10. Tercera parte del conjunto de datos a prueba.
Elaboración propia.

Una vez cargado el archivo .csv, se efectuó la predicción correspondiente, obteniendo para el primer algoritmo los resultados que se aprecian en la ilustración 11, tomando en cuenta que el número de casos va de 0 a 9.

	Modelo
0	11.827206
1	14.141746
2	10.841817
3	16.215384
4	14.444424
5	4.256938
6	19.160525
7	7.622054
8	14.231617
9	7.639874

Ilustración 11. Predicciones del modelo 1.
Elaboración propia.

Para iniciar el análisis del modelo y sus respuestas proporcionadas, se realizó la identificación de las variables sensibles abarcadas por la métrica y aquellas utilizadas por el modelo con ayuda de una tabla. Su elaboración involucra considerar si se hace uso de proxies o variables que reflejen alguna de las categorías protegidas, ya que en muchos modelos es común observarlo. A continuación, se muestran los resultados de ello en la tabla 6.

Tabla 6. Análisis de variables sensibles en el modelo 1.

Variables sensibles	Variables sensibles empleadas en el modelo	Proxies usados en el modelo.
Género/Sexo	Sexo	-
Orientación sexual	-	-
Religión	-	-
Opinión política	-	-
Nacionalidad	-	-
Origen étnico	-	-
Raza	-	-
Idioma	-	-
Discapacidad	-	-
Nivel socioeconómico		Escuela Domicilio del estudiante. Ocupación de la madre. Ocupación del padre.
Escolaridad de los padres	Educación de la madre. Educación del padre.	-
Antecedentes migratorios	-	-

Elaboración propia.

El guion señalado en algunas de las variables, indica que el dato no es solicitado por el modelo ni tampoco se tiene información al respecto en el conjunto de datos. Si bien, desde un inicio se menciona que los estudiantes son de nacionalidad portuguesa, este dato no cobra relevancia debido a que todos pertenecen a la misma nacionalidad y el modelo fue creado e implementado en el mismo contexto.

Como siguiente paso se realizó el análisis de la predicción, comenzando con la ubicación de los datos correspondientes a las variables sensibles en el conjunto empleado para las pruebas, los cuales se observan en la imagen 12.

caso	school	sex	address	Medu	Fedu	Mjob	Fjob	Predicción
1	GP	F	U	5to a 9no grado	Primaria	services	other	11.82721
2	GP	M	U	Superior	Superior	health	services	14.14175
3	GP	M	U	Superior	Secundaria	teacher	other	10.84182
4	GP	M	U	5to a 9no grado	5to a 9no grado	other	other	16.21538
5	GP	F	U	Superior	Superior	health	other	14.44442
6	MS	F	R	Primaria	5to a 9no grado	other	other	4.256938
7	MS	F	R	Superior	Superior	other	other	19.16053
8	MS	F	R	Primaria	Primaria	other	other	7.622054
9	MS	F	U	Superior	5to a 9no grado	health	other	14.23162
10	MS	F	R	Superior	Superior	teacher	services	7.639874

Ilustración 12. Variables sensibles del modelo 1.
Elaboración propia.

Tomando en cuenta cada caso, se analizaron los resultados de las predicciones como se observa en la gráfica 1.



Gráfica 1. Predicciones del modelo 1.
Elaboración propia.

A partir de la gráfica, se obtuvieron las siguientes interpretaciones:

- El caso 6 y 7 corresponden a las calificaciones más bajas y altas respectivamente.
- Ambos casos son de estudiantes de sexo femenino.
- Las dos predicciones son de estudiantes que viven en una zona rural.
- Ambos casos son de alumnos que estudian en la escuela Mousinho da Silveira.
- En el caso 7, los padres de los estudiantes tienen educación superior; en el caso 6, los padres tienen educación primaria y de quinto a noveno grado.
- Las dos predicciones son de estudiantes cuyos padres se dedican a una actividad diferente a las listadas.

Posteriormente se realizaron tablas y gráficas que resumen e ilustran cada variable de interés seleccionada en la tabla 6, así como la cantidad de predicciones reprobatorias, puesto que se trata de la variable objetivo; la elaboración de dichas herramientas tiene la finalidad de evaluar cuántos casos reflejan sesgo ético.

La primera de las variables sensibles es el sexo; en la tabla 7 se destacan en color rojo las calificaciones reprobatorias junto con el sexo del estudiante; de igual forma sobresalen las calificaciones más alta y baja respectivamente en color azul. A partir de estos datos, se elaboró la gráfica 2, donde se distingue que hay tres casos reprobatorios para estudiantes del sexo femenino y ninguno para estudiantes del sexo masculino.

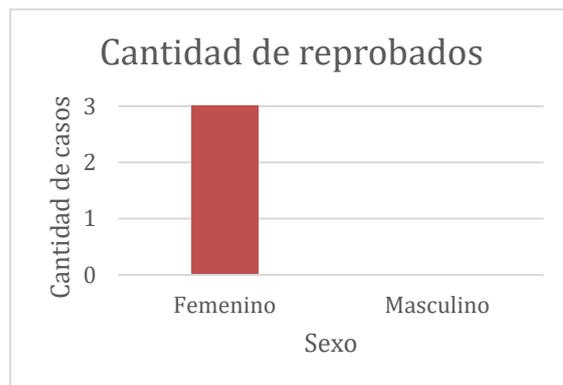
Tabla 7. Análisis del modelo 1 respecto al sexo.

Caso	Predicción	Sexo
1	11.827206	F
2	14.141746	M

3	10.841817	M
4	16.215384	M
5	14.444424	F
6	4.256938	F
7	19.160525	F

8	7.622054	F
9	14.231617	F
10	7.639874	F

Elaboración propia.



Gráfica 2. Reprobados del modelo 1 respecto al sexo.
Elaboración propia.

La siguiente variable sensible analizada fue el nivel socioeconómico, donde destacan tres características: zona donde vive el alumno, escuela donde estudia, y trabajo de la madre y del padre. Se elaboraron las tablas 8, 9 y 10 para cada una con la finalidad de identificar, al igual que en los casos anteriores, las calificaciones reprobatorias en color rojo y las calificaciones más alta y baja respectivamente en color azul.

Tabla 8. Análisis del modelo 1 respecto a la zona donde vive.

Caso	Predicción	Domicilio
1	11.827206	U
2	14.141746	U
3	10.841817	U
4	16.215384	U
5	14.444424	U
6	4.256938	R
7	19.160525	R
8	7.622054	R
9	14.231617	U
10	7.639874	R

Elaboración propia.

Tabla 9. Análisis del modelo 1 respecto a la escuela donde estudia.

Caso	Predicción	Escuela
1	11.827206	GP
2	14.141746	GP
3	10.841817	GP
4	16.215384	GP
5	14.444424	GP
6	4.256938	MS
7	19.160525	MS
8	7.622054	MS
9	14.231617	MS
10	7.639874	MS

Elaboración propia.

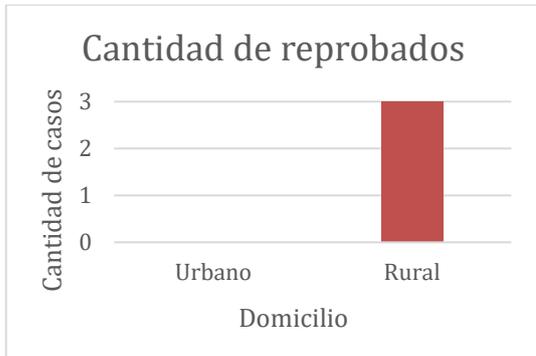
Tabla 10. Análisis del modelo 1 respecto al trabajo de la madre y el padre.

Caso	Predicción	Mjob	Fjob
1	11.827206	services	other
2	14.141746	health	services
3	10.841817	teacher	other
4	16.215384	other	other

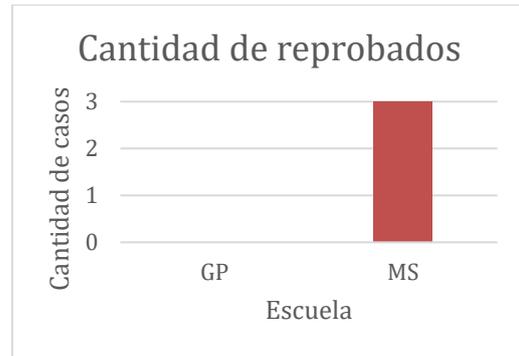
5	14.444424	health	other
6	4.256938	other	other
7	19.160525	other	other
8	7.622054	other	other
9	14.231617	health	other
10	7.639874	teacher	services

Elaboración propia.

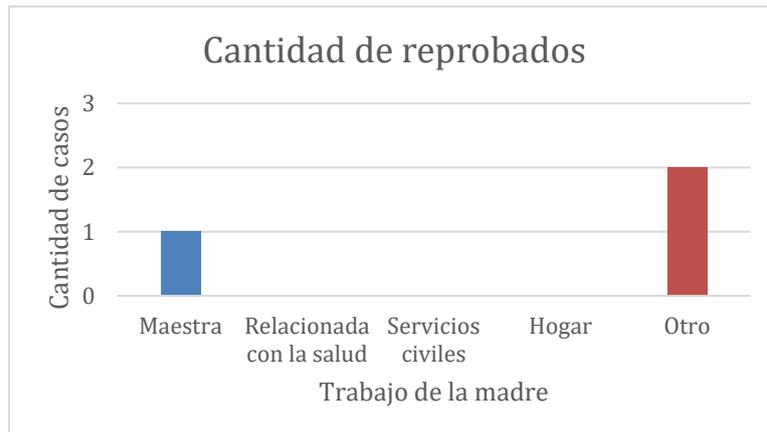
Derivado de los hallazgos encontrados con la realización de las tablas, se construyeron las gráficas 3, 4, 5 y 6.



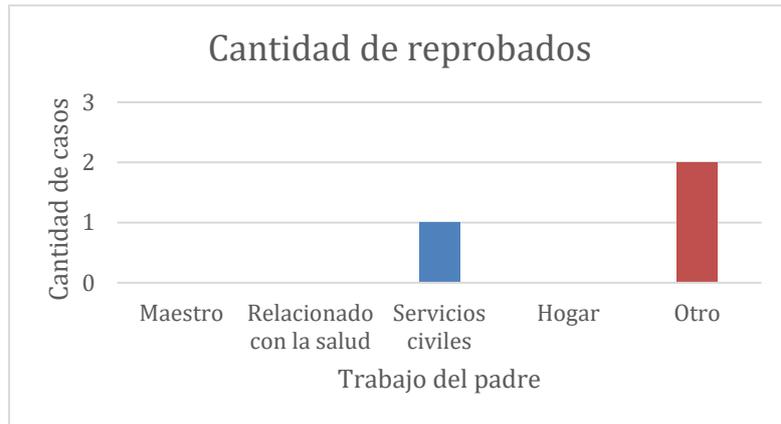
Gráfica 3. Reprobados del modelo 1 respecto a la zona donde vive. Elaboración propia.



Gráfica 4. Reprobados del modelo 1 respecto a la escuela donde estudia. Elaboración propia.



Gráfica 5. Reprobados del modelo 1 respecto al trabajo de la madre. Elaboración propia.



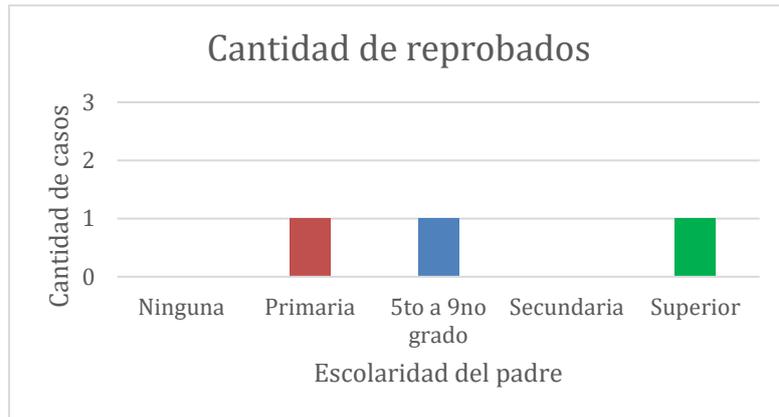
Gráfica 6. Reprobados del modelo 1 respecto al trabajo del padre. Elaboración propia.

Para concluir el análisis se visualizan los datos correspondientes a la escolaridad de los padres. A continuación, se muestra la tabla 11 y gráfica 7 sobre los casos reprobatorios.

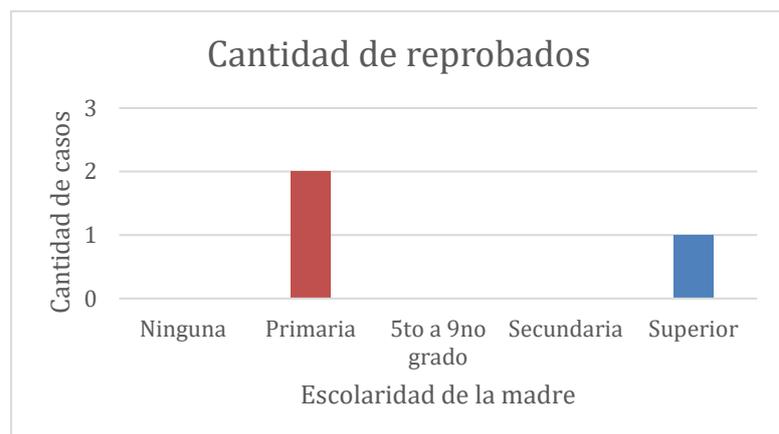
Tabla 11. Análisis del modelo 1 respecto a la escolaridad de los padres.

Caso	Predicción	Medu	Fedu
1	11.827206	5to a 9no grado	Primaria
2	14.141746	Superior	Superior
3	10.841817	Superior	Secundaria
4	16.215384	5to a 9no grado	5to a 9no grado
5	14.444424	Superior	Superior
6	4.256938	Primaria	5to a 9no grado
7	19.160525	Superior	Superior
8	7.622054	Primaria	Primaria
9	14.231617	Superior	5to a 9no grado
10	7.639874	Superior	Superior

Elaboración propia.



Gráfica 7. Reprobados del modelo 1 respecto a la escolaridad del padre. Elaboración propia.



Gráfica 8. Reprobados del modelo 1 respecto a la escolaridad de la madre. Elaboración propia.

De los tres alumnos con calificación reprobatoria, los tres son del sexo femenino; los tres viven en una zona rural; los tres asisten a la escuela Mousinho da Silveira; los padres de dos de ellos se dedican a actividades diferentes a las listadas y solo en un caso, la madre es docente y el padre trabaja en servicios civiles; la madre y padre del estudiante en el primer caso tienen educación primaria y de quinto a noveno grado respectivamente, en el segundo caso ambos tienen educación primaria y en el tercer caso ambos tienen educación superior.

Con base en los datos analizados, es posible observar cierta desventaja hacia variables relacionadas con el sexo, nivel socioeconómico y, en cierta parte, con la escolaridad de los padres. A continuación, se muestra la tabla 12 donde se resumen los hallazgos.

Tabla 12. Reprobados del modelo 1 respecto a las variables sensibles.

Dato de entrada	Valor del dato	Cantidad de reprobados
Sexo	Femenino	3
Zona donde vive	Rural	3
Escuela donde estudia	Mousinho da Silveira	3
Trabajo de la madre	Otro	2
	Docente	1
Trabajo del padre	Otro	2
	Servicios civiles	1
Escolaridad de la madre	Primaria	2
	Superior	1
Escolaridad del padre	5to a 9no grado	1
	Primaria	1
	Superior	1

Elaboración propia.

Utilizando los resultados del análisis, se procedió a responder los ítems de la escala de Likert propuesta para determinar el nivel de sesgo ético dentro del modelo evaluado. La primera dimensión se refiere a la reparación de la desventaja. Las predicciones obtenidas muestran resultados benéficos para diferentes grupos, no obstante, se localizaron algunas excepciones respecto las categorías ya mencionadas. El resto de las variables sensibles, al no ser incluidas dentro de los datos ni especificadas, no representaron problema alguno. Las respuestas se muestran en las imágenes 13, 14 y 15:

Ilustración 13. Primera parte de la contestación de ítems para el modelo 1.

Elaboración propia.

Métrica cualitativa para sesgo ético

5. Nacionalidad

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

6. Origen étnico

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

7. Raza

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

8. Idioma

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

9. Discapacidad

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

Ilustración 14. Segunda parte de la contestación de ítems para el modelo 1.
Elaboración propia.

Métrica cualitativa para sesgo ético

10. Nivel socioeconómico

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

11. Escolaridad de los padres

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

12. Antecedentes migratorios

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

Ilustración 15. Tercera parte de la contestación de los ítems para el modelo 1.
Elaboración propia.

La siguiente dimensión aborda la preferencia negativa hacia alguna característica. Observando los resultados, se distinguen casos reprobatorios que se vinculan al sexo y nivel socioeconómico. Con base en ello, se respondieron los respectivos ítems como se muestra en las ilustraciones 16, 17 y 18.

Métrica cualitativa para sesgo ético

B) Corregir el estigma, los estereotipos y la humillación.
Los resultados obtenidos reflejan una preferencia negativa hacia alguna de las siguientes características del estudiante.

13. Género/ Sexo

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

14. Orientación sexual.

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

15. Religión

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

16. Opinión política

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

17. Nacionalidad

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

Ilustración 16. Cuarta parte de la contestación de los ítems para el modelo 1.
Elaboración propia.

Métrica cualitativa para sesgo ético

18. Origen étnico

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

19. Raza

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

20. Idioma

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

21. Discapacidad

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

22. Nivel socioeconómico

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

Ilustración 17. Quinta parte de la contestación de los ítems para el modelo 1.
Elaboración propia.

Métrica cualitativa para sesgo ético

23. Escolaridad de los padres

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

24. Antecedentes migratorios

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

Ilustración 18. Sexta parte de la contestación de los ítems para el modelo 1.
Elaboración propia.

La tercera dimensión se vincula con la falta de inclusión de estudiantes pertenecientes a diferentes grupos de cada categoría. De acuerdo con los resultados obtenidos, se incluyen a diversos grupos definidos desde la fase de la recopilación de datos. No obstante, hay algunos problemas referentes al sexo, ya que la mayoría de los casos aprobatorios corresponden solo al sexo masculino. Al no limitar ni hacer referencia en datos de los estudiantes a un tipo de orientación sexual, religión, opinión política, nacionalidad, origen étnico, raza, idioma, discapacidad y antecedentes migratorios de los padres, se considera que hay inclusión de cualquier grupo. Las respuestas a esta dimensión se observan en las imágenes 19, 20 y 21.

Métrica cualitativa para sesgo ético

C) Dimensión participativa.
Los resultados obtenidos no benefician a estudiantes de diferentes grupos pertenecientes a las siguientes categorías:

25. Género/Sexo

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

26. Orientación sexual.

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

27. Religión

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

28. Opinión política

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

29. Nacionalidad

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

Ilustración 19. Séptima parte de la contestación de los ítems para el modelo 1.
Elaboración propia.

Métrica cualitativa para sesgo ético

30. Origen étnico

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

31. Raza

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

32. Idioma

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

33. Discapacidad

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

Ilustración 20. Octava parte de la contestación de los ítems para el modelo 1.
Elaboración propia.

Métrica cualitativa para sesgo ético

34. Nivel socioeconómico

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

35. Escolaridad de los padres

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

36. Antecedentes migratorios

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

Ilustración 21. Novena parte de la contestación de los ítems para el modelo 1.
Elaboración propia.

Finalmente, respecto a la última dimensión de la métrica, se evaluó el contexto de implementación. Como los datos utilizados para la prueba son referentes al mismo contexto educativo, sociocultural y geográfico que el original, la mayoría de las respuestas corresponden a la categoría “Muy en desacuerdo”, sin embargo, en el caso del contexto temporal, existe una variación de cinco años, ya que el modelo fue construido en 2019. Los últimos cuatro ítems quedaron contestados como se aprecia en la ilustración 22.

Métrica cualitativa para sesgo ético

D) Evaluación del contexto de implementación

37. Los resultados del modelo se implementan para resolver un problema diferente al que es su finalidad solucionar.

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

38. La época o contexto temporal tiene características diferentes a la época en la que se reunieron los datos con los que el modelo fue entrenado.

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

39. El estado, nación o contexto geográfico que tiene características diferentes al lugar donde se planeó implementarlo cuando fue creado.

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

40. El ambiente socio-cultural que tiene características diferentes al entorno donde se planeó implementar.

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

Reniciar Cuestionario

Calcular Nivel de Sesgo Ético

Ilustración 22. Parte final de la contestación de los ítems para el modelo 1.
Elaboración propia.

Al dar clic en “Calcular el Nivel de Sesgo ético”, el código realizado calculó el nivel de sesgo conforme las condiciones especificadas por los baremos. De acuerdo con ello, se idéntica que el modelo 1 tiene un nivel de sesgo ético muy bajo, con una puntuación de 58 como se muestra en la imagen 23.

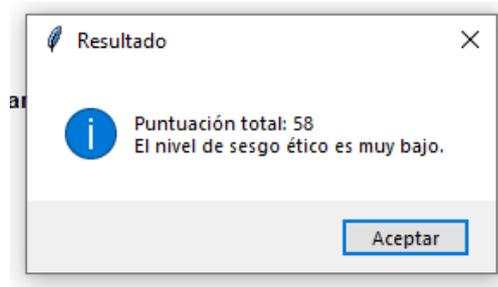


Ilustración 23. Nivel de sesgo obtenido en el modelo 1.
Elaboración propia.

El resto de los modelos que se presentan en el proyecto utilizado para estas pruebas, se encuentran en los Anexos C al G.

El segundo proyecto localizado, se encarga de predecir la calificación semestral de una asignatura dentro del nivel superior. Debido al lenguaje en el que está elaborado el reporte, se identifica que fue desarrollado en Turquía. Los datos de entrenamiento se encontraban en idioma inglés y originalmente se obtuvieron de 145 estudiantes pertenecientes a la Facultad de Ingeniería y la Facultad de Ciencias de la Educación. Las características del estudiante incluidas en el modelo son manejadas mediante un archivo .csv con datos escritos en forma de cadena; como parte del código en el modelo, se realiza el mapeo correspondiente que facilita el procesamiento de los datos. Estas características son:

- ID del estudiante. Este es un identificador que no influye directamente en la predicción.
- Edad del estudiante. Es un dato continuo que va de 18 a más de 26 años. Para representarlos se utiliza el número 1 en el rango de 18-21, 2 en el rango de 22-25, y 3 para mayores de 26.

- Género. Es un dato que indica el género del estudiante. Para el presente modelo, solamente se mencionan dos: femenino, representado con el número 1, y masculino, representado con el número 2.
- Tipo de bachillerato. Se refiere al tipo de escuela al que asistió el alumno antes de ingresar a la universidad. Emplea el número 1 para un bachillerato privado, 2 para público y 3 para indicar otro tipo.
- Tipo de beca. Indica si el estudiante recibe alguna beca, es un dato nominal que se representa con números. Si el valor es 'Ninguna' se utiliza el 1, si es '25%' emplea el 2, si es '50%' se representa con el 3, si es '75%' utiliza el 4 y finalmente 'Completa' se refleja con el 5.
- Trabajo adicional: Es un dato binario que refleja si el alumno trabaja a la par que estudia. Toma los valores sí o no y se representa con el 1 o 0 respectivamente.
- Actividad artística o deportiva regular. Indica si el estudiante realiza actividades deportivas o artísticas extra clase. Se representa de la misma forma que el punto anterior.
- Relaciones afectivas. Es un dato binario que se responde con un sí o un no, representándose con (representándose con 1 y 2 respectivamente).
- Salario total si el alumno trabaja. Se refiere al rango del salario en dólares que el estudiante percibe en caso de trabajar. Puede tomar los valores del 1 al 5 según la cantidad: 135-200, 201-270, 271-340, 341-410 y más de 410 respectivamente.
- Transporte a la universidad. Es un dato nominal que indica el tipo de transporte que utiliza el alumno para acudir a la universidad. Con el número 1 se indica un traslado en autobús, con el 2 en coche privado o taxi, con el 3 en bicicleta y con el 4 refleja que utiliza otra forma no listada.
- Tipo de alojamiento. Se emplea este campo para describir el tipo de lugar donde reside el estudiante. Puede ser en renta o alquiler, como

dormitorio, con la familia u otra opción no listada. De igual forma se hace uso de los números para codificar los datos.

- Escolaridad de la madre. Se tienen seis niveles que van de mayor a menor primaria, secundaria, bachillerato, universidad, maestría, y doctorado; se representan con los números del 1 al 6.
- Escolaridad del padre. Hace uso de los mismos valores que el punto anterior.
- Número de hermanas/hermanos. Se representa del 1 al 4 si el número de hermanos es menor a 5, de lo contrario se simboliza con el 5.
- Estado parental. Es un dato nominal que se refiere al estado de sus padres, casado (1), divorciado (2) y fallecido (3), ya sea uno de ellos o ambos.
- Ocupación de la madre. Incluye las categorías: jubilada, ama de casa, funcionaria pública, empleada del sector privado, trabajadora por cuenta propia u otra y se numeran del 1 al 6 respectivamente.
- Ocupación del padre. Es similar al punto anterior, ya que toma en cuenta las categorías: jubilado, funcionario público, empleado del sector privado, trabajador autónomo u otro. Se codifican del 1 al 5.
- Horas de estudio semanales. Refleja las horas de estudio extra clase que el alumno dedica a la semana. Si el valor es Ninguna se representa con el 1, si es menor a 5 se representa con el 2, si es de 6 a 10 horas se representa con el 3, si es de 11 a 20 horas se utiliza el 4 y si son más de 20 se emplea el 5.
- Frecuencia de lectura sobre libros y revistas no científicas. Es un dato nominal que abarca tres posibles valores: Ninguna, a veces o a menudo; se representa con los números del 1 al 3.
- Frecuencia de lectura (libros/revistas científicas): De igual forma se trata de un dato nominal y se representa de la misma forma que el punto anterior.
- Asistencia a seminarios/congresos relacionados con el departamento. Es un dato de tipo binarios que puede tomar dos valores: Si o No. Puede codificarse con 1 o 2 respectivamente.

- Impacto de sus proyectos/actividades en su éxito. Adquiere tres valores que se mapean de forma numérica: 1 si es positivo, 2 si es negativo o 3 si es neutral.
- Asistencia a clases. Se brindan tres opciones para representarla con su respectivo código numérico para el procesamiento de datos: siempre (1), a veces (2) y nunca (3).
- Preparación para exámenes parciales 1. Este dato nominal indica la forma en que el alumno se prepara para los exámenes, refiriéndose a si desea hacerlo solo, con amigo o si ninguna de las opciones es aplicable. Se representan con el número del 1 al 3 respectivamente.
- Preparación para exámenes parciales 2: Este dato ordinal se dirige a identificar en qué momento el alumno estudia para los exámenes. Puede ser la fecha más cercana al examen, regularmente durante el semestre o nunca. Se mapean de la misma forma que el punto anterior.
- Tomar apuntes en clases. De igual forma se trata de un dato ordinal que toma tres valores: nunca, a veces, siempre. Se presentan con los números del 1 al 3.
- Escuchar en clases. Indica si el estudiante presta atención en las clases escuchando la cátedra. Se representa de la misma forma que el punto anterior.
- La discusión mejora mi interés y éxito en el curso. Es una afirmación que muestra el impacto del debate entre estudiantes para mejorar su rendimiento académico. Se mapea de la misma forma que el punto anterior.
- Aula invertida. Indica si esta técnica pedagógica es no útil, útil o no aplicable en cada caso. Emplea los números del 1 al 3 para el procesamiento de los datos.
- Promedio acumulado de calificaciones del último semestre. Si el promedio es menor a 2.00 se representa con el número 1, si está entre el rango de 2.00 a 2.49 se emplea el número 2, si va de 2.50 a 2.99 se representa con el número 3, si es de 3.00 a 3.49 se mapea con el número 4, y finalmente si es superior a 3.49 se emplea el 5.

- Promedio acumulado esperado de calificaciones al terminar el semestre. Se refiere a la expectativa sobre el promedio final que posee el estudiante sobre sí mismo. Se representa de la misma forma que el punto anterior.
- Identificación del curso. Es un número que identifica a una materia en específico. El conjunto de datos no muestra la equivalencia de los números que los representan.

La variable objetivo del modelo es la calificación semestral de cada alumno en un curso o asignatura. Dentro del contexto abordado, se presentan 7 categorías diferentes representadas por números del 0 al 7.

0. Reprobado

1. DD. Insuficiente. Puede ser aprobado bajo ciertas condiciones.

2. DC. Condicionalmente suficiente.

3. CC. Suficiente.

4. CB. Suficiente bueno.

5. BB. Bueno.

6. BA. Muy bueno.

7. AA. Excelente.

El proyecto incluye cuatro algoritmos distintos para efectuar la predicción, por ende, fueron tomados para las siguientes cuatro pruebas. Al igual que en el anterior, el código recibe los datos desde un archivo .csv con cadenas de texto o números para ser mapeados y procesados adecuadamente según lo indicado. También se añadió código que permitió almacenar los modelos entrenados para hacer las pruebas.

Estos modelos únicamente solicitan los datos mencionados; sin embargo, para contemplar otros elementos que no han sido considerados y forman parte de la evaluación en la métrica, se proponen a manera de un caso hipotético los siguientes supuestos: el archivo .csv tiene datos sobre alumnos

con tres nacionalidades distintas: Argentina, Chilena y Mexicana; los estudiantes mexicanos son de raza indígena; se contempla que la implementación del modelo se lleva a cabo en Latinoamérica, específicamente en Argentina, considerando que los estudiantes de otras nacionalidades están presentes debido a un programa de movilidad estudiantil temporal; aunado a ello, su uso está proyectado para la predicción de deserción escolar, tomando en cuenta que aquellas calificaciones menores a CC están en riesgo de desertar.

Basado en ello, se efectúa el análisis de manera similar a la prueba mostrada previamente. En este modelo se emplea un algoritmo de árboles de decisión a modo de clasificador; en primer lugar, se elaboró un archivo .csv con los datos de diez alumnos acorde a lo solicitado, cuyo contenido se aprecia en las ilustraciones 24, 25 y 26.

STUDENTID	AGE	GENDER	HS_TYPE	SCHOLARSHI	WORK	ACTIVITY	PARTNER	SALARY	TRANSPORT
STUDENT2	22-25	Male	Private	25%	Yes	No	No	\$135-200	Bus
STUDENT3	22-25	Male	Private	50%	Yes	Yes	Yes	\$201-270	Other
STUDENT4	18-21	Female	Private	75%	Yes	No	No	\$201-270	Bus
STUDENT5	22-25	Male	State	25%	Yes	Yes	Yes	\$271-340	Private car/taxi
STUDENT6	22-25	Male	Private	50%	Yes	No	No	\$201-270	Private car/taxi
STUDENT7	18-21	Male	Private	75%	Yes	No	Yes	\$135-200	Bus
STUDENT8	18-21	Female	State	25%	Yes	Yes	Yes	\$201-270	Other
STUDENT9	22-25	Female	State	50%	Yes	No	No	\$135-200	Bus
STUDENT10	22-25	Female	State	25%	Yes	No	Yes	\$135-200	Other
STUDENT11	18-21	Female	State	50%	Yes	No	No	\$135-200	Private car/taxi

Ilustración 24. Primera parte del nuevo conjunto de datos a prueba. Elaboración propia.

LIVING	MOTHER_EDU	FATHER_EDU	#_SIBLINGS	KIDS	MOTHER_JOB	FATHER_JOB	STUDY_HRS	READ_FREQ	READ_FREQ
Rental	High school	University		1 Married	Self-employment	Self-employment	<5 hours	Sometimes	Never
Dormitory	Secondary school	Secondary school		2 Married	Housewife	Self-employment	<5 hours	Often	Often
Dormitory	Primary school	Secondary school		5 Married	Housewife	Government officer	6-10 hours	Sometimes	Sometimes
Rental	High school	High school		2 Married	Government officer	Self-employment	<5 hours	Never	Never
Rental	High school	High school		2 Married	Self-employment	Private sector employee	None	Sometimes	Sometimes
With family	Primary school	Primary school		1 Married	Housewife	Self-employment	<5 hours	Sometimes	Sometimes
With family	University	High school		2 Married	Private sector employee	Private sector employee	None	Often	Often
With family	Secondary school	University		2 Married	Housewife	Self-employment	<5 hours	Sometimes	Sometimes
Dormitory	Primary school	Secondary school		3 Married	Housewife	Private sector employee	<5 hours	Often	Often
With family	Primary school	Secondary school		4 Married	Housewife	Retired	6-10 hours	Never	Never

Ilustración 25. Segunda parte del nuevo conjunto de datos a prueba. Elaboración propia.

ATTEND_DEPT	IMPACT	ATTEND	PREP_STUDY	PREP_EXAM	NOTES	LISTENS	LIKES_DISCU	CLASSROOM	CUML_GPA	EXP_GPA	COURSE ID
Yes	Positive	Yes	With friends	Closest date	Always	Sometimes	Always	Useful	Above 3.49	3.00-3.49	7
Yes	Positive	Yes	Alone	Closest date	Sometimes	Sometimes	Never	Not useful	3.00-3.49	Above 3.49	7
Yes	Positive	No	Alone	Regularly du	Always	Always	Sometimes	Not useful	2.50-2.99	2.00-2.49	7
Yes	Positive	Yes	With friends	Closest date	Always	Always	Sometimes	Not useful	Above 3.49	3.00-3.49	7
Yes	Positive	Yes	Alone	Regularly du	Never	Never	Useful	3.00-3.49	3.00-3.49	7	
No	Positive	No	Alone	Closest date	Sometimes	Always	Always	Not applicat	3.00-3.49	3.00-3.49	7
Yes	Positive	Yes	Not applicat	Closest date	Sometimes	Sometimes	Sometimes	Not useful	<2.00	<2.00	7
Yes	Positive	Yes	Alone	Closest date	Sometimes	Always	Sometimes	Useful	3.00-3.49	2.50-2.99	7
Yes	Positive	No	Alone	Closest date	Sometimes	Sometimes	Sometimes	Useful	<2.00	2.00-2.49	7
No	Positive	No	With friends	Regularly du	Sometimes	Always	Sometimes	Useful	<2.00	2.50-2.99	7

Ilustración 26. Tercera parte del nuevo conjunto de datos a prueba.
Elaboración propia.

Una vez ejecutado el código, se obtuvieron las predicciones mostradas en la imagen 27, considerando que los resultados son arrojados de forma numérica y es necesario consultar su equivalencia para la respectiva interpretación.

```

Model7
7
1
1
7
7
7
1
7
1
1

```

Ilustración 27. Predicciones del modelo 8.
Elaboración propia.

El análisis del modelo y las respuestas proporcionadas requirió la concentración en la tabla 13, de las variables sensibles identificadas por la métrica y aquellas que el modelo emplea.

Tabla 13. Análisis de variables sensibles en el modelo 8.

Variables sensibles	Variables sensibles empleadas en el modelo	Proxies usados en el modelo.
Género/Sexo	Género	-
Orientación sexual	-	-
Religión	-	-
Opinión política	-	-
Nacionalidad*	NA	NA
Origen étnico	NA	NA
Raza*	NA	NA
Idioma	-	-
Discapacidad	-	-
Nivel socioeconómico		Escuela Trabajo del estudiante. Salario del estudiante. Transporte del estudiante a la universidad Tipo de alojamiento. Trabajo del padre. Trabajo de la madre.
Escolaridad de los padres	Escolaridad de la madre. Escolaridad del padre.	-

Antecedentes migratorios*	-	-
----------------------------------	---	---

Elaboración propia.

Es importante considerar que, si bien en el modelo no se toman literalmente ciertas características como la nacionalidad, la raza y antecedentes migratorios, estos sí forman parte del contexto que engloba los casos de cada estudiante, por lo cual es importante tenerlos en consideración al evaluar las respuestas obtenidas y se representan con 'NA' en la tabla.

Como siguiente paso se realizó el análisis de la predicción, comenzando con la ubicación de los datos correspondientes a las variables sensibles en el conjunto empleado para las pruebas, ilustradas en la imagen 28. Con estos resultados, se elaboró la gráfica 9, la cual permite observar el comportamiento del modelo.

Caso	GENDER	HS_TYPE	WORK	SALARY	TRANSPORT	LIVING	MOTHER_EDU	FATHER_EDU	MOTHER_JOB	FATHER_JOB	PREDICCIÓN	Caso	Predicción
1	Male	Private	Yes	\$135-200	Bus	Rental	High school	University	Self-employment	Self-employment	7	1	7
2	Male	Private	Yes	\$201-270	Other	Dormitory	Secondary school	Secondary school	Housewife	Self-employment	1	2	1
3	Female	Private	Yes	\$201-270	Bus	Dormitory	Primary school	Secondary school	Housewife	Government officer	1	3	1
4	Male	State	Yes	\$271-340	Private car/taxi	Rental	High school	High school	Government officer	Self-employment	7	4	7
5	Male	Private	Yes	\$201-270	Private car/taxi	Rental	High school	High school	Self-employment	Private sector employee	7	5	7
6	Male	Private	Yes	\$135-200	Bus	With family	Primary school	Primary school	Housewife	Self-employment	7	6	7
7	Female	State	Yes	\$201-270	Other	With family	University	High school	Private sector employee	Private sector employee	1	7	1
8	Female	State	Yes	\$135-200	Bus	With family	Secondary school	University	Housewife	Self-employment	7	8	7
9	Female	State	Yes	\$135-200	Other	Dormitory	Primary school	Secondary school	Housewife	Private sector employee	1	9	1
10	Female	State	Yes	\$135-200	Private car/taxi	With family	Primary school	Secondary school	Housewife	Retired	1	10	1

Ilustración 28. Variables sensibles del modelo 8.
Elaboración propia.



Gráfica 9. Predicciones del modelo 8.
Elaboración propia.

A partir de ello se obtuvieron las siguientes interpretaciones tomando en cuenta las variables sensibles reflejadas en los datos de entrada del modelo:

- Las predicciones arrojadas únicamente adquieren dos posibles valores: 1 y 7.
- Cinco casos que obtuvieron el valor 1 que representa a la calificación DD, es decir, insuficiente con la posibilidad de ser aprobado bajo ciertas condiciones: 2, 3, 7, 9 y 10.
- Los casos 1, 4, 5, 6 y 8 obtuvieron el valor 7 que representa la calificación AA, es decir, la excelencia.
- Las predicciones de los casos 3, 7, 9 y 10 pertenecen a estudiantes del género femenino; el caso 2 es de un estudiante de género masculino.
- Los casos 1, 4, 5 y 6 son de alumnos de género masculino.
- Las predicciones de los casos 7, 9 y 10 son de estudiantes que acudieron a un bachillerato público; los casos 2 y 3, a uno privado.
- Los casos 1, 5 y 6 son de alumnos provenientes de escuelas privadas; los casos 4 y 8 a uno público.
- Todos los estudiantes trabajan.
- El salario de los estudiantes de los casos 2, 3 y 7, oscila entre los \$201 y \$270 dólares; el de los casos 9 y 10, entre los \$135 y \$200 dólares.
- Los casos 1, 6 y 8, son de estudiantes que perciben un ingreso entre los \$135 y \$200 dólares; el caso 5 entre los \$201 y \$270 dólares; y el caso 4 entre los \$271 y \$340 dólares.
- El tipo de transporte en los casos 2, 7 y 9 es diferente a los señalados en el modelo; en el caso 3 es el autobús y en el caso 10 es en auto particular o taxi.
- En los casos 1, 6 y 8 emplean el autobús; el caso 4 y 5, emplean auto particular o taxi.
- Respecto al tipo de alojamiento, en los casos 2, 3 y 9 se alojan en dormitorios; los casos 7 y 10 son estudiantes que están con familiares.
- Los casos 1, 4 y 5 son de alumnos que rentan y los casos 6 y 8 son de aquellos que viven con su familia.
- Los casos 3, 9 y 10, son de estudiantes cuya madre tiene únicamente la escuela primaria, en el caso 7 tiene escolaridad universitaria y el caso

2 hasta secundaria. A excepción del caso 7 donde la escolaridad del padre es nivel preparatoria, el resto de los casos con la predicción igual a 1, tienen escolaridad secundaria.

- En el caso 1, la madre tiene nivel medio superior y el padre hasta nivel superior; en los casos 4 y 5 ambos tienen nivel preparatoria; en el caso 6 los dos tienen nivel primaria y en el caso 8, la madre tiene nivel secundaria y el padre hasta nivel superior.
- En los casos 2, 3, 9 y 10, la madre se dedica al hogar y el padre es empleado independiente, trabaja en oficina del gobierno, en el sector privado o está retirado respectivamente. En el caso 7, ambos padres trabajan en el sector privado.
- En el caso 1 ambos padres son trabajadores independientes, en el caso 4 la madre trabaja en oficina de gobierno y el padre es trabajador independiente, en el caso 5 la madre es trabajadora independiente y el padre del sector privado, en los casos 6 y 8 la madre es ama de casa y el padre es trabajador independiente.

No obstante, se requiere prestar atención a las características mencionadas de los estudiantes que no fueron solicitadas en el modelo, sin embargo, pueden ayudar a observar casos discriminatorios y deben ser evaluados. En la tabla 14, se observa la nacionalidad y la raza de los estudiantes en cada caso:

Tabla 14. Consideración de otras variables en el modelo 8.

Caso	Nacionalidad	Raza
1	Argentina	Mestizo
2	Mexicana	Mestizo
3	Mexicana	Indígena
4	Argentina	Mestizo
5	Chilena	Mestizo
6	Chilena	Mestizo
7	Mexicana	Indígena
8	Argentina	Mestizo
9	Mexicana	Indígena
10	Mexicana	Indígena

Elaboración propia.

Con ello se pueden determinar otras interpretaciones aunadas a las ya realizadas.

- Los casos 2, 3, 7, 9 y 10 pertenecen a estudiantes de nacionalidad mexicana.
- Los casos 1, 4 y 8 son de estudiantes de nacionalidad argentina y los casos 5 y 6 de nacionalidad chilena.
- El caso 2 es de un alumno de raza mestiza, los casos 3, 7, 9 y 10 son de estudiantes de raza indígena.
- El caso 1, 4, 5, 6 y 8 son de alumnos de raza mestiza.

Tomando como base las interpretaciones realizadas, se realizaron las correspondientes tablas y gráficas de cada variable de interés localizada, y la cantidad de predicciones con la calificación más baja. En este caso no se hace uso de la cantidad de reprobados porque nadie reprobó, sin embargo, es observable que existen muchos casos con calificación DD que es la calificación previa a la reprobatoria.

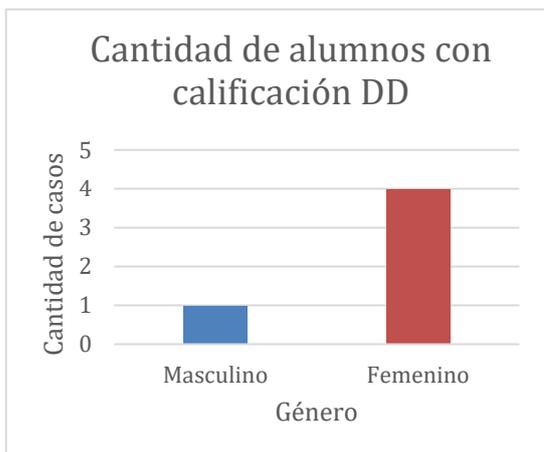
La primera variable sensible fue el género, en la tabla 15 se muestran destacados en color rojo los estudiantes con la calificación equivalente a DD en color negro. El resto de los casos equivalen a AA con la predicción igual a 7. A partir de estos datos, se construye la gráfica 10, en la cual se visualiza que la mayoría de los casos con calificación baja son de género femenino.

Tabla 15. Análisis del modelo 8 respecto al género.

Caso	Predicción	Gender
1	7	Male
2	1	Male
3	1	Female
4	7	Male
5	7	Male
6	7	Male
7	1	Female

8	7	Female
9	1	Female
10	1	Female

Elaboración propia.



Gráfica 10. Calificación DD obtenida del modelo 8 respecto al género.
Elaboración propia.

La siguiente variable sensible analizada fue el nivel socioeconómico, en la cual se abordaron características: trabajo del estudiante, salario del estudiante, tipo de vivienda o alojamiento donde se encuentra, tipo de transporte que emplea para ir a la universidad, tipo de bachillerato del que proviene y la ocupación o trabajo de la madre y el padre. Para cada una se elaboraron las tablas 16, 17, 18, 19, 20 y 21 como en la categoría anterior.

Tabla 16. Análisis del modelo 8 respecto a si el estudiante tiene empleo.

Caso	Predicción	Work
1	7	Si
2	1	Si
3	1	Si
4	7	Si
5	7	Si
6	7	Si
7	1	Si
8	7	Si
9	1	Si
10	1	Si

Elaboración propia.

Tabla 17. Análisis del modelo 8 respecto al salario del estudiante.

Caso	Predicción	Salary
1	7	\$135-200

2	1	\$201-270
3	1	\$201-270
4	7	\$271-340
5	7	\$201-270
6	7	\$135-200
7	1	\$201-270
8	7	\$135-200
9	1	\$135-200
10	1	\$135-200

Elaboración propia.

Tabla 18. Análisis del modelo 8 respecto al tipo de alojamiento.

Caso	Predicción	Living
1	7	Rental
2	1	Dormitory
3	1	Dormitory
4	7	Rental
5	7	Rental

6	7	With family
7	1	With family
8	7	With family
9	1	Dormitory
10	1	With family

Elaboración propia.

Tabla 19. Análisis del modelo 8 respecto al tipo de transporte.

Caso	Predicción	Transport
1	7	Bus

2	1	Other
3	1	Bus
4	7	Private car/taxi
5	7	Private car/taxi
6	7	Bus
7	1	Other
8	7	Bus
9	1	Other
10	1	Private car/taxi

Elaboración propia.

Tabla 20. Análisis del modelo 8 respecto al tipo de bachillerato.

Caso	Predicción	Hs_Type
1	7	Private
2	1	Private
3	1	Private
4	7	State
5	7	Private
6	7	Private
7	1	State
8	7	State
9	1	State
10	1	State

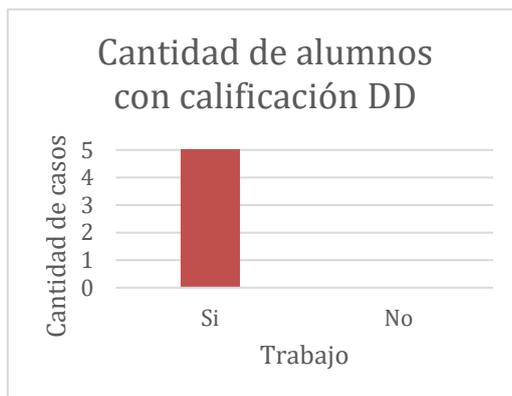
Elaboración propia.

Tabla 21. Análisis del modelo 8 respecto al empleo de los padres.

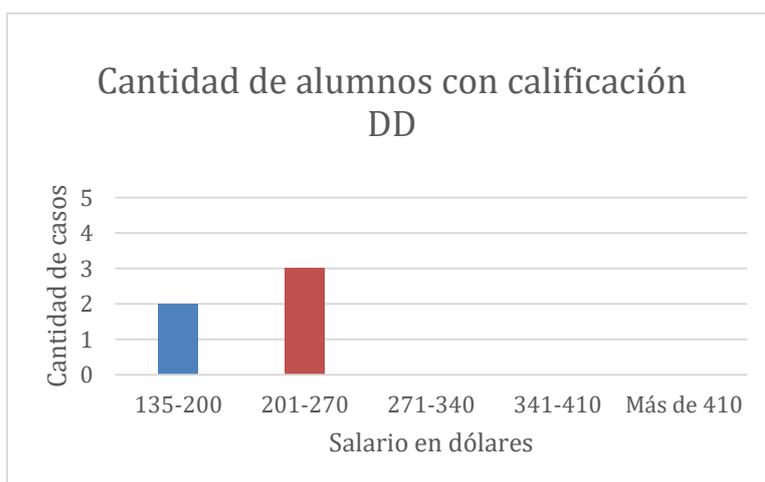
Caso	Predicción	Mother_Job	Father_Job
1	7	Self-employment	Self-employment
2	1	Housewife	Self-employment
3	1	Housewife	Government officer
4	7	Government officer	Self-employment
5	7	Self-employment	Private sector employee
6	7	Housewife	Self-employment
7	1	Private sector employee	Private sector employee
8	7	Housewife	Self-employment
9	1	Housewife	Private sector employee
10	1	Housewife	Retired

Elaboración propia.

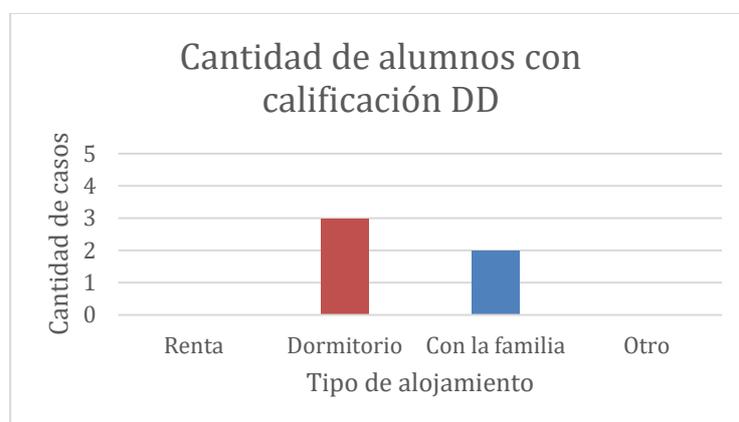
Con base en la tabulación anterior, se elaboraron las gráficas 11, 12, 13, 14, 15, 16 y 17 que reflejan visualmente la situación para cada variable.



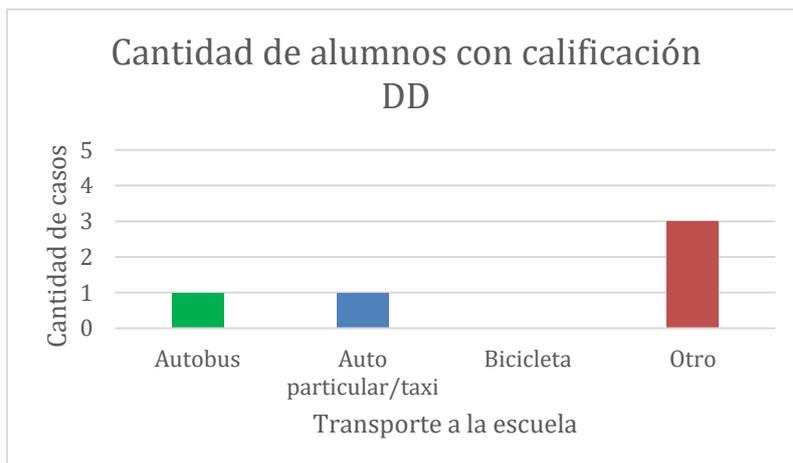
Gráfica 11. Calificación DD obtenida del modelo 8 respecto si el estudiante tiene empleo. Elaboración propia.



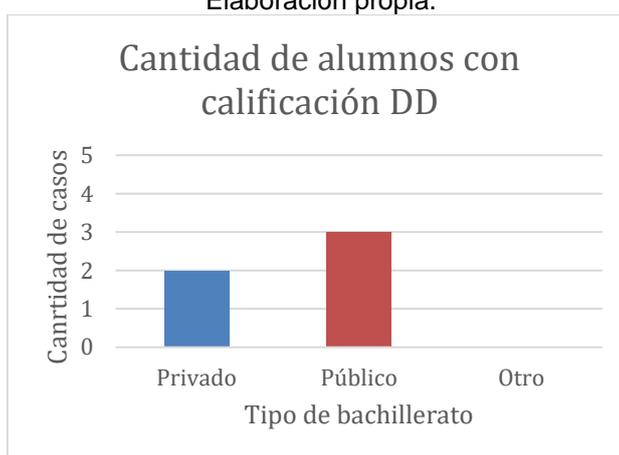
Gráfica 12. Calificación DD obtenida del modelo 8 respecto al salario del estudiante. Elaboración propia.



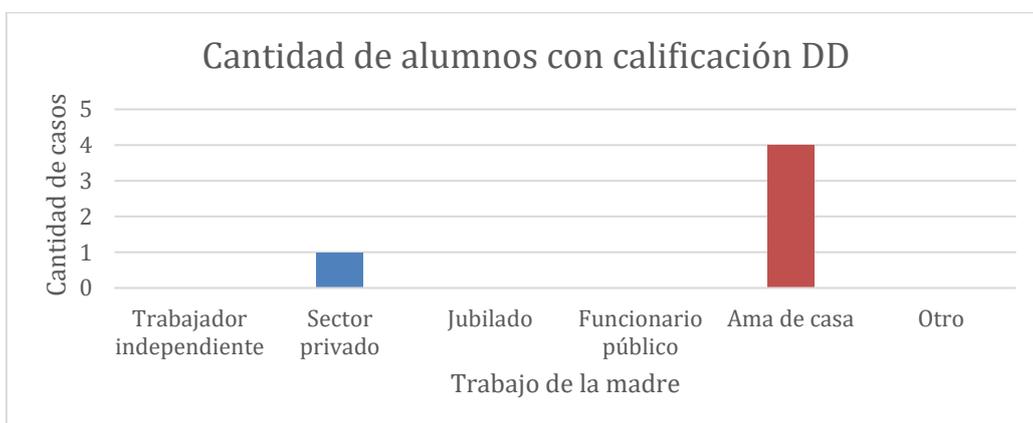
Gráfica 13. Calificación DD obtenida del modelo 8 respecto al tipo de alojamiento. Elaboración propia.



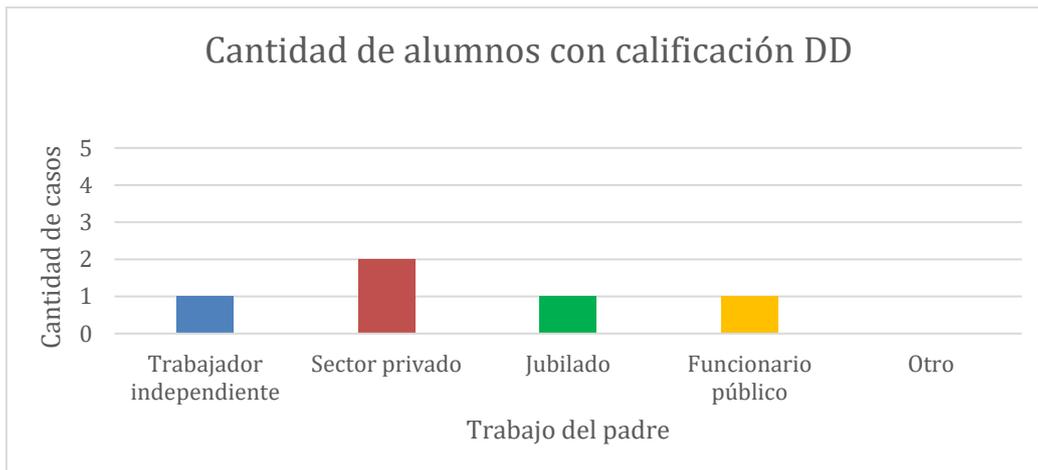
Gráfica 14. Calificación DD obtenida del modelo 8 respecto al tipo de transporte. Elaboración propia.



Gráfica 15. Calificación DD obtenida del modelo 8 respecto al tipo de bachillerato. Elaboración propia.



Gráfica 16. Calificación DD obtenida del modelo 8 respecto al trabajo de la madre. Elaboración propia.



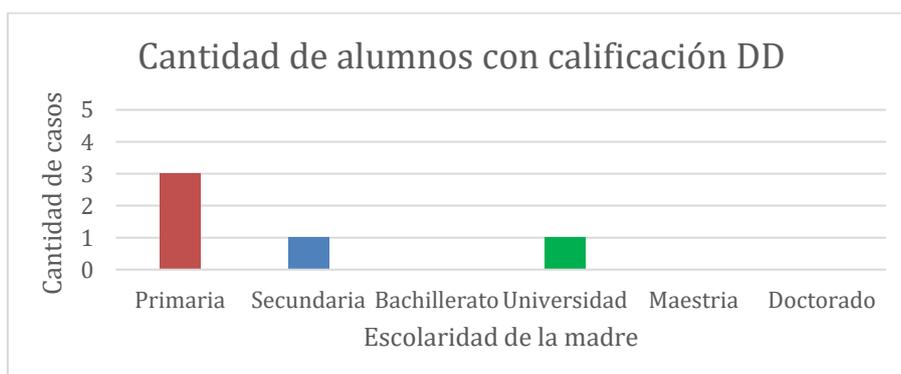
Gráfica 17. Calificación DD obtenida del modelo 8 respecto al trabajo del padre. Elaboración propia.

De igual forma, se analizaron los datos respecto a la escolaridad de los padres. A continuación se muestra la tabla 22 y las gráficas 18 y 19 que ilustran el comportamiento.

Tabla 22. Análisis del modelo 8 respecto a la escolaridad de los padres.

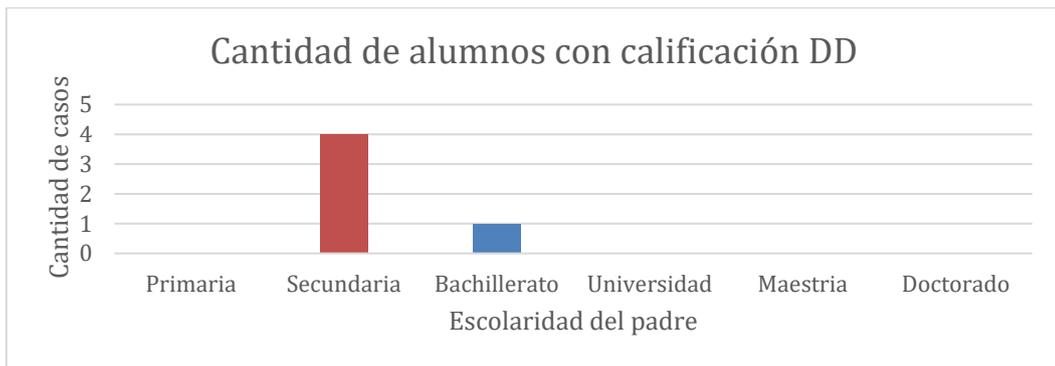
Caso	Predicción	Mother_Edu	Father_Edu
1	7	High school	University
2	1	Secondary school	Secondary school
3	1	Primary school	Secondary school
4	7	High school	High school
5	7	High school	High school
6	7	Primary school	Primary school
7	1	University	High school
8	7	Secondary school	University
9	1	Primary school	Secondary school
10	1	Primary school	Secondary school

Elaboración propia



Gráfica 18. Calificación DD obtenida del modelo 8 respecto a la escolaridad de la madre.

Elaboración propia.



Gráfica 19. Calificación DD obtenida del modelo 8 respecto a la escolaridad del padre. Elaboración propia.

Retomando las indicaciones detectadas que involucran variables sensibles no utilizadas en el modelo, se presentan las tablas 23, 24 y 25 que reflejan discriminación en los resultados, así como las gráficas 20, 21 y 22.

Tabla 23. Análisis del modelo 8 respecto a la nacionalidad.

Caso	Predicción	Nacionalidad
1	7	Argentina
2	1	Mexicana
3	1	Mexicana
4	7	Argentina
5	7	Chilena
6	7	Chilena
7	1	Mexicana
8	7	Argentina
9	1	Mexicana
10	1	Mexicana

Elaboración propia.

Tabla 24. Análisis del modelo 8 respecto a la raza.

Caso	Predicción	Raza
1	7	Mestiza
2	1	Mestiza
3	1	Indígena
4	7	Mestiza
5	7	Mestiza
6	7	Mestiza
7	1	Indígena
8	7	Mestiza
9	1	Indígena
10	1	Indígena

Elaboración propia.

Tabla 25. Análisis del modelo 8 respecto a los antecedentes migratorios.

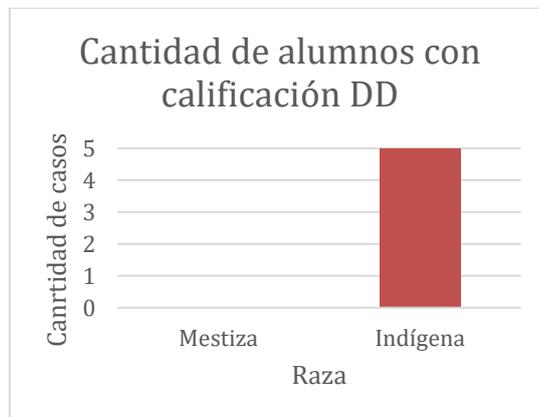
Caso	Predicción	Antecedentes Migratorios
1	7	Ninguno
2	1	Movilidad estudiantil temporal
3	1	Movilidad estudiantil temporal
4	7	Ninguno
5	7	Movilidad estudiantil temporal

6	7	Movilidad estudiantil temporal
7	1	Movilidad estudiantil temporal
8	7	Ninguno
9	1	Movilidad estudiantil temporal
10	1	Movilidad estudiantil temporal

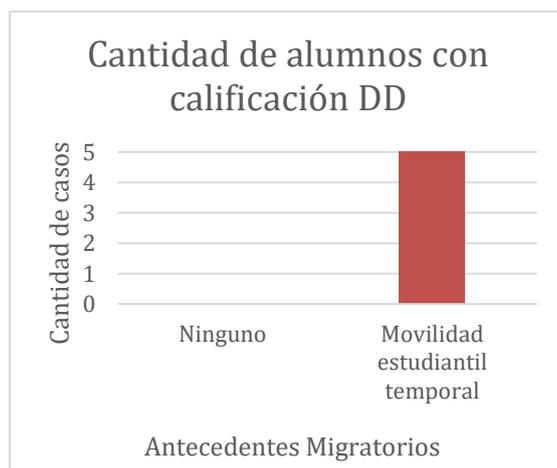
Elaboración propia.



Gráfica 20. Calificación DD obtenida del modelo 8 respecto a la nacionalidad.
Elaboración propia.



Gráfica 21. Calificación DD obtenida del modelo 8 respecto a la raza.
Elaboración propia.



Gráfica 22. Calificación DD obtenida del modelo 8 respecto a los antecedentes migratorios. Elaboración propia.

Como es posible observar, de los cinco estudiantes que recibieron la peor calificación por parte del modelo, cuatro son de género femenino; tres provienen de un bachillerato público y dos de uno privado; todos trabajan; dos de ellos reciben un salario entre \$135-200 y tres de ellos entre \$201-270; uno de ellos utiliza autobús para llegar a la escuela, uno emplea auto particular o taxi y tres utilizan otra forma para transportarse; tres viven en dormitorios y sólo dos con algún familiar; las madres de cuatro de ellos son amas de casa y sólo una se dedica al sector privado; dos de los padres también se dedican al sector privado y el los otros tres responden a ser trabajador independiente, jubilado o funcionario público respectivamente; en el caso de la escolaridad de la madre, en tres estudiantes se tiene el dato de tener nivel primaria, una con nivel secundaria y una con nivel superior, por otra parte, cuatro de los padres tienen escolaridad secundaria y sólo uno tiene bachillerato.

Por otra parte, respecto a las características de los estudiantes como la nacionalidad, raza y antecedentes migratorios, se observan datos alarmantes, ya que los cinco pertenecen a la nacionalidad mexicana, cuatro de ellos son de raza indígena y los cinco están bajo movilidad estudiantil temporal. Esta síntesis puede visualizarse en las tablas 26 y 27.

Tabla 26. Calificaciones DD del modelo 8 respecto a las variables sensibles de entrada.

Dato de entrada	Valor del dato	Cantidad de reprobados
Género	Femenino	4

Tipo de bachillerato	Público	3
	Privado	2
Trabajo adicional	Si	5
Salario	\$135-200	2
	\$201-270	3
Transporte a la escuela	Autobús	1
	Auto particular/taxi	1
	Otro	3
Tipo de alojamiento	Dormitorio	3
	Con la familia	2
Trabajo de la madre	Sector privado	1
	Ama de casa	4
Trabajo del padre	Trabajador independiente	1
	Sector privado	2
	Jubilado	1
	Funcionario Público	1
Escolaridad de la madre	Primaria	3
	Secundaria	1
	Universidad	1
Escolaridad del padre	Secundaria	4
	Bachillerato	1

Elaboración propia.

Tabla 27. Calificaciones DD del modelo 8 respecto al resto de variables sensibles.

Dato identificado	Valor del dato	Cantidad de reprobados
Nacionalidad	Mexicana	5
Raza	Indígena	5
Antecedentes migratorios	Movilidad estudiantil temporal	5

Elaboración propia.

Haciendo uso de los resultados obtenidos en el análisis de las categorías, se respondieron los ítems de la escala de Likert para determinar el nivel de sesgo ético. Respecto a la primera dimensión en la que se aborda el beneficio a individuos de diferentes grupos discriminados, se responde conforme lo localizado en las categorías protegidas correspondientes. En el resto de las variables sensibles al no estar incluidas en el modelo ni especificadas como parte de las características relevantes en el texto, se consideran elementos que no reflejan discriminación. En las imágenes 29, 30, 31, 32, 33 y 34 se observan las respuestas.

Métrica cualitativa para sesgo ético

Conforme los resultados obtenidos por el modelo de Machine Learning aplicado en el área educativa, responde las oraciones presentadas a continuación. Como resultado, se obtendrá el nivel de sesgo ético identificado: Muy Bajo, Bajo, Medio, Alto o Muy alto

A) Reparar la desventaja
Los resultados obtenidos no benefician a individuos pertenecientes a diferentes grupos que han sido discriminados históricamente según su:

1. Género/Sexo

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

2. Orientación sexual.

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

3. Religión

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

4. Opinión política

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

Ilustración 29. Primera parte de la contestación de los ítems para el modelo 8. Elaboración propia.

Métrica cualitativa para sesgo ético

5. Nacionalidad

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

6. Origen étnico

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

7. Raza

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

8. Idioma

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

9. Discapacidad

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

Ilustración 30. Segunda parte de la contestación de los ítems para el modelo 8. Elaboración propia.

Métrica cualitativa para sesgo ético

10. Nivel socioeconómico

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

11. Escolaridad de los padres

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

12. Antecedentes migratorios

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

Ilustración 31. Tercera parte de la contestación de los ítems para el modelo 8. Elaboración propia.

Métrica cualitativa para sesgo ético

B) Corregir el estigma, los estereotipos y la humillación. Los resultados obtenidos reflejan una preferencia negativa hacia alguna de las características del estudiante.

13. **Género/Sexo**

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

14. **Orientación sexual.**

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

15. **Religión**

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

16. **Opinión política**

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

17. **Nacionalidad**

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

Ilustración 32. Tercera parte del contestación de los ítems para el modelo 8. Elaboración propia.

Métrica cualitativa para sesgo ético

18. **Origen étnico**

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

19. **Raza**

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

20. **Idioma**

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

21. **Discapacidad**

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

Ilustración 33. Cuarta parte de la contestación de los ítems para el modelo 8. Elaboración propia.

Métrica cualitativa para sesgo ético

22. **Nivel socioeconómico**

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

23. **Escolaridad de los padres**

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

24. **Antecedentes migratorios**

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

Ilustración 34. Quinta parte de la contestación de los ítems para el modelo 8. Elaboración propia.

Evaluando la tercera dimensión, se identifica que al hablar sobre género en los datos de entrada, no es inclusivo solamente abordar dos de ellos relacionándolos con el sexo, ya que se excluye a personas que no están identificadas con ninguno; aunado a ello, no hay inclusión en los resultados benéficos respecto a una de las nacionalidades, raza y antecedentes migratorios en los estudiantes. Por otra parte, no se excluye a estudiantes de diferente orientación sexual, religión u opinión política, origen étnico, idioma o

con alguna discapacidad debido a que no es estipula ni se observa ninguna limitante. Las respuestas se muestran en las ilustraciones 35, 36 y 37.

Métrica cualitativa para sesgo ético

C) Dimensión participativa.
Los resultados obtenidos no benefician a estudiantes de diferentes grupos por las siguientes categorías:

25. Género/Sexo

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

26. Orientación sexual.

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

27. Religión

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

28. Opinión política

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

29. Nacionalidad

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

Ilustración 35. Quinta parte de la contestación de los ítems para el modelo 8. Elaboración propia.

Métrica cualitativa para sesgo ético

30. Origen étnico

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

31. Raza

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

32. Idioma

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

33. Discapacidad

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

34. Nivel socioeconómico

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

Ilustración 36. Sexta parte de la contestación de los ítems para el modelo 8.

Elaboración propia.

Métrica cualitativa para sesgo ético

35. Escolaridad de los padres

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

36. Antecedentes migratorios

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo
- Muy de acuerdo

Ilustración 37. Séptima parte de la contestación de los ítems para el modelo 8.

Elaboración propia.

Finalmente, en la cuarta dimensión orientada al contexto, se tiene que el modelo será implementado para predecir deserción escolar con base en la calificación obtenida, no obstante, difiere del objetivo original del modelo. De igual forma corresponde al año 2019 y según los supuestos descritos, al implementarse en otra zona geográfica y socioculturalmente distinta de donde se aplicó y se entrenó, esto puede ocasionar sesgo. Los ítems correspondientes se contestaron como se observa en la imagen 38.

Métrica cualitativa para sesgo ético

D) Evaluación del contexto de implementación

37. Los resultados del modelo se implementan para resolver un problema diferente al que es su finalidad solucionar.

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

38. La época o contexto temporal tiene características diferentes a la época en la que se reunieron los datos con los que el modelo fue entrenado.

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

39. El estado, nación o contexto geográfico que tiene características diferentes al lugar donde se planeó implementarlo cuando fue creado.

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

40. El ambiente socio-cultural que tiene características diferentes al entorno donde se planeó implementar.

Muy en desacuerdo
 En desacuerdo
 Ligeramente en desacuerdo
 Ligeramente de acuerdo
 De acuerdo
 Muy de acuerdo

Reiniciar Cuestionario
Calcular Nivel de Sesgo Ético

Ilustración 38. Parte final de la contestación de los ítems para el modelo 8.
Elaboración propia.

El programa arroja de forma automática la sumatoria de la puntuación, la cual equivale a 135, y acorde a los baremos, corresponde a un nivel de sesgo ético medio. Este resultado se visualiza en la ilustración 39.

Resultado

Puntuación total: 135
El nivel de sesgo ético es medio.

Aceptar

Ilustración 39. Nivel de sesgo obtenido en el modelo 8.
Elaboración propia.

En el apartado de anexos H, I y J, se encuentra el procedimiento y evaluación de los modelos restantes del proyecto

Discusión de resultados

Al implementar la métrica propuesta en los modelos seleccionados, se obtuvieron niveles de sesgo equivalentes a la categoría “Muy Bajo” en los primeros seis modelos pertenecientes al primer proyecto explicado; “Bajo” en el modelo 7 y 10; y “Medio” en el modelo 8 y 9.

Aunado a ello, sobresale la relación con el sesgo y las variables sensibles como el género y nivel socioeconómico en todas las evaluaciones realizadas. No obstante, al agregar supuestos respecto a la nacionalidad, la raza y los antecedentes migratorios para la evaluación de los modelos 7, 8, 9 y 10, se observaron variaciones en la puntuación final obtenida, lo cual muestra que entre mayor cantidad de variables sensibles se vean involucradas en un trato discriminatorio, más alto será el indicador de sesgo.

Al analizar los resultados de cada modelo, se observan variaciones respecto al algoritmo aplicado. En los seis primeros modelos, las diferencias son mínimas, puesto que en el modelo 2, se analiza otro dato reprobatorio arrojado como resultado. Sin embargo entre los modelos 7, 8, 9 y 10 existen cambios significativos entre las calificaciones obtenidas.

Referente a la pregunta de investigación, con los resultados obtenidos se observa que el uso de variables sensibles relacionadas con el sesgo ético en aprendizaje automático aplicado a la educación, así como la propuesta de una escala de Likert que muestra los ítems correspondientes al marco de equidad sustantiva y un espacio de resultados que refleja el nivel de acuerdo o de desacuerdo con cada uno, permiten determinar el nivel de presencia de sesgo desde una perspectiva cualitativa, diferente a las métricas ya existentes.

Con los presentes resultados se podría determinar si los modelos analizados deberían o no, implementarse en un contexto educativo a gran escala, puesto que, aunque se observa en cada proyecto que su objetivo era

académico, denotan sesgo en diversas variables sensibles que ocasionarían problemas hacia la comunidad estudiantil. Al incluir supuestos respecto a otras variables en los modelos 7, 8, 9 y 10, observando cambios significativos en el nivel de sesgo, resulta importante analizar la información de cada caso referente a las variables manejadas por la métrica, independiente a su uso como de los datos de entrada. Así también, se muestra importante utilizar los algoritmos de aprendizaje automático adecuados para cada objetivo, con la finalidad de determinar la influencia de ello en resultados errados.

El presente estudio, tiene algunas limitaciones al haber empleado la noción de equidad sustantiva, puesto que para otros marcos propuestos, sería necesario realizar un ajuste en los ítems y dimensiones de la escala de Likert propuesta; por otra parte, el conjunto de datos empleado para la realización de las pruebas, parte de supuestos establecidos con el objetivo de visualizar una posible discriminación hacia determinado sector; finalmente, ésta métrica se centra en detectar el sesgo ético en la fase de post procesamiento, por lo cual, su uso no involucra determinar cuáles son las causas de su aparición.

Conclusiones

Los modelos encontrados y utilizados no hacen uso de librerías que permiten implementar métricas cuantitativas de sesgo como AI Fairness 360; aunado a ello, es notable que dentro de los códigos se abordaba la precisión en las predicciones, por lo cual se observa lo mencionado dentro de la literatura: la lamentable despreocupación por el aspecto ético y los problemas que conlleva el dejar de lado dicho factor. Si bien, se percibe que los proyectos eran a pequeña escala, la ética debe formar parte de los desarrolladores de aprendizaje automático desde su formación

En la búsqueda de modelos, fue notable que los proyectos disponibles van enfocados en su mayoría a la predicción de calificaciones, sin embargo, se requiere una búsqueda exhaustiva para determinar cuál es el objetivo principal que abordan la mayoría de las propuestas dentro de la literatura.

A lo largo de la fase de pruebas, resaltó la importancia de tomar en cuenta si los estudiantes están siendo discriminados por alguna característica que no ingresa directamente en los datos de entrada del modelo ni se observa una relación con proxies, pues esto permite obtener el nivel de sesgo aproximado apegándose a los resultados obtenidos, ya que la fuente del sesgo puede provenir de diferentes elementos. Con ello, se comprueba la relevancia del contexto que rodea al estudiante y su situación respecto a la evaluación de sesgo.

Finalmente, se recomienda que las personas encargadas de la toma de decisiones tomen en cuenta las predicciones de los modelos de aprendizaje automático como una sugerencia, puesto que cada caso estudiantil es diferente y, si bien existen alumnos con resultados reprobatorios que tienen antecedentes de bajo rendimiento académico, es importante analizar si existen los elementos suficientes para sustentar una calificación reprobatoria para la evaluación del proceso en curso y la influencia del contexto del estudiante abordado por la métrica.

Trabajo Futuro

Como líneas de trabajo a futuro se tienen varias opciones. La primera de ellas va enfocada a la inclusión y evaluación de más variables sensibles o categorías protegidas que tengan el sustento teórico y puedan verse involucradas en la discriminación del aprendizaje automático aplicado a la educación.

Por otra parte, se propone observar la implementación de la propuesta tras el uso de librerías dentro de los códigos de los modelos construidos, aplicándola de forma complementaria. Si bien, la métrica propuesta no pretende sustituir a las nociones formales ya existentes, se sugiere evaluar su eficiencia con respecto a métricas cuantitativas expuestas en el apartado del marco teórico. Finalmente se espera aplicar la métrica a proyectos de gran

escala para contribuir a evitar la discriminación dentro del área educativa y tecnológica.

V. REFERENCIAS

- Agarwal, A., Agarwal, H., y Agarwal, N. (2023). Fairness Score and process standardization: framework for fairness certification in artificial intelligence systems. *AI and Ethics*, 3(1), 267-279. <https://doi.org/10.1007/s43681-022-00147-7>
- Akgun, S., y Greenhow, C. (2022). Artificial Intelligence in education Addressing ethical challenges in K-12 settings. *AI and Ethics*, 2, 431-440. <https://doi.org/10.1007/s43681-021-00096-7>
- Akmeşe, O. F., Kör, H., y Erbay, H. (2021). Use of aprendizaje automático techniques for the forecast of student achievement in higher education. . *Information Technologies and Learning Tools*, 82(2), 297-311. <https://doi.org/10.33407/itlt.v82i2.4178>
- Angwin, J., Larson, J., Mattu, S., y Kirchner, L. (2016). Machine Bias: there's software used across the country to predict future criminals. And it's biased against blacks. ProPublica. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- Ashokan, A., y Haas, C. (2021). Fairness metrics and bias mitigation strategies for rating predictions. *Information Processing and Management*. 58(5), 102646-102663. <https://doi.org/10.1016/j.ipm.2021.102646>
- Babic, I. D. (2018). Aprendizaje automático methods in predicting the student academic motivation. *Croatian Operational Research Review*, 8(2), 443-461. <https://doi.org/10.17535/crorr.2017.0028>
- Baker, R. S. y Hawn, A. (2021). Algorithmic Bias in Education. *International Journal of Artificial Intelligence in Education*. 1-14. <https://doi.org/10.1007/s40593-021-00285-9>
- Bogina, V., Hartman, A., Kuflik, T., y Shulner-Tal, A. (2021). Educating Software and AI Stakeholders About Algorithmic Fairness, Accountability, Transparency and Ethics. *International Journal of Artificial Intelligence in Education*. 1-26 <https://doi.org/10.1007/s40593-021-00248-0>
- Cantero, H., (2021). Justicia Algorítmica: El Sesgo de los Algoritmos. [Memoria Proyecto Fin de Grado, Universidad Politécnica de Madrid]. <https://oa.upm.es/69093/>
- Cardona, T., Cudney, E. A., Hoerl, R., y Snyder, J. (2020), Data Mining and Aprendizaje automático Retention Models in Higher Education. *Journal of College. Student Retention-Research Theory & Practice*. <https://doi.org/10.1177%2F1521025120964920>
- Catania, B., Guerrini, G., y Accinelli, C. (2023). Fairness & friends in the data science era. *AI & SOCIETY*, 38(2), 721-731.
- Caton, S., y Haas, C. (2020). Fairness in Aprendizaje automático: A Survey. arXiv. <https://arxiv.org/pdf/2010.04053.pdf>
- Celis, L. E., Keswani, V., y Vishnoi, N. K. (2020). Data preprocessing to mitigate bias: A maximum entropy based approach. *International Conference on Aprendizaje automático*, 1349-1359. <http://proceedings.mlr.press/v119/celis20a/celis20a.pdf>

- Chaudhry, M., y Kazim, E. (2021). Artificial Intelligence in Education (AIEd): a high-level academic and industry note 2021. *AI and Ethics*, 2, 157-165. <https://doi.org/10.1007/s43681-021-00074-z>
- Chen, I. Y., Pierson, E., Rose, S., Joshi, S., Ferryman, K., y Ghassemi, M. (2021). Ethical Aprendizaje automático in Healthcare. *Annual Review of Biomedical Data Science*, 4, 123-144. <https://doi.org/10.1146/annurev-biodatasci-092820-114757>
- Coronado, J. (2007). Escalas de Medición. *Paradigmas*, 2(2). 104-125
- Deena, G., y Raja, K. (2019). Designing an Automated Intelligent e-Learning System to Enhance the Knowledge using Aprendizaje automático Techniques. *International Journal of Advanced Computer Science and Applications*, 10(12), 112-119. <https://doi.org/10.14569/IJACSA.2019.0101215>
- Deho, O., Zhan, C., Li, J., Liu, J., Liu, L., y Duy Le, T. (2022). How do the existing fairness metrics and unfairness mitigation algorithms contribute to ethical learning analytics?. *British Journal of Educational Technology*, 53(4), 822-843. <https://doi.org/10.1111/bjet.13217>
- Dieterle, E., Dede, C., y Walker, M. (2022). The cyclical ethical effects of using artificial intelligence in education. *AI & SOCIETY*. <https://doi.org/10.1007/s00146-022-01497-w>
- Du Boulay, B. (2016). Artificial intelligence as an effective classroom assistant. *IEEE Intelligent Systems*, 31(6), 76-81. <https://doi.org/10.1109/MIS.2016.93>
- Echeverría, J. (2010). Tecnociencia, tecnoética y tecnoaxiología. *Revista Colombiana de Bioética*, 5(1), 142-152. <https://www.redalyc.org/pdf/1892/189217244009.pdf>
- El Naqa, I., Li, R. y Murphy, M. J. (Eds.). (2015). *Aprendizaje automático in radiation oncology: theory and applications*. Switzerland: Springer. <https://link.springer.com/book/10.1007/978-3-319-18305-3>
- Fazelpour, S., y Danks, D. (2021). Algorithmic bias: Senses, sources, solutions. *Philosophy Compass*, 16(8), <https://doi.org/10.1111/phc3.12760>.
- Fredman, S. (2016). Substantive equality revisited. *International Journal of Constitutional Law*, 14(3), 712-738. <https://doi.org/10.1093/icon/mow043>
- Gardner, J., Brooks, C., y Baker, R. (2019). Evaluating the fairness of predictive student models through slicing analysis. In *Proceedings of the 9th international conference on learning analytics & knowledge* (pp. 225-234). <https://doi.org/10.1145/3303772.3303791>
- Garcia, M. (2016). Racist in the Machine: The Disturbing Implications of Algorithmic Bias. *World Policy Journal*, 33(4), 111-117. <https://doi.org/10.1215/07402775-3813015>
- Garnham, A. (2017). *Artificial intelligence: An introduction*. Routledge. https://www.google.com.mx/books/edition/Artificial_Intelligence/qF18DwAAQBAJ?hl=es-419&gbpv=0
- Gaskins, N. (2022). Interrogating Algorithmic Bias: From Speculative Fiction to Liberatory Design. *TechTrends* 67, 417-425. <https://doi.org/10.1007/s11528-022-00783-0>
- Gearhart, D. (2009). Technoethics in education for the twenty-first century. En R. Luppigini & R. Adell (Eds.), *Handbook of research on technoethics*. IGI Global. <https://doi.org/10.4018/978-1-60566-022-6.ch018>
- Giovanola, B., y Tiribelli, S. (2022). Weapons of moral construction? On the value of fairness in algorithmic decision-making. *Ethics and Information Technology*, 24(1), 3.
- Goethals, S., Martens, D., y Calders, T. (2023). PreCoF: counterfactual explanations for fairness. *Aprendizaje automático*, 3111-3142 <https://doi.org/10.1007/s10994-023-06319-8>

- Grote, T. (2024). Fairness as adequacy: a sociotechnical view on model evaluation in aprendizaje automático. *AI and Ethics*, 4(2), 427-440. <https://doi.org/10.1007/s43681-023-00280-x>
- Guil, M. (2006). Escala mixta likert-thurstone. *Anduli*, 5, 81-95. <https://idus.us.es/handle/11441/50616>
- Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines*, 30(1), 99-120. <https://doi.org/10.1007/s11023-020-09517-8>
- Hasan, A., Brown, S., Davidovic, J., Lange, B., y Regan, M. (2022). Algorithmic bias and risk assessments: Lessons from practice. *Digital Society*, 1(2), 14. <https://doi.org/10.1007/s44206-022-00017-z>
- Hashimoto, D. A., Rosman, G., Rus, D., & Meireles, O. R. (2018). Artificial intelligence in surgery: promises and perils. *Annals of surgery*, 268(1), 70-76.
- Helm, J. M., Swiergosz, A. M., Haeberle, H. S., Karnuta, J. M., Schaffer, J. L., Krebs, V. E., Spitzer, A.I., y Ramkumar, P. N. (2020). Aprendizaje automático and artificial intelligence: definitions, applications, and future directions. *Current reviews in musculoskeletal medicine*, 13(1), 69-76. <https://doi.org/10.1007/s12178-020-09600-8>
- Henrique, B. M., Bobreiro, V. A., Kimura, H. (2019). Literature review: Aprendizaje automático techniques applied to financial market prediction. *Expert Systems with Applications*, 124, 226-251. <https://doi.org/10.1016/j.eswa.2019.01.012>
- Henry, J., y Oliver, M. (2022). Who Will Watch the Watchmen? The Ethico-political Arrangements of Algorithmic Proctoring for Academic Integrity. *Postdigital Science and Education*, 4, 330-353. <https://doi.org/10.1007/s42438-021-00273-1>
- Herrmann, D. S. (2007). Complete guide to security and privacy metrics: measuring regulatory compliance, operational resilience, and ROI. Auerbach Publications. https://www.google.com.mx/books/edition/Complete_Guide_to_Security_and_Privacy_M/y5sFSdDFKPcC?hl=es-419&gbpv=1&dq=QUALITATIVE+METRICS+and+scales&pg=PA32&printsec=frontcover
- Hernandez, C. (2023). (Some) algorithmic bias as institutional bias. *Ethics and Information Technology*, 25(24). <https://doi.org/10.1007/s10676-023-09698-7>
- Hilbert, S., Coors, S., Kraus, E., Bischl, B., Lindl, A., Frei, M., Wild, J., Krauss, S., Goretzko, D., y Stachl, C. (2021) Aprendizaje automático for the educational sciences. *Review of Education*, 9(3). <https://doi.org/10.1002/rev3.3310>
- Hussain, M., Zhu, W., Zhang, W., Raza, A. S. M., y Ali, S. (2018). Using aprendizaje automático to predict student difficulties from learning session data. *Artificial Intelligence Review*, 52(1), 381-407. <https://doi.org/10.1007/s10462-018-9620-8>
- Hwang, G. J., Xie, H., Wah, B. W., y Gašević, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, 1, 100001. <https://www.sciencedirect.com/science/article/pii/S2666920X20300011>
- Infante, R. (2021). Rendición de cuentas y políticas públicas aplicadas a la inteligencia artificial en América Latina.
- Jebb AT., Ng V., y Tay, L. (2021). A Review of Key Likert Scale Development Advances: 1995–2019. *Front. Psychol.* 12, 637547. <https://doi.org/10.3389/fpsyg.2021.637547>
- Jeong, H., Wu, M., Dasgupta, N., Médard, M., y Calmon, F. (2022). Who Gets the Benefit of the Doubt? Racial Bias in Aprendizaje automático Algorithms Applied to Secondary School Math Education. 35th Conference on Neural Information Processing Systems (NeurIPS 2021) Workshop on Math AI for Education (MATHAI4ED): Bridging the Gap Between Research and Smart Education.

- Joshi, A., Kale, S., Chandel, S., y Pal, D. (2015). Likert scale: Explored and explained. *British journal of applied science & technology*, 7(4), 396-403. <https://doi.org/10.9734/BJAST/2015/14975>
- Joshi, S., Rambola, R. K., y Churi, P. (2021). Evaluating artificial intelligence in education for next generation. En *Journal of Physics: Conference Series* (Vol. 1714, No. 1, p. 012039). IOP Publishing. <https://iopscience.iop.org/article/10.1088/1742-6596/1714/1/012039>
- Kamiran, F., y Calders, T. (2011). Data preprocessing techniques for classification without discrimination. *Knowledge and information systems*, 33(1), 1-33. <https://doi.org/10.1007/s10115-011-0463-8>
- Kashi, A., Shastri, S., y Deshpande, A. R. (2016). A Score Recommendation System Towards Automating Assessment In Professional Courses. 2016 IEEE 8th International Conference on Technology for Education. 140-143. <https://doi.org/10.1109/T4E.2016.036>
- Kaplan, J. (2016). *Artificial intelligence: What everyone needs to know*. Oxford University Press. <https://books.google.es/books?hl=es&lr=&id=wPvmDAAAQBAJ&oi=fnd&pg=PP1&dq=what+is+the+artificial+intelligence&ots=NylHOBtqLe&sig=BpQldqB5NkiyixRaWtFd2j4Vn8#v=onepage&q=what%20is%20the%20artificial%20intelligence&f=false>
- Khanal, S. S., Prasad, P. W. C., Alsadoon, A., y Maag, A. (2020). A systematic review: aprendizaje automático based recommendation systems for e-learning. *Education and Information Technologies*, 25(4), 2635-2664. <https://doi.org/10.1007/s10639-019-10063-9>
- Kizilcec, R., y Lee, H. (2021). Algorithmic fairness in education. *The ethics of artificial intelligence in education* (pp. 174-202). Routledge.
- Köchling, A., y Wehner, M. (2020). Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Business Research*, 13(3), 795-848. <https://doi.org/10.1007/s40685-020-00134-w>
- Kučak, D., Juričić, V., & Đambić, G. (2018). Aprendizaje automático in Education-A Survey of Current Research Trends. *Annals of DAAAM. Proceedings 29th DAAAM International Symposium on Intelligent Manufacturing And Automation*. 406-410. <https://doi.org/10.2507/29th.daaam.proceedings.059>
- Kunapuli, G. (2023). *Ensemble methods for aprendizaje automático*. Simon and Schuster.
- Lässig, N., Oppold, S., y Herschel, M. (2022). Metrics and algorithms for locally fair and accurate classifications using ensembles. *Datenbank-Spektrum*, 22(1), 23-43. <https://doi.org/10.1007/s13222-021-00401-y>
- Le Quy, T., Friege, G., y Ntoutsis, E. (2023). A review of clustering models in educational data science toward fairness-aware learning. *Educational data science: Essentials, approaches, and tendencies: Proactive education based on empirical big data evidence*, 43-94. https://doi.org/10.1007/978-981-99-0026-8_2
- Lee, N. T. (2018). Detecting racial bias in algorithms and aprendizaje automático. *Journal of Information, Communication and Ethics in Society*. 16(3), 252-260. <https://doi.org/10.1108/JICES-06-2018-0056>
- Lee, W. (2020). *Machine Lawyering—Watch those Hyperconnectivity Threats*. https://www.researchgate.net/profile/Wanbil-Lee/publication/342509744_Machine_Lawyering_-_Watch_those_Hyperconnectivity_Threats/links/5ef834bf92851c52d6040bf6/Machine-Lawyering-Watch-those-Hyperconnectivity-Threats.pdf

- Li, X., y Zhang, T. (2017). An Exploration on Artificial Intelligence Application: From Security, Privacy and Ethic Perspective. 2017 IEEE 2nd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), 416-420. IEEE. <https://doi.org/10.1109/ICCCBDA.2017.7951949>
- Lo, Y., Rensi, S.E., Torng, W., Altman, R.B. (2018). Aprendizaje automático in chemoinformatics and drug discovery. *Drugs discovery Today*. 23(8) 1538-1546. <https://doi.org/10.1016/j.drudis.2018.05.010>
- Lohia, P. (2021). Priority-based post-processing bias mitigation for individual and group fairness. arXiv. <https://arxiv.org/pdf/2010.04053.pdf>
- Lupiccini, R. (2009). The Emerging Field of Technoethics. En Lupiccini, R., y Adell, Rebecca. (Eds). *Handbook of Research on Technoethics*. Hershey: Idea Group Publishing: IGI Global. <https://doi.org/10.4018/978-1-60566-022-6.ch001>
- Madaio, M., Blodgett, S. L., Mayfield, E., y Dixon-Román, E. (2022). Beyond “fairness”: Structural (in) justice lenses on ai for education. In *The ethics of artificial intelligence in education* (pp. 203-239). Routledge.
- Mayfield, E., Madaio, M., Prabhumoye, S., Gerritsen, D., McLaughlin, B., Dixon-Román, E., y Black, A. W. (2019, August). Equity beyond bias in language technologies for education. In *Proceedings of the fourteenth workshop on innovative use of NLP for building educational applications* (pp. 444-460).
- Mehrabani, N., Morstatter, F., Saxena, N., Lerman, K., y Galstyan, A. (2021). A survey on bias and fairness in aprendizaje automático. *ACM computing surveys (CSUR)*, 54(6), 1-35. <https://doi.org/10.48550/arXiv.1908.09635>
- Memarian, B., y Doleck, T. (2023). Fairness, Accountability, Transparency, and Ethics (FATE) in Artificial Intelligence (AI) and higher education: A systematic review, *Computers and Education: Artificial Intelligence*, 5, 100152. <https://doi.org/10.1016/j.caeai.2023.100152>.
- Miller, S. (2019). Aprendizaje automático, Ethics and Law. *Australasian Journal of Information Systems*, 23. <https://doi.org/10.3127/ajis.v23i0.1893>
- Mörch, C. M., Atsu, S., Cai, W., Li, X., Madathil, S. A., Liu, X., Mai, V., Tamimi, F., Dilhac, M. A., y Ducret, M. (2021). Artificial Intelligence and Ethics in Dentistry: A Scoping Review. *Journal of Dental Research*, 100(13), 1452-1460. <https://doi.org/10.1177%2F00220345211013808>
- Mukhopadhyay, S. (2022). InfoGram and admissible aprendizaje automático. *Aprendizaje automático*, 111(1), 205-242. <https://doi.org/10.1007/s10994-021-06121-4>
- Ndukwe, I. G., Daniel, B. K., y Amadi, C. E. (2019). A aprendizaje automático grading system using chatbots. *International Conference on Artificial Intelligence in Education*. 365-368. https://doi.org/10.1007/978-3-030-23207-8_67
- Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., y Nguyen, B. P. T. (2023). Ethical principles for artificial intelligence in education. *Education and Information Technologies*, 28(4), 4221-4241. <https://doi.org/10.1007/s10639-022-11316-w>
- Nemoto, T., y Beglar, D. (2014). Likert-scale questionnaires. En *JALT 2013 conference proceedings* (pp. 1-8). https://jalt-publications.org/sites/default/files/pdf-article/jalt2013_001.pdf
- OCDE. (2013). Síntesis: Diez pasos hacia la equidad en la educación. <https://www.oecd.org/education/school/40043349.pdf>
- OCDE. (2023). Quiénes somos. OCDE. <https://www.oecd.org/acerca/>
- Omand, D., y Phythian, M. (2021). The technoethics of contemporary intelligence practice: A framework for analysis. En (Miller, S., Regan, M., y Walsh, P. (Eds). *National Security Intelligence and Ethics* (pp. 39-60). Routledge. <https://doi.org/10.4324/9781003164197-5>

- Opazo, D., Moreno, S., Álvarez-Miranda, E., y Pereira, J. (2021). Analysis of First-Year University Student Dropout through Aprendizaje automático Models: A Comparison between Universities. *Mathematics*, 9, 2599-2626. <https://doi.org/10.3390/math9202599>
- Padilla, J. C. (2007). Escalas de medición. *Paradigmas*, 2(2), 104-125.
- Pagano, T., Loureiro, R., Lisboa, F., Ramos, G., Matos, R., Aragão de Sousa, G., Lisboa, L., Matos, M., Cruz, M., Lopez, E., Winkler, I., y Sperandio, E. (2022). Bias and unfairness in aprendizaje automático models: a systematic literature review. *arXiv preprint* <https://doi.org/10.48550/arXiv.2202.08176>
- Palacios, C. Reyes-Suárez, J., Bearzotti, L., Leiva, V., y Marchant, C. (2021). Knowledge discovery for higher education student retention based on data mining: Aprendizaje automático algorithms and case study in Chile. *Entropy*, 23(4), 485. <https://doi.org/10.3390/e23040485>
- Patel, L., Shukla, T., Huang, X., Ussery, D. W., Wang, S. (2020). Aprendizaje automático Methods in Drug Discovery. *Molecules* 25(22)5277-5293. <https://doi.org/10.3390/molecules25225277>
- Pessach, D., y Shmueli, E. (2020). Algorithmic fairness. In *Aprendizaje automático for Data Science Handbook: Data Mining and Knowledge Discovery Handbook* (pp. 867-886). Cham: Springer International Publishing.
- Pfeiffer, J., Gutschow, J., Haas, C., Möslein, F., Maspfuhl, O., Borgers, F., y Alpsancar, S. (2023). Algorithmic Fairness in AI: An Interdisciplinary View. *Business & Information Systems Engineering*, 65(2), 209-222. <https://doi.org/10.1007/s12599-023-00787-x>
- Rajendran, S., Chamundeswari, S., Y Sinha, A. (2022). Predicting the academic performance of middle-and high-school students using aprendizaje automático algorithms. *Social Sciences & Humanities Open*, 6(1), 100357.
- Ramesh, D., y Sanampudi, S. (2022). An automated essay scoring systems: a systematic literature review. *Artificial Intelligence Review*, 55(3), 2495-2527. <https://doi.org/10.1007/s10462-021-10068-2>
- Rzepka, N., Simbeck, K., Müller, H. G., y Pinkwart, N. (2022). Fairness of In-session Dropout Prediction. In *CSEDU* (2), 316-326. <https://iug.htw-berlin.de/wp-content/uploads/2022/07/Fairness.pdf>
- Rouhiainen, L. (2018). *Inteligencia artificial*. Madrid: Alienta Editorial, 20-21. https://proassetspdlcom.cdnstatics2.com/usuarios/libros_contenido/arxius/40/39308_Inteligencia_artificial.pdf
- Sahlgren, O. (2021). The politics and reciprocal (re)configuration of accountability and fairness in data-driven education. *Learning, Media and Technology*, 48(1), 95-108. <https://doi.org/10.1080/17439884.2021.1986065>
- Salas-Pilco, S., Xiao, K., Oshima, J. (2022). Inteligencia artificial y nuevas tecnologías en la educación inclusiva para estudiantes minoritarios: una revisión sistemática. *Sostenibilidad*. 14, 13572. <https://doi.org/10.3390/su142013572>
- Saltz, J., Skirpan, M., Fiesler, C., Gorelick, M., Yeh, T., Heckman, R., Dewar, N., y Beard, N. (2019). Integrating ethics within aprendizaje automático courses. *ACM Transactions on Computing Education (TOCE)*, 19(4), 1-26. <https://doi.org/10.1145/3341164>
- Selwyn, N., Hillman, T., Bergviken-Rensfeldt, A., y Perrotta, C. (2023). Making sense of the digital automation of education. *Postdigital Science and Education*, 5(1), 1-14. <https://doi.org/10.1007/s42438-022-00350-z>
- Selwyn, N., Hillman, T., Bergviken-Rensfeldt, A., y Perrotta, C. (2023). Making sense of the digital automation of education. *Postdigital Science and Education*, 5(1), 1-14. <https://doi.org/10.1007/s42438-022-00350-z>

- Shailaja, K., Seetharamulu, B., y Jabbar, M. A. (2018) Aprendizaje automático in Healthcare: A Review. Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), 910-914. <https://doi.org/10.1109/ICECA.2018.8474918>
- Singhal, A., Tanveer, H., y Mago, V. (2023). Towards FATE in AI for Social Media and Healthcare: A Systematic Review. arXiv preprint arXiv:2306.05372. <https://doi.org/10.48550/arXiv.2306.05372>
- Skirpan, M., Beard, N., Bhaduri, S., Fiesler, C., y Yeh, T. (2018). Ethics Education in Context: A Case Study of Novel Ethics Activities for the CS Classroom. En Proceedings of the 49th ACM Technical Symposium on Computer Science Education, 940-945.
- Smith, H. (2020). Algorithmic bias: should students pay the price? AI & society. 35(4) 1077-1078. <https://doi.org/10.1007/s00146-020-01054-3>
- Sjödén, B. (2020). When Lying, Hiding and Deceiving Promotes Learning - A case for Augmented Intelligence with Augmented Ethics. En 21st International Conference, AIED 2020, Ifrane, Morocco, 291-295
- Sun, W., Nasraoui, O., Shafto, P., (2020). Evolution and impact of bias in human and aprendizaje automático algorithm interaction. Plos One, 15(8), <https://doi.org/10.1371/journal.pone.0235502>
- Suresh, H., y Gutttag, J. (2021). A framework for understanding sources of harm throughout the aprendizaje automático life cycle. In Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (pp. 1-9). <https://doi.org/10.1145/3465416.3483305>
- Tiwari, P. (2008). Information Technology And Library Evolution. APH Publishing. <https://bit.ly/3gQajls>
- Tandon, A., Salimath, N., Bhatia, S., Sethi, K., y Tripathy, B. (2019). Introduction to aprendizaje automático. Book Bazooka. https://www.google.com.mx/books/edition/Introduction_to_Machine_Learning/3YTA_DwAAQBAJ?hl=es-419&gbpv=0
- UNESCO (1962). Convención relativa a la Lucha contra las Discriminaciones en la Esfera de la Enseñanza. Adoptada el 14 de diciembre de 1960 en la 11ª reunión de la Conferencia General de la UNESCO. <https://www.unesco.org/es/legal-affairs/convention-against-discrimination-education>
- UNESCO (2023). La educación transforma vidas. UNESCO. <https://www.unesco.org/es/education>
- Vivas, W. J. (2018). Uso seguro y responsable de las TIC: una aproximación desde la tecnoética. Ciencia, docencia y tecnología, (57), 235-255. <https://dialnet.unirioja.es/servlet/articulo?codigo=6673279>
- Wang, P. (2019). On defining artificial intelligence. Journal of Artificial General Intelligence, 10(2), 1-37. <https://sciendo.com/downloadpdf/journals/jagi/10/2/article-p1.pdf>
- Webb, M., Fluck, A., Magenheimer, J., Malyn-Smith, J., Waters, J., Deschênes, M., y Zagami, J. (2020). Aprendizaje automático for human learners: opportunities, issues, tensions and threats. Educational Technology Research and Development, 69, 2109-2130. <https://doi.org/10.1007/s11423-020-09858-2>
- Williams, J. J., Kim, J., Rafferty, A., Maldonado, S., Gajos, K. Z., Lasecki, W. S., Heffernan, N. (2016). AXIS: Generating Explanations at Scale with Learnersourcing and Aprendizaje automático. Proceedings of the Third (2016) ACM Conference on Learning @ Scale. 379-388. <http://dx.doi.org/10.1145/2876034.2876042>
- Xu, W., y Ouyang, F. (2022). The application of AI in STEM education a systematic review from 2011 to 2021. International Journal of STEM Education, 9(59). <https://doi.org/10.1186/s40594-022-00377-5>

- Yu, R., Lee, H., y Kizilcec, R. (2021). Should college dropout prediction models include protected attributes?. In Proceedings of the eighth ACM conference on learning@ scale (pp. 91-100). <https://doi.org/10.1145/3430895.3460139>
- Yu, R., Li, Q., Fischer, C., Doroudi, S., Xu, D. (2020). Towards Accurate and Fair Prediction of College Success: Evaluating Different Sources of Student Data. Proceedings of The 13th International Conference on Educational Data Mining. 292- 301.
- Zhai, X., Krajcik, J., y Pellegrino, J. (2021). On the validity of aprendizaje automático-based Next Generation Science Assessments: A validity inferential network. *Journal of Science Education and Technology*, 30, 298-312. <https://doi.org/10.1007/s10956-020-09879-9>
- Zhai, X., Shi, L., Y Nehm, R. (2021). A meta-analysis of aprendizaje automático-based science assessments: Factors impacting machine-human score agreements. *Journal of Science Education and Technology*, 30, 361-379. <https://doi.org/10.1007/s10956-020-09875-z>
- Zhang, B. H., Lemoine, B., & Mitchell, M. (2018). Mitigating unwanted biases with adversarial learning. Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, 335-340. <https://doi.org/10.1145/3278721.3278779>
- Zhang, F., Xing, W., y Li, C. (2023). Predicting Students' Algebra I Performance using Reinforcement Learning with Multi-Group Fairness. In *LAK23: 13th International Learning Analytics and Knowledge Conference* (pp. 657-662). <https://doi.org/10.1145/3576050.3576104>
- Zook, M., Barocas, S., Boyd, D., Crawford, K., Keller, E., Gangadharan, S. P., Goodman, A., Hollander, R., Koenig, A. B., Metcalf, J., Narayanan, A., Nelson, A., y Pasquale, F. (2017). Ten simple rules for responsible big data research. *PLoS computational biology*, 13(3), <https://doi.org/10.1371/journal.pcbi.1005399>

VI. ANEXOS

Anexo A. Primera versión de la escala de Likert.

Conforme los resultados obtenidos por el modelo de Machine Learning aplicado en el área educativa, responde los siguientes cuestionamientos:

A) Reparar la desventaja

Los resultados obtenidos no benefician a individuos pertenecientes a diferentes grupos que han sido discriminados históricamente según su:

- 1.- Género/Sexo
- 2.- Orientación sexual.
- 3.- Religión
- 4.- Opinión política
- 5.- Nacionalidad
- 6.- Origen étnico
- 7.- Raza
- 8.- Idioma
- 9.- Discapacidad
- 10.- Nivel socioeconómico
- 11.- Escolaridad de los padres
- 12.- Antecedentes migratorios

**Ilustración 40. Primer borrador de la dimensión A en la métrica.
Elaboración propia**

B) Corregir el estigma, los estereotipos y la humillación.

Los resultados obtenidos reflejan una preferencia negativa hacia alguna de las siguientes características del estudiante:

- 1.- Género/Sexo
- 2.- Orientación sexual.
- 3.- Religión
- 4.- Opinión política
- 5.- Nacionalidad
- 6.- Origen étnico
- 7.- Raza
- 8.- Idioma
- 9.- Discapacidad
- 10.- Nivel socioeconómico
- 11.- Escolaridad de los padres
- 12.- Antecedentes migratorios

**Ilustración 41. Primer borrador de la dimensión B en la métrica.
Elaboración propia.**

C) Dimensión participativa: inclusión social y voz política. (Inclusión)

Los resultados obtenidos no benefician a estudiantes de diferentes grupos pertenecientes a las siguientes categorías:

- 1.- Género
- 2.- Orientación sexual.
- 3.- Religión
- 4.- Opinión política
- 5.- Nacionalidad
- 6.- Origen étnico
- 7.- Raza
- 8.- Idioma
- 9.- Discapacidad
- 10.- Nivel socioeconómico
- 11.- Escolaridad de los padres
- 12.- Antecedentes migratorios

**Ilustración 42. Primer borrador de la dimensión C en la métrica.
Elaboración propia.**

D) Evaluación del contexto de implementación

Los resultados obtenidos se implementan en:

- 1.- Un proceso de toma de decisiones análogo al que se considera propósito final.
- 2.- Una época (contexto temporal) con características similares a la que transcurría cuando se obtuvieron los datos con los que fue entrenado.
- 3.- Un estado o nación (contexto geográfico) con características análogas al sitio donde se consideró su uso original.
- 4.- Un ambiente socio-cultural con características análogas al entorno donde se planeó implementar.

**Ilustración 43. Primer borrador de la dimensión D en la métrica.
Elaboración propia.**

Anexo B. Interfaz para la escala de Likert.

Métrica cualitativa para sesgo ético

Conforme los resultados obtenidos por el modelo de Machine Learning aplicado en el área educativa, responde las oraciones presentadas a continuación. Como resultado, se obtendrá el nivel de sesgo ético identificado: Muy Bajo, Bajo, Medio, Alto o Muy alto

A) Reparar la desventaja
Los resultados obtenidos no benefician a individuos pertenecientes a diferentes grupos que han sido discriminados históricamente según su:

1. Género/ Sexo

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo

2. Orientación sexual.

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo

3. Religión

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo

Reiniciar Cuestionario

Calcular Nivel de Sesgo Ético

Ilustración 44. Primera parte de la interfaz creada.
Elaboración propia.

Métrica cualitativa para sesgo ético

B) Corregir el estigma, los estereotipos y la humillación.
Los resultados obtenidos reflejan una preferencia negativa hacia alguna de las siguientes características del estudiante.

8. Idioma

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo

9. Discapacidad

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo

10. Nivel socioeconómico

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo

11. Escolaridad de los padres

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo

12. Antecedentes migratorios

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo

Reiniciar Cuestionario

Calcular Nivel de Sesgo Ético

Ilustración 46. Quinta parte de la interfaz creada.
Elaboración propia.

Métrica cualitativa para sesgo ético

B) Corregir el estigma, los estereotipos y la humillación.
Los resultados obtenidos reflejan una preferencia negativa hacia alguna de las siguientes características del estudiante.

4. Opinión política

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo

5. Nacionalidad

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo

6. Origen étnico

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo

Reiniciar Cuestionario

Calcular Nivel de Sesgo Ético

Ilustración 45. Tercera parte de la interfaz creada.
Elaboración propia.

Métrica cualitativa para sesgo ético

B) Corregir el estigma, los estereotipos y la humillación.
Los resultados obtenidos reflejan una preferencia negativa hacia alguna de las siguientes características del estudiante.

13. Género/ Sexo

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo

14. Orientación sexual.

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo

15. Religión

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo

16. Opinión política

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- Ligeramente de acuerdo
- De acuerdo

Reiniciar Cuestionario

Calcular Nivel de Sesgo Ético

Ilustración 47. Segunda parte de la interfaz creada.
Elaboración propia.

Métrica cualitativa para seguro ético

17. Nacionalidad

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- De acuerdo

18. Origen étnico

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- De acuerdo

19. Raza

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- De acuerdo

20. Idioma

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- De acuerdo

21. Discapacidad

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- De acuerdo

Reiniciar Cuestionario

Calcular Nivel de Seguro Ético

Ilustración 48. Cuarta parte de la interfaz creada.
Elaboración propia.

Métrica cualitativa para seguro ético

29. Nacionalidad

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- De acuerdo

30. Origen étnico

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- De acuerdo

31. Raza

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- De acuerdo

32. Idioma

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- De acuerdo

Reiniciar Cuestionario

Calcular Nivel de Seguro Ético

Ilustración 51. Novena parte de la interfaz creada.
Elaboración propia.

Métrica cualitativa para seguro ético

21. Discapacidad

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- De acuerdo

22. Nivel socioeconómico

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- De acuerdo

23. Escolaridad de los padres

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- De acuerdo

24. Antecedentes migratorios

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- De acuerdo

Reiniciar Cuestionario

Calcular Nivel de Seguro Ético

Ilustración 49. Sexta parte de la interfaz creada.
Elaboración propia.

Métrica cualitativa para seguro ético

33. Discapacidad

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- De acuerdo

34. Nivel socioeconómico

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- De acuerdo

35. Escolaridad de los padres

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- De acuerdo

36. Antecedentes migratorios

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- De acuerdo

Reiniciar Cuestionario

Calcular Nivel de Seguro Ético

Ilustración 52. Décima parte de la interfaz creada.
Elaboración propia.

Métrica cualitativa para seguro ético

C) Dimensión participativa.
Los resultados obtenidos no benefician a estudiantes de diferentes grupos pertenecientes a las siguientes categorías:

25. Género/sexo

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- De acuerdo

26. Orientación sexual

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- De acuerdo

27. Religión

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- De acuerdo

28. Opinión política

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- De acuerdo

Reiniciar Cuestionario

Calcular Nivel de Seguro Ético

Ilustración 50. Octava parte de la interfaz creada.
Elaboración propia.

Métrica cualitativa para seguro ético

D) Evaluación del contexto de implementación

37. Los resultados del modelo se implementan para resolver un problema diferente al que es su finalidad solucionar.

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- De acuerdo

38. La época o contexto temporal tiene características diferentes a la época en la que se reunieron los datos con los que el modelo fue entrenado.

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- De acuerdo

39. El estado, nación o contexto geográfico que tiene características diferentes al lugar donde se planeó implementarlo cuando fue creado.

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- De acuerdo

40. El ambiente socio-cultural que tiene características diferentes al entorno donde se planeó implementar.

- Muy en desacuerdo
- En desacuerdo
- Ligeramente en desacuerdo
- De acuerdo

Reiniciar Cuestionario

Calcular Nivel de Seguro Ético

Ilustración 53. Parte final de la interfaz creada.
Elaboración propia.

Anexo C. Prueba de la métrica al modelo 2.

Las predicciones obtenidas se visualizan en la imagen 54.

Model11
 11.357728
 13.820428
 9.497592
 15.566088
 13.877666
 4.616734
 18.265286
 7.336174
 14.056926
 8.294861

Ilustración 54. Predicciones del modelo 2.
 Elaboración propia.

Las variables sensibles localizadas se aprecian en la ilustración 55. De igual forma se tiene la gráfica 23 acorde a las calificaciones de cada caso.

caso	school	sex	address	Medu	Fedu	Mjob	Fjob	Predicción
1	GP	F	U	5to a 9no grado	Primaria	services	other	11.35773
2	GP	M	U	Superior	Superior	health	services	13.82043
3	GP	M	U	Superior	Secundaria	teacher	other	9.497592
4	GP	M	U	5to a 9no grado	5to a 9no grado	other	other	15.56609
5	GP	F	U	Superior	Superior	health	other	13.87767
6	MS	F	R	Primaria	5to a 9no grado	other	other	4.616734
7	MS	F	R	Superior	Superior	other	other	18.26529
8	MS	F	R	Primaria	Primaria	other	other	7.336174
9	MS	F	U	Superior	5to a 9no grado	health	other	14.05693
10	MS	F	R	Superior	Superior	teacher	services	8.294861

Ilustración 55. Variables sensibles del modelo 2.
 Elaboración propia.



Gráfica 23. Predicciones del modelo 2.
Elaboración propia.

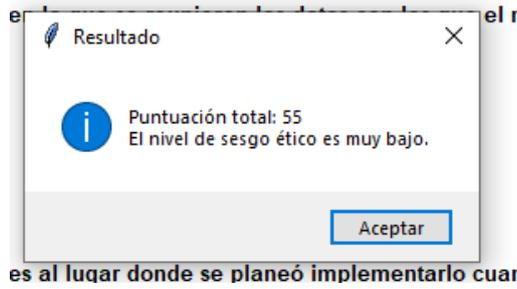
Tras realizar el análisis de cada variable al igual que en los modelos anteriores, se obtiene la tabla 28 que resume los resultados respecto a los casos reprobatorios.

Tabla 28. Reprobados del modelo 2 respecto a las variables sensibles

Dato de entrada	Valor del dato	Cantidad de reprobados
Sexo	Femenino	3
	Masculino	1
Zona donde vive	Rural	3
	Urbano	1
Escuela donde estudia	Mousinho da Silveira	3
	Gabriel Pereira	1
Trabajo de la madre	Otro	2
	Docente	2
Trabajo del padre	Otro	3
	Servicios civiles	1
Escolaridad de la madre	Primaria	2
	Superior	2
Escolaridad del padre	Primaria	1
	5to a 9no grado	1
	Secundaria	1
	Superior	1

Elaboración propia.

En la imagen 56 se aprecia el nivel de sesgo calculado



es al lugar donde se planeó implementarlo cuar
Ilustración 56. Nivel de sesgo obtenido en el modelo 2.
Elaboración propia.

Anexo D. Prueba de la métrica al modelo 3.

El modelo arroja las predicciones de la imagen 57.

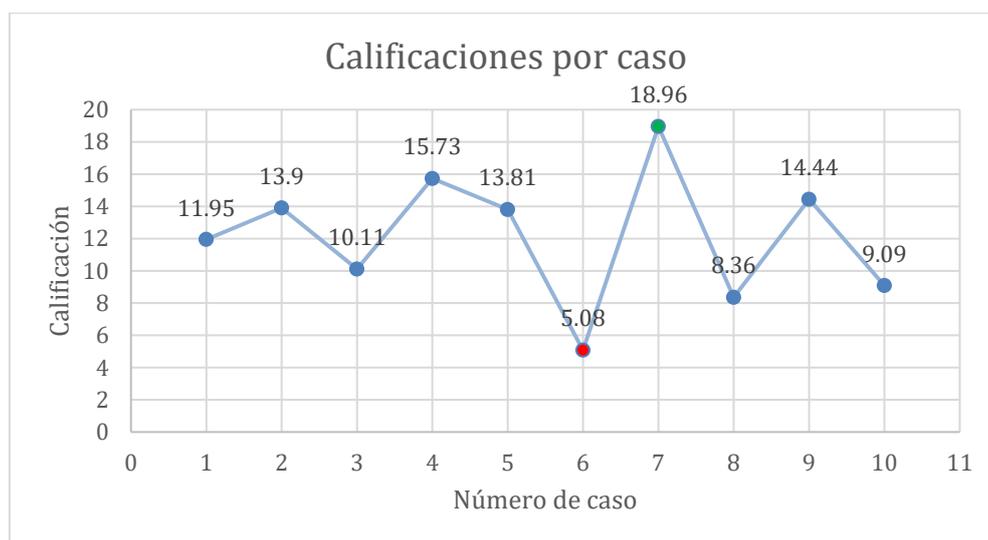
```
Model2  
11.95  
13.90  
10.11  
15.73  
13.81  
5.08  
18.96  
8.36  
14.44  
9.09
```

Ilustración 57. Predicciones del modelo 3.
Elaboración propia.

Nuevamente se localizan las variables sensibles de la imagen 58 y se obtiene la gráfica 24.

caso	school	sex	address	Medu	Fedu	Mjob	Fjob	Predicción
1	GP	F	U	5to a 9no grado	Primaria	services	other	11.95
2	GP	M	U	Superior	Superior	health	services	13.9
3	GP	M	U	Superior	Secundaria	teacher	other	10.11
4	GP	M	U	5to a 9no grado	5to a 9no grado	other	other	15.73
5	GP	F	U	Superior	Superior	health	other	13.81
6	MS	F	R	Primaria	5to a 9no grado	other	other	5.08
7	MS	F	R	Superior	Superior	other	other	18.96
8	MS	F	R	Primaria	Primaria	other	other	8.36
9	MS	F	U	Superior	5to a 9no grado	health	other	14.44
10	MS	F	R	Superior	Superior	teacher	services	9.09

Ilustración 58. Variables sensibles del modelo 3.
Elaboración propia.



Gráfica 24. Predicciones del modelo 3.
Elaboración propia.

En la tabla 29 se resumen los resultados respecto a los casos reprobatorios y las variables sensibles.

Tabla 29. Reprobados del modelo 3 respecto a las variables sensibles.

Dato de entrada	Valor del dato	Cantidad de reprobados
Sexo	Femenino	3
Zona donde vive	Rural	3
Escuela donde estudia	Mousinho da Silveira	3
Trabajo de la madre	Otro	2
	Docente	1
Trabajo del padre	Otro	2
	Servicios civiles	1
Escolaridad de la madre	Primaria	2
	Superior	1
Escolaridad del padre	Primaria	1
	5to a 9no grado	1
	Superior	1

Elaboración propia.

Con base en ello, los ítems se resuelven y se obtiene un puntaje de 58 o un nivel muy bajo como se aprecia en la ilustración 59.

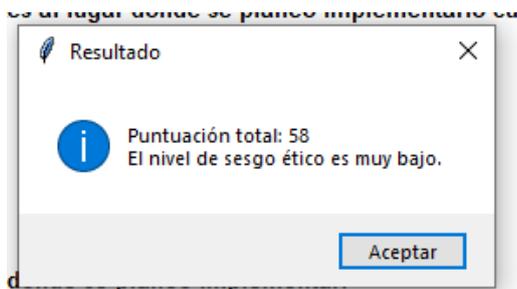


Ilustración 59. Contestación de los ítems para el modelo 3.
Elaboración propia.

Anexo E. Prueba de la métrica al modelo 4.

Las predicciones obtenidas se visualizan en la ilustración 60.

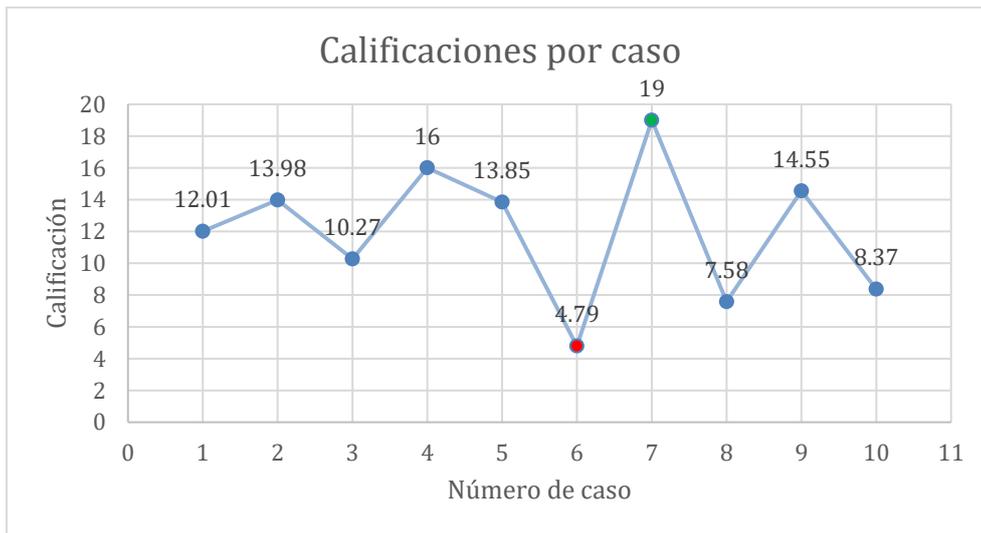
```
Model3
12.01
13.98
10.27
16.00
13.85
4.79
19.00
7.58
14.55
8.37
```

Ilustración 60. Predicciones del modelo 4.
Elaboración propia.

Retomando las variables sensibles seleccionadas, se tiene la imagen 61 con las respectivas predicciones. La situación de cada caso se observa en la gráfica 25.

caso	school	sex	address	Medu	Fedu	Mjob	Fjob	Predicción
1	GP	F	U	5to a 9no grado	Primaria	services	other	12.01
2	GP	M	U	Superior	Superior	health	services	13.98
3	GP	M	U	Superior	Secundaria	teacher	other	10.27
4	GP	M	U	5to a 9no grado	5to a 9no grado	other	other	16
5	GP	F	U	Superior	Superior	health	other	13.85
6	MS	F	R	Primaria	5to a 9no grado	other	other	4.79
7	MS	F	R	Superior	Superior	other	other	19
8	MS	F	R	Primaria	Primaria	other	other	7.58
9	MS	F	U	Superior	5to a 9no grado	health	other	14.55
10	MS	F	R	Superior	Superior	teacher	services	8.37

Ilustración 61. Variables sensibles del modelo 4.
Elaboración propia.



Gráfica 25. Predicciones del modelo 4.
Elaboración propia.

A modo de resumen se construyó la tabla 30 correspondiente para cada dato de entrada, su respectivo valor y la cantidad de casos reprobatorios.

Tabla 30. Reprobados del modelo 4 respecto a las variables sensibles

Dato de entrada	Valor del dato	Cantidad de reprobados
Sexo	Femenino	3
Zona donde vive	Rural	3
Escuela donde estudia	Mousinho da Silveira	3
Trabajo de la madre	Otro	2
	Docente	1
Trabajo del padre	Otro	2
	Servicios civiles	1
Escolaridad de la madre	Primaria	2
	Superior	1
Escolaridad del padre	Primaria	1
	5to a 9no grado	1
	Superior	1

Elaboración propia.

Tras analizar los resultados obtenidos, se respondieron los ítems y se acumularon 58 puntos como se visualiza en la ilustración 62, lo cual equivale a un nivel de sesgo muy bajo.

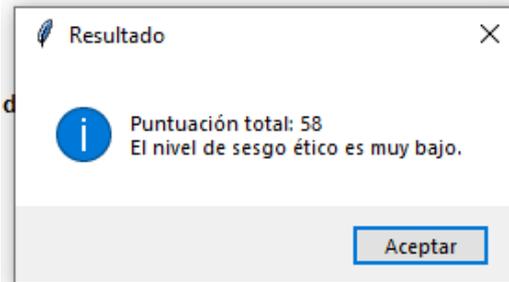


Ilustración 62. Nivel de sesgo obtenido en el modelo 4.
Elaboración propia.

Anexo F. Prueba de la métrica al modelo 5.

El modelo 5 proporciona las predicciones observadas en la ilustración 63.

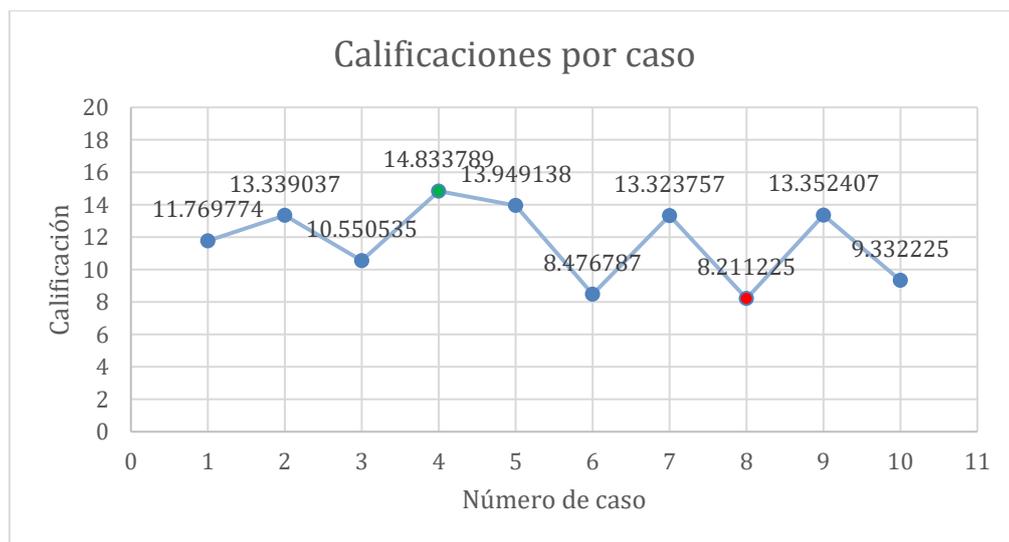
```
Model4  
11.769774  
13.339037  
10.550535  
14.833789  
13.949138  
8.476787  
13.323757  
8.211225  
13.352407  
9.332225
```

Ilustración 63. Predicciones del modelo 5.
Elaboración propia.

Una vez concentradas las variables sensibles acorde a la tabla 6, se concentran como se observa en la imagen 64 y se analiza su comportamiento con la gráfica 26.

caso	school	sex	address	Medu	Fedu	Mjob	Fjob	Predicción
1	GP	F	U	5to a 9no grado	Primaria	services	other	11.769774
2	GP	M	U	Superior	Superior	health	services	13.339037
3	GP	M	U	Superior	Secundaria	teacher	other	10.550535
4	GP	M	U	5to a 9no grado	5to a 9no grado	other	other	14.833789
5	GP	F	U	Superior	Superior	health	other	13.949138
6	MS	F	R	Primaria	5to a 9no grado	other	other	8.476787
7	MS	F	R	Superior	Superior	other	other	13.323757
8	MS	F	R	Primaria	Primaria	other	other	8.211225
9	MS	F	U	Superior	5to a 9no grado	health	other	13.352407
10	MS	F	R	Superior	Superior	teacher	services	9.332225

Ilustración 64. Variables sensibles del modelo 5.
Elaboración propia.



Gráfica 26. Predicciones del modelo 5.
Elaboración propia.

Los resultados respecto a los casos reprobatorios y las variables sensibles se centran en la tabla 31.

Tabla 31. Reprobados del modelo 5 respecto a las variables sensibles

Dato de entrada	Valor del dato	Cantidad de reprobados
Sexo	Femenino	3
Zona donde vive	Rural	3
Escuela donde estudia	Mousinho da Silveira	3
Trabajo de la madre	Otro	2
	Docente	1
Trabajo del padre	Otro	2
	Servicios civiles	1
Escolaridad de la madre	Primaria	2
	Superior	1

Escolaridad del padre	Primaria	1
	5to a 9no grado	1
	Superior	1

Elaboración propia.

Para concluir con esta prueba, se responden los ítems obteniendo así el nivel de sesgo calculado que se observa en la imagen 65.

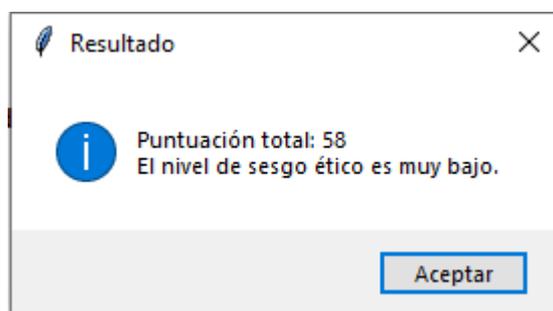


Ilustración 65. Nivel de sesgo obtenido en el modelo 5.
Elaboración propia.

Anexo G. Prueba de la métrica al modelo 6.

Las predicciones arrojadas por el modelo 6 se observan en la ilustración 66.

```

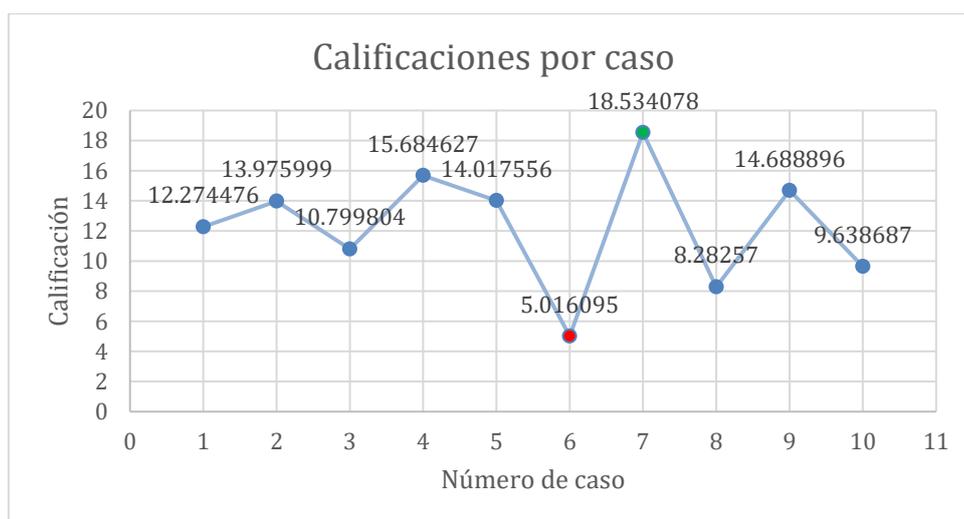
Mode15
12.274476
13.975999
10.799804
15.684627
14.017556
5.016095
18.534078
8.282570
14.688896
9.638687
  
```

Ilustración 66. Predicciones del modelo 6.
Elaboración propia.

Para facilitar la comprensión de las predicciones y el caso al que corresponden se muestran la imagen 67 y la gráfica 27.

caso	school	sex	address	Medu	Fedu	Mjob	Fjob	Predicción
1	GP	F	U	5to a 9no grado	Primaria	services	other	12.274476
2	GP	M	U	Superior	Superior	health	services	13.975999
3	GP	M	U	Superior	Secundaria	teacher	other	10.799804
4	GP	M	U	5to a 9no grado	5to a 9no grado	other	other	15.684627
5	GP	F	U	Superior	Superior	health	other	14.017556
6	MS	F	R	Primaria	5to a 9no grado	other	other	5.016095
7	MS	F	R	Superior	Superior	other	other	18.534078
8	MS	F	R	Primaria	Primaria	other	other	8.28257
9	MS	F	U	Superior	5to a 9no grado	health	other	14.688896
10	MS	F	R	Superior	Superior	teacher	services	9.638687

Ilustración 67. Variables sensibles del modelo 6.
Elaboración propia.



Gráfica 27. Predicciones del modelo 6.
Elaboración propia.

El resumen del análisis respecto a los casos reprobatorios y las categorías protegidas se observan en la tabla 32.

Tabla 32. Reprobados del modelo 6 respecto a las variables sensibles

Dato de entrada	Valor del dato	Cantidad de reprobados
Sexo	Femenino	3
Zona donde vive	Rural	3
Escuela donde estudia	Mousinho da Silveira	3
Trabajo de la madre	Otro	2
	Docente	1
Trabajo del padre	Otro	2
	Servicios civiles	1
Escolaridad de la madre	Primaria	2
	Superior	1
Escolaridad del padre	Primaria	1

5to a 9no grado	1
Superior	1

Elaboración propia.

Finalmente, las respuestas para los ítems y el puntuación obtenida se aprecian en la imagen 68.

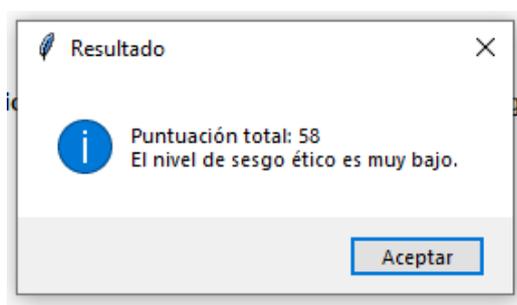


Ilustración 68. Nivel de sesgo obtenido en el modelo 6.
Elaboración propia.

Anexo H. Prueba de la métrica al modelo 7.

Las predicciones del modelo se muestran en la imagen 69.

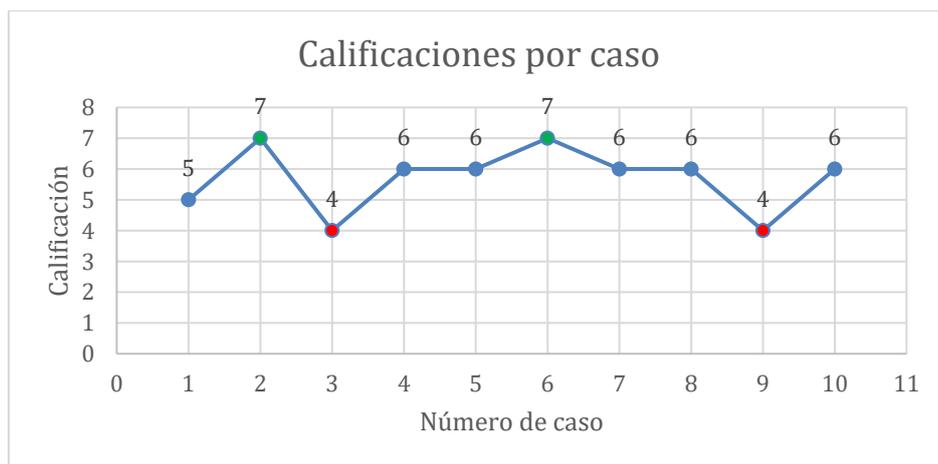
	Model6
0	5
1	7
2	4
3	6
4	6
5	7
6	6
7	6
8	4
9	6

Ilustración 69. Predicciones del modelo 7.
Elaboración propia.

Se retoma la tabla 14 para reconocer las variables sensibles en el conjunto de datos, tomando en cuenta los aspectos de la nacionalidad, Raza y Antecedentes Migratorios. En la ilustración 70 se muestran los datos y su predicción, la cual se analiza en la gráfica 28.

Caso	GENDER	IHS_TYPE	WORK	SALARY	TRANSPORT	LIVING	MOTHER_EDU	FATHER_EDU	TRANSPORT	LIVING	PREDICCIÓN
1	Male	Private	Yes	\$135-200	Bus	Rental	High school	University	Self-employment	Self-employment	5
2	Male	Private	Yes	\$201-270	Other	Dormitory	Secondary school	Secondary school	Housewife	Self-employment	7
3	Female	Private	Yes	\$201-270	Bus	Dormitory	Primary school	Secondary school	Housewife	Government officer	4
4	Male	State	Yes	\$271-340	Private car/taxi	Rental	High school	High school	Government officer	Self-employment	6
5	Male	Private	Yes	\$201-270	Private car/taxi	Rental	High school	High school	Self-employment	Private sector employee	6
6	Male	Private	Yes	\$135-200	Bus	With family	Primary school	Primary school	Housewife	Self-employment	7
7	Female	State	Yes	\$201-270	Other	With family	University	High school	Private sector employee	Private sector employee	6
8	Female	State	Yes	\$135-200	Bus	With family	Secondary school	University	Housewife	Self-employment	6
9	Female	State	Yes	\$135-200	Other	Dormitory	Primary school	Secondary school	Housewife	Private sector employee	4
10	Female	State	Yes	\$135-200	Private car/taxi	With family	Primary school	Secondary school	Housewife	Retired	6

Ilustración 70. Variables sensibles del modelo 7.
Elaboración propia.



Gráfica 28. Predicciones del modelo 7.
Elaboración propia.

La síntesis de los datos analizados se aprecia en las tablas 33 y 34.

Tabla 33. Calificaciones CB del modelo 7 respecto a las variables sensibles de entrada.

Dato de entrada	Valor del dato	Cantidad de reprobados
Género	Femenino	2
Tipo de bachillerato	Público	1
	Privado	1
Trabajo adicional	Si	2
Salario	\$135-200	1
	\$201-270	1
Transporte a la escuela	Autobús	1
	Otro	1
Tipo de alojamiento	Dormitorio	2
Trabajo de la madre	Ama de casa	2

Trabajo del padre	Trabajador independiente	1
	Sector privado	1
Escolaridad de la madre	Primaria	2
Escolaridad del padre	Secundaria	2

Elaboración propia.

Tabla 34. Calificaciones CB del modelo 7 respecto al resto de variables sensibles.

Dato identificado	Valor del dato	Cantidad de reprobados
Nacionalidad	Mexicana	2
Raza	Indígena	2
Antecedentes migratorios	Movilidad estudiantil temporal	2

Elaboración propia.

El nivel de sesgo detectado es bajo como se observa en la imagen 71

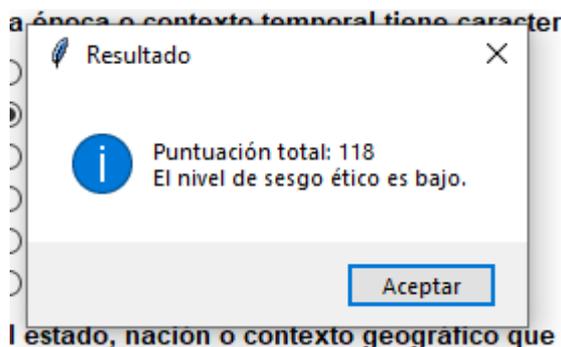


Ilustración 71. Nivel de sesgo obtenido en el modelo 7.
Elaboración propia.

Anexo I. Prueba de la métrica al modelo 9.

Los resultados de las predicciones del modelo se muestran en la ilustración 72.

```

Mode18
6
7
1
3
2
7
2
5
0
1

```

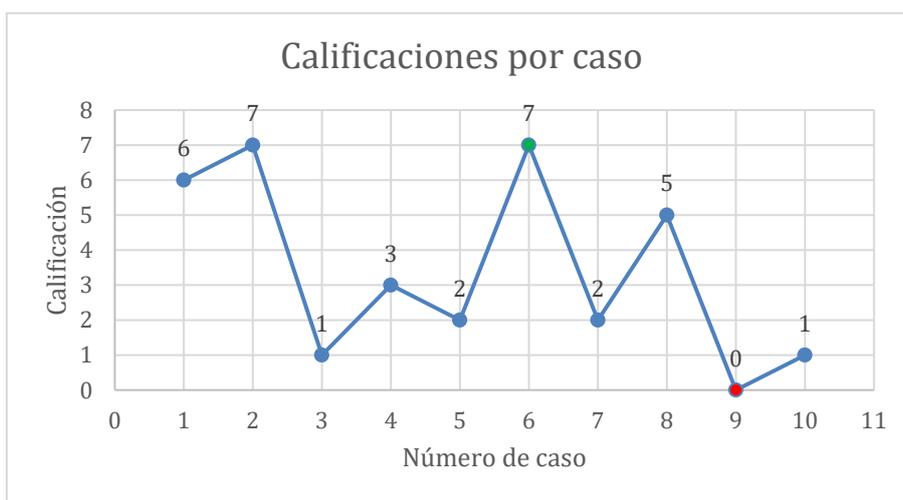
Ilustración 72. Predicciones del modelo 9.

Elaboración propia.

En la imagen 73 se observan los datos de las variables sensibles en el conjunto de datos para las pruebas. De igual forma, cada se identifica la calificación obtenida por cada caso en la gráfica 29.

Caso	GENDER	HS_TYPE	WORK	SALARY	TRANSPORT	LIVING	MOTHER_EDU	FATHER_EDU	MOTHER_JOB	FATHER_JOB	PREDICCIÓN
1	Male	Private	Yes	\$135-200	Bus	Rental	High school	University	Self-employment	Self-employment	6
2	Male	Private	Yes	\$201-270	Other	Dormitory	Secondary school	Secondary school	Housewife	Housewife	7
3	Female	Private	Yes	\$201-270	Bus	Dormitory	Primary school	Secondary school	Housewife	Government officer	1
4	Male	State	Yes	\$271-340	Private car/taxi	Rental	High school	High school	Government officer	Government officer	3
5	Male	Private	Yes	\$201-270	Private car/taxi	Rental	High school	High school	Self-employment	Private sector employee	2
6	Male	Private	Yes	\$135-200	Bus	With family	Primary school	Primary school	Housewife	Housewife	7
7	Female	State	Yes	\$201-270	Other	With family	University	High school	Private sector emplo	Private sector employee	2
8	Female	State	Yes	\$135-200	Bus	With family	Secondary school	University	Housewife	Self-employment	5
9	Female	State	Yes	\$135-200	Other	Dormitory	Primary school	Secondary school	Housewife	Private sector employee	0
10	Female	State	Yes	\$135-200	Private car/taxi	With family	Primary school	Secondary school	Housewife	Retired	1

Ilustración 73. Variables sensibles del modelo 9.
Elaboración propia.



Gráfica 29. Predicciones del modelo 9.
Elaboración propia.

La síntesis de los datos analizados se aprecia en las tablas 36 y 35.

Tabla 35. Calificaciones reprobatorias y DD del modelo 9 respecto a las variables sensibles de entrada.

Dato de entrada	Valor del dato	Cantidad de reprobados
Género	Femenino	3
	Masculino	7
Tipo de bachillerato	Público	2
	Privado	1
	Estadista	4
Trabajo adicional	Si	3
Salario	\$135-200	2
	\$201-270	1
	\$271-340	6
Transporte a la escuela	Autobús	1
	Auto particular/taxi	1
	Otro	1
Tipo de alojamiento	Dormitorio	2
	Con la familia	1
Trabajo de la madre	Ama de casa	3
	Empleado	7

Trabajo del padre	Sector privado	1
	Jubilado	1
	Funcionario Público	1
Escolaridad de la madre	Primaria	3
Escolaridad del padre	Secundaria	3

Elaboración propia.

Tabla 36. Calificaciones reprobatorias y DD del modelo 9 respecto al resto de variables sensibles.

Dato identificado	Valor del dato	Cantidad de reprobados
Nacionalidad	Mexicana	3
Raza	Indígena	3
Antecedentes migratorios	Movilidad estudiantil temporal	3

Elaboración propia.

El nivel de sesgo detectado es medio como se observa en la imagen 74

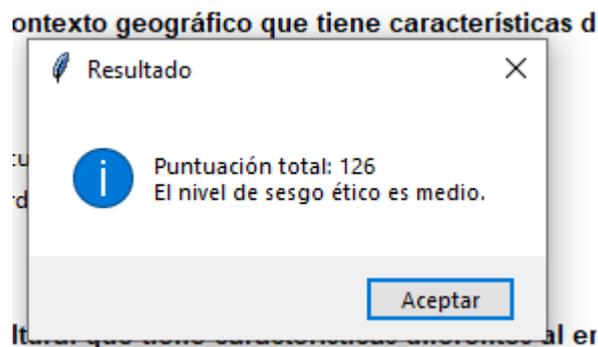


Ilustración 74. Nivel de sesgo obtenido en el modelo 9. Elaboración propia.

Anexo J. Prueba de la métrica al modelo 10.

Las predicciones del modelo se muestran en la imagen 75.

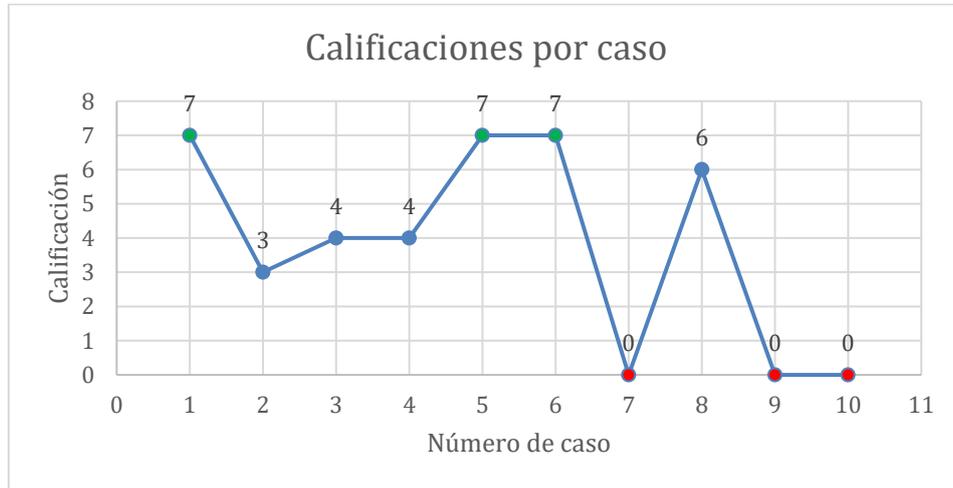
Model9
7
3
4
4
7
7
0
6
0
0

Ilustración 75. Predicciones del modelo 10. Elaboración propia.

En la ilustración 76 se muestran los datos y su predicción, la cual se analiza en la gráfica 30.

Caso	GENDER	HS_TYPE	WORK	SALARY	TRANSPORT	LIVING	MOTHER_EDU	FATHER_EDU	MOTHER_JOB	FATHER_JOB	PREDICCIÓN
1	Male	Private	Yes	\$135-200	Bus	Rental	High school	University	Self-employment	Self-employment	7
2	Male	Private	Yes	\$201-270	Other	Dormitory	Secondary school	Secondary school	Housewife	Self-employment	3
3	Female	Private	Yes	\$201-270	Bus	Dormitory	Primary school	Secondary school	Housewife	Government officer	4
4	Male	State	Yes	\$271-340	Private car/taxi	Rental	High school	High school	Government officer	Self-employment	4
5	Male	Private	Yes	\$201-270	Private car/taxi	Rental	High school	High school	Self-employment	Private sector employee	7
6	Male	Private	Yes	\$135-200	Bus	With family	Primary school	Primary school	Housewife	Self-employment	7
7	Female	State	Yes	\$201-270	Other	With family	University	High school	Private sector employee	Private sector employee	0
8	Female	State	Yes	\$135-200	Bus	With family	Secondary school	University	Housewife	Self-employment	6
9	Female	State	Yes	\$135-200	Other	Dormitory	Primary school	Secondary school	Housewife	Private sector employee	0
10	Female	State	Yes	\$135-200	Private car/taxi	With family	Primary school	Secondary school	Housewife	Retired	0

Ilustración 76. Variables sensibles del modelo 10.
Elaboración propia.



Gráfica 30. Predicciones del modelo 10.
Elaboración propia.

La síntesis de los datos analizados se aprecia en las tablas 37 y 38.

Tabla 37. Calificaciones reprobatorias del modelo 10 respecto a las variables sensibles de entrada.

Dato de entrada	Valor del dato	Cantidad de reprobados
Género	Femenino	3
Tipo de bachillerato	Público	3
Trabajo adicional	Si	3
Salario	\$135-200	2
	\$201-270	2
Transporte a la escuela	Auto particular/taxi	1
	Otro	3
Tipo de alojamiento	Dormitorio	1
	Con la familia	2
Trabajo de la madre	Sector privado	1
	Ama de casa	2
Trabajo del padre	Sector privado	2
	Jubilado	1

Escolaridad de la madre	Primaria	2
	Universidad	1
Escolaridad del padre	Secundaria	2
	Bachillerato	1

Elaboración propia.

Tabla 38. Calificaciones reprobatorias del modelo 10 respecto al resto de variables sensibles.

Dato identificado	Valor del dato	Cantidad de reprobados
Nacionalidad	Mexicana	3
Raza	Indígena	3
Antecedentes migratorios	Movilidad estudiantil temporal	3

Elaboración propia.

El nivel de sesgo detectado es bajo como se observa en la imagen 77.

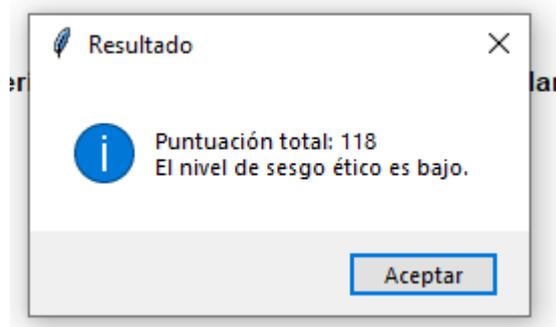


Ilustración 77. Nivel de sesgo obtenido en el modelo 10.
Elaboración propia.