



## Máster Business Intelligence y Big Data

# Inteligencia Financiera - IMEDIA

Proyecto fin de máster - 31 de Marzo, 2019

### Integrantes

Javier Marijuan Robledo

Raúl Andonie Amaya

Ricardo García Chavarry

Ricardo Valle Mayorga

### Tutor

Abel Pacheco Bernabéu

---

# Índice

<b>1. Planteamiento del problema</b>	<b>5</b>
1.1 Introducción	5
1.2 Contexto	5
1.3 Objetivo y oportunidad de negocio	6
1.4 Motivación y viabilidad de la propuesta	7
<b>2. Investigación y toma de datos</b>	<b>7</b>
2.1 Hipótesis a validar	7
2.2 Proceso de validación	8
2.3 Entrevistas realizadas	9
2.4 Conclusiones de las entrevistas	9
<b>3. Análisis DAFO</b>	<b>11</b>
3.1 Fortalezas	12
3.2 Debilidades	12
3.3 Oportunidades	12
3.4 Amenazas	13
<b>4. Plan estratégico</b>	<b>13</b>
4.1 Modelo de negocio	13
4.1.1 Propuesta de valor	14
4.1.2 Socios claves	14
4.1.3 Recursos claves	15
4.1.4 Relación con clientes	16
4.1.5 Segmentos de clientes	16
4.1.6 Canales de distribución	16
4.1.7 Actividades claves	16
4.1.8 Estructura de costes	17
4.1.9 Fuentes de ingresos	18
4.2 Misión, visión y objetivos	18
<b>5. Plan de acción</b>	<b>19</b>
5.1 Alcance del proyecto	19
5.2 Análisis de actividades y tareas	20
5.2.1 Mapa de procesos	20
5.2.2 Procesos estratégicos	20
5.2.3 Procesos de apoyo	21
5.2.4 Procesos clave	23

5.3 Cronograma	25
<b>6. Solución tecnológica</b>	<b>26</b>
6.1 Definición de la solución	26
6.2 Análisis de las fuentes de datos	27
6.2.1 Datos de las noticias	27
6.3 Modelo lógico	29
6.4 Arquitectura	30
6.4.1 Arquitectura Lambda	31
6.4.2 Arquitectura Kappa	32
6.5 Solución propuesta	33
6.5.1 Identificación de nuevas noticias con Change Data Capture (CDC)	36
6.5.2 Preprocesamiento con Kafka	37
6.5.3 Procesado de Flujo	39
6.5.4 Procesamiento del Lenguaje Natural (PLN)	40
6.5.4 Capa de Servicio	48
6.5.5 Capa de aplicación	53
6.5.6 Visualización de los datos	53
<b>7. Optimización de los resultados</b>	<b>59</b>
7.1. Indicadores	59
7.2. Análisis Financiero	62
7.2.1 Cuenta de resultados	62
7.2.2 Flujo de Caja	64
7.2.3 Balance de situación	65
7.2.4 Rentabilidad	65
<b>8. Conclusiones</b>	<b>66</b>
<b>9. Bibliografía</b>	<b>66</b>
<b>Anexos</b>	<b>70</b>
Entrevista a ASBANC	70
Entrevista a Financiera Ohh.	73
Entrevista a Banco de Crédito BCP	77

## Índice de figuras

Figura 1. Análisis DAFO	11
Figura 2. Flujo de trabajo	14

Figura 3. Socios Claves	15
Figura 4. Actividades Claves	17
Figura 5. Mapa de procesos	20
Figura 6. Estructura organizacional	23
Figura 7. Modelo de clasificación	24
Figura 8. Modelo lógico	30
Figura 9. Arquitectura Lambda	32
Figura 10. Arquitectura Kappa	33
Figura 11. Solución propuesta	35
Figura 12. CDC con Debezium	37
Figura 13. Funcionamiento de Kafka	38
Figura 14. Uso de Kafka	39
Figura 15. Publicación del diario Trome 23.03.2019	41
Figura 16. Listado de entidades obtenidos en meaningcloud.com	41
Figura 17. Listado de conceptos obtenidos a partir de meaningcloud.com	42
Figura 18. Listado de Clasificaciones encontradas según las taxonomías IAB e IPTC	43
Figura 19. Listado de polaridades de entidades	43
Figura 20. Análisis de sentimientos considerando personajes conocidos	44
Figura 21. Listado de polaridad de los conceptos	45
Figura 22. Resultado de API Formateado de JSON a CSV	47
Figura 23. Ranking de uso de motores de búsqueda	49
Figura 24. Interfaz de verificación manual	53
Figura 25. Ingreso de información adicional en relación entidad-evento	54
Figura 26. Buscador de personas	55

Figura 27. Resultados de la búsqueda de personas	56
Figura 28. Visualización de noticias con entidades	57
Figura 29. Notificación por correo	58
Figura 30. Proyección del Flujo de Caja Inteligencia Financiera IMEDIA	64

## Índice de tablas

Tabla 1- Estructura de costes	18
Tabla 2- Fuentes de ingresos	18
Tabla 3- Cronograma	26
Tabla 4- Campos de respuesta MeaningCloud	48
Tabla 5- Indicadores (KPI's)	59-60
Tabla 6- Ingresos previstos Inteligencia Financiera IMEDIA	63
Tabla 7- Costes previstos Inteligencia Financiera IMEDIA	63
Tabla 8- Cuadro de Amortización Inteligencia Financiera IMEDIA	63
Tabla 9- Cuenta de resultados Inteligencia Financiera IMEDIA	63
Tabla 10 -Flujo de caja Inteligencia Financiera IMEDIA	64
Tabla 11-Balance de situación Inteligencia Financiera IMEDIA	65
Tabla 12. Rentabilidad	66

# 1. Planteamiento del problema

## 1.1 Introducción

En el mundo actual nos encontramos ante una auténtica explosión de datos. Los factores claves que han originado esta situación son tecnológicos, como el crecimiento de la capacidad de cómputo y abaratamiento del hardware, y también sociales, como la digitalización y la transferencia tecnológica. Esta tendencia se ha manifestado también en el ámbito empresarial, con empresas que cada vez generan más datos y están en constante búsqueda de formas de aprovecharlos.

Ante estas necesidades, han aparecido nuevas tecnologías para el procesamiento y almacenamiento de los datos, así como técnicas para obtener más información a partir de los mismos. Esto ha tenido un gran impacto en la toma de decisiones de negocio para entidades de muy diferente índole (financieras, de seguros, sanitarias, asociadas al sector retail, etc.) con fines diversos como el abaratamiento de los costes o el incremento de los ingresos.

Nuestro proyecto hace referencia a la implementación de tecnologías Big Data, en **IMEDIA**. IMEDIA es una empresa de Perú que se dedica al monitoreo de los medios de comunicación tradicionales (prensa escrita, radio, televisión y diarios on-line), a través de la digitalización de los mismos para luego analizar de manera manual dicha información.

## 1.2 Contexto

Perú es uno de los países del continente sudamericano con mayor proyección de crecimiento fruto de una economía estable y una ubicación prodigiosa para el comercio e intercambio de bienes y servicios. En la actualidad, Perú forma parte de la Alianza del Pacífico junto a México, Colombia y Chile. Esta organización permite intercambiar servicios sin aranceles con el fin de dinamizar las economías de los estados miembros.

Aunque la economía crece, el país se encuentra envuelto en hechos tales como la corrupción, el narcotráfico, el lavado de dinero, la extorsión y otros delitos que la debilitan. Por otro lado, la geografía del país es muy accidentada y las poblaciones rurales son cada vez más dispersas, lo que ocasiona que los medios de comunicación locales y regionales adopten un papel fundamental en la lucha del crimen organizado.

Las leyes aparentan ser drásticas, sin embargo, los funcionarios públicos en muchas ocasiones se prestan a sobornos y manipulaciones de muy bajo nivel ético fomentando la existencia de periodistas que investigan los hechos ilícitos de personas y organizaciones de su localidad para difundirlas a través de los medios de comunicación locales y regionales.

En este entorno, los medios de comunicación ejercen un papel supervisor y de control en la sociedad peruana con la monitorización y el análisis de las noticias de cada localidad. Esta tarea que a primera vista parece ardua se ha simplificado con el uso de las tecnologías de Big Data y la inteligencia de negocios, permitiendo a las agencias que monitorean los medios de comunicación leer, procesar y analizar millones de noticias en escasos segundos favoreciendo el ahorro de costes y la optimización en la toma de decisiones.

Según la Contraloría General de la República, Perú pierde más de diez mil millones de dólares al año solo por delitos relacionados con la corrupción que a su vez están íntimamente relacionados con el lavado de dinero y es ahí, donde las entidades financieras se ven envueltas, pues poseen la responsabilidad de tener que conocer al tipo de clientes/organizaciones con las que operan, siendo multadas por el Estado si realizan operaciones con clientes involucrados en hechos delictivos, lo cual hace que tome especial relevancia los departamentos de riesgos de cada entidad financiera, que actualmente se encuentran muy comprometidos al carecer de información en línea que permita tomar decisiones para evitar comprometer las operaciones de la entidad a la que representan. Es por ello que elaborando una aplicación con la información subyacente al monitoreo tradicional o de gestión de la reputación y utilizando tecnologías emergentes se podría generar información relevante, oportuna y de gran valor para las entidades financieras y la sociedad en general.

### 1.3 Objetivo y oportunidad de negocio

La información del monitoreo de los medios de comunicación que posee **IMEDIA** representa un recurso valioso para identificar hechos delictivos y relacionarlos a personas u organizaciones, sin embargo, el análisis de estos datos digitalizados se realiza de forma manual, lo cual impacta en la disponibilidad de la información y en los costos de las actividades.

Mediante la aplicación de tecnologías **Big Data** se pretende procesar esta gran cantidad de noticias en segundos para posteriormente llevar a cabo el análisis, de forma automática, consiguiendo así un abaratamiento de los costes de la empresa y una disponibilidad casi inmediata de la información para su aprovechamiento.

La información obtenida como resultado del proceso anteriormente descrito, puede ser explotada con tecnologías de visualización y análisis de información (informes de diversa índole), que faciliten el uso de la plataforma a los usuarios, planteando una solución que actualmente no existe en el mercado de Perú.

Considerando el escenario planteado, de datos ya existentes y tecnologías disponibles, se pretenden lograr los siguientes objetivos:

- Crear un repositorio centralizado de información delictiva en medios de prensa de Perú.

- Permitir asociar dichos sucesos a personas e instituciones particulares mediante el acceso a las entidades de registro civil.
- Facilitar a las instituciones financieras el proceso de verificación y validación oportuna de entidades y particulares de Perú.

## 1.4 Motivación y viabilidad de la propuesta

En la actualidad, hasta 1 de cada 20 transacciones financieras de las entidades bancarias de Perú está ligada a personas o instituciones fraudulentas. Esto unido a que la geografía y demografía del país dificulta la existencia de medios de comunicación de carácter global hace que la información de zonas remotas no sea de fácil manejo para las entidades financieras, evitando la acertada toma de decisiones en cuanto a transacciones que involucren gente o instituciones de estas zonas.

Teniendo en cuenta que **IMEDIA** ya está realizando monitoreo de noticias en medios de comunicación de todo Perú para **Banco La Nación**, junto con la necesidad de mitigar el riesgo financiero, y de acuerdo con ellos, hemos detectado la oportunidad de mezclar nuestro trabajo con su necesidad, para utilizar el monitoreo de medios y cruzar la información con el **Registro Nacional de Identificación y Estado Civil (RENIEC)** y la **Superintendencia Nacional de Administración Tributaria (SUNAT)** para así poder ofrecer a la institución financiera una base de datos de personas y entidades potencialmente fraudulentas, basándonos en la información de los medios de comunicación tradicionales, que será monitorizados utilizando tecnología **Big Data** así como **Minería de texto**.

De esta manera la Entidad Financiera **Banco La Nación**, podrá tener información de calidad, la cual guiará a dicha entidad a tomar mejores decisiones y evitar pérdidas por transacciones fraudulentas.

Vemos mucho potencial en el proyecto, dado que no hay un servicio de este tipo actualmente y teniendo en cuenta nuestra experiencia en monitoreo, las herramientas de **Big Data** que hay en la actualidad, junto con la necesidad del cliente y que este tiene la intención de dar soporte a este proyecto.

Con esto pretendemos innovar en este campo, haciendo crecer a **IMEDIA** al mismo tiempo en su negocio.

## 2. Investigación y toma de datos

### 2.1 Hipótesis a validar

El propósito de las entrevistas es validar las siguientes hipótesis:



1. Existe una alta cantidad de transacciones financieras de carácter fraudulento en Perú. Concretamente se estima que hasta una de cada 20 lo son; lo cual supone un 5% del total.
2. La dificultad de acceso a algunas zonas rurales de Perú, hace que sea difícil que las noticias relativas a dichas zonas lleguen a ser conocidas en las grandes urbes del país.
3. Los medios de comunicación locales y regionales en el contexto de Perú adoptan un papel fundamental en la lucha contra el crimen organizado.
4. Las noticias de medios de comunicación tradicionales y digitales, contienen información valiosa y oportuna para la identificación de hechos delictivos. Además, dicha información representa una fuente valiosa para entidades financieras a la hora de evaluar a los clientes reales y potenciales.
5. El desconocimiento por parte de las entidades financieras de realizar operaciones bancarias (fundamentalmente otorgar créditos a clientes involucrados con hechos delictivos o guardar dinero de clientes involucrados en dichos hechos).
6. La predisposición de las entidades financieras a la utilización de herramientas tecnológicas con el fin de mejorar su toma de decisiones en cuanto a la realización de operaciones con clientes reales y potenciales.

## 2.2 Proceso de validación

Con el fin de validar las hipótesis planteadas previamente y recopilar información de valor para mejorar la calidad del proyecto, hemos optado por realizar entrevistas con entidades financieras con presencia sólida en Perú.

Dado que IMEDIA actualmente viene desempeñando la entrega de un informe sobre las personas u organizaciones involucradas con hechos delictivos a dos clientes, se decidió consultar con los mismos el apoyo en esta etapa. El primero es el Banco de la Nación y el segundo es ASBANC (Asociación de Bancos del Perú), ambos permitieron realizar la entrevista con la persona encargada de liderar el proyecto con IMEDIA en cada una de las entidades.

En el caso de ASBANC, el día 14 de enero de 2019 se realizó una reunión con la Sra. Patricia Mongrut Tello (Gerente de Coordinación Institucional) encargada de liderar el proyecto IMEDIA. Así mismo, el día siguiente se llevó a cabo una reunión con el Sr. Mirko Leon de Financiera Oh (Gestor de Riesgos Financieros). Ambas personas con su experiencia pueden brindar información confiable y de gran valor para validar las hipótesis e identificar las actuales necesidades de dichas instituciones.

Adicionalmente se logró agendar una reunión con el Sr. José Marangunich, Presidente del área de Seguridad Corporativa del Banco de Crédito del Perú, quien nos apoyó con la entrevista el día miércoles 30 de febrero de 2019 en la oficina ubicada en la Sede Central de dicho banco.

Es importante resaltar que debido a una serie de contratiempos e impedimentos debido a la temporada de vacaciones por verano en Perú es muy difícil pactar reuniones con altos ejecutivos impidiendo extender el número de entrevistas, a su vez ciertas entidades no son muy abiertas al momento de compartir información de sus procedimientos internos.

## 2.3 Entrevistas realizadas

Se realizaron tres entrevistas a las siguientes instituciones financieras:

- Asociación de Bancos del Perú
- Financiera Oh
- Banco de Crédito del Perú

El detalle de las mismas se incluye en el anexo de este documento.

## 2.4 Conclusiones de las entrevistas

A partir de la información obtenida en las entrevistas, podemos resumir lo siguiente:

- Para las entidades, es necesario delimitar el significado de fraude. Consideran importante incluir en este ámbito: clonaje de tarjetas de crédito, créditos sin abonos (First Payment Default), ocultamiento de efectivo (Kiting), falsificación, lavado de dinero, estafa, narcotráfico, robo, secuestros, asesinatos, raqueteo (asalto al paso cerca a las entidades financieras), marcaje, sicariato, ventas ilegales, tala ilegal y minería ilegal.
- El porcentaje de transacciones fraudulentas en las instituciones entrevistadas oscila entre 1.5% a 2%. Las pérdidas que representan depende del tamaño de la operación del banco, pero sólo con estos porcentajes de ocurrencia, equivalen a pérdidas millonarias cada año.
- Los procedimientos actuales para evitar fraudes incluyen: Auditorías, rotación laboral, control de acceso a información, procedimientos de ética, revisión de antecedentes de personas. Hay capacitaciones durante cada año coordinadas por la Superintendencia de Banca y Seguro (SBS).
- La SBS obtiene información a partir de los medios de comunicación sobre casos de corrupción de personas de alto perfil (políticos, personalidades, empresarios). Esta información es valiosa para las entidades financieras, sin embargo, no es completa para evaluar a personas y organizaciones al momento de ofrecer sus servicios.

- La Asociación de Bancos del Perú (ASBANC) utiliza informes diarios de IMEDIA los cuales brindan información de personas y organizaciones involucradas en hechos delictivos relevantes al sector financiero, a partir de la información en medios de comunicación.
- Hay desconocimiento parcial al momento de otorgar créditos de personas involucradas en hechos delictivos, debido a que ya se considera la información del poder judicial, de las centrales de riesgo (Equifax, Sentinel, SBS) y de los demás bancos. Sin embargo, no es posible determinar si el cliente cumplirá sus pagos. En el caso de la información judicial, sólo identifica clientes con procesos penales en curso, y las centrales de riesgo sólo incluyen a personas que ya han tenido impagos en su historial financiero. Actualmente no hay información para poder identificar a potenciales clientes fraudulentos, ni se consulta información de investigaciones en curso (Policía).
- En el particular de aceptar depósitos de clientes, hay mayor preocupación de las entidades financieras debido a que únicamente se evalúa al cliente mediante listas negras compartidas entre los bancos, datos de procesos judiciales en curso e información provista por la Superintendencia de Banca y Seguro (SBS), en este último, a como se mencionó anteriormente, solo se incluyen casos de corrupción de personas de alto perfil.
- Existen sanciones para las entidades financieras que reciben dinero de hechos ilícitos, estas son gestionadas a través de la Fiscalía Nacional de la República. Estos casos se dan por desconocimiento de la persona u organización que tramita con el banco.
- Las entidades entrevistadas tienen interés en un sistema que permita conocer a las personas u organizaciones involucradas en hechos delictivos como herramienta complementaria a las que ya utilizan en sus procedimientos, sin embargo, no contar con todos los datos de noticias de Perú en tiempo real es uno de los retos actuales para una solución de este tipo.

A partir del análisis del resumen anterior, podemos concluir lo siguiente:

- Las transacciones de entidades relacionadas con actividades fraudulentas oscilan entre el 1.5% y 2%, y representa pérdidas millonarias para las entidades financieras, tanto por capital no recuperado como por multas y sanciones impuestas.
- Actualmente existen herramientas para conocer entidades ó personas con historial impago ó involucradas en casos de corrupción de alto perfil, sin embargo, no se explota adecuadamente la información de medios de comunicación que permitiría conocer a tiempo entidades relacionadas a actividades fraudulentas.
- Hay interés por un sistema complementario a las herramientas actuales para evaluar personas o entidades relacionadas a hechos delictivos en tiempo real.

En resumen, las hipótesis propuestas han sido confirmadas por las entidades que nos permitieron realizar entrevistas. Consideramos importante haber confirmado el valor en la información de los medios de comunicación y la oportunidad existente para explotarla con el fin de complementar las

herramientas con que actualmente disponen las entidades financieras, por lo tanto, existe un alto potencial en una solución que permita acceder a dicha información en tiempo real.

### 3. Análisis DAFO

El análisis DAFO se centra en hacer un análisis de los factores internos al proyecto para reconocer fortalezas y debilidades, así como de hacer un análisis de los factores externos con el fin de analizar las amenazas y oportunidades a las que se expone el proyecto.

El objetivo del mismo es analizar interna y externamente todos los factores que pueden tener un impacto significativo en el proyecto a fin de tomar las mejores decisiones de negocio para que este siga adelante con la mayor rentabilidad posible. Es un análisis vital, que va a determinar las líneas de actuación clave (acciones) a emprender posteriormente para aumentar las posibilidades de éxito del mismo.

A continuación, mostramos un gráfico a modo de resumen, que posteriormente pasaremos a desarrollar.

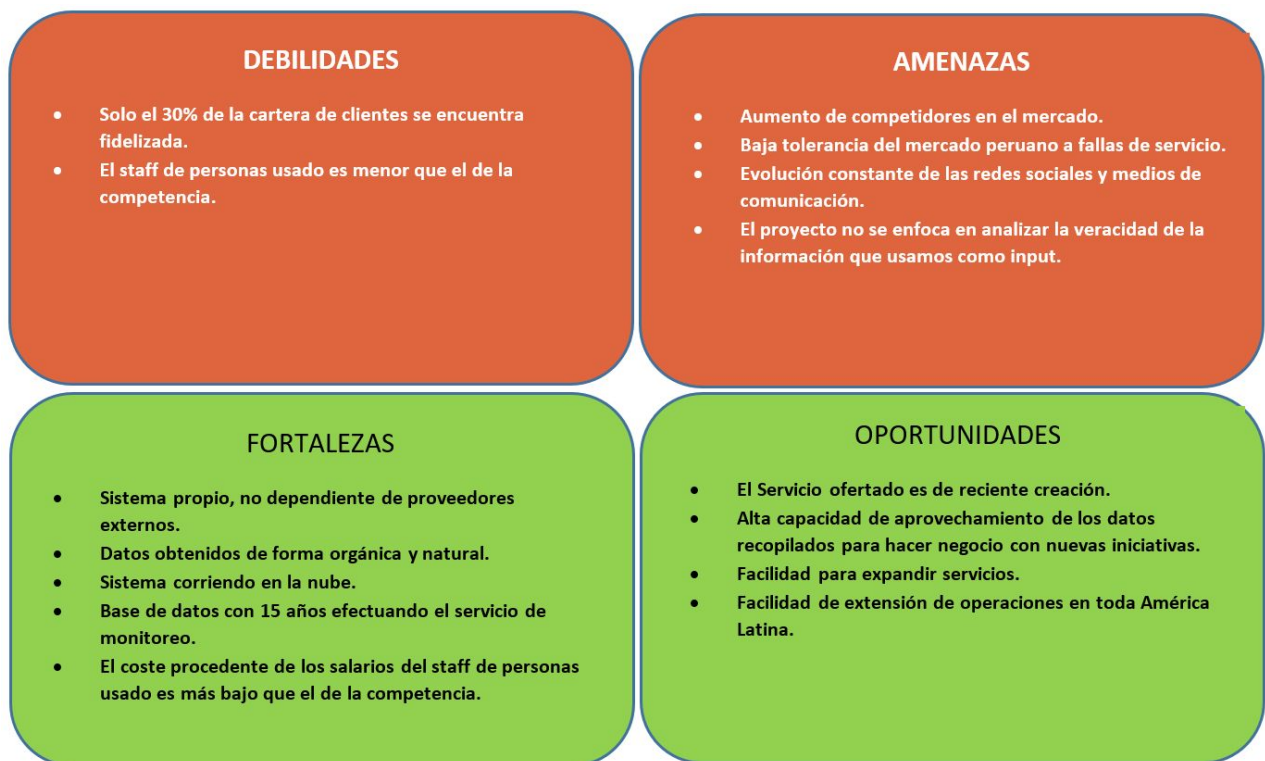


Figura 1. Análisis DAFO

### 3.1 Fortalezas

1. **Sistema propio, no dependiente de proveedores externos:** La empresa IMEDIA, trabaja con un sistema propio y autónomo, sin necesitar módulos de proveedores externos, lo cual facilita tener un control global y total del sistema evitando que este se vea afectado por factores que se escapan del control de la compañía.
2. **Datos obtenidos de forma orgánica y natural:** La obtención de los datos se realiza de forma puramente “natural” y “directa”; es decir, no es necesario de ninguna API para la obtención de los mismos, lo cual favorece y facilita la extracción de los mismos.
3. **Sistema corriendo en la nube:** Lo cual facilita la disponibilidad de los datos y mejora el rendimiento del proceso, obteniendo unos ratios de eficiencia (SLA's) de más del 99% de éxito.
4. **Base de datos con 15 años efectuando el servicio de monitoreo:** El sistema usado como Base de Datos es un sistema maduro con 15 años de uso. Lo cual permite que el sistema de monitorización de los datos y noticias sea robusto y funcione bien ante imprevistos.
5. **El coste procedente de los salarios del staff de personas usado es más bajo que el de la competencia:** Lo cual permite a la empresa ser más competitiva en costes respecto a la competencia.

### 3.2 Debilidades

1. **Solo el 30% de la cartera de clientes se encuentra fidelizada:** Lo cual nos dice que en este sector es difícil mantener relaciones comerciales estables con los clientes finales en el medio y largo plazo.
2. **El staff de personas usado es menor que el de la competencia:** Estimamos que el staff de personas de la empresa es entorno a un 20 % menor que el de la competencia, lo cual nos hace más frágiles frente a nuestros competidores ante posibles crisis comunicacionales con los clientes finales y también, ante nuevas peticiones/requerimientos solicitados por los clientes que demanden un nuevo servicio que debe estar disponible en poco plazo de tiempo.

### 3.3 Oportunidades

1. **El Servicio ofertado es de reciente creación:** Dada esta situación, en la que la solución ofertada no se encuentra con un alto grado de madurez en el mercado, nos hace pensar que el proyecto en sí es una nueva oportunidad para satisfacer una necesidad demandada en el mercado.
2. **Alta capacidad de aprovechamiento de los datos recopilados para hacer negocio con nuevas iniciativas:** Fruto de la implementación de sistemas expertos como el sistema de alerta, determinamos personas y entidades potencialmente relacionadas con hechos

delictivos, lo cual es un output muy aprovechable y útil para clientes finales de diferentes sectores, lo cual potencia y aumenta el volumen de negocio que la empresa puede desarrollar.

3. **Facilidad para expandir servicios:** A partir del sistema de monitorización utilizado actualmente, podemos ofrecer nuevos servicios en un tiempo razonable a los potenciales clientes.
4. **Facilidad de extensión de operaciones en toda América Latina:** El sistema utilizado actualmente para Perú es fácilmente escalable a otros países de América Latina, que poseen un contexto similar a Perú en cuanto al alto volumen de actividades delictivas desconocidas.

### 3.4 Amenazas

1. **Aumento de competidores en el mercado:** Recientemente han entrado al mercado nuevos competidores, tales como newsmonitor (Argentina), litoralpress (Chile), prnewswire (Brasil) y otros más.
2. **Baja tolerancia del mercado peruano a fallas de servicio:** lo cual ocasiona que si una entidad pública no está satisfecha te veta para licitar con nuevas entidades públicas.
3. **Evolución constante de las redes sociales y medios de comunicación:** Lo cual dificulta la obtención y el almacenamiento de las noticias procedentes de dichos medios.
4. **El proyecto no se enfoca en analizar la veracidad de la información que usamos como input:** El proyecto, como anteriormente hemos comentado, trata de obtener potenciales personas o instituciones relacionadas con conductas delictivas a partir de información obtenida de diversos medios de comunicación. Esto supone en sí mismo una amenaza al no analizar la veracidad de las noticias de los medios de comunicación, pudiendo reportar a nuestros clientes, personas o instituciones que realmente no están involucrados en hechos ilícitos.

## 4. Plan estratégico

### 4.1 Modelo de negocio

Inteligencia Financiera en IMEDIA, surge de la unión del proyecto final del “Máster en Business Intelligence y Big Data” y la necesidad de un cliente actual de la compañía en el área de Monitoreo de Medios.

En él, estaremos poniendo en práctica lo aprendido en el máster en un ámbito real, esperando innovar con un nuevo servicio inexistente en el mercado actual, basándonos en tecnología para apoyar la lucha contra delitos financieros.

Seguidamente, expondremos los ítems en los cuales se basará nuestra propuesta de negocio.

#### 4.1.1 Propuesta de valor

Nuestra propuesta surge con el objetivo de desarrollar un sistema funcional capaz de cruzar información de hechos delictivos en medios tradicionales ya sean en formato escrito o digital (no consideramos noticias que provengan de redes sociales) con datos del Registro Nacional de Identificación y Estado Civil (RENIEC) de Perú con el fin de proveer información de calidad a los departamentos de riesgo de las instituciones financieras y evitar que dichas instituciones operen con personas relacionadas con hechos fraudulentos.

Con esto esperamos prevenir problemas y lograr evitar pérdidas y evitar multas por este tipo de transacciones.

El flujo de nuestro trabajo sería:



Figura 2. Flujo de trabajo

#### 4.1.2 Socios claves

Aunque nuestro enfoque es para el sector financiero, siendo los potenciales clientes, bancos y otras entidades financieras, junto a la asociación que los agrupa, no solo estos son clave, sino también, las instituciones del registro civil y los órganos de seguridad del estado.

A continuación detallamos los principales:



Figura 3. Socios claves

- **Asbanc:** La Asociación de Bancos del Perú, fundada en 1967, es una institución gremial que agrupa a los bancos e instituciones financieras privadas de Perú y cuyo principal objetivo es promover el fortalecimiento del sistema financiero privado, proporcionando a sus asociados servicios de información.
- **RENIEC:** El Registro Nacional de Identificación y Estado Civil de Perú, es un organismo público autónomo que cuenta con personalidad jurídica de derecho público interno y goza de atribuciones en materia de registro, técnica, administrativa, económica y financiera. Es el organismo técnico encargado de la identificación de los ciudadanos peruanos, otorga el documento nacional de identidad y registra hechos vitales: nacimientos, matrimonios, defunciones, divorcios y otros que modifican el estado civil.
- **SBS:** La Superintendencia de Banca, Seguros y AFP es el organismo peruano encargado de la regulación y supervisión de los Sistemas Financiero, de Seguros y del Sistema Privado de Pensiones, así como de prevenir y detectar el lavado de activos y financiamiento del terrorismo. Su objetivo principal es preservar los intereses de los depositantes, de los asegurados y de los afiliados al Sistema Privado de Pensiones.
- **Banco de la Nación, Financiera Oh:** Son instituciones financieras peruanas, clientes potenciales con la necesidad de una solución tecnológica para tomar decisiones y evitar transacciones fraudulentas.

#### 4.1.3 Recursos claves

El grueso del material para nuestro proyecto, lo obtenemos de las noticias de los medios de comunicación, que seleccionamos acorde a nuestros criterios, para posteriormente utilizar



técnicas de minería de texto y de analítica, cruzando dicha información con el registro civil de personas y empresas, para obtener como producto final informes que nuestros clientes puedan utilizar en la toma de las decisiones más adecuadas para su negocio.

La información de medios es tanto escrita (impresa y digital), de radio y de televisión.

Requerimos de infraestructura para la extracción de la información, así como de plataformas avanzadas para analizarla y almacenarla y también de personal con experiencia en minería de textos y analítica.

#### 4.1.4 Relación con clientes

Dado que el origen de nuestra solución tiene a un cliente involucrado, estamos dirigiéndola a clientes que son Instituciones Financieras.

El producto a comercializar, es información filtrada, para que los clientes tengan soporte al momento de hacer negocios con personas e instituciones, previniendo así transacciones fraudulentas.

Puesto que la información de origen se genera proporcionalmente al transcurso del tiempo, nuestro producto puede asegurar la continuidad y la regularidad en el negocio.

#### 4.1.5 Segmentos de clientes

Los servicios de IMEDIA, desarrollados en este proyecto, son ofrecidos a instituciones pertenecientes al sector financiero.

#### 4.1.6 Canales de distribución

Puesto que estamos orientando nuestro servicio a las instituciones financieras. Lo que debemos hacer en primer lugar es escuchar sus necesidades, es por ello que, realizaremos visitas personales a nuestros clientes donde explicaremos nuestra propuesta, y la solución que ofertamos, que una vez contratada, será distribuida mediante canales digitales.

Nuestra solución, será promocionada a través de la web, así como en la asistencia a eventos promocionados por el sector financiero de Perú.

#### 4.1.7 Actividades claves

Las principales actividades de Inteligencia Financiera de IMEDIA girarán en torno a:

1. Definición de nuestra solución: Dado que la información en medios es bastante extensa, y de la cual se pueden obtener una gran cantidad de datos, estamos definiendo qué porción de esa información será usada en nuestra solución.

2. Alianzas clave: Además de tener segmentada la información de los medios, es muy importante establecer con quién vamos a cruzar la información para confirmar la veracidad de la misma.
3. Escoger la tecnología más adecuada: Es muy importante estudiar y elegir la tecnología que vamos a implementar tanto para la obtención de datos en el menor tiempo posible, como para procesarlos, analizarlos y cruzarlos con las diferentes bases de datos que tendremos de nuestros aliados.
4. Definición de Indicadores clave: para poder medir cuánto cumple las expectativas de nuestros clientes la solución ofrecida, y enfocarse así en los aspectos principales en los que prestar atención.
5. Desarrollar el producto: después de los pasos previos, nos ponemos en marcha para desarrollar e implementar lo que hemos planificado, con el objetivo de evaluar de forma ágil si la solución que ofrecemos soluciona las necesidad del cliente que la contrata, siguiendo estos pasos:



Figura 4. Actividades claves

6. Investigación + Desarrollo: Es necesario que la solución se actualice en el tiempo, para estar a la vanguardia de las últimas tendencias en tecnología en pro de mantener nuestro solución.

#### 4.1.8 Estructura de costes

Esta estructura está basada principalmente en recursos humanos y tecnológicos junto a la clásica infraestructura física y de servicios.

Hay que remarcar, que estos costes representan únicamente los destinados al proyecto Inteligencia Financiera de IMEDIA, y no representan al total de la compañía:

Concepto	Importe anual	Comentarios
Renta de oficinas	6.000 €	Espacio adicional para el Proyecto
Sueldos y Salarios	6.500 €	Personal de IMEDIA
Asesoría Contable y Fiscal	2.000 €	
Comunicaciones	5.000 €	
Cloud usage	8.000 €	
Hardware	3.000 €	<b>Como inversión, solo primer año</b>
Servidores AWS	6.000 €	Para procesamiento de nuestra solución
Líder de proyecto	10.000,00 €	Contrato por los primeros 6 meses del Proyecto
Programador Back-End	6.000,00 €	
Programador Front-End	12.000,00 €	
Diseñador web	5.000,00 €	
Computadoras	8.500,00 €	
<b>Total Projectado</b>	<b>78.000,00 €</b>	

Tabla 1. Estructura de costes

#### 4.1.9 Fuentes de ingresos

La única fuente de ingresos proyectada, será la venta del propio servicio, ya que nuestra solución está claramente definida.

El nicho de mercado son en las instituciones del sistema financiero peruano, quince de las cuales ya son clientes nuestros en otro servicios lo que nos ayudará en el rollout del proyecto, proyectamos a dos de ellas como inicio en el primer año para la Ingeniería Financiera y crecer anualmente con dos más.

Pronosticamos en el primer año estos ingresos:

Concepto	Importe anual	Comentarios
Venta de Producto a Cliente 1	30.000,00 €	2.500 EUR por mes
Venta de Producto a Cliente 2	30.000,00 €	2.500 EUR por mes
<b>Total Projectado</b>	<b>60.000,00 €</b>	

Tabla 2. Fuentes de ingresos

## 4.2 Misión, visión y objetivos

### Nuestra misión

Proveer un Servicio innovador y exclusivo basado en tecnología Big Data que facilite la lucha contra los delitos financieros en Perú.

### **Nuestra visión**

Que IMEDIA sea visto como un referente en la prestación de servicios basados en Big Data manteniéndose siempre a la vanguardia de la tecnología y apoyando al desarrollo de la sociedad peruana.

### **Objetivos**

1. Implementar fases y tecnologías Big Data en la definición de los servicios para proveer soluciones inteligentes y prácticas a los clientes.
2. Forjar unas sólidas relaciones en la infraestructura del servicio durante el primer año del proyecto con el cliente con el que ya trabajamos.
3. Expandir el servicio a las demás instituciones financieras en los años siguientes confiando en el éxito con el cliente actual.

## **5. Plan de acción**

### **5.1 Alcance del proyecto**

Con la finalidad de poder ofrecer una solución de software que permita monitorear personas u organizaciones se plantea tener una extensión del producto de análisis de reputación y marcas, ya existente en IMEDIA con el objetivo de crear con los mismos datos un sistema especializado en detección de hechos delictivos.

De esta manera los clientes podrán tener una base de personas y organizaciones comprometidas con hechos delictivos que podrían comprometer los activos financieros.

De igual manera también se plantea enviar alertas en tiempo real que sometería a una revisión cotidiana de eventos delictivos de suma importancia para las operaciones de las entidades financieras.

La información estará disponible en la nube otorgándole al cliente los accesos necesarios para ingresar al contenido que suma valor para su organización.

Las alertas serán enviadas por dos mecanismos. El primer mecanismo será vía correo electrónico y el segundo a través de WhatsApp. Estos dos mecanismos de envío se encuentran operativos en los productos base de la compañía y se reutilizará la funcionalidad existente.

La integración con los sistemas de información de las organizaciones financieras sería a través de API restful. El flujo de información sería entregado a través de una clave que permitiría el acceso a la base de datos de IMEDIA y con esto a los datos en tiempo real.

## 5.2 Análisis de actividades y tareas

### 5.2.1 Mapa de procesos



Figura 5. Mapa de procesos

### 5.2.2 Procesos estratégicos

#### Dirección estratégica

Se buscará especializar para el sector bancario y financiero el producto que IMEDIA viene comercializando y así darle mayor utilidad a la solución de monitoreo de reputación y análisis de influencia en los medios de comunicación que desempeña actualmente la compañía.

#### Dirección comercial

IMEDIA actualmente cuenta con más de 15 clientes del sector financiero que podrían optar por adquirir una solución financiera que ayude a otorgar fiabilidad en la entrega de créditos. Así mismo controlar el flujo de capitales de mala procedencia.

El tamaño de mercado consta de más de 50 organizaciones, entre bancos, financieras, cajas municipales, cajas rurales y fondos de inversión.

Partiremos por colocar el producto en la cartera existente de IMEDIA para posteriormente ir a otras entidades financieras.

Se rediseñará la página web de IMEDIA con el objetivo de exponer la vertical financiera desarrollada para contribuir a los indicadores de riesgo crediticio de nuestros clientes. De esta manera los clientes no solo tendrán datos obtenidos del sistema financiero o de casas de riesgo sino también la prensa diaria proporcionará información en tiempo real a los sistemas financieros de los bancos.

Por ser un producto focalizado, con un mercado plenamente especializado, no optaremos por realizar campañas en redes sociales. Sin embargo, si haremos una campaña de marketing email a personas o stakeholders de entidades financieras que puedan contribuir o ser un factor de decisión para la compra de nuestro producto.

Por último, asistirémos a congresos y eventos públicos del mercado financiero como por ejemplo el CADE (Congreso Anual de Ejecutivos) que reúne a casi todas las entidades financieras del sistema bancario nacional.

### **Tecnologías de información**

Las tecnologías aplicadas serán constantemente revisadas de manera que el producto e índice de fiabilidad de encontrar coincidencias respecto a la identificación más inteligente de hechos delictivos será crucial en la determinación y mejora del producto final.

#### **6.2.3 Procesos de apoyo**

##### **Gestión financiera**

Este proceso está destinado a dirigir los flujos de dinero entrante de la organización sobre la mejoría de los servicios de esta vertical de manera que se garantice la continuidad y perfección de los procesos tecnológicos que conlleva la realización y entrega de este nuevo producto.

En lo que se refiere a la estructura de costes, el servicio está basado en costes marginales en su mayoría y en la medida que los clientes ingresen se producirán gastos por almacenamiento y procesamiento de información.

Por otro lado, la junta de accionistas de la compañía acordó aportar 10.000 euros con la finalidad de cubrir los costes fijos de despliegue e implantación del proyecto.

Es importante considerar que no se contará con financiamiento externo ya que como se explicó en apartados anteriores el producto será una extensión de los servicios ya provistos por la compañía.

##### **Gestión de Proveedores**

Es de suma importancia contar con servicios especializados que nos permita escalar la solución sin necesidad de preocuparnos de la infraestructura, investigación y desarrollo de nueva

tecnología asociada a la comprensión natural del lenguaje. Es por ello que decidimos externalizar ciertos servicios de tecnología que nos permitan tener un alto nivel de acuerdo de servicio de manera que esto se refleje en la calidad de producto final.

Para esto se requiere contar con servicios de tecnología como los siguientes:

- Proveedores de plataforma cloud. (IBM, Amazon AWS)
- Proveedores de NLP (MeaningCloud, Watson NLU)
- Proveedores de suministros complementarios (Internet: Optical Networks / Telefónica del Perú, Energía Eléctrica: Luz del sur.)
- Asesoría legal y financiera.

Los operadores actuales cuentan con una enorme confianza y un nivel de calidad bastante elevado. Sin embargo, dado que los cobros son bajo demanda, nos permitirá balancear o tener dos proveedores brindando alta disponibilidad en la entrega del producto final.

### **Gestión de personal**

Todos conocemos que el recurso principal de una organización no es su tecnología, sino que son las personas quienes hacen diferente a una empresa. La empresa IMEDIA es una empresa fundada hace 15 años con gran experiencia en la industria del monitoreo de medios y que busca constantemente la mejora de sus procesos productivos con miras a obtener la excelencia operacional. Esto no es nada fácil pues involucra el reclutamiento, selección y fidelización del personal contratado.

El proyecto se desempeña en un entorno en donde la empresa puede aprovechar los recursos humanos existentes con la finalidad de proveer nuevos servicios a partir de la matriz de análisis de reputación y marcas.

Esto conlleva el tener una ventaja competitiva frente a organizaciones que deseen efectuar lo mismo, lo que permite a IMEDIA sobreponer sus debilidades con los recursos humanos disponibles.

La estructura organizacional de IMEDIA está compuesta por 4 divisiones o Departamentos quienes reportan al director general o CEO.

Además, que la organización cuente con una estructura organizacional horizontal. La organización tiene a cada miembro de la dirección con especializaciones en sus respectivas materias.

A continuación, se muestra un gráfico con la estructura organizacional de la empresa:



Figura 6. Estructura organizacional

Las funciones de cada departamento son:

- Departamento de Finanzas y Administración: Velar por el correcto uso de los recursos financieros. Administrar adecuadamente el patrimonio y los recursos humanos. Gestionar las obligaciones tributarias.
- Departamento de Tecnologías de Información: Desarrollar y emprender proyectos de tecnologías de información. Proporcionar el mantenimiento de las aplicaciones y sistemas de información existentes. Salvaguardar los activos digitales, promover el uso de las TIC en toda la organización.
- Departamento Comercial: Fidelizar a clientes existentes. Encontrar nuevas oportunidades de negocio. Promover líneas de negocio. Capacitar y entrenar a los clientes en el uso de los sistemas de información de la empresa.
- Departamento de Operaciones: Gestionar el comportamiento de las aplicaciones en función a los requerimientos del cliente. Controlar la ingesta de datos. Promover y dar calidad a la entrega de información final. Supervisar y elaborar los reportes analíticos.

#### 5.2.4 Procesos clave

##### Análisis

- **Estudio de los requerimientos del cliente.** Durante esta fase se analiza los hechos delictivos que el cliente desea monitorear de manera que el sistema pueda ser capaz de escuchar, identificar y comprender todas las características relacionadas al delito y si este está realmente relacionado a la persona o solamente se pretende de un hecho aislado sin repercusión en algún individuo.
- **Fuentes de datos.** En esta etapa se pone a disposición otras entradas como:
  - Medios de comunicación online locales y regionales.
  - RENIEC - Registro Nacional de Identidad.
  - Poder Judicial - Expedientes judiciales y penales.



- SUNAT – Consulta de RUC. Registro Único de Contribuyente.
- **Modelo lingüístico.** Durante este proceso seremos capaces de bosquejar los modelos de clasificación de texto para poder ser capaces de comprender personas, organizaciones y hechos delictivos.

## Diseño

- **Modelo de clasificación.** En esta etapa y con los requerimientos del cliente somos capaces de entrenar y determinar los diferentes modelos de clasificación.

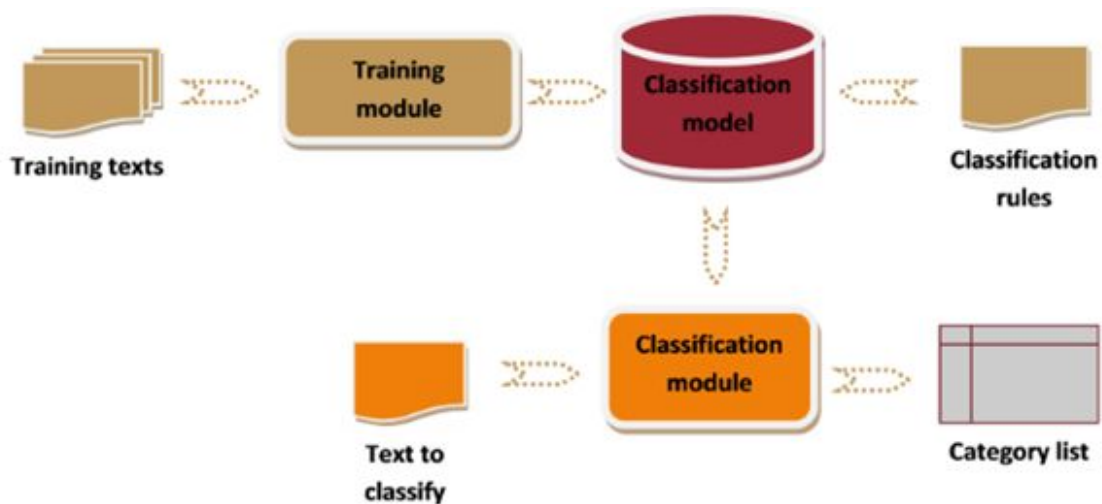


Figura 7. Modelo de clasificación (Fuente: MeaningCloud.com)

- **Definición de indicadores.** Durante este proceso logramos evaluar el nivel de certeza del motor. Determinando con ello el grado de fiabilidad de la herramienta en los contextos de cada tipo de medio.
- **Arquitectura técnica.** Durante esta etapa se identifican las instancias a contratar en los proveedores de servicio de plataforma.

## Implementación

- **Desarrollo de los modelos de clasificación.** Durante esta etapa se generarán los modelos de clasificación previamente establecidas en el ciclo de diseño. Con ello el sistema es capaz de discernir y adoptar las reglas de comportamiento.
- **Integración con soluciones externas.** En este proceso llevaremos a cabo la conexión con los sistemas externos de RENIEC, SUNAT y Poder Judicial.
- **Desarrollo de interfaces.** Se implementará una serie de formularios que permitan a los analistas clasificar la información en donde se encuentren homonimia y tratamiento de datos en conflicto.

## Validación y despliegue

- Pruebas del sistema. En este proceso llevaremos a cabo la fase de pruebas del prototipo. Llevaremos a cabo un ciclo integral de pruebas desde la recepción de datos, transformación y entrega de los mismos.
- Puesta en marcha. Una vez cumplida la etapa de pruebas la solución se desplegará en un entorno que cumpla las expectativas del cliente final.

## Mantenimiento

- **Formación de usuarios.** Durante esta etapa se elaborarán todos los instructivos y guías de usuario necesarias para la correcta configuración y uso del sistema.
- **Soporte a usuarios.** Utilizaremos la mesa de ayuda existente de la compañía. Dicha plataforma cuenta con dos mecanismos de ayuda. El primero es a través de correos electrónicos y el otro es a través del sistema de mesa de ayuda <http://pedidos.imedia.pe>
- **Seguimiento y mantenimiento de la aplicación.** La compañía tiene como practica la visita mensual o trimestral por parte de la ejecutiva de cuenta de manera que pueda resolver los inconvenientes que presenten o revelar información importante para el desenvolvimiento del sistema de información.

## 5.3 Cronograma

El proyecto tendrá una duración aproximada de 14 semanas. Sin embargo, esta duración está sujeta a las necesidades del cliente y a la complejidad con la que se desee someter a comparaciones u otras fuentes de datos externas.

El cronograma propuesto será el siguiente:

PROYECTO: IMEDIA		MES 1				MES 2				MES 3				MES 4				MES 5			
ID	MÓDULO	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4
<b>1</b>	<b>ANÁLISIS</b>	■	■	■	■																
1.1.	Requerimientos del cliente	■	■																		
1.2.	Fuentes de datos			■	■																
1.3.	Modelo de clasificación			■	■																
<b>2</b>	<b>DISEÑO</b>					■	■	■													
2.1.	Modelo de clasificación					■	■	■													
2.2.	Definición de indicadores clave					■															
2.3.	Arquitectura Técnica						■	■													
<b>3</b>	<b>IMPLEMENTACIÓN</b>								■	■	■	■									
3.1.	Desarrollo de Modelo de clasificación								■	■											
3.2.	Integración con soluciones externas								■	■											
3.3.	Desarrollo de interfaces									■	■	■									
<b>4</b>	<b>DESPLIEGUE</b>											■	■	■							
4.1.	Pruebas del sistemas											■	■								
4.2.	Puesta en marcha													■							
<b>5</b>	<b>MANTENIMIENTO</b>																				
5.1.	Formación de usuarios															■					
5.2.	Soporte a usuarios																■	■	■	■	
5.3.	Seguimiento y mantenimiento																	■	■	■	

Tabla 3. Cronograma

## 6. Solución tecnológica

### 6.1 Definición de la solución

El proyecto de Inteligencia Financiera de IMEDIA tiene el objetivo de automatizar el proceso de identificar hechos delictivos, para posteriormente, clasificarlos y relacionarlos a personas e instituciones a partir de los datos recopilados de noticias de medios tradicionales y digitales de Perú (input). La información obtenida a partir de dicho procesamiento se utilizará para crear una solución que permitirá consultar datos históricos y en tiempo real de sucesos delictivos incluyendo la relación de personas e instituciones con los mismos, con el fin de ser explotados por los departamentos de riesgos de las entidades financieras para evaluar tanto a sus clientes actuales, así como a potenciales clientes a los que pueden ofrecer sus servicios.

IMEDIA actualmente cuenta con procesos automáticos y manuales para la digitalización de las noticias de Perú; sin embargo el proceso de identificación, clasificación y relación de las noticias con las personas e instituciones se realiza de forma manual. La solución propuesta persigue automatizar esta tarea para disminuir el tiempo entre la recopilación de noticias y la información que requieren las entidades financieras, además de permitir escalar estos procesos a más clientes y más fuentes de información.

La base de esta solución es la gran cantidad de noticias procedente de distintos medios de Perú (que ya posee IMEDIA) que serán tratadas mediante algoritmos de procesamiento de lenguaje natural y almacenados en estructuras que permiten transmitir la información en tiempo real y realizar consultas de datos históricos desde una plataforma que debe mostrar la información de forma clara y acompañada de gráficos, dashboards y estadísticas de interés.

## 6.2 Análisis de las fuentes de datos

Los medios de comunicación son repositorios valiosos de datos de muy diversa índole, tanto por las distintas fuentes de información existentes así como por el tipo de información que contienen. IMEDIA actualmente realiza la digitalización de las noticias de los siguientes medios:

- Radio
- Noticias de televisión
- Prensa escrita
- Prensa digital

Este proceso de digitalización es constante y utiliza tecnologías de web scraping, RSS, speech-to-text, OCR y APIs de redes sociales, así como de procesos manuales de control de calidad y verificación de datos. El resultado de todo ello son dos bases de datos:

- La información de medios tradicionales (radio, TV y prensa escrita) se almacena en MySQL y se indexa en Apache Solr.
- La información de prensa digital se almacena en una base de datos NoSQL (MongoDB).

El uso de dos bases de datos distintas atiende a una estrategia de modernización del repositorio de datos que se ha iniciado con la información procedente de la prensa escrita.

### 6.2.1 Datos de las noticias

Como hemos mencionado anteriormente, la información obtenida de las noticias es diversa y contiene además metadatos de importancia que serán de utilidad en el procesamiento de las mismas. Las características y atributos más relevantes de las noticias son los siguientes:

- Título de la noticia: El texto que corresponde al título de la noticia, este dato es importante dado que indica de forma concisa, la temática que se aborda en la noticia, e incluso es útil para hacer una clasificación básica de las noticias en categorías.
- Autor: El nombre o pseudónimo completo del autor o locutor de la noticia.

- Contacto del autor: En algunas publicaciones, se puede obtener información de contacto del autor, ya sea su correo electrónico o URL en redes sociales (perfil LinkedIn, link a Twitter).
- Fecha de publicación: La fecha en que se publicó la noticia.
- Resumen de la noticia: En cierto tipo de noticias, se incluye un breve resumen que generalmente se muestra entre el título y el cuerpo de la noticia. Esto es aplicable en la prensa escrita y digital.
- Cuerpo de la noticia: Es el texto principal que brinda los detalles del evento, es una fuente valiosa de información para el posterior análisis de la noticia en cuanto a la extracción de datos valiosos como el sentimiento, palabras claves, intenciones, etc.
- Imágenes de las noticias: Esto corresponde a fotografías o infografías incluidas como complemento al cuerpo de la noticia. Las imágenes, además de complementar visualmente las noticias, pueden proveer información como rostros o lugares de interés que eventualmente podrían ser útiles para un análisis más profundo de las mismas.
- Nombre del medio: El nombre del medio de comunicación que publica la noticia.
- Tipo de medio: Este es un clasificador de los medios de comunicación que incluye: radio, televisión, prensa escrita y prensa digital.
- URL de la noticia: La dirección de Internet donde puede encontrarse la noticia original, en caso que el medio que la publica tenga una versión web.
- URLs referenciados: Si el cuerpo de la noticia contiene enlaces externos, estos se almacenan.
- País origen: El país origen del medio que publica la noticia.
- Región origen: La región geográfica origen del medio que publica la noticia.
- Ubicación en la fuente: En casos donde aplique, se debe especificar la ubicación de la noticia en el medio, con el fin de determinar si se encuentra en la portada ó en otras categorías secundarias.
- ID único de la noticia: Un identificador interno único de la noticia.
- Fecha de carga de la noticia: Corresponde a la fecha en que se cargó la noticia a la base de datos de IMEDIA.
- Audiencia: Este dato aplica para prensa digital, e indica la cantidad estimada de visitas a la noticia.
- No. de veces compartido en Twitter: Indica la cantidad de veces que se ha hecho referencia a la noticia en los tweets públicos en Internet. Este dato permite conocer el

alcance que ha tenido determinada noticia y puede utilizarse como un parámetro para estimar el nivel de credibilidad ó impacto de los eventos.

- No. de likes en Twitter: Las veces que las personas han dado “like” a la noticia en tweets públicos.
- No. de veces compartido en Facebook: Indica la cantidad de veces que se ha hecho referencia a la noticia en posts públicos de Facebook. Este dato permite conocer el alcance que ha tenido determinada noticia y puede utilizarse, al igual que el número de likes en Twitter, como un parámetro para estimar el nivel de credibilidad ó impacto de los eventos.
- No. de likes en Facebook: Las veces que las personas han dado “like” a publicaciones públicas que incluyan la noticia en Facebook.

### 6.3 Modelo lógico

El actual modelo de obtención y almacenamiento de noticias de IMEDIA se realiza mediante servicios de infraestructura en la nube, específicamente de Amazon Web Services. De forma general el proceso actual es el siguiente:

1. Lectura del sitio de noticias o RSS.
2. Procesamiento e identificación de datos de las noticias.
3. Almacenamiento.

Lo anterior se incluye como referencia de la obtención de los datos que ya realiza IMEDIA que alimenta la base de datos a partir de la cual se creará la solución de automatización de la clasificación y asociación de noticias delictivas a personas.

Al incluir el proceso de identificación, clasificación y relación de las noticias con las personas, la automatización de dicho flujo se puede diagramar de forma general en la imagen a continuación:

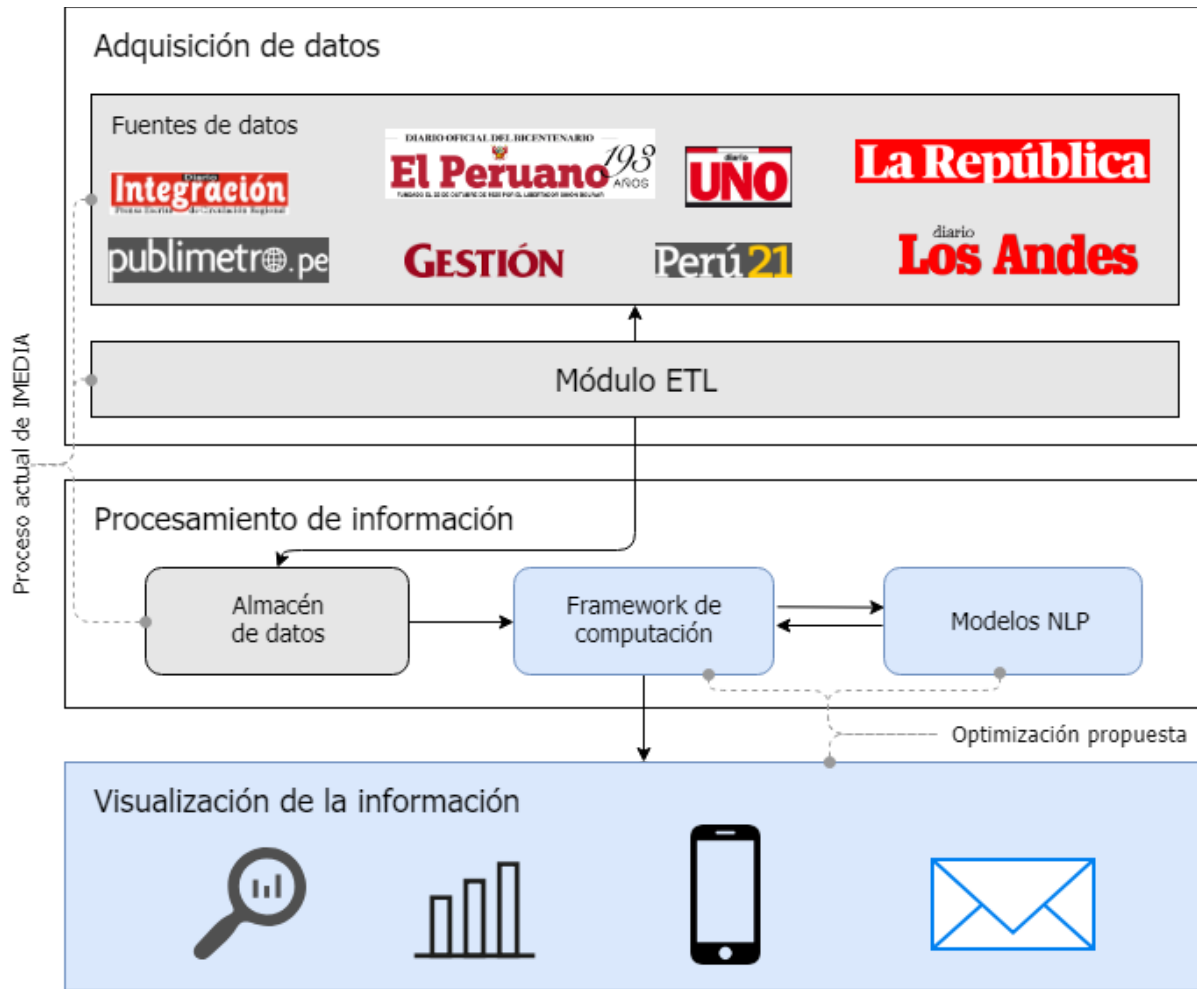


Figura 8. Modelo lógico

La parte más importante de este proceso es el procesamiento del texto de las noticias con los modelos de Natural Language Processing (NLP) con el fin de identificar las entidades que se mencionan, el sentimiento general del artículo, las palabras claves con más frecuencia, que servirán para identificar las noticias que incluyen referencias a actividades delictivas para posteriormente ser verificadas por personal de IMEDIA, con el fin de disminuir las posibilidades de error, pero contando con unas entidades y relaciones ya procesadas de forma ágil.

## 6.4 Arquitectura

Con el fin de definir la arquitectura para el procesamiento de los datos de noticias, evaluaremos dos de los paradigmas más utilizados para el procesamiento de Big Data: Arquitectura Lambda y arquitectura Kappa.

Los criterios que más peso tendrán al momento de evaluar esta arquitectura son los siguientes:

- Escalabilidad: Esta es la características más importante, con el fin de garantizar la operación aunque el volumen, la velocidad o la variedad de los datos aumenten.
- Flexibilidad: Idealmente la arquitectura debe poderse ajustar a nuevos formatos y requerimientos de procesamiento.
- Modularidad/Integración: La arquitectura debe permitir incluir nuevos módulos de procesamiento en casos de necesidades futuras de información.
- Agilidad de procesamiento y consulta: Los datos deben procesarse de forma ágil, y la información debe poderse consultar de forma eficiente.
- Procesamiento distribuido: Para mejor rendimiento y aprovechamiento de recursos.
- Tolerancia a fallos: La arquitectura debe garantizar el funcionamiento constante del procesamiento, gestionando y manejando de la mejor forma los posibles fallos que puedan darse.

#### 6.4.1 Arquitectura Lambda

La arquitectura Lambda, está diseñada para la gestión de grandes cantidades de datos, utilizando un esquema para procesamiento tanto de datos en batch, como en tiempo real. El propósito es lograr un balance entre latencia, volumen de datos y tolerancia a fallos, consolidando la información procesada en una capa de servicios que contiene las vistas de acceso a la información.

La capa de procesamiento por lotes (Batch) de esta arquitectura opera a través de bloques de información y de esta forma ayuda a mantener las vistas de acceso con datos más consistentes, en cambio el procesamiento en tiempo real se basa en un modelo incremental que considera únicamente los datos nuevos que van cambiando, esto compensa la latencia del esquema en Batch.

De forma general esta arquitectura puede representarse de esta forma:



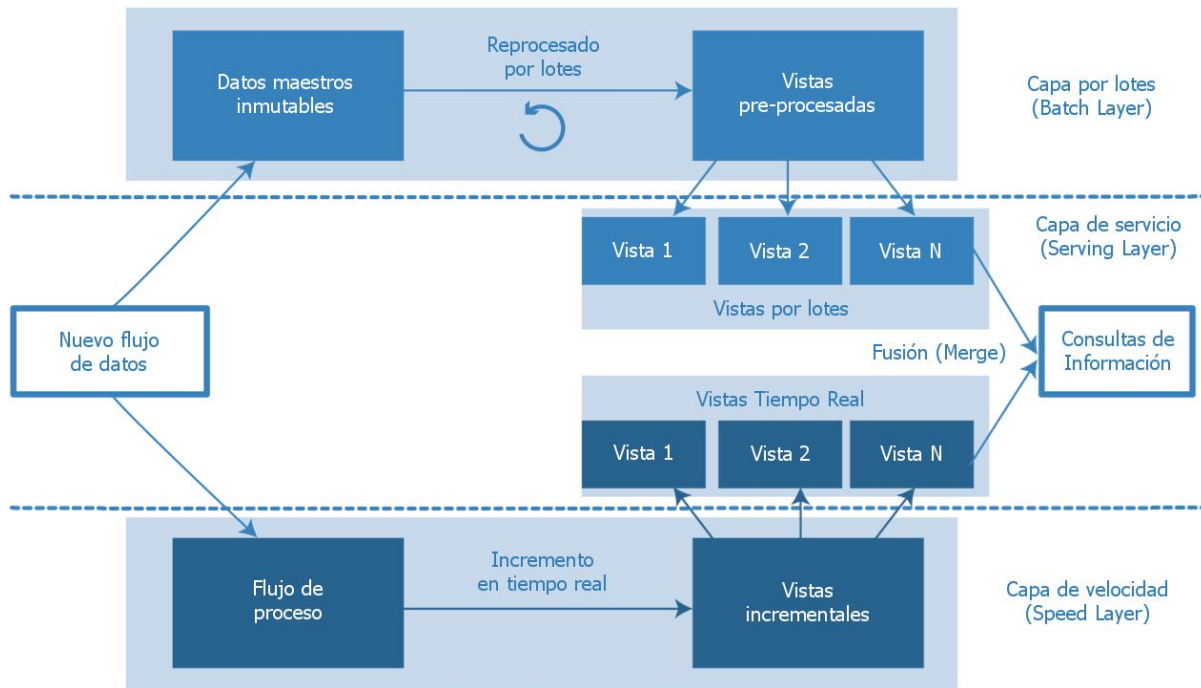


Figura 9. Arquitectura Lambda

Esta capa tiene el inconveniente de contener lógica duplicada debido a que debe existir tanto en la capa de procesamiento por lotes, como en la capa de procesamiento en tiempo real.

#### 6.4.2 Arquitectura Kappa

La arquitectura Kappa se refiere a una simplificación de la arquitectura Lambda, eliminando la capa de procesamiento en lotes, y reemplazando esto con la capa ya existente en tiempo real.

Para lo anterior se auxilia de un almacén de datos inmutable que funciona a manera de log, a partir de estos datos es que realiza el procesamiento y la correspondiente carga para la capa de servicio. De esta forma, cuando es necesario procesar información previa, basta con ejecutar nuevamente el procesamiento en tiempo real a partir del almacén inmutable. Esto facilita el mantenimiento del código fuente de la solución.

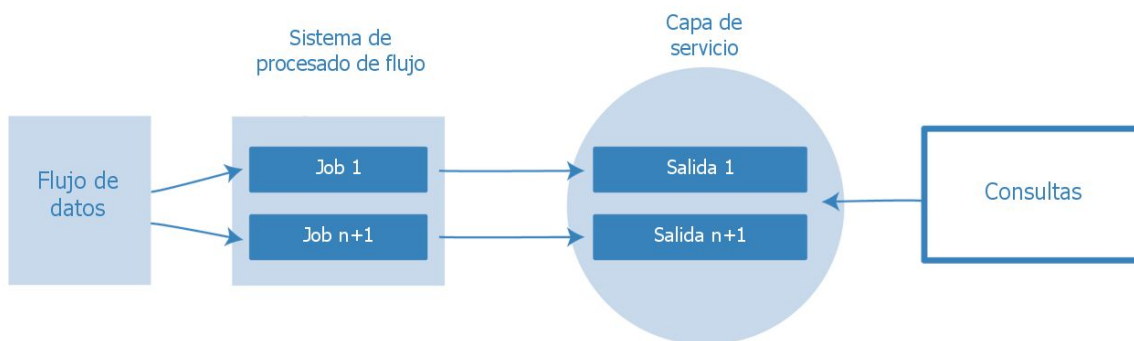


Figura 10. Arquitectura Kappa

## 6.5 Solución propuesta

Para tomar la decisión de arquitectura que mejor se adapta a la solución de Inteligencia Financiera de IMEDIA realizamos una evaluación de la mejor solución para cada uno de los criterios planteados anteriormente:

- Tanto Lambda como Kappa pueden responder a las necesidades de escalabilidad y modularidad del proyecto de Inteligencia Financiera de IMEDIA. A su vez, ambas pueden operar en un entorno distribuido.
- El tipo de análisis que se realiza sobre las noticias está orientado a identificación de personas, análisis de sentimiento e identificación de palabras claves, lo que puede realizarse de forma independiente sobre cada noticia con ambas arquitecturas.
- Consideramos importante facilitar el mantenimiento del código fuente de solución por lo que en este particular, la arquitectura Kappa nos brinda la ventaja de centralizar la lógica de procesamiento.
- Por la naturaleza de los datos de las noticias, en donde cada evento puede analizarse de forma independiente, el proceso de identificación, clasificación y relación de las noticias con las personas es un proceso incremental. Los procesamientos en lote deben de ser únicamente en escenarios especiales donde se incluyan nuevos algoritmos de NLP. Para estos casos, el almacén de datos inmutables de la arquitectura Kappa es suficiente, tomando en cuenta la ventaja que al incluir nuevos algoritmos sólo se debe actualizar una única versión de la lógica de procesamiento.
- La principal desventaja de la arquitectura Kappa es la disponibilidad de procesamiento en tiempo real en los momentos donde se ejecutan procesamientos en lotes. Sin embargo, dado que la plataforma está orientada a operar inicialmente en un único país, en los casos donde se requiera procesamiento masivo de lotes de noticias estos se podrían agendar en horas nocturnas, de tal forma que no se afecten las alertas en tiempo real en los horarios que son requeridas.

De esta forma, aunque ambas arquitecturas son compatibles con la solución a implementar, el hecho de valorar la centralización de la lógica de procesamiento y por la naturaleza de las noticias y la solución a implementar, hemos decidido basarnos en la arquitectura Kappa como base la implementación de la solución de Inteligencia Financiera de IMEDIA.

La arquitectura propuesta con las tecnologías a utilizar es la siguiente:



Figura 11. Solución propuesta

Cada una de las tecnologías a utilizar y la razón de su elección se explican a continuación:

## 6.5.1 Identificación de nuevas noticias con Change Data Capture (CDC)

Change Data Capture son tecnologías y técnicas de bases de datos utilizadas para determinar y rastrear los cambios que ocurren en los repositorios que permiten identificar registros nuevos, modificados o eliminados, es utilizado en almacenes de datos y procesos de integración y ofrece ventajas para ejecución de procesos en tiempo real.

Se implementará esta técnica con el fin de identificar nuevas noticias que se carguen en los repositorios actuales de noticia de IMEDIA (MySQL y MongoDB) y así poder “disparar” el flujo de procesamiento que automatice el proceso de análisis NLP para identificar las entidades y eventos relacionados a hechos delictivos hasta finalizar con su almacenamiento en ElasticSearch.

Existen distintas técnicas para lograr la identificación de cambios en las bases de datos, entre las cuales se encuentran:

- Triggers a nivel de base de datos: Mediante scripts que registren y almacenen los cambios ante cada evento, en un “audit log” o registro de auditoría.
- Triggers a nivel de aplicación: Una variación del modelo anterior, pero gestionado a nivel de la aplicación que interactúa con los datos.
- Log de transacciones del motor de base de datos: Los motores de bases de datos ya cuentan internamente con logs de transacciones (transaction log) que permiten conocer todos los cambios realizados sobre los registros.

Tomando en cuenta que los registros de auditorías representa un duplicado de los logs de transacciones existentes en las bases de datos, se utilizará esta última información para identificar las noticias nuevas que se almacenen en las bases de datos de IMEDIA.

**Debezium** es un proyecto de código abierto de la empresa Red Hat, cuya funcionalidad consiste en integrarse a los logs de transacciones mediante conectores, entre los cuales actualmente están: Oracle, MySQL, PostgreSQL y MongoDB. Provee una plataforma de flujo de baja latencia para CDC, ofreciendo el valor agregado de propagar los cambios identificados a Apache Kafka, quién se encargará de la gestión de estos eventos a como se explicará en la siguiente sección.

Un diagrama representativo del funcionamiento de Debezium es el siguiente:



Figura 12 - CDC con Debezium

### 6.5.2 Preprocesamiento con Kafka

Kafka provee una plataforma flexible, escalable, y confiable para el flujo de datos. Su funcionamiento se basa en recibir y almacenar mensajes de tipo llave-valor que se originan de distintos procesos, denominados **productores**. Estos mensajes son organizados en **particiones** dentro de distintos **tópicos**. Otros procesos, denominados **consumidores** pueden suscribirse a estos mensajes almacenados en las particiones.

Una representación de su funcionamiento es la siguiente:

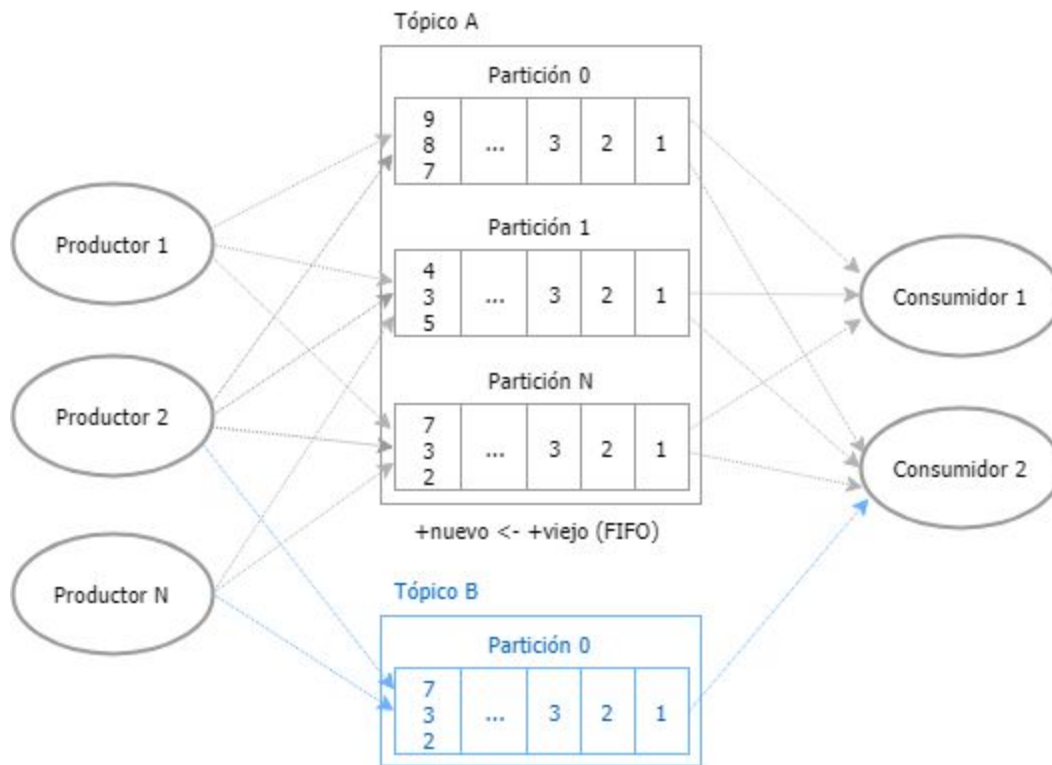


Figura 13. Funcionamiento de Kafka

Está diseñado para operar en clúster de uno o más servidores, distribuyendo las particiones de los tópicos en los nodos del clúster. Adicionalmente, las particiones pueden ser replicadas a otros clústeres. Este tipo de arquitectura lo convierte en una plataforma flexible, escalable y confiable para el flujo de datos.

Actualmente es una solución ampliamente utilizada en arquitecturas big data por su estabilidad, flexibilidad y comunidad open source. Entre las empresas que la utilizan se puede mencionar a Apple, Pay Pal, Spotify, Netflix, Uber, y otras más.

En la arquitectura propuesta para nuestro proyecto, Kafka se ubica como el primer procesamiento de las noticias almacenadas en MongoDB y MySQL, invocado a partir del monitoreo de Debezium al log de transacciones. Kafka es el encargado de la gestión de las colas de los eventos identificados en los repositorios, así como de garantizar la persistencia, consistencia y fluidez para la siguiente etapa de procesamiento en tiempo real con Spark Streaming.

Las principales características que utilizaremos de Kafka son las siguientes:

- Monitoreo de cambios en bases de datos MongoDB y MySQL, mediante el seguimiento del log de transacciones de dichos motores, lo que minimiza el impacto en el rendimiento de

las bases de datos y permite que se generen eventos en cuanto se almacenen nuevas noticias.

- Gestión de mensajes: Al funcionar como intermediario entre las bases de datos y el procesamiento en tiempo real, permite mayor estabilidad de la plataforma al gestionar temas como las colas de mensajes y la retención temporal de datos.



Figura 14 - Uso de Kafka

El proyecto de Inteligencia Financiera de IMEDIA gestionará en Kafka, el tópico “Noticias”, gestionado en 2 brokers, cada uno con 3 particiones. Lo anterior para garantizar la estabilidad de la plataforma, sin necesidad de requerir grandes cantidades de procesamiento:

### 6.5.3 Procesado de Flujo

El flujo de datos ya gestionado mediante Kafka, ingresará a un motor de procesamiento distribuido en donde se deben realizar las principales operaciones con la información de cada noticia. Es en esta etapa donde se realizan las tareas más importantes sobre el texto de las noticias, incluyendo procesamiento de lenguaje natural e identificación y asociación de entidades. Para esta etapa utilizaremos el motor de **Spark Streaming**, que es un sistema de procesamiento de flujo.

Un sistema de procesado de flujo gestiona el procesamiento continuo de datos en cuanto este es producido o recibido. Se adapta muy bien a la realidad de datos de sensores, datos financieros, interacción de usuarios en sitios web o videojuegos, y todo flujo constante de datos generados en el tiempo por una o distintas fuentes. A diferencia de arquitecturas previas donde los datos se consultaban directamente en los repositorios de datos conforme se utilizaban, en los sistemas de procesado de flujo la lógica, analítica y consultas se ejecutan continuamente y los datos fluyen a través de ellas de forma constante, habiendo sido generados por una o más fuentes.

Un aspecto adicional importante de los sistemas de procesado de flujo es que, al gestionar datos provenientes de distintas fuentes de forma activa, incluyen funcionalidades que les permiten escalar y ajustarse a la demanda de procesamiento requerida mediante arquitecturas distribuidas, permitiendo además ser tolerantes a fallos.

Spark Streaming provee una plataforma escalable, eficiente e integrada de procesamiento de flujo y procesamiento por lote, que además ofrece compatibilidad con distintos lenguajes de programación (Java, Scala y Python). Integra además mecanismos de tolerancia a fallos que le permite recuperar el estado de trabajos que han sido abortados, todo esto contemplado de forma



nativa, sin necesidad de gestionarlo en el código de la solución. Permite además el ingreso de datos de distintas fuentes como Kafka, Flume, Kinesis, etc, y a su vez puede cargar información a bases de datos y sistemas de archivo. Todo esto, además de contar con gran soporte de la comunidad y de casos de éxito en empresas de gran prestigio, nos ha motivado a seleccionarlo como motor de procesamiento en el proyecto de Inteligencia Financiera de IMEDIA.

#### 6.5.4 Procesamiento del Lenguaje Natural (PLN)

A continuación, presentamos la forma en que el procesamiento de lenguaje natural cobra protagonismo para el desenvolvimiento del proyecto aportando sustancial información para la determinación de hechos delictivos y la detección automática de personas y organizaciones.

El proyecto utilizará para ello una herramienta existente en el mercado la cual posee la madurez y solidez que requerimos para llevar a cabo el proyecto. En consecuencia, la herramienta a utilizar será la de [meaningcloud.com](http://meaningcloud.com).

Esta API extrae de un texto la información más relevante, como personas, lugares, organizaciones o productos mencionados, conocidos como entidades con nombre. Además, también identifica los conceptos principales y muchos otros datos relevantes como fechas, números de teléfono, cantidades de dinero o direcciones electrónicas (URL, correos electrónicos, hashtags). Estas entidades, conceptos y valores proporcionan una representación semántica de un documento, lo que permite desarrollar aplicaciones inteligentes para procesar contenido. Sin embargo, el análisis no se limita a la identificación de una entidad mencionada: a través de un análisis de referencia se vincula a recursos externos que lo representan, como Wikipedia o Datos Vinculados.

“MeaningCloud identifica esta información en cualquier tipo de texto, ya sea una página web, noticias, contenido de redes sociales o transcripciones de audio y video. Es capaz de llevar a cabo el análisis no solo en diferentes idiomas, sino que también utiliza un conjunto de tipos común (multilingüe). Esta jerarquía de tipos de entidades, que se conoce como ontología, contiene más de 200 clases que permiten decir, por ejemplo, que Google es una organización y una compañía de software al mismo tiempo. “ (Fuente: [Meaningcloud.com](http://Meaningcloud.com))

Este proveedor expone varios endpoints o servicios web a través de API sin embargo es importante aclarar que los servicios que utilizaremos serán los siguientes:

##### **Topic Extraction o Extracción de temas**

Este endpoint nos permitirá conocer los temas identificados en el texto de las publicaciones presentadas en las diferentes noticias y artículos web de internet.

En el siguiente ejemplo se expone un sujeto quien está involucrado en robo de celulares en el distrito de San Martín de Porres, Lima



Figura 15 – Publicación del diario Trome 23.03.2019

El texto a analizar es el siguiente:

“POR ROBO DE CELULARES Atrapan a ‘Chino’ Fue intervenido por la policía un sujeto de 24 años, apodado ‘Chino’ y que, según la policía, se hacía pasar como pasajero para subir a los buses de transporte público y robar celulares a mujeres distraídas, en San Martín de Porres. Luis Tito fue atrapado por agentes del grupo Terna del Escuadrón Verde cuando descendía de un vehículo con el teléfono de una víctima. Un video de seguimiento lo captó al momento de escapar, a la altura del puente ‘Pocitos’, en la Panamericana Norte. (J. Y.)”

Utilizando el motor NLP de Meaningcloud podemos extraer lo siguiente:

Entity	Type	Subtype	Appearances	Relevance
San Martín	Location	City	1	100
San Martín	Location	City	1	100
Porres	Person	LastName	1	100
Luis Tito	Person	FullName	1	100
Terna del Escuadrón Verde	Unknown	-	1	100
Panamericana Norte	Unknown	-	1	100

Figura 16 – Listado de Entidades obtenidos en meaningcloud.com

En dicho proceso somos capaces de detectar la persona "Luis Ttito" y también incurrimos en la detección de un apellido aislado el cual genera ambigüedad por no encontrarse en el diccionario del motor NLP.

Este tipo de ambigüedades serán tratadas posteriormente bajo una capa de negocio que actuará como anti filtro.

Concept	Type	Subtype	Appearances	Relevance
móvil	Product	MobilePhone	2	100
policía	OtherEntity	Vocation	2	100
policía	Organization	Police	2	100
transporte público	? Unknown	-	1	70
transporte público	? Process	Transportation	1	70

Figura 17 – Listado de conceptos obtenidos a partir de meaningcloud.com

De igual manera la determinación o detección de conceptos clave relacionados a delitos como la palabra "policía" con relevancia al 100% dotarán a nuestro motor de búsqueda un importante criterio de detección de falsos positivos o negativos.

### **Text Classification o Clasificación de texto**

Este endpoint nos permitirá conocer la temática o taxonomía semántica del texto inmerso en el análisis. De esta manera el motor será capaz de comprender semánticamente el contexto en que se desenvuelve la noticia y si ésta tiene un ponderado relevante respecto a nuestro foco de análisis el cual se asocia a delitos, crimen y comportamiento inadecuado en la sociedad.

IPTC		
Code	Label	Relevance
02003000	policía y justicia - policía	100

IAB		
Code	Label	Relevance
Automotive	Automoción	100
LawGovt&Politics	Ley, gobierno y política	100
Society	Sociedad	100
Technology&Computing>CellPhones	Tecnología y programación>Móviles	100

Figura 18 – Listado de Clasificaciones encontradas según las taxonomías IAB e IPTC.

Tomando como referencia la clasificación IPTC el artículo es claramente relacionado al código de taxonomía 02003000 el cual se refiere a hechos policiales y de justicia.

### Sentiment Analysis o Análisis de sentimiento.

Este endpoint nos permite comprender el sentimiento de la nota de prensa de manera que seamos capaces de evaluar si la noticia contribuye positivamente o negativamente al análisis contextual de la historia.

This text is **Positive** with a confidence of a **86** percent. The polarities detected in it are in **disagreement**. The text is **subjective** and **without irony**.

#### Feature-level sentiment

Entities Concepts

Entity	Type	Subtype	Sentiment
San Martín	Location	City	None
San Martín	Location	City	None
Luis Ttito	Person	FullName	Very positive
Panamericana Norte	Unknown	-	None
Terna del Escuadrón Verde	Unknown	-	None
Porres	Person	LastName	None

Figura 19 – Listado de polaridad de entidades.

El motor asume que las personas no conocidas por Wikipedia y que no poseen un reconocimiento público aportan por defecto una tonalidad positiva al contexto de la historia. En consecuencia, el reconocimiento público de las actividades del personaje aporta un sentido a la reputación de la persona.

Por ejemplo, si analizamos la frase: “Alberto Fujimori ex presidente del Perú fue sentenciado a 20 años de prisión.”

El motor será capaz de comprender que noticia expone o tiene un sentido negativo en el contexto.



Figura 20 – Análisis de sentimiento considerando personajes conocidos.

En consecuencia, dado que nuestro producto aporta valor a los clientes del rubro financiero que desconocen de personas u organizaciones nuevas involucradas en hechos delictivos es importante dar mayor relevancia a la determinación de sentimiento a partir de conceptos.

Entities		Concepts	
Concept	Type	Subtype	Sentiment
agente	Person	Person	Very positive
móvil	Product	MobilePhone	Neutral
autobús	Product	Vehicle	None
víctima	Person	Person	Very negative
policía	OtherEntity	Vocation	Negative
grupo	Organization	Organization	None
agente	Person	Person	Very positive
altura	Unit	PhysicalExtentUnit	None
puente	LivingThing	BodyPart	None
pasajero	Person	Person	None
mujer	Person	Person	None
teléfono	Id	PhoneNumber	Neutral
vehículo	Product	Vehicle	Very positive
robo	OtherEntity	Offence	Neutral
agente	OtherEntity	Vocation	Very positive
transporte público	Unknown	-	None
policía	Organization	Police	Negative
transporte público	Process	Transportation	None
puente	Location	Line	None

Figura 21 – Listado de polaridad de los conceptos.

En la figura apreciamos como el motor NLP es capaz de identificar qué conceptos contribuyen positivamente o negativamente al valor semántico de la noticia.

De esta manera seremos capaces de poder identificar hechos delictivos a través de la identificación de sustantivos clave involucrados con situaciones delictivas.

Haciendo un análisis más exhaustivo del resultado por frase podemos comprender por qué el score de sentimiento arrojado es positivo.

```

"status": {
  "code": "0",
  "msg": "OK",
  "credits": "1",
  "remaining_credits": "2581576",
  "model": "general_es",
  "score_tag": "P",
  "agreement": "DISAGREEMENT",
  "subjectivity": "SUBJECTIVE",
  "confidence": "86",
  "irony": "NONIRONIC",
  // ...
}

```

Level	Text	Score tag	Agreement	Confidence
Global	-	P	DISAGREEMENT	86
Sentence	POR ROBO DE CELULARES Atrapan a 'Chino' Fue intervenido por la policía un sujeto de 24 años, apodado 'Chino' y que, según la policía, se hacía pasar como pasajero para subir a los buses de transporte público y robar celulares a mujeres distraídas, en San Martín de Porres. Luis Ttito fue atrapado por agentes del grupo Terna del Escuadrón Verde cuando descendía de un vehículo con el teléfono de una víctima.	P	DISAGREEMENT	86
Segment	POR ROBO DE CELULARES Atrapan a 'Chino' Fue intervenido por la policía un sujeto de 24 años, apodado 'Chino'	N	DISAGREEMENT	97
Segment	que, según la policía, se hacía pasar como pasajero para subir a los buses de transporte público y robar celulares a mujeres distraídas	NONE	AGREEMENT	100

Segment	en San Martín de Porres	NONE	AGREEMENT	100
Segment	Luis Ttito fue atrapado por agentes del grupo Terna del Escuadrón Verde cuando descendía de un vehículo con el teléfono de una víctima	P+	AGREEMENT	92
Sentence	Un video de seguimiento lo captó al momento de escapar, a la altura del puente 'Pocitos', en la Panamericana Norte.	NONE	AGREEMENT	100
Segment	Un video de seguimiento lo captó al momento de escapar, a la altura del puente 'Pocitos', en la Panamericana Norte	NONE	AGREEMENT	100
Sentence	(J. Y.)	NONE	AGREEMENT	100
Segment	(J. Y.)	NONE	AGREEMENT	100

Figura 22 – Resultado del API Formateado de Json a CSV.

Para poder interpretar el resultado es importante conocer los campos de la respuesta.

Nombre	Descripción
status	<p>Contiene información sobre el proceso de extracción y si ha finalizado correctamente. Está formado por cuatro valores diferentes:</p> <ul style="list-style-type: none"> <li>• code: un número natural que identifica el estado devuelto.</li> <li>• msg: una cadena que describe el estado.</li> <li>• credits: Contiene un número natural que indica los créditos consumidos en la solicitud.</li> <li>• remaining_credits: muestra el número de créditos que le quedan al usuario para alcanzar el límite de crédito.</li> </ul>
model	Este campo contiene el modelo utilizado en la evaluación seguido de un guión bajo y el idioma en el que se realizó el análisis.
score_tag	<p>Esta etiqueta indica la polaridad encontrada (o no encontrada) en el elemento al que se refiere. Puede aplicarse a varios elementos: polarity_term, sentimented_concept, sentimented_entity, segment, sentence o la polaridad global del texto.</p> <p>Los valores posibles son los siguientes:</p> <ul style="list-style-type: none"> <li>• P+: muy positivo</li> <li>• P: positivo</li> <li>• NEU: neutro</li> <li>• N: negativo</li> <li>• N+: muy negativo</li> <li>• NONE: sin sentimiento</li> </ul>



agreement	Este campo marca el acuerdo entre los sentimientos detectados en el texto, la oración o el segmento al que se refiere. Tiene dos valores posibles.: <ul style="list-style-type: none"> <li>• AGREEMENT: Los diferentes elementos tienen la misma polaridad..</li> <li>• DISAGREEMENT: Hay desacuerdo entre la polaridad de los diferentes elementos.</li> </ul>
subjectivity	Este campo marca la subjetividad del texto. Tiene dos valores posibles: <ul style="list-style-type: none"> <li>• OBJECTIVE: El texto no tiene marcas de subjetividad.</li> <li>• SUBJECTIVE: El texto tiene marcas subjetivas.</li> </ul>
confidence	Este campo representa la confianza asociada con el análisis de sentimiento realizado en el texto. Su valor es un número entero en el rango 0-100.
irony	Este campo indica la ironía del texto. Tiene dos valores posibles: <ul style="list-style-type: none"> <li>• NONIRONIC: El texto no tiene marcas irónicas.</li> <li>• IRONIC: El texto tiene marcas irónicas.</li> </ul>
sentence_list	Lista de frases en las que se divide el texto. Cada oración está representada por un elemento. <a href="#">sentence</a> .

Tabla 4 - Campos de respuesta MeaningCloud. Fuente: Meaningcloud.com

Podemos concluir que el resultado es positivo dado que han conseguido atrapar al ladrón de celulares, esto porque en la noticia se mezcla el score de un evento negativo (el robo), con una conclusión positiva (capturaron al ladrón), resultando en mayor peso el evento que MeaningCloud identifica como evento final del texto.

La herramienta de sentimiento contribuye significativamente a la determinación de la importancia de la noticia respecto a la probabilidad del sector involucrado. En ese sentido somos capaces de determinar que la noticia está relacionada a un hecho delictivo y si aún más tiene relación directa con los intereses de nuestros clientes de forma positiva o negativa.

#### 6.5.4 Capa de Servicio

Los datos ya procesados en la plataforma de Spark Streaming utilizando el API de MeaningCloud deben almacenarse para quedar disponibles para su posterior verificación y consulta por el usuario final, para este fin utilizaremos el motor de base de datos **ElasticSearch**.

**ElasticSearch** es una base de datos NoSQL distribuida y optimizada para el almacenamiento de texto. Es una tecnología de código abierto, escalable y de amplio uso empresarial como motor de búsqueda debido a su alto rendimiento y API orientado a búsquedas y análisis de datos.

Entre sus principales características se encuentran:

- Distribuida y escalable: Está diseñada para poder operar ya sea en un solo equipo de cómputo hasta en un clúster de múltiples nodos, mediante escalamiento horizontal, gestionando automáticamente la distribución de índices y consultas.

- Alto rendimiento: Su arquitectura está orientada a realizar operaciones de lectura y escritura de forma rápida, reduciendo el tiempo requerido para almacenar un documento, indexarlo y ponerlo disponible para consulta. Además, su naturaleza distribuida le permite procesar grandes volúmenes de datos en paralelo, obteniendo rápidamente los resultados de las consultas.
- Alta disponibilidad: El motor de ElasticSearch incluye mecanismos para garantizar la disponibilidad de los datos en caso que un nodo falle o no esté funcionando correctamente, redireccionando los datos a otro recurso disponible.
- Motor de búsqueda a través de una interfaz web HTTP, utilizando documentos JSON como estructura de datos tanto para los parámetros de consultas como para los resultados de las mismas. Ofrece además librerías oficiales para los siguientes lenguajes: Java, Groovy, PHP, Ruby, Perl, Python, .NET, y Javascript.

De acuerdo al sitio DB-Engines, dedicado al seguimiento de estadísticas de uso de motores de bases de datos, ElasticSearch es el motor de búsqueda empresarial más popular:

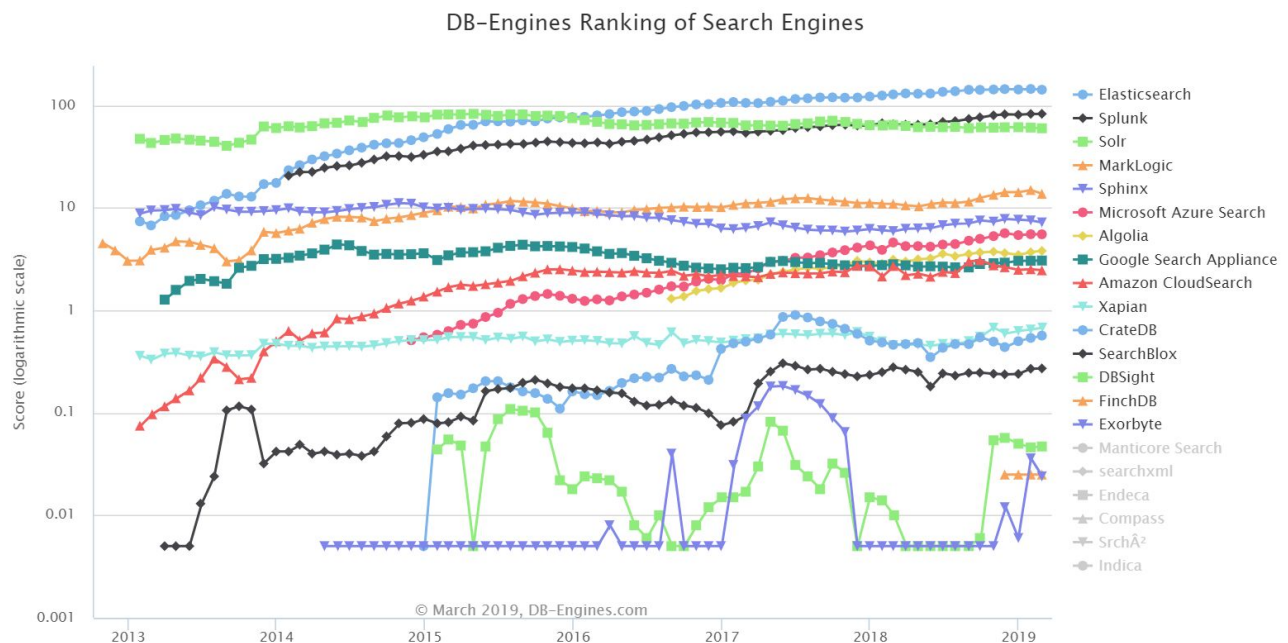


Figura 23, [ranking de uso de motores de búsqueda](#)

Esta tendencia, y las características expuestas anteriormente, son las razones por la que decidimos utilizar este motor como base de datos de la información procesada de las noticias.

Un ejemplo de un documento de tipo “noticia” en ElasticSearch es el siguiente:

```

{
  "_id": "123456789",
  "titulo": "Atrapan a 'Chino'",
  "fecha_hora_publicacion": "2019-09-16 10:39",
  "resumen": "El detenido fue puesto a disposición de la comisaría de Barboncitos. Las cámaras de seguridad captaron sus actos delictivos",
  "cuerpo": "POR ROBO DE CELULARES Atrapan a 'Chino' Fue intervenido por la policía un sujeto de 24 años, apodado 'Chino' y que, según la policía, se hacía pasar como pasajero para subir a los buses de transporte público y robar celulares a mujeres distraídas, en San Martín de Porres...",
  "imagenes":
  ["https://img.elcomercio.pe/files/article_content_ec_fotos/uploads/2018/09/16/5b9e6ef511114.jpeg"],
  "tipo_medio": "El Comercio Perú",
  "url":
  "https://elcomercio.pe/lima/smp-capturan-ladrones-dedicados-robar-celulares-pasajeros-video-noticia-nndc-558253",
  "url_externos": [
    {
      "texto": "San Martín de Porres",
      "url": "https://elcomercio.pe/noticias/san-martin-de-porres"
    }
  ],
  "pais_origen": "Perú",
  "region_origen": "San Martín de Porres",
  "ubicacion_en_fuente": "",
  "fecha_hora_carga": "2019-03-31 23:59:59",
  "entidades": [
    {
      "descripcion_entidad": "Luis Tito",
      "subtipo": "FullName",
      "sentimiento": "Very positive",
      "DNI": "XYZ-123",
      "nombres": "Luis",
      "apellidos": "Tito",
      "origen": "Lima",
      "match_manual": "1",
      "comentario": "",
      "usuario_clasificacion": "Juan López",
      "fecha_clasificacion": "2019-04-01 08:01:00"
    }
  ],
  {

```

```

    "_id": "12346",
    "descripcion_entidad": "Porres",
    "subtipo": "LastName",
    "sentimiento": "Very positive",
    "DNI": "",
    "nombres": "",
    "apellidos": "",
    "origen": "",
    "match_manual": "0",
    "comentario": "Porres corresponde a una ubicación",
    "usuario_clasificacion": "Juan López",
    "fecha_clasificacion": "2019-04-01 08:00:05"
  }
]
}

```

En cada documento de noticia, se almacenará un arreglo de las entidades que se logren identificar con MeaningCloud con el tipo “Person” y “LastName”. Este arreglo deberá ser definido como tipo “nested” para garantizar que pueda ser consultado de forma independiente (para las búsquedas por entidad). En este arreglo se encuentran los siguientes campos:

- Descripción\_Entidad: Contiene el texto identificado por MeaningCloud.
- Subtipo: Define el subtipo de la entidad, también por MeaningCloud.
- Sentimiento: Especifica el sentimiento identificado por MeaningCloud.
- DNI: Es el documento nacional de identidad del Perú. Este se ingresará en el proceso de verificación manual que realizará el analista ya en la capa de aplicación.
- Nombres: Obtenido manualmente al separar el campo “Descripción\_Entidad”.
- Apellidos: Obtenido manualmente al separar el campo “Descripción\_Entidad”.
- Origen: Ubicación de la entidad, también se debe ingresar manualmente.
- Match\_Manual: Este campo por defecto se carga vacío, y se actualiza una vez que el analista confirma o rechaza la relación de la noticia con la entidad.
- Comentario: El comentario ingresado por el analista.
- Usuario\_Clasificacion: El usuario del analista que revisó la relación.
- Fecha\_Clasificacion: La fecha de la revisión de la relación.

A partir del ejemplo anterior se puede generar la siguiente consulta para obtener todas las noticias relacionadas a entidades con apellido “Tito” que ya han sido verificadas por el analista de IMEDIA:

```
// GET IMEDIA/_search
```

```

{
  "query": {
    "nested": {
      "path": "entidades",
      "query": {
        "bool": {
          "must": [
            { "match": { "entidades.apellidos": "Tito" } },
            { "match": { "entidades.match_manual": "1" } }
          ]
        }
      }
    }
  }
}

```

Por su arquitectura, Elasticsearch puede procesar consultas como la anterior de forma eficiente, siendo de gran ayuda en proyectos de este tipo.

### 6.5.5 Capa de aplicación

El proceso de verificación de las relaciones entre entidades y noticias que realizará el analista de IMEDIA, se realizará en una aplicación web que consumirá y actualizará los campos relacionados, en los documentos de Elasticsearch.

Para esta etapa se utilizarán las siguientes tecnologías:

- Django: Este es un framework de desarrollo en Python de gran aceptación con una comunidad muy activa. Está implementado con el patrón de desarrollo Modelo, Vista, Controlador (MVC) que permite la separación de las partes de la aplicación basado en la forma en que los usuarios interactúan con la misma.
- HTML, CSS y JavaScript: Este trinomio de tecnologías se han convertido en el estándar del desarrollo web al lado del cliente. Ofrecen la ventaja de estar basados en estándares de la industria, con gran cantidad de librerías y la posibilidad de ser renderizados por casi cualquier dispositivo electrónico que cuente con un navegador de Internet.

Además de las ventajas técnicas que ofrecen, actualmente son tecnologías ya utilizadas a lo interno por el equipo de desarrolladores de IMEDIA.

Esta capa de aplicación cuenta de dos partes importantes:

- Un módulo administrativo, orientado a ser utilizado por analistas de IMEDIA que realizarán el proceso de verificación. En este módulo está destinado a la gestión de todas las noticias que ya han sido procesadas por la solución pero que, para garantizar la mayor fiabilidad de la información, deben ser verificadas y completadas de forma manual.
- El módulo de acceso a clientes: En este módulo se podrán consultar las entidades y noticias y se podrá explorar la relación de cada una, permitiendo que los clientes de IMEDIA puedan realizar búsquedas de personas (por nombre, apellidos, ubicación o DNI) para identificar las noticias relacionadas. A su vez, cada noticia permitirá visualizar las entidades identificadas y verificadas en la misma. Este módulo también es el responsable de generar los correos electrónicos de alertas que llegarán a los clientes en situaciones de importancia o como informes consolidados por períodos de tiempo fijados (cada semana, cada quince días, etc).

### 6.5.6 Visualización de los datos

#### Verificación de relación evento-entidad

**imedia**  
INTELIGENCIA DE MEDIOS

**Atrapan a 'Chino'** 16/Sep/2019 10:39

El detenido fue puesto a disposición de la comisaría de Barboncitos. Las cámaras de seguridad capturaron sus actos delictivos

POR ROBO DE CELULARES Atrapan a 'Chino' Fue intervenido por la policía un sujeto de 24 años, apodado 'Chino' y que, según la policía, se hacía pasar como pasajero para subir a los buses de transporte público y robar celulares a mujeres distraídas, en San Martín de Porres. Luis Tito fue - atrapado por agentes del grupo Terna del Escuadrón Verde cuando descendía de un vehículo con el teléfono de una víctima. Un video de seguimiento lo captó al momento de escapar, a la altura del puente 'Pocitos', en la Panamericana Norte. (J. Y.)

Fujimori realizó crímenes en Perú **La República**

**Entidades identificadas**

**Positivas**

Entidad	% sentimiento	Acciones
Luis Tito	90%	✓ - ✕
Porres	80%	✓ - ✕

**Negativas**



Entidad	% sentimiento	Acciones
Fujimori	100%	✓ - ✕

% sentimiento      Acciones posibles

Siguiete Noticia

Figura 24 - Interfaz para verificación manual

La información de relación de entidades-eventos deberá pasar por una etapa de verificación manual realizada por analistas de IMEDIA. En esta etapa se muestra la noticia en el panel izquierdo, y en la parte derecha las entidades identificadas y su “% de sentimiento”, el analista

debe definir cada entidad como “Positiva” o “Negativa”, además podrá simplemente eliminar la relación en casos donde se generen referencias a entidades distintas de personas. Al presionar el ícono  o el ícono , al usuario le aparecerá una ventana como la siguiente:



**Luis Tito** Tiene relación negativa con noticia

Especifique el resto de información:

DNI

Nombres

Apellidos

Ciudad

Comentario

Si el DNI existe el resto de campos se completan automáticamente

Aceptar Cancelar

Figura 25 - Ingreso de información adicional en relación entidad-evento

En esta interfaz el analista de IMEDIA podrá ingresar información para identificar la entidad y asociarla a una persona, mediante la definición del número del Documento Nacional de Identidad (DNI). Si el DNI se ha ingresado anteriormente, los datos de la personas se completarán automáticamente.

### Buscador de personas



Figura 26 - Buscador de personas

Mediante el buscador, los clientes de IMEDIA podrán consultar los datos ya procesados por la solución y verificados por los analistas. En esta interfaz el cliente puede buscar por número de DNI, nombres o apellidos de la persona.

### **Visualización de resultados**



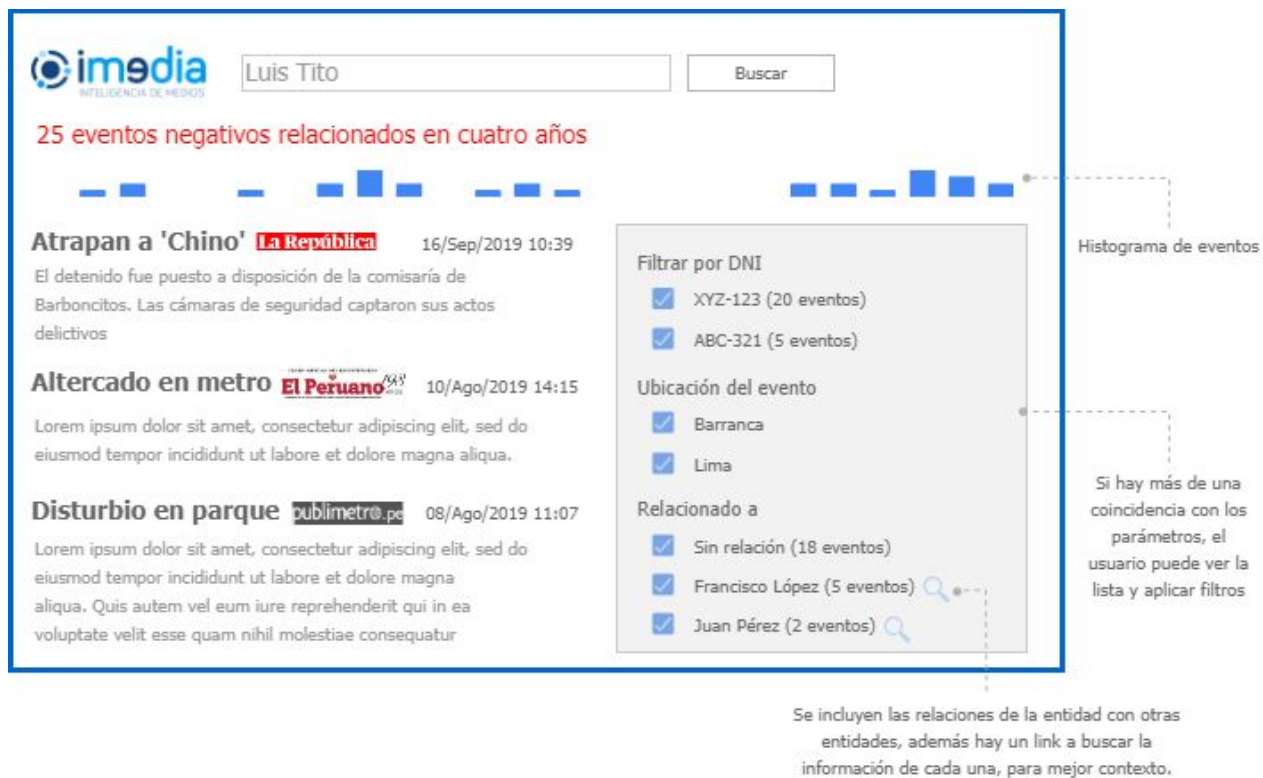


Figura 27 - Resultados de búsqueda de personas

Al desplegar los resultados de las noticias relacionadas a los parámetros de búsqueda, se muestran todas las noticias relacionadas, y además, en la sección de la parte derecha, se muestra una herramienta para filtrar los resultados en caso que se encuentren relaciones con más de una persona.

Las herramientas de filtrado son las siguientes:

- Filtro por DNI: Incluye la lista de los DNI de las entidades principales que coinciden con los parámetros de búsqueda. Al filtrar solo una entidad se le permite al usuario delimitar los resultados encontrados.
- Filtro por ubicación: En caso que no se conozca el DNI de la persona, o que se desee conocer las noticias por ubicación geográfica, se podrá filtrar por las ubicaciones.
- “Relacionado a”: Esta sección permite visualizar las personas relacionadas a las coincidencias encontradas. Al delimitar los resultados por DNI, se podría conocer las personas relacionadas y ayudar en el rastreo de redes de delincuencia.

## Visualizador de eventos

**imedia**  
INTELIGENCIA DE MEDIOS

**Atrapan a 'Chino'** 16/Sep/2019 10:39

El detenido fue puesto a disposición de la comisaría de Barboncitos. Las cámaras de seguridad capturaron sus actos delictivos

POR ROBO DE CELULARES Atrapan a 'Chino' Fue intervenido por la policía un sujeto de 24 años, apodado 'Chino' y que, según la policía, se hacía pasar como pasajero para subir a los buses de transporte público y robar celulares a mujeres distraídas, en San Martín de Porres. **Luis Tito** fue atrapado por agentes del grupo Terna del Escuadrón Verde cuando descendía de un vehículo con el teléfono de una víctima. Un video de seguimiento lo captó al momento de escapar, a la altura del puente 'Pocitos', en la Panamericana Norte. (J. Y.)

**Fujimori** realizó crímenes en Perú **La República**

**Luis Manuel Tito López**  
Lima, Perú DNI: XYZ-123

1 referencia **negativa** en la noticia.

Existen además referencias a las siguientes entidades:

**Negativas**

**Fujimori**

Feedback

El cliente puede dar retroalimentación para conocer si el resultado le ha sido útil.

Periódico fuente

Siguiente Noticia

Figura 28 - Visualización de noticia con entidades

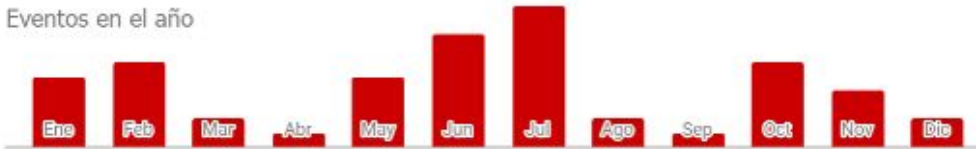
Al presionar cualquiera de los enlaces a las noticias de los resultados del buscador, se mostrará el detalle de misma en dos secciones:

- La sección izquierda mostrará la información de la noticia, y resalta en rojo la entidad que se busca. Además se resaltan las otras entidades identificadas en la noticia.
- La parte derecha mostrará un resumen de la relación entre la entidad buscada y la noticia, además del listado de entidades adicionales en la noticia. El cliente puede dar retroalimentación de los resultados encontrados mediante las opciones de “Me gusta”, “No me gusta” y “Comentario”. La información recopilada de esta retroalimentación será revisada por IMEDIA para garantizar la calidad de la información, además será importante en una etapa futura donde se integren otros mecanismos para automatizar las clasificaciones.

## Notificación por correo

## Cuatro eventos nuevos en los medios el 30 de Marzo 2019

Eventos en el año



Histograma de total de eventos por mes

Listado de eventos

1. **Atrapan a 'Chino'** **La República**
2. **Altercado en metro** **El Peruano**
3. **Disturbio en parque** **publimetro.pe**
4. **Asalto en supermercado** **El Peruano**

Listado de eventos con su fuente

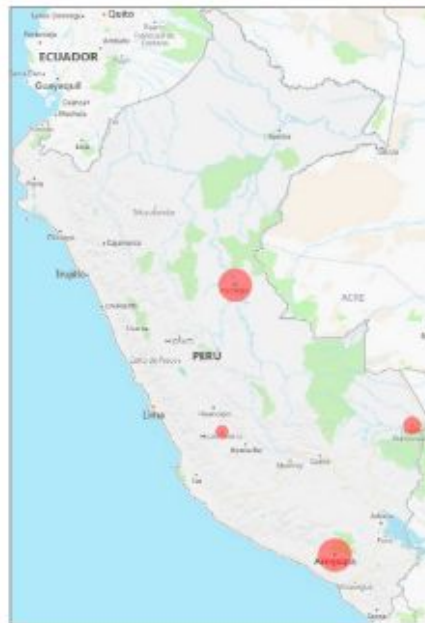
Nuevas entidades:

1. **Luis Manuel Tito López** **15 eventos en la última semana**
2. **Juan Ramiro Pérez** **2 eventos en la última semana**

Entidades identificadas y resumen de eventos recientes

Ubicaciones

- Arequipa
- Pucallpa
- Puerto Maldonado
- Huancavelica



Mapa con indicador de las fuentes con eventos

Figura 29 - Notificación por correo

El cliente recibirá diariamente un resumen de los nuevos hallazgos identificados en las noticias, en dicha notificación se incluirá lo siguiente:

- Cantidad de eventos identificados en el último día.
- Comportamiento de los eventos de noticias relacionadas a fraude del último año.
- Detalle de los eventos identificados, las personas encontradas y la ubicación geográfica de las noticias.

La notificación tiene el objetivo de alertar al departamento de riesgo ante eventos relevantes que deban ser considerados y comunicados a lo interno del banco.

## 7. Optimización de los resultados

### 7.1. Indicadores

Los siguientes indicadores nos ayudan a conocer el estado actual del negocio de nuestra empresa, de cómo ésta va evolucionando, así como de prever potenciales tendencias futuras. Los utilizamos con el fin de seguir el camino adecuado para ir logrando nuestros objetivos:

Objetivo	Estrategia	KPI	Periodicidad
Entender la evolución financiera de la empresa	Estructura general de costes/ingresos	% de aumento de los ingresos	Mensual
		% de disminución de costes	Mensual
	Estructura específica de costes	% de costes fijos sobre los variables	Mensual
Entender el funcionamiento del mercado	Análisis del mercado	% de crecimiento del mercado	Trimestral
		% Cuota de mercado	Trimestral
	Análisis de la competencia	% de crecimiento de la competencia	Trimestral
Adquirir cartera de clientes	Plan de Marketing	ROI campañas de marketing	Trimestral
	Fidelización de clientes	% de retención de clientes	Anual
	Adquisición de nuevos clientes	% de nuevos clientes	Anual
Conocer la situación de nuestros clientes	Satisfacción de los clientes	Nivel de Satisfacción de los clientes	Anual

Calidad del producto/solución ofertada por cliente	Análisis de falsos positivos	% de falsos positivos	Trimestral
	Análisis de falsos negativos	% de falsos negativos	Trimestral
	Transacciones descartadas	% de transacciones descartadas	Mensual

Tabla 5. Indicadores (KPI's)

A continuación, pasamos a explicar cómo calcular cada uno de los KPI's propuestos:

### **% de aumento de los ingresos**

Forma de calcularlo:

$(\text{Ingresos obtenidos en el mes actual} - \text{Ingresos obtenidos en el mes anterior}) / \text{Ingresos obtenidos en el mes anterior}$ .

### **% de disminución de costes**

Forma de calcularlo:

$(\text{Costes del mes pasado} - \text{Costes del mes actual}) / \text{Costes del mes pasado}$ .

### **% de costes fijos sobre los variables**

Forma de calcularlo:

$\text{Costes fijos} / \text{Costes variables}$ .

### **% de crecimiento del mercado**

Forma de calcularlo:

$(\text{Ingresos del mercado en últimos tres meses} - \text{Ingresos del mercado en los anteriores tres meses}) / \text{Ingresos del mercado en los anteriores tres meses}$ .

NOTA: Los ingresos del mercado se pueden consultar vía Internet.

### **% Cuota de mercado**

Forma de calcularlo:

$\text{Ingresos de la empresa} / \text{Ingresos de todo el mercado}$ .

NOTA: Los ingresos del mercado se pueden consultar vía Internet.

## **% de crecimiento de la competencia**

Forma de calcularlo:

(Ingresos de la competencia en los últimos tres meses – Ingresos de la competencia en los anteriores tres meses)/ ingresos de la competencia en los anteriores tres meses.

NOTA: Los ingresos de la competencia de forma genérica, no para una empresa en concreto, se pueden calcular a partir de los ingresos del mercado y de los ingresos de nuestra empresa, de la siguiente manera: Ingresos de la competencia = Ingresos de todo el mercado – Ingresos de nuestra empresa.

## **ROI campañas de marketing**

Forma de calcularlo:

(Beneficio – Inversión realizada en la campaña)/Inversión realizada en la campaña.

NOTA: Para saber si el beneficio obtenido (Contratación por parte de un cliente) es debido a alguna actividad relacionada con las campañas de marketing realizadas, se le preguntará directamente al cliente, como ha conocido nuestra empresa para saber si es por causa de alguna campaña/acción de marketing o no.

## **% de retención de clientes**

Forma de calcularlo:

Número de clientes renovados/número de clientes totales.

## **% de nuevos clientes.**

Forma de calcularlo:

Número de nuevos clientes/número de clientes totales.

## **Nivel de Satisfacción de los clientes**

Forma de calcularlo:

Realización de encuestas.

## **% de falsos positivos**

Forma de calcularlo:

Número de falsos positivos/Número total de instituciones o personas analizados.

## **% de falsos negativos**

Forma de calcularlo:

Número de falsos negativos/Número total de instituciones o personas analizados.

## **% de transacciones útiles**

Forma de calcularlo:

Porcentaje de transacciones marcadas como “útiles” por el cliente, utilizando la herramienta de retroalimentación.

## **7.2. Análisis Financiero**

A continuación mostramos el análisis financiero para poder ver la viabilidad económica de Inteligencia Financiera y con el fin de prever puntos que sea necesario reforzar con el suficiente tiempo antes de la puesta en marcha.

### **7.2.1 Cuenta de resultados**

Puesto que el proyecto de Inteligencia Financiera de IMEDIA es una extensión de los actuales servicios de la compañía, hace que la complejidad de la estructura financiera de la misma sea muy simple y sin necesidad de financiación, nada más un aporte como saldo inicial de 10 mil Euros por parte de los socios.

En esta cuenta estamos considerando los ingresos mostrados en el apartado 2.1.9 considerando como crecimiento 2 clientes más por año.

Tenemos la ventaja de que el crecimiento de clientes no es directamente proporcional al de los costos, ya que consiste en replicar el servicio a los nuevos sin tener importantes incrementos de gastos, aun así, estamos considerando un 2% de incremento en los costes por año, siendo éstos básicamente de Estructura.

Para ser competitivos, no estamos proyectando el incremento en el precio del servicio los primeros 5 años.

En cuanto a la amortización, solo lo aplicamos al Hardware utilizado.

No tenemos intereses involucrados por no tener financiamiento, y el impuesto sobre la renta a utilizar será del 29.5%.

## **Ingresos Proyectados**

Concepto	Año 1	Año 2	Año 3	Año 4	Año 5
Número de Clientes	2	4	6	8	10
Precio anual del servicio	30.000 €	30.000 €	30.000 €	30.000 €	30.000 €
<b>Total ingresos por Ventas</b>	<b>60.000 €</b>	<b>120.000 €</b>	<b>180.000 €</b>	<b>240.000 €</b>	<b>300.000 €</b>

Tabla 6. Ingresos Previstos Inteligencia Financiera IMEDIA

## Costes Projectados

Concepto	Año 1	Año 2	Año 3	Año 4	Año 5
Renta de oficinas	6.000 €	6.120 €	6.242 €	6.367 €	6.495 €
Sueldos y Salarios	6.500 €	6.630 €	6.763 €	6.898 €	7.036 €
Asesoría Contable y Fiscal	2.000 €	2.040 €	2.081 €	2.122 €	2.165 €
Comunicaciones	5.000 €	5.100 €	5.202 €	5.306 €	5.412 €
Cloud usage y Servicio AWS	14.000 €	14.280 €	14.566 €	14.857 €	15.154 €
Equipo de Programadores	41.500 €	- €	- €	- €	- €
Mantenimiento e Incidencias	- €	5.500 €	5.500 €	5.500 €	5.500 €
<b>Suma</b>	<b>75.000 €</b>	<b>39.670 €</b>	<b>40.353 €</b>	<b>41.050 €</b>	<b>41.761 €</b>

Tabla 7. Costes Previstos Inteligencia Financiera IMEDIA

## Amortización

Concepto	Inversión	% Anual	Valor Anual
Hardware	3.000 €	25%	750 €

Tabla 8. Cuadro de Amortización Inteligencia Financiera IMEDIA

## Cuenta de Resultados

Concepto	Año 1	Año 2	Año 3	Año 4	Año 5
Total ingresos por Ventas	60.000 €	120.000 €	180.000 €	240.000 €	300.000 €
Gastos por Estructura	75.000 €	39.670 €	40.353 €	41.050 €	41.761 €
Margen Operativo (EBITDA)	- 15.000 €	80.330 €	139.647 €	198.950 €	258.239 €
Amortización	750 €	750 €	750 €	750 €	- €
EBIT	- 15.750 €	79.580 €	138.897 €	198.200 €	258.239 €
Intereses	- €	- €	- €	- €	- €
EBT	- 15.750 €	79.580 €	138.897 €	198.200 €	258.239 €
Impuestos	- €	23.476 €	40.974 €	58.469 €	76.180 €
<b>BENEFICIO NETO</b>	<b>-15.750 €</b>	<b>56.104 €</b>	<b>97.922 €</b>	<b>139.731 €</b>	<b>182.058 €</b>
<b>MARGEN DE BENEFICIO</b>	<b>-26,25%</b>	<b>46,75%</b>	<b>54,40%</b>	<b>58,22%</b>	<b>60,69%</b>

Tabla 9. Cuenta de Resultados Inteligencia Financiera IMEDIA

Como vemos la cuenta de resultados, el proyecto es rentable desde el segundo año.



## 7.2.2 Flujo de Caja

Este Flujo de Caja nos muestra una proyección de los movimientos de efectivo pronosticados durante la puesta en marcha de proyecto; un dato sumamente importante en la toma de decisiones para la realización del proyecto.

Concepto	Año 1	Año 2	Año 3	Año 4	Año 5
Aporte de los socios	10.000 €	- €	- €	- €	- €
Saldo Inicial	10.000 €	8.000 €	48.854 €	147.526 €	288.007 €
Total ingresos por Ventas	60.000 €	120.000 €	180.000 €	240.000 €	300.000 €
Total Ingresos	60.000 €	120.000 €	180.000 €	240.000 €	300.000 €
Hardware	3.000 €	- €	- €	- €	- €
Total Inversiones	3.000 €	- €	- €	- €	- €
Gastos por Estructura	75.000 €	39.670 €	40.353 €	41.050 €	41.761 €
Impuestos	- €	23.476 €	40.974 €	58.469 €	76.180 €
Total Gastos	75.000 €	63.146 €	81.328 €	99.519 €	117.942 €
<b>SALDO FINAL</b>	<b>- 8.000 €</b>	<b>48.854 €</b>	<b>147.526 €</b>	<b>288.007 €</b>	<b>470.065 €</b>

Tabla 10. Flujo de Caja Inteligencia Financiera IMEDIA

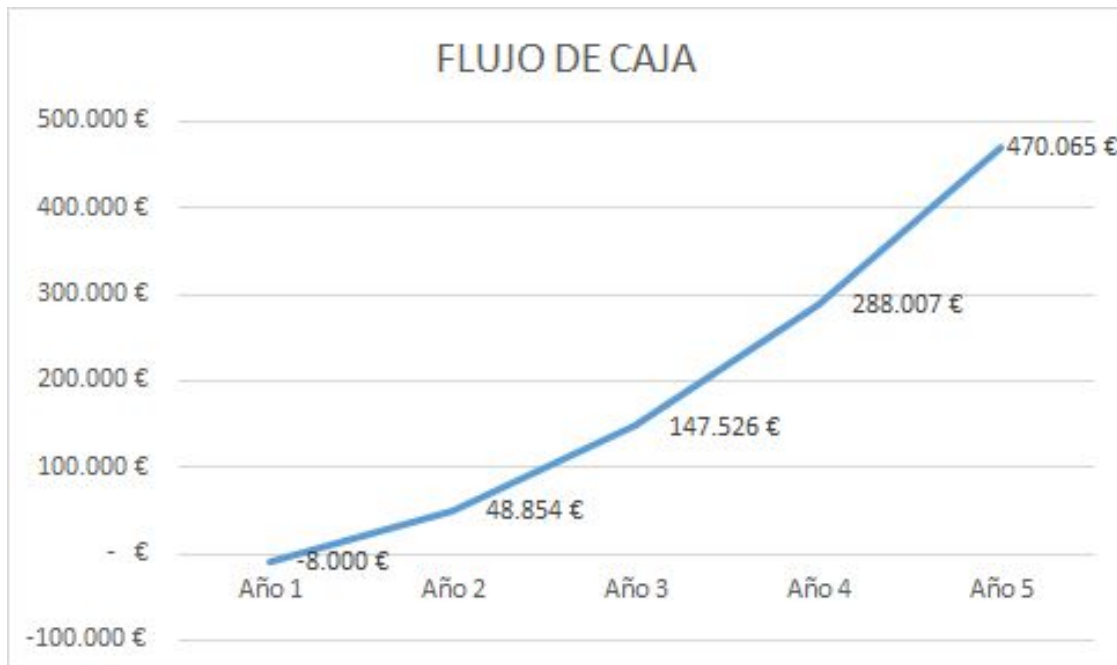


Figura 30. Proyección del Flujo de Caja Inteligencia Financiera IMEDIA

### 7.2.3 Balance de situación

El presente balance es una proyección únicamente basada en los datos del proyecto Inteligencia Financiera de IMEDIA, es decir, no representa el balance de la compañía en global.

ACTIVO	Año 1	Año 2	Año 3	Año 4	Año 5
Inversión	3.000 €	3.000 €	3.000 €	3.000 €	3.000 €
Amortización	- 750 €	- 1.500 €	- 2.250 €	- 3.000 €	- 3.000 €
Activo No Corriente	2.250 €	1.500 €	750 €	- €	- €
Efectivo	- 8.000 €	48.854 €	147.526 €	288.007 €	470.065 €
Activo Corriente	- 8.000 €	48.854 €	147.526 €	288.007 €	470.065 €
<b>TOTAL ACTIVO</b>	<b>- 5.750 €</b>	<b>50.354 €</b>	<b>148.276 €</b>	<b>288.007 €</b>	<b>470.065 €</b>
PASIVO	Año 1	Año 2	Año 3	Año 4	Año 5
Capital	10.000 €	10.000 €	10.000 €	10.000 €	10.000 €
Reservas	- €	15.750 €	40.354 €	138.276 €	278.007 €
Resultado del Ejercicio	- 15.750 €	56.104 €	97.922 €	139.731 €	182.058 €
Patrimonio Neto	- 5.750 €	50.354 €	148.276 €	288.007 €	470.065 €
<b>TOTAL PASIVO</b>	<b>- 5.750 €</b>	<b>50.354 €</b>	<b>148.276 €</b>	<b>288.007 €</b>	<b>470.065 €</b>

Tabla 11. Balance de Situación Inteligencia Financiera IMEDIA

### 7.2.4 Rentabilidad

A continuación, vamos a definir los indicadores de rentabilidad que vamos a medir:

- VAN (Valor Actual Neto): es un criterio de inversión que consiste en actualizar los cobros y pagos de un proyecto o inversión para conocer cuánto se va a ganar o perder con esa inversión. Hemos asumido una tasa de descuento del 10% para efectos de nuestro análisis.
- TIR (Tasa Interna de Rentabilidad): es la tasa de interés o rentabilidad que ofrece una inversión. Es decir, es el porcentaje de beneficio o pérdida que tendrá una inversión para las cantidades que no se han retirado del proyecto.
- TIRM (Tasa Interna de Rentabilidad Modificada): propone la definición de proyectos de inversión con dos únicos flujos de caja. En el momento inicial, los desembolsos y en el momento final, los cobros. Hemos asumido que la tasa de financiación es del 0% y la de reinversión del 10%.
- Pay-Back (Plazo de Recuperación): es un criterio para evaluar inversiones que se define como el periodo de tiempo requerido para recuperar el capital inicial de una inversión. Es un método estático para la evaluación de inversiones.

<b>TIR</b>
<b>777%</b>
<b>VAN</b>
<b>632.527 €</b>
<b>TIRM</b>
<b>237%</b>
<b>PAYBACK</b>
<b>1,114</b>
<b>Después del 1er año</b>

Tabla 12. Rentabilidad

## 8. Conclusiones

Mediante la investigación, análisis y prototipado de tecnologías de Business Intelligence y Big Data, y el aprendizaje adquirido durante los módulos del Máster, hemos podido profundizar y entender mejor los casos de uso de estas tecnologías, así como las ventajas que representan para la optimización de procesos y el descubrimiento de datos e indicadores que, sin ellos, es difícil obtener. En el particular del NLP, hemos identificado una gran oportunidad de integrarlo en más procesos del día a día de IMEDIA.

A su vez, hemos identificado oportunidades de mejora a futuro, como experimentar con tecnologías de machine learning en el tema de la asociación de entidades a noticias y la disminución de falsos positivos y falsos negativos, esta es una de las razones por la hemos integrado opciones de retroalimentación por parte del cliente en el concepto del desarrollo del producto.

El proceso de desarrollo de este proyecto ha sido además un buen aprendizaje sobre metodologías de trabajo en equipo, considerando además que cada integrante de este grupo vive en un país distinto.

## 9. Bibliografía

VAN:

<https://economipedia.com/definiciones/valor-actual-neto.html>

TIR:

<https://economipedia.com/definiciones/tasa-interna-de-retorno-tir.html>

TIRM:

<https://raimon.serrahima.com/tirm-tasa-rendimiento-interno-modificada/>

Payback:

<https://economipedia.com/definiciones/payback.html>

<https://excelforo.blogspot.com/2016/04/calculo-del-payback-o-del-periodo-de.html>

Cálculo VAN, TIR:

<https://www.zonaeconomica.com/excel/van-tir>

Impuestos:

<http://orientacion.sunat.gob.pe/index.php/empresas-menu/impuesto-a-la-renta-empresas/regimen-general-del-impuesto-a-la-renta-empresas/calculo-anual-del-impuesto-a-la-renta-empresas/2900-03-tasas-para-la-determinacion-del-impuesto-a-la-renta-annual>

Referencia Asbanc:

<https://www.asbanc.com.pe/Paginas/Conocenos/Conocenos.aspx>

Referencia Reniec:

<https://www.reniec.gob.pe/portal/institucional.htm#>

Referencia SBS:

<http://www.sbs.gob.pe/acercadelasbs>

Lambda y Kappa:

<https://unpocodejava.com/2013/09/07/arquitectura-lambda-principios-de-arquitectura-para-sistemas-big-data-en-tiempo-real/>

<http://milinda.pathirage.org/kappa-architecture.com/>

<http:// analisisdedatos.net/bigData/eco/lambdaKappa.php>

<https://blog.gft.com/es/2017/12/05/que-es-una-arquitectura-de-referencia-de-datos-y-que-utilidad-tiene/>

<https://www.paradigmigital.com/techbiz/de-lambda-a-kappa-evolucion-de-las-arquitecturas-big-data/>

<https://www.blue-granite.com/blog/a-different-way-to-process-data-kappa-architecture>

<https://towardsdatascience.com/a-brief-introduction-to-two-data-processing-architectures-lambda-and-kappa-for-big-data-4f35c28005bb>

Change Data Capture (CDC):

<https://vladmihalcea.com/a-beginners-guide-to-cdc-change-data-capture/>

<https://medium.com/blablacar-tech/streaming-data-out-of-the-monolith-building-a-highly-reliable-cdc-stack-d71599131acb>

Debezium:

<https://iamninad.com/how-debezium-kafka-stream-can-help-you-write-cdc/>

<https://debezium.io/docs/faq/>

<https://rmoff.net/2018/03/27/streaming-data-from-mongodb-into-kafka-with-kafka-connect-and-debezium/>

Kafka:

<https://kafka.apache.org/uses>

<https://dzone.com/articles/mongodb-amp-data-streaming-implementing-a-mongodb>

<https://sookocheff.com/post/kafka/kafka-in-a-nutshell/>

Spark Streaming:

<https://spark.apache.org/docs/latest/streaming-programming-guide.html>

<https://dzone.com/articles/lambda-architecture-with-apache-spark>

<https://databricks.com/glossary/what-is-spark-streaming>

<https://www.ververica.com/what-is-stream-processing>

<https://www.sigmoid.com/spark-streaming/>

<https://www.sigmoid.com/getting-data-into-spark-streaming-2/>

<https://www.sigmoid.com/spark-streaming-internals/>

<https://blog.cloudera.com/blog/2014/08/building-lambda-architecture-with-spark-streaming/>

ElasticSearch:

<https://www.elastic.co/products/elasticsearch>

[https://www.elastic.co/guide/en/elasticsearch/reference/6.1/the\\_search\\_api.html](https://www.elastic.co/guide/en/elasticsearch/reference/6.1/the_search_api.html)

<https://www.atlantic.net/hipaa-compliant-database-hosting/elasticsearch-distributed-nosql-database/>

<https://dzone.com/articles/what-is-elasticsearch-and-how-it-can-be-useful>

[https://db-engines.com/en/ranking\\_trend/search+engine](https://db-engines.com/en/ranking_trend/search+engine)

<https://qbox.io/blog/what-is-elasticsearch>

<https://logz.io/blog/solr-vs-elasticsearch/>

<https://www.elastic.co/blog/how-the-world-is-using-elasticsearch>

## Anexos

### Entrevista a ASBANC

Atendida por Patricia Mongrut (Supervisor de Investigación de Estadísticas y Estudios Económicos).

1. Existe una alta cantidad de transacciones financieras de carácter fraudulento en Perú.

Patricia:

El fraude tiene varias aristas y deben de definir a qué tipo de fraude se refiere el objeto de estudio. Los fraudes más comunes son el kiting (ocultamiento de efectivo), la falsificación, el cambio de beneficiarios y el blanqueo de dinero.

- a. ¿Qué porcentaje de las transacciones pueden ser fraudulentas o asociadas a personas de poca fiabilidad?

Patricia:

Aproximadamente el 2% de las entidades bancarias reportan transacciones que están asociadas a fraude de diferente tipo.

- b. ¿Cuántos miles o millones de soles puede perder un banco por este tipo de fraude?

Patricia:

Esa cifra no te la puedo dar porque dependerá del banco y no podemos brindar cifras de nuestros agremiados.

2. Los medios de comunicación locales y regionales adoptan un papel fundamental en la lucha contra el crimen organizado.

a. ¿Tienen algún procedimiento para contrarrestar los fraudes?

Patricia:

Sí, tenemos una serie de medidas que desplegamos para disminuir los fraudes. Como por ejemplo la realización de auditorías, la rotación laboral, controles de acceso a la información privilegiada, procedimientos éticos laborales y empresariales, revisar antecedentes de personal y los controles periódicos y sorpresivos que han dado resultados increíbles sobre nuestros agremiados.

b. ¿Tienen algún procedimiento asociado a obtener información de los medios de comunicación?

Patricia:

Sí, contamos con informes diarios de IMEDIA sobre los principales hechos noticiosos del país. Además, tenemos un informe diario sobre las personas y organizaciones involucradas con hechos delictivos relevantes a la finanzas y banca en general.

c. ¿Qué delitos en particular asociados a activos/pasivos financieros le interesa rastrear?

Patricia:

Lo que primordialmente buscamos es lavado de dinero, estafa, narcotráfico y robo sin embargo entendemos que existen otros delitos que preocupan a nuestro sector como por ejemplo el asesinato a cambistas, el raqueteo (asalto al paso cerca de las entidades financieras o bancarias), el marcaje (asalto después de haber retirado dinero de los bancos).



3. Sobre el desconocimiento de parte de las entidades financieras referente a la viabilidad de otorgar créditos a clientes involucrados con hechos delictivos.

a. ¿Existe un desconocimiento total o parcial de parte de las entidades financieras para otorgar créditos a personas involucradas con hechos delictivos?

Patricia:

Solo es parcial, pues todas las entidades financieras cuentan con centrales de riesgo que posibilitan una detección de personas preventiva de personas con comportamiento impago de sus obligaciones financieras.

Sin embargo, dichas centrales de riesgo no están conectadas con el poder judicial y la policía nacional del Perú que permita conocer los hechos de gran trascendencia para la ciudadanía.

4. Sobre el desconocimiento de parte de las entidades financieras de guardar el dinero procedente de hechos ilícitos.

a. ¿Existe sanción alguna para las entidades financieras que reciben dinero procedente de hechos ilícitos?

Patricia:

Definitivamente sí, y es de gran preocupación para nuestros agremiados pues el estado peruano efectúa las demandas a través de la Fiscalía Nacional de la República acusando de blanqueo de fondos.

En muchas ocasiones son acusados y es por desconocimiento de la persona u organización, hecho que compromete seriamente a nuestros agremiados.

b. ¿Existe un desconocimiento total o parcial de parte de las entidades financieras para recibir depósitos de personas involucradas con hechos delictivos?

Patricia:

Es solo parcial porque nosotros contamos con una base procedente de cada banco (lista negra) en donde imputamos si la persona u organización ha estado involucrado en hechos delictivos.

5. La predisposición de las entidades financieras a la utilización de herramientas tecnológicas con el fin de mejorar su toma de decisiones para los clientes reales y potenciales.
  - a. ¿Existe o existiría la predisposición de parte de las entidades financieras a adquirir un sistema que permita conocer las personas u organizaciones involucradas con hechos delictivos?

Patricia:

Sí, de hecho, lo conversamos con Uds. (IMEDIA) pues gran parte de nuestros agremiados piden una solución de este tipo, pero no tenemos los datos noticiosos de todo el Perú. Y los bancos no cuentan con dichos datos en tiempo real.

## Entrevista a Financiera Ohh.

Atendida por Mirko León (Gestor de riesgos financieros).

1. Existe una alta cantidad de transacciones financieras de carácter fraudulento en Perú.

Mirko:

Primero es recomendable que se defina bien a que se refiere con fraude. Fraude puede ser cuando clonan las tarjetas de crédito y hacen cargos a la tarjeta que no son reconocidas por el cliente. Así también puede llamarse fraude a los préstamos que hace un cliente pero que nunca se pagan. A este hecho se le llama First Payment Default.

- a. ¿Qué porcentaje de las transacciones pueden ser fraudulentas o asociadas a personas de poca fiabilidad?

Mirko:

Aproximadamente el 1.5% de las transacciones en nuestra financiera son de carácter fraudulento.

- b. ¿Cuántos miles o millones de soles puede perder un banco por este tipo de fraude?

Mirko:

Dependiendo del tamaño de operación de cada banco, este podrá perder entre 2 millones a 6 millones de soles.

2. Los medios de comunicación locales y regionales adoptan un papel fundamental en la lucha contra el crimen organizado.

- a. ¿Tienen algún procedimiento para contrarrestar los fraudes?

Mirko:

Sí, las entidades financieras son normadas por la Superintendencia de banca y seguro (SBS). Es esta entidad quien despliega entre 3 y 4 capacitaciones al año para poder detectar posibles fraudes.

- b. ¿Tienen algún procedimiento asociado a obtener información de los medios de comunicación?

Mirko:

Entiendo que la SBS lo hace y que lo comparte con cada entidad financiera. Pero solo lo hace para políticos y personas que están envueltas ya en escándalos de corrupción.

- c. ¿Qué delitos en particular asociados a activos/pasivos financieros le interesa rastrear?

Mirko:

Robo, sicariato, narcotráfico, lavado de dinero, venta ilegal, estafa, tala ilegal, minería ilegal y otros más que no recuerdo.

3. Sobre el desconocimiento de parte de las entidades financieras referente a la viabilidad de otorgar créditos a clientes involucrados con hechos delictivos.

- a. ¿Existe un desconocimiento total o parcial de parte de las entidades financieras para otorgar créditos a personas involucradas con hechos delictivos?

Mirko:

Existe un desconocimiento parcial pues algunos bancos se conectan con el poder judicial para obtener el registro de personas u organizaciones que llevan procesos civiles y penales en curso.

Sin embargo, no existe una base de datos que permita prever si el cliente podrá ser un buen o mal cliente.

Solo podemos dar o no crédito según los datos provistos por las centrales de riesgos existentes. (Equifax, Sentinel, SBS).

4. Sobre el desconocimiento de parte de las entidades financieras de guardar el dinero procedente de hechos ilícitos.

- a. ¿Existe sanción alguna para las entidades financieras que reciben dinero procedente de hechos ilícitos?

Mirko:

Primero que nada, las entidades financieras están reguladas por la SBS y si las entidades financieras reciben dinero procedente de hechos ilícitos, estas son multadas y sancionadas por esta entidad reguladora por blanqueamiento de fondos.

- b. ¿Existe un desconocimiento total o parcial de parte de las entidades financieras para recibir depósitos de personas involucradas con hechos delictivos?

Mirko:

Es parcial pues ya recibimos datos del poder judicial y de la SBS sin embargo estos datos son solo de personajes políticos o empresarios famosos.

5. La predisposición de las entidades financieras a la utilización de herramientas tecnológicas con el fin de mejorar su toma de decisiones para los clientes reales y potenciales.

- a. ¿Existe o existiría la predisposición de parte de las entidades financieras a adquirir un sistema que permita conocer las personas u organizaciones involucradas con hechos delictivos?

Mirko:

Sí, sería interesante tener una herramienta. Sin embargo sería una herramienta complementaria a los datos provistos por la SBS y las centrales de riesgo actual.

## Entrevista a Banco de Crédito BCP

Atendida por José Marangunich, Gerente de Compliance.

1. Y del 100% de transacciones del BCP se puede decir 0.01% es fraudulento?

José:

Hasta menos, es un ratio de clase mundial. Estamos en un punto base.

2. Y de cuantas transacciones estaríamos hablando, mensual, si se puede saber? ese 0.01% involucra una fuerte suma de dinero?

José:

Son 100 millones de transacciones por mes.

No, como tenemos buenos sistemas de monitoreo, una cosa es el phishing, lo que tu escuchas, otra cosa es el fraude bancario, no es lavado de dinero, yo te hablo de fraude.

3. Pero si sabes datos del lavado de dinero?

José:

Sí, pero no soy el ente autorizado para darte los números correctos. Es más en lavado de dinero he estado una vez hace tiempo, entonces si es que el lavado de dinero uno lo conoce sin duda, pero no lo veo, no está en mi gestión hoy. Es más, en la asociación de bancos donde tenemos comités allí tampoco creo que está separado un comité de lavado de dinero y otro comité de fraudes bancarios. En ese caso esto es cumplimiento o compliance.

4. Compliance ósea Entonces no se puede perder ni 1,000,000 de soles ni 100,000 dólares nada, entonces no se pierde nada por fraude?

José:

No, no tampoco he dicho eso. Seguro se pierde, pero este digamos lo que se pierde, no pierde el cliente. A ver en la banca no existe fraude cero no hay riesgo cero es una actividad de riesgo diferentes diferentes lo principal es el crediticio. No es cierto, pero también hay entes más fraudulentos, entonces yo te invito a que mires los reportes de Visa y los reportes de American Express sobre cómo son los números en el mundo sobre el tema del fraude y donde normalmente donde más toma segura es en la modalidad del comercio electrónico. Comercio electrónico así es, pero comercio electrónico tiene sus propias reglas quiere decir que tú puedes contra cargar la operación, ósea el cliente puede decir yo no lo reconozco, tú se lo contra cargas. Y en el comercial si tú has vendido más de esa cosa entonces no la reconozco te la devuelvo y devuelvo la plata inmediatamente. Pero el 85 o el 80% del fraude en el mundo en tarjeta o es por comercio electrónico. O sea fuerte el porcentaje de la tipología, pero en cuantía en Perú no porque este país no está muy desarrollado el comercio electrónico.

5. Y qué acerca de esas personas que por ejemplo estos sacan un crédito y nunca lo pagan?

José:

Aquí hay 2 escenarios tú vas a tener el que no lo paga de pronto porque lo despidieron o de lo contrario simulan quieren aprovecharse de esa circunstancia cierto, allí es donde haces análisis si es que es riesgo de crédito o finalmente riesgo operativo fraude en originación. Ya hay modelos que se usan para detectar fraude originación y nosotros lo utilizamos no, pocos bancos creo que en el Perú utilizan sistemas técnicos no.

Lo que haces es que tienes una serie de variables en el cual empiezas a entender digamos donde debes apuntar tus... lo cual no significa que no hay, seguro que hay gente por allí la pasa como tema de crédito se lo reporta cosa por el estilo y si es fraude se le denuncia y se le pone en un archivo de activo.

6. La segunda hipótesis que tengo es: Que los medios de comunicación locales y regionales adoptan un papel fundamental en la lucha contra el crimen organizado asociado evidentemente al fraude y a los delitos financieros.

José:

A ver, aquí hay varias frases en tu hipótesis haber primero estás hablando de crimen organizado ya entonces y hablas de fraude yo lo que te voy hablar es técnicamente.

7. Cómo infieren si ustedes usan los medios de comunicación o de pronto saben que no miramos o no?

José:

Si escuchamos nuestra obligación es escuchar los medios de comunicación, pero creo que los medios de comunicación lo que más presentan son casuística, el caso de un cliente, caso de un reclamante, alguien que no escucharon un caso así ahh pero normalmente no veo que se vean actividades de crimen organizado el día que la prensa empiece a tomar una acción más del orden preventivo, llámelo así más de reactivo por impulso ante un riesgo llamémoslo así.

8. Por ejemplo cuando llamen al RPP llaman y dicen que esa persona está haciendo tala ilegal de árboles y ese dinero se va al Banco del Crédito de la región. Es un poco el escenario que te estoy planteando para que se entienda, o pide prestado para maquinarias, o pide prestado un crédito para comprar maquinaria para ese hecho que va hacer.

José:

Es otro lavado de dinero, es fraguado con otra persona, yo te voy hablar del fraude que tú me has dicho. Por eso cuando tú hablas de fraude y hablas de crimen organizado está hablando de dos cosas, estás hablando de una condición de fraude que tiene que estar en un súper escenario donde tienes crimen organizado probado



no estás hablando el que aisladamente pide o saca un crédito no paga eso no es crimen organizado ya, crimen organizado son bandas muy grandes que están en sistemas o redes. La prensa no te ayuda a detectar bandas de crimen organizado, sin embargo, sí hacen difusión segura, de noticias de seguridad y seguridad ciudadana. Ok pero crimen organizado bancario no, lo que hace es darle cobertura capturaron a una banda de clonadores.

Le está dando publicidad porque la policía llama y la policía le dice tengo esta noticia y voy a presentar y voy a dar conferencia de prensa. Eso funciona así, pero no es que los investigadores o periodistas de investigación como sí ocurre en el campo político, están investigando el crimen organizado en tema de fraudes bancarios, no hay, yo no he visto no conozco.

9. Hay mucha especulación en los periodistas nada más en los regionales no sé si ha escuchado, noticias de regionales

José:

La verdad que no mucho, hay entes de seguridad que me reportan a mí que tienen presencia allá, este entonces seguro sí temas todo lo regional termina siendo más que la persona tal o algo, pero no crimen organizado.

El crimen organizado más que nada en varias partes de Latinoamérica son como bandas internacionales, entonces podrías tú sí entender banda del país del norte vecino del Perú, o de otro lado por allí entonces en determinado cual nacionalidad. Por ejemplo, cuando esto y si está claro cuando detuvieron a esa gente del tren de Aragua no es que hubo investigación y eso si era crimen organizado, es una organización de 100,000 integrantes en la policía y que son muy feroces, algo como las maras y vinieron acá comenzaron a delinquir y no sé si se han quedado no lo sé tampoco los puedo descartar. Comprendes.

El otro día habían detenido a 6 tipos venezolanos con armas y cosas por el estilo no sabes si son del mismo clan o no, pero a lo que iba es que esto sale como consecuencia de la conferencia de prensa.

10. Labor de inteligencia preventiva para usted claro.

José:

Preventiva del periodismo. No viene por allí, con el tema cibernético por ejemplo si nosotros recibimos más que nada información, pero de empresas en el mundo que dedican a ver este tema como foros de hackers, sí, pero no por la prensa.

11. Pero si tienen un procedimiento bastante claro respecto?

José:

Más sin embargo sí creo que la prensa es un cuarto poder el día que lo tenga va a ser una oportunidad interesante cuando lo haga bien no sensacionalista mente porque al final lo que tenemos que decir como ligo esta noticia qué parte de la noticia es realidad y qué parte es la ficción aumentada me comprendes. Entonces en la medida que tu encuentres que se tenga un centro de comunicación con información bien cuadrada, bien sentada, en otras palabras, información de valor. Que no sea que te comuniquen en conferencia de prensa, sino que te comuniquen elementos en donde ellos creen que se está trabajando estos riesgos por ejemplo no, qué pasa si empiezan hacer un trabajo con respecto a la presencia de colombianos cerca de la estación de los buses que hay un tema de colombianos que algunos se quedan allí que están haciendo estos préstamos esa es una labor que deben estar haciendo ellos. Lo que están haciendo en Ecuador en Tube, con colombianas si pasan o no pasan o venezolanos que se han quedado en Ecuador o Tube, y que está pasando por allí me comprendes es cuestión de investigación no es que detuvieron a... y ya me dan una noticia y conferencia de prensa y sobre eso yo empiezo a escribir.

No hay un periodismo que yo conozca especializado, en tema de riesgo de seguridad financiera, de repente, no sé si me entiende.

Entonces el día que exista, o que exista una central que tenga información que sea de valor, va hacer de especial interés, para esto habrá que tener sistemas internos, de parametrización de riesgos, pesos en el modelo que te diga esto es un escenario

crítico, este no, este es una anécdota de una noticia aislada, este no. Cuando empieces haber todas estas noticias que coinciden, orientación x cosa, este es un foco de investigación. Entonces allí lo va hacer, pero va a tener que tener cosas como modelo de análisis.

12. Ahora con respecto al procedimiento de contrarrestar el fraude, ustedes tienen procedimientos estrictos en cuanto a la determinación contrarrestar los fraudes bancarios?

José:

Eso es una pregunta?

Sí hay, se tiene mucho sistema de inteligencia contratado, lo cual vuelvo a decir no significa que no pueda irse una u otra operación, por qué, porque la nueva tipología de fraude vamos hablar del mundo de las tarjetas por ejemplo básicamente va por la suplantación de identidad, alguien que dice soy cliente y opero con el nombre de él físicamente o virtualmente lo que pasa es que ya se busca, se trata de mimetizar con el comportamiento al cliente. Entonces te pongo un ejemplo la gente piensa que un fraude bancario es responsabilidad del banco que el banco ha sido hackeado. El no entiende el que ha sido hackeado es él su máquina que tiene phishing que tiene el troyan. O tiene su teléfono celular sin antivirus yo te aseguro el 98% o 99% de los teléfonos del mercado nacional están sin antivirus porque si así lo venden? Claro que quieres ir a una marca no es necesario antivirus, o sea como asunto de speech de posicionamiento vacan. Pero yo quiero comprar y vender entonces la gente todavía no asocia el robo de información con el dispositivo esté por allí esta. Hay un tema de educación y un tema del esquema de valoración de riesgo. Nosotros hicimos en diciembre una campaña que se llama: “Juntos somos responsables” porque si es cierto invertimos en seguridad, sistema de monitoreo, sistema de inteligencia todo eso debe acompañarse con actividades seguras por parte del cliente, la pregunta y el ejemplo es bien sencillo, si yo salgo de mi casa para trabajar, si yo abro la puerta de mi casa cuando salgo, la respuesta es sí, entonces si eso es así porque no tomo las mismas seguridades en tu teléfono o con el dispositivo donde está tu información. Es más, la gente piensa ahora le roban información no firme no me robaron. Ya le robaron la información. Entonces allí hay un tema un

proceso de auto concientización porque los bancos estamos haciendo esfuerzos, ahora entras a cualquier web dice campo de seguridad consejos que estén enterados y entonces sí tiene que ver que pase algo fuerte para que digas tomó conciencia este concepto de juntos somos responsables.

13. Es muy bueno perfecto ahora otro tema. Qué delitos en particular están interesados en seguir en el caso que evidentemente opten por seguir las noticias asociadas?

José:

O sea perdón, la respuesta interesado en general es todo, no es cierto ese es el primer punto pero el orden de prioridades de la atención, ese es distinto nosotros vamos a priorizar por criticidad, es decir todo aquello que vaya contra la intimidad del cliente o información general debe ser más protegido; el otro lado es, ya tienes un mundo muy importante el mundo de la tarjeta bancaria después también es interesante ver el otro es el mundo de la corrupción es interesante porque alrededor de la corrupción vienen muchas cosas asociadas entonces en los bancos por ejemplo suele ser muy sensible este tipo de situaciones.

14. Cómo se enteran ustedes de un hecho de corrupción? Si no es por el periodismo, es viable o tienen algún mecanismo?

José:

Sí claro empresas que tú le dices hágame screening de tal persona ellos escanean. Hay empresas de fuera que buscan en fuentes públicas en google, sentencias publicadas por cosas.

15. Por ejemplo con eso de sentencias publicadas ustedes tienen conexión con el poder judicial para notificar los hechos delictivos por las faltas que haya caído un ciudadano?

José:

hay empresas que se dedican a ver solo sentencias publicadas y cosas por el estilo y llenan esa información y ese servicio tú lo puedes contratar. Se dedican a ver eso o también fallecidos.

16. Conoce la empresa de casualidad?

José:

Hay unas, pero tenemos limitante de decir.

17. Y para dar ese préstamo no se tiene ese mismo nivel para darle prestada plata a un seudo empresario?

José:

Ingresas archivo negativo si esta en archivo negativo no se da. Como te digo allí viene la diferencia también si es sentenciado o no sentenciado o depende las condiciones de ley el tema del expediente no del expediente me refiero a la noticia. No que leen los expedientes sino la noticia asociados a sentencias.

18. Nosotros le vendemos un poco este rigió es el logo de la nación porque ellos querían un reporte de todos los hechos delictivos asociados a personas categorizamos esas personas que están asociadas a corrupción, robo, lavado de dinero todos los delitos. Nosotros tenemos que analizar los textos de cada noticia y si es que esa persona estaba identificada

deben de colocar su número de DNI y si es una organización su rubro entonces tipificamos y hacemos un repositorio de datos eso era un poco.

José:

Estas unidades ven eso y ven fallecidos. Al decir esta persona está fallecida entonces haces un bloqueo por fallecimiento y con eso evitas que vengan a suplantarte.

19. La otra hipótesis es el desconocimiento por parte de las entidades financieras referente a la viabilidad de otorgar créditos a clientes involucrados en hechos delictivos me había respondido que sí existe el cruce de información con estas empresas..

José:

No es un tema de cruce de información yo no te doy crédito porque sales en una noticia o no, en cualquier sitio yo debo evaluar la fuente de re pago el crédito se ve con la fuente de repago y la fuente de repago seguro si tienes una línea que diga que es una fuente lícita, yo te podría dar una súper garantía me das el crédito, pero no yo no te lo doy por la garantía eso es un colateral sino te lo doy por la fuente de repago por estudiar tu actividad económica no solamente tu comportamiento. Yo puedo tener un comportamiento de pago perfecto limpio normal normal no tengo clasificación alguna pero la actividad que está atrás como es o sea lo que tu declaras en impuesto guarda relación con lo que tienes declarado por acá esos cruces los hacen los modelos. Los modelos no solo de pago de impuestos sino también los modelos de créditos hay empresas en el mundo que te venden modelos para crédito tienen una serie de variables, 5000 o no sé cuántas, entre esas cosas que son aplicables y ese es el punto, pero imagínate si yo te diera crédito por si tienes un juicio o no, lo primero que haríamos es ir a buscar gente que no tiene trabajo, estudio ni tengo antecedentes y con eso yo soy sujeto a crédito.

20. La otra hipótesis que tengo es el desconocimiento de parte de las entidades financieras de guardar el dinero procedente de hechos ilícitos de guardar de depositar.

José:

Acá en la banca existe lo que se llama secreto a la reserva oficial y a la bancaria, o sea no te puedo decir si Juan o si yo tengo cuentas en este banco, ese es el tema, no podemos citar o afirmar si tal persona tiene cuentas, sino puedo afirmar no te puedo acompañar en tu hipótesis, no se puede por eso la afirmación, en tu caso tendría que tener algún elemento para que llegues alguna conclusión; si tú me dices que has hecho un estudio la mayoría de casos tal y tal. Lo que yo te digo es que los sistemas de compliance funcionan y para mí funcionan bien, entonces consecuentemente si ellos tienen esa información que reciben de fuentes públicas, privadas, ellos tienen sus modelos que extraen la información y que inmediatamente dice esto debe estar en un tema de riesgo. Tienen sus políticas y lo mismo en fraudes, en fraudes tienen que generar cierto nivel de inteligencia para saber si alguna persona por allí digo, tampoco es blanco y negro tu puedes decir el delito es de mil soles y tienes otro delito de un diez millones de dólares entonces no tienes acá un esquema, esto es asimétrico y de repente esta plata yo la metí en un banco y este banco lo pasó para acá estas metiendo transferencia de otro banco me entiendes. Pero lo que sí te digo que los bancos siempre abren y cierran cuentas si ven algo que no está bien cierran cuentas eso se hace regularmente bajo el protocolo que la ley indica.

21. Ya compliance bloquea esa cuenta y dicen esa cuenta ya no se mueve?

José:

No, hay un protocolo siempre la ley, pero te dice tu para bloquearle la cuenta alguien, para cerrarle la cuenta alguien, tienes mandar una comunicación para que te expliquen en lo que tienes dudas, en lo que no está claro y que quieres que te aclaren si no es suficiente este tema, le mandas una carta señor tal le hemos cerrado

su cuenta. Pero ya previamente ha habido esta comunicación de ida y vuelta explica tal y tal cosa. Pero eso es en todos los bancos.

22. Pero entonces el banco no puede ser unilateralmente, ok yo sospecho que tú tienes actividades ilícitas fundamentadas en estas noticias, si se basan en eso y bloquea entonces son multados por el INDECOPI?

José:

Puedo decir que está la multa del Indecopi, pero son 2 cosas, uno es el tema administrativo y el otro tema el judicial, entonces si tú hablas de fraudes o hablas de estos eventos, tienes que cumplir ciertos protocolos la denuncia o si es por lavado de dinero la comunicación a la UIF, pero antes de eso tú tienes que llamar a la persona, por ejemplo las cuentas mulas que existen en el mundo, lo que se buscan son personas sin antecedentes y abre la cuenta pero mañana le cayó una transferencia de un fraude interbancario de otro país y ya tiene antecedentes. Me comprendes, eso pasa en todo el mundo. Hoy día hay un desocupado a la vuelta de tu casa y ese señor tiene 5 meses de no trabajar de pronto viene alguien y le dice gánate 1,000 soles por abrir una cuenta, eso pasa en todo el mundo pero tienes que darme tu tarjeta y tu clave y él se gana los 1000 soles el tema de la necesidad, el triángulo del fraude: oportunidad, necesidad y racionalidad entonces el tema de la necesidad es tan fuerte que hay gente que se prestan para eso una vez que el soltó su tarjeta y soltó su tema ya le pueden hacer phishing a la gente y transferir esas cuentas para retirar.

Pero a lo que voy es esa persona que abrió esa cuenta que no tiene nada es como cualquier ciudadano, pero una vez que sucede tu notas y dices esto esta raro vamos a llamarlo y le vamos a decir señor Ramírez explíqueme porque esta esté depósito y de repente no viene no da la cara, él no sabe que pasa en su cuenta, están usando su nombre lo llamas a él no contesta. De repente viene y te dice me preste para eso tu vienes denuncias el hecho y cancelas la cuenta. Primero tienes que cumplir con la formalidad de llamarlo y venir a conversar. Ese es el punto porque si prejuzgamos y empezamos a llenar denuncia de todo, también te pones a lo que se llama las contra-denuncias penales, principio de presunción de inocencia.



23. La última pregunta para finalizar la última hipótesis que tenemos es que hay una predisposición de las entidades financieras de utilizar evidentemente herramientas tecnológicas para mejorar la toma de decisiones referentes a clientes reales y potenciales; replanteando la pregunta, existe o existiría predisposiciones por las entidades financieras adquirir un sistema que permita conocer las personas u organizaciones involucradas en hechos delictivos?

José:

Sin duda que sí, todo lo que sea información siempre y cuando sea información de calidad o sea yo quiero información que sea seria porque luego puedo presentarme a tomar decisiones equivocadas, lo que le agregaría a tu pregunta es herramienta con información de calidad. La seriedad lo va a dar la empresa que lo vende que te da el servicio en función a como hizo quienes son sus fuentes, pero si tu fuente va hacer un chisme entonces lo que vas hacer es que primero ya no te voy a leer y luego no te voy a renovar el contrato. Así de sencillo. Entonces al final tú dices, tú me estás diciendo que esto tiene una fuente confiable, si tus fuentes no son confiables doy sentado, chao contrato y se acabó el servicio, por eso es que yo le agregaría eso de información de calidad y que tenga una propuesta de valor. Tú me puedes decir, el comercio me dice, Gestión me dice, la prensa me dice, pero tú has hecho algún análisis de data cruzada, un tema que me diga esto es publicable, esto no, para saber si quiera que el ciudadano existe siquiera, para empezar, cuál es la homonimia. Te puedo decir 20 cosas ahorita, Si solo me vas a poner un nombre no me ayudas. Es como en INFOCORP, si quieres información mía te va a dar mía y no de mi hermana, cada uno tiene su llave de conexión en estos sistemas distintos que normalmente están asociados al número de DNI o algo por el estilo, si tu no me das ese alcance de individualización para empezar de la gente que me estás comunicando no me estás diciendo elementos en actividad o en ese tipo sencillamente te puedo juntar información de los periódicos y te la cuento. Es distinto a que tú dices es que estoy triangulando información con cierto nivel de inteligencia entre lo que publica el twitter del ministerio público, la entidad del poder judicial, El Peruano. También tienes que entender que cosa es el ruido que estás vendiendo, tú me vendes un registro no más, pero a mí no me interesa que me llenes de nombres, a

mí me interesa que tu segmentes, a ver, no me interesa contratar tanto, me interesa contratar la crema, normalmente esto es un Pareto: 20% dónde está el 80% más del ruido reputacional y el 80% donde el 20% está el ruido reputacional ese 20% de abajo con el 80% yo lo veo a lo interno pero tu consígueme la información de la crema y nata grande. Si tú al final me dices vendo por vender ...recuerda cuales son los elementos no es cierto.

