

CAIO OLIVEIRA QUINAMO
DATA SCIENTIST



DETECCIÓN DE ANOMALÍAS EN TRANSACCIONES FINANCIERAS PARA LA IDENTIFICACIÓN DE FRAUDE MEDIANTE AUTOENCODERS Y RANDOM FOREST



CAIO QUINAMO
ANALYTICS SOLUTIONS
MACHINE LEARNING &
DATA SCIENCE FOR BUSINESS IMPACT

DOCUMENTACIÓN DEL INFORME TÉCNICO

ÍNDICE

0. Introducción	Página 02
1. Comprensión del Negocio	
1.1 Determinar los objetivos empresariales	Página 04
1.2 Evaluar la situación	Página 04
1.3 Determinar los objetivos de la minería de datos	Página 05
1.4 Elaborar el plan del proyecto	Página 05
2. Comprensión de los Datos	
2.1 Recopilar los datos iniciales	Página 06
2.2 Describir los datos	Página 06
2.3 Explorar los datos	
2.3.1 Análisis de la variable objetivo	Página 06
2.3.2 Análisis de correlaciones con la variable objetivo	Página 07
2.3.3 Análisis de las variables «Time» y «Amount»	Página 07
2.3.4 Visualización de los datos (antes de la reconstrucción del Autoencoder)	Página 08
2.3.5 Detección univariada de valores atípicos mediante normalización Z-Score	Página 09
2.3.6 Detección multivariada de valores atípicos mediante la distancia de Mahalanobis	Página 10
2.4 Verificar la calidad de los datos	Página 11
3. Preparación de los Datos	
3.1 Seleccionar los datos	Página 12
3.2 Limpiar los datos	Página 12
3.3 Construir los datos	Página 13
3.4 Integrar y formatear los datos	Página 13
4. Modelado	
4.1 Selección de técnicas de modelado	Página 14
4.2 Generar el diseño de pruebas	Página 15
4.3 Configurar los parámetros del modelo	Página 17
4.4 Evaluación del modelo	Página 18
5. Evaluación	
5.1 Evaluar resultados	Página 20
5.2 Proceso de revisión	Página 20
5.3 Modelos aprobados	Página 21
5.4 Determinar los próximos pasos	Página 21
6. Referencias	Página 22
7. Licencia y Nota Final del Autor	Página 23

Página 04
Página 04
Página 05
Página 05

Página 06
Página 06
Página 06
Página 06
Página 07
Página 07
Página 07
Página 08
Página 09
Página 10
Página 11

Página 12
Página 12
Página 13
Página 13

Página 14
Página 15
Página 17
Página 18

Página 20
Página 20
Página 21
Página 21

Página 22

Página 23

CAIO OLIVEIRA QUINAMO
DATA SCIENTIST
caioquinamocontact@gmail.com





0. INTRODUCCIÓN

En los últimos años, las compras online han experimentado un crecimiento vertiginoso. Este nuevo patrón de consumo, ya integrado en el mercado moderno, ha impulsado la aparición de modelos de negocio digitales. Las grandes empresas, ante este cambio, se han visto obligadas a migrar hacia canales digitales, dejando atrás su dependencia exclusiva de medios tradicionales.

A nivel global, las ventas de comercio electrónico pasaron de 1.34 billones de dólares en 2014 a una estimación de 6.33 billones en 2024, con proyecciones de alcanzar los 8.03 billones en 2027. Esto representa un crecimiento de casi seis veces en apenas trece años. En 2023, las billeteras digitales representaron el 49% de las transacciones online, seguidas por las tarjetas de crédito con un 21%. Se espera que los ingresos generados por pagos digitales alcancen los 14.79 billones para 2027. Actualmente, dos tercios de los adultos en el mundo utilizan este tipo de pagos, con una adopción del 95% en países desarrollados.

En Estados Unidos, las ventas online pasaron de 79.02 mil millones de dólares en el cuarto trimestre de 2014 a 308.91 mil millones en el mismo periodo de 2024. En 2023, el 69% de los adultos en línea en EE.UU. usaron algún método de pago digital en los últimos tres meses. Entre los más populares destacan PayPal (40%), Apple Pay (24%) y Venmo (16%). El 89% de los estadounidenses utilizan pagos digitales y un 62% emplea al menos dos métodos distintos.

Este auge digital también ha abierto espacio para que el fraude online crezca como un negocio paralelo impulsado por ciberdelincuentes. Aunque la ciberseguridad ha avanzado, los atacantes han refinado sus técnicas para burlar los sistemas de detección. Entre los métodos más comunes están el robo de tarjeta (carding), el phishing (captura de datos en sitios fraudulentos), el fraude por identidad sintética (combinación de datos reales y ficticios), el chargeback fraud (reembolsos falsos tras recibir el producto) y la apropiación de cuentas (account takeover).

El fraude en transacciones online representa un riesgo significativo tanto para los usuarios como para las empresas. Los consumidores pueden sufrir pérdidas económicas, robo de identidad y pérdida de confianza en el entorno digital. Las empresas enfrentan consecuencias como pérdidas financieras, daño reputacional, mayores costes operativos por seguridad adicional y pérdida de clientes. A medida que aumentan las operaciones digitales, también lo hacen las oportunidades para los actores maliciosos, lo que exige soluciones cada vez más sofisticadas para la detección y prevención del fraude.

Este proyecto tiene un enfoque didáctico y representativo. Se analizará un historial de transacciones con el objetivo de identificar patrones que caractericen las operaciones fraudulentas, apoyándose en visualizaciones explicativas y un modelo de Machine Learning capaz de señalar los factores clave para la clasificación. El objetivo principal será detectar el mayor número posible de fraudes, minimizando los falsos positivos.

Todo el proceso se desarrollará siguiendo la metodología CRISP-DM, que organiza el trabajo en seis etapas: comprensión del negocio, comprensión de los datos, preparación, modelado, evaluación y despliegue. Esta última fase no será tratada por estar fuera del alcance del objetivo del proyecto.

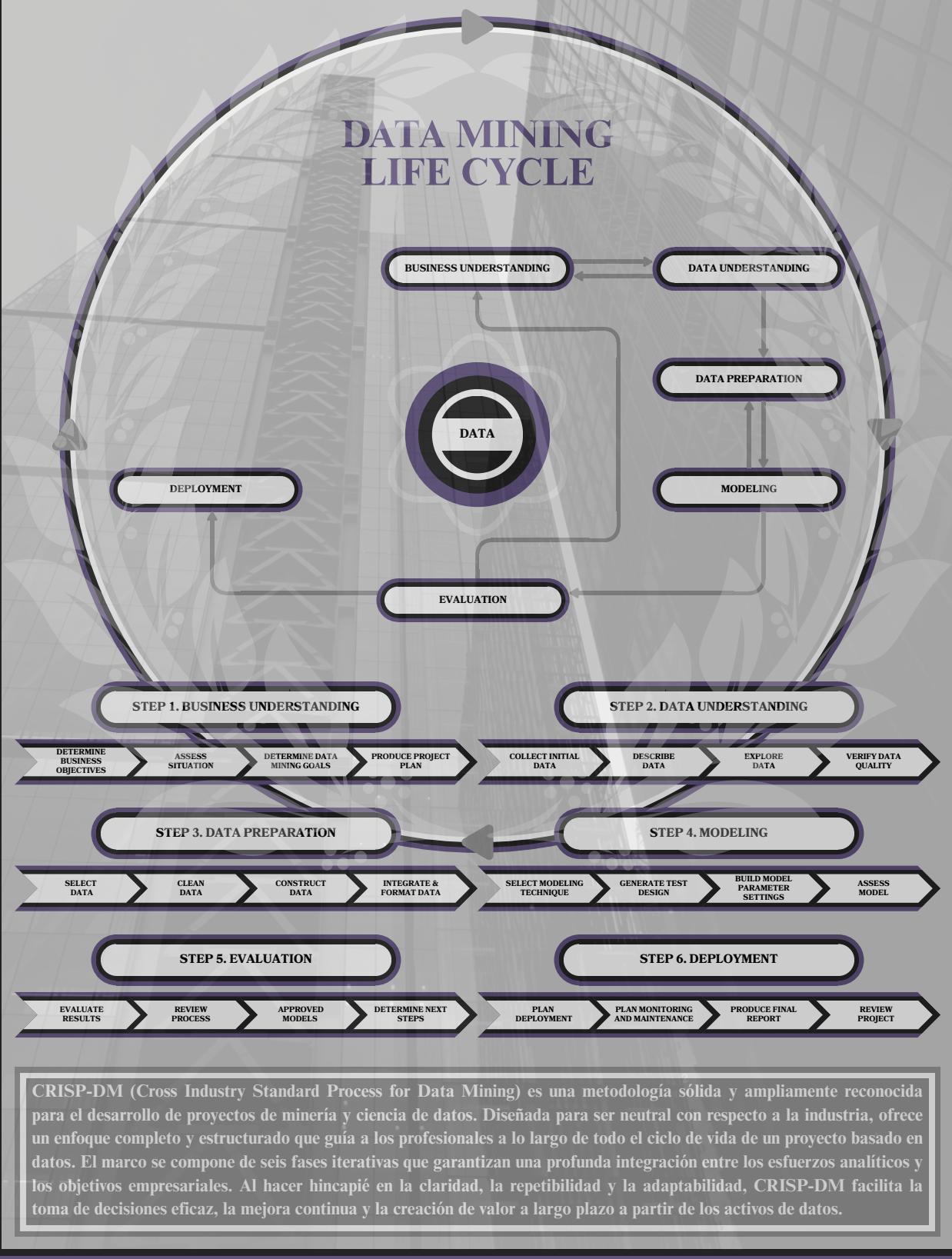
La base del enfoque será un autoencoder entrenado exclusivamente con transacciones verídicas (clase mayoritaria), para aprender la representación latente de los datos normales. Al no haber sido expuesta a fraudes durante el entrenamiento, se espera que su capacidad de reconstrucción sea deficiente frente a transacciones anómalas, generando así un error de reconstrucción elevado. Este error se utilizará como una nueva característica para alimentar un modelo de clasificación supervisado basado en Random Forest. Esta combinación busca unir las fortalezas del aprendizaje no supervisado (detección de anomalías mediante el autoencoder) con la solidez de un clasificador supervisado, mejorando así la precisión global del sistema. Sin embargo, uno de los principales retos será que ciertas transacciones legítimas puedan presentar patrones similares a los fraudes, generando confusión en el modelo. Por ello, será clave un estudio adecuado para los valores atípicos y del desbalance de clases, aspecto inherente a este tipo de problemas.

CAIO OLIVEIRA QUINAMO
DATA SCIENTIST
caioquinamocontact@gmail.com





CRISP-DM METHODOLOGY



CRISP-DM (Cross Industry Standard Process for Data Mining) es una metodología sólida y ampliamente reconocida para el desarrollo de proyectos de minería y ciencia de datos. Diseñada para ser neutral con respecto a la industria, ofrece un enfoque completo y estructurado que guía a los profesionales a lo largo de todo el ciclo de vida de un proyecto basado en datos. El marco se compone de seis fases iterativas que garantizan una profunda integración entre los esfuerzos analíticos y los objetivos empresariales. Al hacer hincapié en la claridad, la repetibilidad y la adaptabilidad, CRISP-DM facilita la toma de decisiones eficaz, la mejora continua y la creación de valor a largo plazo a partir de los activos de datos.





1. COMPRENSIÓN DEL NEGOCIO

1.1 DETERMINAR LOS OBJETIVOS EMPRESARIALES

¿Es posible detectar de forma automatizada y precisa las transacciones fraudulentas en un sistema de pagos digitales, minimizando al mismo tiempo el número de operaciones legítimas clasificadas erróneamente como fraude?

Esta es la pregunta central que motiva este proyecto, y cuya respuesta puede marcar una diferencia sustancial para las pequeñas y medianas empresas (pymes) en el actual ecosistema digital. Hoy en día, las pymes enfrentan un entorno en el que las transacciones electrónicas fraudulentas crecen en volumen, complejidad y sofisticación. A diferencia de grandes corporaciones, estas empresas a menudo carecen de infraestructura tecnológica avanzada y recursos humanos especializados para implementar soluciones efectivas de detección de fraude. Como resultado, dependen de pasarelas de pago o servicios externos, que suelen basarse en modelos genéricos, poco adaptables a la realidad operativa específica de cada negocio.

Este contexto plantea una necesidad urgente: disponer de una herramienta predictiva eficaz, asequible y adaptada al contexto real de la empresa, que permita tomar decisiones más inteligentes, en tiempo real, sobre posibles fraudes sin perjudicar a usuarios legítimos.

Los objetivos empresariales son los siguientes: **Mejorar la precisión en la detección de fraude.** Desarrollar un sistema capaz de identificar de manera automatizada las transacciones fraudulentas con alta precisión, minimizando tanto los falsos negativos como los falsos positivos. **Proteger la experiencia del cliente.** Reducir los rechazos injustificados de operaciones válidas, lo que ayuda a mantener la confianza del cliente en la plataforma y evita pérdidas por fricción comercial. **Minimizar pérdidas económicas.** Disminuir el impacto financiero directo asociado a fraudes no detectados, así como los costos indirectos (cargos por contracargos, penalizaciones, deterioro reputacional). **Ganar autonomía y competitividad.** Contar con una solución ajustada a las particularidades del negocio, que permita tomar decisiones informadas y rápidas, sin depender exclusivamente de filtros genéricos de terceros. **Asegurar escalabilidad y adaptabilidad.** Implementar una solución que pueda crecer junto con el negocio y adaptarse a nuevos patrones de fraude conforme evolucionen las amenazas digitales.

1.2 EVALUAR LA SITUACIÓN

Este proyecto tiene un propósito académico y demostrativo, por lo que no se trabajará con datos reales de una empresa. En su lugar, se utilizará un conjunto de datos público, ampliamente referenciado en estudios sobre fraude digital, que contiene transacciones con tarjetas de crédito realizadas por titulares europeos durante dos días de septiembre de 2013. Aunque los datos han sido anonimizados, mantienen patrones representativos del problema y permiten simular un entorno realista.

Una característica crítica de este dataset es su fuerte desbalance de clases, donde las transacciones fraudulentas representan una mínima fracción del total. Este desequilibrio plantea desafíos importantes en la construcción de modelos, ya que obliga a buscar estrategias que preserven la sensibilidad sin disparar la tasa de falsos positivos. Si bien esta limitación técnica complica la tarea de clasificación, también ofrece una oportunidad valiosa para poner a prueba enfoques robustos en contextos adversos.

El desarrollo se llevará a cabo en una estación de trabajo personal de gama media-alta, suficiente para análisis exploratorio y entrenamiento de modelos ligeros. El entorno elegido es Jupyter Notebooks con Python, apoyado en librerías estándar como Pandas, Numpy, Scikit-Learn, Matplotlib, Seaborn y TensorFlow/Keras.

Entre las principales limitaciones del proyecto se encuentran la ausencia de datos empresariales específicos, la falta de acceso a transacciones en tiempo real y la exclusión del despliegue en un entorno productivo. Sin embargo, se simulará un flujo de trabajo completo, siguiendo buenas prácticas de desarrollo en ciencia de datos, con el objetivo de demostrar la viabilidad del enfoque propuesto y su potencial adaptación a contextos reales.

1.3 DETERMINAR LOS OBJETIVOS DE LA MINERÍA DE DATOS

Desde la perspectiva de la minería de datos, el objetivo principal de este proyecto es construir un sistema predictivo capaz de identificar transacciones anómalas en un entorno digital de pagos con tarjetas, maximizando la detección de fraudes reales y minimizando la clasificación errónea de transacciones legítimas. La solución debe ser lo suficientemente robusta para enfrentar escenarios con un fuerte desbalance de clases, típicos en problemas de fraude financiero, y permitir una interpretación clara de los resultados para facilitar su aplicación en contextos reales.

CAIO OLIVEIRA QUINAMO
DATA SCIENTIST
caioquinamocontact@gmail.com





Esto implica transformar el problema de negocio en tareas específicas de minería de datos, como la detección de anomalías y la clasificación binaria, con un enfoque orientado a la optimización de métricas que reflejen de manera más precisa la calidad del sistema en contextos desbalanceados, como el área bajo la curva precisión-recall (AUPRC) y el F1-score. Se prioriza, por tanto, el desarrollo de un modelo que no solo sea técnicamente preciso, sino también funcional desde el punto de vista operativo: que reduzca al mínimo los falsos positivos que afectan la experiencia del usuario legítimo y que mantenga una sensibilidad alta frente a patrones de fraude, incluso cuando estos se presenten en volúmenes marginales.

Para alcanzar este objetivo, la estrategia de minería de datos combinará técnicas de aprendizaje supervisado y no supervisado, permitiendo capturar patrones sutiles en el comportamiento transaccional. Asimismo, se incorporarán métodos de selección de características para mejorar la eficiencia del modelo, reducir la redundancia y favorecer su capacidad de generalización. El proceso de evaluación será riguroso y centrado en métricas adecuadas al problema, buscando un equilibrio óptimo entre precisión operativa y riesgo de pérdida.

1.4 ELABORAR EL PLAN DEL PROYECTO

El desarrollo del proyecto seguirá las seis fases establecidas por la metodología CRISP-DM, adaptadas al contexto específico de la detección de fraude en transacciones digitales. Cada etapa incluirá tareas concretas y objetivos definidos para garantizar un enfoque riguroso y coherente.

En la fase de comprensión del negocio se analizará el impacto del fraude digital, con especial atención a las limitaciones que enfrentan las pequeñas y medianas empresas. A partir de este análisis se definirá una pregunta central que guiará todo el proceso de minería de datos. En la fase de comprensión de los datos se realizará una exploración inicial del dataset, identificando su estructura, patrones generales, valores atípicos y, especialmente, el grado de desbalance entre clases, lo que condicionará las decisiones posteriores de modelado.

La preparación de los datos abarcará tareas de limpieza, transformación y creación de nuevas variables relevantes, orientadas a maximizar la calidad de la información disponible para el modelo. Se establecerán esquemas de partición del dataset acordes a los requerimientos de las técnicas utilizadas, asegurando una separación adecuada entre los datos de entrenamiento y evaluación.

Durante la fase de modelado se desarrollará un sistema híbrido que combine técnicas de detección de anomalías y clasificación supervisada. Se aplicarán métodos de validación cruzada y selección de características para optimizar el rendimiento predictivo y reducir el riesgo de sobreajuste.

En la fase de evaluación se medirán distintas métricas de desempeño, como F1-score, AUPRC o AUROCC, priorizando aquellas más adecuadas para contextos desbalanceados. El modelo será validado no solo en términos cuantitativos, sino también considerando el impacto práctico de sus errores, en especial la proporción de transacciones legítimas marcadas incorrectamente como fraudulentas.

Finalmente, si bien la fase de despliegue no será abordada en este proyecto, se considera su rol esencial en aplicaciones reales, donde la integración del modelo con sistemas operativos, la monitorización en tiempo real y la capacidad de adaptación ante nuevos patrones de fraude son aspectos clave para su efectividad y sostenibilidad.

A lo largo del proyecto se prestará especial atención al diseño del flujo de trabajo de datos, con el objetivo de evitar errores comunes que pueden comprometer la validez de los modelos, como la fuga de datos. Este fenómeno, que ocurre cuando información del conjunto de prueba influye inadvertidamente en el entrenamiento, puede generar una falsa percepción de rendimiento y es especialmente difícil de detectar en procesos con múltiples etapas de transformación.

Para mitigar este riesgo, el conjunto de datos será dividido cuidadosamente en particiones separadas para entrenamiento, validación y prueba, manteniendo el conjunto de prueba completamente aislado hasta la fase final. Las transformaciones aplicadas, tanto en el preprocessamiento como en la generación de nuevas variables, se realizarán exclusivamente dentro de los conjuntos de entrenamiento y validación. Esta estrategia garantizará que las métricas obtenidas reflejen el comportamiento real del sistema ante datos no vistos, asegurando la integridad del proceso de evaluación.

Aunque el flujo detallado será abordado en secciones posteriores, desde esta etapa se establece como principio fundamental la separación estricta de los datos y la replicabilidad del proceso, siguiendo buenas prácticas profesionales en ciencia de datos.

CAIO OLIVEIRA QUINAMO
DATA SCIENTIST
caioquinamocontact@gmail.com





2. COMPRENSIÓN DE LOS DATOS

2.1 RECOLLECTAR LOS DATOS INICIALES

La detección temprana de fraudes con tarjetas de crédito es esencial para proteger tanto a los usuarios como a las entidades financieras, evitando cargos indebidos y pérdidas económicas. Para el presente proyecto se ha empleado un conjunto de datos públicos ampliamente utilizado en investigaciones académicas sobre fraude digital.

El dataset recoge 284.807 transacciones realizadas por titulares europeos durante dos días de septiembre de 2013. De ellas, solo 492 fueron clasificadas como fraudulentas (aproximadamente el 0,172 % del total), lo que refleja un fuerte desbalance entre clases, característico de este tipo de problemas.

Las variables predictoras han sido transformadas mediante Análisis de Componentes Principales (PCA) con fines de anonimización. Esto da lugar a 28 componentes principales, denominados V1 a V28. Solo dos variables no han sido transformadas: Time, que mide los segundos desde la primera transacción, y Amount, que representa el valor monetario de cada operación. La variable objetivo es Class, con valor 1 para fraude y 0 para transacción legítima.

Este dataset ha sido elegido por su representatividad en escenarios reales de detección de fraude y su idoneidad para aplicar técnicas avanzadas de modelado en contextos con alta desproporción entre clases. Dado ese desbalance, se priorizará el uso de métricas como AUPRC para una evaluación más adecuada del desempeño del sistema. Se respetarán las atribuciones correspondientes conforme a su licencia ([Open Database License](#)).

Conjunto de datos de referencia: [Credit Card Fraud Detection](#)

2.2 DESCRIBIR LOS DATOS

El conjunto de datos está compuesto por 284.807 registros y 31 columnas, organizadas en un DataFrame. La mayoría de las variables predictoras son continuas, de tipo float64, y han sido transformadas mediante Análisis de Componentes Principales (PCA), lo que dio lugar a 28 componentes principales anónimos (V1 a V28) sin interpretación directa. Esta transformación impide conocer el significado exacto de estas variables, aunque su relación estadística con la clase objetivo puede ser explorada.

Las tres columnas que no han sido transformadas mediante PCA son:

- **Time:** Representa los segundos transcurridos desde la primera transacción registrada. Aunque tiene una naturaleza temporal, su interpretación directa es limitada. Su utilidad será evaluada empíricamente durante el análisis y modelado.
- **Amount:** Indica el importe monetario de la transacción. Es la única variable con significado explícito y se considerará relevante para el modelo, aunque se le aplicará una transformación de escala.
- **Class:** Variable binaria objetivo. Toma el valor 1 si la transacción es fraudulenta y 0 en caso contrario. Es la etiqueta que el sistema debe aprender a predecir.

El dataset ha sido previamente anonimizado y formateado, lo que reduce la necesidad de preprocessamiento en esta etapa. Las transformaciones requeridas (principalmente relacionadas con el escalado de variables) se aplicarán en fases posteriores del flujo de trabajo. En este sentido, el preprocessamiento inicial no supone un desafío técnico significativo, ya que se trata de un dataset específicamente diseñado para experimentación en tareas de detección de fraude.

2.3 EXPLORAR LOS DATOS

2.3.1 Análisis de la variable objetivo

El conjunto de datos original, sin ningún tipo de modificación previa, presenta una estructura de 284.807 registros y 31 variables, sin valores nulos en ninguna de ellas. Esto permite trabajar directamente con los datos sin necesidad de aplicar técnicas de imputación. No obstante, se identificaron 1.081 registros duplicados, de los cuales 1.062 pertenecen a la clase negativa (0) y 19 a la clase positiva (1). Dado el fuerte desequilibrio de clases existente (donde la clase positiva representa menos del 0,2% del total), se ha tomado la decisión de eliminar únicamente los duplicados pertenecientes a la clase negativa, manteniendo aquellos correspondientes a la clase positiva. Esta decisión se justifica por la necesidad de conservar la mayor cantidad posible de información sobre transacciones fraudulentas, ya que su escasez podría limitar la capacidad del modelo para aprender patrones representativos de fraude. Aunque estos duplicados podrían ser fruto de errores en la recogida de datos en tiempo real, también es posible que representen intentos reiterados de fraude por parte de un atacante en un corto período de tiempo. Por tanto, su presencia podría introducir cierto ruido, pero también aportar evidencia crítica sobre comportamientos anómalos. Dado ese posible valor informativo, se ha optado por retenerlos en esta fase del análisis.

CAIO OLIVEIRA QUINAMO
DATA SCIENTIST
caioquinamocontact@gmail.com





2.3.2 Análisis de correlaciones con la variable objetivo

Para comprender mejor las relaciones entre las variables independientes y la variable objetivo (Class), se ha realizado un análisis de correlación. Las cinco variables que presentan mayor correlación (en valor absoluto) con la clase positiva son: V17, V14, V12, V10 y V16. Todas ellas muestran una correlación negativa significativa, lo que sugiere que, a medida que disminuyen sus valores, aumenta la probabilidad de que una transacción sea fraudulenta. Por otro lado, las variables Time y Amount, que son las únicas con significado explícito (fuera del espacio transformado por PCA), presentan una correlación muy baja con la variable objetivo. A pesar de su bajo valor correlacional, se decidió profundizar en su análisis individual, dado que podrían contener patrones relevantes que no son capturados directamente por la correlación lineal. En el notebook adjunto se explora con mayor profundidad el comportamiento de estas variables, tanto desde una perspectiva estadística como visual, evaluando su posible utilidad dentro del pipeline de modelado.

2.3.3 Análisis de las variables «Time» y «Amount»

En términos de distribución, la variable Time muestra una forma relativamente simétrica, mientras que Amount presenta una fuerte asimetría positiva, con un coeficiente de asimetría de 16.978, lo cual evidencia una concentración de valores bajos y una larga cola hacia valores altos. Este hallazgo plantea una inquietud metodológica relevante: **¿hasta qué punto la asimetría afecta al resto de las variables del conjunto de datos?** Características con asimetría pronunciada (ya sea positiva o negativa) pueden suponer un desafío significativo para el rendimiento de los modelos, especialmente en arquitecturas sensibles a la escala y la forma de la distribución, como los autoencoders.

Para abordar esta problemática, se exploró un enfoque de preprocesamiento basado inicialmente en una transformación Yeo-Johnson y una transformación cuantílica, diseñada para corregir la asimetría sin requerir datos estrictamente positivos. Posteriormente, se aplicó un escalado Min-Max para normalizar las características al rango [0, 1]. Esta estrategia pretendía mejorar el aprendizaje del modelo reduciendo la distorsión causada por distribuciones sesgadas y estableciendo una escala uniforme.

No obstante, tras varios experimentos y evaluaciones, se observó que la aplicación exclusiva del escalado Min-Max en todas las características producía mejores resultados y además sobre una transformación cuantílica de la variable Amount (la variable con mayor coeficiente de asimetría en valores absolutos), tanto en términos del error de reconstrucción del autoencoder como en las métricas de clasificación del modelo Random Forest posterior. Este resultado sugiere que el autoencoder fue capaz de adaptarse a la asimetría inherente de la mayoría de las variables o, alternativamente, que la preservación de la forma original de las distribuciones permitió conservar patrones relevantes para la tarea de detección de fraude.

La elección del MinMaxScaler frente a otras alternativas como StandardScaler o RobustScaler se fundamentó en dos razones principales: (1) su capacidad para preservar la forma original de la distribución de los datos (crítica en contextos donde las transformaciones agresivas degradan el rendimiento), y (2) su idoneidad para modelos neuronales, dado que los autoencoders son sensibles a la escala de entrada y se benefician de trabajar con datos normalizados dentro de un rango fijo. Aunque el escalado Min-Max es sensible a valores extremos, en este estudio se optó por conservar los outliers en lugar de recortarlos o eliminarlos. Esta decisión responde al hecho de que los valores atípicos, lejos de ser errores, podrían representar comportamientos anómalos genuinos, relevantes para el aprendizaje del modelo. En tareas de detección de anomalías, eliminar patrones poco frecuentes podría debilitar la capacidad del modelo para identificar precisamente esas situaciones excepcionales.

Se realizó un análisis descriptivo comparativo entre las transacciones legítimas y fraudulentas en las variables Time y Amount. Un hallazgo destacado fue que el valor medio de las transacciones fraudulentas es un 38,23% superior al de las legítimas. Sin embargo, el 50% de las transacciones fraudulentas tienen un valor inferior a 9,25 €, confirmando la fuerte asimetría de la variable Amount. Además, se identificó una diferencia considerable entre los valores máximos de cada clase: la clase 0 (no fraudulenta) alcanza un máximo de 25.691 €, mientras que la clase 1 (fraudulenta) no supera los 2.126 €.

Desde una perspectiva temporal, se observó que la primera transacción fraudulenta se registró 406 segundos después del inicio del conjunto de datos, y la última ocurrió 2.444 segundos antes del final. Esto sugiere una dispersión temporal no uniforme de los eventos de fraude, aunque no se identificaron patrones temporales claros que aportaran valor al modelo.

Adicionalmente, se constató que, en ambas clases, el percentil 99 de Amount es muy inferior a su valor máximo, indicando la existencia de outliers extremos que influyen de forma desproporcionada sobre la media, especialmente en la clase no fraudulenta.

CAIO OLIVEIRA QUINAMO
DATA SCIENTIST
caioquinamocontact@gmail.com





Tras experimentar con la inclusión de la variable Time como entrada del modelo, se concluyó que su presencia degradaba el rendimiento, por lo que fue finalmente excluida del flujo de trabajo. Su comportamiento parecía añadir más ruido que información útil para la tarea de clasificación.

Aunque los coeficientes de asimetría no se utilizaron como criterio directo (sin contar con la variable Amount) para las decisiones de preprocesamiento, se realizó un análisis informativo complementario. Este reveló que aproximadamente el 60% de las variables presentan una asimetría extremadamente elevada, mientras que cuatro variables se encuentran en un rango de asimetría moderada y el resto en niveles aceptables. En la figura siguiente se ilustran estos resultados junto con los coeficientes de asimetría correspondientes a cada variable.

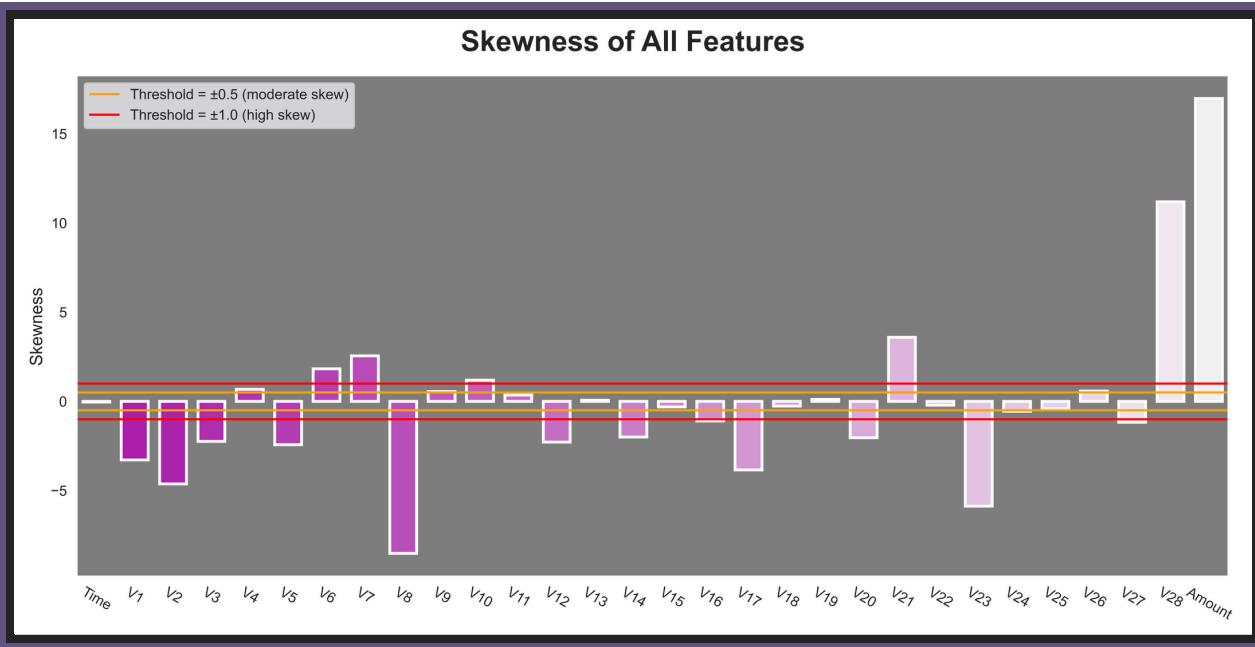


Figura 1. Coeficiente de Asimetría de todas las características

Este diagrama de barras muestra los valores de asimetría de cada característica del conjunto de datos. La asimetría positiva indica una cola derecha más larga, mientras que la negativa indica una cola izquierda más larga. Las características con una asimetría absoluta elevada pueden beneficiarse de la transformación para mejorar el rendimiento del modelo.

2.3.4 Visualización de los datos (antes de la reconstrucción del autoencoder)

Dado que este conjunto de datos contiene numerosas características, no es posible representar visualmente la relación entre todas ellas y la variable objetivo mediante un gráfico bidimensional directo. La alternativa más adecuada consiste en aplicar una técnica de reducción de dimensionalidad que transforme el espacio original en dos dimensiones, facilitando así la exploración visual de la estructura interna del conjunto de datos.

Entre las técnicas disponibles (como PCA (Análisis de Componentes Principales), UMAP (Uniform Manifold Approximation and Projection) y LLE (Locally Linear Embedding)) se optó por t-SNE (t-distributed Stochastic Neighbor Embedding), debido a su capacidad superior para preservar la estructura local de los datos, incluso al proyectarlos en espacios de baja dimensión. A diferencia de PCA, que realiza una transformación lineal y tiende a preservar la varianza global, t-SNE se enfoca en conservar la proximidad entre observaciones similares. Esta propiedad resulta especialmente relevante en tareas como la detección de anomalías o el análisis de clases poco representadas, donde pequeñas diferencias pueden ser decisivas.

Aunque t-SNE no es útil como técnica de clasificación directa ni permite reconstruir las variables originales, es una herramienta poderosa para la exploración visual, ya que permite detectar posibles agrupamientos, superposiciones o estructuras subyacentes. En este caso, se aplicó t-SNE sobre una muestra aleatoria del 10% de las transacciones verídicas, mientras que del conjunto de transacciones fraudulentas se incluyeron todos los casos disponibles. Esta estrategia busca reducir la carga computacional sin comprometer la representatividad de la distribución general de los datos.

CAIO OLIVEIRA QUINAMO
DATA SCIENTIST
caioquinamocontact@gmail.com



Caio Oliveira Quinamo
Data Scientist
caioquinamocontact@gmail.com



El propósito de esta visualización es explorar si existe algún tipo de separación o agrupamiento natural entre las transacciones legítimas y fraudulentas. Esta información es útil para fundamentar decisiones sobre la aplicación de técnicas más sofisticadas, como el uso de autoencoders. En la visualización resultante, si bien no se observa una separación claramente definida entre ambas clases, sí se identifican zonas con mayor densidad de transacciones fraudulentas, así como regiones más dispersas que podrían indicar diferencias sutiles en la estructura interna de los datos. No obstante, muchas transacciones de ambas clases aparecen superpuestas y distribuidas de forma caótica, lo que motivó el diseño de un autoencoder con el objetivo de aprender una representación interna más discriminativa del conjunto.

En este contexto, resulta conveniente generar la misma visualización tras el entrenamiento del autoencoder, utilizando el espacio latente aprendido. Esto permitirá analizar visualmente la capacidad del modelo para mejorar la separación entre clases al transformar la estructura interna de los datos, ofreciendo así un criterio adicional para evaluar su efectividad.

2.3.5 Detección univariada de valores atípicos mediante normalización Z-Score

A partir del análisis exploratorio de las variables Time y Amount, surgió el interés por estudiar la presencia de valores atípicos (outliers). Si bien estos no fueron eliminados del modelo final, esta decisión fue tomada tras una serie de pruebas sistemáticas, incluyendo el ajuste de umbrales de Z-Score para suprimir únicamente los outliers más extremos, así como experimentos que mantuvieron ciertos valores atípicos de menor magnitud dentro de la clase no fraudulenta. No se realizaron intervenciones sobre los outliers pertenecientes a la clase fraudulenta, dado que eliminar observaciones de la clase minoritaria podría comprometer seriamente la representatividad del modelo.

Los resultados mostraron que la exclusión de outliers aumentaba la sensibilidad del modelo combinado (Autoencoder + Random Forest), especialmente en presencia de datos con comportamiento atípico. En particular, cuando se entrenaba el modelo exclusivamente con datos "típicos", se observaba un aumento significativo en los falsos positivos al aplicarlo sobre transacciones legítimas pero atípicas. Esto sugiere una pérdida de capacidad de generalización frente a la variabilidad legítima presente en situaciones reales. Si bien mantener los outliers puede elevar ligeramente el número de falsos negativos, permite una reducción considerable de los falsos positivos, lo cual resulta preferible en un contexto donde los falsos positivos pueden tener un alto coste operativo.

Además, se observó que al entrenar el autoencoder sin outliers, se obtenía una representación latente más separada entre clases, aunque esto implicaba una menor cobertura del espectro real de comportamiento no fraudulento. Este compromiso entre separación estructural y generalización motivó la decisión de conservar los outliers moderados dentro del conjunto de entrenamiento.

El análisis de valores atípicos se orientó, principalmente, a detectar patrones característicos en la clase positiva. Dado que las transacciones fraudulentas buscan imitar operaciones legítimas, pero suelen contener alguna desviación puntual, se hipotetiza que muchas de ellas presentan al menos un valor atípico univariado.

Para la detección de estos valores se aplicó una estandarización mediante Z-Score, calculando la distancia de cada observación con respecto a la media de su variable, normalizada por su desviación estándar. Se estableció un umbral de ± 3 unidades z, clasificando como atípicos aquellos valores que lo superaban. Sin embargo, eliminar todas las transacciones que presentan al menos un outlier univariado habría llevado una pérdida excesiva de datos legítimos, introduciendo además un sesgo de muestreo considerable. Esta limitación se deriva del enfoque univariado, que analiza cada variable de forma independiente e ignora las correlaciones entre características.

Para facilitar el análisis visual de estos outliers, se aplicó una técnica de reducción de dimensionalidad mediante PCA (Análisis de Componentes Principales), elegida por su capacidad de preservar la varianza global del conjunto de datos. A diferencia de métodos como t-SNE, que priorizan la estructura local, PCA proporciona una representación más coherente con las distancias estadísticas originales, además de ser más eficiente computacionalmente.

Las variables más predictivas de fraude (V14, V17, V10, V16, V3 y V12), identificadas previamente por su alta correlación con la variable objetivo, presentan una marcada concentración de valores atípicos dentro de la clase fraudulenta. Este comportamiento sugiere que las transacciones fraudulentas tienden a desviarse de los patrones estadísticos normales, generando anomalías en características clave. Al mismo tiempo, estas variables también presentan una cantidad moderada de outliers en la clase no fraudulenta, lo que indica que no todos los outliers son fraude, pero sí que muchos fraudes implican outliers. Esta dualidad justifica el uso de dichas variables tanto como predictoras directas como para la ingeniería de nuevas características, por ejemplo, mediante la creación de indicadores binarios de outlier o conteos de z-scores extremos, siempre respetando la necesidad de no excluir transacciones legítimas con comportamientos atípicos válidos.

CAIO OLIVEIRA QUINAMO
DATA SCIENTIST
caioquinamocontact@gmail.com







2.3.6 Detección multivariada de valores atípicos mediante la distancia de Mahalanobis

Tras la detección univariada de valores atípicos con Z-Score, se implementó un enfoque complementario para identificar outliers multivariados mediante la distancia de Mahalanobis. A diferencia del análisis univariado, este método permite capturar relaciones entre múltiples variables, al considerar la covarianza entre ellas. Así, evalúa cuán lejos se encuentra una observación del centro multivariado del conjunto de datos.

La elección de esta técnica se debe a su capacidad para identificar observaciones que, aunque no destaque en ninguna variable por separado, resultan atípicas en combinación. En el contexto del fraude financiero, esto resulta especialmente útil: las transacciones fraudulentas pueden aparentar normalidad en cada variable individual, pero desviarse al considerar su interacción conjunta.

Para establecer un umbral de decisión, se adoptó un criterio basado en percentiles, concretamente el valor correspondiente al percentil 99.9 de las distancias de Mahalanobis calculadas sobre los datos estandarizados con Z-Score. Este umbral permite enfocar la detección en los casos más extremos (aproximadamente el 0.1% de las observaciones), equilibrando la sensibilidad ante anomalías sin comprometer en exceso el volumen de datos legítimos. Esto resulta crucial en conjuntos con clases desbalanceadas, donde eliminar en exceso observaciones no fraudulentas pero inusuales podría introducir sesgos.

Cabe subrayar que tanto en la detección univariada como en la multivariada, la identificación de outliers es en gran medida subjetiva. Depende del contexto, los objetivos del análisis y el criterio del analista. La elección del umbral debe adaptarse al caso de uso, considerando su impacto en la interpretación y la capacidad de generalización del modelo. Un umbral más estricto puede favorecer la detección de fraudes de alto impacto, mientras que uno más laxo puede ser preferible en entornos donde se prioriza la reducción de falsos positivos.

Entre las principales fortalezas de la distancia de Mahalanobis destacan:

- Su consideración de la estructura multivariada del conjunto de datos, integrando correlaciones entre variables.
- Su carácter estadísticamente sólido bajo la suposición de normalidad multivariante.
- Su simplicidad computacional y capacidad de escalar razonablemente bien en datasets de tamaño medio.

No obstante, presenta limitaciones. Es sensible a la presencia de valores extremos (aunque en este caso se mitigó mediante estandarización) y depende de una matriz de covarianza bien condicionada, lo que puede ser problemático en datos con alta multicolinealidad o muchas dimensiones.

Existen otras técnicas para la detección multivariada de outliers, como Isolation Forest, un método basado en árboles que no requiere supuestos de distribución y es eficaz con grandes volúmenes de datos; One-Class SVM, útil cuando se dispone de una sola clase mayoritaria y se busca modelar su frontera; y LOF (Local Outlier Factor), que mide la densidad local de cada observación en comparación con sus vecinos, siendo útil para detectar estructuras complejas. Sin embargo, para este análisis se optó por Mahalanobis por su interpretabilidad estadística, su coherencia con la estandarización previa y su eficacia en un conjunto bien estructurado tras la reducción con PCA.

Aplicando este enfoque, se observó que aproximadamente el 22% de los casos de fraude fueron detectados como valores atípicos multivariados. Este resultado es esperable, ya que muchas transacciones fraudulentas no se comportan como atípicas globales: los atacantes suelen camuflarse dentro del comportamiento general, especialmente cuando el fraude no implica montos elevados ni desviaciones extremas. Mahalanobis resulta eficaz para detectar outliers globales, pero no aquellos que dependen de un contexto individual, rompen patrones personales o presentan relaciones no lineales. Así, ciertas transacciones pueden no alejarse del centro multivariado, pero siguen siendo anómalas bajo otras perspectivas que este método lineal no capta.

CAIO OLIVEIRA QUINAMO
DATA SCIENTIST
caioquinamocontact@gmail.com





2.4 VERIFICAR LA CALIDAD DE LOS DATOS

El conjunto de datos analizado presenta una estructura válida y adecuada para análisis académicos y demostrativos. Aunque cuenta con un número considerable de características, la mayoría están codificadas y carecen de un contexto explícito. La falta de información contextual relevante, como la dirección IP, el país de origen o el tipo de dispositivo, limita la posibilidad de realizar un análisis conductual más profundo sobre el fraude, especialmente desde una perspectiva centrada en el comportamiento del usuario.

Desde un punto de vista técnico, la calidad estructural del dataset es elevada. No se detectaron valores faltantes, lo cual simplificó el diseño del flujo de trabajo al eliminar la necesidad de técnicas de imputación. En cuanto a duplicados, se eliminaron las filas repetidas de la clase negativa (no fraude), mientras que los registros duplicados en la clase positiva (fraude) se conservaron, dada su escasez y la importancia de capturar la mayor variabilidad posible en una clase minoritaria.

Las variables presentan distribuciones altamente asimétricas, algunas con sesgos hacia valores extremos. Sin embargo, no se aplicaron transformaciones, ya que los modelos utilizados (Autoencoder y Random Forest) demostraron ser robustos frente a estas características. Tampoco se eliminaron outliers, ya que se consideró que aportaban valor al proceso de aprendizaje del Autoencoder al incluir tanto patrones comunes como transacciones legítimas poco frecuentes.

Dado el elevado número de variables, se implementó una selección de características mediante el método SelectKBest con la función de puntuación mutual_info_classif. Esta técnica permitió reducir el conjunto a las 25 variables más relevantes para el modelo Random Forest. La selección se realizó tras la codificación con Autoencoder, considerando que esta arquitectura transforma la representación latente de los datos. Por tanto, se optó por que el modelo final determinara qué variables codificadas eran más útiles para la clasificación. Durante este proceso se descartaron 6 variables, entre ellas Time, y se incorporó el Reconstruction Error como nueva característica clave.

Aunque el fuerte desbalance entre clases representa de forma realista el comportamiento típico del fraude en la práctica, la baja proporción de la clase positiva limitó la capacidad del modelo para aprender profundamente de ella. Aun así, se decidió mantener la distribución original y aplicar ajustes adaptativos, como la estrategia class_weight='balanced' y el ajuste personalizado del umbral de decisión, en lugar de recurrir a técnicas de sobremuestreo.

En términos generales, el conjunto de datos resulta especialmente útil con fines didácticos: permite experimentar con distintos enfoques, aplicar técnicas variadas de ingeniería de características y evaluar múltiples estrategias para la detección de fraude. Sin embargo, para un uso operativo, su aplicabilidad es limitada debido a la falta de contexto real del comportamiento del usuario, lo que restringe tanto el análisis exploratorio como el diseño de modelos más especializados.

Según la evaluación automática proporcionada por Kaggle, este conjunto recibió una puntuación de 8,53 sobre 10 en términos de usabilidad. Esta valoración se basó en los siguientes criterios:

Completeness · 100%

Check: Subtitle, Check: Tag, Check: Description, Check: Cover Image

Credibility · 33%

Close: Source/Provenance, Check: Public Notebook, Close: Update Frequency

Compatibility · 100%

Check: License, Check: File Format, Check: File Description, Check: Column Description

CAIO OLIVEIRA QUINAMO
DATA SCIENTIST
caioquinamocontact@gmail.com





3. PREPARACIÓN DE DATOS

3.1 SELECCIONAR LOS DATOS

Antes del entrenamiento del modelo Autoencoder, se consideraron todas las variables disponibles, con excepción de la variable Time, que fue descartada al no representar una métrica temporal interpretable ni aportar información relevante al problema. Esta variable posee una escala acumulativa sin correspondencia directa con unidades cronológicas reales, lo que limita su utilidad para modelar patrones transaccionales en el tiempo.

La variable objetivo utilizada fue Class, codificada como 0 para transacciones legítimas y 1 para transacciones fraudulentas. El conjunto de datos presenta una fuerte desproporción de clases, con la clase positiva representando solo el 0,18% del total de observaciones. Para abordar este desbalance, la partición en conjuntos de entrenamiento, validación y prueba se realizó mediante estratificación sobre la variable objetivo, lo que garantiza la presencia de ejemplos de fraude en cada subconjunto. Esta decisión implicó prescindir del orden temporal en la partición, priorizando la representatividad estadística por encima de la secuencia cronológica, dado que mantener dicha secuencia podría haber derivado en conjuntos sin presencia de la clase minoritaria, afectando negativamente el cálculo de métricas sensibles como el área bajo la curva de precisión-sensibilidad (AUPRC).

Antes del entrenamiento final del modelo Random Forest, se aplicó un proceso de selección de características mediante SelectKBest, optimizando el valor de k a través de validación cruzada estratificada, utilizando AUPRC como métrica objetivo. El valor óptimo fue $k = 20$, que se utilizó como hiperparámetro fijo en la integración del pipeline SelectKBest + RandomForestClassifier.

Se exploraron también alternativas como RFE (Recursive Feature Elimination), pero se optó por SelectKBest por su mayor simplicidad, su interpretabilidad directa respecto a las variables originales (sin necesidad de introspección sobre un estimador), y su integración nativa en scikit-learn, lo que facilitó su uso en flujos reproducibles y auditables. Estas propiedades resultan especialmente relevantes en contextos como el de la detección de fraude, donde la transparencia del proceso de selección de variables es un requerimiento crítico.

Para una descripción visual de este flujo de preparación, incluyendo los criterios de selección de variables y la lógica de las particiones, se remite al apartado 3 del notebook, donde se incluye un diagrama esquemático que resume la arquitectura de datos previa al modelado.

3.2 LIMPIAR LOS DATOS

Se identificaron filas duplicadas dentro de la clase no fraudulenta, las cuales fueron eliminadas para evitar sesgos innecesarios durante el aprendizaje. Su impacto estadístico era marginal, dado que representaban un porcentaje muy bajo del total. En cambio, en la clase fraudulenta se conservaron todos los registros, incluidos los duplicados, debido a la extrema escasez de observaciones, donde cada muestra resulta valiosa para el aprendizaje del modelo.

No se encontraron valores faltantes en ninguna variable. Se detectó una alta presencia de valores atípicos en múltiples características, los cuales no fueron eliminados, ya que podrían corresponder a comportamientos extremos válidos, particularmente relevantes en el contexto del fraude financiero. Estos outliers fueron aprovechados durante el entrenamiento del Autoencoder, que solo utilizó la clase legítima, ya que aportaban mayor variabilidad dentro de lo considerado "normal", favoreciendo así la capacidad del modelo para detectar anomalías.

Se evaluaron distintos métodos de escalado: RobustScaler, PowerTransformer y MinMaxScaler. Este último fue seleccionado por ofrecer el mejor desempeño en términos de reconstrucción, debido a las siguientes razones: Las funciones de activación utilizadas (tanh en las capas ocultas y ReLU en la capa de salida) operan de manera más eficiente con datos en el rango [0, 1]. MinMaxScaler conserva las proporciones originales entre valores, lo que favorece la fidelidad estructural del aprendizaje. A diferencia de otras transformaciones que generan valores negativos, MinMaxScaler evita conflictos con ReLU, cuya salida está restringida a valores no negativos.

Adicionalmente, se aplicó un QuantileTransformer únicamente sobre la variable Amount, que presentaba una asimetría extrema. Bajo el escalado con MinMaxScaler, su distribución se aplanaba, generando valores relativos muy bajos y reconstrucciones cercanas a cero para todas las filas. Esto anulaba su capacidad predictiva al eliminar cualquier correlación con la variable objetivo. La transformación mediante cuantiles permitió preservar mejor su relevancia informativa.

Tanto el escalador como el transformador fueron entrenados exclusivamente sobre los datos de la clase legítima ($X_{0\text{train}}$) para evitar data leakage. Como consecuencia, algunos valores de validación y prueba quedaron fuera del rango [0, 1], aunque el modelo demostró una buena capacidad de generalización ante estas situaciones.

CAIO OLIVEIRA QUINAMO
DATA SCIENTIST
caioquinamocoutac@gmail.com



Caio Oliveira Quinamo





Finalmente, aunque la mayoría de las variables presentaban asimetrías elevadas, no se aplicaron transformaciones como logaritmos o Box-Cox, dado que su impacto fue nulo o negativo en el rendimiento general del modelo, con la única excepción ya señalada de Amount.

3.3 CONSTRUIR LOS DATOS

Con el objetivo de enriquecer el conjunto de datos antes del entrenamiento supervisado, se construyó una nueva variable derivada del error de reconstrucción generado por un autoencoder entrenado exclusivamente sobre transacciones legítimas. Esta variable captura el grado de disconformidad de cada transacción respecto a los patrones normales aprendidos por el modelo, y fue añadida como una característica adicional al conjunto original.

La utilidad de esta nueva variable se evaluó mediante el coeficiente de correlación de Pearson respecto a la variable objetivo, obteniéndose un valor moderadamente positivo. Esto sugiere que el error de reconstrucción aporta información útil para discriminar entre transacciones legítimas y fraudulentas, aunque no fue empleado como umbral de decisión directa.

Además, se aplicó la técnica de reducción de dimensionalidad t-SNE para visualizar el espacio de características antes y después de la transformación realizada por el autoencoder. La comparación mostró una mayor separación estructural entre clases tras la reconstrucción, lo que respalda la utilidad del autoencoder como una etapa previa de compresión y enriquecimiento del espacio de representación, facilitando así el aprendizaje supervisado posterior.

3.4 INTEGRAR Y FORMATEAR LOS DATOS

Se definió una estrategia de integración de datos que garantiza la independencia entre los subconjuntos de entrenamiento, validación y prueba, evitando cualquier tipo de fuga de información a lo largo del flujo de trabajo. La Figura 1 del código (apartado 3) ilustra este esquema general de particiones, preservando la integridad experimental desde la preparación inicial hasta la evaluación final.

Durante el desarrollo del modelo no se implementó un pipeline automatizado mediante sklearn. En su lugar, se optó por una integración manual y progresiva de los distintos pasos de transformación y limpieza, lo que permitió un mayor control y trazabilidad en cada fase del proceso. Esta decisión fue especialmente relevante dada la naturaleza secuencial del flujo de trabajo: el conjunto de datos fue enriquecido y adaptado iterativamente a medida que se aplicaban distintas técnicas, como la ingeniería de características o la compresión mediante autoencoders.

Si bien no se aplicó un flujo automatizado unificado durante la fase de construcción del modelo, se diseñó posteriormente una función modular capaz de reproducir este proceso completo de preprocesamiento. Dicha función se describe en el apartado 5 (Evaluación), donde se utiliza para transformar nuevos datos de entrada y aplicar el modelo entrenado de forma coherente y segura.



RATIO ET INTEGRITAS DUCUNT AD
PROSPERITATEM





4. MODELADO

4.1 SELECCIÓN DE TÉCNICAS DE MODELADO

La elección de la técnica de modelado surgió durante el análisis exploratorio de los datos (EDA), particularmente al intentar visualizar la estructura del conjunto en dos dimensiones. Dado el volumen elevado de datos y la disponibilidad limitada de recursos computacionales, se optó por aplicar una reducción de dimensionalidad mediante la técnica t-SNE sobre una muestra del dataset. Para ello, se seleccionó aleatoriamente un 10% de las transacciones no fraudulentas y se incluyeron todas las transacciones fraudulentas.

El análisis del gráfico reveló que las transacciones seguían una distribución con forma espiral, donde las instancias fraudulentas se encontraban completamente superpuestas con las no fraudulentas. Esta observación planteó una cuestión central: **¿existe alguna forma de transformar el espacio de representación de manera que se facilite la separación entre ambas clases, incluso si no es visualmente evidente en la dimensión reducida?**

Como respuesta, se decidió implementar un autoencoder con el objetivo de transformar los datos de modo que se enfatizaran las diferencias sutiles entre las clases. La idea fue entrenar el autoencoder exclusivamente con datos de la clase mayoritaria (no fraude), para que este aprendiera su estructura con el mayor detalle posible. De este modo, al enfrentar datos fraudulentos (que no había visto durante el entrenamiento) el autoencoder generaría reconstrucciones deficientes, aumentando el error de reconstrucción para estos casos. Posteriormente, este error sería utilizado como una nueva variable, combinada con la salida reconstruida del autoencoder, como entrada para un clasificador basado en Random Forest.

Durante la fase experimental del autoencoder, se evaluó el impacto de eliminar valores atípicos univariados de la clase negativa. Esta eliminación se realizó aplicando una estandarización mediante Z-score, con un umbral más conservador de ± 5.5 unidades z, buscando excluir solo los outliers extremos sin eliminar información relevante. Se observó que, al eliminar estos valores, el rango del error de reconstrucción para las transacciones fraudulentas aumentaba, lo que en teoría incrementaba la capacidad discriminativa del modelo.

Sin embargo, pruebas posteriores demostraron que al entrenar el autoencoder con datos que excluían estos outliers, el clasificador Random Forest tendía a etiquetar muchos casos verídicos atípicos como fraude, afectando negativamente la capacidad de generalización. Aunque se mejoraba la detección de fraudes, se generaban demasiados falsos positivos, lo que comprometía la utilidad del modelo en un contexto real. Por tanto, se tomó la decisión de conservar los valores atípicos en el conjunto de entrenamiento, a fin de construir un modelo más equilibrado, capaz de identificar un mayor número de fraudes sin comprometer la clasificación correcta de los casos negativos atípicos.

Una vez identificado que la inclusión de valores atípicos en el conjunto de entrenamiento reducía ligeramente la capacidad discriminativa del autoencoder, se consideró necesario compensar esta pérdida optimizando al máximo el modelo de clasificación. Para ello, se evaluaron diferentes algoritmos de clasificación binaria con el objetivo de maximizar el poder predictivo tanto en el conjunto de entrenamiento como en datos no vistos (conjunto de prueba).

Entre los modelos considerados, los dos principales candidatos fueron XGBoost y Random Forest Classifier (RFC). En las pruebas iniciales, XGBoost obtuvo una puntuación ligeramente superior en la métrica área bajo la curva de precisión-sensibilidad (AUPRC), con una mejora absoluta de 0,33% respecto a RFC, además de mostrar una desviación estándar levemente menor (aunque estadísticamente poco significativa). A pesar de esta ligera ventaja, se optó por utilizar Random Forest como clasificador principal. Esta decisión se tomó por criterio del analista, valorando la estabilidad, interpretabilidad, y facilidad de integración del modelo RFC dentro del flujo de trabajo, sin desestimar que XGBoost también sería una alternativa muy válida para este problema.

El enfoque propuesto integra tres algoritmos complementarios que operan en conjunto para optimizar el desempeño del sistema de clasificación:

- 1. Autoencoder:** entrenado exclusivamente con datos de la clase negativa (no fraude), su propósito es reconstruir las entradas y generar el error de reconstrucción como una feature adicional, lo que permite cuantificar las discrepancias entre clases en términos de reconstrucción, facilitando la detección de anomalías.
- 2. SelectKBest:** algoritmo de selección de características basado en la métrica de información mutua para clasificación. Esta técnica permite identificar las variables más relevantes para el modelo final, optimizando el rendimiento sin introducir complejidad innecesaria.





3. Random Forest Classifier: modelo de clasificación robusto y eficiente, utilizado para realizar predicciones finales sobre los datos procesados por las etapas anteriores. Este clasificador fue ajustado posteriormente mediante optimización de hiperparámetros, con el foco específico en mejorar el rendimiento bajo la métrica AUPRC.

Inicialmente, se utilizó SelectKBest sin ajuste de hiperparámetros, con el objetivo de seleccionar las mejores variables antes de construir un estimador inicial. Esta configuración preliminar permitió establecer una base sobre la cual se desarrolló posteriormente un flujo de trabajo completo. Aunque se contempló el uso de RFE (Recursive Feature Elimination) con Random Forest como método alternativo de selección, se optó por SelectKBest debido a su bajo costo computacional, facilidad de integración en etapas de procesamiento secuencial en scikit-learn y buen desempeño observado en las validaciones cruzadas.

En resumen, el modelo completo está compuesto por tres componentes interdependientes: el autoencoder para transformar y enriquecer los datos, el método de selección de características para reducir la dimensionalidad de forma efectiva, y el clasificador Random Forest como núcleo del sistema predictivo. Esta arquitectura fue diseñada para maximizar la eficiencia y la capacidad de detección del modelo, especialmente en escenarios donde la clase positiva (fraude) es escasa y difícil de distinguir.

4.2 GENERAR EL DISEÑO DE PRUEBAS

Con el objetivo de abordar el problema de manera rigurosa, se definió que la métrica principal a optimizar sería el Área Bajo la Curva de Precision-Recall (AUPRC), ya que el foco está en la predicción eficaz de la clase positiva (fraude). Esta métrica es especialmente adecuada en escenarios de clasificación desbalanceada, dado que se centra exclusivamente en el comportamiento del modelo respecto a los verdaderos positivos, sin estar influida por la proporción elevada de verdaderos negativos, como ocurre con métricas como la precisión o AUROC.

AUPRC penaliza directamente los falsos positivos a través de la métrica de precisión, lo cual resulta crucial en este contexto, donde una predicción incorrecta puede conllevar consecuencias significativas (por ejemplo, el bloqueo de transacciones legítimas). Además, proporciona una evaluación detallada sobre la calidad del ranking del modelo, ya que un AUPRC elevado no solo implica detección efectiva de positivos, sino que también indica que estos se están ordenando correctamente según su probabilidad estimada. Esta propiedad es particularmente valiosa en escenarios donde se desea aplicar un umbral dinámico o establecer una priorización de alertas.

Por estas razones, AUPRC fue seleccionado como métrica objetivo tanto para la etapa de selección de características como para la optimización de los hiperparámetros del clasificador Random Forest, con el propósito de maximizar su capacidad discriminativa en la clase minoritaria.

Dado que AUPRC no requiere un umbral de decisión para su cálculo, una vez entrenado el modelo se llevó a cabo una etapa posterior de optimización del umbral de clasificación, utilizando como criterio F1-score. Esta métrica, al representar el promedio armónico entre precisión y sensibilidad, es adecuada en contextos con fuerte desbalance, ya que permite capturar simultáneamente la capacidad del modelo para detectar correctamente los positivos (sensibilidad) y su capacidad para evitar errores al predecirlos (precisión).

La optimización del umbral cumple una función correctiva frente a la asimetría natural del dataset, permitiendo ajustar el punto de corte de la predicción probabilística del modelo para lograr un equilibrio más eficaz entre los errores tipo I (falsos positivos) y tipo II (falsos negativos). Así, se traduce la salida continua del modelo en una decisión binaria robusta, adaptada al coste diferencial de los errores en este tipo de problema, sin comprometer la evaluación global proporcionada por AUPRC.

Para garantizar la robustez y fiabilidad de las métricas de evaluación del modelo, se implementaron múltiples estrategias de selección de modelos, siendo esta una de las fases más demandantes a nivel computacional. En total, se aplicaron técnicas de validación en tres momentos clave del flujo de trabajo.

La primera correspondió al proceso de selección de características mediante validación cruzada estratificada, manteniendo la proporción original de clases en cada pliegue. Se utilizó un esquema de 10 pliegues, evaluando la métrica AUPRC para diferentes valores de k en el algoritmo SelectKBest. La selección del número óptimo de características se realizó empleando un modelo base de Random Forest, permitiendo así elegir las variables que mejor contribuían a la detección de la clase positiva.

CAIO OLIVEIRA QUINAMO
DATA SCIENTIST
caioquinamocontact@gmail.com









La segunda instancia se centró en la optimización de hiperparámetros del clasificador Random Forest, utilizando el método RandomizedSearchCV con 50 iteraciones, cada una evaluada mediante validación cruzada estratificada de 10 pliegues. La métrica de optimización fue nuevamente AUPRC, garantizando que los hiperparámetros seleccionados maximizaran la capacidad del modelo para distinguir correctamente la clase minoritaria. Se mantuvo una semilla fija durante todo el proceso para asegurar la reproducibilidad y la comparabilidad entre ejecuciones. Cabe destacar que la estratificación fue una constante en todos los procesos de evaluación, dado el alto grado de desbalance entre clases y la necesidad crítica de que el modelo aprenda representaciones significativas de ambas. Para la realización del ajuste fino de los ajustes de los hiperparámetros correspondientes a la profundidad máxima de cada árbol y el número mínimo de muestras para dividir las ramas de cada árbol, el método de selección de modelos para ambos ha sido GridSearchCV con la misma validación cruzada utilizada anteriormente.

En una tercera etapa, y con el modelo ya optimizado, se procedió a realizar una última validación cruzada con el fin de determinar el mejor umbral de decisión. Para ello, se utilizaron las probabilidades de salida del modelo en el conjunto de entrenamiento, seleccionando aquel umbral que maximice el F1-score. Esta métrica, al depender directamente del umbral, complementa adecuadamente a AUPRC y permite ajustar la salida probabilística del modelo a una predicción binaria que equilibre precisión y sensibilidad, lo cual es crítico en el contexto del problema.

Finalmente, se elaboró un esquema visual integrador que resume las fases del proceso de modelado y validación, incluyendo los puntos donde se aplicaron cada una de las técnicas de evaluación. Este esquema permite visualizar de forma clara la arquitectura del pipeline y la lógica detrás de cada decisión metodológica.

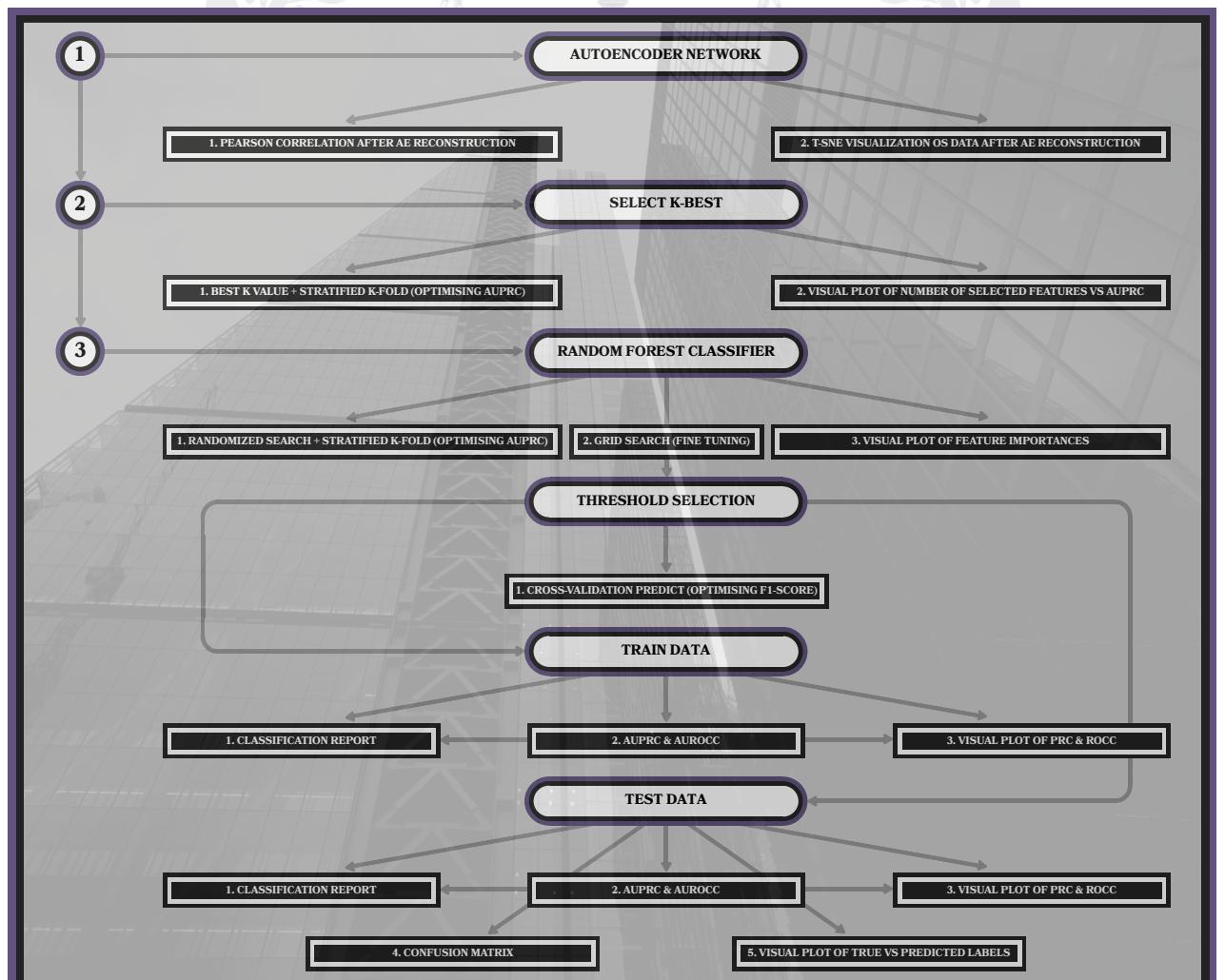


Figura 2. Diseño de pruebas del proceso de modelado propuesto

Diagrama del flujo de pruebas aplicadas en cada etapa del proceso de modelado, incluyendo visualizaciones, métricas y validaciones utilizadas para evaluar el rendimiento del autoencoder, la selección de características y el clasificador Random Forest.

CAIO OLIVEIRA QUINAMO
DATA SCIENTIST
caioquinamocontact@gmail.com





4.3 CONSTRUIR LA CONFIGURACIÓN DE LOS PARÁMETROS DEL MODELO

ARQUITECTURA DEL AUTOENCODER

El autoencoder fue implementado mediante la API funcional de Keras con TensorFlow como backend, adoptando una arquitectura simétrica codificador-decodificador. Esta estructura ($128 \rightarrow 64 \rightarrow 128 \rightarrow 64$) fue diseñada para realizar una compresión progresiva seguida de una reconstrucción en espejo, lo que facilita el aprendizaje de representaciones latentes compactas sin comprometer significativamente la integridad de la información original.

Se utilizaron funciones de activación Tanh en todas las capas ocultas debido a su naturaleza centrada en cero, lo cual mejora la estabilidad durante el entrenamiento y aunque tanh produce salidas en $[-1, 1]$, su simetría centrada en cero aporta estabilidad al entrenamiento en capas ocultas, complementando la entrada previamente escalada a $[0, 1]$ mediante MinMaxScaler.. En contraste, la capa de salida empleó ReLU, ya que los datos están acotados a valores positivos. Esta elección asegura que la reconstrucción no genere valores negativos fuera del dominio esperado y refuerza la función del autoencoder como filtro de anomalías.

Se introdujo una regularización mediante Dropout (30%) tras cada capa del codificador, lo cual resultó crucial dada la naturaleza unilateral del entrenamiento (solo con clase legítima). Esta técnica ayuda a prevenir el sobreajuste y fuerza al modelo a aprender representaciones robustas al desactivar aleatoriamente neuronas durante el entrenamiento, promoviendo así la generalización ante desviaciones fuera de la distribución entrenada.

El entrenamiento se llevó a cabo utilizando el optimizador Adam, ideal por su adaptabilidad y eficiencia en problemas de regresión. La función de pérdida fue el error cuadrático medio (MSE) entre la entrada y su reconstrucción, permitiendo una cuantificación directa del error dimensional. Para la validación, se reservó el 30% de los datos legítimos del conjunto de entrenamiento, lo que permitió monitorear el comportamiento del modelo sobre ejemplos no vistos. Tras varios ensayos, se definieron los siguientes parámetros: batch size de 256 y 25 épocas, encontrándose que el modelo convergía adecuadamente sin evidencia de sobreajuste, fijando semillas aleatorias para asegurar la reproducibilidad total del experimento.

El error de reconstrucción generado por el autoencoder se utilizó como variable principal en la detección de fraude, resultando ser la característica de mayor relevancia según la importancia atribuida por el modelo Random Forest posterior. Este hallazgo refuerza la utilidad del autoencoder como filtro inicial basado en anomalías.

Respecto a la generalización fuera del rango $[0, 1]$, se validó que el modelo mantenía su rendimiento al enfrentarse a datos de validación y prueba, a pesar de que el escalador se ajustó únicamente sobre X0train. No se aplicó early stopping, ya que se optó por controlar el sobreentrenamiento mediante regularización explícita y evaluación cruzada posterior.

En conjunto, el diseño del autoencoder se basó en principios de robustez, interpretabilidad y capacidad de generalización, alineándose plenamente con el objetivo de detección de anomalías. Cada componente de su arquitectura fue validado empíricamente y se integró de manera coherente en el flujo de trabajo global del modelo.

SELECCIÓN DE CARACTERÍSTICAS CON SELECTKBEST

Para reducir la dimensionalidad del conjunto de datos y mejorar la eficiencia del modelo, se aplicó la técnica de selección univariada SelectKBest, utilizando como función de puntuación la información mutua (mutual_info_classif). Si bien esta métrica está diseñada principalmente para variables discretas, su aplicación ha demostrado ser eficaz también en escenarios con variables continuas, ya que permite capturar relaciones no lineales entre las características y la variable objetivo.

La selección del número óptimo de características se realizó empíricamente mediante validación cruzada estratificada de 10 pliegues, buscando maximizar la métrica de precisión promedio (AUPRC). Para este fin, se empleó un flujo de trabajo sencillo con un modelo Random Forest sin optimización de hiperparámetros, salvo la fijación de una semilla aleatoria y el peso de las clases definido en balanceado. Como resultado, se identificó que el mejor rendimiento se alcanzaba utilizando 20 características, configurando así la dimensionalidad final del conjunto de entrada al modelo.

RANDOM FOREST, OPTIMIZACIÓN DE HIPERPARÁMETROS Y ARQUITECTURA FINAL

El clasificador Random Forest fue seleccionado por su idoneidad frente a conjuntos de datos con alta dimensionalidad, relaciones no lineales y ruido inherente, características comunes en escenarios de detección de fraude. Este algoritmo no requiere normalización previa de los datos y destaca por su robustez frente al sobreajuste, su capacidad de manejar distribuciones desbalanceadas y su interpretabilidad a través de la evaluación de la importancia de variables.

CAIO OLIVEIRA QUINAMO
DATA SCIENTIST
caioquinamocoutac@gmail.com





El modelo fue entrenado sobre un conjunto de datos enriquecido, incluyendo como variable adicional el error de reconstrucción generado por el autoencoder. Esta nueva característica resultó clave al aportar una dimensión adicional sobre la desviación de cada observación respecto al patrón dominante aprendido, fortaleciendo así la capacidad del modelo para identificar casos atípicos.

Para optimizar el rendimiento del clasificador, se aplicó RandomizedSearchCV sobre un espacio definido de hiperparámetros, con validación cruzada estratificada. Se exploraron las siguientes dimensiones: **n_estimators**: Se ajustó el número de árboles del bosque para lograr un balance entre rendimiento predictivo y coste computacional. Si bien más árboles incrementan la estabilidad, también llevan rendimientos marginales decrecientes. **max_depth**: La profundidad máxima de los árboles se limitó para evitar que el modelo se sobreajuste al ruido, especialmente en la clase minoritaria. **min_samples_split** y **min_samples_leaf**: Estos parámetros permitieron prevenir divisiones excesivamente específicas que podrían capturar ruido o valores atípicos, favoreciendo así una mejor generalización. **max_features**: Se optimizó el número de variables consideradas en cada división para aumentar la diversidad entre árboles y reducir la correlación interna del modelo, mejorando su capacidad de generalización.

Tras la primera fase de optimización mediante búsqueda aleatoria, se decidió profundizar en dos hiperparámetros clave: la profundidad máxima de cada árbol (**max_depth**) y el número mínimo de muestras requeridas para dividir un nodo (**min_samples_split**). Para ello, se definió una rejilla de búsqueda centrada en los valores óptimos detectados por RandomizedSearchCV, explorando cinco valores hacia arriba y cinco hacia abajo en torno al valor inicial. Esta segunda fase de ajuste fino se llevó a cabo con GridSearchCV, utilizando validación cruzada estratificada y fijando el resto de los hiperparámetros en los valores previamente seleccionados.

La elección de RandomizedSearchCV frente a una búsqueda exhaustiva permitió una exploración eficiente del espacio de búsqueda, reduciendo el tiempo computacional sin comprometer el rendimiento. Este proceso resultó fundamental no solo para optimizar métricas clave como el AUPRC, sino también para conservar la interpretabilidad del modelo, un aspecto esencial en entornos donde la trazabilidad y la validación del sistema son críticas.

La arquitectura definitiva del modelo se implementó como un flujo de trabajo de clasificación, cuyo núcleo es un clasificador Random Forest configurado específicamente para el reto de detección de fraude en un contexto altamente desbalanceado. La configuración final, fruto de la optimización mediante RandomizedSearchCV y validación cruzada estratificada, presenta los siguientes parámetros clave: **n_estimators = 300**: Se estableció un total de 300 árboles para asegurar una predicción robusta y reducir la varianza mediante el voto agregado de múltiples estimadores. **max_depth = 18**: Esta restricción de profundidad limita la complejidad de cada árbol, evitando el sobreajuste especialmente en una clase positiva (fraude) escasa. **min_samples_split = 8**: Se exige un mínimo de 8 muestras para permitir una división, lo que fortalece la generalización del modelo frente a patrones específicos de baja frecuencia. **class_weight = 'balanced'**: Esta configuración permite al algoritmo asignar automáticamente pesos inversamente proporcionales a la frecuencia de cada clase, mejorando la sensibilidad ante el fraude. **random_state = 4**: Se fija una semilla para garantizar la reproducibilidad de los resultados. **n_jobs = -1**: Se habilita el uso de todos los núcleos disponibles para acelerar el proceso de entrenamiento.

Esta combinación de hiperparámetros da lugar a una arquitectura robusta y eficiente, capaz de adaptarse a los desafíos propios del aprendizaje con clases desbalanceadas. El modelo final no solo logra una alta capacidad predictiva, sino que también mantiene un nivel de interpretabilidad y reproducibilidad adecuado para su uso en entornos críticos como la detección de fraude financiero.

4.4 EVALUACIÓN DEL MODELO

COMPORTAMIENTO DEL AUTOENCODER

El autoencoder seleccionado, tras múltiples pruebas empíricas, mostró un comportamiento coherente con su propósito como componente de detección no supervisada. Aunque se optó por conservar valores atípicos durante el entrenamiento (lo que redujo la diferenciación del error de reconstrucción entre clases en algunos casos), se observó una tendencia constante: las transacciones fraudulentas tienden a presentar errores de reconstrucción mayores en comparación con las legítimas.

Este patrón valida la hipótesis fundamental del enfoque: al estar entrenado exclusivamente con datos verídicos, el autoencoder reconstruye deficientemente las observaciones que se desvían del patrón dominante, convirtiendo el error de reconstrucción en una señal efectiva para la detección de anomalías. Más allá del error como variable explícita, el autoencoder también influyó positivamente en la estructura de los datos transformados. En particular, variables como Amount, que habían perdido capacidad discriminativa tras ser sometidas a una normalización, recuperaron relevancia una vez pasadas por la reconstrucción. Se detectó un incremento en su correlación absoluta con la variable objetivo, lo que sugiere que el proceso de reconstrucción no solo conserva, sino que puede amplificar patrones útiles para la clasificación posterior.



A handwritten signature in black ink, appearing to read "Caio Oliveira Quinamo".





Para evaluar la estructura interna generada por el autoencoder, se aplicó una reducción de dimensionalidad mediante t-SNE sobre las reconstrucciones. Los resultados mostraron la emergencia de dos conglomerados diferenciados: uno predominantemente compuesto por transacciones legítimas, y otro donde se agrupan una proporción considerable de fraudes. Esta segmentación es especialmente relevante dado que el modelo no fue entrenado con datos fraudulentos. El hecho de que los fraudes se proyecten en una región distinta del espacio latente refuerza la idea de que el autoencoder ha aprendido una representación estructuralmente sensible a desviaciones anómalas, lo cual implica una forma implícita de detección no supervisada.

No obstante, también se identificaron casos de fraude que permanecen superpuestos con las transacciones legítimas en el espacio reducido, lo que indica que la separación lograda es alta, aunque no perfecta. Este resultado es consistente con los objetivos de diseño del autoencoder y evidencia su eficacia como filtro inicial que contribuye significativamente al rendimiento del modelo global sin necesidad de supervisión directa. En conjunto, el autoencoder no solo cumple su función principal como generador de una métrica de reconstrucción útil, sino que también actúa como transformador de características, promoviendo estructuras latentes más informativas y útiles para la clasificación final. Su integración al flujo de trabajo fortalece la arquitectura general del sistema de detección, aportando robustez, interpretabilidad y capacidad de generalización.

COMPORTAMIENTO DE SELECTKBEST

La selección de variables se realizó mediante SelectKBest con puntuación por información mutua y validación cruzada estratificada (10 pliegues), utilizando un Random Forest como estimador base. El mejor rendimiento AUPRC (0.873) se alcanzó al seleccionar 20 características.

El desempeño fue bajo al usar menos de 10 variables, pero se estabilizó significativamente a partir de dicho punto. Esto indica que el modelo es robusto incluso con muchas variables, aunque el subconjunto óptimo permite reducir ruido, mejorar generalización y facilitar la posterior búsqueda de hiperparámetros. En conjunto, SelectKBest mejoró tanto el rendimiento como la eficiencia del modelo sin comprometer su poder predictivo.

COMPORTAMIENTO DEL CLASIFICADOR RANDOM FOREST

Tras el proceso de optimización de hiperparámetros, el modelo final seleccionado fue un clasificador Random Forest configurado con penalización de clases balanceada, compuesto por 300 árboles, una profundidad máxima de 18, y una división mínima de 8 muestras. Para acelerar el entrenamiento, se habilitó el uso de todos los núcleos disponibles (`n_jobs = -1`) y se fijó una semilla aleatoria (valor 4), garantizando así la reproducibilidad del experimento.

Antes de la predicción final, se procedió a ajustar el umbral de decisión, optimizándolo en función del F1-score. Esta etapa consistió en explorar múltiples valores de umbral y seleccionar aquel que maximizaba el equilibrio entre precisión y sensibilidad. Inicialmente, el modelo mostró métricas casi perfectas sobre el conjunto de entrenamiento, lo cual sugería un potencial sobreajuste. Sin embargo, tras aplicar el umbral óptimo, se alcanzaron resultados más realistas y generalizables: **F1-score (fraude): 0.8913. Precisión (fraude): 0.9567. Sensibilidad (fraude): 0.8343.** Cabe destacar que las métricas AUPRC y AUROCC, al basarse en el comportamiento del modelo a lo largo de distintos umbrales, no se ven afectadas por el ajuste del umbral. Los valores obtenidos fueron: **AUPRC: 0.8726. AUROCC: 0.9631.**

Estos resultados reflejan una excelente capacidad discriminativa del modelo en un escenario altamente desbalanceado. El valor de AUPRC, muy superior al porcentaje de la clase positiva (fraudes), confirma que el modelo mantiene una alta precisión incluso cuando se incrementa la sensibilidad, lo cual es crucial en contextos donde los falsos positivos tienen un costo manejable comparado con los falsos negativos.

Si bien el modelo detecta aproximadamente el 83% de las transacciones fraudulentas, lo que representa una buena sensibilidad, aún deja escapar algunos fraudes. Esta limitación puede ser crítica dependiendo del nivel de tolerancia al riesgo definido por el negocio. En consecuencia, si se prioriza la detección exhaustiva, podrían considerarse estrategias complementarias como el ensamblado de modelos o el ajuste de umbrales más sensibles.

En cuanto a la importancia de las variables, el análisis reveló que la más influyente fue V14, seguida por la variable derivada del error de reconstrucción del autoencoder, y luego V10, V4 y V12. Estas cinco variables constituyen el núcleo predictivo del modelo. Si bien se esperaba que la variable Amount tuviera un mayor impacto, su contribución, aunque menor, también fue significativa, especialmente tras su transformación con QuantileTransformer, que mejoró su correlación con la clase objetivo. En conjunto, el clasificador Random Forest demostró ser robusto, preciso y altamente interpretable, con un excelente equilibrio entre rendimiento y trazabilidad. No obstante, como se abordará en el apartado 5, será fundamental verificar estas métricas sobre el conjunto de prueba para evaluar la capacidad de generalización del modelo y descartar cualquier efecto residual de sobreajuste.

CAIO OLIVEIRA QUINAMO
DATA SCIENTIST
caioquinamocontact@gmail.com



Caio Oliveira Quinamo
Data Scientist
caioquinamocontact@gmail.com





5. EVALUACIÓN

5.1 EVALUAR RESULTADOS

En esta fase final, se evaluó el rendimiento del modelo sobre el conjunto de prueba, nunca utilizado durante el entrenamiento ni la optimización de hiperparámetros. Esto permite estimar con mayor fidelidad su capacidad de generalización y su posible comportamiento en un entorno real.

Los datos fueron transformados siguiendo el flujo de preprocessamiento previamente definido, garantizando compatibilidad estructural con el conjunto de entrenamiento. Posteriormente, se generaron las probabilidades de clase y se aplicó el umbral de decisión optimizado, descartando el valor por defecto (0.5), con el objetivo de lograr un mejor equilibrio entre precisión y sensibilidad.

Los resultados obtenidos fueron consistentes con los de entrenamiento y validación, confirmando la robustez del modelo. La precisión para la clase de fraude alcanzó un 93,89%, reflejando una tasa muy baja de falsos positivos, mientras que la sensibilidad fue del 83,11%, lo que indica que el modelo logra detectar más de 8 de cada 10 fraudes. El F1-score correspondiente fue de 0,8817, confirmando un excelente balance entre ambas métricas. A nivel global, el modelo mostró una capacidad discriminativa muy elevada, con un AUPRC de 0,8941 y un AUROCC de 0,9771, métricas especialmente relevantes en contextos de clases desbalanceadas. La matriz de confusión respalda esta conclusión, mostrando una tasa de error mínima en transacciones legítimas: apenas un error cada aproximadamente 3.185 casos.

Complementariamente, una proyección en 2D mediante t-SNE del espacio latente reveló una segmentación parcial de los fraudes, con muchos de ellos agrupados en una región bien diferenciada del espacio. El modelo fue particularmente efectivo para detectar la mayoría de los fraudes que se separaron estructuralmente de las transacciones legítimas. Sin embargo, mostró limitaciones a la hora de identificar los fraudes que permanecen solapados con los casos verídicos dispersos en el espacio latente, lo que sugiere una oportunidad de mejora en la arquitectura del autoencoder o en sus hiperparámetros. Abordar esta zona gris del espacio latente podría traducirse en una mejor detección de fraudes sutiles o estructuralmente mimetizados.

Conclusión: El modelo demuestra un rendimiento sólido, generaliza bien y mantiene coherencia entre validación y prueba. Aunque existe margen para mejorar la sensibilidad frente a fraudes más difíciles de distinguir, el sistema es viable para una implementación operativa, siempre que se encuentre alineado con el umbral de riesgo definido por el negocio. El flujo de trabajo completo (desde la detección temprana con autoencoder hasta la clasificación con Random Forest calibrado) ha demostrado ser efectivo. A futuro, se recomienda iterar sobre el autoencoder para potenciar su capacidad de separar casos solapados en el espacio latente, mejorando así la sensibilidad sin sacrificar la precisión.

5.2 PROCESO DE REVISIÓN

Durante la fase de evaluación, se realizó una revisión integral para verificar el cumplimiento de los objetivos definidos en las etapas iniciales del proyecto. Esta revisión no se limitó al análisis de métricas cuantitativas, sino que también incorporó una evaluación cualitativa del comportamiento del modelo en distintos escenarios operativos, con especial atención a la robustez, interpretabilidad y capacidad de generalización.

Para mitigar el riesgo de sobreajuste, se utilizó validación cruzada estratificada durante el entrenamiento, manteniendo proporciones consistentes entre clases en todos los pliegos. Además, se trabajó con un conjunto de prueba completamente independiente, que no participó en ninguna fase previa. A este conjunto se le aplicó el mismo flujo de preprocessamiento completo, incluyendo la normalización, la reconstrucción de datos a través del autoencoder, el cálculo del error de reconstrucción y la concatenación de características originales e ingenieras, garantizando así consistencia estructural con el entorno de validación.

Los resultados obtenidos sobre este conjunto fueron consistentes con los valores esperados. Las métricas globales de precisión, sensibilidad, F1-score, AUPRC y AUROCC reflejaron un alto nivel de generalización, sin indicio de deterioro respecto al rendimiento observado en los datos de entrenamiento. De forma complementaria, se aplicó un análisis visual del espacio latente mediante t-SNE, que permitió observar una separación razonable entre clases. Esta visualización confirmó la capacidad del modelo para capturar patrones estructurales útiles, validando cualitativamente los resultados numéricos.

No obstante, se llevó a cabo un análisis de errores para comprender mejor las limitaciones del modelo. Se identificó que los fraudes solapados con casos legítimos en regiones difusas del espacio latente siguen representando un desafío, pues tienden a escapar al clasificador. Este hallazgo sugiere un posible foco de mejora en el autoencoder o en los métodos de enriquecimiento del espacio de representación.





Desde una perspectiva práctica, también se consideraron aspectos clave para la puesta en producción. El modelo, aunque robusto, requiere una monitorización periódica para garantizar su estabilidad ante cambios en los patrones de transacción o derivaciones concept drift. Asimismo, la actualización de los umbrales o el reentrenamiento con nuevos datos podría formar parte de un plan de mantenimiento responsable.

En conclusión, el proceso de revisión evidenció que el modelo no solo cumple con los criterios de éxito definidos, sino que los supera con holgura. Se trata de un sistema confiable, con buen equilibrio entre precisión y sensibilidad, y con fundamentos sólidos tanto en su diseño como en su validación. Aun así, se reconoce que existen márgenes de mejora, en particular en lo que respecta a la detección de fraudes sutiles o mimetizados, por lo que futuras iteraciones podrían centrarse en refinar las capacidades del autoencoder o explorar técnicas de ensamblado para cubrir estos casos residuales.

5.3 MODELOS APROBADOS

Tras concluir el proceso de validación sobre el conjunto de prueba, se aprobó como solución final el modelo compuesto por un clasificador Random Forest entrenado sobre las representaciones reconstruidas por el autoencoder, incluyendo el error de reconstrucción como una variable explicativa adicional. Esta arquitectura fue seleccionada no solo por su excelente desempeño cuantitativo, sino también por su robustez, capacidad de generalización y facilidad de interpretación frente a opciones más complejas.

La elección del modelo se sustentó en su rendimiento sostenido en datos no vistos, demostrando coherencia con los objetivos definidos desde el inicio del proyecto. Durante la fase de evaluación, se aplicó el flujo completo de preprocesamiento, que incluyó la normalización de las variables originales, la transformación de los datos mediante autoencoder, el cálculo del error de reconstrucción, y la selección de características relevantes mediante SelectKBest, antes de la etapa de clasificación. Este encadenamiento metodológico se consolidó como una solución integrada, efectiva de extremo a extremo.

El umbral de decisión fue ajustado de forma precisa para maximizar el F1-score, lo cual permitió optimizar la relación entre precisión y sensibilidad, logrando un equilibrio adecuado entre la identificación de fraudes reales y la minimización de falsos positivos. Esta calibración aportó estabilidad al comportamiento del modelo en contextos operativos donde los errores de clasificación pueden tener costos significativamente distintos.

5.4 DETERMINAR LOS PRÓXIMOS PASOS

El proyecto concluyó exitosamente con el desarrollo de un modelo robusto y generalizable para la detección de fraude, validado exhaustivamente sobre un conjunto de prueba independiente. Aunque no se contempló una fase de despliegue operativo, se identificaron varias líneas de mejora que podrían guiar futuros desarrollos e implementaciones.

Una prioridad futura es revisar la arquitectura del autoencoder. Si bien ha contribuido a una separación efectiva de muchas observaciones fraudulentas en el espacio latente, algunos casos permanecen solapados con la clase legítima. Optimizar esta representación podría mejorar la sensibilidad del sistema ante fraudes más sutiles. De forma complementaria, se propone adaptar el flujo de trabajo a entornos de detección en tiempo real, lo que implica abordar retos como la latencia, la recalibración dinámica del umbral de decisión y la monitorización continua del desempeño en producción.

Durante el desarrollo, se enfrentaron desafíos significativos que aportaron valor formativo. Uno de los más críticos fue la fuga de datos, producto de aplicar transformaciones previas a la partición. Esto se resolvió encapsulando cada paso del preprocesamiento (incluyendo la normalización, el autoencoder y el cálculo del error) dentro del conjunto de entrenamiento y replicándolo posteriormente en validación y prueba. También se abordó el sobreajuste, detectado al observar discrepancias entre entrenamiento y validación, mediante regularización, validación cruzada estratificada y selección de características con SelectKBest.

Asimismo, se corrigió un error conceptual importante al abandonar el uso del umbral fijo de 0.5, inadecuado para conjuntos de datos desbalanceados. En su lugar, se empleó un análisis de la curva precisión-recall para encontrar el punto que maximizara el F1-score, optimizando así la detección sin comprometer la precisión. Este ajuste fue fundamental para mejorar la sensibilidad del sistema.

En términos estructurales, se diseñó un flujo de trabajo robusto y reproducible que garantiza consistencia entre entrenamiento e inferencia. Este diseño mitigó problemas derivados de aplicar predicciones fuera del flujo completo de procesamiento. Finalmente, se superaron desafíos propios de la naturaleza desbalanceada del problema, asegurando una distribución representativa en cada conjunto de datos y empleando métricas como AUPRC, AUROCC y visualizaciones t-SNE para comprender mejor la estructura del espacio latente y detectar oportunidades de mejora.

CAIO OLIVEIRA QUINAMO
DATA SCIENTIST
caioquinamocontact@gmail.com



Caio Oliveira Quinamo





6. REFERENCIAS

CRECIMIENTO DEL COMERCIO ELECTRÓNICO GLOBAL

- [Retail e-commerce sales worldwide from 2014 to 2027](#)
- [51 eCommerce Statistics In 2025 \(Global and U.S. Data\)](#)

USO DE BILLETERAS DIGITALES Y MÉTODOS DE PAGO ONLINE

- [Digital Wallet Statistics](#)
- [Top US Payment Methods \(2023–2027\)](#)

TIPOS DE FRAUDE EN TRANSACCIONES ONLINE

- [What Is Carding? How It Works, Prevention Methods, and Examples](#)
- [6 Types of Credit Card Fraud & How Businesses Can Stop Them](#)
- [The Ultimate Guide to Fraud Detection and Prevention](#)

METODOLOGÍA CRISP-DM

- [The CRISP-DM Process: A Comprehensive Guide](#)

AUTOENCODERS Y DETECCIÓN DE FRAUDE

- [Sehrawat, D., & Singh, Y. \(2023\). Auto-Encoder and LSTM-Based Credit Card Fraud Detection. SN Computer Science, 4\(557\).](#)
- [Anónimo. \(2023\). Credit Card Fraud Detection Using an Autoencoder Model with New Loss Function. OAJI](#)
- [Semi Supervised Classification using AutoEncoders](#)
- [How Autoencoders Work: Intro and UseCases](#)

RANDOM FOREST Y CLASIFICACIÓN SUPERVISADA

- [Anónimo. \(2023\). Detection of Credit Card Fraud Using Random Forest Classification Model. ResearchGate](#)
- [Anónimo. \(2023\). A Random Forest Classifier Approach to Payment Fraud Detection. IJISRT](#)

SELECCIÓN DE CARACTERÍSTICAS CON MUTUAL INFORMATION

- [Anónimo. \(2023\). The Effect of Feature Selection on the Accuracy of X-Platform User Identification. Electronics, 13\(1\), 205](#)
- [Bellani, C., Kraev, E., & Shestopaloff, A. \(2023\). Feature Selection with Neural Estimation of Mutual Information. ICLR 2024 Submission](#)

VISUALIZACIÓN Y SEPARACIÓN DE CLASES CON T-SNE

- [Anónimo. \(2023\). Credit Card Fraud Detection using Logistic Regression Compared with t-SNE to Improve Accuracy. ResearchGate](#)

OTROS TRABAJOS DE REFERENCIA

- [Credit Fraud || Dealing with Imbalanced Datasets.](#)
- [Best techniques and metrics for Imbalanced Dataset](#)
- [SMOTE with Imbalance Data](#)

CAIO OLIVEIRA QUINAMO
DATA SCIENTIST
caioquinamocoutocontact@gmail.com





7. LICENCIA Y NOTA FINAL DEL AUTOR

LICENCIA

Este informe técnico realizado por **Caio Quinamo** está bajo licencia **Creative Commons Reconocimiento 4.0 Internacional (CC BY 4.0)**.

Usted es libre de compartir, copiar, redistribuir y adaptar el material para cualquier propósito, incluso comercial, siempre que se otorgue el crédito apropiado al autor y se indique claramente si se realizaron cambios.

El código fuente que acompaña este informe está distribuido bajo la **Licencia MIT**, lo que permite su uso, reutilización, modificación y redistribución, siempre que se cite debidamente al autor original.

Nota sobre el uso del nombre comercial: Aunque el contenido puede ser reutilizado con fines comerciales bajo la licencia CC BY 4.0, queda estrictamente prohibido el uso del nombre "**Caio Quinamo Analytics Solutions**", el logotipo, o cualquier otro elemento identificativo asociado al autor, salvo autorización expresa y por escrito del titular. La concesión de esta licencia no implica apoyo, patrocinio ni aprobación de usos derivados por parte del autor original.

Para consultar los textos completos de la licencia: [CC BY 4.0](#), [MIT License](#)

NOTA FINAL DEL AUTOR

El autor agradece sinceramente el tiempo y el esfuerzo de todas las personas que han tenido la oportunidad de revisar y comentar este trabajo. Su compromiso es muy valioso, ya que contribuye al crecimiento de las ideas presentadas y fomenta un intercambio significativo de conocimientos.

El autor valora el debate abierto y constructivo y cree en el poder de compartir ideas y perspectivas diversas. Tanto si está de acuerdo como si no lo está con los puntos planteados en este trabajo, sus opiniones y comentarios son siempre bienvenidos, ya que ayudan a perfeccionar la comprensión del tema y fomentan el crecimiento intelectual.

No dude en ponerte en contacto con nosotros por correo electrónico para cualquier pregunta, sugerencia de mejora o para participar en debates sobre el trabajo. Siempre se aprecia y se agradece que comparta sus ideas y opiniones, ya se trate de críticas, puntos de vista alternativos o sugerencias para seguir investigando.

Su genuina participación en estos debates contribuye a enriquecer el acervo de conocimientos y favorece el avance colectivo en la materia. Gracias por sus valiosas contribuciones.

— CAIO O. QUINAMO

caioquinamocontact@gmail.com



RATIO ET INTELLIGENTIAS DUCUNT AD
PROSPERITATEM



"EL OBJETIVO ES CONVERTIR LOS DATOS EN
INFORMACIÓN Y LA INFORMACIÓN EN CONOCIMIENTO"

— CARLY FIORINA

"DECIR LA VERDAD CON DATOS"

— HANS ROSLING



© 2025 CAIO O. QUINAMO. ESTE INFORME TÉCNICO ESTÁ DISPONIBLE BAJO LA LICENCIA CREATIVE COMMONS ATRIBUCIÓN 4.0 INTERNACIONAL (CC BY 4.0)



[LinkedIn](#) [GitHub](#) [kaggle](#) [M](#) [Gmail](#)

CONTACTO Y REDES SOCIALES

