



Universidad César Vallejo

**FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

**Plataforma de Reclutamiento 4.0: Aplicación de Machine Learning
para la Evaluación Autónoma de Perfiles Laborales, Lima, 2025**

Autor(es):

Alcedo Javier Carlos José (orcid.org/0000-0002-7248-6949)

Pachas Luicho Freddy Amós (orcid.org/0000-0001-7704-9150)

Asesor(a)(es):

Dr. Daza Vergaray, Alfredo(orcid.org/0000-0002-2259-1070)

Línea de investigación:

Sistema de información y comunicaciones

Línea de responsabilidad social universitaria:

Desarrollo económico, empleo y emprendimiento

LIMA – PERÚ

2025

Declaratoria de autenticidad del asesor

Declaratoria de originalidad del/los autor/es

Índice de contenidos

Declaratoria de autenticidad del asesor	ii
Declaratoria de originalidad del/los autor/es	iii
Índice de contenidos	iv
Índice de tablas	v
Índice de figuras	vi
I. Introducción	1
II. Metodología	14
III. Aspectos administrativos	21
3.1. Recursos	21
3.2. Financiamiento	23
3.3. Cronograma de ejecución	24
Referencias	27
Anexos	

Índice de tablas

Tabla 1. Recursos no monetarios del proyecto de investigación	21
Tabla 2. Recursos monetarios del proyecto de investigación	22
Tabla 3. Financiamiento de recursos del proyecto de investigación	23
Tabla 4. Cronograma de proyecto de investigación	24
Tabla 5. Cronograma de ejecución del proyecto de investigación	25

Índice de figuras

I. Introducción

El proceso de evaluación de perfiles laborales, en la actualidad ha tomado una relevancia significativa en toda organización, debido a que es clave para garantizar que los equipos de trabajo en una entidad, sean capaces de asumir retos y responsabilidades organizacionales que el mercado presenta [1]; sin embargo, este proceso enfrenta grandes obstáculos, pues existen serias dificultades para encontrar al candidato óptimo con los mecanismos actuales de evaluación, que poseen una brecha entre la oferta y demanda de talento, generando frustración en ambas partes [2]. de modo que no se tiene la seguridad de que los candidatos seleccionados cuenten con el talento, la pasión y la responsabilidad necesaria para cumplir con las expectativas de las organizaciones. [3]

En el ámbito internacional, los procesos de selección abordan diferentes desafíos que afectan de forma específica a la eficiencia y efectividad de la contratación de nuevos colaboradores, puesto que el mercado volátil y la necesidad de contar con habilidades diferenciadoras, ha conllevado que los modelos de captación de talento humano tradicionales, sean insuficientes [4]. Al respecto, se señala que el 50% de los líderes empresariales enfrenta diferentes dificultades con el modelo actual que poseen para captar talento en el proceso de evaluación de personal, que impide a sus organizaciones tener personal que satisfaga la demanda del mercado actual [5]; por ello, este proceso ha dejado de ser un aspecto clave en la gestión empresarial a una necesidad estratégica de toda organización a nivel mundial.

A nivel nacional, la situación de este proceso clave no es ajena a esta realidad, pues según el estudio “Tendencias de HR 2025” presentado en el diario Perú 21 [6], se ha podido identificar que el 65% de los reclutadores en nuestro país presentan serias dificultades para contratar el talento que estos desean, debido a la falta de información y datos claros de los conocimientos que estos poseen y una tendencia subjetiva al ejecutar las pruebas de selección, ocasionan barreras y limitaciones. Asimismo, se ha podido conocer que el 36% de las organizaciones formales utilizan instrumentos de inteligencia artificial dentro de sus procesos de

evaluación, sin embargo, se prevé que para años posteriores su empleo aumentará de forma significativa, debido a su capacidad para garantizar equidad y transparencia en un proceso crucial. Por otro lado, un estudio desarrollado por PwC, señala que el índice promedio de rotación de personal en nuestro país alcanza el 20.7%, lo que lo posiciona en las primeras posiciones de Latinoamérica, por el alto índice que este posee [7]. Este índice elevado tiene como uno de los factores principales, la ineficiencia en el proceso de selección, ya que se captan colaboradores que no poseen todas las habilidades y competencias necesarias para el puesto al cual fueron asignados, demostrando así que existe una brecha, que debe ser abordada con prontitud.

Dentro de las causas que originan este problema, se encuentran la existencia de sesgos en los procesos de evaluación de personal y el empleo de criterios subjetivos, que conlleva la existencia de estereotipos y discriminación en el proceso [8]. Además, la falta de eficiencia durante el proceso de selección, reflejada en la falta de consistencia en los criterios de selección y la mala gestión de tiempo genera serios limitantes para un proceso clave, como lo es la evaluación de perfiles laborales [9].

En consecuencia, estos factores generan que el proceso de evaluación de perfiles sea deficiente, generando en los candidatos un desequilibrio en la equidad que se tiene durante el proceso de evaluación y pierden la confianza en el proceso tradicional [8], además de que propician una mayor índice de rotación y afectan de forma directa al crecimiento organizacional. [10].

La entidad del rubro tecnológico que se tomará como referencia en el presente estudio, no es ajena a esta problemática, ya que presenta serias limitaciones y dificultades en su proceso de selección de personal, debido a que este no está definido y estructurado, lo que conlleva a que existan inconsistencias en la contratación y una grave dificultad para identificar qué candidatos cumplen con los perfiles requeridos para un puesto determinado. Además, el mecanismo de selección que se posee la entidad en la actualidad es subjetivo, ya que se basa en criterios personales de los encargados del área de reclutamiento, que genera la prevalencia

de sesgos, que afectan de manera directa en el proceso; asimismo, el tiempo empleado para este proceso de evaluación de perfiles demora más de lo previsto, lo que genera obstáculos en los procesos adyacentes, es así como esto impacta de forma negativa en la eficiencia de este proceso y repercute de forma directa en la rotación de personal y la cohesión organizacional, generando un entorno deficiente para el crecimiento y desarrollo de talento, así como el cumplimiento de los objetivos estratégicos.

Es así como frente a esta situación, es necesario explorar soluciones que enmarcan dos aspectos fundamentales: la innovación tecnológica y la visión estratégica de la gestión de talento humano. Por lo que, la incorporación de tecnologías con machine learning representan una alternativa eficaz para resolver las deficiencias del proceso de evaluación y mejorar la toma de decisiones dentro del área de gestión de talento humano, pues se brinda el margen suficiente para centrarse en cuestiones más importantes relacionadas con la mejora y el desarrollo del rendimiento. [11]

Es debido a las implicancias descritas que este estudio se encuentra orientado al Objetivo de Desarrollo Sostenible (ODS) número 9 propuesto por la ONU: Industria, innovación e infraestructura, que promueve la incorporación de tecnologías innovadoras en los procesos organizacionales para fomentar el desarrollo sostenible y la competitividad. Asimismo, está ligado al eje 3 del lineamiento de Responsabilidad Social Universitaria (RSU), donde se promueve el desarrollo económico, el empleo y el emprendimiento, ya que el presente estudio propone una solución tecnológica que optimiza el proceso de gestión de talento humano en las organizaciones y fortalece en cierta medida el entorno laboral, pues tiene un enfoque sostenible e inclusivo.

Por otra parte, la investigación tiene como eje responder a la siguiente interrogante: ¿Cómo una plataforma inteligente basada en algoritmos de machine learning automatiza la evaluación de perfiles laborales?, ya que esta interrogante surge de la necesidad inherente de optimizar los métodos tradicionales de reclutamiento, mediante la incorporación de tecnologías innovadoras, de modo que

se posibilite la mejora en la toma de decisiones, se garantice la transparencia en el proceso y se reduzca de forma clara los tiempos asignados a la evaluación de los potenciales candidatos, para un puesto de trabajo en una entidad.

Consecuentemente, para ahondar en el estudio, es necesario responder tres preguntas específicas que encaminan esta investigación, pues estas abordan de manera precisa los problemas adyacentes; en primer lugar, ¿el desarrollo de una plataforma inteligente basada en algoritmos de machine learning puede disminuir el nivel de subjetividad en la evaluación de perfiles labores?, debido a que esta interrogante permite analizar de la manera en cómo la automatización de este proceso puede disminuir en cierta medida la subjetividad e impulsar procesos más justos y equitativos; en segundo lugar, ¿el desarrollo de una plataforma inteligente puede reducir el tiempo empleado en la evaluación de perfiles laborales?, pues, esta interrogante nos permite entender cómo el empleo de nuevas tecnologías pueden acelerar las etapas que engloba el proceso de evaluación de perfiles; por último, ¿una plataforma inteligente basada en algoritmos de machine learning puede mejorar la eficiencia en la clasificación y evaluación de perfiles laborales?, ya que esta interrogante permite evaluar el impacto de esta herramientas en el procesamiento y análisis de información, de modo que se obtenga una elección de candidatos idónea, reduciendo en gran medida los errores que puedan suscitarse.

Por consiguiente, es necesario determinar la justificación que posee la presente investigación, en lo que refiere al ámbito operativo, se ha identificado la necesidad de reducir los tiempos empleados en el proceso, así como aumentar la eficiencia del proceso de evaluación de perfiles, a fin de tomar decisiones precisas, coherentes e inteligentes; pues, la integración de un sistema inteligente en las organizaciones, permite alcanzar cada aspecto señalado. Por lo que, usar herramientas basadas en inteligencia artificial para evaluar habilidades y perfiles laborales ayuda mucho a tomar decisiones sustentadas, ya que se procesan y clasifican los datos de manera rápida y efectiva, brindando al equipo responsable información significativa que facilite la toma de decisiones [12], pues a corto plazo será necesario alcanzar una automatización integral de este proceso, debido a su

relevancia en entornos empresariales cambiantes y su impacto directo en el desempeño organizacional. Uno de los principales beneficios de emplear estas herramientas es que favorecen a las entidades a encontrar y contratar al mejor talento, pues analizar y estudiar a profundidad estas tecnologías permitirá brindar a los gerentes ideas de cómo se puede mejorar el proceso de contratación [13]. Por ende, el construir una plataforma inteligente basada en algoritmos de machine learning permite una mejora operativa del proceso, así como genera una ventaja para las entidades que se apropien de esta herramienta.

Asimismo, desde el aspecto metodológico, esta investigación se sustenta en la construcción de una solución tecnológica cuya base está sujeta a fundamentos científicos, debido a que se emplean herramientas tecnológicas innovadoras que en la actualidad están revolucionando los procesos organizacionales; por ello, el emplear algoritmos de machine learning resulta necesario en la problemática abordada en este estudio, debido a que poseen la capacidad de gestionar de forma eficiente grandes cantidades de datos y establecer patrones complejos dentro de estos, a fin de predecir ciertos comportamientos. Es así como, el utilizar algoritmos de machine learning permiten una mayor consistencia en los resultados obtenidos en una evaluación de perfil, pues se disminuye los niveles de subjetividad y se procesa la información de forma más eficiente; en línea con esta perspectiva, el empleo de herramientas que incluyan IA sugiere un potencial significativo dentro de la gestión de recursos humanos, pues su flexibilidad y capacidad de procesamiento son claves para optimizar procesos dentro de esta área; por lo que es importante profundizar en investigaciones de este tipo para determinar las buenas prácticas en la aplicación de IA en este tipo de procesos [14]. Por ello, el enfoque propuesto no solo permitirá construir un sistema inteligente válido, sino que también contribuye a generar conocimientos aplicables a la mejora del proceso de evaluación de los perfiles laborales dentro de entornos organizacionales.

Desde el ámbito social, la presente investigación se justifica en el impacto positivo que conlleva la construcción de una plataforma inteligente con algoritmos de machine learning, debido a que esta, favorece en gran medida a la evaluación de

perfíles laborales desarrollado en las entidades sea equitativo y justo, además mejora la experiencia en los potenciales colaboradores, al eliminar sesgos inconscientes y aplicar criterios homogéneos y objetivos, que nutren y refuerzan el proceso. Al respecto, en un estudio se demostró que el empleo de una plataforma de inteligencia artificial dentro del proceso de reclutamiento de personal (evaluación de perfíles laborales), generó un impacto positivo en los participantes, pues calificaron que la plataforma demostró un comportamiento justo, confiable y transparente, destacando la diferencia significativa que existe entre esta herramienta y los métodos tradicional [15]. Por ende, aplicar la solución propuesta permitirá promover ambientes laborales diversos, donde el talento humano sea el eje central de la evaluación y se dejen de lado percepciones subjetivas y se favorezca al desarrollo organizacional.

Dado que, la presente investigación posee un sustento verídico, es necesario establecer objetivos claros que permitan responder de forma eficiente a cada una de las interrogantes planteadas, pues los objetivos deberán no solo abordar la creación de una solución tecnológica funcional para la problemática descrita, sino que deberán permitir la integración de aspectos teóricos y prácticos que faciliten la comprensión cable del proceso de selección en una entidad. En esta línea, se sostiene que el objetivo general es desarrollar una plataforma inteligente basada en algoritmos de machine learning para la automatización y mejora del proceso de evaluación de perfíles laborales, a fin de mejorar la eficiencia de este proceso clave que influye de forma directa en la selección de personal.

A partir de este objetivo, que es el eje central de la presente investigación, se derivan 3 objetivos específicos, estos son los siguientes: i) desarrollar un modelo de evaluación objetiva que reduzcan los sesgos en la selección, mediante el empleo de datos anonimizados y balanceados ii) reducir el tiempo empleado en el proceso de evaluación de perfíles laborales mediante la implementación de algoritmos de machine learning que optimicen cada una de las etapas de este proceso; iii) mejorar la eficiencia del proceso de evaluación de perfíles laborales mediante la implementación de algoritmos de machine learning. De esta manera, los objetivos planteados permiten la construcción de una solución tecnológica que abarque las

diferentes dificultades identificadas y se optimice de forma eficiente el proceso en estudio, para así generar una ventaja competitiva en la gestión de talento humano.

Por otro lado, este capítulo permite sustentar la presente investigación, mediante la revisión de investigaciones anteriores que han sido indexadas en base datos, y que han abordado la aplicación de herramientas basadas en machine learning dentro del proceso de selección de personal. Por lo que, en líneas posteriores se revisa el aporte de las investigaciones más relevantes, que abordaron desde diversas perspectivas esta aplicación, a fin de entender el impacto del empleo de esta tecnología en este proceso esencial y reconocer su importancia en nuestro contexto actual, donde el mercado es cada vez más competitivo.

En el ámbito internacional, Pessach et al. [16] realizaron una investigación que buscó desarrollar una herramienta de apoyo para la toma de decisiones en el proceso de reclutamiento haciendo uso del aprendizaje automático y programación matemática, para ello se basaron en 2 procesos: la predicción del éxito de la contratación y la optimización del proceso de contratación; y usaron alrededor de 700,000 registros para poder entrenar su modelo, de modo que se recopiló toda la información para poder ejecutar la fase inicial del proyecto, haciendo uso de las redes bayesianas de orden variable (VOBN); por consiguiente, para la segunda fase, aplicaron programación matemática con extensiones como el Rglpk en R, para poder así obtener modelos de optimización. En este sentido, los resultados evidenciaron que el modelo VOBN logró una precisión más alta en la predicción del éxito de la contratación, con un Área bajo la Curva (AUC) de 0.73 y el modelo de optimización permitió la mejora de la diversidad en las contrataciones en un 40%. Por lo que, el estudio concluyó que es factible combinar modelos de predicción y optimización.

Así como también, Oviedo et al. [17] en su investigación, plantearon el desarrollo de una herramienta que automatice el proceso de evaluación de personal, de manera que está efectúe una evaluación más eficiente y transparente de los postulantes, y esté alineada con las expectativas de la organización. Por lo que, la investigación se dividió en varias etapas pasando por la recolección de datos, preprocesamiento de datos, diseño de la red y la última etapa entrenamiento y

validación. Además, usaron enfoques de inventarios de personalidad (IPIP-NEO-FFI) y las redes neuronales artificiales de perceptrón multicapa, siendo este último el que permitió un óptimo desempeño en el análisis y predicción. Posteriormente a todas las etapas aplicadas, obtuvieron una precisión del 85% en la predicción de posibles colaboradores adecuados para la organización, así como también una reducción del 40% en el tiempo empleado para el proceso de selección. Por lo que, la herramienta desarrollada permitió la mejora de la eficiencia y efectividad durante el proceso de selección de personal.

En esta misma línea, Delecraz et al. [18] en su estudio buscaron desarrollar una solución de tipo algorítmica que mejore y automatice la contratación de los colaboradores, para ello su solución, se basó en tres principios fundamentales: ofrecer recomendaciones relevantes y responsables, reducir el tiempo, y asegurar un procesamiento de datos actualizado; para lo que se ejecutó una combinación de procesamiento de lenguaje natural y funciones cognitivas integradas, que potenciaron el algoritmo y lo hicieron más robusto; emplearon para esta solución un dataset con campos clave etiquetados (género, nacionalidad, lugar de nacimiento, edad, educación y requisito de RP); que permitieron que el modelo pueda clasificar y seleccionar al personal de forma óptima. Efectuada la construcción del algoritmo y una vez desarrolladas las pruebas, obtuvieron que la precisión para evaluar y ejecutar el proceso de selección fue de 80%, y el tiempo empleado en este proceso disminuyó. Por lo que, la investigación llegó a concluir que el algoritmo construido permite un mayor nivel de imparcialidad y mejora la transparencia y la explicabilidad de las decisiones ejecutadas.

En lo que refiere al ámbito nacional, Alvarado [19] realizó una investigación donde buscó la identificación de técnicas para el procesamiento de datos y videos en el proceso de evaluación de personal, por medio de redes neuronales y lógica difusa, para su construcción, se basó en un enfoque de estudio aplicado. Además, para el agrupamiento de datos, tomó en cuenta técnicas como el K-Means, DBSCAN, Clustering Jerárquico, Redes Neuronales y Lógica Difusa, mientras que, para los videos y su correcto procesamiento empleó Redes Neuronales Convolucionales,

Recurrentes, Generativas Adversariales y Encoder-Decoder. Siendo así que los resultados que obtuvo, demostraron una precisión de un 80% y un recall de 94.12%, teniendo así una reducción en el tiempo de selección, una mejora en la precisión de la selección y la facilitación del proceso para la toma de decisiones. Por lo que, concluyó que la aplicación de lógica difusa y las redes neuronales convolucionales, permite una correcta evaluación de personal.

En igual medida, Ogosi [20] desarrolló su investigación bajo el objetivo de implementar y evaluar una aplicación móvil con Machine Learning para la evaluación y selección de personal; aplicó un enfoque cuantitativo aplicado, teniendo una muestra de 5000 casos seleccionados aleatoriamente, así como utilizó una nueva tecnología llamada CSKT que considera los factores más importantes de las metodologías CRISP-DM, KDD, SEMMA y TDSP, para construir una solución enmarcada en la optimización y eficiencia. Luego de aplicar la solución, obtuvo resultados favorables, pues la metodología aplicada junto con el ML mejoró significativamente la eficiencia de tiempo y costo, pues se obtuvo un aumento del 40,3% en el índice de los postulantes y un aumento del 80% en la selección de mejores candidatos mediante la evaluación de sus perfiles laborales, así como un 50% en el índice de contratación. Por lo que, concluyó, que el uso de Machine Learning incrementa la eficiencia en la contratación, reduce tiempos y esfuerzo requerido en el proceso de selección y facilita la toma de decisiones.

Por otra parte, se ejecutó la búsqueda del dataset que será insumo para entrar correctamente al modelo de machine learning a construir, teniendo en cuenta la complejidad, escalabilidad e información requerida para poder tener así un rendimiento eficiente durante la toma de decisiones del modelo. Es por ello que, se optó por utilizar el conjunto de datos denominado FairCVdb, el cual está conformado por 24,000 perfiles en curriculums vitae (CVs), que servirán de insumo para analizar, procesar y establecer patrones de evaluación, que nos permitan reducir los sesgos y niveles de subjetividad dentro de la plataforma inteligente a desarrollar.

En este sentido, Ortega [21], utilizó el dataset señalado en conjunto con una programación inductiva LFIT para desarrollar una evaluación de perfiles laborales,

que permitió el reclutamiento automático de personal, de modo que se establecieron reglas importantes de clasificación que permitieron la detección de sesgos y una eficiencia significativa del sistema. Por lo que, en el estudio en cuestión, la aplicación del dataset y el modelo de programación proporcionaron una herramienta fiable para abordar contextos semejantes a la problemática de la presente investigación. De la misma forma, Sogancioglu et al. [22], hicieron uso de dicho dataset en su investigación para poder dar uso de la explicabilidad para la mitigación de sesgos dentro del proceso de evaluación en el reclutamiento de personal, mediante un algoritmo de mitigación de sesgos, llamado ProxyMute, el cual emplea la explicabilidad para poder detectar atributos sensibles y reducir los defectos en las decisiones al deshabilitarlas durante el tiempo de predicción.

Consecuentemente, a la revisión de los diferentes estudios que evidencian el impacto del machine learning en los procesos de la evaluación de perfiles laborales, es importante definir un marco conceptual el cual argumenta teóricamente la presente investigación. Por ello, se abordarán las principales teorías y conceptos relacionados al objeto de estudio, en líneas siguientes.

El machine learning es un concepto clave, en la presente investigación, puesto que actúa como variable independiente; y se entiende como un área dentro de la inteligencia artificial, que se orienta al desarrollo de algoritmos y técnicas que posibiliten a los equipos tecnológicos puedan aprender y mejorar su forma de actuar al ejecutar tareas a partir de datos [23]; por lo que, en lugar de delimitar funciones definidas, se da la información a los equipos para que estos puedan desarrollar sus propias predicciones, tomen decisiones y lleven a cabo tareas, basadas en la experiencia que han adquirido en sus entrenamiento, estableciendo de este modo sus propios patrones y construyendo su propio conocimiento.

Dado que el empleo de machine learning en la evaluación de los perfiles laborales implica abordar de forma categórica diversos escenarios, es necesario conocer teóricamente algunas teorías y algoritmos aplicables que posibiliten la automatización de este proceso.

En esta línea, las redes neuronales artificiales (RNA) pueden ser entendidos

como modelos matemáticos cuya base está centrada en la respuesta biológica del cerebro humano y el modo de aprender, pues estos modelos no lineales semi-paramétricos integran variables y unidades de procesamiento que se ajustan, entran y responden a diversos escenarios, de modo que posibilitan el manejo de forma eficiente de grandes volúmenes de datos [24]. Además, las RNA son modelos conexionistas que representan la estructura neuronal del cerebro y poseen un funcionamiento similar al de las neuronas biológicas, al momento de ejecutar procesos de aprendizaje, generación y abstracción, pues aplican un comportamiento lógico -racional semejante al ser humano [25].

Asimismo, el aprendizaje supervisado es aquel que posee un conjunto de entrenamiento, donde cada dato está delimitado por variables explicativas o características delimitadas y una variable de respuesta, para construir mediante datos un modelo de predicción, que genere una respuesta eficiente a un nuevo conjunto de datos no conocidos en base a las características que estas poseen [26]. Este tipo de aprendizaje tiende a aproximar la función de mapeo de forma precisa, para que los resultados obtenidos concuerden con la respuesta deseada, por ello el algoritmo va ejecutando predicciones de manera iterativa, hasta alcanzar predicciones aceptables, que cumplan con los estándares de validez. [27]

En igual medida, el aprendizaje automático (ML) es un método de análisis de datos que posibilita automatizar la creación de modelos analíticos, mediante el aprendizaje autónomo basado en experiencias [28], para que de esta forma, se establezcan patrones complejos que permitan al modelo actuar de forma autónoma frente a los escenarios planteados.

Por consiguiente, se considera importante definir las dimensiones que se abordan en la variable independiente, siendo la primera de estas la métrica de precisión en los modelos de machine learning [29], que comprende la parte de las observaciones positivas predichas adecuadamente respecto al total de positivos predichos por el modelo evaluado: cuya aplicación es útil principalmente cuando existe un alto coste de los falsos positivos.

Del mismo modo, se toma en cuenta a la métrica de recall o sensibilidad la

cual implica la proporción de observaciones positivas predichas de forma correcta con respecto al total de observaciones de la clase real, esta es indispensable en casos donde los falsos negativos son costosos o de alto riesgo. [29]

Así como también, el accuracy o exactitud ya que representa al rendimiento de un modelo sobre una tarea en específico, puede considerarse también como el grado de predicción que coincide con los datos reales. [29]

En lo referido a la variable dependiente, se define que para la presente investigación, la Evaluación Autónoma de Perfiles Laborales desempeñará este rol, de forma conceptual esta evaluación comprende agilizar el proceso de toma de decisiones, mediante el análisis de datos, bajo criterios establecidos que reduzcan los sesgos y mejoren la eficiencia y precisión para la identificación del candidato más adecuado.[30] [31] Esta variable dependiente se fundamenta teóricamente con las siguientes definiciones las cuales permiten conocer a mayor cabalidad las implicancias de este proceso y su influencia en el desempeño organizacional.

En ese sentido, es esencial comprender aspectos clave del proceso en el que aplicaremos las técnicas y algoritmos de machine learning; por lo que es necesario entender que son recursos humanos dentro de una entidad. En esta línea, Armijos et al. [32] nos mencionan que son un conjunto de conocimientos, motivaciones, experiencia, capacidades, competencias, habilidades y técnicas que tienen y aportan las personas a una organización, como un factor de ventaja más importante en el mercado actual.

Dentro de esta área, se desarrolla el proceso de reclutamiento y selección de personal, que conlleva una evaluación de los perfiles laborales de los potenciales colaboradores, por lo que este proceso es un desafío para las organizaciones, dado que involucra la identificación, clasificación y selección de candidatos que se ajusten a lo requerimientos de la organización [4]; asimismo como comprende la recopilación de datos de los candidatos potenciales para medir y evaluar los conocimientos, capacidades y habilidades laborales, que permitan al evaluador conocer cómo será su desempeño laboral dentro de la entidad [33].

Con lo mencionado anteriormente, es necesario conocer las dimensiones que

abordaremos para evaluar la variable dependiente señalada, por lo que se determinó que la primera dimensión serán los sesgos, que se entienden como un fenómeno común que es inevitable y puede afectar al proceso de evaluación que se lleva a cabo por el equipo de recursos humanos ya que este se basa en la evaluación no correspondida hacia las personas de manera inconsciente [34]. Además, esta dimensión puede presentar como estereotipos sexistas, racistas o tratos con personas racializadas de forma desigual en función de su origen o color [35].

Por consiguiente, se estableció a la reducción de tiempo como una dimensión, que implica emplear la inteligencia artificial para la selección de personal y evaluación de perfiles, por lo que se conceptualiza como el agregado de más recursos a actividades para acortar la duración de las mismas. [36]. Por último, se considera como tercer dimensión a la eficiencia en la evaluación de los postulantes, que es entendida como el valor que permite obtener los resultados esperados haciendo uso de la menor cantidad de recursos disponibles [37].

Por otra parte, es necesario abordar dentro de esta revisión teórica, aquellas herramientas que nos ayudarán con el desarrollo de la solución, siendo la primera de estas el lenguaje de programación Python el cual se define como un lenguaje de programación interpretado, siendo este de alto nivel y propósito general, además de contar tipado dinámico y recolección de elementos no utilizados, este lenguaje es muy popular en la investigación gracias a un sintaxis simple y directo, sin embargo lo que en verdad hace excelente a este lenguaje es su versatilidad dado que puede abordar diferentes tareas como la ciencia de datos, aprendizaje automático y computación científica [38].

En ese sentido, se consideró el uso de pandas una librería moderna de programación de alto nivel, está orientada a objetos los cuales contienen una alta colección de información, ya que a su vez se integra con otra biblioteca llamada Numpy para el manejo adecuado de los datos, ya que Numpy básicamente trabaja con datos tipo array para poder realizar indexaciones, ordenamiento, remodelación, etc [39].

Pasando a un entorno relacionado con la inteligencia artificial, se estará

considerando la biblioteca de Scikit-learn la cual es la más ideal para el aprendizaje automático en Python, ya que esta ofrece una diversidad de instrumentos para el modelado estadístico y aprendizaje automático, incluyendo la clasificación, regresión y agrupamiento [40]. Asimismo, se usará el sistema de TensorFlow el cual se trata de un sistema para el aprendizaje automático, pero a gran escala el cual puede usarse en diversas situaciones, ya que es excelente para realizar pruebas en nuevas técnicas de entrenamiento y optimizaciones, además de que esta cubre una gran gama de aplicaciones las cuales están centradas en la inferencia y el entrenamiento de redes neuronales profundas. [41]. Por último, dentro de las herramientas de desarrollo, se tiene en consideración a la biblioteca de Fairlearn el cual se define como un proyecto de código abierto que permite a los profesionales realizar evaluaciones y mejorar la equidad de los sistemas de inteligencia artificial, ya que esta permite evaluar los resultados de los modelos y aplicar diversos algoritmos para poder mitigar problemas de equidad. [42].

En cuanto a los algoritmos, optaron por el Random Forest Classifier ya que este se trata de un modelo de aprendizaje automático el cual es usado en la clasificación y previsión, principalmente para entrenar algoritmos de aprendizaje automático y modelos de inteligencia artificial, en ese sentido se debe considerar el uso de una cantidad sustancial de datos de alta calidad para tener una mejor replicación de datos. [43]

Y para poder abordar todas estas herramientas y algoritmos se planteó el uso de la metodología CRISP-DM, [44] Se define como un estándar y modelo de procesos que es independiente de la industria para la aplicación en los proyectos de minería de información o datos, contando así con 6 fases desde la comprensión del negocio hasta la fase de implementación.

Finalmente, se planteó una hipótesis general que responda a la pregunta central formulada, esta es: la construcción de una plataforma con algoritmos de machine learning permitirá optimizar de forma significativa el proceso de evaluación de perfiles laborales; y de igual forma se formularon las hipótesis específicas que responde a cada interrogante subyacente a la pregunta central; en primer lugar, el

desarrollo de una plataforma inteligente basada en algoritmos de machine learning reducirá el nivel de subjetividad en la evaluación de perfiles laborales, en segundo lugar, la construcción de una plataforma inteligente basada en algoritmos de machine learning disminuirá el tiempo empleado para la evaluación de perfiles laborales; y por último, el empleo de una plataforma basada en algoritmos de machine learning mejorará la eficiencia en la clasificación y evaluación de perfiles laborales; de modo que se beneficie al proceso de selección de personal.

II. Metodología

En esta sección se definirán los principales enfoques metodológicos que se tomarán en la presente investigación, así como también las variables a considerar, la población sobre la que se aplicará, los instrumentos considerados para la recolección de información o datos, diversos métodos para el análisis de datos y los aspectos éticos que se considerarán para garantizar el correcto desarrollo de la investigación.

Es por ello que, definimos el tipo de la presente investigación como aplicada dado que se centra en la solución del problema que afecta a los postulantes y las organizaciones o empresas frente al proceso tradicional de selección de personal. Así como también Castro et al. [45] mencionan que una investigación aplicada se centra en la adquisición de nuevos conocimientos para dar la solución de los problemas específicos en diversos períodos de tiempo. Con lo mencionado anteriormente, se buscaron diversas soluciones para poder cumplir con los objetivos mencionados en la investigación referentes a la reducción de sesgos en la selección, reducción de tiempo y la mejora de eficiencia del proceso selección de personal mediante el machine learning

Consecuentemente, se tomará un enfoque cuantitativo dado que se busca la evaluación de métricas numéricas de carácter positivo lo que permitirá la evaluación del cambio. Del mismo modo Acosta [46], se refiere al enfoque cuantitativo como el análisis y medición para los datos numéricos, los cuales son recolectados mediante encuestas, cuestionarios y experimentos controlados, para posteriormente ser llevados a procesos estadísticos. Con lo cual, podemos realizar la medición de impacto de los objetivos.

En base a lo mencionado anteriormente, se optara por un diseño de investigación del tipo experimental con el esquema pre experimental y un alcance de investigación de tipo explicativo dado que, se buscará el desarrollo de los objetivos a partir del dataset o conjunto de datos para el entrenamiento del modelo y así obtener el pre estímulo y el post estímulo. Es importante señalar que el diseño experimental busca que la variable independiente pueda generar un cambio en la variable dependiente [47].

En ese sentido, la variable independiente del presente estudio es “Machine Learning”, la cual se define de manera conceptual como una rama de la inteligencia artificial basada en modelos y métodos que favorezcan el proceso toma de decisiones, mediante el aprendizaje automático por medio de datos para predecir con precisión, exactitud y sensibilidad el comportamiento de datos [13] [23]; esta variable posee 3 dimensiones: en primer lugar, métrica de precisión, cuyo indicar es el nivel de precisión que posee el modelo en el proceso de evaluación; en segundo lugar, la métrica de recuperación, que será evaluado bajo el indicador de rendimiento del modelo; finalmente, se posee a la métrica de accuracy como última dimensión a evaluar, mediante el nivel de exactitud del modelo para la clasificación y evaluación. Mientras que la variable dependiente es la “Evaluación Autónoma de Perfiles Laborales”, que comprende agilizar el proceso en la toma de decisiones, a través del análisis de datos, bajo criterios establecidos que reduzcan los sesgos y mejoren la eficiencia y precisión para la identificación del candidato más adecuado. [30][31]. Esta variable posee 3 dimensiones: la reducción de sesgos, con los indicadores de tasa de reducción y paridad demográfica; la reducción de tiempo, con los indicadores de variación porcentual y tiempo promedio de evaluación de perfiles; y la eficiencia en el proceso de evaluación, que tiene como indicadores al rendimiento del modelo y el área bajo la curva, que determinan el nivel de confiabilidad de la herramienta para el proceso en el cual ha sido empleado; asimismo la información estructurada de forma visual con un nivel mayor de detalle se encuentra organizada en la tabla de operacionalización de variables, adjuntada en el Anexo 1.

La población se trata de las personas, muestras científicas, documentación, objetos, animales o empresas sobre las que se aplicará la metodología para poder medir correctamente los indicadores mencionados anteriormente. De igual manera Ocampo et al. [48] nos mencionan que, población o población humana se refiere a la unidad dinámica la cual está constituida por relaciones biológicas y sociales. En ese sentido, en ese sentido la población estará conformada por los candidatos o postulantes a un puesto de trabajo del dataset FairCvdb siendo un total de 24,000, además de considerar los respectivos modelos para el Machine Learning, los que

permiten la medición de la reducción de sesgos, tiempo y la mejora en la eficiencia en el proceso de evaluación de perfiles laborales.

En cuanto a los criterios de inclusión de la investigación son en principio, los conjunto de datos o dataset, los cuales deben estar exclusivamente orientados al contexto o problemática que abordaremos para poder garantizar así un correcto desarrollo y generación del modelo de Machine Learning, considerando que este dataset esté correctamente etiquetado y con información válida para poder realizar las fases de entrenamiento y prueba dentro del modelo. Mientras que, en los criterios de exclusión se toman en cuenta a aquellos que no fueron diseñados o desarrollados específicamente para el escenario en el que se realizara la investigación, además de también considerar excluidos a aquellos que no hayan sido probados o utilizados en algún estudio de la misma índole.

Teniendo en cuenta, a la muestra que se considera a una parte generalmente más pequeña de la población sobre la cual se usa para la aplicación del estudio o investigación. Es así que Bhandari [49], nos indica de manera similar que la muestra se trata de un grupo específico del cual se recopila información y que este siempre suele ser más pequeño que la población total. Para el presente estudio, se está considerando una sola muestra la cual está conformada por el conjunto de datos o dataset FairCvdb el cual fue obtenido de investigación descritas en el primer capítulo.. Dicho conjunto de datos cuenta con alrededor de 24,000 CV registrados con información previamente analizada y evaluada, de los cuales se usarán 100 CV del dataset como muestra.

Con base en lo mencionado anteriormente, podemos definir el muestreo el cual se trata de un procedimiento el cual escoge a un grupo de personas de una población de tal manera que permita estudiarlos. Definido de manera similar, algunos autores indican que se trata no solo de la población escogida para el estudio sino que se usa también para describir una porción de la población [50]. Por lo que dentro del presente estudio se aplicará un muestreo probabilístico aleatorio simple, ya que los datos no requieren de un filtro en específico o criterios especiales para ser partícipes de la muestra, siendo así uniformes permitiendo la selección aleatoria de

estos. Bajo la misma idea, el muestreo probabilístico permite al experimentador saber la probabilidad con la que un individuo puede ser escogido para formar parte de la muestra sobre la que se aplicará la investigación [51]

En ese sentido, podemos definir la unidad de análisis es un elemento crucial dentro de una investigación ya que este indicara lo que se está investigando. De la misma forma Damsa y Jornet [52] mencionan que la unidad de análisis es un elemento central dentro de cualquier metodología la cual determina el objeto de investigación. Por lo que dentro de la presente investigación se tomará como unidad de análisis a un modelo de Machine Learning el cual permitirá una correcta evaluación de los perfiles laborales mediante el uso de métricas de desempeño, eficiencia en el entrenamiento del modelo y la adaptabilidad de este, es así que a través de la implementación y análisis del modelo de Machine Learning se podrá brindar una respuesta al problema general y a los problemas específicos .

Dentro de las técnicas e instrumentos de recolección, Medina et al. [53] definen instrumentos de recolección como una herramienta utilizada para la recopilación y análisis de información en el transcurso del proceso de investigación. De la misma manera, se refiere a las técnicas de recolección como aquellas que aplican procedimientos y actividades las cuales proporcionan la obtención de datos necesarios por el investigador de tal manera que le permita dar una respuesta a las preguntas de investigación planteadas. Por lo que en la presente investigación se hará uso de las fichas de registro, mientras que para la técnica de recolección de datos se estará haciendo uso del fichaje lo que permitirá la visualización de las métricas del modelo, dado que se examinará la evaluación de perfiles laborales a partir del conjunto de datos para poder comprender la información que se recopiló a través de este, obteniendo así el número de evaluaciones aceptadas, la duración del proceso de evaluación y el índice de acierto en la preselección.

Se emplean diversos procedimientos para poder representar a los datos de los indicadores en las fichas de registro, esto mediante el modelo de Machine Learning propuesto frente a la evaluación de perfiles laborales, concentrándose más en la medida de los indicadores de tal manera que se pueda obtener la primera instancia

de los datos y posteriormente el análisis de los mismos. Dichos indicadores dentro de la ficha de registro hacen referencia al número de evaluaciones aceptadas, la duración del proceso de evaluación y el índice de acierto en la preselección. Bajo este procedimiento se espera encontrar una diferencias positivas entre el antes y después en el proceso de evaluación de perfiles laborales y poder formular conclusiones por cada objetivo.

Luego del proceso de obtención de datos, se llevará a cabo el análisis de los mismos con la herramienta SPSS considerando como técnica de análisis la estadística inferencial, la cual permite determinar parámetros de una población, mediante los datos de la muestra para poder contrastar con las hipótesis, Proaño [55] definen a la estadística inferencial como un conjunto de métodos que sirven de apoyo al investigador para poder obtener los datos de la población teniendo como base la información que proporciona la muestra.

En lo referido a los aspectos éticos, resulta importante destacar que estos son ejes centrales de una investigación científica, por ello su empleo debe ser de forma constante y con sustento, debido a que su aplicación garantiza que los estudios desarrollados sean íntegros y verificables. En este sentido, para el desarrollo de la investigación presente se ha considerado emplear un conjunto de datos fidedigno que posibilite la evaluación precisa de las variables desde un perspectiva objetiva; asimismo, se aplicó los principios de integridad y respeto de propiedad intelectual, de modo que la información empleada para la construcción de la presente investigación sea verídica y se respete la autoría de otros autores en fragmentos del sustento teórico empleado en la investigación, de modo que se cumpla con lo establecido en el artículo 3 del Código de Ética en Investigación de la Universidad César Vallejo aprobado en la RESOLUCIÓN DE CONSEJO UNIVERSITARIO N° 0659-2024/UCV, lo señalado en la Normativa Interna para asegurar la originalidad de las investigaciones de la Universidad César Vallejo aprobada en la RESOLUCIÓN DE CONSEJO UNIVERSITARIO No 0157-2025/UCV, y lo sostenido en el Código Nacional de Integridad Científica aprobado por CONCYTEC; donde se determinan las bases éticas que guiaron el desarrollo de esta investigación.

III. Aspectos administrativos

3.1. Recursos

Tabla 1
Recursos no monetarios del proyecto de investigación

Recursos humanos				
Apellidos y nombres de los investigadores	Costo unitario mensual en soles (U)	Cantidad de meses destinados al desarrollo del proyecto (Q)	Costo total (U*Q)	
Alcedo Javier Carlos José	1000	4	4000	
Pachas Luicho Freddy Amos	1000	4	4000	
			Subtotal	0
Servicios de terceros				
Nombre del servicio a ser adquirido	Unidad de medida	Costo unitario (U)	Cantidad (Q)	Costo total (U*Q)
Servicio de revisión de similitud (Turnitin)	Servicio	0	3	0
Asesoría de Tesis	Servicio	0	8	0
			Subtotal	0
Equipos y bienes duraderos				
Nombre del bien a ser adquirido	Unidad de medida	Costo unitario (U)	Cantidad (Q)	Costo total (U*Q)
Laptop Intel(R) Core(TM) i7-7700HQ, RAM 16.0 GB, SSD 512 GB, Tarjeta Gráfica Nvidia GeForce GTX 1060	Unidad	2900	1	2900
			Subtotal	2900
Pasajes y viáticos				
Descripción	Unidad de medida	Costo unitario (U)	Cantidad (Q)	Costo total (U*Q)

Movilidad local para reuniones de coordinación	Unidad	0	10	0
Subtotal				0
Materiales e insumos				
Nombre del material o insumo	Unidad de medida	Costo unitario (U)	Cantidad (Q)	Costo total (U*Q)
Dataset FairCVdb	Unidad	0	1	0
Repositorio de GitHub	Unidad	0	1	0
Visual Studio Code	Programa	0	2	0
Google Drive (Google Meet, Google Documents)	Cuenta	0	2	0
IBM SPSS v	Programa	290.72	1	290.72
PostgreSQL	Programa	0	1	0
Subtotal				290.72
Total de recursos no monetarios				11190.72

Tabla 2.

Recursos monetarios del proyecto de investigación

Recursos humanos				
Apellidos y nombres de los *investigadores	Costo unitario mensual en soles (U)	Cantidad de meses destinados al desarrollo del proyecto (Q)	Costo total (U*Q)	
Subtotal				
Servicios de terceros				
*Denominación	Unidad de medida	Costo unitario (U)	Cantidad (Q)	Costo total (U*Q)
Asesoría personalizada en integración de sistemas de ML	Servicio	S/. 200.00	1	S/. 200.00
Subtotal				S/ 200.00
Equipos y bienes duraderos				

*Denominación	Unidad de medida	Costo unitario (U)	Cantidad (Q)	Costo total (U*Q)
Laptop Intel(R) Core(TM) i9-13900H, 14 Procesadores, RAM 16.0 GB, SSD 1 TB	Unidad	S/. 4699.00	1	S/. 4699.00
Cooler Bowmann BWC4V máx. 16", 4 ventiladores	Unidad	S/. 49.99	1	S/.49.99
Teclado alámbrico membrana, layout español + Mouse 3600 dpi,	Unidad	S/.69.90	1	S/. 69.90
				Subtotal S/.4818.89

Pasajes y viáticos

Descripción	Unidad de medida	Costo unitario (U)	Cantidad (Q)	Costo total (U*Q)
				Subtotal
Materiales e insumos				
*Denominación	Unidad de medida	Costo unitario (U)	Cantidad (Q)	Costo total (U*Q)
Conexión a internet	Mensual	S/. 40.00	4	S/. 160.00
Consumo eléctrico	Mensual	S/. 30.00	4	S/. 120.00
				Subtotal S/.280.00
Total de recursos monetarios				S/. 5298.89

3.2. Financiamiento

Tabla 3.

Financiamiento de recursos del proyecto de investigación

	Recursos	*Monto	Entidad financiadora
No monetarios	Recursos humanos Servicios de terceros Equipos y bienes duraderos	S/. 11190.72	Financiamiento Propio

	<u>Pasajes y viáticos</u>	
	<u>Materiales e insumos</u>	
	<u>Subtotal</u>	S/. 11190.72
Monetarios	<u>Recursos humanos</u>	
	<u>Servicios de terceros</u>	
	<u>Equipos y bienes duraderos</u>	S/. 5298.89
	<u>Pasajes y viáticos</u>	Financiamiento Propio
	<u>Materiales e insumos</u>	
	<u>Subtotal</u>	S/. 5298.89
TOTAL		S/. 16,480.61

3.3. Cronograma de ejecución

Tabla 4

Cronograma de proyecto de investigación

Actividad	Semana															
	1	2	3	4	5	6	7	8	9	10	11	12	12	14	15	16
Determinar problemática de investigación																
Definir un título tentativo para investigación																
Desarrollar el planteamiento del problema																
Determinar preguntas de investigación y objetivos																
Investigar sobre antecedentes de investigación y bases teóricas																
Delimitar aspectos metodológicos de investigación																
Determinar																

instrumentos de investigación														
Definir recursos monetarios y no monetarios de investigación														
Determinar modo de financiamiento														
Construir cronograma de proyecto														
Sustentar proyecto de investigación														
Levantar y ajustar observaciones														

Tabla 5
Cronograma de ejecución del proyecto de investigación

Actividad	Semana															
	1	2	3	4	5	6	7	8	9	10	11	12	12	14	15	16
Adaptar formato de informe al establecido para tesis																
Desarrollar modelo de análisis y selección propuesto																
Entrenar y testear modelo																
Construir interfaz gráfica de plataforma																
Integrar interfaz con modelo de selección																
Recopilar datos e información con modelo construido																
Analizar datos recopilados en																

pruebas
desarrolladas

Discutir resultados
obtenidos y
contrastarlos

Elaborar
conclusiones y
recomendaciones
del proyecto

Elaborar artículo
científico

Postular a revista
científica
indexada

Sustentar informe
de tesis

Referencias

- [1] J. Naranjo Gaibor. Inteligencia artificial y el proceso de reclutamiento y selección de recursos humanos. *Journal of Science and Research.* vol.10, n°1, ene. - mar., 2025. Consultado 29 abr. 2025. doi: [10.5281/zenodo.14612811](https://doi.org/10.5281/zenodo.14612811)
- [2] M. Navarrete, Y. Andramunio, P. Tadeo y H. Narvaez. Modelación de perfiles por competencias en la gestión del talento humano. *Revista Latinoamericana de Ciencias Sociales y Humanidades.* vol. 5 n°5, oct., 2024. doi: [10.56712/latam.v5i5.2743](https://doi.org/10.56712/latam.v5i5.2743)
- [3] G. Serrano, L. Guerrero, V. Zamudio, M. González. Redes neuronales: nueva estrategia de IA para implementar dentro del proceso de reclutamiento y selección de personal. *Revista Internacional de Tecnología, Conocimiento y Sociedad.* vol. 9, n°2, 2022. Consultado 29 abr. 2025. doi: [10.18848/2474-588X/CGP/v09i02/45-54](https://doi.org/10.18848/2474-588X/CGP/v09i02/45-54)
- [4] I. Gonzabay Quiñoñez y S. Pacheco Mendoza, El rol de la Inteligencia Artificial en los procesos de reclutamiento y selección en la Gestión del Talento Humano . vol. 3, n°6, jul-dic 2024. doi: [10.59282/reincisol.V3\(6\)3880-3902](https://doi.org/10.59282/reincisol.V3(6)3880-3902)
- [5] Mercer. Beat the crisis. How executives are responding to economic shocks and talent shortages. Mercer, 2023. [En línea] Disponible : <https://www.mercer.com/en-ca/insights/talent-and-transformation/attracting-and-retaining-talent/how-executives-are-responding-to-economic-shocks-and-talent-shortages/>
- [6] PERÚ21, “El 65% de los reclutadores en Perú enfrentan dificultades para encontrar talento calificado”, Perú21, Consultado 29 abr. 2025. <https://peru21.pe/economia/el-65-de-los-reclutadores-en-peru-enfrentan-dificultades-para-encontrar-talento-calificado/>
- [7] PERÚ21, “Empresas: ¿Cómo evitar la pérdida de dinero en recursos por la rotación de personal?”, Perú21, Consultado 29 abr. 2025. <https://peru21.pe/cheqa/empresas-rotacion-personal-como-evitar-la-perdida-de-dinero-en-recursos-por-la-rotacion-de-personal-noticia/>
- [8] L. Roumbanis, *On the present-future impact of AI technologies on personnel*

- selection and the exponential increase in meta-algorithmic judgments.* Futures, vol. 166, 2025, doi: [10.1016/j.futures.2025.103538](https://doi.org/10.1016/j.futures.2025.103538)
- [9] H. Suen y K. Hung. Revealing the influence of AI and its interfaces on job candidates' honest and deceptive impression management in asynchronous video interviews. *Technological Forecasting and Social Change.* vol. 198, nov., 2023. doi: [10.1016/j.techfore.2023.123011](https://doi.org/10.1016/j.techfore.2023.123011)
- [10] I. Méndez y M. Sotolongo. CONTRIBUCIÓN AL PERFECCIONAMIENTO DEL PROCESO DE SELECCIÓN DE PERSONAL BASADO EN COMPETENCIAS. *Rev. Cub. de Admin. Pub y Emp.* vol.6 n°2, may.--ago. 2022. doi: [10.5281/zenodo.6997287](https://doi.org/10.5281/zenodo.6997287)
- [11] J. Fraij y V. A Laszlo, Literature Review: Artificial Intelligence Impact on the Recruitment Process. *International Journal of Engineering and Management Sciences.* vol. 6, n°1, 2021. doi: [10.21791/IJEMS.2021.1.10](https://doi.org/10.21791/IJEMS.2021.1.10).
- [12] R. Deepa, S. Sekar, A. Malik, J. Kumar y R. Attri. Impact of AI-focussed technologies on social and technical competencies for HR managers – A systematic review and research agenda. *Technological Forecasting and Social Change,* vol. 202, may. 2024. doi: [10.1016/j.techfore.2024.123301](https://doi.org/10.1016/j.techfore.2024.123301)
- [13] T. Fernandes, H. Sao, J. Pereira y V. Pereira. Artificial intelligence applied to potential assessment and talent identification in an organisational context. *Heliyon.* vol. 9, n°4, abr.v2024. doi: [10.1016/j.heliyon.2023.e14694](https://doi.org/10.1016/j.heliyon.2023.e14694)
- [14] N. Sakib, M. A. Jabber, M. Younus, M. Bithee, M. Sannyamat, A. Saha y A. Guha, Mapping the impacts of neural networks on human resource management research: a bibliometric analysis. *Future Business Journal .* vol. 11, n° 1, abr. 2025. doi: [10.1186/s43093-025-00515-9](https://doi.org/10.1186/s43093-025-00515-9)
- [15] G. Ogunniye, B. Legastelois, M. Rovatsos, L. Dowthwaite, V. Portillo, E. Perez Vallejos, J. Zhao, M. Jirotka. Understanding User Perceptions of Trustworthiness in E-Recruitment Systems. *IEEE Internet Computing .* vol. 25, no. 6, nov-dic 2021. doi: [10.1109/MIC.2021.3115670](https://doi.org/10.1109/MIC.2021.3115670)
- [16] D. Pessach, G. Sinbger, D. Avrahami, H. Chalutz Ben-gal, E. Shmueli y I. Ben-gal, Employees recruitment: A prescriptive analytics approach via

- machine learning and mathematical programming. Decision Support Systems. vol. 134, jul. 2024. doi: [10.1016/j.dss.2020.113290](https://doi.org/10.1016/j.dss.2020.113290)
- [17] J. A. Oviedo Monroy, A. Mauricio Arciniegas y F. Camacho Garcia,. Desarrollo de una herramienta de selección de personal basada en reclutamiento predictivo e inteligencia artificial. 1 ed. Colombia: Centro de Comercio y Servicios, 2021. Consultado: 10 may. 2025] . [en línea]. Disponible : <https://hdl.handle.net/11404/8358>
- [18] S. Delecraz, L. Eltarr, M. Becuwe, H. Bouxin, N. Boutin y O. Oillier. Responsible Artificial Intelligence in Human Resources Technology: An innovative inclusive and fair by design matching algorithm for job recruitment purposes. Journal of Responsible Technology. vol. 11, oct. 2022. doi: [10.1016/j.jrt.2022.100041](https://doi.org/10.1016/j.jrt.2022.100041)
- [19] E. J. Alvarado Alva. Sistema inteligente para el proceso de reclutamiento y selección de miembros en Aiesec Chiclayo. Tesis de Titulación en Ingeniería. Perú: Universidad Católica de Santo Toribio. Chiclayo, Perú, 2025. Disponible: https://tesis.usat.edu.pe/bitstream/20.500.12423/8388/1/TL_AlvaradoAlvaEmil.pdf
- [20] J. A. Ogosi Auqui. IMPLEMENTACIÓN DE UNA APLICACIÓN MÓVIL CON MACHINE LEARNING APlicando LA NUEVA METODOLOGÍA CSKT PARA EL PROCESO DE CONTRATACIÓN DE PERSONAL EN EL SECTOR PRIVADO. Tesis de Doctorado en Ingeniería: Universidad Nacional Federico Villarreal, Lima, Perú, 2025 . Disponible : <https://hdl.handle.net/20.500.13084/10220>
- [21] A. Ortega, J. Fierrez, A. Morales, y Z. Wang. Symbolic AI for XAI: Evaluating LFIT Inductive Programming for Fair and Explainable Automatic Recruitment. Computers . Vol. 10, n° 11, dic, 2020. doi: [10.3390/computers10110154](https://doi.org/10.3390/computers10110154)
- [22] G. Sogancioglu, H. Kaya, A. A. Salah. *Using Explainability for Bias Mitigation: A Case Study for Fair Recruitment Assessment*. En: Proceedings of the 25th International Conference on Multimodal Interaction (ICMI '23). Paris, France: Association for Computing Machinery, pp. 631–639, 2023. doi:

10.1145/3577190.3614170

- [23] D. Rodriguez, ANÁLISIS DE DATOS Y MACHINE LEARNING. Tesis de Titulación en Estadística. Salamanca: Universidad de Salamanca, 2023. Disponible:
<https://gredos.usal.es/bitstream/handle/10366/157350/TFG%20Damia%cc%81n%20Rodri%cc%81quez%20Go%cc%81mez.pdf?sequence=1>
- [24] A. Viloria, M. Mendieta, L. A. Borrero y O. B. Pineda. Prediction of Mandibular Morphology through Artificial Neural Networks. Procedia Computer Science. Vol. 170, Abril 2020. doi: [10.1016/j.procs.2020.03.064](https://doi.org/10.1016/j.procs.2020.03.064)
- [25] D. Restrepo, J. Viloria y C. Robles. El camino a las redes neuronales artificiales [en linea]. 1 ed. Colombia: Universidad de Magdalena, 2021. Consultado: 10 de may. 2025, .Disponible:
https://books.google.com.pe/books?hl=es&lr=&id=hTSDEAAAQBAJ&oi=fnd&pg=PT12&dq=redes+neuronales+artificiales&ots=ytpL0b0SaS&sig=Ws8soUHrg0f29r9YwkuJ7k-Tld4&redir_esc=y#v=onepage&q=redes%20neuronales%20artificiales&f=false
- [26] G. Valenzuela, Aprendizaje Supervisado: Métodos, Propiedades y Aplicaciones. Tesis Bachiller en Matemáticas. España: Universidad de Málaga, 63 pp, 2022. Disponible
https://riuma.uma.es/xmlui/bitstream/handle/10630/25147/TFG_Aprendizaje_Supervisado_GVG.pdf?sequence=4&isAllowed=y
- [27] A. Serra. Comparación de algoritmos de clasificación supervisada. Tesis Maestría en Ingeniería. España: Escuela Técnica Superior de Ingeniería Industrial de Barcelona, 99 pp, 2020. Disponible:
<https://upcommons.upc.edu/bitstream/handle/2117/330482/tfm-mueo-alexander-e-serra.pdf?sequence=1>
- [28] K. Sharifani, y M. Amini. Machine Learning and Deep Learning: A Review of Methods and Applications. World Information Technology and Engineering Journal vol. 10, n°7, Julio 2023. Disponible:
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4458723

- [29] T. Adhikari, "How to Evaluate the Accuracy of an AI Model?", 2024. Disponible: https://www.researchgate.net/publication/384233756_How_to_Evaluate_the_Accuracy_of_an_AI_Model
- [30] G. Cortina Robles, K. Valencia Borja y V. Bayona Galindo. "Automatización en los procesos de selección y reclutamiento, para simplificar el procedimiento de contratación", Tesis de Titulación, Fac. Cien. Hum. y Soc., Univ. Cooperativa de Colombia, Bogotá, Colombia, 2022. [En línea]. Disponible: <https://hdl.handle.net/20.500.12494/44984>
- [31] C. Gonzales García. "Impacto de la inteligencia artificial en la toma de decisiones de selección de personal: Una revisión integrativa", Tesis de Maestría, Univ. Miguel Hernandez, España, 2023. [En línea]. Disponible: <https://hdl.handle.net/11000/32183>
- [32] F. B. Armijos Mayon, A. I. Bermudez Burgos y N. V. Mora Sanchez,. Gestión de administración de los Recursos Humanos. Universidad y Sociedad . vol.11, n.4, sep. 2019. Disponible: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2218-36202019000400163&lng=es&nrm=iso.
- [33] E. Campion y M. Campion, Impact of machine learning on personnel selection. Organizational Dynamics. vol 53, n°1, ene-mar 2024. doi: [10.1016/j.orgdyn.2024.101035](https://doi.org/10.1016/j.orgdyn.2024.101035)
- [34] D. D. Boateng, Effect of Bias on Selection and Recruitment, Academia Letters, ago. 2021, doi: [10.20935/al3187](https://doi.org/10.20935/al3187).
- [35] I. Kekez, L. Lauwaert, y N. Begićević Ređep, Is artificial intelligence (AI) research biased and conceptually vague? A systematic review of research on bias and discrimination in the context of using AI in human resource management, Technol Soc, vol. 81, jun. 2025, doi: [10.1016/j.techsoc.2025.102818](https://doi.org/10.1016/j.techsoc.2025.102818).
- [36] C. R. Ramos A. y C. A. Flores B., REDUCCIÓN DEL TIEMPO DE FINALIZACIÓN DEL PROYECTO DE UNA PLANTA DE CONSERVAS DE PESCADO UTILIZANDO UN MODELO DE PROGRAMACIÓN LINEAL",

- Anales Científicos, vol. 77, núm. 1, p. 110, jun. 2016, doi: [10.21704/ac.v77i1.480](https://doi.org/10.21704/ac.v77i1.480).
- [37] B. W. Najar, “Efficiency and/or Effectiveness in Managing Organizations”, J Educ Cult Stud, vol. 4, núm. 2, p. p131, may 2020, doi: [10.22158/jecs.v4n2p131](https://doi.org/10.22158/jecs.v4n2p131).
- [38] D. G. Abu Rayhan, “The Rise of Python: A Survey of Recent Research”, sep. 2023, doi: 10.13140/RG.2.2.27388.92809.
- [39] R. Snehkunj y K. Vachiyatwala, “Data Analysis Using Pandas Library of Python”, Acta Scientific Computer Sciences 4.3, vol. 4, 2022, Consultado: el 5 de junio de 2025. [En línea]. Disponible en: <https://actascientific.com/ASCS/pdf/ASCS-04-0236.pdf>
- [40] Z. Karimi, “Scikit-learn-Quick-Review”, 2021, doi: 10.13140/RG.2.2.14605.67043.
- [41] M. Ramchandani et al., “Survey: Tensorflow in Machine Learning”, en Journal of Physics: Conference Series, Institute of Physics, 2022. doi: 10.1088/1742-6596/2273/1/012008.
- [42] H. Weerts, M. Dudík, R. Edgar, A. Jalali, R. Lutz, y M. Madaio, “Fairlearn: Assessing and Improving Fairness of AI Systems”, mar. 2023, [En línea]. Disponible en: <http://arxiv.org/abs/2303.16626>
- [43] H. A. Salman, A. Kalakech, y A. Steiti, “Random Forest Algorithm Overview”, Babylonian Journal of Machine Learning, vol. 2024, pp. 69–79, jun. 2024, doi: 10.58496/bjml/2024/007.
- [44] C. Schröer, F. Kruse, y J. M. Gómez, “A systematic literature review on applying CRISP-DM process model”, en Procedia Computer Science, Elsevier B.V., 2021, pp. 526–534. doi: 10.1016/j.procs.2021.01.199.
- [45] J. J. Castro Maldonado, L. K. Gómez Macho, and E. Camargo Casallas, “La investigación aplicada y el desarrollo experimental en el fortalecimiento de las competencias de la sociedad del siglo XXI,” Tecnura, vol. 27, no. 75, pp. 140–174, Jan. 2023, doi: 10.14483/22487638.19171.
- [46] S. Acosta Faneite, Los enfoques de investigación en las Ciencias Sociales.

- Revista Latinoamericana Ogmios. vol. 3, n.º 8, p. 82–95, jul 2023. doi: 10.53595/rlo.v3.i8.084
- [47] C. Ramos Galarza, Diseños de investigación experimental. CienciAmérica. vol. 10, n.º 1, 2021. doi: [10.33210/ca.v10i1.356](https://doi.org/10.33210/ca.v10i1.356)
- [48] C. Ocampo Mallou, A. Delvitto, y F. M. Di Pasquo. Discusiones sobre el concepto de población humana en la investigación epidemiológica del dengue en Argentina. vol 22, n°44, jun. 2022.doi: [10.18270/rcfc.v22i44.3486](https://doi.org/10.18270/rcfc.v22i44.3486)
- [49]] P. Bhandari, Population vs Sample | Definitions, Differences & Examples. may. Disponible:
<https://www.scribbr.co.uk/research-methods/population-versus-sample/>
- [50] M. Cortes, E. Moraga y D. Silva, Técnicas de muestreo probabilístico para investigación en ciencias de la salud, 1 ed. Brasil: Ed. Atenea, 2023. [En línea]. Disponible: <http://dx.doi.org/10.22533/at.ed.8802311102>.
- [51] B. Almaguer y E. Cossio. “Métodos de Muestreo para la Optimización de Diseño de Experimentos”. Academia Journals, vol. 14, no 22, 2022.Disponible: <https://ciateq.repositorioinstitucional.mx/jspui/bitstream/1020/644/1/Metodos%20de%20muestreo%20para%20la%20optimizacion%20de%20dise%c3%b1o.pdf>
- [52] C. Damsa y A. Jornet. “The unit of analysis in learning research: Approaches for imagining a transformative agenda”. Learning, Culture and Social Interaction, vol. 31, no 2, dic. 2021. doi: [10.1016/j.lcsi.2020.100407](https://doi.org/10.1016/j.lcsi.2020.100407)
- [53] M. Medina, R. Rojas, W. Bustamante, R. Loaiza, C.Martel y R. Castillo, Metodología de la investigación: Técnicas e instrumentos de investigación. Perú: Inst. Univ. de Innovación Ciencia y Tecnología, 2023. doi: [10.35622/inudi.b.080](https://doi.org/10.35622/inudi.b.080)
- [54] W. B. Proaño-Rivera, Estadística descriptiva e inferencial. Casa Editora Universidad del Azuay, 2020. [doi: 10.33324/ceuazuay.127](https://doi.org/10.33324/ceuazuay.127).
- [55] S. Diaz, W. Sanyer. “ Selección de candidatos para encuestas mediante técnicas de machine learning”, Tesis de Titulación, Fac. Cien. Natu., Esc. Superior Politécnica del Litoral, Guayaquil, Ecuador, 2021 [En línea]

- Disponible: <http://www.dspace.espol.edu.ec/handle/123456789/54068>
- [56] J. Cook y V. Ramadas, "When to consult precision-recall curves". The Stata Journal, vol. 20, n° 1, mar. 2020. Consultado: 10 may. 2025. doi: [10.1177/1536867X20909693](https://doi.org/10.1177/1536867X20909693)
- [57] M. Sokolova, N. Japkowicz y S. Szpakowicz, Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation, 1 ed. Berlín, Alemania: LNCS, 2006. [En línea] doi: [10.1007/11941439_114](https://doi.org/10.1007/11941439_114)
- [58] D. Chicco y G. Jurmen, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation". BMC Genomics", BMC Genomics, vol. 21, no 6, pp 1-13 , ene. 2020. doi: [10.1186/s12864-019-6413-7](https://doi.org/10.1186/s12864-019-6413-7)
- [59] J. S. Barajas Pérez, R. Montes Belmont, F. Castrejón Ayala, H. E. Flores Moctezuma, y M. Á. Serrato Cruz, Propiedades antifúngicas en especies del género Tagetes, Rev. Mex. Mic vol.34 , pp. 85–91, jun. 2011. Disponible: https://www.scielo.org.mx/scielo.php?pid=S0187-31802011000200010&script=sci_arttext
- [60] Z. Tang, J. Zhang, y K. Zhang, "What-is and How-to for Fairness in Machine Learning: A Survey, Reflection, and Perspective", ACM Comput Surv, vol. 55, núm. 13s, pp. 1–37, dic. 2023, doi: [10.1145/3597199](https://doi.org/10.1145/3597199).
- [61] D. Kelmasky. Estadísticas para todos: Estrategias de pensamiento y herramientas para la solución de problemas, 1a ed. Buenos Aires, Argentina: INET, 2009. [En línea]. Disponible: <http://www.bnm.me.gov.ar/giga1/documentos/EL001858.pdf>
- [62] L. da Fontoura Costa, "By Means of the Means: Arithmetic, Harmonic, Geometric ...," Jul. 2023. [Online]. Available: <https://hal.science/hal-04152919>
- [63] Z. C. Lipton, C. Elkan, y B. Narayanaswamy, "Thresholding Classifiers to Maximize F1 Score", 2014. [En línea]. Disponible en: <https://arxiv.org/abs/1402.1892>
- [4] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms", Pattern Recognit, vol. 30, núm. 7, pp.

1145–1159, jul. 1997, [doi: 10.1016/S0031-3203\(96\)00142-2](https://doi.org/10.1016/S0031-3203(96)00142-2).

Anexos

Anexo 1. Tabla de operacionalización de variables

Variable	Definición Conceptual	Definición Operacional	Dimensiones	Indicadores	Escala de medición
Machine Learning	<p>El machine learning es una rama de la inteligencia artificial que se basa en modelos y métodos que favorezcan el proceso toma de decisiones, mediante el aprendizaje automático por medio de datos para predecir con precisión, exactitud y sensibilidad el comportamiento de datos.</p> <p>(Díaz y Sanyer [55] y Ogosi [20])</p>	<p>El machine learning es el área dedicada a la construcción de modelos de aprendizaje autónomo, por medio de datos. Por lo que, para el desarrollo del algoritmo de Machine Learning se tomará en cuenta las dimensiones de <i>precisión</i>, <i>recall</i> y <i>accuracy</i> del modelo.</p>	Métrica de Precisión	<p>Precisión de modelo</p> <p>Precision = $\frac{TP}{TP+FP}$</p> <p>(Cook y ramadas [56])</p> <ul style="list-style-type: none"> • Total de positivos • Falsos positivos 	Intervalos
			Métrica recuperación de	<p>Recall del modelo</p> <p>$recall = \frac{tp}{tp + fn}$</p> <p>(Sokolova et al [57])</p> <ul style="list-style-type: none"> • Total de positivos • Falsos negativos 	
			Métrica de Accuracy	<p>Exactitud de modelo</p> <p>$accuracy = \frac{TP + TN}{n^+ + n^-} = \frac{TP + TN}{TP + TN + FP + FN}$</p> <p>(Chico y Jurman, [58])</p> <ul style="list-style-type: none"> • Total de positivos • Total de negativos • Falsos negativos • Falsos positivos 	

Evaluación Autónoma de Perfiles Laborales	<p>La evaluación autónoma de perfiles laborales comprende agilizar el proceso de toma de decisiones, mediante el análisis de datos, bajo criterios establecidos que reduzcan los sesgos y mejoren la eficiencia y precisión para la identificación del candidato más adecuado.</p> <p>(Cortina [30] y González [31])</p>	<p>La evaluación de perfiles comprende el análisis y revisión de información de los candidatos de forma autónoma. Por lo que se medirá la evaluación autónoma de perfiles laborales bajo las dimensiones de: reducción de sesgos , tiempo en el proceso de evaluación y el grado de eficiencia durante el proceso de evaluación mediante el empleo de fichas de registro</p>	Reducción de Sesgos	<p>Tasa de reducción</p> $RRT (\%) = \frac{TCT - TCt}{TCT} \times 100$ <p>(Barajas , et al.,[59])</p>
				<p>Paridad demográfica</p> $\forall a, a' \in \mathcal{A} : P(\widehat{Y} = 1 A = a) = P(\widehat{Y} = 1 A = a')$ <p>(Tang, Zhang y Zhang,[60])</p>
			Reducción de Tiempo	<p>Tasa de variación de tiempo</p> $\text{Variacion} = \frac{VF - VI}{VI} \times 100$ <p>(Kelmasky, [61])</p>
				<p>Razón</p> <p>Tiempo promedio de evaluación de perfiles</p> $\mu = \frac{1}{N} \sum_{i=1}^N s_i$ <p>(Costa, [62])</p>
				<p>Eficiencia durante el proceso de evaluación</p> <p>Rendimiento de modelo</p> $F1 = \frac{2tp}{2tp + fp + fn}$ <p>(Lipton, Elkan y</p>

				Narayanaswamy,[63])	
				Área bajo la curva	
				$AUC = \sum_i \left((1 - \beta_i \cdot \Delta\alpha) + \frac{1}{2}(\Delta(1 - \beta) \cdot \Delta\alpha) \right)$	

Anexo 2. Instrumentos de recolección de datos

Instrumento de recolección de datos para la evaluación del modelo de machine learning

Ficha de registro									
La presente ficha, tiene como objeto recolectar datos que permitan calcular cada uno de los indicadores del modelo elaborado, en las diferentes épocas de entrenamiento, a fin de garantizar la calidad del mismo.									
Nombre de Proyecto									
Plataforma de reclutamiento con Machine Learning para la Evaluación de Perfiles Laborales, Lima 2025									
Investigadores									
<ul style="list-style-type: none"> • Alcedo Javier Carlos José • Pachas Luicho Freddy Amos 									
Código de Ficha				Fecha aplicación					
Fórmulas aplicadas									
<ul style="list-style-type: none"> • Precisión $\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$ <ul style="list-style-type: none"> • Recall $\text{recall} = \frac{tp}{tp + fn}$ <ul style="list-style-type: none"> • Accuracy $\text{accuracy} = \frac{\text{TP} + \text{TN}}{n^+ + n^-} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$									
Donde:									
<ul style="list-style-type: none"> • TP: Total de positivos • TN: Total de negativos • FN: Falsos negativos • FP: Falsos positivos 									
Tabla de registro									
n° Epocas	Verdaderos Positivos	Falsos positivos	Total Positivos	Verdaderos negativos	Falsos negativo	Total negativos	Precisión	Recall	Accuracy

Comentarios									
Evidencias de fórmulas									
<p>a) Precisión</p> <p>The Stata Journal Volume 20, Issue 1, March 2020, Pages 131-148 © 2020 StataCorp LLC, Article Reuse Guidelines https://doi.org/10.1177/1536867X20909693</p> <p>Article and Columns</p> <p>When to consult precision-recall curves</p> <p>Jonathan Cook¹ and Vikram Ramadas²</p> <p>Abstract Receiver operating characteristic (ROC) curves are commonly used to evaluate predictions of binary outcomes. When there is a small percentage of items of interest (as would be the case with fraud detection, for example), ROC curves can provide an inflated view of performance. This can cause challenges in determining which set of predictions is better. In this article, we discuss the conditions under which precision-recall curves may be preferable to ROC curves. As an illustrative example, we compare two commonly used fraud predictors (Bensish's [1999, <i>Financial Analysts Journal</i> 55: 24–36] M score and Dechow et al.'s [2011, <i>Contemporary Accounting Research</i> 28: 17–82] F score) using both ROC and precision-recall curves. To aid the reader with using precision-recall curves, we also introduce the command <i>prcurve</i> to plot them.</p> <p>Keywords st0591, prcurve, precision-recall curves, classifier evaluation, ROC curves</p> <hr/> <p>¹Public Company Accounting Oversight Board Washington, DC, jacob@uci.edu ²Public Company Accounting Oversight Board Washington, DC, vramadas@ucdavis.edu</p> <p>1 Introduction Recent developments in machine learning have increased interest in predictive modeling. An important component of building a predictive model is evaluating model efficacy. For evaluating predictions of binary outcomes, receiver operating characteristic (ROC) curves are the most common tool. In this article, we discuss when it may be advisable to consult an alternative tool—precision-recall (PR) curves—and introduce a command, <i>prcurve</i>, for doing so.</p> <p>In some settings, we may be interested in predicting an outcome that is relatively rare (for example, fraud). In these settings with a rare outcome, ROC curves can be shifted outward relative to what would be found under a more balanced distribution. This outward shift can hinder comparisons of predictors. Our suggestion, and that of some recent literature (for example, Saito and Rehmsmeier [2015]), is to compare the PR plots for these predictors. There are, of course, other reasons for preferring PR curves to ROC, for example, having a loss function (or objective function) that better aligns with the output provided by the PR curve.</p> <p>2 Review of PR and ROC curves We assume that each observation belongs to one of two classes: positive or negative. In most economic applications, positive is coded as 1 and negative is coded as 0. To make predictions, we have a continuous “score”. For example, the predictive probabilities from a logistic regression could be used as a score. We do not require that the score be a probability. Instead, we focus on how our score ranks the instances. Throughout this article, we refer to a set of ordinal scores as a “classifier”.</p> <p>Our task is to evaluate how well our classifier predicts class. Given a threshold, we could predict that all observations with a value that is above the threshold are positive and all observations below the threshold are negative. To see how well the rating works in combination with the threshold predict class, we define precision, recall, and the false-positive rate as</p> <div style="border: 1px solid red; padding: 2px; display: inline-block;">Precision = $\frac{TP}{TP+FP}$</div> (1) <div style="display: inline-block; margin-left: 20px;">Recall = $\frac{TP}{P}$</div> <div style="display: inline-block; margin-left: 20px;">False-positive rate = $\frac{FP}{N}$</div> (2) <p>where the confusion matrix in table 1 defines true positives (TP), false positives (FP), negatives (N), and positives (P). In other words, precision measures how many of the items that were predicted to be positive cases are truly positive cases. Recall is the percent of positive cases that were identified. The false-positive rate is the percent of negative cases that were incorrectly predicted to be positive.</p>									

b) Recall

Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation*

Marina Sokolova¹, Nathalie Japkowicz², and Stan Szpakowicz³

¹ DIRO, Université de Montréal, Montreal, Canada

sokolova@iro.umontreal.ca

² SITE, University of Ottawa, Ottawa, Canada

nat@site.uottawa.ca

³ SITE, University of Ottawa, Ottawa, Canada

ICS, Polish Academy of Sciences, Warsaw, Poland

szpak@site.uottawa.ca

Abstract. Different evaluation measures assess different characteristics of machine learning algorithms. The empirical evaluation of algorithms and classifiers is a matter of on-going debate among researchers. Most measures in use today focus on a classifier's ability to identify classes correctly. We note other useful properties, such as failure avoidance or class discrimination, and we suggest measures to evaluate such properties. These measures – Youden's index, likelihood, Discriminant power – are used in medical diagnosis. We show that they are interrelated, and we apply them to a case study from the field of electronic negotiations. We also list other learning problems which may benefit from the application of these measures.

1 Introduction

Supervised Machine Learning (ML) has several ways of evaluating the performance of learning algorithms and the classifiers they produce. Measures of the quality of classification are built from a confusion matrix which records correctly and incorrectly recognized examples for each class. Table 1 presents a confusion matrix for binary classification, where tp are true positive, fp – false positive, fn – false negative, and tn – true negative counts.

Table 1. A confusion matrix for binary classification

Class \ Recognized	as Positive	as Negative
Positive	tp	fn
Negative	fp	tn

c) Accuracy

RESEARCH ARTICLE

Open Access



The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation

Davide Chicco^{1,*} and Giuseppe Jurman³

Abstract

Background: To evaluate binary classifications and their confusion matrices, scientific researchers can employ several statistical rates, according to the goal of the experiment they are investigating. Despite being a crucial issue in machine learning, no widespread consensus has been reached on a unified elective chosen measure yet. Accuracy and F_1 score computed on confusion matrices have been (and still are) among the most popular adopted metrics in binary classification tasks. However, these statistical measures can dangerously show overoptimistic inflated results, especially on imbalanced datasets.

Results: The Matthews correlation coefficient (MCC), instead, is a more reliable statistical rate which produces a high score only if the prediction obtained good results in all of the four confusion matrix categories (true positives, false negatives, true negatives, and false positives), proportionally both to the size of positive elements and the size of negative elements in the dataset.

Conclusions: In this article, we show how MCC produces a more informative and truthful score in evaluating binary classifications than accuracy and F_1 score, by first explaining the mathematical properties, and then the asset of MCC in six synthetic use cases and in a real genomics scenario. We believe that the Matthews correlation coefficient should be preferred to accuracy and F_1 score in evaluating binary classification tasks by all scientific communities.

Keywords: Matthews correlation coefficient, Binary classification, F_1 score, Confusion matrices, Machine learning, Biostatistics, Accuracy, Dataset imbalance, Genomics

Background

Given a clinical feature dataset of patients with cancer traits [1, 2], which patients will develop the tumor and which will not? Considering the gene expression of neuroblastoma patients [3], can we identify which patients are going to survive, and which will not? Evaluating the metagenomic profiles of patients [4], is it possible to discriminate different phenotypes of a complex disease? Answering these questions is the aim of machine learning and computational statistics, nowadays pervasive in analysis of biological and health care datasets, and

many other scientific fields. In particular, these binary classification tasks can be efficiently addressed by supervised machine learning techniques, such as artificial neural networks [5], k -nearest neighbors [6], support vector machines [7], random forest [8], gradient boosting [9], or other methods. Here the word *binary* means that the data element statuses and prediction outcomes (class labels) can be twofold: in the example of patients, it can mean healthy/sick, or low/high grade tumor. Usually scientists indicate the two classes as the negative and the positive class. The term *classification* means that the goal of the process is to attribute the correct label to each data instance (sample); the process itself is known as the classifier, or classification algorithm.

2 Commonly-accepted Performance Evaluation Measures

The vast majority of ML research focus on the settings where the examples are assumed to be identically and independently distributed (IID). This is the case we focus on in this study. Classification performance without focussing on a class is the most general way of comparing algorithms. It does not favour any particular application. The introduction of a new learning problem inevitably concentrates on its domain but omits a detailed analysis. Thus, the most used empirical measure, *accuracy*, does not distinguish between the number of correct labels of different classes:

$$\text{accuracy} = \frac{tp + tn}{tp + fp + fn + tn} \quad (1)$$

Conversely, two measures that separately estimate a classifier's performance on different classes are sensitivity and specificity (often employed in biomedical and medical applications, and in studies which involve image and visual data):

$$\text{sensitivity} = \frac{tp}{tp + fn}; \text{specificity} = \frac{tn}{fp + tn} \quad (2)$$

Focus on one class prevails in text classification, information extraction, natural language processing and bioinformatics, where the number of examples belonging to one class is often substantially lower than the overall number of examples. The experimental setting is as follows: within a set of classes there is a class of special interest (usually *positive*). Other classes are either left as is – multi-class classification – or combined into one – binary classification. The measures of choice calculated on the positive class are:

$$\text{precision} = \frac{tp}{tp + fp}; \text{recall} = \frac{tp}{tp + fn} = \text{sensitivity} \quad (3)$$

$$F\text{-measure} = \frac{(beta^2 + 1) * \text{precision} * \text{recall}}{\beta^2 * \text{precision} + \text{recall}} \quad (4)$$

This partition can be presented in a 2×2 table called *confusion matrix* $M = \begin{pmatrix} TP & FN \\ FP & TN \end{pmatrix}$ (expanded in Table 1), which completely describes the outcome of the classification task.

Clearly $TP + FN = n^+$ and $TN + FP = n^-$. When one performs a machine learning binary classification, she/he hopes to see a high number of true positives (TP) and true negatives (TN), and less false negatives (FN) and false positives (FP). When $M = \begin{pmatrix} n^+ & 0 \\ 0 & n^- \end{pmatrix}$ the classification is perfect.

Since analyzing all the four categories of the confusion matrix separately would be time-consuming, statisticians introduced some useful statistical rates able to immediately describe the quality of a prediction [22], aimed at conveying into a single figure the structure of M . A set of these functions act *classwise* (either actual or predicted), that is, they involve only the two entries of M belonging to the same row or column (Table 2). We cannot consider such measures fully informative because they use only two categories of confusion matrix [39].

Accuracy. Moving to global rates, having three or more entries of M as n , many researchers consider computing the accuracy as the standard way to go.

Accuracy, in fact, represents the ratio between the correctly predicted instances and all the instances in the dataset:

$$\text{accuracy} = \frac{TP + TN}{n^+ + n^-} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Table 1. The standard confusion matrix M

	Predicted positive	Predicted negative
Actual positive	True positives TP	False negatives FN
Actual negative	False positives FP	True negatives TN

True positives (TP) and true negatives (TN) are the correct predictions, while false negatives (FN) and false positives (FP) are the incorrect predictions

$$\text{Perfect misclassification } M = \begin{pmatrix} 0 & n^+ \\ n^- & 0 \end{pmatrix}.$$

As anticipated (Background), accuracy fails in providing a fair estimate of the classifier performance in the class-imbalanced datasets. For any dataset, the proportion of samples belonging to the largest class is called the *no-information error rate* $ni = \max(n^+, n^-)$; a binary dataset is (perfectly) balanced if the two classes have the same size, that is, $ni = \frac{1}{2}$, and it is unbalanced if one class is much larger than the other, that is $ni \gg \frac{1}{2}$. Suppose now that $ni \neq \frac{1}{2}$, and apply the trivial majority classifier: this algorithm learns only which is the largest class in the training set, and attributes this label to all instances. If the largest class is the positive class, the resulting confusion matrix is $M = \begin{pmatrix} n^+ & 0 \\ 0 & n^- \end{pmatrix}$, and thus accuracy = ni . If the dataset is highly unbalanced, $ni \approx 1$, and thus the accuracy measure gives an unreliable estimation of the goodness of the classifier. Note that, although we achieved this result by means of a trivial classifier, this is not the only possible result: as stated by Blagus and Lounsbury [99], several classifiers are biased towards the largest class in unbalanced studies.

Finally, consider another trivial algorithm, the coin tossing classifier: this classifier randomly attributes to each category, the label positive or negative with probability $\frac{1}{2}$. Applying the coin tossing classifier to any binary dataset gives an accuracy with expected value $\frac{1}{2}$, since $(M) = \begin{pmatrix} n^+/2 & n^-/2 \\ n^-/2 & n^+/2 \end{pmatrix}$.

Matthews correlation coefficient (MCC). As an alternative measure unaffected by the unbalanced datasets issue, the Matthews correlation coefficient is a contingency matrix method of calculating the *Pearson product-moment correlation coefficient* [22] between actual and predicted values. In terms of the entries of M , MCC reads as follows:

*Correspondence: davide.chicco@dal.academy

¹Kreml Research Institute, Toronto, Ontario, Canada

³Peter Munk Cardiac Centre, Toronto, Ontario, Canada

Full list of author information is available at the end of the article

Instrumento de recolección de datos para los indicadores de dimensión reducción de sesgos de evaluación

Comentarios

Evidencia de fórmulas

a) Tasa de reducción

Redalyc
Sistema de Información Científica
Red de Revistas Científicas de América Latina, el Caribe, España y Portugal

Barajas Pérez, Juan Saúl; Montes-Belmont, Roberto; Castrejón Ayala, Federico; Flores-Moctezuma, Hilda Elizabeth; Serrato Cruz, Miguel Ángel

Propiedades antifúngicas en especies del género Tagetes
Revista Mexicana de Micología, núm. 34, diciembre, 2011, pp. 83-88
Sociedad Mexicana de Micología
Xalapa, México

Disponible en: <http://www.redalyc.org/ar/Inicio/ArtPDFRed.jsp?iCve=80321339008>



Revista Mexicana de Micología
ISSN (Versión impresa): 0187-3180
gerardo.mata@inecol.edu.mx
Sociedad Mexicana de Micología
México

«Cómo citar» | «Número completo» | «Más información del artículo» | «Página de la revista»

www.redalyc.org
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

disco de 5 mm de diámetro de *S. rolfsii* o *M. fruticola* en cada experimento. Se incubaron a 25-27 °C por un máximo de 30 días.

En ambas especies se evaluó el área de crecimiento micelial hasta que el testigo llenó la caja Petri. Se tomaron fotografías digitales diariamente, y con ayuda del programa ImageJ® (versión 1.4), para análisis de imágenes (Instituto Nacional de la Salud Mental de Estados Unidos), se calculó el área de crecimiento en milímetros cuadrados. Con el programa Sigma Stat V. 3.5 se realizaron regresiones lineales para calcular la tasa de crecimiento. Con los datos de la tasa de crecimiento se calculó el porcentaje de reducción del crecimiento con respecto al tratamiento testigo (RRT) mediante la fórmula:

$$\text{RRT (\%)} = \frac{\text{TCT} - \text{TCt}}{\text{TCT}} \times 100$$

Donde: TCT= tasa de crecimiento en el testigo; TCt= tasa de crecimiento en el tratamiento.

Se cuantificó el total de esclerosones producidos por *S. rolfsii* después de 30 días de incubación.

Para evaluar la sporulación de *M. fruticola*, una vez que en el testigo se obtuvo el máximo crecimiento, se le adicionaron 2 mL de agua destilada estéril y se raspó el micelio, se obtuvo una suspensión de la que se tomaron 10 µL que se depositaron en una cámara de Neubauer (Marienfeld

Co., Alemania) y con el objetivo 40x del microscopio óptico, se realizó el conteo de esporas. Se efectuaron conteos en 4 campos microscópicos por unidad experimental en cada tratamiento.

Con los datos obtenidos (promedio de dos repeticiones) de todos los experimentos se realizaron análisis de varianza con el programa SAS versión 8.0 y se aplicó el procedimiento PROC GLM. La comparación de medias se llevó a cabo por la diferencia mínima significativa DMS.

Para aceites esenciales con *S. rolfsii* se encontró que el tratamiento con *T. filifolia* al 0.1 % fue el más notable porque tuvo un efecto fungicida sobre todos los aislamientos y en los demás tratamientos con aceites esenciales se encontró que la tasa de crecimiento se redujo respecto al testigo, pero varió entre aislamientos de *S. rolfsii* y entre tratamientos con aceites esenciales de *Tagetes* (Tabla 1). El porcentaje de reducción en la tasa de crecimiento con respecto al testigo también varió. *T. coronopifolia* presentó una inhibición mayor al 50 % en todos los aislamientos, en tanto que *T. lucida* tuvo ese mismo comportamiento en siete aislamientos y se redujo a cinco y a cuatro aislamientos en *T. erecta* y *T. foetidissima*, respectivamente. El ensuficiente Tween no interferió en el crecimiento micelial y estimuló ligeramente el crecimiento en algunos aislamientos como el 1 y el 27. Los aislamientos más sensibles fueron el 4, 5 y el 27 que tuvieron

NOTA CORRIENTE

Burgos J.S. et al. Propiedades antifúngicas en especies del género Tagetes

Tabla 1. Efecto de los aceites esenciales de *Tagetes* al 0.1%, sobre la tasa de crecimiento micelial (mm/día) de diferentes aislamientos de *S. rolfsii* y porcentaje de reducción (RRT) del crecimiento de cada tratamiento respecto al testigo

	A	T1	T2	T0.1T	T3	RRT	T4	RRT	T5	RRT	T6	RRT	T7	RRT	T8	RRT	DMS
4	16.4 ^a	15.1 ^a	7.9 ^b	0 ^c	100	11.1 ^b	12.5	0.2 ^c	52.2 ^b	9.2 ^c	4.9 ^c	10.3 ^b	37.2 ^b	2.9 ^c	68.4 ^b	2.5 ^c	
5	11.7 ^a	12.1 ^a	3.4 ^b	0	100	2.9 ^c	7.5 ^b	3.3 ^b	54.7 ^b	1.9 ^c	8.3 ^b	3.7 ^b	6.1 ^b	1.3 ^c	31.3 ^b	3.2 ^c	
12	14.5 ^a	11.1 ^a	27.6 ^b	0	100	3.5 ^c	4.8 ^b	0.1 ^c	22.4 ^b	1.9 ^c	1.4 ^c	11.3 ^b	7.7 ^b	93.8 ^b	32.2 ^b	3.2 ^c	
16	12.5 ^a	8.5 ^a	32.0 ^b	0	100	1.6 ^c	8.7 ^b	0.1 ^c	96.2 ^b	0.1 ^c	9.8 ^b	3.0 ^b	76.0 ^b	1.4 ^c	11.3 ^b	1.6 ^c	
17	12.5 ^a	10.0 ^a	10.0 ^a	0	100	13.2 ^b	12.5 ^b	0.1 ^c	52.0 ^b	0.1 ^c	5.5 ^c	11.3 ^b	5.5 ^c	1.6 ^c	61.7 ^b	2.5 ^c	
18	14.5 ^a	16.0 ^a	7.4 ^b	0	100	9.5 ^b	36.2 ^b	7.4 ^b	90.3 ^b	6.4 ^b	5.7 ^c	5.7 ^c	61.7 ^b	2.5 ^c	11.3 ^b	1.6 ^c	
20	14.5 ^a	16.0 ^a	10.0 ^a	0	100	12.4 ^b	12.5 ^b	0.1 ^c	52.0 ^b	0.1 ^c	5.5 ^c	10.3 ^b	50.0 ^b	1.4 ^c	11.3 ^b	1.6 ^c	
25	15.1 ^a	15.7 ^a	4.0 ^b	0	100	8.7 ^b	42.4 ^b	0.1 ^c	99.3 ^b	2.8 ^c	8.1 ^b	10.3 ^b	33.8 ^b	1.7 ^c	11.3 ^b	1.6 ^c	
27	14.5 ^a	15.4 ^a	10.0 ^a	0	100	7.5 ^b	47.9 ^b	0.1 ^c	94.4 ^b	2.8 ^c	8.4 ^b	3.3 ^c	53.3 ^b	2.7 ^c	49.8 ^b	2.7 ^c	
X	14.5 ^a	13.5 ^a	6.35 ^b	0	100	7.65 ^b	47.9 ^b	5.15 ^b	64.0 ^b	5.08 ^b	65.65 ^b	7.33 ^b	49.8 ^b	40.87 ^b			

*Valores seguidos de la misma letra en cada línea no difieren estadísticamente ($P>0.0001$). A: Aislamiento de *S. rolfsii*; T1: Testigo sin AE.

T2: Testigo Tween 20 al 0.02%; T3: AE de *T. filifolia*; T4: AE de *T. foetidissima*; T5: AE de *T. lucida*; T6: AE de *T. coronopifolia*; T7: AE de *T. erecta*.

**Valores no incluidos en el análisis estadístico, por ser cincuenta.

85

b) Paridad Demográfica



What-is and How-to for Fairness in Machine Learning: A Survey, Reflection, and Perspective

ZEYU TANG, Carnegie Mellon University, United States
JIJI ZHANG, The Chinese University of Hong Kong, Hong Kong
KUN ZHANG, Carnegie Mellon University, United States

We review and reflect on fairness notions proposed in machine learning literature and make an attempt to draw connections to arguments in moral and political philosophy, especially theories of justice. We survey dynamic fairness inquiries and further consider the long-term impact induced by current prediction and decision. We present a flowchart that encompasses implicit assumptions and expected outcomes of different fairness inquiries on the data-generating process, the predicted outcome, and the induced impact, respectively. We demonstrate the importance of matching the mission (what kind of fairness to enforce) and the means (which appropriate fairness spectrum to analyze) to fulfill the intended purpose.

CCS Concepts: • Computing methodologies → Artificial intelligence; Machine learning;

Additional Key Words and Phrases: Algorithmic fairness, causality, bias mitigation, dynamic process, fair machine learning

ACM Reference format:

Zeyu Tang, Jiji Zhang, and Kun Zhang. 2023. What-is and How-to for Fairness in Machine Learning: A Survey, Reflection, and Perspective. *ACM Comput. Surv.* 55, 13s, Article 299 (July 2023), 37 pages.
<https://doi.org/10.1145/3597199>

1 INTRODUCTION

With the widespread utilization of machine learning models in our daily life, researchers have been thinking about the potential social consequences of the prediction/decision made by algorithms. To date, there is ample evidence that machine learning models have resulted in discrimination against certain groups of individuals under many circumstances, for instance, the discrimination in ad delivery when searching for names that can be predictive of the race of an individual [174]; the gender discrimination in job-related ads push [47]; stereotypes associated with gender in word embeddings [22]; the bias against certain ethnic groups in the assessment of recidivism

Kun Zhang also with Mohamed bin Zayed University of Artificial Intelligence, United Arab Emirates.
The work was supported in part by the NSF Convergence Accelerator Track-D Award No. 2134901, by the National Institutes of Health (NIH) under Contract No. R01MH1515805, by grants from Apple Inc., KDDI Research, Qiris AI, and IBT, and by generous gifts from Amazon, Microsoft Research, and Salesforce. J.Z.'s research was supported in part by the RGC of Hong Kong under Grant No. 9040322.
Authors' addresses: Z. Tang and K. Zhang, Department of Philosophy, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213 USA; emails: jeyutang, kanz1@cmu.edu. J. Zhang, Department of Philosophy, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong; email: jjzhang@cuhk.edu.hk.
Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
© 2023 Copyright held by the owner/author(s).
0360-0290/2023/07-ART299 \$15.00
<https://doi.org/10.1145/3597199>

ACM Computing Surveys, Vol. 55, No. 13s, Article 299. Publication date: July 2023.

4.1 Demographic Parity

Demographic Parity, also known as Statistical Parity, is one of the earliest fairness notions proposed in the literature [25, 51, 61, 197]. In the context of binary classification ($Y = \{0, 1\}$), Demographic Parity requires that the ratio of positive decisions among different groups is equal:

$$\forall a, a' \in \mathcal{A} : P(\hat{Y} = 1 | A = a) = P(\hat{Y} = 1 | A = a') \quad (1)$$

In general contexts, Demographic Parity is characterized via the independence between the prediction \hat{Y} and the protected feature A .

Definition 4.1 (Demographic Parity). We say that a predictor \hat{Y} is fair in terms of Demographic Parity with respect to the protected feature A , if \hat{Y} is independent from A , i.e., $\text{Minipoled } A$.

While it is intuitive to characterize fairness through the aforementioned independence, the notion has significant drawbacks [51]. For instance, when there is unobjectionable dependence between the ground truth Y and the protected feature A , i.e., $Y \perp\!\!\!\perp A$ by definition the perfect predictor is also dependent on A ($\hat{Y} \not\perp\!\!\!\perp A$ since $\hat{Y} = Y$). It is not intuitive why we should rule out the perfect predictor (although this might not be achievable in reality) for the sake of satisfying the Demographic Parity fairness requirement on the prediction even if we allow $Y \perp\!\!\!\perp A$ in the data.

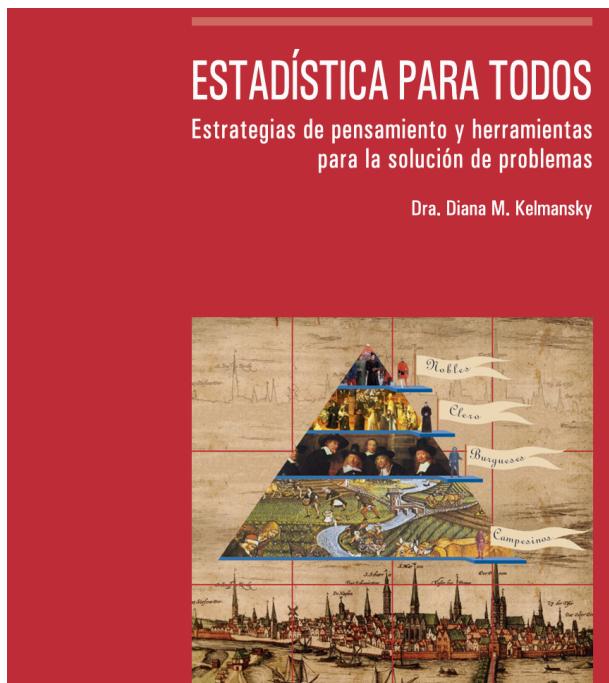
299

Instrumento de recolección de datos para los indicadores de dimensión reducción de tiempo de evaluación

Ficha de registro					
<p>La presente ficha, tiene como objeto recolectar datos que permitan calcular la reducción de tiempo efectuada en el proceso de evaluación de perfiles al aplicar el modelo elaborado.</p>					
Nombre de Proyecto					
<p>Plataforma de reclutamiento con Machine Learning para la Evaluación de Perfiles Laborales, Lima 2025</p>					
Investigadores					
<ul style="list-style-type: none"> Alcedo Javier Carlos José Pachas Luicho Freddy Amos 					
Código de Ficha		Fecha de aplicación:			
Fórmula aplicada					
<ul style="list-style-type: none"> Tiempo promedio de evaluación $\mu = \frac{1}{N} \sum_{i=1}^N s_i$ <ul style="list-style-type: none"> Variación $\text{Variacion} = \frac{\text{VF} - \text{VI}}{\text{VI}} \times 100$					
<p>Donde:</p> <ul style="list-style-type: none"> VF: Tiempo con modelo VI: Tiempo tradicional 					
Tabla de registro					
Nº de CVs	Tiempo modelo	Tiempo tradicional	Nº de CVs seleccionados	Tiempo promedio de evaluación	% de Variación

Comentarios					
Evidencia de fórmulas					
<p>a) Tiempo promedio de evaluación</p> <div style="display: flex; align-items: center;">  </div> <p>By Means of the Means: Arithmetic, Harmonic, Geometric ...</p> <p>Luciano da Fontoura Costa</p> <p>► To cite this version:</p> <p>Luciano da Fontoura Costa. By Means of the Means: Arithmetic, Harmonic, Geometric 2023 hal-04152919</p> <p>HAL Id: hal-04152919 https://hal.science/hal-04152919v1 Preprint submitted on 5 Jul 2023</p> <p>HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.</p> <p>Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License</p>					
<p>Given some numeric data (discrete or continuous), such as a set S of numbers, it is possible to assign a respective scalar value to that set, which will henceforth be considered to be real. The mapping that implements this association is often called a <i>functional</i>.</p> <p>A general functional F can be mathematically expressed as:</p> $F : f(x) \longrightarrow F(f(x)) \in \mathbb{R} \quad (7)$ <p>As already observed, the means of a set S of values are functions on S. However, not every functional is a mean, including the mode, median, extremes and the range of S, to name but a few cases among infinite possibilities.</p> <p>When considered from the perspective of statistics and probability (e.g. [2, 13, 3, 4, 5, 6, 7, 8]), the concept of mean relates directly to the <i>expectance</i> of a random variable X, which constitutes a functional that can be expressed in discrete form as:</p> $E[X] = \sum_{i=1}^N x_i p_i(x) \quad (8)$ <p>where x_i are N the possible values of the discrete random variable X and p_i corresponds to the respective relative frequencies.</p> <p>where κ is often taken as corresponding to a positive integer value.</p> <p>Of particular interest are the <i>central moments</i>, which take into account a translation by the respective expected value $E[X]$, being defined as:</p> $\begin{aligned} M_k[X] &= E[(X - E[X])^\kappa] = \\ &= \int_{-\infty}^{\infty} (x - \mu_x)^\kappa p(x) dx \end{aligned} \quad (13)$ <p>The important quantity known as <i>variance</i> of a random variable X is a particular case of a central moment, obtained for $k = 2$:</p> $\sigma_X^2 = M_2[X] = \int_{-\infty}^{\infty} (x - \mu_x)^2 p(x) dx \quad (14)$ <p>The <i>standard deviation</i> of X can then be defined as:</p> $\sigma_X = +\sqrt{\sigma_X^2} \quad (15)$ <h2>6 Arithmetic Mean</h2> <p>The arithmetic mean of a set of sampled values s_i, $i = 1, 2, \dots, N$, can be defined as follows:</p> $\mu = \frac{1}{N} \sum_{i=1}^N s_i \quad (16)$					

b) Variación porcentual



Una **tasa** (o velocidad) es un **cociente que refleja una cierta cantidad por unidad**. Por ejemplo, un automóvil se desplaza a 45 km por hora (la unidad es una hora), o la tasa de robos en un barrio, 3 robos por cada 1.000 hogares (la unidad es 1.000 hogares).

Un **porcentaje** es un **número entre 0 y 100** que mide la **proporción** de un total. Por ejemplo, cuando decimos que una camisa tiene un 10% de descuento, si el precio original (el total) es \$ 90, el descuento es de \$ 9. Si decimos que el 35% de la población está a favor de un período de cuatro días de trabajo a la semana, y la población tiene 50.000 habitantes, entonces son 17.500 ($50.000 \times 0,35 = 17.500$) los que están a favor. La proporción de los que están a favor es 0,35.

- Un porcentaje del 35% es lo mismo que una proporción de 0,35.
- Para convertir un porcentaje en una proporción, se divide al porcentaje por 100.
- Para convertir una proporción en un porcentaje, se multiplica la proporción por 100.

3.3.2 Variaciones relativas

Cuando un porcentaje se utiliza para determinar un aumento o reducción relativa (relativa al valor inicial), se denomina **variación porcentual**.

Supongamos que la cantidad de accidentes por año en una ciudad pasó de 50 a 60, mientras que la cantidad de accidentes en otra ciudad pasó de 500 a 510. Ambas ciudades tuvieron un **aumento** de 10 accidentes por año, pero para la primera ciudad, esta diferencia como porcentaje del número inicial de accidentes, es mucho mayor.

Variación porcentual: se toma el valor "después de" y se le resta el "antes de", luego se divide ese resultado por el "antes de". Así, se obtiene una proporción. Para transformarla en un porcentaje se multiplica el resultado por 100.

Para la primera ciudad, esto significa que la cantidad de accidentes aumentó en un

Instrumento de recolección de datos para los indicadores de la dimensión eficiencia de evaluación

Ficha de registro					
<p>La presente ficha, tiene como objeto recolectar datos que permitan calcular el nivel de eficiencia en la evaluación que conlleva la aplicación del modelo de machine learning.</p>					
Nombre de Proyecto					
Plataforma de reclutamiento con Machine Learning para la Evaluación de Perfiles Laborales, Lima 2025					
Investigadores					
<ul style="list-style-type: none"> • Alcedo Javier Carlos José • Pachas Luicho Freddy Amos 					
Código de Ficha		Fecha aplicación			
Fórmulas aplicadas					
<ul style="list-style-type: none"> • Rendimiento de modelo (F1 Score) $F1 = \frac{2tp}{2tp + fp + fn}.$ <p>Donde:</p> <ul style="list-style-type: none"> - TP: Total de positivos - FN: Falsos negativos - FP: Falsos positivos <ul style="list-style-type: none"> • Área bajo la curva $AUC = \sum_i \left((1 - \beta_i \cdot \Delta \alpha) + \frac{1}{2} (\Delta(1 - \beta) \cdot \Delta \alpha) \right)$					
Tabla de registro					
n° Epochs	Falsos positivos	Total Positivos	Falsos negativo	F1 Score	Área Bajo la Curva

Comentarios					
Evidencias de fórmulas					
<p>a) Rendimiento</p> <div style="border: 1px solid black; padding: 10px;"> <p>Thresholding Classifiers to Maximize F1 Score</p> <p>Zachary C. Lipton, Charles Elkan, and Balakrishnan Narayanaswamy University of California, San Diego, La Jolla, California, 92093-0404, USA {zlipton,celkan,muralib}@cs.ucsd.edu</p> <p>Abstract. This paper provides new insight into maximizing F1 scores in the context of binary classification and also in the context of multilabel classification. The harmonic mean of precision and recall, F1 score is widely used to measure the success of a binary classifier when one class is rare. Micro average, macro average, and per instance average F1 scores are used in multilabel classification. For any classifier that produces a real-valued output, we derive the relationship between the best achievable F1 score and the decision-making threshold that achieves this optimum. As a special case, if the classifier outputs are well-calibrated conditional probabilities, then the optimal threshold is half the optimal F1 score. As another special case, if the classifier is completely uninformative, then the optimal behavior is to classify all examples as positive. Since the actual prevalence of positive examples typically is low, this behavior can be considered undesirable. As a case study, we discuss the results, which can be surprising, of applying this procedure when predicting 26,853 labels for Medline documents.</p> <p>Keywords: machine learning, evaluation methodology, F1-score, multilabel classification, binary classification</p> <h3>1 Introduction</h3> <p>Performance metrics are useful for comparing the quality of predictions across systems. Some commonly used metrics for binary classification are accuracy, precision, recall, F1 score, and Jaccard index [15]. Multilabel classification is an extension of binary classification that is currently an area of active research in supervised machine learning [18]. Micro averaging, macro averaging, and per instance averaging are three commonly used variants of F1 score used in the multilabel setting. In general, macro averaging increases the impact on final score of performance on rare labels, while per instance averaging increases the importance of performing well on each example [17]. In this paper, we present theoretical and experimental results on the properties of the F1 metric.¹</p> </div>					

Fig. 1: Confusion Matrix

of each label applying to each instance given the feature vector. For a batch of data of dimension $n \times d$, the model outputs an $n \times m$ matrix C of probabilities. In the single-label setting, $m = 1$ and C is an $n \times 1$ matrix, i.e. a column vector. A decision rule $D(C) : \mathbb{R}^{n \times m} \rightarrow \{0, 1\}^{n \times m}$ converts a matrix of probabilities C to binary predictions P . The gold standard $G \in \mathbb{R}^{n \times m}$ represents the true values of all labels for all instances in a given batch. A performance metric M assigns a score to a prediction given a gold standard:

$$M(P|G) : \{0, 1\}^{n \times m} \times \{0, 1\}^{n \times m} \rightarrow \mathbb{R} \in [0, 1].$$

The counts of true positives tp , false positives fp , false negatives fn , and true negatives tn are represented via a confusion matrix (Figure 1).

Precision $p = tp/(tp + fp)$ is the fraction of all positive predictions that are true positives, while recall $r = tp/(tp + fn)$ is the fraction of all actual positives that are predicted positive. By definition the F1 score is the harmonic mean of precision and recall: $F1 = 2/(1/r + 1/p)$. By substitution, F1 can be expressed as a function of counts of true positives, false positives and false negatives:

$$F1 = \frac{2tp}{2tp + fp + fn}. \quad (1)$$

The harmonic mean expression for F1 is undefined when $tp = 0$, but the translated expression is defined. This difference does not impact the results below.

2.1 Basic Properties of F1

Before explaining optimal thresholding to maximize F1, we first discuss some properties of F1. For any fixed number of actual positives in the gold standard, only two of the four entries in the confusion matrix (Figure 1) vary independently. This is because the number of actual positives is equal to the sum $tp + fn$ while the number of actual negatives is equal to the sum $tn + fp$. A second basic property of F1 is that it is non-linear in its inputs. Specifically, fixing the number fp , F1 is concave as a function of tp (Figure 2). By contrast, accuracy is a linear function of tp and tn (Figure 3).

As mentioned in the introduction, F1 is asymmetric. By this, we mean that the score assigned to a prediction P given gold standard G can be arbitrarily different from the score assigned to a complementary prediction P^c given complementary gold standard G^c . This can be seen by comparing Figure 2 with Figure 5. This asymmetry is problematic when both false positives and false negatives are costly. For example, F1 has been used to evaluate the classification of tumors as benign or malignant [1], a domain where both false positives and false negatives have considerable costs.

¹ For concreteness, the results of this paper are given specifically for the F1 metric and its multilabel variants. However, the results can be generalized to $F\beta$ metrics for $\beta \neq 1$.

b) Área bajo la curva

The Use of the Area Under the ROC Curve in the
Evaluation of Machine Learning Algorithms

Andrew P. Bradley*

method is to plot $P(T_p)$ against $P(F_p)$ as the *decision threshold* is varied. Selecting the operating point (decision threshold) that most closely meets the requirements for $P(F_n)$ and $P(F_p)$. The plotted values of $P(T_p)$ and $P(F_p)$ as the decision threshold is varied is called a Receiver Operating Characteristic (ROC) curve.

There is still, however, a problem with specifying performance in terms of a single operating point (usually a $P(T_p)$, $P(T_n)$ pair), in that there is no indication as to how these two measures vary as the decision threshold is varied. They may represent an operating point where sensitivity ($P(T_p)$) can be increased with little loss in specificity ($P(T_n)$), or they may not. This means that the comparison of two systems can become ambiguous. Therefore, there is a need for a *single* measure of classifier performance (often termed accuracy, but not to be confused with $P(C)$) that is invariant to the decision criterion selected, prior probabilities, and is easily extended to include cost/benefit analysis. This paper describes the results of an experimental study to investigate the use of the area under the ROC curve (AUC) as such as a measure of classifier performance.

When the decision threshold is varied and a number of points on the ROC curve ($P(F_p) = \alpha$, $P(T_p) = 1 - \beta$) have been obtained the simplest way to calculate the area under the ROC curve is to use trapezoidal integration,

$$AUC = \sum_i \left((1 - \beta_i) \cdot \Delta\alpha + \frac{1}{2} (\Delta(1 - \beta) \cdot \Delta\alpha) \right). \quad (7)$$

Where,

$$\Delta(1 - \beta) = (1 - \beta_i) - (1 - \beta_{i-1}), \quad (8)$$

$$\Delta\alpha = \alpha_i - \alpha_{i-1}. \quad (9)$$

*The author is with the Cooperative Research Centre for Sensor Signal and Information Processing (CSSIP) at the Dept. of Electrical and Computer Engineering, The University of Queensland, QLD 4072, Australia.
E-mail bradley@elec.uq.edu.au.

Anexo 3. Reporte de similitud en software Turnitin

Feedback Studio - Personal: Microsoft Edge
https://ev.turnitin.com/app/carta/es/?lang=es&u=1118232096&ro=103&s=1&o=2698009564&student_user=1

feedback studio CARLOS JOSE ALCEDO JAVIER | LIMA-NORTE_PI_ALCEDO_PACHAS (1).docx

Resumen de coincidencias 13 %

Universidad César Vallejo

FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

Plataforma de Reclutamiento 4.0: Aplicación de Machine Learning para la Evaluación Autónoma de Perfiles Laborales, Lima, 2025

Autor(es):
Alcedo Javier Carlos José ([orcid.org/0000-0002-7248-6949](#))
Pachas Lucho Freddy Amós ([orcid.org/0000-0001-7704-9150](#))

Asesor(a)s:
Dr. Daza Vergaray, Alfredo ([orcid.org/0000-0002-2259-1070](#))

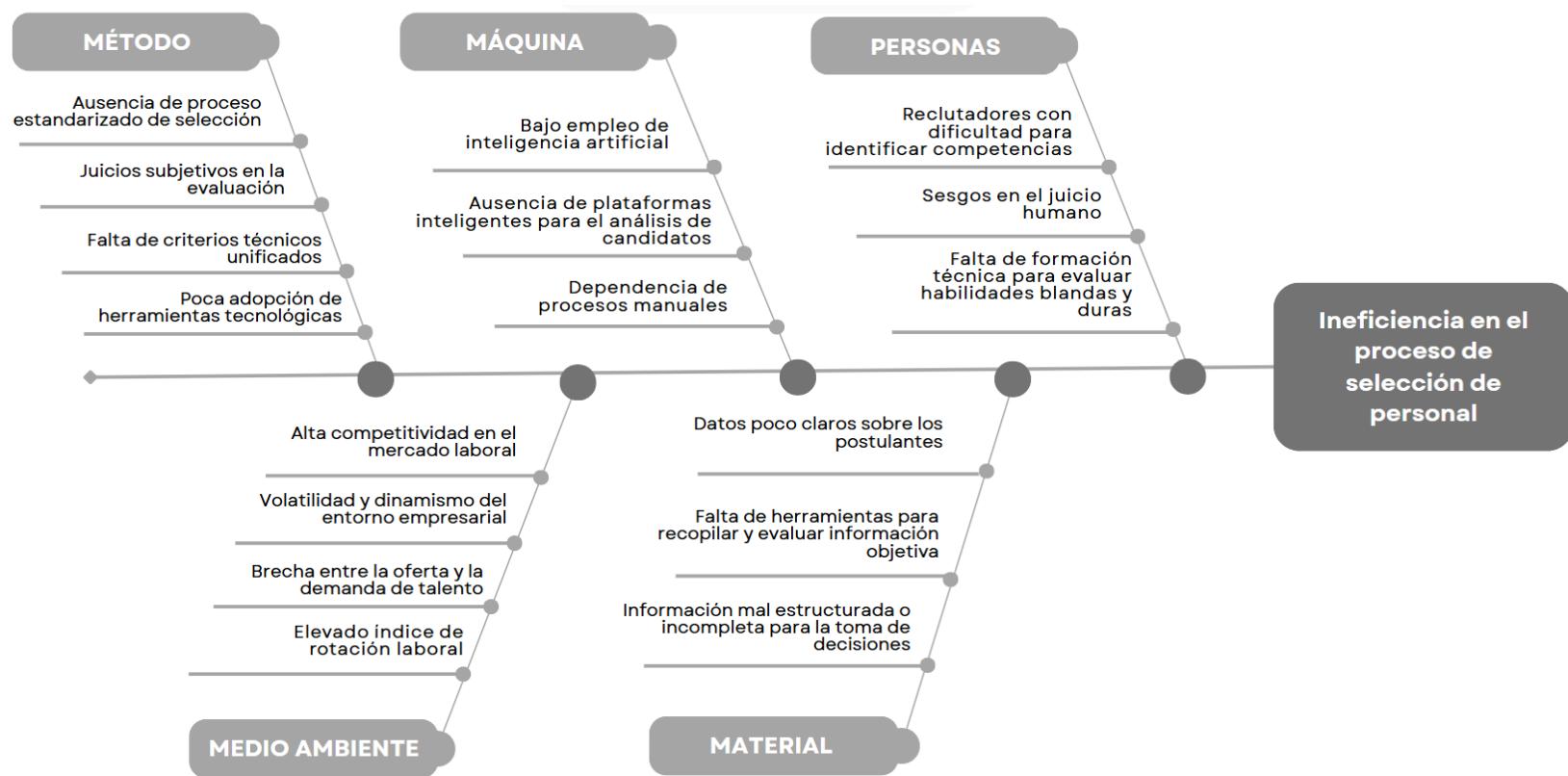
Línea de investigación:
Sistema de información y comunicaciones

Línea de responsabilidad social universitaria:
Desarrollo económico, empleo y emprendimiento

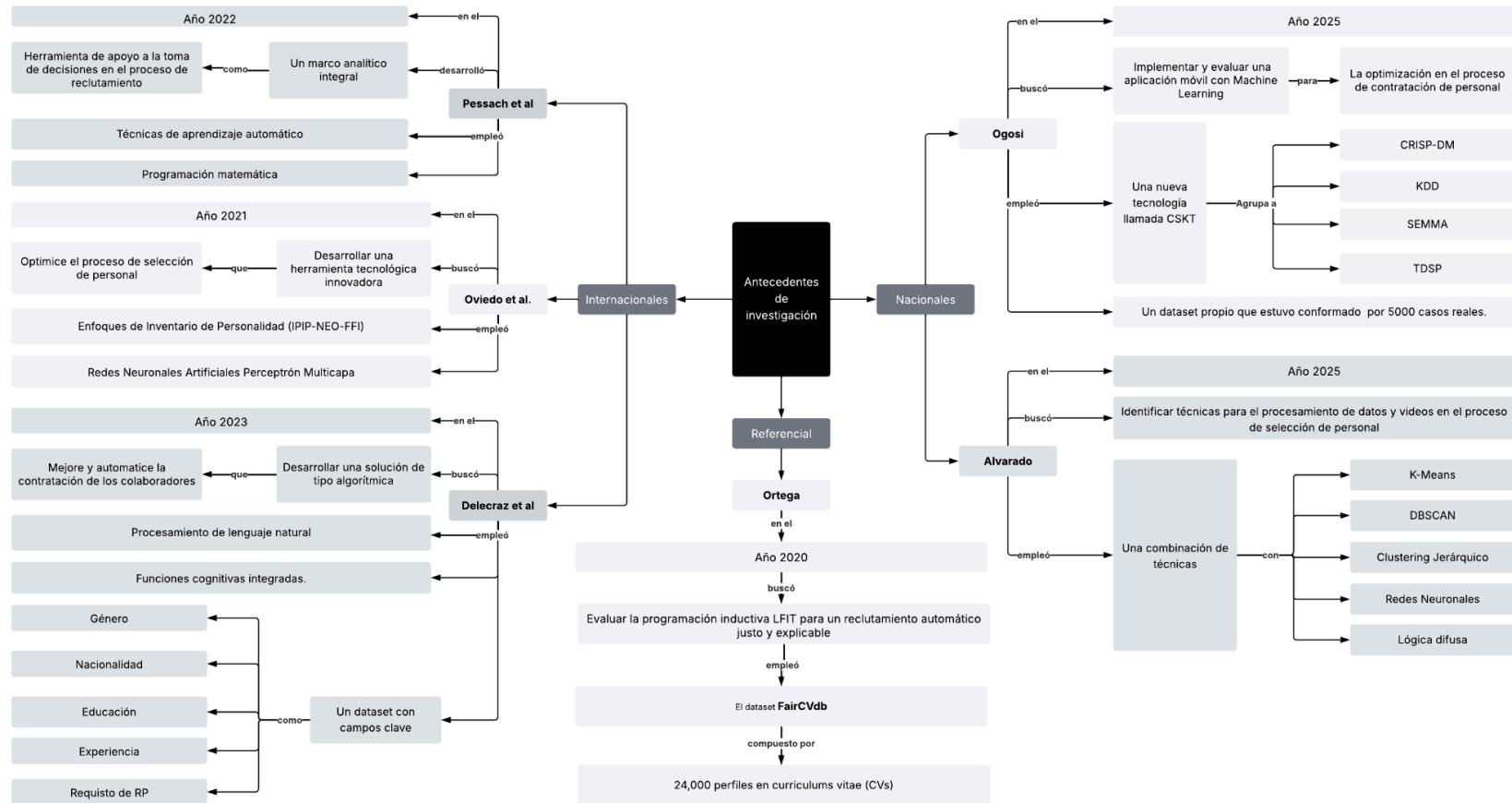
LIMA-PERÚ
2025

RANK	SOURCE	COINCIDENCE (%)
1	Entregado a Universidad...	6 %
2	www.coursehero.com	2 %
3	repositorio.ucv.edu.pe	1 %
4	tesis.usat.edu.pe	<1 %
5	Alejandro Castro Solan...	<1 %
6	"Inter-American Yearbo...	<1 %
7	Entregado a Universidad...	<1 %
8	ane-environment.net	<1 %
9	catalogo.uns.edu.ar	<1 %
10	Entregado a Universidad...	<1 %

Anexo 4. Diagrama de ishikawa



Anexo 5. Mapa mental de Antecedentes de Investigación



Anexo 6. Mapa mental de Marco Teórico de Investigación

