



Implementación de una aplicación móvil con Machine Learning aplicando la nueva metodología CSKT para el proceso de contratación de personal en el sector privado

Implementation of a mobile application with Machine Learning applying the new CSKT methodology for the recruitment process of personnel in the private sector

 José A. Ogosi-Auqui
jogosi@unfv.edu.pe
Universidad Nacional Federico Villarreal, Perú

Resumen

Este trabajo se evalúa una Máquina modelo de Aprendizaje desarrollado utilizando el CSKT metodología para mejorar la asignación de una pensión en proyectos de ingeniería. Se aborda los retos en la adopción de tales modelos, siguiendo un enfoque estructurado a partir de la comprensión del negocio el modelo de presentación. El modelo de árbol de decisión inicial de la fase de entrenamiento produce una F1-Score de 0.699825, indicando equilibrado de precisión y recall. Individuales de precisión es 0.650974, y el recuerdo es 0.756603, lo que demuestra la predicción exacta de los positivos. La curva ROC tiene un valor de 0.739770, la evaluación de la capacidad discriminatoria del modelo. Los resultados incluyen la definición de los parámetros del modelo, la creación de un árbol de decisión, y la realización inicial de las simulaciones. La conclusión destaca CSKT la eficacia en la asignación de una pensión y las tensiones de continua colaboración del cliente en proyectos de ingeniería. Los resultados numéricos de relieve el modelo de la contribución a la mejora de la asignación de una pensión de procesos en proyectos de ingeniería.

Palabras claves: Aprendizaje automático, metodología CSKT, aplicaciones móviles, gestión de contratación de personal.

Abstract


This paper evaluates a Machine Learning model developed using the CSKT methodology to improve pension allocation in engineering projects. It addresses the challenges in adopting such models by following a structured approach from understanding the business model to presenting it. The initial decision tree model training phase yields an F1-Score of 0.699825, indicating balanced precision and recall. Individual precision is 0.650974, and recall is 0.756603, demonstrating accurate prediction of positive outcomes. The ROC curve has a value of 0.739770, assessing the discriminatory ability of the model. The results include defining the model parameters, creating a decision tree, and performing initial simulations. The conclusion highlights CSKT's effectiveness in pension allocation and the stresses of ongoing client collaboration in engineering projects. The numerical results highlight the model's contribution to improving pension allocation processes in engineering projects.

Keywords: Machine learning, CSKT methodology, mobile applications, recruitment management.



Publicado: 18/06/2024
Aceptado: 14/06/2024
Recibido: 12/03/2024

Open Access
Article scientific

 <https://doi.org/10.47422/jstri.v5i1.46>

Este artículo es publicado por la Journal of Scientific and Technological Research Industrial, Professionals On Line SAC. Este es un artículo de acceso abierto, distribuido bajo los términos de la Licencia Creative Commons Atribución 4.0 Internacional (<https://creativecommons.org/licenses/by/4.0/>) que permite compartir (copiar y redistribuir el material en cualquier medio o formato) y adaptar (remezclar, transformar y construir a partir del material) para cualquier propósito, incluso comercialmente.





INTRODUCCIÓN

En la actualidad, el desafío de reclutamiento de los individuos a menudo conduce a complicaciones, ya sea debido a los criterios establecidos por las empresas en función de sus necesidades específicas o debido a las restricciones relacionadas. Como se señaló en un informe sobre la Discriminación por razones de edad [1], que se adentra en un estudio sobre la materia, pone de relieve las luchas que enfrentan los trabajadores de entre 40 y anteriormente durante el proceso de contratación. El informe indica que el 54% de estas personas sufren discriminación cuando el proceso de solicitud se basa en correos electrónicos y llamadas telefónicas. Además, el 48% afirma que la discriminación se produce durante las entrevistas de trabajo llevado a cabo dentro de las organizaciones, mientras que el 28% afirman que es menos frecuente cuando se considera la promoción de los empleados existentes. A pesar del potencial de las personas jóvenes a ser los roles asignados, es importante tener en cuenta que esto no garantiza su idoneidad para el cargo en términos de reclutamiento inicial o posterior de las promociones [2].

La pandemia ha sumido en la incertidumbre de la duración de sus efectos, en especial, sobre contratación de personal en el sector privado, que se enfrenta a la rigidez. Por el contrario, el sector público ha mostrado la agilidad en la contratación, la rápida integración de los trabajadores con prestaciones por cesantía. La actividad económica sufre de contratación y de las relaciones laborales, la rigidez, agravada por las restricciones reglamentarias y una reticencia a las metodologías ágiles. Perú el marco legal de los somete a escrutinio en medio de la condición laboral de cambios, que afectan a la contratación de los empleados la flexibilidad. La demanda para el personal de seguridad sobretensiones, especialmente durante el fin de año de las campañas, que requiere de personal capacitado, especialmente en el comercio al por menor, según lo informado por Eulen grupo de Perú. El Instituto de Economía y Desarrollo Empresarial destaca un aumento significativo en el sector formal del empleo de los jóvenes desde el mes de Marzo, impactando de selección de candidatos. Sobre el sistema de pensiones, más del 25% de los Peruanos de la fuerza laboral está afiliada, sin embargo, carecen de suficientes contribuciones o de ahorros para la jubilación. Aproximadamente el 60% de SPP afiliados no han ahorrado lo suficiente para una vida de la anualidad. Sólo el 13,7% de la población económicamente activa a los Peruanos entre 15 y 64 años realmente puede aspirar a una pensión de jubilación en el SNP o SPP en la vejez. Existen legítimas razones para la presencia de empleo, y podría manifestarse durante las horas de trabajo y contribuir a la satisfacción de los empleados. Sin embargo, en la práctica,

la conciencia del tiempo se intensifica, y el ambiente de trabajo se vuelve menos adaptable, que conduce a un aumento de la carga sobre los empleados y el personal de la empresa. Esto se atribuye a la contratación y el despido de prácticas que disminuyen la competencia y fomentar la informalidad. En consecuencia, la nación se caracteriza por tener uno de los más exigentes de los sistemas que emplean, según lo indicado por los datos compilados [3]. En la gestión de los recursos humanos de nivel, el Perú es actualmente limitada a las tareas de transacción, centrado principalmente en la nómina de la generación. Sin embargo, el talento humano en el sistema de las necesidades del país para avanzar hacia actividades de transformación. Esto implica de forma proactiva anticipar y atender las necesidades de los ciudadanos, más allá de un enfoque reactivo. La realización de esta transformación requiere de la integración de la tecnología para mejorar y apoyar el desarrollo y desempeño de las actividades de recursos humanos, como se resalta en el estudio [4].

Este estudio explora los desafíos que enfrentan los proyectos de ingeniería en la aplicación de los modelos de Aprendizaje automático, como el campo cada vez más adopta un enfoque tecnológico. Para solucionar esto, la CSKT metodología se introdujo, proporcionando un marco estructurado para la efectiva implementación del modelo. Este método tiene como objetivo mejorar la eficiencia, la información de apoyo para la adopción de decisiones, y mejorar en general la calidad de los proyectos mediante la utilización de la Máquina de Aprendizaje predictivo y capacidades analíticas. Dada la complejidad de la labor del paisaje y de las deficiencias en el sistema de pensiones, un enfoque holístico es necesario. Esta investigación se centra en la implementación y evaluación de una aplicación móvil utilizando la Máquina de Aprendizaje y la CSKT metodología para optimizar el sector privado en los procesos de contratación. Más allá de abordar temas como la discriminación y la contratación de la inflexibilidad, el enfoque faculta a los trabajadores para tomar decisiones de pensiones de decisiones. Mediante la integración de soluciones innovadoras en el trabajo de parto y de las relaciones financieras, la investigación tiene por objetivo adaptar la CSKT metodología para un Aprendizaje automático.

El uso de la Máquina de Aprendizaje (ML) en los proyectos de ingeniería, ha adquirido importancia debido a su capacidad para resolver tareas complejas, el uso de algoritmos. ML es una subdisciplina de la inteligencia artificial, que se destaca por su adaptabilidad y capacidad para interpretar datos nunca antes visto [5]. También destaca la diferencia fundamental entre la programación tradicional y ML, siendo este último esencial para el



análisis de grandes conjuntos de datos y la detección de patrones objetivamente [6].

Varios estudios han demostrado la aplicabilidad de ML en diferentes campos [7]. Por ejemplo, el aprendizaje automático puede predecir observables características microscópicas tales como la dureza y la flexibilidad de los materiales [8]. Además, la máquina de aprendizaje, mejora el rendimiento de los programas de ordenador a través de la experiencia, siendo especialmente útil para grandes y complejos conjuntos de datos [9].

Otro aspecto crucial de la integración de la Máquina de Aprendizaje en los proyectos de ingeniería es la elección de los algoritmos. Los algoritmos de aprendizaje supervisado, tales como los árboles de decisión, máquinas de soporte vectorial, y las redes neuronales, son ampliamente utilizados debido a su capacidad para modelar las relaciones entre las variables de entrada y salida [10]. Estos algoritmos son particularmente eficaces cuando la etiqueta de datos es abundante, ya que pueden hacer predicciones precisas y aprender de los ejemplos más conocidos [11]. Por otro lado, la supervisión de las técnicas de aprendizaje, como de la agrupación y reducción de dimensionalidad, son esenciales cuando se carece de datos explícitos etiquetas. Estos métodos descubrir relaciones y patrones ocultos en los datos, lo cual es especialmente valioso en el exploratorio de las fases de los proyectos de ingeniería [12].

La creciente tendencia de emplear ML en ingeniería es más apoyada por los avances en la potencia de cálculo y la disponibilidad de grandes volúmenes de datos. Las plataformas de Cloud computing, como AWS y en la Nube de Google, proporcionar escalables infraestructuras que permiten a los ingenieros de tren de modelos complejos sin la necesidad de costosas de hardware local [13]. Por otra parte, la integración de ML con otras tecnologías como el Internet de las Cosas (IoT) ha ampliado sus aplicaciones, permitiendo a los datos en tiempo real y análisis predictivo de mantenimiento en sectores como la manufactura, la energía y el transporte [14].

El CSKT metodología es un marco innovador que optimiza la aplicación de ML modelos en proyectos de ingeniería. Esta metodología sigue un proceso iterativo y de ciclo adaptativo que se ajusta a las necesidades de la empresa y la calidad de los datos disponibles. CSKT ha sido validado por la Universidad Internacional de La Rioja, en el Perú, destacando su eficacia en la mejora de los procesos de contratación en el sector privado [15].

La implementación de ML modelos se enfrenta a importantes desafíos, tales como la calidad de los datos y la necesidad de un amplio preprocesamiento. La importancia de las métricas tales como la precisión, recall,

y F1 score se destacó en la evaluación de desempeño del modelo [12, 13]. Estos indicadores proporcionan una comprensión detallada del modelo de capacidad para identificar los positivos y negativos de instancias correctamente.

El estudio también aborda el impacto de ML en la administración de pensiones, poniendo de relieve cómo el CSKT metodología puede mejorar la asignación de una pensión en los proyectos de ingeniería en el Perú. Los resultados iniciales muestran una mejora en la precisión y la capacidad de predicción del modelo, lo que subraya la importancia de la continuación de la colaboración con los clientes y la optimización de ML algoritmos.

La pregunta de investigación es: "¿Cómo puede una Máquina modelo de Aprendizaje desarrollado utilizando el CSKT metodología de mejorar la asignación de una pensión en los proyectos de ingeniería en el Perú?" Los objetivos son: en primer lugar, para evaluar la efectividad del modelo en la mejora de la asignación de una pensión; en segundo lugar, para medir el desempeño de la modelo utilizando métricas tales como F1-Score, la precisión y la curva ROC; y en tercer lugar, para abordar los desafíos en la adopción de los modelos de Aprendizaje automático en proyectos de ingeniería, de la comprensión del contexto de negocios para presentar el modelo.

MATERIALES Y MÉTODO

La Máquina De Aprendizaje

La máquina de aprendizaje es un subconjunto de la inteligencia artificial, que resuelve tareas complejas, el uso de algoritmos. La calidad de los datos es crucial en este proceso y la capacidad de adaptación y generalización de la máquina de aprendizaje, que va más allá de la programación tradicional, se pone de relieve [5].

La diferencia fundamental entre la programación tradicional y la máquina de aprendizaje [6] también se pone de relieve: la última capacidad para interpretar datos nunca antes visto. De aprendizaje de la máquina también es esencial a la hora de analizar grandes y complejos conjuntos de datos, lo que permite la identificación de patrones de una manera objetiva [7].

Aprendizaje automático puede no sólo predicen la existencia de un compuesto, sino también directamente anticipar varias observables características microscópicas, tales como la dureza, ductilidad, tenacidad, y la temperatura de Curie [8]. En su trabajo de 2009, se ilustra el uso de PCA para la replicación de mapas estructurales que representan la correlación entre la estructura y las propiedades (incluyendo las propiedades y los parámetros de la estructura cristalina) a través de la aplicación de la



minería de datos. Posteriormente, en el año 2018, se utilizó un Bosque Aleatorio (RF) algoritmo para explorar las características como la capacidad de calor, vibración, la entropía y la energía libre del espectro. Este enfoque dirigido a mejorar la exactitud de la predicción de finito de temperatura estabilidad termodinámica. Además, el modelo es adaptable a incluir aleación de condiciones como parte de la función de vector para predecir el procesamiento de los resultados.

Similar a los árboles de decisión, Bosque Aleatorio y PCA, existen otros algoritmos de aprendizaje automático, tales como Máquinas de Vectores Soporte (SVM), que se utilizan para la clasificación y regresión, la búsqueda de la óptima hiperplano que separa los datos con la mayor margen, y el uso de técnicas tales como núcleo truco para manejar no lineal de datos [16]; K-Vecinos más cercanos (K-NN), que clasifica o retrocede de valores basados en la K más cercano puntos de datos en el espacio de características, utilizando la distancia de medidas tales como la Euclidiana [17]; Naive Bayes, una familia probabilístico de los clasificadores que se basan en el teorema de Bayes con una suposición de independencia entre características, siendo especialmente eficaz en el texto problemas de clasificación [18]; y K-means Clustering, un método de aprendizaje no supervisado que los grupos de datos en K grupos, la minimización de la variación dentro de los grupos y de la maximización de la variación entre los grupos de iterar sobre los centroides y los datos de las asignaciones [19].

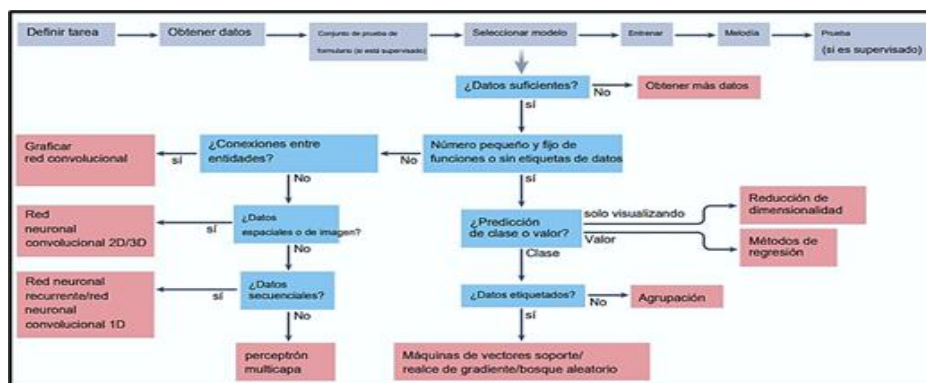
En comparación con los árboles de decisión [20], que utiliza una estructura jerárquica para dividir los datos en subconjuntos basados en reglas de decisión, los otros mencionados algoritmos tienen diferentes enfoques. Máquinas de Vectores soporte (SVM) la búsqueda de un óptimo hiperplano para maximizar el margen entre las clases, siendo más eficaz en la alta dimensionalidad de los problemas y no lineales de datos gracias a la utilización de

núcleos. K-Vecinos más cercanos (K-NN) no se puede construir un modelo explícito, sino que clasifica los datos basados en sus vecinos más cercanos en el espacio de características, lo cual puede resultar en costosas computacionalmente rendimiento con grandes conjuntos de datos. Naive Bayes asume la independencia entre las funciones, que puede ser una limitación, pero también permite una rápida y eficaz de la clasificación, especialmente en el texto de los problemas. K-means clustering, a diferencia de los métodos supervisados, es un algoritmo de agrupamiento no supervisado que no requieren etiquetas de clase y tiene como objetivo dividir los datos en grupos homogéneos basados en la proximidad de los centroides. Mientras que los árboles de decisión son interpretables y fáciles de visualizar, los otros métodos ofrecen ventajas en términos de complejo manejo de los datos, predicciones probabilísticas, o de agrupamiento no supervisado, dependiendo del problema específico [20].

Aprendizaje automático (ML) generalmente se refiere a mejorar el rendimiento de un programa de ordenador a través de la experiencia en la realización de determinados tipos de tareas [9]. El término "aprendizaje automático" en términos generales abarca el proceso de ajuste de los modelos de predicción de los datos y la detección de clústeres de la información dentro del conjunto de datos [10]. En el ámbito de aprendizaje de la máquina, cálculos se empleó principalmente para aproximar o emular la capacidad humana para identificar patrones, aunque de una manera objetiva. La máquina de aprendizaje resulta especialmente valiosa cuando se trata con conjuntos de datos que son demasiado extensas, involucrando a numerosos puntos de datos individuales, o complejo, que abarca una multitud de características para un efectivo análisis humano. Este enfoque establece una replicable y eficaz flujo de trabajo a lo largo del tiempo. Los pasos involucrados en la máquina de entrenamiento de las técnicas de aprendizaje se ilustran en la siguiente Figura 1.

Figura 1

Máquina de la guía de aprendizaje para los biólogos [10]





estructura y adaptación de la metodología que permite la eficiente aplicación de los modelos de Aprendizaje automático en proyectos de ingeniería. Esta metodología se centra en la comprensión del negocio, la preparación y selección de datos, la creación y modelos de formación, garantizando la continuidad de la colaboración con el cliente y la mejora continua del modelo.

En la comparación de metodologías CRISP-DM es ampliamente aceptado y estructurado en seis fases de un proceso iterativo, centrándose en la comprensión del negocio, compresión de datos, preparación de datos, modelado, evaluación e implementación. Su principal ventaja radica en la claridad de su estructura, a pesar de que puede ser rígido y menos adaptable a los cambios rápidos. SEMMA, centrado en la manipulación de datos y análisis para el modelado predictivo, hace hincapié en la preparación de los datos y la exploración, sino que pone menos énfasis en la comprensión del negocio y la colaboración del cliente. KDD es un fin-a-fin del proceso de descubrimiento de conocimiento en bases de datos, haciendo hincapié en la extracción de información útil, pero puede ser complejo y menos práctico para la implantación del modelo. TDP es un proceso de colaboración diseñado para la ciencia de datos de los equipos, desde la planificación del proyecto para el modelo de entrega, promoviendo el trabajo en equipo y la herramienta de integración, pero puede ser más adecuado para equipos más grandes y complejos proyectos.

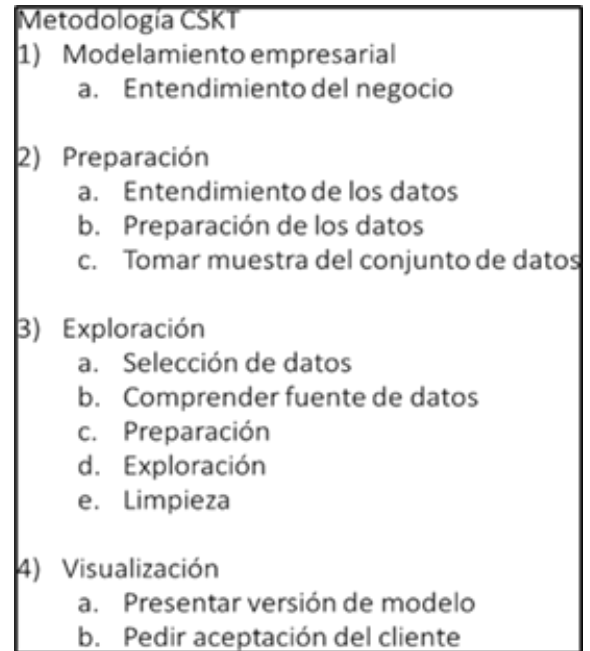
Con respecto a la justificación para la selección de la CSKT metodología, se destaca por su adaptabilidad y enfoque iterativo, que permite una rápida y continua de los ajustes basados en las necesidades del cliente y la calidad de los datos disponibles. A diferencia de otras metodologías, CSKT hace hincapié en la colaboración constante con el cliente y mejora continua del modelo, lo cual es crucial para la ingeniería de proyectos en los que los requisitos pueden cambiar rápidamente. Además, su estructura clara y definida de las fases de asegurar una aplicación eficiente y efectiva de la Máquina modelo de Aprendizaje, optimizando tanto el sector privado en el proceso de contratación y la asignación de una pensión.

Posteriormente, la versión del modelo que se presenta, y si los resultados son factibles, aceptación del cliente se solicita, por lo tanto, el cierre de la CSKT metodología de proceso. De lo contrario, volvemos a la comprensión de los datos hasta que se cumplen los requisitos previos. Cabe señalar que la Universidad Internacional de La Rioja en el Perú [15] aceptó el diseño de esta metodología. Este enfoque integral de la CSKT metodología busca optimizar el proceso de contratación en el sector privado a través de un proceso iterativo y de ciclo adaptativo que se ajusta a las

necesidades de la empresa y la calidad de los datos disponibles.

Figura 4

La estructura de la CSKT metodología [15]



En la figura 4, una evaluación exhaustiva de la actual organización del proceso de contratación se lleva a cabo para identificar las áreas clave de mejora a través de la integración de la CSKT metodología. Este enfoque comienza con un profundo conocimiento de la empresa, llevando a una fase de toma de decisiones. Si las nuevas ideas son descubiertos, están perfectamente integrados en el proyecto. De lo contrario, el foco se desplaza a la comprensión de los datos y preparación, con condicionales controles para garantizar alteraciones necesarias. Modificación de los puntos de datos registrados, y una significativa muestra se extrae del conjunto de datos.

Si no se requieren cambios en la preparación de los datos, el proceso implica una comprensión detallada de la fuente de datos a través de una meticulosa preparación, exploración y limpieza para obtener una clara y refinado conjunto de datos. Refinado de los puntos de datos seleccionados, que conduce a la adquisición de una muestra significativa. La metodología, a continuación, avanza a presentar el modelo de la versión. Si los resultados son viables, aceptación del cliente se solicita, concluyendo el proceso. Sin embargo, si los resultados no son factibles, el procedimiento devuelve a la comprensión de los datos hasta que se cumplen los requisitos. Este modelo de recibo de la validación en la Universidad de La Rioja y de expertos de la aprobación para su evaluación de la eficacia de los instrumentos.



RESULTADOS

En esta sección se muestra la Máquina modelo de Aprendizaje, la cual fue adaptada a la nueva CSKT metodología, la presentación de los siguientes aspectos se muestra en la Figura 5, lo que indica la definición de los parámetros y variables consideradas en la Máquina modelo de Aprendizaje bajo la CSKT metodología.

Figura 5

Parámetros y variables de la Máquina modelo de Aprendizaje con metodología CSKT

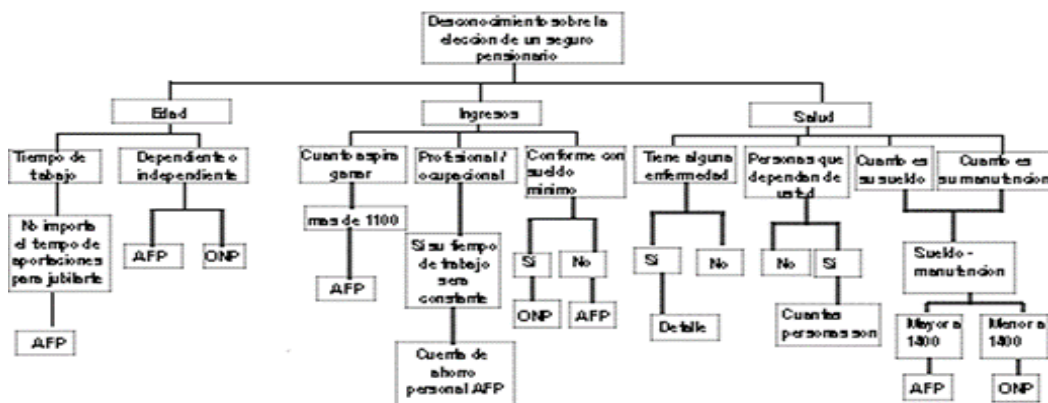
ML	CSKT	Modelo Empresarial	Preparación	Exploración	Visualización
Obtención de la DATA		Entendimiento del negocio	Entendimiento de los datos	Selección de datos	
ETL		SNP y SBS	Preparación de los datos	Preparación - Exploración - Limpieza	
Entrenamiento del modelo			Toma muestra de DATASETS	Comprender fuentes de datos	
Prueba de la DATA		Usuario	SNP ↔ SBS	Visualización de datos	Modelo preparado
Optimización					Presentar versión del modelo
					Aceptación del cliente

Los parámetros del modelo ML:

- ✓ Datos adquirentes: La fuente de datos desde el sistema nacional de pensiones y el SBS serán obtenidos.
- ✓ Extracción, Transformación y Carga de Los datos desde el sistema nacional de pensiones y el SBS se filtran.
- ✓ Modelo de formación: vamos a proceder con la formación de la ML modelo dentro del sistema nacional de pensiones.

Figura 6

Árbol de decisión en la elección de los seguros de pensiones, cada una de las situaciones relacionadas con el sistema de pensiones se muestra



- ✓ Los DATOS de las Pruebas: Las pruebas correspondientes de los datos seleccionados en el sistema nacional de pensiones y el SBS se realizan.
- ✓ Optimización: se concluye con los cambios respectivos para su entrega futura del cliente.

CSKT variables:

- ✓ Negocios modelo: descripción de la empresa en la cual la máquina modelo de aprendizaje estará centrado.
- ✓ La comprensión de los datos: El tipo de datos de la información extraída de la fuente del sistema nacional de pensiones y el SBS está definido.
- ✓ Selección de datos: Después de haber identificado los datos extraídos, una simplificación que se realiza mediante la obtención de los datos necesarios de la muestra.
- ✓ Preparación: Los datos del sistema nacional de pensiones y SBS fuente están preparados para el ETL.
- ✓ Preparación de Exploración de la Limpieza: Los datos se clasifican, ordenan y se filtra para obtener una más clara y definida de la muestra.
- ✓ Conjunto de datos de muestreo: Después de que el proceso de ETL, el final de la fuente de datos desde el sistema nacional de pensiones y el SBS se visualizan.
- ✓ Comprender las fuentes de datos: una vez que el filtrado de datos está disponible, un análisis que se realiza para confirmar si es el adecuado para presentar el modelo o el mismo proceso se repite.
- ✓ Presente versión del modelo: La primera versión de la ML presentación del modelo con los datos introducidos y seleccionados para las pruebas correspondientes.
- ✓ Solicitud de aceptación del cliente: una vez que se cumplen los criterios, es presentado al cliente para la validación del modelo creado.

Diseño del esquema preliminar de la modelo ML

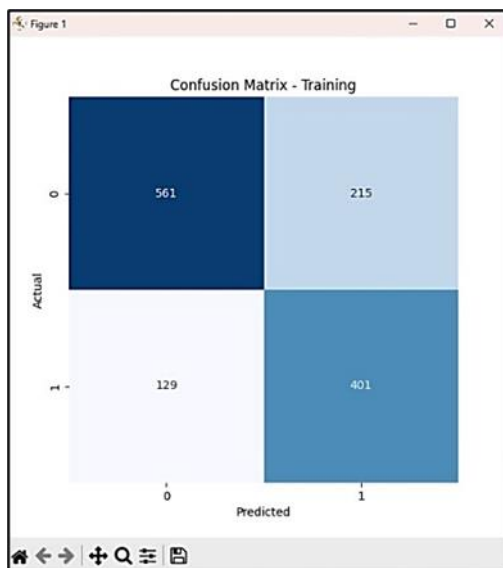


Dentro del árbol de decisión se muestra en la figura 6, hay los puntos de la Edad, los Ingresos y la Salud, en la parte de salud hay dos opciones que son el tiempo de Trabajo y Dependiente o Independiente, en el tiempo de Trabajo se muestra que el tiempo de contribuciones que no le importa a retirar para que proceda a una AFP, mientras que en el otro caso, usted tiene que seleccionar la AFP o a la ONP, En la parte de Ingreso hay tres opciones, que son la cantidad que el usuario aspira a ganar, si es profesional o laboral, y si se está de acuerdo con el salario mínimo, con respecto a la primera opción sigue una secuencia en la que pone más de 1100 soles y se va a la AFP, en el segundo caso hay una condición que, si el usuario del tiempo de trabajo va a ser constante, entonces usted tendrá una cuenta en una AFP.

Tenemos el modelo de formación, donde el código, cuando se ejecuta, los verdaderos positivos y los negativos y falsos positivos y negativos se muestran, siendo las respuestas que se obtiene ahora de partida de la simulación.

Figura 7

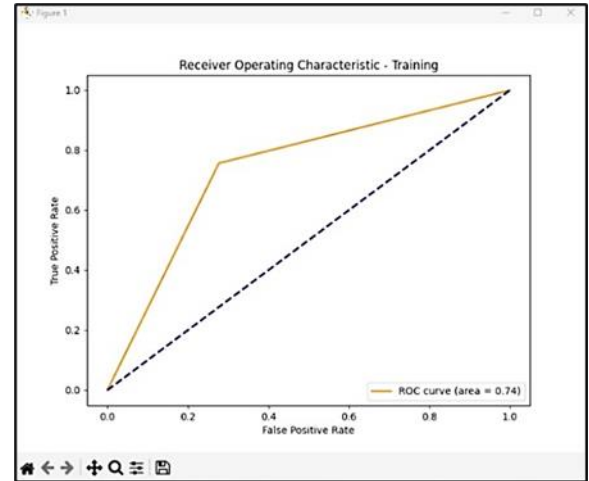
Matriz de confusión para la formación de la Máquina modelo de Aprendizaje



Basado en la matriz de confusión de la Máquina de Aprendizaje de entrenamiento del modelo se muestra en la figura 7, los dos valores de 1 y 0 se presentan en la que el 1 representa la ONP mientras que 0 es la AFP; esto es en referencia a la asignación de los seguros de pensiones, que está asignado el usuario, ya sea un ONP o AFP donde se muestran los datos en la escala actual que presenta un importe total de 129 que representa los falsos negativos y 215 el valor de los Falsos positivos. Mientras que los Verdaderos positivos presentan una cantidad de 561, y los Verdaderos negativos presentan un valor de 401 ahora de la formación de la presente modelo.

Figura 8

ROC esquema de la máquina modelo de aprendizaje

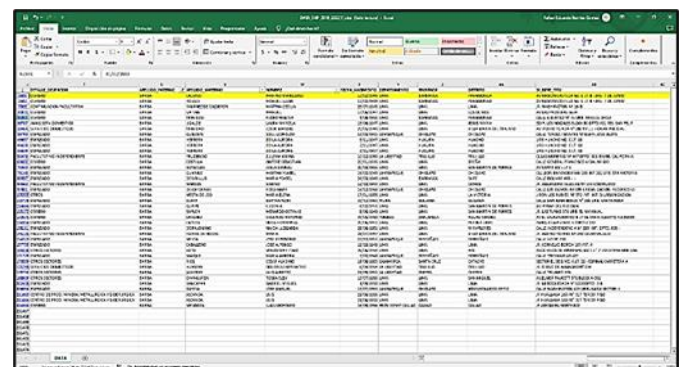


Después de ejecutar el código con los datos suministrados se muestra en la figura 8, la curva ROC de la parcela del actual modelo, que está siendo entrenado con la información mencionada anteriormente, se muestra. Este gráfico muestra la proyección de los falsos positivos a lo largo de la curva, presentando un área bajo la curva (AUC) de 0,74.

Inicialmente, la base de datos DATA_SNP_2010_2022_data.xls se ha tomado como referencia, información que se ha acumulado y se envía por la Superintendencia de Banca y Seguros. Estos datos podrían haber omisiones en algunos registros, ya que los sistemas transaccionales de la ONP y AFP podría presentar incidentes en la migración de la información, así como la atención durante la alternativa de horas de trabajo.

Figura 9

Datos de SNP



Para el presente ajuste se muestra en la figura 9, hemos tomado como referencia el caso de la BARBA CALSINO MARINO MARCELINO, que tiene la siguiente información en JSON código.



Figura 10

Jubilado datos de perfil

```
{
  "perfiles": {
    "nombre_apellido": "barba calsino marino marcelino",
    "edad": 20,
    "sexo": "M",
    "rango_edad_jubilarse": "60",
    "sueldo": 1236.54,
    "aspiracion_ganar_mas": "si"
  }
}
```

En la figura 10, código se refiere a que el usuario seleccionó para la formación de este modelo donde los respectivos datos necesarios para la selección de los seguros de pensiones son centrados, y al final, los resultados se muestran cuando el pensionado se le asigna una AFP, como el sistema de pensiones.

Figura 11

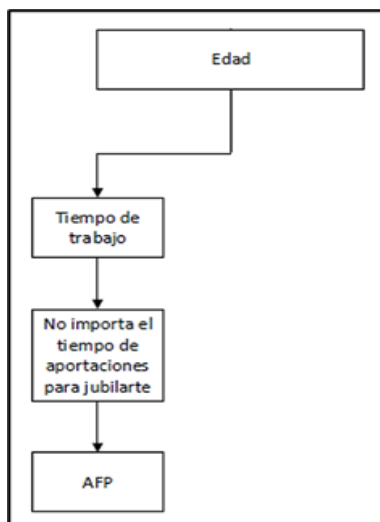
Resultado de la predicción del modelo

```
Body Cookies Headers (5) Test Results
Pretty Raw Preview Visualize JSON
1 {
2   "decision": "Predicción para barba calsino marino marcelino: Se elige AFP."
3 }
```

En la figura 11, el resultado de la indicada en caso de que se dedujo que se trataba de AFP o que se adapte a la AFP, porque es introducido por el criterio de la edad, el tiempo de trabajo entró y el salario que se muestra, esto es corroborado por medio del árbol de decisión donde la siguiente secuencia fue tomada con base en los criterios y los datos introducidos por el presente caballero.

Figura 12

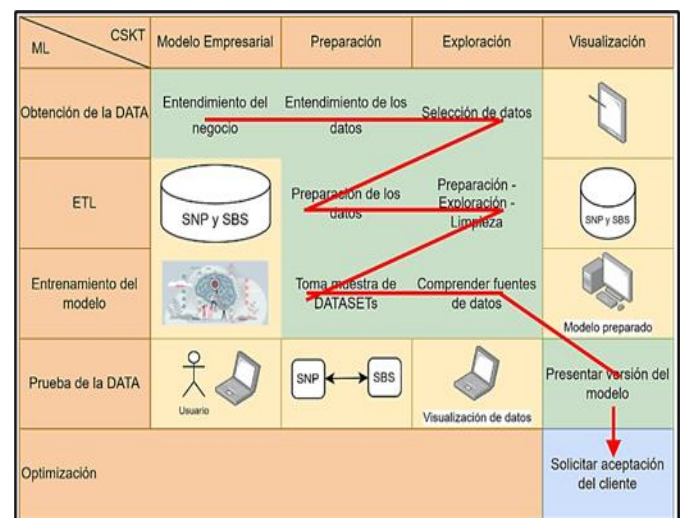
Secuencia de el árbol de decisión para los jubilados



En la figura 12, el modelo sigue el camino de la CSKT metodología de lograr la aceptación de la decisión evaluados para el caso del Señor de la BARBA CALSINO MARINO MARCELINO donde primero el monto de la estimación del tiempo de trabajo de la presente se determinó, seguido por la siguiente, que es independiente del tiempo de las contribuciones, de modo que el caballero puede retirarse dando como producto la asignación de una AFP.

Figura 13

Metodología de la secuencia



En la figura 13, la metodología se muestra en la Figura que sigue una secuencia estructurada, a partir de un enfoque en la comprensión de la empresa para captar sus necesidades y objetivos del fondo. Posteriormente, pasa a la comprensión de los datos, la recopilación de la información pertinente.

Una selección deliberada de una porción específica del total de los datos se realiza, seguido por la meticulosa preparación de los datos. Esta etapa consiste en la aplicación de filtros para eliminar incompleta puntos de datos o de usuarios duplicados.

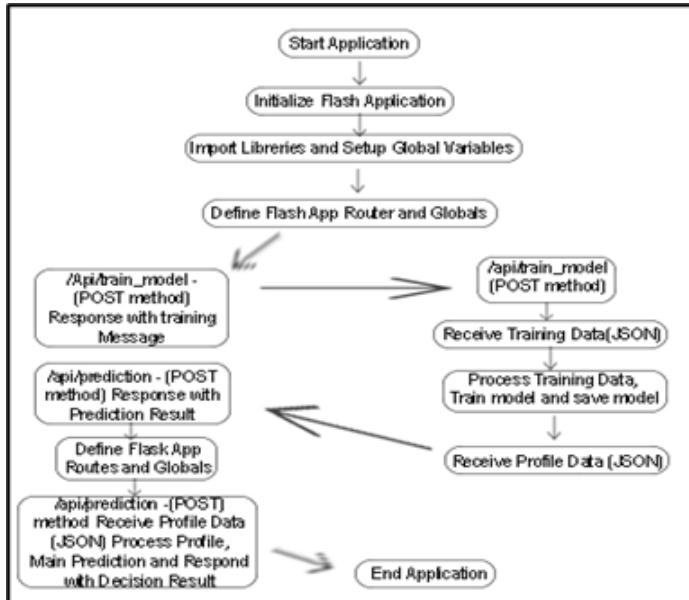
Después de que el filtrado de la fase, el limpiado de la muestra es presentada para su posterior revisión. Si es necesario, una nueva selección de la muestra se realiza para garantizar la calidad de los datos. Una vez que la información está lista, es enviado a la Máquina modelo de Aprendizaje para la evaluación. Los resultados obtenidos son utilizados para presentar la primera versión del modelo para el cliente.

En esta etapa, la aprobación del cliente del modelo desarrollado se solicita. Este enfoque secuencial asegura la aplicación efectiva y continua colaboración con el cliente durante todo el proceso.



Figura 14

Modelo funcional diseño



En la figura 14, el a partir de la secuencia de la Máquina modelo de Aprendizaje, que va desde la ejecución de la aplicación para el proceso que va desde la importación de las bibliotecas, las rutas que el modelo debe seguir hasta la ejecución de los resultados que el modelo se expresa a través de los datos que el usuario introduce culminando en el cierre de la aplicación.

Figura 15

Máquina de aprendizaje algoritmo de entrenamiento - Python script

```

from flask import Flask, request, jsonify
import pandas as pd
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import precision_score, recall_score, f1_score, roc_auc_score
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
import matplotlib.pyplot as plt
import joblib

app = Flask(__name__)

# Definir x_train y x_val_scaled a nivel global para que sea accesible en todo el script
x_train = None
y_train = None
model = None

# Crear un clasificador de árboles de decisión
model = DecisionTreeClassifier(random_state=0)

@app.route('/api/train_model', methods=['POST'])
def train_model():
    global x_train, y_train, model # Inicializar que rotamos utilizando las variables globales
    try:
        # Recibir los datos de entrenamiento aquí...
        training_data = request.json
        df_train = pd.DataFrame(training_data["training_examples"])

        # Dividir los datos en conjuntos de entrenamiento y prueba
        x_train = df_train[['edad', 'sueldo', 'rango_edad_beneficio', 'sueldo', 'aspiracion_ganar_mas']]
        x_train = pd.get_dummies(x_train, columns=['sexo', 'aspiracion_ganar_mas'], drop_first=True)
        y_train_scaled = StandardScaler().fit_transform(x_train)
    
```

En la figura 15, el estudio emplea un algoritmo de Árboles de Decisión para clasificar las opciones de pensión (AFP u ONP) con base en criterios tales como la edad, los ingresos, y la duración de los trabajos. Este método ofrece varias ventajas. Los árboles de decisión son muy interpretables y fáciles de visualizar, haciendo que el modelo de decisiones comprensible. Ellos se encargan de ambos datos categóricos y numéricos sin la necesidad de un amplio preprocesamiento. Además, son computacionalmente

eficiente, permitiendo la rápida formación y la predicción, lo cual es beneficioso para aplicaciones en tiempo real.

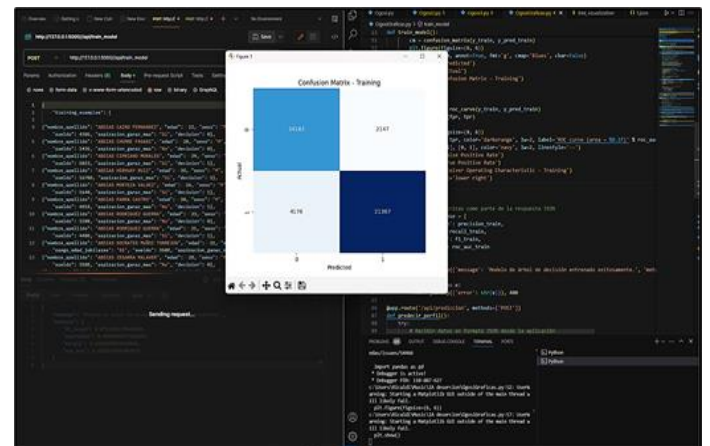
Sin embargo, los árboles de decisión también tienen limitaciones. Son propensos a sobreajuste, especialmente si no se podan. También son sensibles a pequeñas variaciones en los datos, lo cual puede resultar en cambios significativos en la estructura de árbol.

A pesar de estos inconvenientes, los árboles de decisión son adecuados para el problema en cuestión. La tarea consiste en la clasificación de la pensión de opciones basadas en criterios múltiples, que los árboles de decisión de gestionar de forma eficaz debido a su capacidad para manejar las decisiones complejas y múltiples variables. Además, el requisito para interpretar y justificar el modelo de las decisiones de los clientes se alinea bien con la interpretación ofrecida por los árboles de decisión, haciendo de ellos una elección apropiada para este estudio.

Esta secuencia de comandos de Python está diseñada para implementar la CSKT (Captura, Seleccionar, Kreate, Tren) metodología, específicamente de un árbol de decisión clasificador. El CSKT metodología se centra en la captura de datos, la selección de características relevantes, la creación del modelo, y por último curso de formación. La estación ('/api/prediccion') hace predicciones basadas en el modelo de aprendizaje. Procesa los datos del perfil, realiza las transformaciones necesarias y se hace la predicción. La decisión final (AFP u ONP) se determina con base en el modelo de salida. El Matraz se inicia la aplicación en modo de depuración para permitir la ejecución de código y pruebas. Este código Python que implementa la CSKT metodología para entrenar a un árbol de decisión clasificador a través de una API de web, lo que permite que tanto el modelo de formación y hacer predicciones basadas en perfiles específicos.

Figura 16

Modelo entrenado con los Aumentada de datos

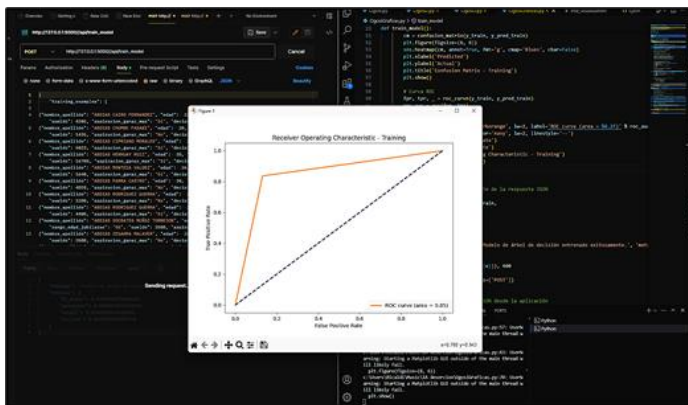




En la figura 16, basados en la matriz de confusión de la formación de la Máquina modelo de Aprendizaje adaptado con la aumentada SBS datos, dos valores de 1 y 0 se presentan; donde 1 representa a la ONP y 0 representa la AFP, en referencia al seguro de pensiones de la asignación a la que está asignado el usuario ONP o AFP, mostrando los datos en la escala de corriente que presenta un número total de 4176 como falsos negativos y 2147 como falsos positivos, mientras que en los verdaderos positivos que presenta una cantidad de 14182 y en los verdaderos negativos que presenta una cantidad de 21367 en el momento de la formación del modelo actual.

Figura 17

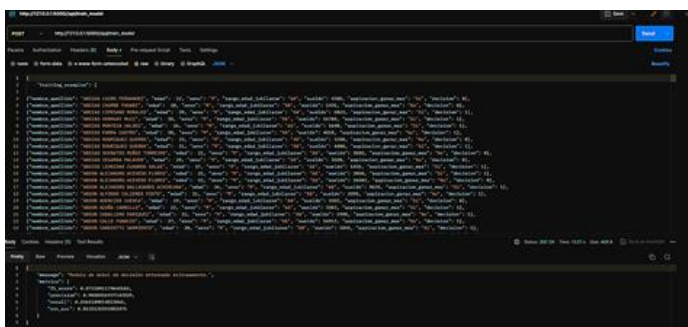
La curva ROC se centró en el aumento de los datos



En la figura 17, después de ejecutar el código, usando los datos proporcionados por el modelo adaptado con el aumento de la SBS de datos, el diagrama de la curva ROC del modelo actual se muestra, que está siendo entrenado con la información mencionada anteriormente. Esta representación gráfica se muestra la proyección de los falsos positivos a lo largo de la curva, la presentación de un área equivalente (AUC) de 0,85 bajo la curva.

Figura 18

Métrica de resultados centrados en el aumento de los datos

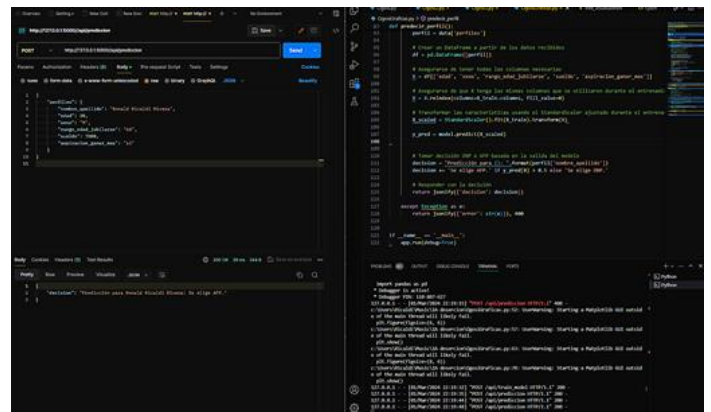


En la figura 18, en base al examen de la formación de las métricas del modelo de árbol de decisión con la aumentada SBS datos, cabe destacar que la F1-Score presenta un valor

específico de 0.87110911, lo que indica un equilibrio entre la precisión y el recall. El individuo de la precisión del modelo se encuentra en 0.90869269, destacando la capacidad de predecir con precisión positivos. Respecto de recordar, un valor de 0.83651098 se observa, que muestra la proporción de ejemplos positivos identificados correctamente por el modelo. Finalmente, la curva ROC muestra un valor de 0.85251355, proporcionando una medición completa de la capacidad de discriminación del modelo en diferentes puntos de corte.

Figura 19

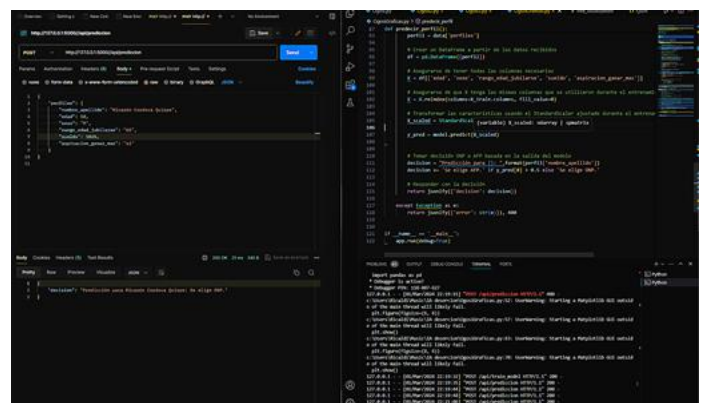
Las pruebas de los primeros datos de la modelo



En la figura 19, as pruebas de las predicciones del modelo ML, fue probado con el registro de uno de los trabajadores como los primeros datos, que podrían ser obtenidos, que marcó la AFP la selección correcta.

Figura 20

Segundo caso para el modelo de formación

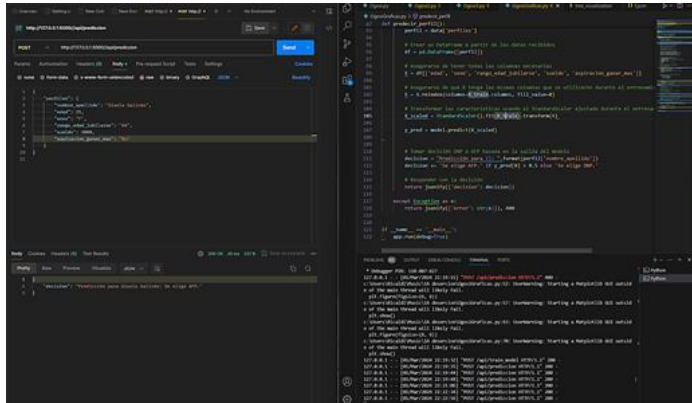


En la figura 20, en base a este segundo caso, los datos de un joven de 18 años de edad el joven que tiene la posición de Operador se considera el lugar donde, a través de sus datos introducidos en el modelo, es posible obtener el resultado que él eligió ONP basa en los mismos campos introducidos. para el primer caso.



Figura 21

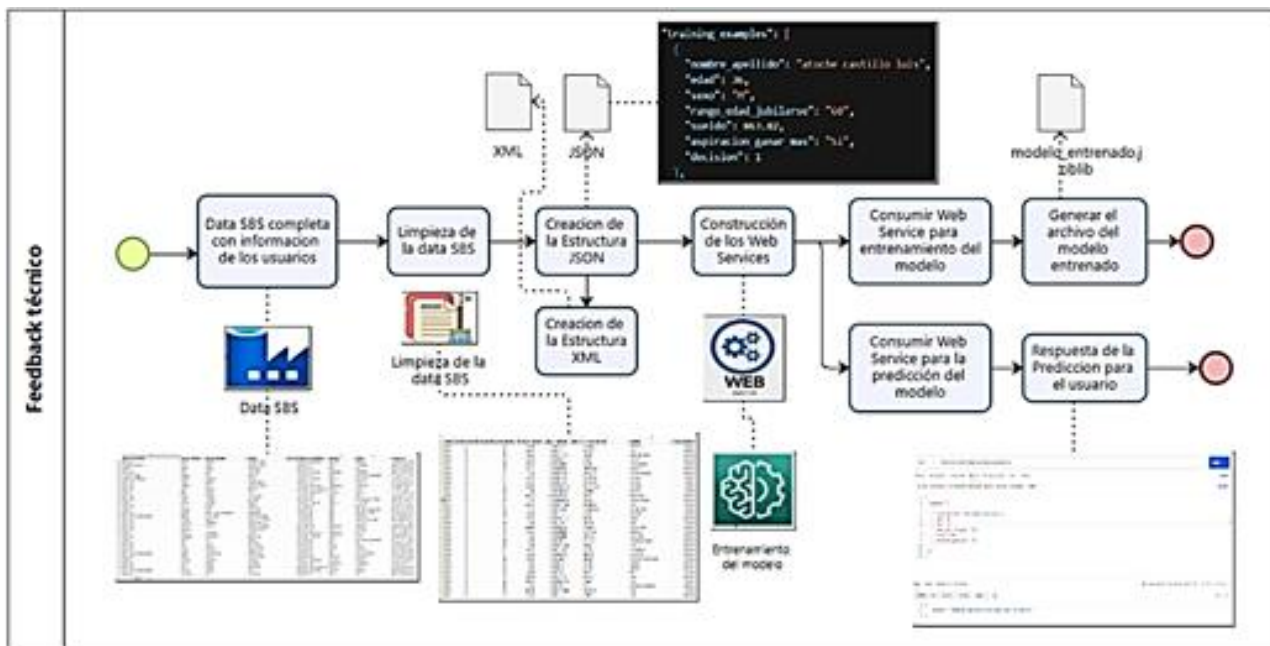
Tercer caso para el modelo de formación



En la figura 21, en base al tercer caso, tenemos los datos de un programador Jr., que a la hora de elegir las opciones que se ofrecen a través de ella los datos introducidos, el modelo, a través de dichos datos introducidos, fue el que predijo que su mejor opción para ella fue a la AFP.

Figura 22

Técnica de retroalimentación



En la figura 22, basados en la información técnica que se muestra en la Figura, se determina que algunos de los ajustes se deben hacer como se define para el entrenamiento de la ML-basado en el modelo. En este caso, de la SBS de datos debe seguir para ser limpiado antes de pasar al escenario de producción, debido a los posibles

errores que pueden surgir durante el procesamiento de las épocas. Otro ajuste que debe hacerse es el consumo de Servicios Web, y esto es debido a que el aumento en los datos en formato JSON, que consistirá en el aumento de la información para tener una mayor precisión en el modelo de aprendizaje.

Tabla 1

Comparación entre los resultados de la adaptación del modelo

Métricas de	Valor alcanzado en la adaptación del modelo con el CSKT metodología	Alcanzado el valor de la adaptación del modelo con el aumento de la SBS de datos
Precisión	0.69982547	0.90869269
Recall	0.65097402	0.83651098
F1 - Score	0.75660377	0.87110911
AUC - ROC	0.73977095	0.85251355





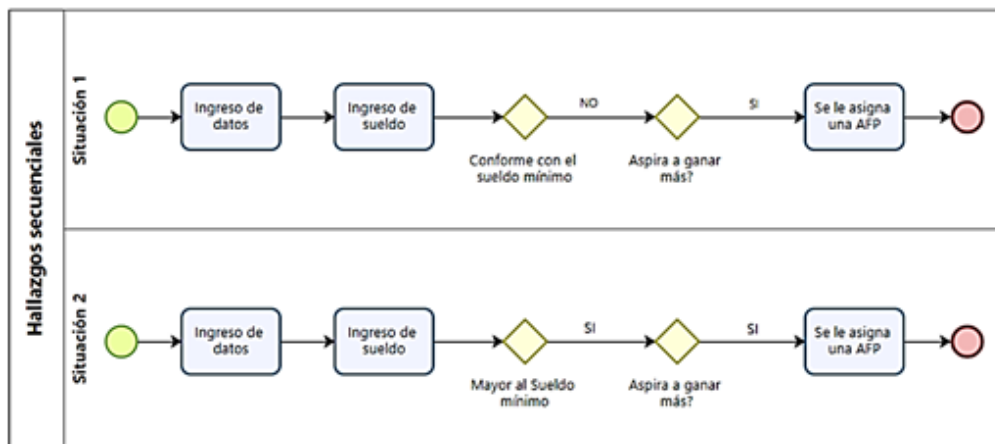
La tabla 1 muestra que el modelo mejoró significativamente después de la adición de SBS de datos. El aumento de precisión de 0.6998 a 0.9087, y el recuerdo de 0.6509 a 0.8365, lo que indica una mayor capacidad para identificar correctamente a los ejemplos positivos. El F1-Score también aumentó de 0.7566 a 0.8711, lo que refleja un mayor rendimiento equilibrado, y el AUC-ROC

aumento de 0.7398 a 0.8525, lo que demuestra una mejor discriminación entre clases.

Basado en las pruebas realizadas en el prototipo, la secuencia de los dos casos de trabajadores que tienen muy diferentes de los datos se consideró, por lo cual tenemos el siguiente esquema.

Figura 23

Aprendizaje automático modelo de caso de la secuencia de Flujo



En la figura 23, Basado en los casos explicados anteriormente, esta secuencia fue establecido en función de las decisiones tomadas en el momento de introducir sus datos, en el que uno de los casos indica que su sueldo es menor que el mínimo, pero que aspira a ganar más capital, mientras que el otro caso indica que su salario es superior

al salario mínimo, pero quiere ganar más capital, lo que resulta en ambos casos se coloca en el modelo que se asigna a la AFP, por lo tanto, se supone que, independientemente de la cantidad que ganan, si es que quieres ganar más contribuciones, a continuación, el modelo genera la AFP basado en los datos obtenidos a partir de dicho registro.

Tabla 2

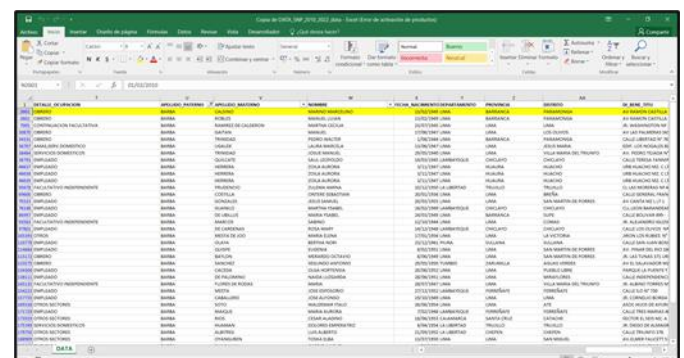
El éxito de los criterios

Las métricas Usadas	Criterios de aceptación	Alcanzado el Valor de	los Criterios de Éxito de
F1 – Score	Alcalde al 0.7 o 0.8	0.87110911	SI
Precisión	Alcalde al 0.7 o 0.8	0.90869269	SI
Recall	el Alcalde al 0.7 o 0.8	0.83651098	SI
AUC - ROC	Alcalde al 0.7 o 0.8	0.85251355	SI

En la tabla 2, basados en la tabla de criterios de la muestra, se determinó que en el caso de la F1 – Score tiene un valor de 0.87110911, que es mayor que la propuesta de aceptación criterio, entonces, por la precisión métrica, que tiene un valor de 0.90869269, que se acepta con respecto a los criterios de éxito, en el caso de recordar, presenta un valor de 0.83651098 que determina como criterio de éxito se considera después de la obtención de los datos respectivos en el modelo ML, finalmente, dado el ROC métrica, se determinó que se presenta el valor de 0.85251355, siendo mayor que el criterio de aceptación, que puede ser demostrado que este modelo presenta buenos niveles.

Figura 24

Colección histórica de la contratación de los datos



En la figura 24, unos se muestra en esta Figura, limpieza de datos se lleva a cabo, es decir, las columnas que no son relevantes se eliminan y sólo los datos respectivos se considera para la formación de la Máquina modelo de Aprendizaje.

Figura 25

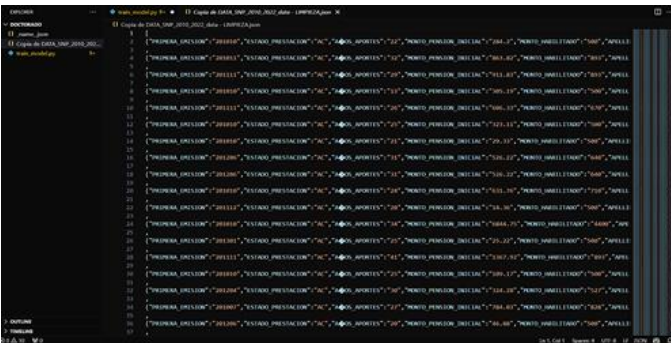
Excel para JSON Herramienta de Transformación



En la figura 25, una después de la limpieza de los datos respectivos para el modelo, usando un formato en línea de la herramienta, los datos se transforman en el formato de Excel en el formato JSON.

Figura 26

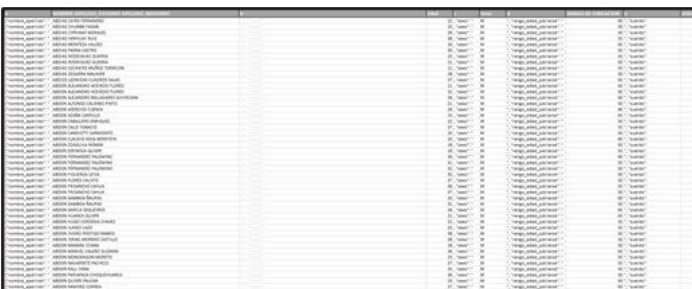
La visualización de los datos en formato JSON



En la figura 26, Después de proceder con la transformación del formato de los datos, esta Figura muestra todos los datos que se muestran a través de Visual Studio herramienta de Código en el archivo en formato JSON.

Figura 27

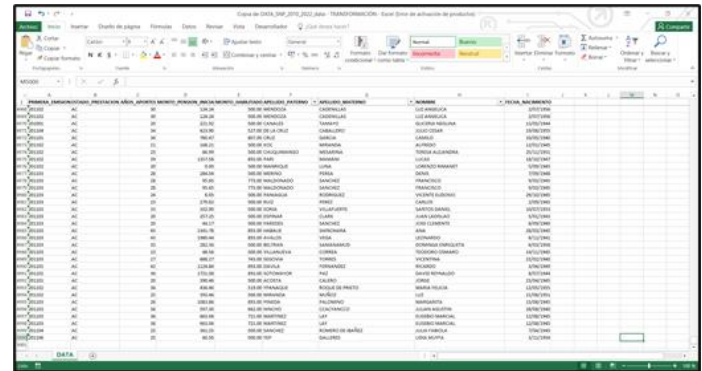
Datos de la purificación



En la figura 27, data la limpieza se realiza para asegurar que el conjunto de datos es correcto, así como fiables. Esto implica la corrección de errores de entrada, eliminando duplicados o inconsistencia en los datos, y la estandarización de formatos para evitar la confusión. Por la limpieza de datos, eliminar las barreras que pudieran distorsionar los resultados o llevar a conclusiones erróneas. Con los datos purificada y precisa, usted confiar en la información presentada para apoyar sus acciones, esto es crucial para la planificación estratégica y la gestión eficaz de los resultados que se esperan para este trabajo.

Figura 28

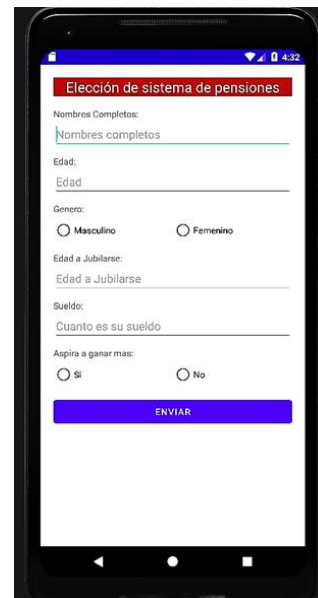
Datos ajustados



En la figura 28, basados en los datos recogidos, un total de 50.000 datos fueron presentados, en el que debió ser simplificado a un total de 5.000 registros, porque incompleta registros que presentan redundancia o también que los datos no eran correctos fueron descartados. Por lo tanto, se realizaron modificaciones a la hoja de cálculo.

Figura 29

Entrenamiento para el control de los datos

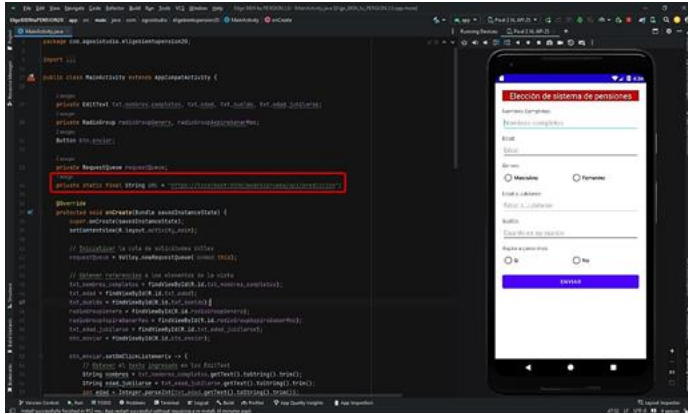




En la figura 29, de datos, se ejecuta mediante el control de la información, el cual está organizado en formato JSON, con el fin de determinar la exactitud de los índices de uso de la aplicación móvil.

Figura 30

Entorno de prueba para evaluar el modelo de proyección



En la figura 30, in el Android Studio entorno de prueba, la configuración ha sido establecida con el propósito de evaluar el modelo de proyección y se basa en él para ser capaz de determinar el valor de precisión.

Figura 31

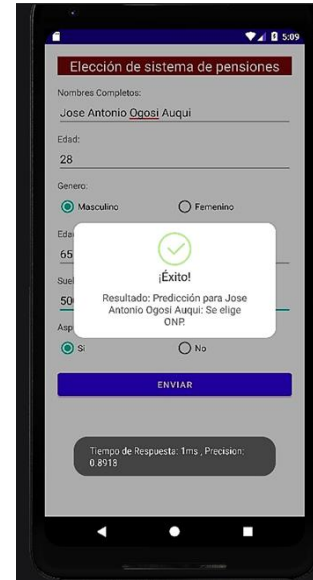
La medición de la eficiencia operativa



En la figura 31, el envío de la información de un usuario se muestra para medir la eficiencia operativa, la cual muestra un mensaje de confirmación, lo que indica que los datos han sido recibidos, y en base a esto, se observa que la precisión es 0.7129.

Figura 32

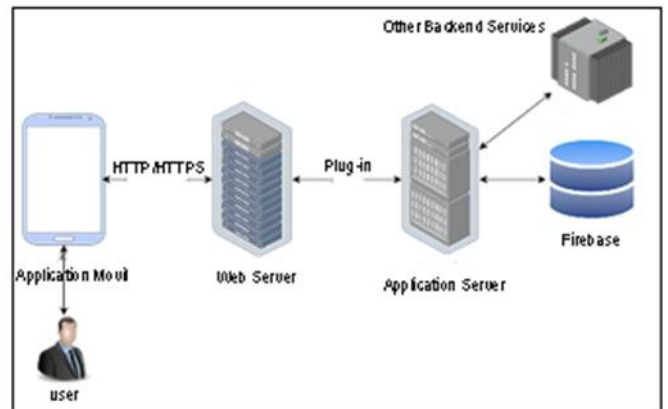
La medición de operativos optimizados eficiencia



En la figura 32, la transmisión de datos de un usuario es representado para medir la optimización de la eficiencia operativa, la cual muestra un mensaje de confirmación, lo que indica que los datos han sido recibidos, y en base a esto, se observa que la precisión es 0.8918.

Figura 33

Aplicación móvil de la implementación de la infraestructura



En la figura 33, basados en la infraestructura de la muestra, el usuario comienza con la interacción de la aplicación móvil en la que él envía instrucciones al servidor web agrega a la aplicación de servidor que está conectado a otros servicios de backend y a la base de datos de la Firestore servicio. Firebase, donde los dos últimos enviarle la petición realizada y la respuesta desde el servidor de la aplicación, seguido por la respuesta desde el servidor Web para luego ser proyectada en la aplicación móvil en vista del usuario que envía dicha solicitud.



Se crea un formulario en Google Formas de medir la satisfacción del usuario, basado en la aplicación móvil con la máquina de aprendizaje de la aplicación de la CSKT metodología. Cabe señalar que la fiabilidad, privacidad y confidencialidad de la información proporcionada por los usuarios, es respetado, así como la seguridad de los datos que se proporcionan. El formulario está alojado en: <https://docs.google.com/forms/d/e/1FAIpQLScg45fSCdgGR3xsqHcM6cLLoXS0H7jJFz6iQGDWBGGBb2b29Q/viewform>.

Figura 34

Respuesta del usuario final para evaluar la satisfacción de la aplicación móvil - Pregunta 1



En la figura 34, La recopila las respuestas de los encuestados revelan un significativo nivel de satisfacción entre los encuestados.

La mayoría de los encuestados, representados por el 67%, indicaron que estaban "Totalmente Satisfecho", subrayando una muy percepción positiva respecto a la eficiencia de la interfaz para facilitar la gestión de los recursos y mejorar la eficiencia del proceso de contratación. Además, el 29% expresó estar "Satisfechos", apoyando la conclusión general de que la interfaz cumple con las expectativas y es satisfactorio para una porción considerable de los usuarios. Lo que sugiere que el 96% de los participantes de la experiencia de satisfacción con la interfaz de la aplicación. Estos resultados indican que un impacto positivo que se ha logrado, según la perspectiva de los encuestados.

DISCUSIÓN

La aplicación de CSKT la metodología de los modelos de Aprendizaje automático ha sido demostrado ser eficaz en la mejora de procesos tales como la dotación de personal, modelo de supervisión en la producción y análisis de negocio. La importancia de la selección de representante y de la calidad de los datos se destacó, junto con el modelo de optimización mediante técnicas tales como la validación cruzada y los parámetros de ajuste. Los resultados demuestran una mejora significativa en la asignación de

una pensión, con una precisión de 69.98% y la recuperación de la 65.10%. Sin embargo, las áreas de mejora se identifican en la limpieza de los datos y optimización, lo que podría aumentar aún más la eficacia del modelo. El uso de la infraestructura tecnológica como base avanzada y Google Cloud para la implementación de aplicaciones móviles que optimizar la gestión de los recursos de los procesos también se pone de relieve. A pesar de los beneficios, es esencial para hacer frente a desafíos tales como la dependencia de la calidad de los datos y la necesidad de conocimientos técnicos para su implementación.

CONCLUSIÓN

El estudio valida la eficacia de la CSKT metodología para la aplicación de los modelos de Aprendizaje automático, la mejora de la eficiencia en los procesos de contratación, análisis de datos y la asignación de una pensión. Continua optimización y el uso de avanzadas técnicas permiten mejoras sustanciales en el rendimiento y la precisión, destacando el valor de la metodología en la transformación organizativa de los procesos. Sin embargo, el potencial de los datos de los prejuicios y las implicaciones éticas que deben ser considerados, así como la infraestructura necesaria para su implementación. Los beneficios incluyen el ahorro de tiempo, reducción de costos y una mayor satisfacción en la gestión de los recursos humanos.

REFERENCIAS BIBLIOGRÁFICAS

1. Editorial (2021). Ageism: half of those over 40 have felt disadvantaged by their age in a hiring process. RRHH Digital. <https://www.rrhhdigital.com/secciones/seleccion/150135/Edadismo-la-mitad-de-los-mayores-de-40-se-ha-sentido-en-desventaja-por-su-edad-en-un-proceso-de-contratacion/>, accessed on 13 Feb. 2024.
2. Management. (2023). Demand for security personnel will increase 18% due to the end-of-year campaign. Management. <https://gestion.pe/economia/empresas/manda-de-personal-de-seguridad-aumentara-18-por-campana-de-fin-de-ano-noticia/>
3. W. Sanguinetti Raymond, The employment contract in the face of economic, social and productive transformations. Peru: El Buhó E.I.R.L., 2016. <http://trabajodigno.pe/wp-content/uploads/2016/02/Los-contratos-de-trabajo-de-duraci%C3%B3n-determinada.pdf>
4. Zafeiriadou, A. (2021). Candidates' reaction on the modern methods of personnel selection processes and the influence of word-of-mouth on employer branding. <https://repository.ihu.edu.gr/xmlui/bitstream/handle/1>



- 1544/29891/a.zafeiriadou_mng_23-06-2021.pdf?sequence=1
5. Campesato, O. (2020). Artificial intelligence, machine learning, and deep learning. <https://styluspub.presswarehouse.com/browse/book/9781683924678/Artificial-Intelligence-Machine-Learning-and-Deep-Learning>.
 6. Chinnamgari, S.K. (2019). R Machine Learning Projects: Implement supervised, unsupervised, and reinforcement learning techniques using R 3.5. <https://search.worldcat.org/es/title/R-Machine-Learning-Projects-:-Implement-Supervised-Unsupervised-and-Reinforcement-Learning-Techniques-Using-R-3.-5/oclc/1083465396>
 7. Greener, J.G., Kandathil, S.M., Moffat, L., Jones, D.T. (2022). A guide to machine learning for biologists. *Nature Reviews. Molecular Cell Biology*, 23, 40-55. <https://doi.org/10.1038/s41580-021-00407-0>
 8. Hart, G., Mueller, T., Toher, C., Curtarolo, S. (2021). Machine learning for alloys. *Nature Reviews. Materials*, 6(8): 730-755. <https://doi.org/10.1038/s41578-021-00340-w>
 9. Janiesch, C., Zschech, P., Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, 31(3): 685-695. <https://doi.org/10.1007/s12525-021-00475-2>
 10. Nasiriany, S., Thomas, G., Wang, W., Yang, A., Listgarten, J., & Sahai, A. (2019). A comprehensive guide to machine learning. Department of Electrical Engineering and Computer Sciences University of California, Berkeley. <https://sourestdeeds.github.io/pdf/intro-to-ml-berkeley.pdf>
 11. Mahesh, B. (2020). Machine learning algorithms - A review. *International Journal of Science and Research (IJSR)*, 9: 381-386. <https://doi.org/10.21275/art20203995>
 12. Bozzi, M.E. (2023). Machine learning models applied to the estimation of the probability of default of banks in the Argentinean banking system. University of San Andrés. <http://hdl.handle.net/10908/23089>
 13. González, P., Andrés, G. (2021). Detection of fraudulent credit card transactions using Machine Learning models. Universidad de los Andes. <http://hdl.handle.net/1992/53571>
 14. Borja-Robalino, R., Monleon-Getino, A., Rodellar, J. (2020). Standardisation of performance metrics for Machine and Deep Learning classifiers. *Revista Ibérica de Sistemas e Tecnologías de Informação*, E30: 184-196. https://www.researchgate.net/profile/Antonio-Monleon-Getino/publication/342009715_Estandarizacion_de_metricas_de_rendimiento_para_clasificadores_Machi
 15. Ogosi Auqui, J.A., Camargo, J.L., Vera Tito, F.S. (2022). New CSKT methodology to improve Machine Learning implementation projects in Industrial Engineering. <http://45.5.56.102/bitstream/123456789/181/1/Nueva%20metodolog%C3%ADa%20CSKT%20para%20mejorar%20los%20proyectos%20de%20implementaci%C3%B3n%20de%20Machine%20Learning%20en%20Ingenier%C3%ADa%20Industrial.pdf>
 16. Pisner, D.A., Schnyer, D.M. (2020). Support vector machine. *Machine learning*, 101-121. <https://doi.org/10.1016/B978-0-12-815739-8.00006-7>
 17. Hidayati, N., Hermawan, A. (2021). K-Nearest Neighbor (K-NN) algorithm with Euclidean and Manhattan in classification of student graduation. *Journal of Engineering and Applied Technology*, 2(2): 86-91. <http://dx.doi.org/10.21831/jeatech.v2i2.42777>
 18. Blanquero, R., Carrizosa, E., Ramírez-Cobo, P., Sillero-Denamiel, M.R. (2021). Variable selection for Naïve Bayes classification. *Computers & Operations Research*, 135: 105456. <https://doi.org/10.1016/j.cor.2021.105456>
 19. Sinaga, K.P., Yang, M.S. (2020). Unsupervised K-means clustering algorithm. *IEEE access*, 8, 80716-80727. <https://doi.org/10.1109/ACCESS.2020.2988796>
 20. Blockeel, H., Devos, L., Blockeel, B., Nanfack, G., Nijssen, S. (2023). Decision trees: From efficient prediction to responsible AI. *Frontiers in Artificial Intelligence*, 6. <https://doi.org/10.3389/frai.2023.112455>