**Propuesta de Proyecto**

**Seminario de Analítica y Ciencia de Datos**

**Presentación del Proyecto**

En un mundo cada vez más conectado y, al mismo tiempo, más propenso al desarrollo de actividades ilícitas, el lavado de activos representa una amenaza creciente para la integridad del sistema financiero colombiano, considerando el gran volumen de transacciones ejecutadas diariamente tanto en el mercado interno como global. Hoy en día, el lavado de dinero afecta a todas las economías del mundo y es responsable de generar flujos financieros ilegales entre 1,6 y 2,85 billones de dólares al año, equivalente al 2,1% - 4% del Producto Bruto Mundial.

Nuestra propuesta de proyecto se encamina a abordar el desafío del lavado de activos mediante el uso de técnicas de Machine Learning y a reducir los procesos manuales que actualmente se utilizan en dichas instituciones. Con esta iniciativa, buscamos refinar el proceso de detección de lavado de activos (Anti-Money Laundering) y contribuir de manera técnica al desarrollo de modelos y al mejoramiento de los sistemas de prevención de lavado de activos y financiación del terrorismo (SARLAFT) en Colombia.

**Objetivo del Proyecto**

El objetivo del proyecto es aprender y aplicar técnicas de Machine Learning a datos financieros con el fin de desarrollar un aplicativo capaz de predecir posibles transacciones que puedan estar relacionadas con el lavado de activos. Para lograr esto, llevamos a cabo una evaluación experimental en un conjunto de datos sintéticos. Este conjunto de datos incluye tanto transacciones normales como patrones anómalos que pueden estar vinculados a posibles actividades de lavado de dinero. Evaluaremos la eficacia de los modelos desarrollados mediante el análisis de estos conjuntos de datos históricos.

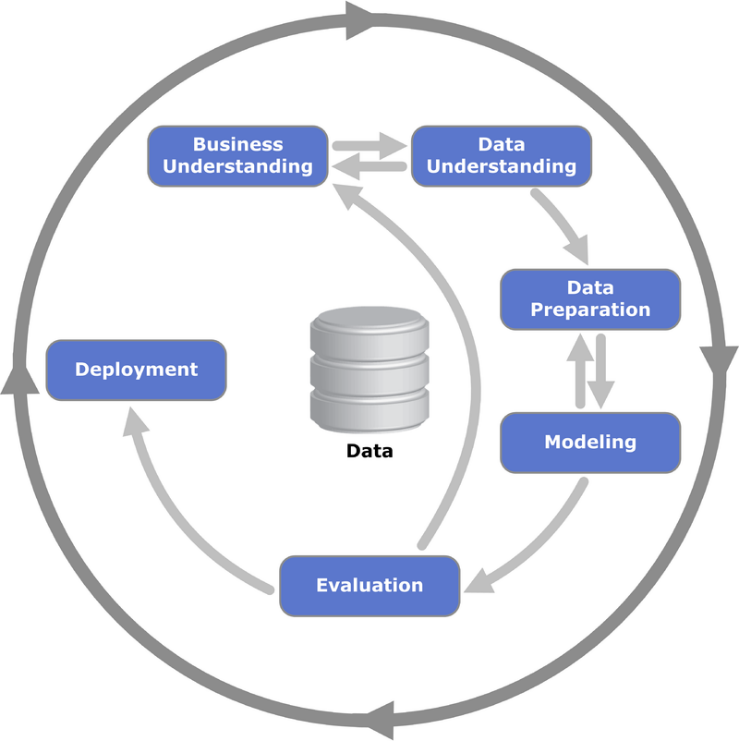
**Contexto del Problema**

El lavado de activos es una actividad que ha afectado la economía global durante muchos años. Grandes cantidades de dinero obtenido de actividades ilegales se utilizan para hacer que los fondos adquiridos ilegalmente parezcan legales o legítimos. Colombia no es ajena a este fenómeno; el auge del negocio del lavado de activos en Colombia ha sido significativo y ha marcado la historia económica del país. Diferentes grupos criminales, como las guerrillas, grupos paramilitares y narcotraficantes, han desarrollado alcances internacionales en esta actividad, generando más recursos que en el pasado. Este aumento de dineros ilícitos ha creado la necesidad de contar con sistemas, controles y prácticas eficaces contra el lavado de dinero (AML) para gestionar el riesgo de las actividades de lavado de dinero, principalmente en el sector financiero colombiano.

Monitorear las transacciones financieras es una obligación crítica contra el lavado de dinero (AML) para las instituciones financieras. Dichas entidades están obligadas a supervisar y reportar operaciones sospechosas (ROS) a la Unidad de Información y Análisis Financiero (UIAF). Para ello, a fin de evitar el riesgo legal y reputacional, el sector financiero enfoca sus esfuerzos contra el lavado de activos generalmente hacia la aplicación de sistemas de leyes como los sistemas SARLAFT. Sin embargo, se puede suponer que las vulnerabilidades surgen de la simplicidad de estos sistemas y de la vigilancia manual que se realiza sobre las transacciones, lo que da lugar a altas tasas de falsos positivos y bajas tasas de detección.

**Propuesta de Solución y Alcance**

Para realizar esta investigación, nos basaremos en el diseño de investigación mixta, es decir, se utilizará la inteligencia artificial (IA), específicamente modelos de Machine Learning, en conjunto con la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining), que es un enfoque estándar para proyectos de minería de datos que nos guía a través de las etapas de un proyecto de análisis de datos, como la comprensión del negocio, la comprensión de los datos, la preparación de los datos, el modelado, la evaluación y la implementación. La combinación de estos enfoques puede ser muy efectiva para abordar problemas complejos y obtener conclusiones significativas.



Para abordar el desarrollo del proyecto, proponemos utilizar técnicas de Machine Learning combinadas con herramientas como Python y la biblioteca scikit-learn. Esta combinación nos permitirá desarrollar modelos predictivos capaces de identificar patrones y comportamientos sospechosos en los datos financieros.

En primer lugar, utilizaremos técnicas de procesamiento de datos para limpiar y preparar los conjuntos de datos históricos proporcionados. Luego, aplicaremos algoritmos de Machine Learning supervisado, como Support Vector Machines (SVM) y Random Forest, para entrenar modelos que puedan clasificar transacciones como sospechosas o no sospechosas de lavado de activos. Además, exploraremos técnicas de aprendizaje no supervisado, como clustering, para identificar grupos de transacciones que puedan indicar actividades fraudulentas. Utilizaremos la metodología de evaluación de modelos para medir la eficiencia y precisión de nuestros modelos, ajustando los parámetros según sea necesario para mejorar su desempeño.

Posteriormente, se procederá a la evaluación de la precisión y eficacia de los modelos utilizando métricas relevantes, como precisión, recall y F1-score. Esto nos permitirá determinar la viabilidad y eficiencia de los modelos en la detección de lavado de activos.

**Definición del alcance**

Se pretende lograr un modelo que arroje resultados significativos en cuanto a eficiencia y precisión, siendo capaz de identificar patrones de comportamiento en la información transaccional y permita generar alertas tempranas que indiquen posibles riesgos de lavado de activos dentro de la organización, de acuerdo con las normativas de la Superintendencia Financiera de Colombia. Se buscará desarrollar y validar este modelo utilizando datos simulados que reflejen de manera realista la actividad financiera, asegurando así la robustez y confiabilidad del sistema. Se busca un modelo 'genérico' que pueda ser aplicado en la mayoría de instituciones financieras colombianas para la detección de transacciones fraudulentas y que a su vez sea capaz de destacar y mejorar en problemáticas como el alto número de falsos negativos.

Los conjuntos de datos financieros son cruciales para investigadores que buscan enfoques innovadores, particularmente para nuestro proyecto de investigación en detección de transacciones sospechosas, sin embargo, debido a la naturaleza intrínsecamente privada de las transacciones financieras, se produce una escasez de conjuntos de datos estandarizados y disponibles especialmente en el ámbito de las transacciones financieras, lo que convierte la detección de lavado de activos en una de las tareas más desafiantes en el campo de la detección de anomalías. Dicho esto, se buscará desarrollar y validar este modelo utilizando datos simulados que reflejen de manera realista la actividad financiera, asegurando así la robustez y confiabilidad del sistema.

Finalmente, se pretende señalar qué técnicas faltaron o no fueron abordadas lo suficiente en la investigación existente, con el propósito de indicar direcciones para futuras investigaciones.

**Riesgos e Impacto del Negocio**

La implementación de nuestra propuesta de detección de lavado de activos en transacciones financieras puede tener un impacto positivo significativo en varios aspectos clave del negocio. En primer lugar, contribuye a la reducción de los riesgos asociados al lavado de activos al identificar patrones de comportamiento sospechosos, fortaleciendo así la integridad y reputación de la institución financiera ante clientes y reguladores. Además, al permitir una detección temprana de posibles transacciones fraudulentas, mejora la eficiencia operativa al evitar pérdidas financieras importantes y minimizar interrupciones en la actividad financiera. Por último, al cumplir con las normativas establecidas por la Superintendencia Financiera de Colombia y específicamente los lineamientos de los sistemas de prevención de lavado de activos y financiación del terrorismo (SARLAFT) en Colombia, demuestra el compromiso de la institución con el cumplimiento regulatorio, generando así confianza y credibilidad en el mercado financiero. En conjunto, estos aspectos respaldan el crecimiento y la sostenibilidad del negocio, al tiempo que garantizan la protección de los intereses de los clientes y las partes interesadas.

Sin embargo, la implementación de esta solución también conlleva ciertos riesgos potenciales para el negocio. En primer lugar, aunque los datos simulados pueden ser útiles para desarrollar y validar el modelo de detección de fraudes, es importante asegurarse de que reflejen de manera precisa y realista la actividad financiera. Si los datos simulados no son representativos o no capturan adecuadamente la complejidad y variabilidad de las transacciones reales, existe el riesgo de que el modelo de detección de fraudes no funcione eficientemente en un entorno de producción real.

Por otra parte, no existe una regulación mundial común que establezca estándares sobre qué transacciones son sospechosas con respecto a actividades de lavado de dinero. Además, los procesos de blanqueo de dinero implican múltiples transacciones entre diferentes contrapartes utilizando diversos instrumentos monetarios. Por lo tanto, las técnicas de detección de anomalías que analizan transacciones pueden resultar ineficaces según la normativa específica de cada país o región.

Finalmente, existe la posibilidad de que el modelo de detección de fraudes genere falsos positivos o falsos negativos, lo que podría afectar negativamente la experiencia y satisfacción del cliente, en este caso, la institución financiera. Esto, a largo plazo, podría tener un impacto en la reputación de la institución financiera. Además, los costos asociados con la implementación y mantenimiento de la solución podrían ser significativos y representar una carga financiera para la organización, especialmente si los beneficios no superan estos costos.

**Ejemplo de éxito:**

Como ejemplo de éxito nos podemos referir a la implementación de sistemas de detección de fraudes en transacciones financieras que implemento la compañía PayPal. PayPal ha desarrollado sofisticados sistemas de detección de fraudes que utilizan algoritmos avanzados y análisis de datos (en ocasiones acompañados con un conjunto de filtros personalizados predefinidos que están diseñados para cada empresa teniendo en cuenta diversos atributos, como la categoría de la empresa, el volumen medio de pagos y las devoluciones de cargo anteriores) en tiempo real para identificar y prevenir actividades fraudulentas. Gracias a estas soluciones, PayPal ha logrado reducir significativamente las pérdidas por fraude y mantener la confianza de sus usuarios y socios comerciales. Este caso demuestra el impacto positivo que una sólida estrategia de detección de fraudes puede tener en un negocio, así como la importancia de la inversión continua en tecnología y recursos para mantener la eficacia de dichos sistemas.