



Tecnológico de Monterrey

Evidencia 1

Entendimiento del negocio y de los datos

CD3002C.600 Inteligencia artificial y analítica de datos
con impacto empresarial

21 de marzo del 2022

Equipo #3:

Carolina Velarde Díaz A01720509

Chantal Aimeé Simó García A00827554

Kízari Hernández Huerta A00828451

Ximena Araceli Martínez Flores A00829670

Docente

Profe. David Saucedo De La Fuente

Índice

1. Introducción	1
2. Segmento de mercado	1
2.1. Industria de banca y servicios financieros	1
3. Industria	1
3.1. Tendencias de la industria	1
3.2. Plataformas	1
3.3. Herramientas	2
3.4. Tecnologías	2
4. Cool Ideas	2
4.1. Ideas, propuestas e impacto general	2
4.1.1. Datos estructurados	3
4.1.1.1. Sistema de análisis de riesgo en la detección de fraude	3
4.1.1.1.1. Problemática y propuesta	3
4.1.1.1.2. Beneficios para los clientes de Softtek	4
4.1.1.1.3. Valor agregado y oportunidades de desarrollo	4
4.1.1.1.4. Glosario	4
4.1.2. Datos no estructurados	7
4.1.2.1. Chatbot identificador de mensajes fraudulentos (<i>Spam</i> o <i>Ham</i>)	7
4.1.2.1.1. Problemática y propuesta	7
4.1.2.1.2. Beneficios para los clientes de Softtek	8
4.1.2.1.3. Valor agregado y oportunidades de desarrollo	8
4.1.2.1.4. Glosario	8
5. Diagrama de Gantt	10
6. Conclusión	11
7. Bibliografía	12
8. Anexos	13
8.1. Análisis exploratorio de los datos	13
8.1.1. Datos estructurados	13
8.1.2. Datos no estructurados	13

1. Introducción

La Inteligencia Artificial (IA) es una herramienta que ha cobrado relevancia en los últimos años llevando así la analítica de datos al siguiente nivel para tomar mejores decisiones de negocio. Como parte del reto se busca brindar soluciones al socio formador **Softtek** basado en modelos de inteligencia artificial, aprendizaje automático y aprendizaje estadístico ante los desafíos y problemáticas de negocios que se presentan en la industria financiera.

Ahora, la **industria financiera** con frecuencia es blanco de diversas formas de delitos financieros y fraudes. Los escenarios han cambiado con el paso del tiempo y los actores maliciosos han actualizado sus tácticas para adaptarse mejor al mundo digital. Con esto en mente, a lo largo de este reporte se estará presentando el contexto y forma de desarrollo de las soluciones como propuesta de valor para Softtek y las entidades financieras.

2. Segmento de mercado

2.1. Industria de banca y servicios financieros

Para el reto se busca dar soluciones a las situaciones problemas y desafíos dentro de la industria de la banca y servicios financieros, específicamente, las instituciones financieras que requieran transformación en *FinTech* y soluciones bancarias de experiencia del cliente.

Buscamos mejorar la experiencia del usuario dentro del sector financiero al ofrecer servicios más seguros, confiables y personalizados que lo beneficiará y diferenciará con valor agregado.

3. Industria

3.1. Tendencias de la industria

- Banca digital: Ofrece una mejor experiencia al cliente con servicios más rápidos y eficientes.
- Pagos habilitados por voz: La voz se convierte en una nueva forma confiable para que las personas y las empresas realicen operaciones bancarias rutinarias.
- Plataformas de desintermediación: Pagos en línea, firmas digitales y disponibilidad en el *smartphone* de absolutamente todo.

3.2. Plataformas

- Infosys NIA: Ayuda a la automatización de procesos complejos, reduciendo los esfuerzos humanos y ahorrando tiempo para procesos comerciales redundantes.
- IBM Watson: Es capaz de aplicar procesamiento de lenguaje natural de nivel avanzado, información, recuperación, presentación de conocimiento, razonamiento automatizado y aprendizaje automático.

Evidencia 1. Entendimiento del negocio y de los datos

- TensorFlow: Se utiliza para diseñar modelos de aprendizaje profundo e implementa el cálculo numérico con la ayuda de un gráfico de flujo de datos, también conocido como gráfico computacional.

3.3. Herramientas

- Chatbots: Herramienta que pueden utilizar los bancos para brindar una experiencia personalizada a sus usuarios.
- Evaluación de riesgos automatizada: Herramienta que permite a los bancos evaluar los riesgos en la concesión de un préstamo sin tener que recurrir a otros intermediarios, logrando un ahorro importante de tiempo y dinero.
- Inteligencia de decisiones: Herramienta que incrementa la precisión en la autorización de transacciones auténticas en tiempo real, reduciendo el número de declinaciones falsas.

3.4. Tecnologías

- Procesamiento de lenguaje natural: Es un campo dentro de la inteligencia artificial y la lingüística aplicada que estudia las interacciones mediante el uso del lenguaje natural entre los seres humanos y las máquinas.
- Aprendizaje automático: Es una disciplina del campo de la inteligencia artificial que, a través de algoritmos, dota a los ordenadores de la capacidad de identificar patrones en datos masivos y elaborar predicciones.
- Big Data: Son datos que contienen una mayor variedad y que se presentan en volúmenes crecientes y a mayor velocidad.
- Inteligencia artificial: Es la habilidad de una máquina de presentar las mismas capacidades que los seres humanos, como el razonamiento, el aprendizaje, la creatividad y la capacidad de planear.

4. Cool Ideas

4.1. Ideas, propuestas e impacto general

La inteligencia artificial está llamada a ser la tecnología que transforme a la industria financiera, no sólo en términos de creación de nuevos productos y servicios, sino también en materia de funcionalidad y usabilidad, mejorando así la relación entre el cliente y el banco (*Onespan s.f.*). Utilizando la inteligencia artificial, se busca dar soluciones ante las principales problemáticas que vive esta industria como lo son los riesgos relacionados con la suplantación de identidad y otras formas de fraude financiero (*Rodríguez, 2018*).

Mediante la detección de eventos asociados con posibles ataques fraudulentos, estas soluciones pueden adoptar estrategias de evaluación de riesgos, seguidas de mecanismos de autenticación inteligentes y menos invasivos, que tienen como objetivo reducir el riesgo de

una transacción digital determinada, evitando que se realicen delitos o la exposición de datos sensibles.

Dicho esto, se considera que estas propuestas pueden impactar a la cartera de productos y servicios de Softtek, principalmente en sus servicios *Fintech* y la experiencia bancaria de sus clientes, puesto que representan una solución moderna ante los problemas modernos que vivimos hoy en día. Al guiarnos de las megatendencia del sector y entendiendo la evolución y transformación del entorno digital, estas ideas representarán una adición valiosa para mantenerse a la vanguardia ante las necesidades del mercado y del sector financiero.

4.1.1. Datos estructurados

4.1.1.1. Sistema de análisis de riesgo en la detección de fraude

4.1.1.1.1. Problemática y propuesta

Cada año hay más personas que utilizan la banca en línea o que gestionan sus finanzas por internet. En 2020, el bloqueo mundial debido al COVID-19 convenció a un número aún mayor de clientes de utilizar la banca en línea para al menos una parte de sus actividades financieras. El fraude en línea, que ya aumenta año tras año, ha seguido el mismo camino. El fraude financiero se ha vuelto particularmente popular, donde se ha disparado más del 280% entre el segundo trimestre de 2019 y el segundo trimestre de 2020 (*OneSpan, s.f*).

El fraude en línea evoluciona constantemente y a medida que los bancos ponen en marcha medidas de corrección, aparecen nuevas amenazas. Los sistemas tradicionales de prevención del fraude basados en reglas estáticas no pueden seguir el ritmo. Con esto en mente, se propone una solución que se basa en la creación de un sistema de análisis de riesgo en la detección de fraude, mismo que utilizará aprendizaje automático e inteligencia artificial para predecir, detectar y bloquear el fraude financiero y adaptarse a las nuevas amenazas, todo sin dañar la experiencia del usuario en el proceso (*OneSpan, s.f*).

Aplicando técnicas de aprendizaje automático con modelos supervisados y no supervisados se examinarán todos los datos pertinentes relativos a una transacción y se asignará una puntuación de riesgo a la misma. Con base a la puntuación de riesgo, el sistema recomendará permitir la transacción, bloquearla o solicitar una autenticación avanzada antes de autorizarla (*OneSpan, s.f*).

Se buscará que todo esto pueda hacerse en tiempo real con o sin intervención humana, lo que proporcionará a las entidades financieras una mayor prevención del fraude sin causar una fricción indebida en la sesión del cliente.

4.1.1.1.2. Beneficios para los clientes de Softtek

Algunos beneficios para los clientes de Softtek al implementar esta tecnología son:

- Las herramientas de inteligencia artificial permitirán encontrar patrones que sean sospechosos al momento de hacer alguna transacción de cualquier tipo.
- Se reduce la cantidad de falsas alarmas al detectar transacciones irregulares.
- Menos fricciones para los usuarios al momento de realizar transacciones.
- Disminución en la pérdida de usuarios debido a problemas al momento de interactuar y hacer operaciones en sitios de comercio electrónico o banca.

4.1.1.1.3. Valor agregado y oportunidades de desarrollo

- Soluciones precisas: Contar con estas aumenta los ingresos, puesto que los clientes obtienen experiencias personalizadas sin fricciones, las empresas aceptan más pedidos y detienen las pérdidas por fraude.
- Automatización de decisiones: Evaluación más precisa de los riesgos en tiempo real, pues el 25% de los consumidores no vuelve a un sitio de comercio electrónico si su transacción fue rechazada.
- Reducción en los costos de operación: Las empresas al no tener que gastar en revisiones manuales, ahora tendrán más tiempo para enfocarse en sus objetivos empresariales estratégicos.
- Mayor seguridad y confianza: Colaborar en el desafío de equilibrar los requisitos de cumplimiento, con la necesidad de proteger a los clientes contra las amenazas del fraude, todo esto sin comprometer la experiencia general del cliente.
- Servicios de asesoramiento financiero personalizado: Con los algoritmos de inteligencia artificial a desarrollar se brindará ayuda en el análisis de la cartera de los clientes para que se pueda recibir la información que se necesita lo antes posible, además de asesoramiento personalizado de su situación de fraude.

4.1.1.1.4. Glosario

Para el desarrollo de esta solución se estarán utilizando datos estructurados, mismos que tienen la característica de tener esquema determinado como los de una tabla organizada. Investigando los datos en fuentes secundarias se encontraron tres bases de datos que son de gran utilidad para la creación de los modelos de aprendizaje automático.

Las fuentes principales fueron *Kaggle* (plataforma de ciencia de datos) y *Opensim* (sistema de inteligencia artificial que simula datos financieros). Con los datos

Evidencia 1. Entendimiento del negocio y de los datos

recabados de estas fuentes se podrá entrenar un modelo supervisado presentándole tanto eventos fraudulentos como legítimos y ejecutándose para desarrollar un conjunto de instrucciones, buscando así que el modelo identifique patrones tanto desconocidos como conocidos para producir una puntuación de riesgo precisa para una transacción solicitada.

Variables demográficas			
#	Variables	Descripción	Tipo
1	Nombre y apellido	Nombre y apellido del cliente	Carácter
2	ID del cliente	Clave de identificación del cliente	Numérico
3	Nombre del vendedor	Nombre y apellido del vendedor	Carácter
4	ID del vendedor	Clave de identificación del vendedor	Numérico
5	Categoría del vendedor	Categoría de productos de venta del vendedor	Carácter
6	Género	Género del cliente	Carácter
7	Edad	Edad del cliente	Numérico
8	Calle	Calle donde vive el cliente	Carácter
9	Ciudad	Ciudad donde vive el cliente	Carácter
10	Estado	Estado donde vive el cliente	Carácter
11	País	País donde vive el cliente	Carácter
12	ZIP	Código postal del cliente	Numérico
13	Trabajo	Puesto de trabajo que ejerce el cliente	Carácter
Fuente: <i>Kaggle</i>			

Evidencia 1. Entendimiento del negocio y de los datos

Variables del historial de transacción			
#	Variables	Descripción	Unidades
1	Monto	Monto de la transacción en moneda local	Numérico
2	Tipo de transacción	Si la transacción fue <i>cash-in</i> , <i>cash-out</i> , débito, pagos o transferencia	Factor
3	Número de transacción	Número de transacción	Numérico
5	ID de la cuenta de transferencia	Identificación de la cuenta de transferir	Numérico
6	ID de la cuenta a transferir	Identificación de la cuenta a transferir	Numérico
7	**Considerado fraude o no**	Si la transacción fue considerada como fraude o no fraude	Factor
Fuente: <i>Kaggle</i> y <i>OpenSim</i>			

Variables del dispositivo usado			
#	Variables	Descripción	Unidades
1	<i>Sources</i>	Fuente utilizada para la compra	Carácter
2	<i>Browser</i>	Navegador web o buscador utilizada	Carácter
3	ID del dispositivo	Identificación del dispositivo	Numérico
4	Dirección IP	Dirección IP que identifica una red o dispositivo en Internet.	Numérico
5	Fecha y hora del <i>Log In</i>	Fecha y hora en la que se ingresó al portal	Date time
6	Fecha y hora de la compra	Fecha y hora en la que se ingresó a realizar la operación	Date time
Fuente: <i>Kaggle</i>			

4.1.2. Datos no estructurados

4.1.2.1. Chatbot identificador de mensajes fraudulentos (*Spam* o *Ham*)

4.1.2.1.1. Problemática y propuesta

El *spam* son mensajes de correo electrónico no solicitados que saturan las bandejas de entrada, estos constituyen un desafío para los usuarios de internet, para las empresas y también para los formuladores de políticas. Hay diferentes estimaciones, sin embargo, hay quienes sugieren que cada día se envían más de 100 mil millones de mensajes de *spam*, lo que representaría hasta un 85% del tráfico diario de correo electrónico a nivel mundial (*Internet Society, 2017*).

Dentro de una investigación realizada sobre el *spam*, se encontró que, a nivel global, se mandan 122.33 miles de millones de correos tipo *Spam* todos los días, los cuales se dividen de la siguiente manera:

- El 36% son relacionados con la mercadotecnia.
- El 31.7% son relacionados a contenido adulto
- El 26.5% son relacionados a temas financieros.

El *spam* generalmente son comunicaciones electrónicas no solicitadas que se envían indiscriminadamente y de manera masiva, ya sea manual o con el uso de *bots*. Si bien la actividad de *spam* toma mayormente la forma de mensajes de correo electrónico (*phishing*), no obstante, esta es una amenaza que ha evolucionado y se ha extendido a prácticamente todos los tipos de mensajes electrónicos, incluso a los mensajes SMS (*smishing*), a las publicaciones en los medios sociales, a los sistemas de mensajería instantánea (*bots*) y a los foros en línea. Más allá de la molestia y el tiempo que se pierde a causa de los mensajes no deseados, el *spam* puede causar daños significativos, infectando las computadoras de los usuarios con *software* malicioso capaz de dañar los sistemas y robar información personal (*Internet Society, 2017*). Esto ha causado que anualmente se pierdan alrededor de 20.5 miles de millones USD en las empresas (*dataprot, 2022*).

Entendiendo la problemática del *spam*, con el uso del procesamiento del lenguaje natural se identificará como varía el lenguaje usado en ambos medios de comunicación, para, así, poder detectar cuando un mensaje es *spam* o *ham* (correos que no son *spam*). Este sistema será incorporado en un *chatbot*, el cual podrá ser integrado en las plataformas de entidades financieras mediante el uso de *APIs*. El objetivo está basado en cuando un usuario reciba un mensaje con una oportunidad de inversión o de ganar un premio monetario, podrá reenviar el mensaje al *chatbot*, mismo que realizará la tarea de analiza el mensaje y su veracidad y, luego, podrá hacer una recomendación.

4.1.2.1.2. Beneficios para los clientes de Softtek

Conforme aumenta el uso de tecnología inteligente en la vida cotidiana, asimismo aumentará el cibercrimen y los esfuerzos de los cibercriminales en hacer que la población sea víctima de fraude.

Algunos beneficios para los clientes de Softtek al implementar esta tecnología son:

- Los clientes de Softtek tendrán un sistema específico para la protección de sus usuarios referente a tecnologías emergentes.
- Los clientes de Softtek gozarán de una reputación como una entidad financiera que protege y cuida a sus clientes, lo cual atraerá a nuevos usuarios.
- Será una importante fuente de ventaja competitiva puesto que les impedirá fuertes pérdidas millonarias que, seguramente, impactarán negativamente en el crecimiento de sus empresas.
- Los clientes de Softtek evitarán caídas de valores accionarios, afectaciones en la capacidad para capitalizarse en la Bolsa de Valores, menoscabo en el patrimonio total de sus empresas y desconfianza en los distintos mercados.

4.1.2.1.3. Valor agregado y oportunidades de desarrollo

- El chatbot puede ser incorporado a páginas web y aplicaciones móviles, lo que representa un fácil acceso para los usuarios.
- Este chatbot puede ser adaptado a diferentes tipos de estafa, como lo es el *vishing*.
- Debido a la alta reputación que esta medida de seguridad traerá, se espera que este beneficio atraiga a nuevos usuarios.

4.1.2.1.4. Glosario

Se cuenta con dos bases de datos, ambas se obtuvieron de la plataforma *Kaggle* y siguen el mismo formato; en la primera base de datos, la primera columna declara si el mensaje es *spam* o *ham* y la segunda columna contiene el mensaje. Por otro lado, la segunda base de datos es de correos electrónicos donde el *spam* es de contenido, incluyendo fraudes de 419 o del “príncipe nigeriano”.

Se considera que ambas bases de datos son relevantes para el desarrollo de la propuesta, esto debido a que ya cuentan con mensajes tanto catalogados como *spam* como *ham*. Con esto se busca que los modelos y el desarrollo del *chatbot* identifique patrones tanto desconocidos como conocidos para que esta lo pueda reconocer de manera precisa. Esto permitirá un análisis más completo, lo cual representa un producto de calidad para la cartera de productos de Softtek.

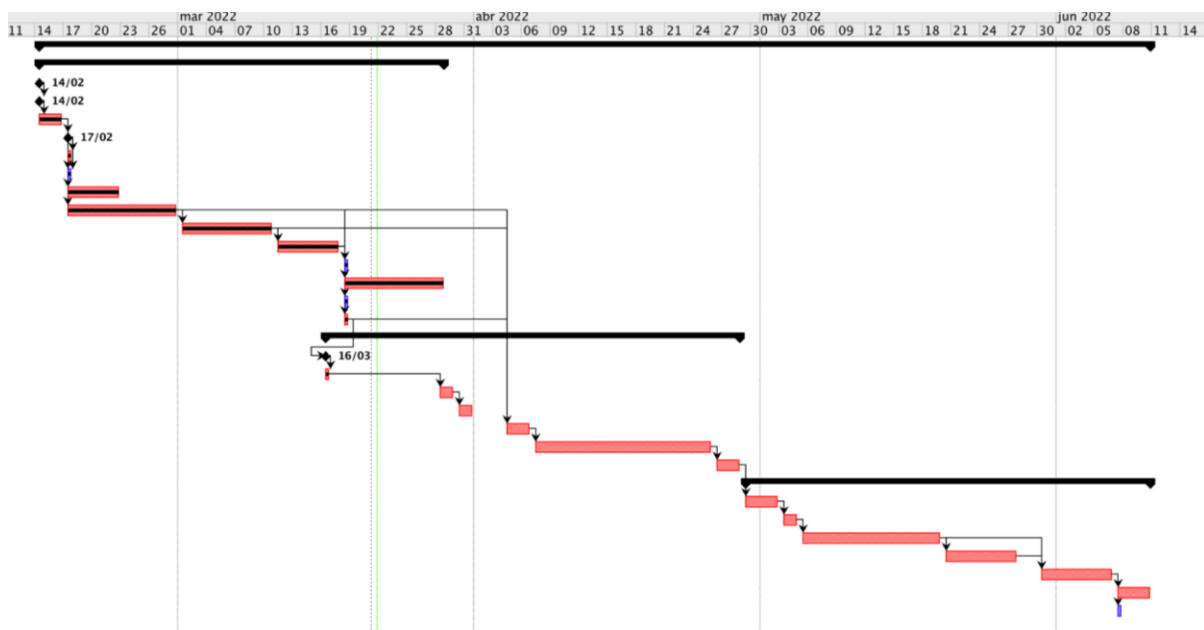
Evidencia 1. Entendimiento del negocio y de los datos

SMS			
#	Variables	Descripción	Tipo
1	**Spam o Ham**	Define si el mensaje es <i>spam</i> o <i>ham</i>	Factor
2	SMS	El mensaje a analizar	Carácter
Correo electrónico			
#	Variables	Descripción	Tipo
1	**Spam o Ham**	Define si el mensaje es <i>spam</i> o <i>ham</i>	Factor
2	<i>Email</i>	El mensaje a analizar	Carácter
Fuente: <i>Kaggle</i>			

5. Diagrama de Gantt

Para el desarrollo del proyecto se ha creado el siguiente diagrama de Gantt, el cual presentará todas las actividades claves que deben llevarse a cabo junto a su duración de días para el desarrollo de las soluciones. Esta guía brinda un claro lineamiento, de manera visual, de los componentes básicos de los proyectos y su organización en tareas más pequeñas y gestionables.

		Nombre	Duración	Inicio	Terminado	Predecesore
1		☐ P1. Proyecto solución basada en modelos de IA para una problemática de negocios	70 days	14/02/22 13:00	10/06/22 21:00	
2	☐	1. Entendimiento del negocio y de los datos	26 days	14/02/22 13:00	28/03/22 21:00	
3	✓	1.1. Comprensión del contexto y el segmento del negocio	0 days	14/02/22 13:00	14/02/22 13:00	
4	✓	1.2. Selección de industria y mercado meta	0 days	14/02/22 13:00	14/02/22 13:00	3
5	✓	1.3. Brainstorming de posibles ideas/soluciones	3 days	14/02/22 13:00	16/02/22 21:00	4
6	✓	1.4. Impactos de las soluciones para los clientes de Softtek	0 days	17/02/22 13:00	17/02/22 13:00	5
7	✓	1.5. Investigación de tendencias de plataformas, herramientas y tecnologías de IA	1 day	17/02/22 13:00	17/02/22 21:00	6
8	✓	1.6. Adición valiosa al portafolio de Softtek	1 day	17/02/22 13:00	17/02/22 21:00	5;6
9	✓	1.7. Propuesta de valor para la solución de la cool idea 1 y 2	4 days	17/02/22 13:00	22/02/22 21:00	5
10	✓	1.8. Investigación de fuentes para el glosario de las bases de datos est y no est	8 days	17/02/22 13:00	28/02/22 21:00	5
11	✓	1.9. Recolección de datos primarios	8 days	1/03/22 13:00	10/03/22 21:00	10
12	✓	1.10. Ideación concreta de las cool ideas 1 y 2	5 days	11/03/22 13:00	17/03/22 21:00	11
13	✓	1.11. Situación problema a solucionar	1 day	18/03/22 13:00	18/03/22 21:00	12
14	✓	1.12. Tendencias de plataformas, herramientas y tecnologías de IA que se utilizarán	2 days	18/03/22 13:00	28/03/22 21:00	12
15	✓	1.13. Propuesta de cool ideas 1 y 2	1 day	18/03/22 13:00	18/03/22 21:00	12
16	✓	1.14. Glosario de datos	1 day	18/03/22 13:00	18/03/22 21:00	10
17	☐	2. Preparación y modelados de datos	22 days	16/03/22 13:00	28/04/22 21:00	
18	✓	2.1. Enfoque al segmento de mercado	0 days	16/03/22 13:00	16/03/22 13:00	16
19	✓	2.2. Creación del value proposition canvas	1 day	16/03/22 13:00	16/03/22 21:00	18
20		2.3. Creación de marca y propuesta de valor	2 days	28/03/22 13:00	29/03/22 21:00	19
21		2.4. Investigación de plataformas para la propuesta (API)	2 days	30/03/22 13:00	31/03/22 21:00	20
22		2.5. Análisis exploratorio de datos	3 days	4/04/22 13:00	6/04/22 21:00	10;11;16
23		2.6. Limpieza e imputación de datos	8 days	7/04/22 13:00	25/04/22 21:00	22
24		2.7. Definición y prueba de modelos de clasificación y de agrupamiento	3 days	26/04/22 13:00	28/04/22 21:00	23
25	☐	3. Evaluación del modelo de analítica y despliegue del proyecto	26 days	29/04/22 13:00	10/06/22 21:00	
26		3.1. Definición del modelo a utilizar para el prototipo	2 days	29/04/22 13:00	2/05/22 21:00	24
27		3.2. Evaluación del desempeño final de los modelos	2 days	3/05/22 13:00	4/05/22 21:00	26
28		3.3. Generación del plan de implementación	6 days	5/05/22 13:00	19/05/22 21:00	27
29		3.4. Definición de la plataforma de integración del prototipo (API)	6 days	20/05/22 13:00	27/05/22 21:00	28
30		3.5. Definición del monitoreo y mantenimiento que el modelo requerirá	6 days	30/05/22 13:00	6/06/22 21:00	28;29
31		3.6. Generación de una presentación final	4 days	7/06/22 13:00	10/06/22 21:00	30
32		3.7. Creación de tres líneas de acción estratégica	1 day	7/06/22 13:00	7/06/22 21:00	30



6. Conclusión

Para concluir, los fraudes financieros han existido siempre, pero los avances tecnológicos han facilitado, en muchos casos, su alcance y éxito en los últimos tiempos. Dependiendo de la escala los efectos de un fraude financiero pueden ser muy graves, y afectar no solo a nivel interno en las empresas sino repercutir en los clientes y en el entorno social en el que actúan las compañías.

La prevención y la detección temprana de errores es el mejor camino para acabar con el fraude empresarial. Invertir en auditorías y en herramientas como las que se han mencionado es un freno para no entrar en un ciclo de infracciones que desemboque en desastres.

Dicho lo anterior, se podrán posicionar a los clientes de Softtek, que son las entidades financieras públicas y privadas, como agentes del cambio en cuanto a la prevención y protección de sus usuarios en contra de los ciberdelincuentes.

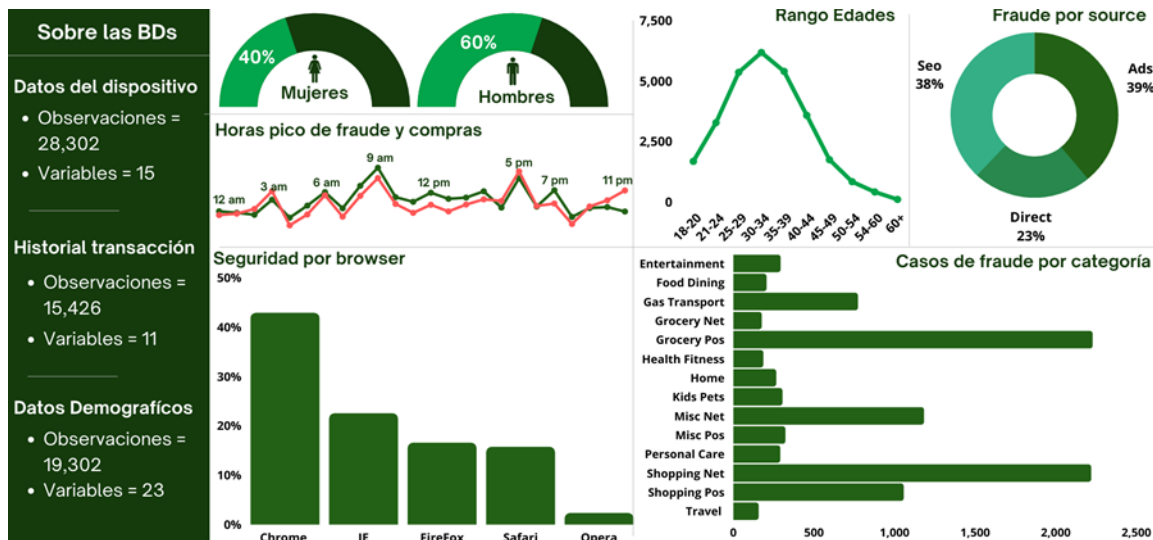
7. Bibliografía

- Astorga, P. (2020). *Procesamiento del Lenguaje Natural (PLN): qué es y para qué se utiliza*. Decide <https://decidesoluciones.es/procesamiento-del-lenguaje-natural-pln-o-nlp-que-es-y-para-que-se-utiliza/>
- Cveticanin, N. (2022). *What's On the Other Side of Your Inbox - 20 SPAM Statistics for 2022*. Dataprot. <https://dataprot.net/statistics/spam-statistics/>
- Deyan, G. (2022). *67+ Revealing Smartphone Statistics for 2022*. Techjury. <https://techjury.net/blog/smartphone-usage-statistics/#gref>
- Dreamer, F. (2021). *Ham v Spam: what's the difference?*. Journey Notes <https://blog.barracuda.com/2013/10/03/ham-v-spam-whats-the-difference/>
- FBI. (2016). *Nigerian Letter or "419" Fraud*. Federal Bureau of Investigation. <https://www.fbi.gov/scams-and-safety/common-scams-and-crimes/nigerian-letter-or-419-fraud>
- Iberdrola. (s.f.). *Descubre los principales beneficios del Machine Learning*. Iberdrola. <https://www.iberdrola.com/innovacion/machine-learning-aprendizaje-automatico>
- Internet Society. (2017). *Informe de políticas: El desafío del spam*. Internet Society. <https://www.internetsociety.org/es/policybriefs/spam/#:~:text=El%20t%C3%A9rmino%20spam%20generalmente%20se,de%20mensajes%20como%20correo%20basura.>
- Michael. (2020). *Security at the University of Portsmouth*. Security blog <https://securityblog.port.ac.uk/?p=1335>
- OneSpan. (s.f.). *Análisis del fraude*. OneSpan. <https://www.onespan.com/es/topics/analisis-del-fraude#:~:text=La%20anal%C3%ADtica%20del%20fraude%20es,actividad%20fraudulenta%20en%20tiempo%20real.C%20for%20confetti>
- Oracle. (s.f.). *¿Qué es el big data?*. Oracle Mx. <https://www.oracle.com/mx/big-data/what-is-big-data/>
- Parlamento Europeo. (2021). *¿Qué es la inteligencia artificial y cómo se usa?*. Sociedad. <https://www.europarl.europa.eu/news/es/headlines/society/20200827STO85804/que-es-la-inteligencia-artificial-y-como-se-usa>
- Quest, L. (2021). *Los riesgos y beneficios de usar la IA para detectar crímenes*. El Financiero. <https://www.elfinanciero.com/gerencia/innovacion/los-riesgos-y-beneficios-de-usar-la-ia-para/FABRSVOCUVAGVFGY6Y3RVHRIRO/story/>
- Rodríguez, A. (2021). *Inteligencia Artificial para detectar fraudes en empresas*. Blog Club del Emprendimiento. <https://www.clubdeemprendimiento.com/blog/pymes/inteligencia-artificial-detectar-fraudes-empresas/>
- Rodríguez, E. (2018). *4 ejemplos de cómo la inteligencia artificial está transformando al sector financiero*. blog.cobiscorp. <https://blog.cobiscorp.com/inteligencia-artificial-sector-financiero>
- 99firms. (2021). *How Many Email Users Are There in 2022* | 99firms. <https://99firms.com/blog/how-many-email-users-are-there/#gref>
- 99 firms. (2021). *Spam Statistics 2022* | 99firms. 99firms. <https://99firms.com/blog/spam-statistics/#gref>

8. Anexos

8.1. Análisis exploratorio de los datos

8.1.1. Datos estructurados



8.1.2. Datos no estructurados

[illegible]