

Uso de Datos Sintéticos para garantizar la protección de datos personales por diseño en el entrenamiento de modelos de detección de fraude de identidad

Leandro Guzman, *MDS2-2022, UAI*, Carlos Muñoz, *MDS2-2022, UAI*,
 Francisco Guzman, *MDS2-2022, UAI*, Cristopher Lincoleo, *MDS2-2022, UAI*
 and Andrés Pumarino, *Profesor Guía*

Resumen—La creciente adopción del onboarding digital por parte del sector financiero minorista ha agilizado la incorporación de nuevos clientes, ofreciendo una experiencia fácil y rápida para acceder a sus servicios. No obstante, el aumento de casos de fraude por suplantación de identidad ha presionado a las empresas a reforzar sus mecanismos de verificación y detección. Además, entrenar modelos de detección cumpliendo con las nuevas normativas de protección de datos personales presenta desafíos: por un lado, entrenar modelos altamente complejos requiere una gran cantidad de imágenes, y por otro lado, la protección de datos personales, particularmente datos biométricos, requiere la autorización de los titulares para el uso específico de este tipo de datos. Este trabajo presenta una metodología basada en cuatro elementos clave: Una metodología similar a CRISP-DM que integra la Evaluación de Impacto en la Protección de Datos (DPIA) para garantizar el cumplimiento de las normativas de protección de datos. El uso de datos sintéticos para proporcionar suficiente información para entrenar modelos equilibrados. El uso de modelos simples que se centran en las fortalezas de los algoritmos y las debilidades de cada tipo de fraude en lugar de una única solución altamente compleja. Una tubería modular secuencial optimizada para los costos de errores Tipo I y Tipo II, que permite la integración de múltiples modelos especializados. La solución propuesta se aplicó a través de un caso de negocio en una empresa del sector minorista financiero chileno enfocada en el nivel socioeconómico medio-bajo, la cual tiene implementado el onboarding digital, lo que permitió probar la metodología y evaluar su impacto financiero.

Index Terms—Biometric Validation, Facial Recognition, Fraud Detection, Deep Learning, Data Privacy, Synthetic Data, Generative Adversarial Networks (GANs), CRISP-DM, Regulatory Compliance

I. INTRODUCCIÓN

La transformación digital en el sector financiero minorista ha impulsado significativamente la eficiencia operativa y la accesibilidad de sus clientes con la incorporación del proceso de onboarding digital, pero también ha generado nuevos desafíos en términos de seguridad. A medida que las empresas han sumado este proceso de incorporación digital de nuevos clientes, han debido reforzar sus mecanismos de verificación y detección para prevenir el fraude por suplantación de identidad.

Esto ha llevado al desarrollo de sistemas de validación biométrica basados en la verificación del rostro, los cuales buscan equilibrar la capacidad de detección con la experiencia

del usuario. Por un lado, al fallar en la detección de un fraude, se produce una pérdida económica limitada por el nivel de crédito asignado. Por otro lado, al identificar como potencial fraude a un cliente legítimo, se afecta significativamente la experiencia del usuario, arriesgando la pérdida del cliente con el correspondiente costo de oportunidad.

Históricamente, los sistemas de validación biométrica han dependido del entrenamiento de modelos de Machine Learning basados en Deep Learning, los cuales tienen varios millones de parámetros y, por lo tanto, han requerido millones de imágenes para su entrenamiento. Esta necesidad ha llevado a que muchos de los modelos existentes hayan sido entrenados usando imágenes capturadas desde la web sin el *consentimiento explícito y expreso* de los usuarios[1], [2], [3]. Aunque se han basado en la captura de imágenes públicas, las exigencias actuales de protección de datos personales requieren que el uso de datos personales sensibles (como el rostro) cuente con autorizaciones explícitas y no genéricas[4].

En este contexto, el uso de datos biométricos para el entrenamiento de modelos de detección de fraude presenta un dilema crucial: la necesidad de grandes volúmenes de datos de alta calidad para lograr precisión y la obligación de proteger estos datos sensibles para cumplir con las normativas de privacidad.

Los nuevos marcos regulatorios[4] complican el desarrollo de modelos eficaces de detección de fraude, ya que el acceso a datos reales está restringido y la privacidad debe ser garantizada en todo momento. Las técnicas tradicionales de anonimización no son suficientes, ya que pueden comprometer la calidad y la utilidad de los datos.

Asimismo, la complejidad técnica de desarrollar soluciones eficaces para la detección de fraude de identidad es considerable. Los modelos avanzados, como las redes neuronales convolucionales y las técnicas de normalización de pose y expresión en 3D, requieren un procesamiento intensivo y grandes volúmenes de datos etiquetados. Aunque estos modelos complejos pueden ofrecer alta precisión, también presentan desventajas significativas, como el alto costo computacional, la dificultad de implementación y la falta de interpretabilidad.

Para abordar estos desafíos, este estudio propone una metodología innovadora que emplea datos generados artificialmente mediante técnicas avanzadas como las Redes Generativas

Adversariales (GANs) [5]. Estos *datos sintéticos* imitan las propiedades estadísticas de los datos reales sin comprometer la privacidad ni revelar información sensible, ofreciendo así una solución que equilibra la necesidad de detalles con la protección de la privacidad.

Esta metodología se basa en cuatro pilares fundamentales: la integración de la Evaluación de Impacto en la Protección de Datos (DPIA) dentro del marco CRISP-DM para garantizar el cumplimiento normativo, el uso de datos sintéticos para entrenar modelos equilibrados, la implementación de modelos simples y especializados que aprovechan las fortalezas de los algoritmos y las debilidades de los intentos de fraude, y un pipeline modular secuencial optimizado en función de minimizar los costos de errores Tipo I y Tipo II, equilibrando la capacidad de detección con la experiencia de usuario, al mismo tiempo que proporciona una solución adaptable y escalable.

Para validar esta metodología, se desarrolló un caso de negocio en una empresa del sector financiero minorista enfocada en el sector socioeconómico medio-bajo. Este caso práctico permitió aplicar la metodología a intentos de fraudes reales. Los resultados demostraron que es posible desarrollar sistemas de detección de fraude que sean precisos y respetuosos con la privacidad mediante el uso adecuado de datos sintéticos y modelos de machine learning optimizados.

La metodología propuesta busca servir como un marco de referencia para futuras investigaciones y aplicaciones en diversos sectores que enfrentan problemas similares de protección de datos y necesidades de modelos avanzados.

II. DESARROLLO DE LA METODOLOGÍA

II-A. Relevancia del problema: Onboarding Digital, innovación empresarial y desafío regulatorio

Es importante entender el proceso de onboarding digital no solo como proceso remoto para la captura de nuevos clientes, sino como un proceso que debe asegurar una relación de confianza entre el cliente y la empresa, garantizando la legitimidad de los usuarios a incorporar y cumpliendo con normativas como KYC (Know Your Client), AML (Anti Money Laundry), CDD (Customer Due Diligence) y regulaciones de protección de datos como GDPR (General Data Protection Regulation). En un mundo cada vez más digitalizado, ofrecer una experiencia rápida y fácil puede hacer la diferencia al momento de atraer a nuevos usuarios, sin embargo, esto conlleva desafíos que las empresas han debido enfrentar al utilizar este nuevo proceso. De acuerdo al reporte Consumer Sentinel Network Data Book 2023 de la Comisión Federal de Comercio de Estados Unidos (FTC, sigla en inglés) [6], durante 2023 se recibieron 5,5 millones de reportes de fraude, encabezando la lista el robo de identidad con un 19,2% de los registros. El aumento de incidentes de fraude por suplantación de identidad ha llevado a las empresas a invertir en soluciones de verificación de identidad que aseguren su proceso de onboarding sin tener que afectar la experiencia de usuario. Según la consultora Mordor Intelligence, en su reporte Identity Verification Market - Growth, Trends, and Forecasts (2024 - 2029) [7] indica que se estima que las

empresas invertirán durante 2024 en soluciones de verificación de identidad un valor cercano a MUSD 13 000, mientras que se espera que en 2029 esta cifra aumente a un valor cercano a MUSD 23 000. Por otro lado, las soluciones de verificación son variadas, desde la validación del documento de identidad oficial hasta la verificación biométrica. Ésta última es la que más se está utilizando; por ejemplo, según estadísticas expuestas en Comisión de Hacienda del Senado de Chile por Comisión para el Mercado Financiero (CMF) Chile en marzo de 2024 [8], de las soluciones de verificación de identidad utilizadas por las empresas chilenas, un 62% de las cuentas digitales aperturadas fue previo a un proceso de validación biométrica. Sin embargo, trabajar con este tipo de datos personales sensibles es complejo, ya que se debe cumplir con las normativas relacionadas a protección de datos, ya sea obteniendo el consentimiento explícito del titular de los datos para la recopilación y procesamiento de los datos, como en el cumplimiento de los principios de legalidad, imparcialidad y transparencia en el tratamiento de éstos.

II-B. Contexto Normativo

Tanto por razones éticas como normativas es necesario proteger la identidad de las personas, en efecto, en la última década la protección de datos personales ha tomado relevancia convirtiéndose en un derecho fundamental (por ejemplo, en la Carta de los Derechos Fundamentales de la Unión Europea, también en la Constitución chilena). Luego de la entrada en vigencia del Reglamento General de Protección de Datos de la Unión Europea (GDPR, sigla en inglés), los países han ido adecuando sus leyes sobre protección de datos personales considerando como marco de referencia esta regulación. Por lo tanto, es importante recordar la definición de dato personal según el GDPR [4]: "toda información de una persona física identificada o identifiable. Se considerará persona física identifiable toda persona cuya identidad pueda determinarse, directa o indirectamente, en particular mediante un identificador, como por ejemplo un nombre, un número de identificación, datos de localización, un identificador en línea o uno o varios elementos propios de la identidad física, fisiológica, genética, psíquica, económica, cultural o social de dicha persona".

De igual manera, el reglamento en su Considerando 78 indica que el responsable del "tratamiento" (por ejemplo, el uso de datos para modelos de Machine Learning) debe adoptar las medidas técnicas y organizativas apropiadas para garantizar la "protección de los derechos y libertades de las personas físicas con respecto al tratamiento de datos personales", debiendo aplicar los principios de la protección de datos desde el diseño y por defecto. Estos principios son adoptados por la definición realizada por Cavoukian[9], quien describe siete principios fundamentales, los que se comentan brevemente, a continuación:

1. Proactividad, no reactividad; prevención, no corrección. Anticipar y prevenir eventos invasivos de privacidad antes de que sucedan.
2. Privacidad como configuración predeterminada. Asegurar que los datos personales son automáticamente protegidos

en cualquier sistema de TI o negocio. La privacidad de una persona permanece intacta, sin que esta deba realizar acción alguna.

3. Privacidad incorporada en el diseño: La privacidad debe estar incorporada en el diseño de la arquitectura de los sistemas de TI y prácticas de negocio, como parte integral del sistema, sin que este pierda funcionalidad.
4. Funcionalidad total: Suma positiva, todos ganan (win-win). Evitar falsas dicotomías, como privacidad versus seguridad o privacidad versus funcionalidad, es posible y deseable alcanzar ambos.
5. Seguridad de extremo a extremo: Protección durante todo el ciclo de vida de los datos. Asegura que los datos se retengan de forma segura y al final del proceso sean destruidos también de forma segura y oportuna.
6. Visibilidad y transparencia: Mantenimiento de un entorno abierto. Asegura que las políticas, prácticas y procedimientos vinculados con datos de las personas sean claros, accesibles y comprensibles por ellas.
7. Respeto por la privacidad del usuario: Centrado en el usuario. Insta a que los arquitectos y operadores mantengan en cuenta los intereses de los usuarios, y estos puedan tener un papel activo en la gestión de sus propios datos.

Resumiendo a lo anterior, se pueden extraer tres ideas clave:

- Restringir al acceso de los datos, con el fin de garantizar el correcto tratamiento de éstos y proteger los derechos de los interesados.
- Utilizar los datos según el uso específico declarado mediante un consentimiento explícito del titular de los datos.
- Definir el período de conservación de los datos asegurando su eliminación posterior o anonimización.

En el caso de Chile, la Ley N°19.628 sobre protección de la vida privada está próxima a ser actualizada mediante reforma de ley para regular el tratamiento de los datos personales, incorporando varios elementos provenientes de la regulación europea, robusteciendo la protección de datos personales. Por otra parte, la legislación ya está tomando medidas respecto a la verificación de identidad, actividad crítica del proceso de onboarding. Por ejemplo, la Resolución exenta N°566 de abril de 2024 establece los requisitos mínimos de verificación de identidad y estándares de seguridad aplicables por proveedores de servicios de telecomunicaciones [10], la que entrará en vigencia en octubre del mismo año.

II-C. Metodologías para la gestión de proyectos de ciencia de datos

El marco metodológico actual, con los cuales se dirigen los proyectos de ciencia de datos, se enfoca principalmente en extraer patrones no obvios y útiles desde grandes conjuntos de datos, los que aportan información valiosa para la toma de decisiones, utilizando algoritmos y procesos de estadística y machine learning. Para resumir, se describen tres de las metodologías más populares: Knowledge Discovery in Databases (KDD), Cross-Industry Standard Process for

Data Mining (CRISP-DM) y Sample, Explore, Modify, Model, Assess (SEMMA).[11]

El proceso KDD: es un método utilizado para extraer patrones o información desde las bases de datos, se puede resumir en cinco pasos:

1. **Selección de datos**, ya sea un subset de determinadas variables o una muestra que resulte relevante.
2. **Procesamiento**, limpieza y preparación de los datos.
3. **Transformación** de los datos, ingeniería de características que faciliten la implementación de algoritmos de minería de datos.
4. **Minería de datos**, aplicación de técnicas para extraer patrones.
5. **Interpretación**, evaluación de los resultados.

El proceso SEMMA fue desarrollado por el Instituto SAS, consta de cinco etapas:

1. **Sample**: selecciona una muestra representativa de los datos, suficientemente grande para extraer información y suficientemente pequeña para ser manipulada rápidamente.
2. **Explore**: explorar los datos en busca de patrones y relaciones para un mejor entendimiento.
3. **Modify**: Modificar datos, incluye limpieza y transformación.
4. **Model**: aplicación de modelos estadísticos y de machine learning para la tarea propuesta.
5. **Assess**: evaluar la validez y utilidad de los modelos generados.

El proceso CRISP-DM, desarrollado en 1999 por IBM [12], es un proceso que ha sido ampliamente utilizado en la industria y no ha tenido modificaciones importantes desde que se creó. El ciclo de vida del CRISP-DM consta de seis etapas [13]:

1. **Comprensión del negocio**, se trata de entender los objetivos del proyecto en el marco del negocio en que se desarrolla, sus necesidades comerciales.
2. **Comprensión de los datos**, revisión de los datos que se tienen a disposición.
3. **Preparación de los datos**, limpieza, selección y transformación de los datos.
4. **Modelado**, aplicación de técnicas de modelado de datos.
5. **Evaluación** de los modelos desarrollados según las necesidades del negocio.
6. **Despliegue**, implementación del modelo en un entorno empresarial y monitoreo periódico de su rendimiento.

II-D. Selección de un marco metodológico adaptado

Las metodologías convencionales no contemplan los riesgos normativos a los que los datos pueden estar sometidos, sino que más bien se concentran en establecer distintas etapas para extraer información relevante de éstos en beneficio de

la toma de decisiones. Este aspecto puede ser problemático para las organizaciones, tanto públicas como privadas, las que incluso podrían incurrir en infracciones legales si no incorporan dentro de sus procesos un análisis de riesgos que involucre la protección de datos personales.

En efecto, el GDPR (Art. 25) establece la obligación de realizar una Evaluación de Impactos Relativa a la Protección de Datos (Data Protection Impact Assessment, DPIA), cuando un tipo de tratamiento pueda implicar un “alto riesgo” para los derechos y libertades de las personas físicas. Aplicar el DPIA permite evaluar los riesgos en etapas tempranas del proceso, lo cual se alinea con garantizar que los proyectos cumplan con la protección de datos desde el diseño. Los pasos del DPIA no están claramente definidos en el GDPR, sin embargo, es común esquematizarlo de la siguiente manera[14]:

- Identificar si es necesario realizar una DPIA.
- Definir las características del proyecto para permitir una evaluación de los riesgos.
- Identificar los riesgos relacionados con la protección de datos.
- Identificar soluciones de protección de datos para reducir o eliminar los riesgos.
- Aprobar los resultados de la DPIA.
- Integrar las soluciones de protección de datos en el proyecto.

En función de lo anteriormente expuesto, nuestra propuesta fue integrar el DPIA al proceso CRISP-DM, el cual cuenta con la flexibilidad suficiente para ampliar sus dos primeras etapas a una perspectiva de la protección de datos y la privacidad por diseño. Esto es absolutamente necesario en la actividad de validación de identidad en un proceso de onboarding digital, en el cual, a menudo se solicita entregar información personal para validar la identidad, como datos biométricos (reconocimiento facial) y una imagen de un documento de identificación oficial.

El análisis demuestra que uno de los principales riesgos detectados en el proceso de validación de identidad por reconocimiento facial es que implica la captura y almacenamiento de datos biométricos de las personas, lo que requiere de un consentimiento explícito por parte del usuario, a quien también se debe informar claramente con qué fines se utilizará su información personal y por cuanto tiempo. Además, existe el riesgo de violación de los datos y el consiguiente mal uso de éstos. Por otra parte, la detección de fraudes por suplantación de identidad requiere el entrenamiento de modelos complejos, necesitando éstos una gran cantidad de imágenes de rostros, las cuales serán difíciles de conseguir de acuerdo con las tendencias normativas actuales.

CRISP-DM con integración de DPIA, constará de ocho pasos, sin embargo, dependiendo del análisis de riesgos de privacidad y protección de datos, y las soluciones a integrar, el proceso podría ser ampliado. En el caso de la detección de fraude en el onboarding digital, una solución que minimiza los riesgos asociados a información personal y garantiza la protección de datos desde el diseño, es el uso de **datos sintéticos**.

Etapas del CRISP-DM con integración de DPIA 1:

1. Comprendión del negocio
2. Comprendión del marco normativo
3. DPIA
4. Comprendión de los datos
5. Preparación de los datos
6. Modelado
7. Evaluación
8. Despliegue

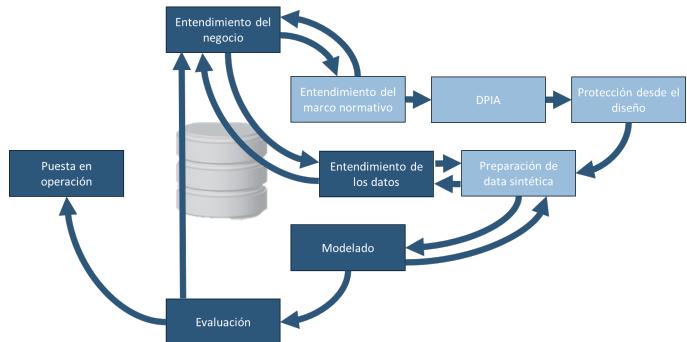


Figura 1: Marco metodológico adaptado para integrar protección de datos personales.

Fuente: Elaboración propia

II-E. Uso de data sintética para protección de identidad

El uso de datos creados para la imputación ha sido una herramienta estándar en la ciencia de datos; sin embargo, la generación de datos sintéticos en el contexto de datos asociados a personas es algo más elaborado. Esta necesidad surge de la demanda de acceso a datos detallados que puedan ser utilizados para la investigación y la toma de decisiones, al mismo tiempo que se protege la confidencialidad de los individuos. Los primeros desafíos en la creación de datos sintéticos comenzaron con la distribución de datos censales, donde era esencial equilibrar la accesibilidad con la protección de la privacidad. Si bien históricamente han existido técnicas de anonimización, estas pueden comprometer la calidad y la utilidad de los datos. Esto llevó a la idea de crear conjuntos de datos sintéticos como una alternativa [15]. Estos conjuntos consisten en datos artificiales que retienen las propiedades estadísticas de los datos reales sin revelar información sensible. Este enfoque permite que los datos sean utilizados para análisis complejos y detallados, asegurando al mismo tiempo la privacidad de los individuos involucrados en los conjuntos de datos originales.

Sin embargo, la creación de datos sintéticos no es trivial y debe cumplir con una serie de características:

- Propiedades Estadísticas: Los datos sintéticos deben replicar las propiedades estadísticas de los datos originales para garantizar que los análisis realizados sobre ellos sean válidos. Esto incluye mantener las distribuciones, correlaciones y otras relaciones estadísticas presentes en el conjunto de datos original [16].
- Privacidad y Confidencialidad: Una de las ventajas más significativas de los datos sintéticos es su capacidad

para proteger la privacidad. Al no contener datos reales, la posibilidad de identificar a individuos específicos se reduce drásticamente, cumpliendo con regulaciones de privacidad y protegiendo a los sujetos de datos [16].

- Utilidad Analítica: Para que los datos sintéticos sean útiles, deben permitir análisis detallados y complejos que sean comparables a los realizados con datos reales. Esto significa que deben ser lo suficientemente realistas para ser utilizados en modelos predictivos y otros tipos de análisis [15].
- Flexibilidad y Escalabilidad: Los datos sintéticos pueden generarse en grandes volúmenes y adaptarse a diferentes necesidades de investigación. Esto permite a los investigadores trabajar con conjuntos de datos grandes y diversos sin las limitaciones impuestas por la disponibilidad de datos reales [16].

Si bien la técnica presenta múltiples ventajas, se debe considerar que su implementación no está exenta de desafíos:

- Complejidad en la Generación: Crear modelos precisos para la generación de datos sintéticos puede ser complicado y requiere un entendimiento profundo de las relaciones y estructuras en los datos originales.
- Potencial de Sesgo: Si los modelos generadores no están bien especificados, pueden introducir sesgos o errores en los datos sintéticos, afectando la validez de los análisis [16].
- Aceptación: Existe una desconfianza inicial hacia los datos sintéticos, ya que los usuarios pueden dudar de su precisión y relevancia comparados con los datos reales [16].

En el contexto del uso de datos sintéticos para la validación biométrica, la complejidad es aún mayor. La generación de imágenes sintéticas es significativamente más compleja debido al alto nivel de detalle y a la alta dimensionalidad de los datos visuales. Las imágenes sintéticas no solo deben reproducir características visuales finas genéricas como texturas, colores y formas, sino que también deben generar diversidad en aspectos como género, color de piel y cabello, uso de accesorios y muchos otros elementos asociados a las características de una persona que definen una identidad biométrica. Al mismo tiempo, deben garantizar que no permitan la identificación de personas reales.

Por otro lado, la necesidad de contar con datos sintéticos en este ámbito se hace evidente: los modelos de reconocimiento facial históricamente han dependido de grandes conjuntos de datos de imágenes faciales recolectadas de la web, a menudo sin el consentimiento explícito de los individuos. Esto plantea serios problemas éticos y legales, además de introducir sesgos significativos debido a la falta de diversidad en los datos recolectados [3]. Adicionalmente, las imágenes recolectadas limitan el entrenamiento de modelos al no considerar características potenciales o emergentes.

Para mitigar estos problemas, en los últimos años se ha desarrollado la generación de datos sintéticos mediante técnicas avanzadas de gráficos por computadora y modelos generativos. Los modelos generativos de imágenes, como las Redes Generativas Adversariales (GANs) [5], funcionan utilizando dos

redes neuronales que compiten entre sí: una red generadora que crea imágenes sintéticas y una red discriminadora que evalúa la autenticidad de estas imágenes. A través de este proceso competitivo, la red generadora mejora continuamente sus capacidades para producir imágenes cada vez más realistas.

Esta metodología ha sido usada en proyectos como "DigiFace-1M: 1 Million Digital Face Images for Face Recognition" [3], que han buscado generar conjuntos de datos sintéticos a gran escala para el reconocimiento facial, generados mediante un pipeline de gráficos por computadora. Este enfoque permite un control total sobre las características de las imágenes, como la pose, la expresión, los accesorios y las texturas, reduciendo así el sesgo y mejorando la precisión de los modelos de reconocimiento facial. Si bien este sistema permite un amplio control, las imágenes generadas no pueden ser clasificadas como fotorealistas (figura 2).

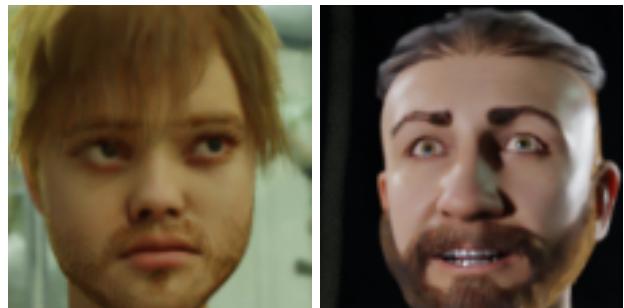


Figura 2: Ejemplos de imágenes sintéticas del proyecto DigiFace-1M.

Fuente: DigiFace1M GitHub repository.

Si bien existen diversos sistemas (figura 3), StyleGAN, desarrollado por NVIDIA, se ha establecido como una herramienta líder en la generación de datos sintéticos, especialmente por su calidad de imagen y realismo (figura 4).

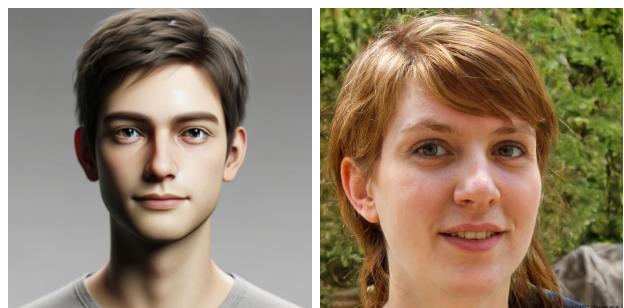


Figura 3: Ejemplos de imágenes sintéticas generadas con DALL-E (izquierda) vs. StyleGAN (derecha)

Fuente (izquierda): Elaboración propia usando DALL-E y el prompt "persona aleatoria, mirando al frente, con un fondo gris, debe ser lo más fotorrealista posible"

Fuente (derecha): Generada aleatoriamente por la página <https://thispersondoesnotexist.com/>

Sin embargo, el uso de StyleGAN no es sencillo, especialmente cuando se utilizan modelos preentrenados no etiquetados, como stylegan2-ffhq-512x512.pkl [17]. Manipular características específicas, como la edad, el género y las expresiones faciales, requiere la búsqueda y modificación de vectores latentes. Este proceso implica identificar representaciones latentes

que preserven la apariencia del sujeto (por ejemplo, identidad, iluminación, peinados) mientras se permiten manipulaciones significativas. Para abordar este desafío, se utilizan técnicas como el Análisis de Componentes Principales (PCA) o clasificadores supervisados que identifican direcciones específicas en el espacio latente, permitiendo ediciones precisas y controladas [18].



Figura 4: Ejemplo de imagen sintética generada con StyleGAN3. Se buscó generar una imagen de una persona mirando al frente con un fondo gris-neutro.

Fuente: Elaboración propia usando STYLEGAN3 y el modelo preentrenado stylegan2-ffhq-512x512.pkl

II-F. Desarrollo de una estrategia para la elaboración de modelos de Machine Learning

Uno de los puntos más importantes a considerar, cuando se desarrolla un sistema con cumplimiento normativo desde el diseño, es que *las condiciones y restricciones dejan de ser definidas por la necesidad técnica y pasan a ser dependientes de la normativa*. Para la metodología propuesta, quién determina las restricciones para la data disponible es el resultado del análisis DPIA. Esta es una característica fuerte que limita en gran medida los enfoques tradicionales, ya que éstos requieren de grandes cantidades de datos reales.

Bajo esta restricción se evaluó el uso de redes profundas que pudieran ser entrenadas para detectar los distintos tipos de fraude en un único paso:

- Redes CNN profundas, VGG 16/19
 - Ventajas: Alto rendimiento en clasificación de imágenes.
 - Requerimiento: Alta demanda de hardware y gran número de parámetros (138-144 millones).

- ResNet
 - Ventajas: Captura características complejas y profundas en imágenes.
 - Requerimientos: Alta demanda de hardware (GPU potente) y gran cantidad de imágenes para entrenamiento (ResNet-152 tiene más de 60 millones de parámetros).

Sin embargo éstas presentan los siguientes problemas fundamentales:

- Cantidad de datos: Los fraudes más elaborados afectan sólo unos pocos pixels de las imágenes, por lo que se requiere de grandes números de imágenes para que las redes profundas capturen los patrones. Esto atenta contra los principios de la protección de datos (si se usan datos

reales) o aumentan la carga de la generación de imágenes sintéticas.

- Capacidad de actualización: Cada día están apareciendo nuevos tipos de fraude, al principio no se cuenta con suficientes ejemplos reales. Si bien se pueden generar datos sintéticos, la actualización implica re-entrenar todo el sistema, afectando la calidad de detección de fraudes previos para adaptar la red completa al fraude nuevo.

Lo anterior llevó a desarrollar una estrategia de solución basada en segmentar el problema en clases individuales, en lugar de un modelo único, proponiendo combinar una serie de modelos sencillos, con bajo número de parámetros, que se focalicen en la fortaleza de cada modelo y en la debilidad de cada fraude.

II-G. Integración de modelos

Para la integración de los modelos individuales, se propone una cañería modular secuencial, donde los modelos se van ejecutando en orden. Si uno de los modelos indica un potencial fraude (clase positiva) se detiene la ejecución.

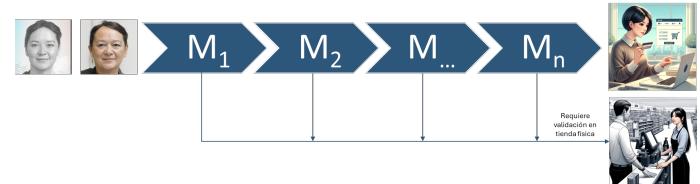


Figura 5: Ejecución modular secuencial. Recibe una imagen de identificación y una de validación, devuelve positivo ante la posibilidad de fraude, negativo en caso contrario.

Fuente: Elaboración propia

Para asegurar la mejor capacidad integrada, los modelos se prueban de manera cruzada y se realiza una optimización del orden en función del costo de los errores tipo I y tipo II de cada modelo, lo que permite tener un sistema integrado que funciona no según indicadores técnicos, sino en función del equilibrio de costos entre no detectar una imagen fraudulenta (que se materializaría en fraude) versus rechazar una imagen legítima, lo que podría implicar perder al cliente.

III. ESTUDIO DE CASO: EMPRESA DE SERVICIOS FINANCIEROS

III-A. Contexto empresarial y desafíos

La compañía analizada pertenece al sector minorista financiero chileno. Su mercado objetivo es el nivel socioeconómico medio-bajo. Su producto principal es la tarjeta de crédito, la cual tiene un cupo inicial del orden de USD 250. Como parte de su estrategia digital, ha incorporado el onboarding digital para la captura de nuevos clientes. Si bien esto le ha permitido asegurar mayor participación del marketshare al que apunta, también le ha impactado en el aumento de fraudes por suplantación de identidad. De 120 mil procesos de onboarding iniciados durante el 2022, solo 43 mil finalizaron correctamente, sin embargo, 9 % de este grupo resultaron ser fraudes, presentando pérdidas del orden de MUSD 1, representando este valor un 6 % de sus ventas. La compañía tomó medidas de remediación como la contratación de un SaaS

de verificación de identidad con un costo anual del orden de kUSD 300, no obstante, la efectividad del proceso no mejoró según esperado, por lo que debió disponer de un equipo de 7 validadores para chequear manualmente los potenciales casos de fraude, con un costo anual del orden de kUSD 67. Esto puede generar un delay time de hasta 48 horas para validar identidad. Si hay dudas de verificación, se le solicita al cliente validar su identidad en una sucursal física de la tienda, lo que puede generar en algunos casos que los clientes desistan del proceso. Por lo tanto, un alto costo por la mantención de un SaaS de verificación de identidad junto con un equipo dedicado de personas ha sido una solución costosa que no ha garantizado la disminución de la tasa de fraudes (tasa error tipo II).

III-B. DPIA

III-B1. Primer paso, identificar la necesidad de realizar DPIA

Actualmente, la compañía del caso en estudio está reco-giendo, almacenando y usando datos personales en la etapa del proceso de validación de identidad para la apertura de cuentas digitales. Esta forma de operar expone a riesgos la información de las personas, los cuales se ven aumentados por el hecho de que un equipo humano está validando identidades y detectando intentos de fraude manualmente. Por lo tanto, se cumple la necesidad de realizar un DPIA.

III-B2. Definir las características del proyecto

Básicamente, la compañía en cuestión tiene el siguiente flujo de información: mediante una aplicación de celular se captura la imagen de la cédula de identidad y una fotografía del rostro de la persona. Esta información es procesada por un sistema SaaS para validar la identidad. Sin embargo, también es procesada y comparada por un equipo de siete personas que confirman si la identidad de la persona corresponde o si existe algún tipo de adulteración (intento de apertura fraudulenta). En el caso de que no lo puedan determinar, se invita al cliente a realizar una apertura presencial.

III-B3. Identificación de riesgos asociados a la protección de datos personales

El análisis del flujo de información descrito permitió detectar riesgos para la compañía (normativos) y riesgos para las personas. Cabe destacar que la operación de la compañía se desarrolla en Chile, donde la Ley de Protección de Datos Personales está *ad portas* de ser actualizada y se basa en gran medida en el GDPR. Los riesgos detectados son:

- No se cuenta con el consentimiento explícito del usuario para capturar, almacenar y procesar datos personales biométricos, tampoco señala el periodo de uso o el ciclo de vida de los datos, es decir, podrían ser almacenados indefinidamente. Esto resulta en un incumplimiento legal de acuerdo a la nueva normativa.
- Con respecto al uso del SaaS, no es claro de qué forma se están utilizando las imágenes personales para entrenar el modelo clasificador.
- La validación de identidad se está realizando de forma

manual por un equipo de validadores. Si no se aplican medidas de control exigentes, existe el riesgo de divulgación inapropiada dentro de la organización y sesgos en la clasificación de los casos.

- Para los usuarios existe el riesgo de que las imágenes del rostro (e incluso otros datos personales) sean accedidos ilícitamente debido a brechas de seguridad.
- Al no existir una declaración de uso explícita, las imágenes capturadas podrían ser utilizadas para otros fines, distintos a la validación de identidad, por ejemplo, perfilado o vigilancia.
- Los daños causados a las personas por un mal manejo de la protección de datos personales pueden derivar en demandas legales hacia la compañía.

III-B4. Medidas para mitigar el riesgo

De acuerdo con el análisis de riesgo, el proceso de validación de identidad no cumple con los preceptos de la privacidad por diseño y por defecto, de modo que se deberían hacer modificaciones importantes para mitigar el riesgo, tanto para la compañía como para las personas. Algunas medidas son:

- Obtener el consentimiento explícito del usuario para utilizar sus datos biométricos, señalando claramente cómo se utilizarán sus datos, por cuánto tiempo se utilizarán y permitiéndole realizar consultas sobre el estado de uso de sus datos.
- Reducir el tiempo de uso de los datos personales y biométricos al mínimo necesario para la verificación de identidad y eliminarlos una vez terminado el proceso.
- Utilizar técnicas de cifrado de datos que permitan proteger la información biométrica.
- Fortalecer la validación automatizada de identidad y hacer que esta prevalezca sobre la revisión manual.

No es tarea fácil implementar cada una de estas medidas, en particular, la automatización de la validación de identidad, dado que los modelos de machine learning y deep learning utilizados para este tipo de clasificaciones necesitan ser entrenados con datos biométricos (una enorme cantidad de estos). Además, en tal caso, se debería tener el consentimiento de cada usuario para que su imagen sea utilizada en el entrenamiento de uno o más modelos y los datos deberían ser utilizados por un tiempo acotado, lo que se traduce en una limitación importante para el proceso de validación en el onboarding digital y, aún así, persistiría el riesgo de que los datos fuesen accedidos indebidamente.

De acuerdo al análisis realizado se recomienda el uso de datos sintéticos para entrenar los modelos de validación de identidad. Al entrenar los modelos con rostros de personas que no existen en la realidad, no es necesario almacenar las imágenes de los usuarios, eliminando así el riesgo de fuga de información sensible. Además, esto agiliza el proceso de validación y supera las restricciones legales, ofreciendo una excelente medida para proteger tanto a la compañía como a los usuarios.

III-C. Definición y límites de batería

Para el escenario evaluado, se consideró que el proceso de captura de la imagen de la cédula, la prueba de vida y el pre-processing son parte del pipeline de captura de la aplicación de la empresa. El punto de entrada para la validación consiste en dos imágenes de rostros, una correspondiente al documento oficial y otra obtenida desde la prueba de vida para su validación. El producto esperado es una clasificación binaria, donde la clase positiva indica un potencial fraude. La información del clasificador entra nuevamente al proceso de la empresa para la toma de acción.



Figura 6: Límites de batería del caso analizado.

Fuente: Elaboración propia

III-D. Objetivos para el caso analizado

■ GENERALES

1. Equilibrar los errores tipo I y tipo II en la validación de la imagen de la cédula de identidad.
2. Cumplir con las regulaciones de protección de datos al utilizar **datos sintéticos** para el entrenamiento.

■ ESPECÍFICOS

1. Desarrollar un modelo de *machine learning* que cumpla con:
 - **KPI - Precisión del modelo > 85 %**
 - **KPI - Tasa de error tipo II < 9 %**
 - **KPI - Reducción de costo operacional anual < kUSD\$ 367**

■ ENTREGABLE

- Sistema de validación que permita tomar 2 imágenes, una de ID y una de validación, y detecte un potencial fraude (caso verdadero).

■ ALCANCE

- Este trabajo llega hasta la entrega de las rutinas y modelos en Python realizados con data sintética listos para ser ajustados (fine-tuning con data real) y probados por el cliente.
- Se encuentra fuera del alcance el pre-processing necesario para preparar las imágenes y la implementación final.

III-E. Análisis Exploratorio de los Datos (EDA)

Para el desarrollo de esta investigación, se entrevistó al Gerente de Riesgo de Crédito de la empresa en estudio, con la finalidad de entender en detalle la problemática de la suplantación de identidad en el proceso de onboarding digital. Según informó, durante el período de abril y agosto

de 2023, de un total cercano a 14 000 procesos de onboarding completados, se detectaron aproximadamente 450 intentos de fraude. Estos incidentes estaban relacionados principalmente con inconsistencias entre la foto del rostro de la cédula de identidad y la selfie como prueba de vida. En un menor porcentaje, las adulteraciones eran otros sectores de la cédula de identidad. Por ende, la validación del rostro de la cédula de identidad comparada con la imagen del rostro (prueba de vida) constituyeron el foco principal de este estudio.



Figura 7: Cédula de Identidad de la República de Chile vigente 2024.

Fuente: Registro Civil de Chile



Figura 8: Área de la fotografía del titular de la Cédula de Identidad de Chile vigente 2024.

Fuente: Elaboración propia.

Es importante señalar que las adulteraciones y/o discrepancias que se observan en las cédulas de identidad identificadas como fraude son burdas y no presentan ningún grado de sofisticación, no están hechas por personas especializadas en fraudes de falsificación, por lo que pueden detectarse a simple vista. Complementando la entrevista realizada al Gerente de Riesgo de Crédito, también se entrevistó a su equipo, responsable de la revisión manual de cada cédula de identidad enviada para la apertura de cuentas. Este equipo se encarga de separar aquellas cédulas que presentan dudas sobre su autenticidad, basándose en los criterios mencionados previamente. El equipo compartió que, de los aproximadamente 450 intentos de fraude detectados en el período de abril a agosto de 2023, la mitad correspondía a adulteraciones en la zona de la fotografía del rostro del titular. Estas adulteraciones fueron clasificadas en Clases de acuerdo con el tipo del que se trate:

- Clase 0: No es la persona. La fotografía del rostro que aparece en la cédula de identidad no corresponde al titular de esta.
- Clase I: No tiene la edad correspondiente. La fotografía del rostro no corresponde a una persona mayor de edad (18 años o más) habilitada para abrir una cuenta.
- Clase II: Fotografía a color. La fotografía de la cédula de

identidad presenta colores diferentes al formato original en blanco y negro con escala de grises.

- Clase III: Área del rostro no corresponde. La fotografía de la cédula de identidad ha sido reemplazada en el área del rostro (óvalo).
- Clase IV: Colores extraños. En el área de la fotografía del titular aparecen zonas con colores que no corresponden al formato original.

# Clase	Descripción	Ejemplo sintético
0	No es la persona	
I	No tiene la edad correspondiente	
II	Fotografía a color	
III	Área del rostro no corresponde	
IV	Colores extraños	

Figura 9: Ejemplos sintéticos de adulteraciones de la zona del rostro de la Cédula de Identidad por Clase.

Fuente: Elaboración propia.

Según se informó, dentro del período de abril a agosto de 2023, la distribución para cada Clase dentro del total de adulteraciones en la zona de la fotografía del rostro del titular de la cédula de identidad es la siguiente:

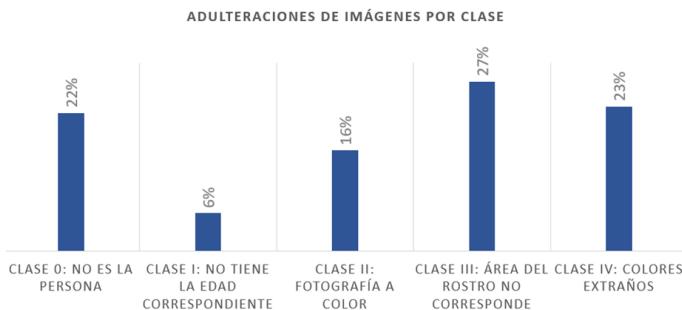


Figura 10: Participación porcentual por Clase en adulteraciones de la zona de la fotografía del rostro de la Cédula de Identidad.

Fuente: Elaboración propia.

III-F. Generación de datos sintéticos utilizando StyleGAN

Para este caso específico, la data sintética debe cumplir no sólo con ser fotorrealista y diversa, sino que debe cumplir con una necesidad específica. En el caso se busca entrenar modelos para detección de fraude, lo que implica usar data sintética consistente con el flujo del caso de negocio, el cual toma dos imágenes, una equivalente a un documento de identidad y otra

equivalente a una foto de validación. Por otro lado, se requiere emular que las capturas de estas imágenes no sean perfectas (ruido, iluminación, desenfoque, otros).

Las características clave de la data sintética son:

- Imagen directa de stylegan: Se requiere de una imagen con una persona aleatoria, pero la persona debe estar centrada, mirando al frente, además el color del fondo debe ser un gris-neutro que pueda ser diferenciado fácilmente de la persona, para poder transformar la imagen en una identificación. Este paso es complejo ya que requiere entrenar o contar con un modelo pre-entrenado capaz de generar imágenes que no sólo sean fotorrealistas, sino que cumplan con las condiciones definidas. Dado que el entrenamiento desde cero de un modelo de este tipo requiere decenas de miles de imágenes y un alto nivel de recursos, se optó por usar StyleGAN2-ffhq-512x512.pkl el cual es un modelo preentrenado de StyleGAN2 desarrollado por NVIDIA para generar imágenes de alta calidad de rostros humanos con una resolución de 512x512 píxeles. Este modelo fue entrenado con el conjunto de datos FFHQ (Flickr-Faces-HQ), que contiene 70 000 imágenes de alta calidad de caras humanas en diversas edades, etnias y condiciones de iluminación. Sin embargo, no basta con contar con el modelo, se debe navegar el espacio latente para conseguir generar imágenes de las características específicas. En este caso, se requería generar un set de personas distintas, con la pose correcta y el fondo neutro. Ésto se logró mediante una exploración del espacio latente y determinación de conjuntos de coordenadas que cumplieran con las condiciones requeridas.



Figura 11: Imágenes generadas en la exploración del espacio latente.

Izquierda, búsqueda de distancia para diferenciar identidad. Derecha, búsqueda de vector semántico asociado al fondo neutro

Fuente: Elaboración propia

- Imagen modificada para representar la de una cédula de identidad: Se requiere emular las características distintivas de una cédula, incluyendo colores, tonos, sellos. Además, se deben aleatorizar sus características para reproducir la variabilidad de una imagen capturada, las cuales tienen cierta rotación, desenfoque, distintos niveles de brillo, etc.



Figura 12: Cédula referencia (primera de izquierda a derecha) versus sintéticas. Las sintéticas tienen brillo, ruido y enfoque aleatorizado para emular el proceso de captura.

Fuente: Elaboración propia

- Imagen de validación: Debe ser de la misma persona, pero con más edad, para reproducir la diferencia temporal entre la cédula y la validación. También debe aleatorizar sus características, para ello se realizó una búsqueda de las coordenadas que permitieran cambiar la edad sin modificar la identidad.



Figura 13: Búsqueda del vector semántico para control de edad

Fuente: Elaboración propia

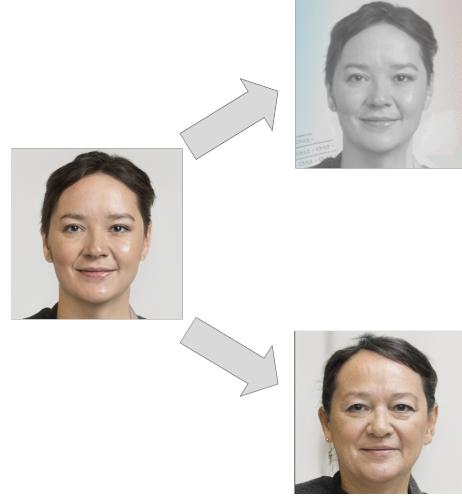


Figura 14: Imagen creada con stylegan y su transformación en cédula y validación, donde la validación tiene una edad distinta a la cédula.

Fuente: Elaboración propia

- Imagen adulterada: Se deben reproducir las condiciones de fraude detectadas en el análisis exploratorio de datos (EDA) descrito previamente. Esto implica seleccionar una identidad sintética “robada” y una identidad que intenta cometer el fraude, luego se debe crear la cédula adulterada para intentar usar la imagen de validación.



Figura 15: Proceso creación fraude sintético.

Fuente: Elaboración propia

El proceso completo se resume en el siguiente esquema:

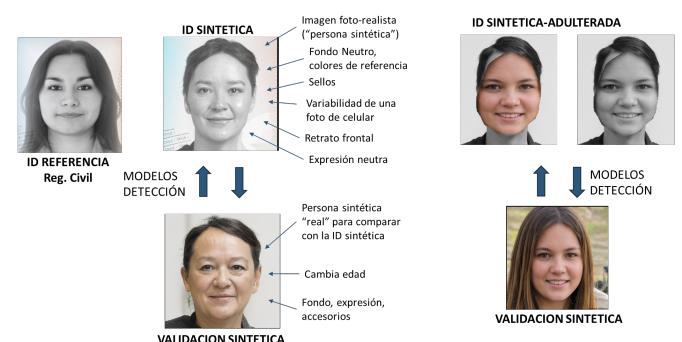


Figura 16: Proceso de generación imágenes sintéticas.

Fuente: Elaboración propia

La automatización del proceso implicó un flujo de trabajo que consistió en la definición de un conjunto de vectores latentes, cada uno con una persona distinta, pero cumpliendo con las condiciones necesarias para su uso según lo indicado.

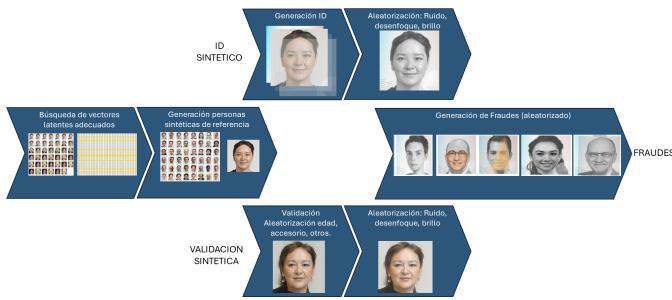


Figura 17: Flujo de trabajo generación imágenes sintéticas.
Fuente: Elaboración propia

Con el flujo previo, se creó un set de 15 000 imágenes aleatorias, desde las cuales se crearon las respectivas cédulas y validaciones. Luego fueron separadas en conjuntos de 5 000 imágenes: uno de referencia (imágenes que constituyen la clase negativa, es decir imágenes no adulteradas) y otro de validaciones adulteradas (clase positiva). El tercer grupo de 5 000 se usa para la creación de las adulteraciones, permitiendo no repetir personas del primer o segundo grupo.

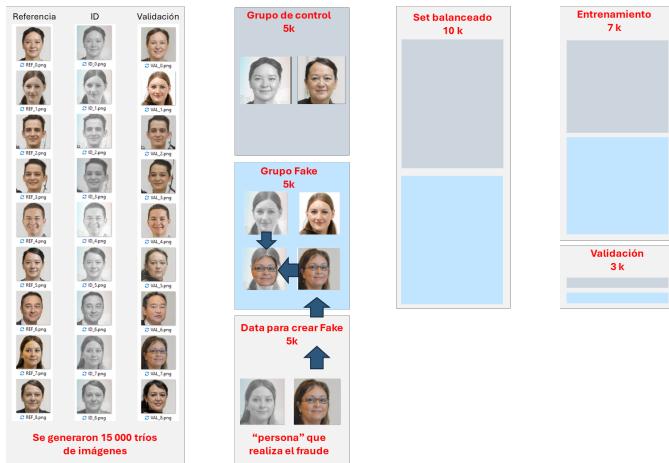


Figura 18: Conjunto de imágenes y su división para entrenamiento y validación.

Fuente: Elaboración propia

III-G. Desarrollo y entrenamiento de modelos individuales

Tal como se mostró en el EDA, los tipos de fraude detectados pueden ser agrupados según características específicas. Bajo la metodología desarrollada, se busca desarrollar modelos simples que aprovechen las vulnerabilidades de cada tipo de fraude mediante modelos específicos en lugar de un único modelo. Dentro de esta lógica, cada clase de fraude se resuelve mediante un modelo ad-hoc y el modelo se encapsula en una clase de python para posteriormente realizar la integración final.

Los modelos se desarrollaron considerando que la clase positiva corresponde a un potencial fraude y la clase negativa a una validación legítima. Dado esto, el error tipo I corresponde a un falso positivo, es decir, cuando una validación

legítima es declarada fraudulenta. Inversamente, el error tipo II corresponde a un intento de fraude que no es detectado por el sistema.

III-G1. Clase de Fraude 0: Cuando la persona no es la misma

Este caso es el punto base del proceso de validación de identidad (por eso se denominó clase 0), es decir, asegurar que la persona es la misma al comparar la imagen de ID (cédula de identidad) y la imagen de validación (19).



Figura 19: Clase 0, validación de identidad.
Fuente: Elaboración propia

Hoy en día, el proceso de comparar identidad se encuentra bastante estandarizado, especialmente mediante el uso de landmarks [19]. Uno de los focos de la metodología es la modularidad, en particular por su capacidad de re-utilizar modelos existentes o sistemas pre-entrenados, lo que permite acelerar el desarrollo y bajar el costo de la solución. Para el presente caso se utilizó la biblioteca DLIB [20] con los modelos pre-entrenados de detección de rostro y landmarks [21], [22].

El modelo pre-entrenado permite codificar un rostro en su vector de landmarks. Para el proceso de validación de la identidad se calculó la distancia L2 entre ambos vectores. Para determinar el mejor punto de corte se realizó un barrido de las distancias, realizando una optimización de los errores tipo I y tipo II. Dado que se contaba con una estimación del costo diferenciado por error, se realizó una evaluación técnica y económica para determinar el mejor corte.

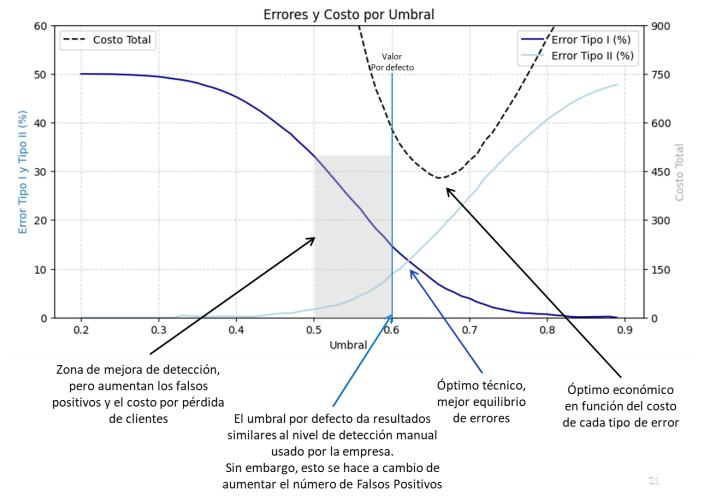


Figura 20: Punto de corte en función de mejor equilibrio técnico/económico de los errores tipo I/II

Fuente: Elaboración propia

III-G2. Clase de Fraude I: Cuando la persona es menor de edad

Esta clase corresponde al uso de la imagen de un menor para adulterar la ID. Por lo general este caso corresponde a un escenario mixto donde se usa la cédula de un menor y la validación de un adulto (clase 0) o se usa la imagen de un menor para adulterar la ID (clases II y III).

Esta clase no es trivial de resolver, por un lado estimar la edad de una persona es un proceso complejo no sólo para una máquina. Por un lado no todas las personas envejecen de la misma manera, por otro, personas de distintas ascendencias presentan rasgos de envejecimiento de manera diferente.[23]. Por otro lado, se deben considerar una serie de aspectos éticos y legales, especialmente en los potenciales sesgos que se puedan injectar en los sistemas de IA.



Figura 21: Clase I, el ID corresponde a un menor de edad (nota: el menor de la imagen es una persona sintética)

Fuente: Elaboración propia

Para la solución se usó una red pre-entrenada basada en pytorch [24]. Según el desarrollador, el modelo tiene una precisión del orden de 65 %. En cuanto a la estimación de edad, se tiene un error absoluto medio de 3 años en la edad, lo que indica una precisión razonable.

Para la implementación del modelo, se intentó evaluar un punto de corte de la misma manera que en el caso previo; sin embargo, los resultados no fueron tan buenos. Esto se debió a que, al crear menores sintéticos con stylegan, no se tiene pleno control sobre la edad en el vector latente, lo que implica que se debe generar un conjunto de imágenes que luego son etiquetadas manualmente. Este proceso limitó el número de imágenes y, al mismo tiempo, introdujo un sesgo, ya que resulta difícil para los etiquetadores detectar imágenes que se encuentran en la zona de quiebre (18 años).

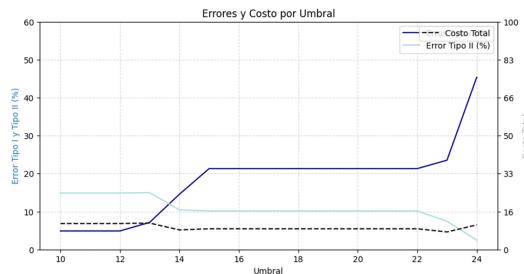


Figura 22: Punto de corte para clase I

Fuente: Elaboración propia

Dado lo anterior y en función de las especificaciones técnicas del modelo, se optó por usar 18+3 años como punto de corte del modelo. Si bien las capacidades de detección en esta clase no son las mejores, se debe considerar que la naturaleza mixta de esta clase permite que en la integración se mezclen

las capacidades de detección.

III-G3. Clases de Fraude II y IV: Cuando se adultera, pero se modifica la firma de colores

Este caso corresponde a la inserción de un rostro a color para adulterar la imagen o cuando se detectan colores específicos (esto último corresponde a una especificación del proveedor de la muestra de fraudes).



Figura 23: Clases II y IV, corresponden a fraudes distinguibles por el uso de color

Fuente: Elaboración propia

Siguiendo la metodología, se realizó una evaluación de alternativas para el desarrollo de un modelo simple de detección. Es importante considerar que el perfil de colores de una cédula se encuentra bien definido por las especificaciones del documento emitido por el Registro Civil. Sobre esta base, se calcularon los histogramas por canal de las imágenes de referencia (cédulas no adulteradas) versus imágenes modificadas. Al revisarlos, se observa cómo las ID de referencia tienen formas consistentes para cada canal. Para simplificar la comparación, se calculó un histograma único a partir de los 3 canales. El cálculo se basó en las diferencias entre el valor porcentual de cada bin entre los tres canales. Al revisar el vector creado, se observó que las diferencias entre la firma de color de las imágenes legítimas y las adulteradas se encuentran principalmente en los primeros 30 bins del histograma.

Esta ingeniería de características permitió reducir una imagen a un vector de pequeño tamaño donde la separación de clases es fácilmente distinguible.

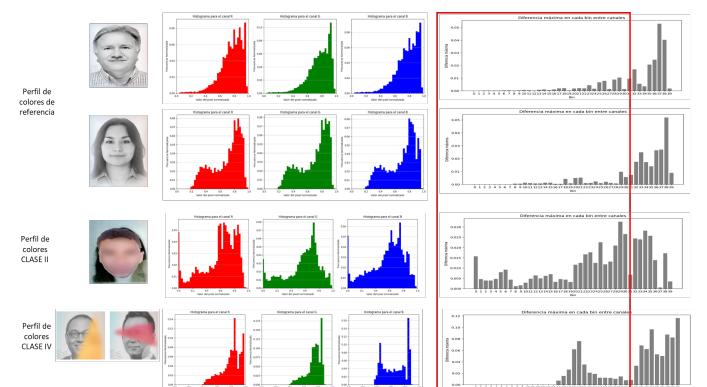


Figura 24: Detección de fraude mediante diferencias en la firma de colores.

Fuente: Elaboración propia

Como primera aproximación se evaluó un conjunto de imágenes (pares ID y validación) balanceado mediante Auto Machine Learning (AML), en particular autogluon. El objetivo de este paso es tener una primera idea de las capacidades de detección de los distintos modelos para este problema en específico. Dado que la ingeniería de características permitió

transformar las imágenes en un vector simple y bien diferenciado, los resultados fueron muy buenos.

Model	Score (Val)	Eval Metric
NeuralNetFastAI	0.988333	Accuracy
WeightedEnsemble_L2	0.988333	Accuracy
LightGBM	0.987500	Accuracy
XGBoost	0.986667	Accuracy
LightGBMXT	0.986667	Accuracy
CatBoost	0.985833	Accuracy
NeuralNetTorch	0.985833	Accuracy
KNeighborsDist	0.985000	Accuracy
KNeighborsUnif	0.985000	Accuracy
LightGBMLarge	0.984167	Accuracy
ExtraTreesGini	0.982500	Accuracy
ExtraTreesEntr	0.981667	Accuracy
RandomForestGini	0.981667	Accuracy
RandomForestEntr	0.981667	Accuracy

Cuadro I: Ranking Resultados AML

Dada la similitud en los resultados en la prueba preliminar, la selección del modelo a usar se basó en las fortalezas propias de cada uno. En este caso, dado que se está usando data sintética, uno de los requerimientos base es la capacidad de generalizar y la dependencia de los hiperparámetros, es por ello que se tomaron en cuenta los aspectos de robustez y estabilidad. Dado esto se optó por RANDOMFOREST, en particular por[25]:

- Robustez y reducción de sobre ajuste: RandomForest es conocido por su capacidad para reducir el sobre ajuste debido a que combina múltiples árboles de decisión independientes. Cada árbol en el bosque se entrena con un subconjunto diferente de datos y características, lo que mejora la capacidad del modelo para generalizar a nuevos datos.
- Estabilidad y precisión: RandomForest tiende a ser más estable y preciso en comparación con modelos individuales como árboles de decisión, debido a su enfoque de votación mayoritaria entre múltiples árboles.
- Versatilidad: RandomForest Es menos sensible a la variabilidad en los datos, lo que lo hace adecuado para una amplia gama de problemas.

Para el entrenamiento del modelo, se realizó una optimización de hiperparámetros, se configuró una grilla de búsqueda para explorar diversas combinaciones de hiperparámetros clave. Estos hiperparámetros incluyan el número de árboles en el bosque (n estimators) con valores de 10, 50, 100 y 500; la profundidad máxima de cada árbol (max depth) con opciones de ninguno, 10 y 20; el número mínimo de muestras requeridas para dividir un nodo (min samples split) con valores de 2 y 10; el número mínimo de muestras requeridas en una hoja (min samples leaf) con valores de 1 y 4; y el número de características a considerar al buscar la mejor división (max features) con el valor de 'sqrt'.

Se realizaron un total de 480 ajustes mediante validación cruzada de 10 pliegues para cada una de las 48 combinaciones de hiperparámetros.

Los mejores hiperparámetros encontrados fueron: max depth de 10, max features de 'sqrt', min samples leaf de 1, min samples split de 2 y n estimators de 50.

Los resultados de la evaluación del modelo mostraron una

precisión, un recall y una puntuación f1 de 0.99 tanto para la clase 0 como para la clase 1, con una precisión global del 99 % sobre un conjunto de 8000 muestras, lo que indica un rendimiento excepcional del modelo optimizado. Este resultado hace sentido si se considera el tamaño reducido del conjunto de características, la buena separación lograda en la ingeniería de características y el tamaño del conjunto de entrenamiento.

III-G4. Clases de Fraude III: Cuando se adultera, pero se mantiene la firma de colores

Este caso es más complejo, ya que no puede ser detectado simplemente mediante la firma de colores, por lo que se debió realizar la detección mediante los patrones de la imagen misma.



Figura 25: Clase III, se adultera validación mediante reemplazo de óvalo o imagen completa

Fuente: Elaboración propia

Siguiendo la metodología, se trató de evitar el uso de un modelo que analizara la imagen completa. El motivo de esto es que la imagen de la cara de una persona contiene mucha información que varía entre individuos, lo cual se transforma en distractores al usar un modelo para detectar los patrones que corresponden a un fraude. Los modelos que intentan hacer esto inevitablemente requieren un alto número de parámetros para poder capturar la complejidad y separar los elementos que identifican un fraude de las diferencias legítimas de cada persona, lo que lleva a un elevado número de imágenes de entrenamiento y una gran capacidad de cómputo. Por ejemplo, ImageNet fue entrenado con más de un millón de imágenes [2].

Bajo esta lógica, se analizaron las muestras de fraude, observándose que *la debilidad de esta clase está en los puntos de unión entre la imagen original y la modificación*. Este aspecto permite focalizar el entrenamiento de la detección en los puntos clave, eliminando los elementos distractores. Para realizar la ingeniería de características, se utilizó el modelo de detección de landmarks y las coordenadas de la imagen para extraer las zonas críticas. Al revisar los píxeles de la zona capturada, se puede observar cómo el procedimiento permite ver a simple vista los puntos adulterados, algo que es complejo de detectar en la imagen completa.



Se genera una imagen que codifica el borde
Se observa como las irregularidades se resaltan.

Figura 26: Clase III, Captura de zonas vulnerables y transformación a imagen procesada para entrenamiento

Fuente: Elaboración propia

Para la creación del modelo, se optó por el enfoque clásico de clasificadores de imágenes mediante una red neuronal convolucional. La arquitectura se basó en la red del Visual Geometry Group (VGG)[1] debido a su reconocida capacidad para la clasificación. Esto consistió en los siguientes elementos clave:

- **Profundidad:** Múltiples capas de convolución para aumentar la profundidad de la red, lo que permite captar características complejas.
- **Convoluciones pequeñas:** Filtros pequeños de 3x3 en las capas de convolución, lo que ayuda a reducir el número de parámetros y la carga computacional.
- **Uniformidad:** Estructura uniforme a lo largo de la red, facilitando el diseño y el ajuste de hiperparámetros.
- **Capas de pooling:** Para reducir la dimensión espacial de las características, lo que ayuda a controlar el sobreajuste y a reducir la complejidad computacional.

Para el desarrollo de la arquitectura, se comenzó con una red básica y se fue incrementando su complejidad hasta lograr una buena clasificación. La red final consistió en capas de convolución iniciales con 32 filtros de tamaño 3x3 y activación ReLU, seguidas por una capa de MaxPooling de tamaño 2x2 y una capa de Dropout con una tasa de 0.25 para prevenir el sobreajuste.

Este patrón se repitió, aumentando a 64 filtros en las capas de convolución subsiguientes. Posteriormente, la red se aplano mediante una capa Flatten, conectada a una capa densa de 128 unidades con activación ReLU, seguida de otra capa de Dropout con una tasa de 0.5.

Finalmente, la salida del modelo se logró mediante una capa densa con una única unidad y activación sigmoide, adecuada para tareas de clasificación binaria.

Esta configuración permitió captar y procesar las características de las imágenes, manteniendo un equilibrio entre la capacidad del modelo y la mitigación del riesgo de sobreajuste.

Para la optimización, la red se entrenó dos veces: primero, con 40 épocas para determinar el punto de corte óptimo donde detener el entrenamiento, y luego, una segunda vez utilizando

ese punto óptimo, deteniéndose en 17 épocas.

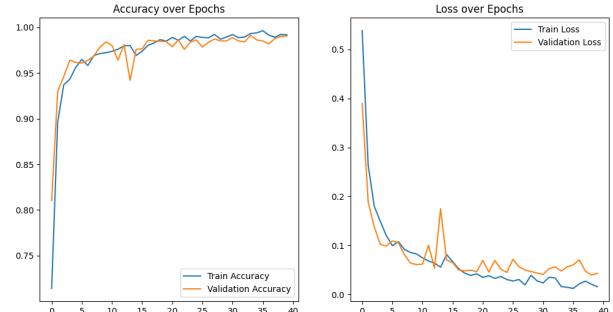


Figura 27: Entrenamiento de la red neuronal convolucional.

Fuente: Elaboración propia

Para la clase 0, la precisión alcanzada fue de aproximadamente 92.2 %. La puntuación F1 fue de 95.8 %. En cuanto a la clase 1, la precisión fue de aproximadamente 91.7 % con una puntuación F1 de 95.5 %. La exactitud general del modelo fue de 95.6 %. Estos indicadores demuestran que el modelo es altamente efectivo en la clasificación de ambas clases, con un rendimiento equilibrado entre precisión y recall.

III-H. Pipeline de modelos modulares para la detección de fraude

Como se estableció en la metodología, la integración se llevó a cabo mediante una cañería modular secuencial simple que va ejecutando cada modelo uno a uno. Para poder determinar el orden óptimo se hace necesario evaluar la capacidad cruzada de detección, esto es, la capacidad de cada modelo individual respecto de todos los tipos de fraude. Los resultados de esta evaluación se muestran en las tablas II y III.

El planteo de la optimización se basó en el costo de cada clasificador en el contexto del negocio, esto es, la potencial pérdida económica al no detectar un fraude (error tipo II) versus el costo de oportunidad de pérdida de un cliente que es clasificado como fraude y que no termina el proceso de on boarding.

Lo anterior se plantea como: para cada modelo j considerando la probabilidad de no detectar un fraude (Falso Negativo) versus detectar erróneamente a alguien (Falso Positivo) en función de la frecuencia del tipo de fraude específico i y la probabilidad de perder el cliente (Falso Positivo) :

$$Z = F_i \cdot (C_f \cdot TFN_{ij} + P_{perdida} \cdot C_{op} \cdot TFP_{ij})$$

Lo que se traduce en lo siguiente función de optimización:

$$\min Z = \sum_i \sum_j x_j \cdot (F_i \cdot (C_f \cdot TFN_{ij} + P_{perdida} \cdot C_{op} \cdot TFP_{ij}))$$

Sujeto a:

Variables de decisión:

x_j : Variable binaria que indica si se utiliza el Modelo j

Parámetros:

- F_i : Frecuencia de ocurrencia del fraude tipo i .
- TFN_{ij} : Probabilidad de no detectar un tipo de fraude i por parte del modelo j

Tipo de fraude i	Modelo	Frecuencia F_i	TFN_M0	TFN_MI	TFN_MII	TFN_MIIIA	TFN_MIIIB	TFN_MIV
0	Compara rostros ID/Foto	$F_0 = 23,4\%$	29,3 %	100,0 %	100,0 %	100,0 %	100,0 %	100,0 %
I	Estima si edad es mayor a 18 años	$F_I = 6,4\%$	100,0 %	21,3 %	100,0 %	100,0 %	100,0 %	100,0 %
II	Detecta adulteración: pegado de foto color sobre ID	$F_{II} = 17,0\%$	100,0 %	100,0 %	0,0 %	3,4 %	99,9 %	0,0 %
III-A	Detecta adulteración: pegado de foto gris sobre ID	$F_{III} = 14,4\%$	100,0 %	100,0 %	0,8 %	1,4 %	98,2 %	95,7 %
III-B	Detecta adulteración: ID sobre ID	$F_{III} = 14,3\%$	100,0 %	100,0 %	94,0 %	99,9 %	2,6 %	94,0 %
IV	Detecta colores inusuales en ID	$F_{IV} = 24,5\%$	100,0 %	100,0 %	1,9 %	93,2 %	99,9 %	1,9 %

Cuadro II: Tabla de tipos de fraude y modelos (Parte 1)

Tipo de fraude i	Modelo	Frecuencia F_i	TFP_M0	TFP_MI	TFP_MII	TFP_MIIIA	TFP_MIIIB	TFP_MIV
0	Compara rostros ID/Foto	$F_0 = 23,4\%$	7,1 %	0,0 %	0,0 %	0,0 %	0,0 %	0,0 %
I	Estima si edad es mayor a 18 años	$F_I = 6,4\%$	0,0 %	10,2 %	0,0 %	0,0 %	0,0 %	0,0 %
II	Detecta adulteración: pegado de foto color sobre ID	$F_{II} = 17,0\%$	0,0 %	0,0 %	2,3 %	3,0 %	0,1 %	2,3 %
III-A	Detecta adulteración: pegado de foto gris sobre ID	$F_{III} = 14,4\%$	0,0 %	0,0 %	1,1 %	1,1 %	1,1 %	1,1 %
III-B	Detecta adulteración: ID sobre ID	$F_{III} = 14,3\%$	0,0 %	0,0 %	2,3 %	0,1 %	0,1 %	2,3 %
IV	Detecta colores inusuales en ID	$F_{IV} = 24,5\%$	0,0 %	0,0 %	2,3 %	3,0 %	0,1 %	2,3 %

Cuadro III: Tabla de tipos de fraude y modelos (Parte 2)

- TFP_{ij} : Probabilidad de detectar erróneamente un tipo de fraude i por parte del modelo j
- C_f : Pérdida promedio por fraude no detectado.
- $P_{perdida}$: Porcentaje de clientes que se pierden cuando son identificados erróneamente.
- C_{op} : Pérdida promedio por cliente perdido.

Se debe seleccionar sólo un modelo, por lo que:

$$\sum_j X_j = 1$$

Se debe tener en consideración que este problema de optimización no se puede resolver directamente, ya que la lógica de cañería implica que, cuando un modelo detecta un fraude, los siguientes no se ejecutan. Para poder determinar el orden óptimo completo y no sólo el del primer modelo de la cañería, la optimización previa se resolvió mediante la lógica de un meta-modelo que va detectando iterativamente el orden.

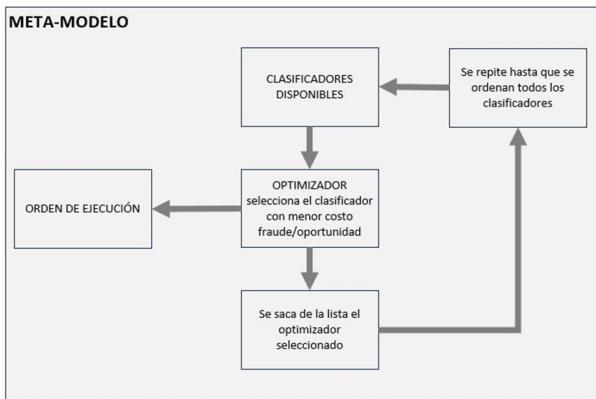


Figura 28: Lógica cálculo meta-modelo para optimización orden cañería ejecución modelos detección.

Fuente: Elaboración propia

El resultado de ejecución optimizado fue:

- Modelo MII**, Valor objetivo: 118279.64
- Modelo MIV**, Valor objetivo: 152443.64
- Modelo MIIIA**, Valor objetivo: 177738.82
- Modelo MIIIB**, Valor objetivo: 215298.95
- Modelo M0**, Valor objetivo: 218550.40
- Modelo MI**, Valor objetivo: 241357.44

El orden resultante es el esperado en función de las capacidades de detección, poniendo primero los con más altas capacidades de detección cruzada y los modelos más débiles al final.

III-I. Resultados y métricas de desempeño técnico e impacto económico

Para la evaluación final, se preparó un set sintético que reproduce las proporciones de tipos de fraude definidas en el EDA. Este set se ejecutó en la cañería completa, por lo que la evaluación integra las capacidades cruzadas del sistema. Los resultados fueron:

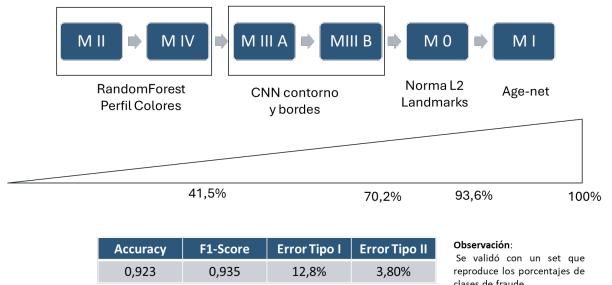


Figura 29: Resultados cañería ejecución optimizada.

Fuente: Elaboración propia

Al evaluar la capacidad de detección del modelo desarrollado y compararlo con la situación actual se tiene:

Costos anuales		
	Situación actual	Situación proyectada
Licencia Software	kUSD 300	-
Dotación (7 personas)	kUSD 67	kUSD 67
Tasa Error Tipo II	9 %	3,8 %
Costo Error Tipo II	MUSD 1	kUSD 420
TOTAL	kUSD 1.367	kUSD 487

Cuadro IV: Comparación de costos anuales entre la situación actual y la proyectada con modelo ML

Del análisis se observa que el modelo de machine learning generado permitiría prescindir del servicio de validación de imágenes actual (kUSD 25 mensuales), mantener el equipo actual y aprovechar que la estructura modular tiene una tasa de error de falsos negativos en 5,2 puntos porcentuales, es decir, bajar de 9 % a 3,8 % respectivamente y se pueden incorporar nuevos módulos sin necesidad de re-entrenar el

modelo completo.

De lo anterior se tiene un ahorro potencial anual del orden de kUSD 880.

IV. DISCUSIÓN

IV-A. Elementos distintivos

Este trabajo presenta cuatro puntos innovadores:

- **Metodología CRISP-DM Integrada con DPIA:** La integración de DPIA en la metodología CRISP-DM asegura el cumplimiento de las normativas de protección de datos, proporcionando un marco sólido para gestionar los riesgos de privacidad a lo largo del ciclo de vida del proyecto.
- **Uso de Datos Sintéticos:** Los datos sintéticos nos permitieron generar los grandes volúmenes necesarios para entrenar modelos equilibrados sin comprometer la privacidad. Las aleatorizaciones añadidas mejoran la capacidad de generalización de los modelos, lo que explica en parte los buenos resultados obtenidos.
- **Modelos Simples y Efectivos:** Utilizar modelos simples enfocados en las fortalezas de los algoritmos y las debilidades de cada intento de fraude ofrece una solución innovadora y efectiva. En lugar de depender de una única solución compleja, esta estrategia modular permite una adaptación y actualización más flexibles.
- **Pipeline Modular Secuencial:** Implementar un pipeline modular secuencial optimizado para los costos de error Tipo I y Tipo II permite integrar múltiples modelos especializados. Este enfoque mejora la precisión y la robustez del sistema de detección de fraude y facilita su mantenimiento y escalabilidad.

IV-B. Interpretación y Comparación de Resultados

Los resultados obtenidos superan los métodos actuales de la empresa. Los datos sintéticos generados, con las aleatorizaciones adecuadas, han demostrado ser efectivos para mejorar la generalización de los modelos. Sin embargo, los resultados en producción del sistema deberían estar más cerca de los números actuales de la empresa.

IV-C. Implicaciones y Relevancia

La metodología aborda un problema crítico de robo de identidad en el sector financiero, proponiendo una solución innovadora. Con la inminente emisión de una nueva Cédula de Identidad en Chile, una solución basada en un solo paso se habría vuelto obsoleta. Este enfoque metodológico permite una fácil modificación del sistema, utilizando los datos existentes para entrenar elementos adicionales del pipeline.

IV-D. Limitaciones y Futuras Investigaciones

Aunque los resultados son prometedores, el sistema debe ser validado con datos específicos del cliente para un ajuste adecuado. Investigaciones futuras podrían mejorar los algoritmos de generación de datos sintéticos para hacerlos aún más realistas y extender el enfoque modular a otras aplicaciones de verificación biométrica y detección de fraude.

V. CONCLUSIÓN

La transformación digital en los sectores minorista y de crédito ha impulsado la eficiencia operativa y la accesibilidad, pero también ha presentado nuevos desafíos en términos de seguridad y protección de datos personales. Este estudio demuestra que es posible desarrollar sistemas de detección de fraude precisos y respetuosos con la privacidad mediante el uso adecuado de datos sintéticos y modelos de machine learning optimizados.

La metodología propuesta establece un marco de referencia sólido para futuras investigaciones y aplicaciones en diversos sectores que enfrentan problemas similares de protección de datos y necesidades de modelos avanzados.

V-A. Principales Contribuciones

- **Resultado del modelo integrado:** La metodología desarrollada permitió obtener una precisión de 92,3 %, superando el objetivo inicial en 7,3 puntos porcentuales, lo que demuestra una mejora significativa en la capacidad de detección de fraudes.
- **Reducción de Errores Tipo II:** El modelo entregó una tasa de error tipo II de 3,8 %, reduciendo en 5,2 puntos porcentuales la tasa de error actual de la empresa, lo cual implica un ahorro económico considerable.
- **Cumplimiento Normativo:** La integración de la Evaluación de Impacto en la Protección de Datos (DPIA) dentro del marco CRISP-DM asegura el cumplimiento de las normativas legales y éticas, garantizando la privacidad de los datos personales.
- **Generación de Datos Sintéticos:** El uso de Redes Generativas Adversariales (GANs) para crear datos sintéticos permitió la generación de imágenes faciales realistas sin comprometer la privacidad, facilitando el entrenamiento de modelos de machine learning.
- **Implementación Modular:** La estructura modular del sistema permite una implementación más rápida y un menor costo computacional, facilitando la adaptación y el escalado del sistema para abordar nuevos tipos de fraudes.

Aunque los resultados son prometedores, el sistema debe ser validado con datos específicos del cliente para asegurar un ajuste óptimo. Futuras investigaciones podrían enfocarse en mejorar los algoritmos de generación de datos sintéticos para hacerlos aún más realistas y extender el enfoque modular a otras aplicaciones de verificación biométrica y detección de fraude.

AGRADECIMIENTOS

Los autores desean agradecer a:

- Sus familias, por su paciencia y constante apoyo, brindados durante todo el programa.
- A nuestro profesor guía, cuya enseñanza inspiró la idea de orientar este proyecto hacia una visión integral, combinando los enfoques técnico y normativo.
- A las empresas que auspiciaron este programa:
 - CODELCO
 - CAP S.A.
 - Dole Chile S.A.

REFERENCIAS

- [1] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015. [Online]. Available: <https://arxiv.org/abs/1409.1556>
- [2] C. Sun, A. Shrivastava, S. Singh, and A. Gupta, "Revisiting unreasonable effectiveness of data in deep learning era," 2017. [Online]. Available: <https://arxiv.org/abs/1707.02968>
- [3] G. Bae, M. de La Gorce, T. Baltrušaitis, C. Hewitt, D. Chen, J. Valentin, R. Cipolla, and J. Shen, "Digiface-1m: 1 million digital face images for face recognition," in *2023 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2023.
- [4] G. D. P. R. GDPR, "General data protection regulation," *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC*, 2016.
- [5] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014. [Online]. Available: <https://arxiv.org/abs/1406.2661>
- [6] F. T. Commission, "Consumer sentinel network data book 2023," <https://www.ftc.gov/reports/consumer-sentinel-network-data-book-2023>, Feb. 2024, [Accessed: Jul. 16, 2024].
- [7] M. Intelligence, "Identity verification market - growth, trends, and forecasts (2024 - 2029)," <https://www.mordorintelligence.com/industry-reports/identity-verification-market>, 2024, [Accessed: Jul. 16, 2024].
- [8] C. para el Mercado Financiero (CMF) Chile, "Resolución 566 exenta (06-abr-2024) m. de transportes y telecomunicaciones; subsecretaría de telecomunicaciones," <https://www.cmfchile.cl/portal/prensa/615/w3-article-79527.html>, 2024, [Accessed: Jul. 16, 2024].
- [9] A. Cavoukian *et al.*, "Privacy by design: The 7 foundational principles," *Information and privacy commissioner of Ontario, Canada*, vol. 5, p. 12, 2009.
- [10] B. del Congreso Nacional de Chile, "Resolución 566 exenta (06-abr-2024) m. de transportes y telecomunicaciones; subsecretaría de telecomunicaciones," <https://www.bcn.cl/leychile/navegar?idNorma=1202428&idVersion=2024-10-05>, 2024, [Accessed: Jul. 16, 2024].
- [11] U. Shafique and H. Qaiser, "A comparative study of data mining process models (kdd, crisp-dm and semma)," *International Journal of Innovation and Scientific Research*, vol. 12, no. 1, pp. 217–222, 2014.
- [12] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth, "The crisp-dm user guide," in *4th CRISP-DM SIG Workshop in Brussels in March*, vol. 1999. sn, 1999.
- [13] J. D. Kelleher and B. Tierney, *Data science*. MIT press, 2018.
- [14] Data Protection Commission, "Data protection impact assessments," <https://www.dataprotection.ie/en/organisations/know-your-obligations/data-protection-impact-assessments>, accessed: 2024-07-16.
- [15] D. B. Rubin, "Statistical disclosure limitation," *Journal of Official Statistics*, vol. 9, no. 2, pp. 461–468, 1993.
- [16] J. Drechsler, *Synthetic Datasets for Statistical Disclosure Control: Theory and Implementation*, ser. Lecture Notes in Statistics. New York, USA: Springer Science & Business Media, 2011, vol. 201.
- [17] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "stylegan2-ffhq," <https://github.com/NVlabs/stylegan2>, 2019.
- [18] Y. Yu, G. Kamran, W. HsiangTao, Y. Jiaolong, T. Xi, and F. Yun, "Expanding the latent space of stylegan for real face editing," 2022. [Online]. Available: <https://arxiv.org/abs/2204.12530>
- [19] Z. Zhu, Z. Lei, J. Yan, D. Yi, and S. Z. Li, "High-fidelity pose and expression normalization for face recognition in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 787–796.
- [20] D. E. King, "Dlib-ml: A machine learning toolkit," *Journal of Machine Learning Research*, vol. 10, no. Jul, pp. 1755–1758, 2009.
- [21] ——, "Dlib-ml: A machine learning toolkit," http://dlib.net/files/dlib_face_recognition_resnet_model_v1.dat.bz2, 2009, available at: http://dlib.net/files/dlib_face_recognition_resnet_model_v1.dat.bz2.
- [22] ——, "Dlib-ml: A machine learning toolkit," http://dlib.net/files/shape_predictor_68_face_landmarks.dat.bz2, 2009, available at: http://dlib.net/files/shape_predictor_68_face_landmarks.dat.bz2.
- [23] G. Guo, G. Mu, Y. Fu, and T. S. Huang, "Human age estimation using bio-inspired features," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 112–119.
- [24] M. Cuong, "Pytorch-age-estimation," <https://github.com/manhcuong02/Pytorch-Age-Estimation>, 2022.
- [25] A. Liaw and M. Wiener, "Classification and regression by randomforest," *Forest*, vol. 23, 11 2001.