

Universidad Católica de Santa María
Facultad de Ciencias e Ingenierías Físicas y Formales
Escuela Profesional de Ingeniería de Sistemas



**Aplicación de técnicas de data science para demostrar el nivel de
rendimiento del uso del sistema de créditos por parte del personal de una
empresa Financiera.**

Tesis presentada por la Bachiller:

Camilo Llayqui, Anita Graciela

ORCID :0009-0006-0513-5408

Para optar el Título Profesional de

Ingeniera de Sistemas con Especialidad en Sistemas de Información

Asesor:

Mg Montesinos Murillo, Ángel Felipe

ORCID :0000-0003-2116-1535

Arequipa - Perú

2024

UNIVERSIDAD CATÓLICA DE SANTA MARÍA
INGENIERIA DE SISTEMAS
CON ESPECIALIDAD EN SISTEMAS DE INFORMACION
TITULACIÓN CON TESIS
DICTAMEN APROBACIÓN DE BORRADOR

Arequipa, 17 de Mayo del 2024

Dictamen: 011308-C-EPIS-2024

Visto el borrador del expediente 011308, presentado por:

2011200762 - CAMILO LLAYQUI ANITA GRACIELA

Titulado:

**APLICACIÓN DE TÉCNICAS DE DATA SCIENCE PARA DEMOSTRAR EL NIVEL DE RENDIMIENTO
DEL USO DEL SISTEMA DE CRÉDITOS POR PARTE DEL PERSONAL DE UNA EMPRESA
FINANCIERA**

Nuestro dictamen es:

APROBADO

Título Profesional/Título de Segunda Especialidad/Grado Académico a optar:

**INGENIERO DE SISTEMAS CON ESPECIALIDAD EN
SISTEMAS DE INFORMACION**

**29393323 - ZUÑIGA CARNERO MANUEL MARIANO
DICTAMINADOR**



**29601217 - ROSAS PAREDES KARINA
DICTAMINADOR**



**43635330 - ESQUICHA TEJADA JOSE DAVID
DICTAMINADOR**



Aplicación de técnicas de data science para demostrar el nivel de rendimiento del uso del sistema de créditos por parte del personal de una empresa Financiera.

INFORME DE ORIGINALIDAD

6%

INDICE DE SIMILITUD

5%

FUENTES DE INTERNET

2%

PUBLICACIONES

3%

TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1	Submitted to Universidad Europea de Madrid Trabajo del estudiante	1%
2	repositorio.ucsg.edu.ec Fuente de Internet	1%
3	Submitted to Pontificia Universidad Catolica del Peru Trabajo del estudiante	1%
4	ubr.universia.net Fuente de Internet	1%
5	www.fceqyn.unam.edu.ar Fuente de Internet	1%
6	ciateq.repositorioinstitucional.mx Fuente de Internet	1%
7	upc.aws.openrepository.com Fuente de Internet	1%

PRESENTACIÓN

**SEÑOR DECANO DE LA FACULTAD DE CIENCIAS E
INGENIERÍAS FÍSICAS Y FORMALES.**

**SEÑOR DIRECTOR DE LA ESCUELA PROFESIONAL DE INGENIERÍA DE
SISTEMAS.**

SEÑORES MIEMBROS DEL JURADO DICTAMINADOR:

De conformidad con las disposiciones del reglamento de grados y títulos de la Escuela Profesional de Ingeniería de Sistemas, pongo a su consideración el presente trabajo de tesis titulado: “APLICACIÓN DE TÉCNICAS DE DATA SCIENCE PARA DEMOSTRAR EL NIVEL DE RENDIMIENTO DEL USO DEL SISTEMA DE CREDITOS POR PARTE DEL PERSONAL DE UNA EMPRESA FINANCIERA”, el cual merece su aprobación, que me permita optar el título Profesional de Ingeniera de Sistemas con especialidad en Sistemas de Información.

Deseo manifestar mi agradecimiento a las autoridades de la Facultad de Ciencias e Ingeniería físicas y Formales y en especial de la Escuela Profesional de Ingeniería de Sistemas, por el apoyo en el presente trabajo.

Arequipa, Mayo del 2024.

Atentamente.

Camilo Llayqui, Anita Graciela

N° de Matrícula: 2011200762

Bachiller de Ingeniería de Sistemas.

DEDICATORIA

A Dios y la virgen María por un día más de vida y estar en cada paso que doy. A mi familia por enseñarme a mejorar cada día y estar siempre cuando los he necesitado.

Para Dios nada es imposible. (Lucas 1:37)



AGRADECIMIENTO

Agradezco a la Facultad de Ingenierías Físicas y Formales de la Universidad Católica de Santa María por compartir conmigo su experiencia y conocimientos.

Agradezco a mi familia, sobre todo a mi madre por enseñarme a luchar por cada objetivo que tengo en la vida y a mi abuela por la paciencia que tengo para poder servir al prójimo cuando más lo necesita.



RESUMEN

La información que generan las organizaciones es relevante para una buena toma de decisiones para una organización. La clave para determinar el nivel de rendimiento del uso del sistema de créditos por parte del personal de una empresa financiera radica en que un personal pueda ser polifuncional. Por ello, se encontró la necesidad de comprobar si un usuario está completamente capacitado en el sistema o hay un bajo rendimiento, lo que provoca un uso inadecuado del sistema que genera inconsistencias que ocasionan falencias que generan retrasos en la atención. Por ello, se realizó la aplicación de técnicas de data science como regresión lineal y clasificación, para poder evaluar el rendimiento del personal y se comprobó el bajo rendimiento. La información que se utilizó se obtendrá de la base de datos de los tickets registrados por el personal a través de la plataforma Sysaid, el cual ayudó a generar información necesaria para poder hacer una toma de decisiones necesaria de las acciones que se tomarán de manera oportuna en la financiera sobre el personal del área de negocios y operaciones.

Para este proyecto se utilizó Python en Jupyter Notebooks en la versión 6.3.0, con la información obtenida de los 5 años de una empresa financiera, la que ayudó a determinar el nivel de rendimiento que se dieron dentro del tiempo indicado. También se utilizaron técnicas de visualización en Python donde se obtuvieron los siguientes resultados. En el año 2020 se encontró que se registraron más tickets, y también que hay una diferencia total de tickets ingresados por motivo de error de sistema con el 50% y con el de proceso un 43%. La diferencia de tickets de uno a otro es corta. Con lo que se obtuvo es que hay una gran cantidad de tickets de proceso por el personal, que ocasionaron que no se cumplan con los acuerdos de nivel de servicio Service Level Agreement (SLA) de la financiera; con esos resultados obtenidos, se realizó una encuesta al personal donde se concluyó que el 70% de ellos no está capacitado en todos los flujos del sistema. También se observa que el mismo porcentaje indica que los mensajes de alertas del sistema no son claros. Por lo tanto, con los resultados se obtuvo la detección de un bajo rendimiento del uso del sistema, para tomar acciones sobre los resultados obtenidos y así resolver dicha problemática que se está presentando.

Palabras clave: Nivel de rendimiento, sistema de créditos, ciencia de datos.

ABSTRACT

The information generated by organizations is relevant for good decision making for an organization. The key to determining the level of performance of the use of the credit system by the staff of a financial company is that staff can be multifunctional. Therefore The need was found to check if a user is fully trained in the system or there is poor performance, which causes inappropriate use of the system that generates inconsistencies, causing failures that generate delays in care. For this reason, data science techniques such as linear regression and classification were applied to evaluate staff performance and low performance was verified. The information that was used will be obtained from the database of tickets recorded by the staff through the Sysaid platform, which helped generate the necessary information to be able to make the necessary decision-making on the actions to be taken in a timely manner. in the financial area on the personnel of the business and operations area.

For this project, Python was used in Jupyter Notebooks in version 6.3.0, with the information obtained from the 5 years of a financial company, which helped determine the level of performance that occurred within the indicated time, techniques were also used. visualization in Python where the following results were obtained. In 2020 it was found that more tickets were registered, also that there is a total difference in tickets entered due to system error with 50% and with process error 43%, the difference in tickets from one to the other is short , with what was obtained is that there are a large number of process tickets by the staff, which caused the service level agreements Service Level Agreement (SLA) of the financial company to not be met, with these results obtained, it was carried out A staff survey concluded that 70% of them are not trained in all the flows of the system. It is also observed that the same percentage indicates that the system alert messages are not clear. Therefore,

with the results, the detection of poor performance in the use of the system was obtained, to take actions on the results obtained and thus resolve the problem that is occurring.

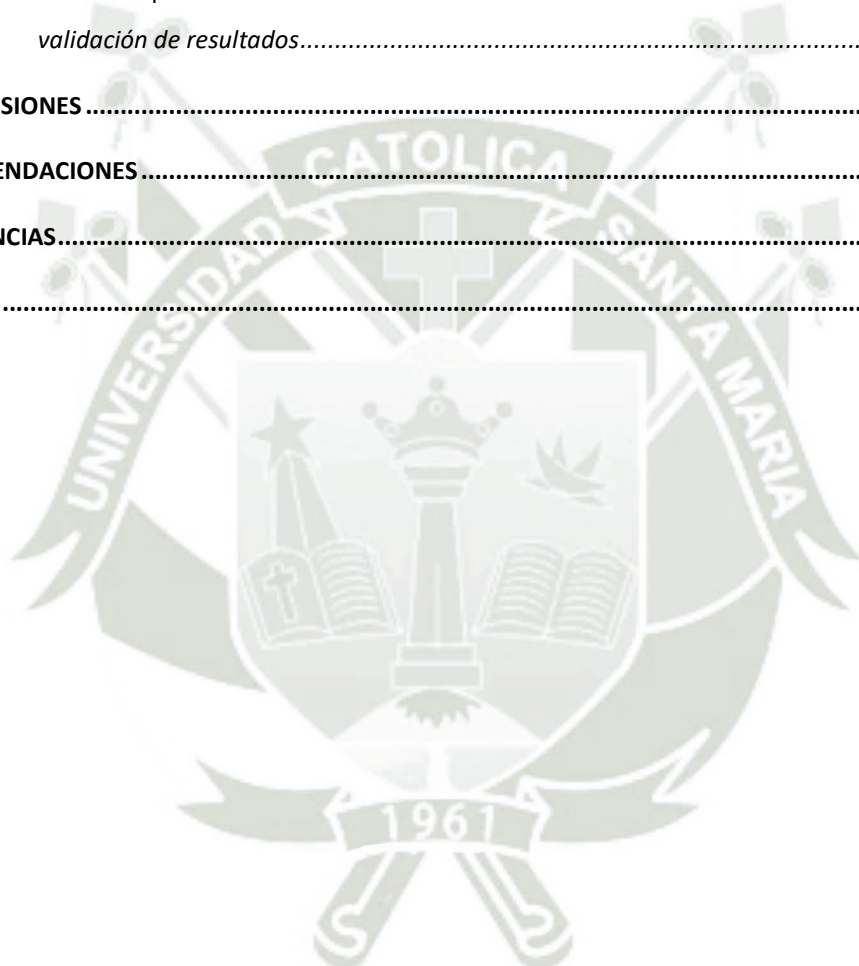
Keywords: Performance level, credit system, data science.



ÍNDICE

PRESENTACIÓN	iii
DEDICATORIA.....	iv
AGRADECIMIENTO	v
RESUMEN	vi
ABSTRACT	vii
ÍNDICE	ix
LISTA DE FIGURAS	xi
LISTA DE TABLAS	xiii
INTRODUCCIÓN.....	xiv
CAPITULO I: PLANTEAMIENTO DE LA INVESTIGACIÓN	1
1. PLANTEAMIENTO DE LA INVESTIGACIÓN	1
1.1. <i>Objetivos de la investigación</i>	2
1.1.1. General.....	2
1.1.2. Específicos	2
1.2. <i>Preguntas de la Investigación</i>	3
1.2.1. Pregunta General.....	3
1.2.2. Preguntas Específicas.....	3
1.3. <i>Alcances y Limitaciones</i>	3
1.3.1. Alcance	3
1.3.2. Limitaciones.....	4
1.4. <i>Población y muestra</i>	4
1.5. <i>Fundamentos teóricos</i>	5
1.5.1. Antecedentes de la investigación.....	5
1.5.2. Bases teóricas de la investigación	9
1.5.3. Técnicas y Herramientas	21
CAPITULO II: DOCUMENTACIÓN TÉCNICA.....	22
2. PLAN DE LA INVESTIGACIÓN INFORMÁTICO.....	22
2.1. <i>Elaboración de la arquitectura de data science</i>	22
2.2. <i>Descripción del modelo</i>	23
2.3. <i>Determinación de datos</i>	24
2.4. <i>Ingeniería de Características</i>	25
2.4.1. Insumos	28
2.4.2. Procesamiento.....	30
2.4.3. Almacenamiento.....	30
2.4.4. Destino	30
2.5. <i>Revisión y depuración de datos</i>	30

CAPITULO III: RESULTADO DE LA INVESTIGACIÓN.....	34
3. RESULTADO DE LA INVESTIGACIÓN	34
3.1. <i>Pre procesamiento de la data.....</i>	34
3.2. <i>Instancias de prueba</i>	35
3.3. <i>Visualización de datos.....</i>	36
3.4. <i>Aplicación del método de Machine Learning</i>	40
3.4.1. <i>Aplicación del método de Regresión Lineal</i>	40
3.4.2. <i>Aplicación del método de Clasificación</i>	44
3.4.3. <i>Comparación de Modelos</i>	48
3.5. <i>validación de resultados.....</i>	48
CONCLUSIONES	51
RECOMENDACIONES.....	53
REFERENCIAS.....	54
ANEXOS.....	57



LISTA DE FIGURAS

Figura 1.	Tipos de entes económicos	15
Figura 2.	Causa de la problemática.	18
Figura 3.	Árbol de objetivos de una organización.	19
Figura 4.	Arquitectura de datos.....	23
Figura 5.	Base de datos.....	24
Figura 6.	Cantidad de ticket.....	26
Figura 7.	Resumen de data	27
Figura 8.	Incidencias	27
Figura 9.	Extracción de datos	28
Figura 10.	Descarga de datos.....	29
Figura 11.	Base de datos sysaid	29
Figura 12.	Filtración de información	30
Figura 13.	Ticket por motivo.....	31
Figura 14.	Ticket por motivo.....	32
Figura 15.	Ticket por motivo.....	32
Figura 16.	Ticket por motivo.....	33
Figura 17.	Ticket por motivo.....	33
Figura 18.	Librerías y procesamiento de data.....	35
Figura 19.	Procesamiento de data	35
Figura 20.	Procesamiento	36
Figura 21.	Cantidad de ticket registrados.....	36
Figura 22.	Resumen de ticket por año.....	37
Figura 23.	Gráfico de ticket por año	37
Figura 24.	Resumen por motivo.....	37
Figura 25.	Gráfico de motivo.....	38
Figura 26.	Resumen de Sla	38
Figura 27.	Gráfico de Sla.....	39
Figura 28.	Promedio por motivo.....	39

Figura 29.	Modelo Regresión Lineal	40
Figura 30.	Datos.....	40
Figura 31.	Regresión lineal.....	42
Figura 32.	Regresión lineal.....	42
Figura 33.	Mean Absolute Error	43
Figura 34.	Mean Squared Error	43
Figura 35.	Root Mean Squared Error.....	43
Figura 36.	Regresión logística	45
Figura 37.	Reporte de clasificación.....	46
Figura 38.	Reporte de clasificación.....	47
Figura 39.	Encuesta 1	48
Figura 40.	Encuesta 2	49
Figura 41.	Encuesta 3	50
Figura 42.	Encuesta 4	50

LISTA DE TABLAS

Tabla 1.	Tipos de co-creación.....	11
Tabla 2.	Archivo por años	26
Tabla 3.	Conversión de datos Motivo.....	41
Tabla 4.	Conversión de datos Cumplimiento	41
Tabla 5.	Comparativo de modelos	48



INTRODUCCIÓN

La importancia de que el personal esté bien capacitado radica en que podrá cumplir con sus labores adecuadamente de una financiera; podrá garantizar que sean polifuncionales, que puedan tener un proceso bueno y ágil en la atención del usuario final, que viene a ser todo cliente que requiera realizar alguna actividad en la agencia, y además se logre reducir los tiempos de atención al cliente brindándole un servicio de calidad. Por eso, esta tesis tiene como objetivo determinar el nivel de rendimiento del personal; por consiguiente, se está utilizando la base de datos de registros de tickets ingresados por el personal de una empresa financiera para poder ser analizados, lo que genera que mientras una financiera empieza a crecer no se dé abasto con el tema de capacitación en todos los procesos del personal para un sistema bancario o que el sistema no sea interactivo con el usuario. En este caso encontraremos 3 tipos de motivos por los que se clasifica cada ticket; por ejemplo, por proceso, son los usuarios que se capacitan nuevamente, ya teniendo una capacitación anterior o un manual sobre su flujo; también tenemos por capacitación, donde los usuarios no saben nada del proceso y se capacitan o se hace un refuerzo sobre el nuevo flujo por una actualización del sistema y, por último, tenemos por error de sistema, que son errores mismos de software y son corregidos con una actualización por base de datos o una corrección del sistema; con ello se procederá a demostrar con la información generada el nivel de rendimiento del uso del sistema por parte del personal.

La presente tesis se encuentra conformada por los siguientes capítulos:

El primer capítulo está compuesto por el planteamiento del problema, donde la principal problemática es poder determinar el nivel de rendimiento del personal con los registros de tickets por el personal en agencia.

El segundo capítulo está compuesto por antecedentes, marco conceptual y el marco teórico donde se desarrolló teóricamente las metodologías de data science.

El tercer capítulo se expresa el marco metodológico, donde se incluyen apartados como alcances, limitaciones, tipo de la investigación, además de métodos, técnicas e instrumentos empleados en la presente tesis.

El cuarto capítulo está compuesto por los resultados de la investigación, cumpliendo con los objetivos indicados, aplicando técnicas de data science.

Finalmente encontraremos las conclusiones, recomendaciones, referencias bibliográficas y anexos del presente proyecto.

CAPITULO I: PLANTEAMIENTO DE LA INVESTIGACIÓN

1. PLANTEAMIENTO DE LA INVESTIGACIÓN

Según la Superintendencia de Banca, Seguros y AFP SBS, "el crecimiento sostenido de la red de atención, tanto a nivel nacional como departamental del país, se ha visto reflejado en un mayor acceso de la población a los servicios financieros" (SBS, 2022, párr. 16-18). Por lo cual, el rubro financiero evoluciona con el pasar de los tiempos, lo que genera que las financieras se expandan de una forma notable y descentralizada geográficamente en el sistema financiero. Lo que conlleva a que si las financieras crecen, el personal que contrata también se incrementa, ya sea por ser nuevo, antiguo o que va a cubrir un puesto nuevo. Esto ocasiona un bajo nivel de rendimiento del uso del sistema de créditos por parte del personal.

Por lo tanto, la falta de capacitación del personal en el sistema genera un rendimiento inadecuado en un cierto porcentaje, provocando que el flujo del proceso no continúe con normalidad, ocasionando que se generen reprocesos en el flujo por falta de capacitación y no pueda realizar todas las actividades. Esto hace que existan diferentes causas del mal manejo inadecuado del software que usa el personal de la empresa, generando inconsistencias, lo que provoca que los procesos se retrasen por el usuario y piensen que sea un incidente con el sistema y de lugar a un bajo rendimiento de calidad de atención al cliente en ciertas ocasiones.

Así como menciona Gómez (2017, p. 8) se identifica que existe la necesidad de que todo el personal que labora en la empresa sea polifuncional, es decir, que todos aprendan a realizar los diferentes procesos del estudio de créditos con el fin de alivianar las cargas de procesos que se registran cuando se presenta alguna ausencia de algún personal o para cubrir las vacaciones y licencias de ellos mismos. Esto da lugar a que la capacitación en los usuarios sea primordial y así no producir inconsistencias en el proceso del uso del software. Para no provocar una pérdida de tiempo en la atención de incidentes por proceso, para así identificar

los errores con el sistema a tiempo y que no se pierda con los otros incidentes, cumpliéndose a cabalidad las políticas de capacitación en la empresa, para no originar a su vez una baja calidad de atención en todo el flujo.

Con la información que se genera cada día por los ticket registrados por los usuarios esto nos ayudara a poder aplicar técnicas de DATA SCIENCE y así poder hacer una toma de decisiones demostrando así la falta de proceso que se presentaron semanalmente o mensualmente, esto ayudará a poder obtener una estadística y poder hacer a continuación una encuesta al personal para poder ver los puntos a reforzar en una capacitación de proceso en el flujo o brindar la información necesaria en los manuales, como resultado tendremos una reducción de tiempo en la atención de incidentes y atender los verdaderos incidentes de error con el sistema y puedan ser resueltos a tiempo cumpliendo con los acuerdos del nivel de servicio (*SLAs*).

1.1. Objetivos de la investigación

1.1.1. General

Demostrar el nivel de rendimiento del uso del sistema de créditos por parte del personal de una empresa financiera utilizando técnicas de data science.

1.1.2. Específicos

- A. Recolectar la información que se va a utilizar para demostrar el nivel de rendimiento.
- B. Validar los modelos adecuados para la ejecución del plan de muestreo
- C. Comparar los modelos de regresión lineal y de clasificación para la aplicación en esta investigación.
- D. Determinar el nivel de rendimiento del uso del sistema por parte del personal de una empresa financiera.

1.2. Preguntas de la Investigación

1.2.1. Pregunta General

¿Cómo se demostrará el nivel de rendimiento del uso del sistema de créditos de una empresa financiera, al utilizar técnicas de data science?

1.2.2. Preguntas Específicas

- A. ¿Cómo se recolectará la información que se va a emplear para demostrar el nivel de rendimiento?
- B. ¿Cuáles serán los resultados de validar los modelos adecuados para la ejecución del plan de muestreo?
- C. ¿Cómo comparar los modelos de regresión lineal y de clasificación para la aplicación en esta investigación?
- D. ¿Se podrá determinar el nivel de rendimiento del uso del sistema de créditos por parte del personal de una empresa financiera?

1.3. Alcances y Limitaciones

1.3.1. Alcance

En el siguiente proyecto se usará la base de datos de una financiera en donde se tomarán los datos de 5 años, que abarcan los años del 2017 al 2021; con ello se podrá hacer un análisis respectivo para poder determinar las causas del nivel de rendimiento de los usuarios. El software que nos ayudará para poder tener resultados será Anaconda Copyright © 2016 Anaconda, Inc. versión 2.0.3, con ayuda de este software se podrán obtener resultados de la empresa de las estadísticas de cada año sobre el motivo de incidente ingresado por los usuarios; para ello se utilizarán las técnicas de visualización, regresión lineal y clasificación en el lenguaje de Python versión 3.

Para poder analizar estadísticamente los datos, después confirmar con las encuestas realizadas a los usuarios si realmente están capacitados, y de acuerdo con ello tomar las medidas correctivas para no generar pérdidas en la empresa y generar una mejor atención al usuario final en su atención.

1.3.2. Limitaciones

Acceso limitado al área de reportaría que utiliza la herramienta de Power BI que tiene datos sensibles, al tratarse de un área diferente de la empresa la que está a cargo de este software.

1.4. Población y muestra

- Universo: Personal operador del sistema de créditos.
- Muestra: Como la población será mayor a 35 personas, se debe aplicar la fórmula de poblaciones finitas.

Según Gemmell (1977) menciona que, para determinar el tamaño de la muestra, se utiliza la fórmula de población finita para calcular el tamaño de la muestra en función del tamaño de la población, la precisión deseada y otros factores relevantes:

$$n = \frac{Z^2 * p * q * N}{e^2 * (N - 1) + Z^2 * p * q}$$

N.C = 95 %

Donde:

N.C = Nivel de confianza, grado de seguridad que los datos son representativos de toda la población.

Z = Parámetro estadístico ligado al nivel de confianza, al tener un 95 % de nivel de confianza, Z=1.96.

p = Valor estadístico de la probabilidad, es decir que las unidades de muestra tienen un 0.5 o 50 % de probabilidades de que respondan al instrumento.

q = Es el otro 0.5 o 50 % de unidades que decidan no responder o participar de la investigación.

e = Representa el error que se asume por cualquier sesgo que se pueda generar. 5% en este caso.

N = Tamaño de la población

n = Es el tamaño de la muestra calculada

Para obtener los datos del tamaño del universo y de la muestra, estos fueron proporcionados por la oficina general de estudios de la Unasam. Los cuales son los siguientes:

- Universo: 5600
- Población: 5600

Reemplazando los valores en la fórmula:

$$n = \frac{(1.96)^2(0.50)(0.50)(5600)}{(5600 - 1)(0.05)^2 + (1.96)^2x(0.50)x(0.50)}$$
$$n = 359$$

Se obtiene una muestra de 359.

1.5. Fundamentos teóricos

1.5.1. Antecedentes de la investigación

A. Rua y Gomez (2017), La importancia de la capacitación en los asesores de confiar para el buen funcionamiento del proceso de análisis de los créditos, Tecnológico de Antioquia, Institución Universitaria, Medellín-Colombia.

El objetivo del siguiente proyecto es diseñar herramientas de capacitación para el personal para que puedan ser polifuncionales. Para el desarrollo se procedió a emplear capacitaciones virtuales como presenciales y cursos didácticos; con ello se llevó a la conclusión que si hay retroalimentación de la nueva información se podrán generar nuevos conocimientos constantes para sus funciones. Este proyecto es importante porque podemos darnos cuenta de que la mayor parte del personal en una financiera no son polifuncionales y lo que plantea esta tesis es mejorar la producción en la entrega de créditos.

- B. Douglas y Diocelina (2021). Técnicas e indicadores de rendimiento financiero aplicados al estado de resultados en empresas comerciales y de servicios colombianas, Colombia.

El objetivo principal que tuvo la investigación fue identificar técnicas e indicadores del rendimiento financiero. Para el desarrollo de esta investigación se aplicó un cuestionario, el cual fue aplicado a 55 empresas comerciales y 35 empresas de servicios. El propósito de esta investigación fue apoyar la toma de decisiones empresariales. Esta investigación es importante porque nos ayudará a aplicar técnicas para poder determinar el rendimiento de forma cuantitativa para aplicar un cuestionario correcto.

- C. Bravo et al. (2021) Aplicación de ciencia de datos para incrementar la efectividad del número de operaciones de la base de clientes tácticos de Mibanco-Agencia Zárata, Universidad Peruana de Ciencias Aplicadas, Lima-Perú.

El presente trabajo de investigación su principal objetivo es incrementar el nivel de efectividad de número de operaciones. Para poder desarrollar la siguiente investigación se aplicó ciencia de datos estableciendo un modelo para así poder

predecir el perfil de los clientes para una financiera. Se procedió a aplicar en este caso a una agencia en específica que fue la agencia Zarate de Lima que forma parte de la financiera y se llegó con la conclusión de que si aplicamos ciencia de datos se podrá analizar de manera más rápida para poder mejorar la efectividad de más operaciones para encontrar los perfiles adecuados de cada cliente. Esta investigación resulta ser importante porque nos damos cuenta de que el avance tecnológico nos ayuda a aplicar diferentes técnicas de data science para poder analizar la información de gran cantidad y poder así predecir sobre una base de datos de una institución.

- D. Eckert (2019) Modelo basado en la toma de decisiones con criterios múltiples para la elección de metodologías de Data Science, Universidad Nacional de Misiones, Posadas, Argentina.

El objetivo principal que tuvo la investigación es establecer un modelo comparativo de metodologías de data science mediante métodos para la toma de decisiones. Para el desarrollo se realizaron dos casos de validación, uno por el motivo por el que se producen las roturas de las autopartes de automóviles y el otro caso fue aplicado en una universitaria en la carrera de Licenciatura en Sistemas de la Universidad Nacional de Río Negro, para el período comprendido entre 2009 y 2015. Este proyecto logró fundar un modelo basado en la toma de decisiones con criterios múltiples con el objetivo de comparar metodologías de data science. Esta investigación es importante porque se pudo utilizar una de las dos metodologías utilizadas para una mejor opción en la toma de decisiones.

- E. Gutierrez (2020) Modelo adoptado para la toma de decisiones efectivas en el fondo de garantías de instituciones financieras, Universidad Santo Tomás Tunja, Boyacá, Colombia.

El procesamiento de la información para la toma de decisiones efectivas para tener un proceso más ágil y confiable mediante herramientas tecnológicas para no generar pérdidas de tiempo innecesarias y tener la información precisa y actualizada. Para el desarrollo de la investigación, se utilizó Microsoft Power BI para tener la información en la nube para que la información esté disponible de manera inmediata. Esta herramienta fue aplicada para entidades financieras. Contar con una herramienta tecnológica con la información organizada permite una toma de decisiones oportuna, ya que los datos que tiene son de un análisis confiable, generando una mejor productividad para las empresas. Esta investigación resulta importante porque ayuda a darse cuenta de que las herramientas tecnológicas nos ayudarán a una buena toma de decisiones.

- F. Pignano Pino (2021) La generación de valor mediante el uso de data science en la toma de decisiones comerciales de tiendas por departamento, Pontificia Universidad Católica del Perú, Lima-Perú.

El siguiente trabajo buscó conocer las necesidades del consumidor para así facilitar la toma de decisiones o encontrar ciertas características para mejorar sus procesos organizacionales y construir un marco analítico para que permita hacer el uso de data science para una mejor toma de decisiones. En esta investigación, se contó con un reporte bibliométrico de las siguientes páginas SCOPUS y WEB OF SCIENCE para poder analizar los datos, para poder hacer un marco analítico. Este proyecto se aplicó a empresas de Perú. Esta investigación ayudará a mejorar los procesos de toma de decisiones de las organizaciones para poder generar un valor

añadido que genere un entendimiento con el cliente para ver sus necesidades y expectativas, para así tener un diferenciado frente ante la competencia para poder predecir futuros acontecimientos. Mejorar el flujo de capacitación de los usuarios y conocerlos bien para ver sus necesidades para que mejoren en el uso de la herramienta.

1.5.2. Bases teóricas de la investigación

La ciencia de datos busca extraer conocimiento sobre una base de datos por ello engloba teorías, métodos científicos, procesos y técnicas.

1.5.2.1. Técnicas

Las principales técnicas en data science, por ello nombraremos 3 técnicas importantes. Según AWS Amazon Web Services (2023) es una plataforma en la nube el cual ofrece diferentes servicios y con una amplia variedad en el mundo empresarial.

- Clasificación:

En esta técnica podremos encontrar los datos ordenados en grupos o categorías específicas. En el cual el ordenador estará entrenado para identificar y ordenar datos a través de algoritmos de decisión.

- Regresión:

Es un método para ayudar a encontrar una relación entre dos puntos que aparentemente no podrían relacionarse. Aquí se utilizará una fórmula matemática que será representada por un gráfico. En esta técnica, si nosotros conocemos un punto, podemos predecir el otro punto de datos, los cuales serán importantes puesto que se identificará en el diagrama si existen puntos fuertes o débiles.

- Clústeres

Este método permitirá agrupar datos estrechamente relacionados para poder encontrar patrones y anomalías. Aquí los datos no se pueden clasificar con precisión, en donde se pueden agrupar en relaciones más probables donde se pueden descubrir nuevas relaciones

1.5.2.2. Tipos

El autor Arevalo (2007) nombra 3 tipos de gestión de la información que veremos a continuación.

- Gestión del conocimiento:

Menciona que en la gestión del conocimiento está relacionado con la información que tiene el personal de una empresa sobre todo el conocimiento, flujo y proceso que se realiza en la empresa o en una institución. También veremos que los procesos de transacciones en los sistemas se registrarán diariamente en una base de la organización.

- Gestión de la información:

El autor indica que el análisis de la explotación de la información dará una calidad, exactitud y claridad que ayudará a la organización. Al igual que el autor Bustelo (2000) esto ayudará a controlar, almacenar y tener la información adecuada y poder tener el control de los procesos del negocio.

- Gestión de contenidos:

La gestión de contenidos es el que permite la administración de información para su evolución, el cual será fundamental para la empresa para la toma de decisiones. Por otra parte, Arcila et al. (2016) indica que para hacer un análisis de contenido es mejor mejorar las técnicas utilizadas para así poder analizar de forma analítica y poder obtener nuevos descubrimientos. Se han puesto en

contexto para ayudar a controlar toda la información gestionada en las organizaciones.

1.5.2.3. Teorías

Como lo menciona Sanchez et al., (2018) en el mundo de hoy se mira que los negocios han cambiado puesto que el cliente o el usuario final, ahora están más informados, activos y puede tener acceso al alcance de su mano en cualquier lugar y en el momento que lo necesite. Por lo cual las organizaciones hoy en día piensan que los clientes pasan a ser cocreadores o co-productores como se clasifica en la tabla 1.

Tabla 1. *Tipos de co-creación*

Tipo	Concepto	Referencia
Codiseño	Incluye un conjunto de prácticas orientadas a alinear las soluciones con los problemas o necesidades del consumidor.	Russo-Spena & Mele (2012)
Co-evaluación de ideas	Las organizaciones recaban la opinión de agentes externos fundamentalmente a través de comentarios y votaciones.	Russo-Spena & Mele (2012)
Co-generación de ideas	La generación de ideas se alimenta de las aportaciones de una red externa de actores que puede incluir usuarios, seguidores, socios, profesionales e intermediarios que participan de manera activa.	Russo-Spena & Mele (2012)
Colanzamiento	Participación de agentes externos en la difusión y lanzamiento del producto al mercado.	Russo-Spena & Mele (2012)
Co-test	Actividades cuyo objetivo es validar los prototipos de productos o servicios de manera previa a su lanzamiento al mercado	Russo-Spena & Mele (2012)
Cofinanciación	Participación de diferentes actores en la financiación del producto o servicio	Ordanini et al. (2011) Payne et al. (2008);
Co-consumo	Prácticas de colaboración para el uso o consumo del producto o servicio.	Grönroos & Ravald (2011); Gebauer et al. (2010)

Nota: En la siguiente investigación se valida que el cliente ya no es el objetivo de las empresas si no paso hacer un elemento activo para ello, para así generar valor de forma conjunta para la venta de productos y servicios diarios. Adaptado de “La co-creación y los nuevos retos de generación de valor que enfrentan las organizaciones”, por C. Sanchez y D. Prada, 2018. p.174.

Por otro lado, Delgado (2021) indica que la teoría organizacional valida el comportamiento de individuos o grupos para encontrar las necesidades organizativas para así poder lograr el objetivo y poder adaptarse en un ambiente cambiante. Por ello encuentra 6 principales teorías organizacionales, las cuales son:

A. Teoría clásica

En esta teoría, tiene más énfasis en las organizaciones, que en los trabajadores, por lo cual sus principales características son. Detectar errores y los corrige una vez cometidos en el instante. Está más pendiente de la cantidad de producción que genera una empresa que por los seres humanos. Se asume que los empleados son estables frente a algún cambio en términos que realice una modificación en la organización.

B. Teoría de la gestión científica:

En esta teoría, Delgado habla de la teoría de Fredrick Winslow Taylor, menciona la gestión científica que a menudo se denomina "taylorismo" los cuales define de la siguiente manera.

Separa la planificación de la acción. Los supervisores deben ser claros en dar instrucciones en sus respectivas áreas. Después de varios estudios, se utilizan los tiempos, movimientos y fatiga para determinar la cantidad justa de trabajo realizado. Mejora las condiciones de trabajo y estandariza las herramientas para el personal para poder validar el tiempo y el costo de producción. Debería realizar una formación de los trabajadores. Los incentivos económicos deben darse a los trabajadores para aumentar su productividad y motivarlos a desempeñarse para poder mejorar su rendimiento.

C. Teoría administrativa:

En esta teoría, el objetivo es realizar diferentes actividades con los trabajadores para lograr un propósito en común para identificar y clasificar y que puedan ayudar a la organización.

Según Delgado, la teoría administrativa de Henri Fayol creía que debería poner más énfasis en la gestión organizacional, como ver los factores humanos y de comportamiento en la gestión. A diferencia de estas dos teorías de la gestión científica, Taylor da hincapié a mejorar la eficiencia y minimizar el tiempo de tarea del trabajador para que el principal enfoque estructure la gestión de las organizaciones.

D. Teoría burocrática:

E. Está relacionado a la estructura y el proceso administrativo dada por Max Weber.

Una organización burocrática es una jerarquía de autoridad, que ayudará a reducir la complejidad del funcionamiento de la organización por las reglas y regulaciones. Es el medio más racional para poder ejercer un control de los trabajadores.

F. Teoría neoclásica:

En esta teoría, el principal factor son las acciones humanas que realizan los trabajadores de una organización, lo cual hace hincapié para determinar la eficiencia de la organización, lo cual hace que las variables sean positivas, pero no se vean reflejadas en el comportamiento laboral en su día a día de una organización.

G. Teoría Moderna:

En esta teoría indica el comportamiento, el cual cambia con su entorno, tanto en el interno como el externo. Hay varias teorías que se distinguen de otras teorías organizacionales.

La teoría moderna es considerada un sistema abierto; lo que hace es que interactúe constantemente con su entorno para poder mejorar organizacionalmente. Lo cual se diferencia del clásico donde se considera un sistema cerrado. El sistema abierto es el crecimiento que se da por los cambios en el entorno que son cambiantes y deben adaptarse al entorno en el que están.

Considerada como un sistema dinámico. Es probabilista cuando los resultados están predeterminados y no se determinan, por lo que no son exactos y cuando los modelos probabilísticos son inciertos y depende lo que suceda o de la posibilidad que pueda ocurrir en el transcurso.

En esta teoría vemos como la organización cubre tanto lo externo como lo interno de la organización. También es considerada multivariable; hace que la causa y el efecto no sean fenómenos puesto que puede ser de variables interrelacionadas o ser interdependientes. En las finanzas también podemos encontrar diferentes tipos de entes económicos, como se muestra en la figura 1 (Soto et al., 2017).

Figura 1. *Tipos de entes económicos*

Nota: Estos tipos ayudarán que los entes sean más eficaces y eficientes con los recursos financieros. Adaptado de “Gestión Financiera Empresarial”, por M. Fajardo y C. Soto, 2017, p.17.

1.5.2.4. *Procesamiento*

El procesamiento de los datos se realizará para el análisis y así sacar conclusiones o recomendaciones que ayudarán al proceso de extracción para el cumplimiento de los objetivos empresariales (Rubiños et al., 2020).

Según la empresa estadounidense que analiza contenido para crear un análisis instantáneo, Mdcloud (2018) indica que para poder procesar los datos de una empresa, para que permitan extraerlos de uno o varios lugares, se debe validar que los datos estén actualizados para poder utilizar la información, de ahí validar las operaciones de registro, verificación, separación, clasificación y así finalmente poder extraer una nueva base de datos. En consecuencia, obtendremos nuevos datos limpios para tener un acceso rápido, sencillo y ágil. Además, se nombrarán las fases del proceso de ETL que se deben dividir en las siguientes fases.

A. Fase 1: Extraer

La extracción de información se puede extraer de diferentes sitios que puede almacenar una empresa desde una base de datos que tiene una empresa o también de datos estadísticos resultados de una clasificación de datos.

La extracción de datos se puede clasificar para poder adaptarlo en cada dato de información; de ahí se cree una nueva fuente de datos que pueda adaptarse en un formato estándar y uniforme, para que esté estandarizado y se gestione el resultado más adelante.

B. Fase 2: Transformar

Después de realizada la primera fase, se tendrá datos preparados y limpios, el cual estará listo para ser utilizado en esta segunda etapa. Los datos obtenidos se pueden convertir en datos finales que formarán una base de datos nueva que se transformó de la que se extrajo. Podremos utilizar solo los datos útiles que serán adecuados para el desarrollo de esta tesis. Por ello, esta fase nos ayudará a eliminar información que puede ser innecesaria, desfasada o duplicada.

C. Fase 3: Cargar

En esta fase de carga, que es la última, se realizará la carga para subir los datos limpios obtenidos en la anterior fase. Esta fase también es buena para mantener la base de datos antigua como la nueva para cuando sea necesario y pueda ser requerido en su momento para la empresa; también se puedan usar estos datos cuando se necesiten y con ello pueda contribuir a ayudar a la organización.

1.5.2.5. Requisitos, comprensión y preparación de datos

A. Determinación de datos

La La extracción de datos será fundamental para poder determinar el nivel de rendimiento del personal; por ello se procederá a usar los datos de una empresa financiera para su análisis.

B. Fuente de datos

La fuente de datos será utilizada de una financiera, que serán datos utilizados como extracción de datos como encuestas a los mismos usuarios.

C. Presentación de datos recopilados

En este caso se extraerá información de una base de datos la cual se procesará para su respectivo análisis y pueda ser presentada para la siguiente fase de depuración y limpieza.

D. Revisión y depuración de datos

En esta última fase se realizará la carga de los datos limpios para subir los datos obtenidos en la anterior fase.

1.5.2.6. Nivel de rendimiento del personal de una empresa financiera

A. Causas

En el rubro financiero es importante realizar un diagnóstico integral para administrar las operaciones y tener una la relación de causa-efecto para que afecte a todos los aspectos de toda la organización (Terrazas, 2009).

Figura 2. Causa de la problemática.

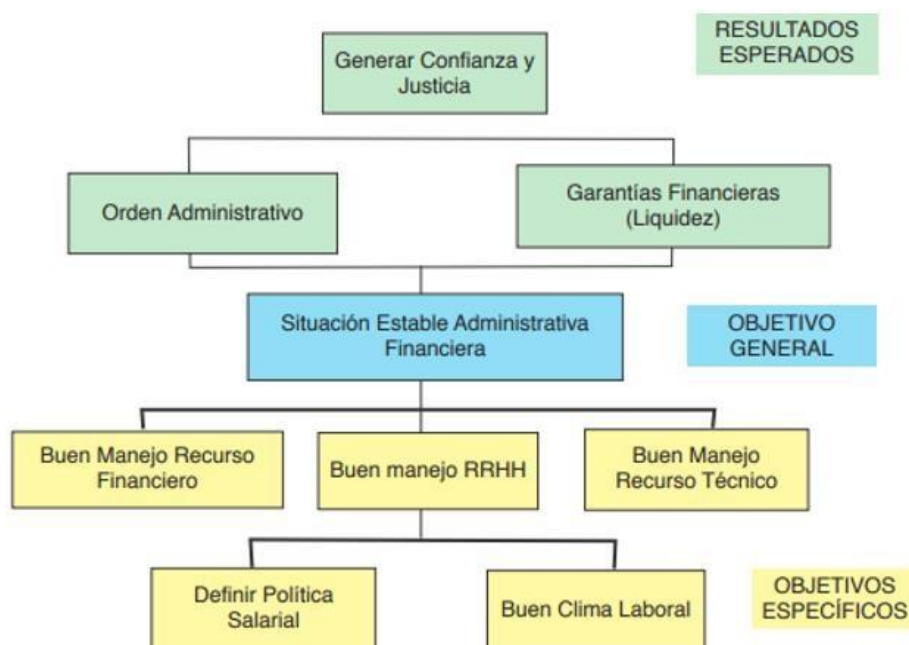


Nota: En esta causa problemática se muestra la causa y efecto de los aspectos de una organización. Adaptado de “Modelo de gestión financiera para una organización”, por R. Terrazas, 2009, p.60.

B. Consecuencias

En la siguiente figura 3 se muestra el planteamiento de un modelo, en el cual ayudará a una solución de las consecuencias generada para poder dar alcances positivos.

Figura 3. *Árbol de objetivos de una organización.*



Nota: Este árbol de objetivos mejorar la toma de decisión de una organización. Adaptado de “Modelo de gestión financiera para una organización”, por R. Terrazas, 2009, p.61.

C. Como evaluar el desempeño del personal

Medir el desempeño del personal será beneficioso para la empresa, sino también para el mismo personal, puesto que de acuerdo con su evolución se podrán obtener mejores oportunidades (QuestionPro, 2022).

D. Cuáles son los indicadores de desempeño de un trabajador

Esto servirá para poder medir los diferentes aspectos del personal y ver un aspecto global del talento a continuación, veremos 5 factores según (GDM, 2022).

- Indicador de trabajo

Para medir los indicadores por la calidad y el tiempo que toma cada actividad, si hay un desnivel amplio en el reporte podemos ver que hay una problemática o impedimento que interfiere en los resultados.

- Indicador financiero
En este indicador podemos ver cuánto cuesta un trabajador y los beneficios económicos que recibe la empresa para ello debe incluirse un KPI y de ahí medir los ingresos por roles y así poder implementar acciones que motiven a continuar a cada trabajador.
- Indicador de satisfacción del cliente
Otro de los indicadores que se debe poner en hincapié debe ser la satisfacción de los clientes puesto que pone en riesgo la estabilidad de la organización y reputación de la empresa.
- Indicador de procesos
En este indicador se evaluará al personal si cumple con todos los reglamentos de la empresa el cual permitirá encontrar los cuellos de botella que se puedan estar generando en algunas ocasiones para poner darles una solución anticipada.
- Indicador de desempeño
En este indicador se consideran 4 tipos de métricas
 - ✓ Productividad
 - ✓ Eficacia
 - ✓ Eficiencia
 - ✓ Formación

1.5.2.7. *Motivo de incidentes*

A. Proceso

Es cuando el personal no sabe su flujo ya especificado y que realiza todos los días.

B. Capacitación

Es cuando es una actualización en el sistema un cambio nuevo reciente en el cual aún el personal no sabe aún aplicarlo.

C. Error

Es cuando el personal no puede continuar con su flujo y cuando entra en revisión hay datos incongruentes con el sistema que no se tuvieron contemplados.

1.5.3. Técnicas y Herramientas

1.5.3.1. Herramientas

- Plataforma sysaid

Es una plataforma de gestión de servicios que ayudará a una empresa a dar soporte sobre los incidentes y solicitudes generados por el personal, para poder medir así el tiempo de servicios que le entrega a un usuario.

- Power BI

Es una herramienta que ayuda a la visualización de datos de grandes cantidades de información, es una reportaría muy eficiente para la gestión de mediación de indicadores y ayuda una mejor toma de decisiones sobre los datos.

- Python

Es un lenguaje de programación para poder hacer data science en este caso se está utilizando la versión Python 3.

- Jupyter Notebooks

Es una aplicación web, el cual permitirá realizar la ejecución de código Python o diferentes lenguajes.

- SQL server

Es una herramienta para la gestión de base de datos de una empresa, que almacena información en grandes cantidades, en la cual vamos a poder manejar los datos como crear una base de datos y poder hacer la gestión de información.

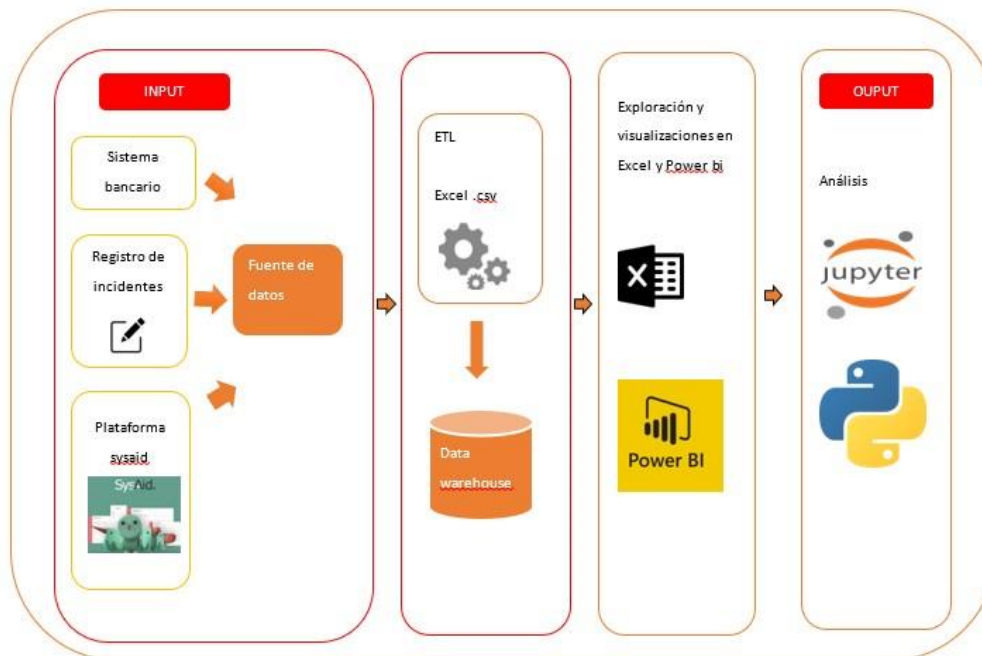
CAPITULO II: DOCUMENTACIÓN TÉCNICA

2. PLAN DE LA INVESTIGACIÓN INFORMÁTICO

2.1. Elaboración de la arquitectura de data science

Para poder identificar los datos y poder determinar el nivel de rendimiento del personal, se elaborará el siguiente modelo de la figura 4, en el cual se extraerá los datos de los registros que realizan los usuarios como un incidente en la plataforma sysaid, estos incidentes son causados por el sistema por una inconsistencia o por un mal uso del software en su flujo, lo que genera una interrupción en su proceso, el cual se convertirá en una fuente de datos que se procederá a procesar como se muestra en la siguiente figura que presentaremos, luego se procederá a hacer el flujo de la extracción de datos y el procesamiento para la gestión previa, a continuación se procederá eliminar datos para tener una base de datos nueva y limpia con los datos necesarios para poder determinar el nivel de rendimiento, entonces para poder obtener una información clara y de calidad, para asegurar la adecuada información, se utilizarán las siguientes herramientas como Excel y Power BI a través de gráficos donde se mostrarán los primeros datos de la investigación.

En la figura 4, se observa la arquitectura de datos de un sistema financiero en el cual veremos cómo se gestionaran los datos para poder hacer el proyecto, en donde el sistema financiero es el principal factor donde empieza el flujo del personal, el cual con lleva que si no saben una parte del flujo o tienen un inconveniente generen un ticket, el cual será un registro de datos importante para el proyecto, el cual estará registrado en la plataforma sysaid, a continuación se hará una extracción de la información, el cual se convertirá en una nueva información que se convertirá en una nueva base de datos, en la cual se podrá hacer una exploración y visualización de dichos datos, el cual se convertirá en una salida puesto que se tendrá el resultado de análisis realizado con la herramienta Python sobre la información recopilada de datos.

Figura 4. *Arquitectura de datos*

Nota: Elaboración propia.

2.2. Descripción del modelo

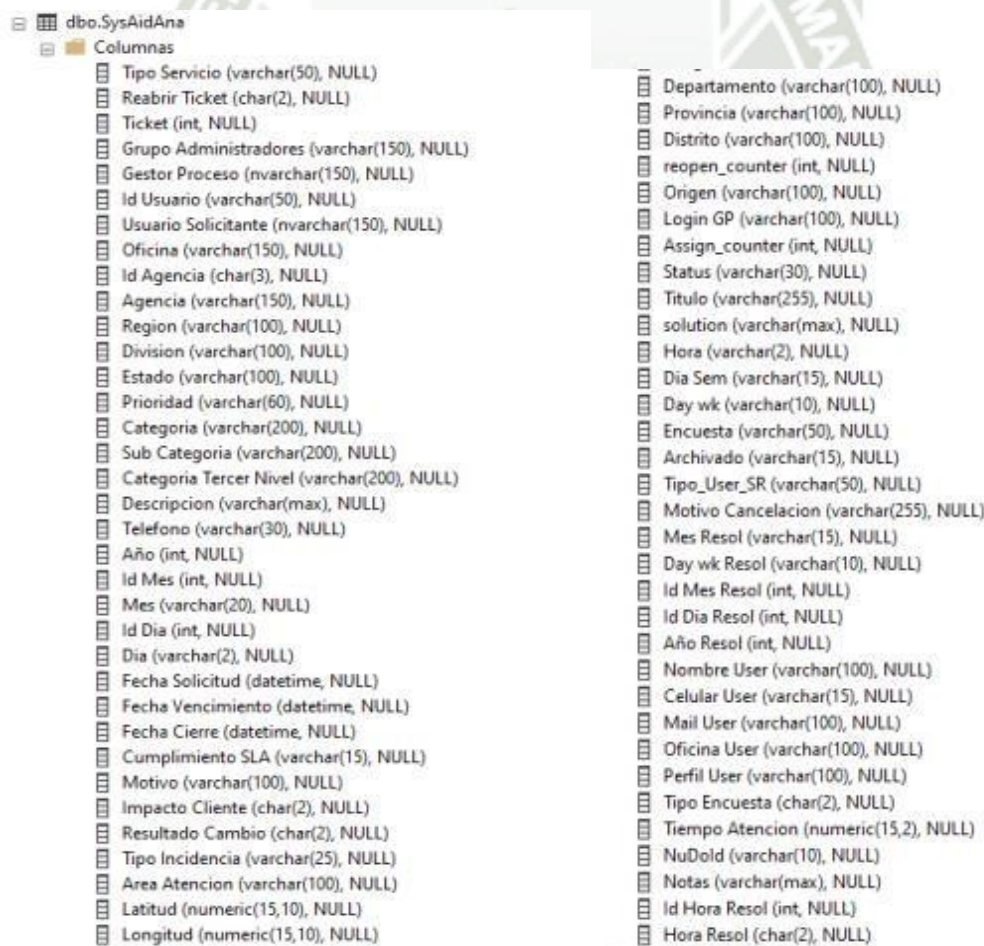
Para el presente trabajo de proyecto se usará datos históricos de varios años de una empresa financiera, la cual utiliza la plataforma sysaid para el registro de incidentes, donde cada usuario genera un ticket, por cada inconveniente que presenta en el día a día, éste después se clasificará y se guardará en un backup lo que permitirá tener una base de datos original como una nueva, después estará en un servidor llamado datawarehouse que luego se podrá visualizar en los reportes de Power BI y en la base de datos, después se hará la filtración de datos para ver las columnas que se utilizarán para el análisis en Python, luego se procederá a extraer en un Excel con las variables que se utilizarán para poder aplicar el modelo de regresión utilizando Python en Jupyter Notebooks y se guardará en un archivo de formato. CSV en Excel para abrirlo y para poder determinar las causas de los datos que se han tenido de una base histórica de atención de tickets que ingresan los usuarios. Se clasificará por proceso, capacitación y error de sistema; esto ayudará a determinar el nivel de rendimiento.

2.3. Determinación de datos

En esta etapa haremos la recolección de datos, donde se recolectarán datos de la empresa para poder determinar los resultados y sean fiables y así poder determinar el nivel de rendimiento del personal. Para ello, estaremos aplicando regresión lineal para poder predecir cómo poder reducir la cantidad de ticket por proceso.

En la figura 5 se muestra la base de datos de sysaid en la cual mostramos todos los campos que contiene dicha base de datos con el nombre *SysAidAna*, el cual está compuesto por las siguientes columnas.

Figura 5. Base de datos



Column Name	Data Type	Nullability
Tipo Servicio	varchar(50)	NULL
Reabrir Ticket	char(2)	NULL
Ticket	int	NULL
Grupo Administradores	varchar(150)	NULL
Gestor Proceso	nvarchar(150)	NULL
Id Usuario	varchar(50)	NULL
Usuario Solicitante	nvarchar(150)	NULL
Oficina	varchar(150)	NULL
Id Agencia	char(3)	NULL
Agencia	varchar(150)	NULL
Region	varchar(100)	NULL
Division	varchar(100)	NULL
Estado	varchar(100)	NULL
Prioridad	varchar(60)	NULL
Categoria	varchar(200)	NULL
Sub Categoria	varchar(200)	NULL
Categoria Tercer Nivel	varchar(200)	NULL
Descripcion	varchar(max)	NULL
Telefono	varchar(30)	NULL
Año	int	NULL
Id Mes	int	NULL
Mes	varchar(20)	NULL
Id Dia	int	NULL
Dia	varchar(2)	NULL
Fecha Solicitud	datetime	NULL
Fecha Vencimiento	datetime	NULL
Fecha Cierre	datetime	NULL
Cumplimiento SLA	varchar(15)	NULL
Motivo	varchar(100)	NULL
Impacto Cliente	char(2)	NULL
Resultado Cambio	char(2)	NULL
Tipo Incidencia	varchar(25)	NULL
Area Atencion	varchar(100)	NULL
Latitud	numeric(15,10)	NULL
Longitud	numeric(15,10)	NULL
Departamento	varchar(100)	NULL
Provincia	varchar(100)	NULL
Distrito	varchar(100)	NULL
reopen_counter	int	NULL
Origen	varchar(100)	NULL
Login GP	varchar(100)	NULL
Assign_counter	int	NULL
Status	varchar(30)	NULL
Titulo	varchar(255)	NULL
solution	varchar(max)	NULL
Hora	varchar(2)	NULL
Dia Sem	varchar(15)	NULL
Day wk	varchar(10)	NULL
Encuesta	varchar(50)	NULL
Archivado	varchar(15)	NULL
Tipo_User_SR	varchar(50)	NULL
Motivo Cancelacion	varchar(255)	NULL
Mes Resol	varchar(15)	NULL
Day wk Resol	varchar(10)	NULL
Id Mes Resol	int	NULL
Id Dia Resol	int	NULL
Año Resol	int	NULL
Nombre User	varchar(100)	NULL
Celular User	varchar(15)	NULL
Mail User	varchar(100)	NULL
Oficina User	varchar(100)	NULL
Perfil User	varchar(100)	NULL
Tipo Encuesta	char(2)	NULL
Tiempo Atencion	numeric(15,2)	NULL
NuDold	varchar(10)	NULL
Notas	varchar(max)	NULL
Id Hora Resol	int	NULL
Hora Resol	char(2)	NULL

Nota: Elaboración propia.

Aquí se tendrá la base de datos donde se estarán depurando y limpiando, lo que significa que se tendrá una nueva base, en donde permitirá utilizar la información necesaria, solo con las columnas importantes como el número de incidente, oficina, prioridad, año, motivo y cumplimiento Sla, para poder determinar las causas y así tener los datos apropiados.

En esta etapa, es necesario contar con información necesaria de varios periodos de años en la empresa, para tener un muestreo y determinar el tipo de motivo del incidente, la cantidad de incidentes generados y otros que ayuden con la productividad para poder responder la pregunta de investigación.

La data recopilada que se tendrá será valiosa para poder determinar las causas después de un análisis, el cual será de gran utilidad para poder conocer el comportamiento del personal de la agencia, tanto nuevos como antiguos.

2.4. Ingeniería de Características

En este punto, recopilaremos la información necesaria de una financiera para poder desarrollar el propósito de la investigación.

En las siguientes figuras mostraremos el detalle de la información extraída de la cantidad de datos que se utilizarán para el desarrollo de la investigación, donde podremos encontrar ya una data limpia y filtrada de los campos que vamos a utilizar y nos servirán para poder determinar el nivel de rendimiento del personal. Se mostrarán los datos por cada año de los tickets ingresados por los mismos usuarios de la financiera, indicando que solo se están tomando incidencias.

En la tabla 2 podremos ver los datos que utilizaremos por cada año donde se visualiza la cantidad de tickets registrados. En el año 2017, encontramos 10644 registros, en el año 2018, encontramos 12551 registros, en el año 2019 encontramos 20821, en el año 2020 encontramos 46937 y por último en el año 2021 encontramos 43675.

Tabla 2. Archivo por años

Etiquetas de fila	Cuenta de Ticket
2017	10644
2018	12551
2019	20821
2020	46937
2021	43675
Total general	134628

Nota. Elaboración propia.

Figura 6. Cantidad de ticket



Nota. Elaboración propia.

En la figura 7 se procede a determinar la cantidad de tickets ingresados en todos estos 5 años. Según los datos, se interpreta que el total de tickets es de 134628 tickets registrados, en la cual el mínimo de ticket ingresado es de 39594 en el 2017 y el máximo 520304 en el 2021. También encontramos el valor promedio (media) de 295910.03 teniendo una desviación estándar 137322.55.

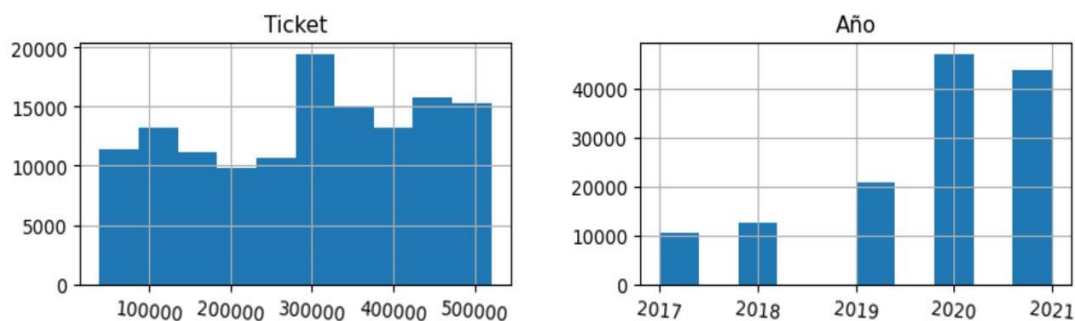
Figura 7. *Resumen de data*

	Ticket	Año
count	134628.000000	134628.000000
mean	295910.036538	2019.746115
std	137322.555870	1.224374
min	39594.000000	2017.000000
25%	174579.500000	2019.000000
50%	304017.500000	2020.000000
75%	413570.000000	2021.000000
max	520304.000000	2021.000000

Nota: Elaboración propia.

En la siguiente figura 8 podemos interpretar la cantidad de tickets que pueden generarse en los transcurso de los años donde unos se mantienen ligeramente y en otros incrementa imprevisiblemente, lo que nos puede indicar que está sucediendo algo inesperado o fortuito, ya sea por una capacitación al usuario por una actualización o un error por la actualización. En la siguiente imagen podemos ver la cantidad de tickets ingresados por año cómo incrementa con pasar de los años. Con ello inferimos que, si el personal incrementa y la financiera también, por lo cual la cantidad de tickets por año también incrementará.

Figura 8. *Incidencias*



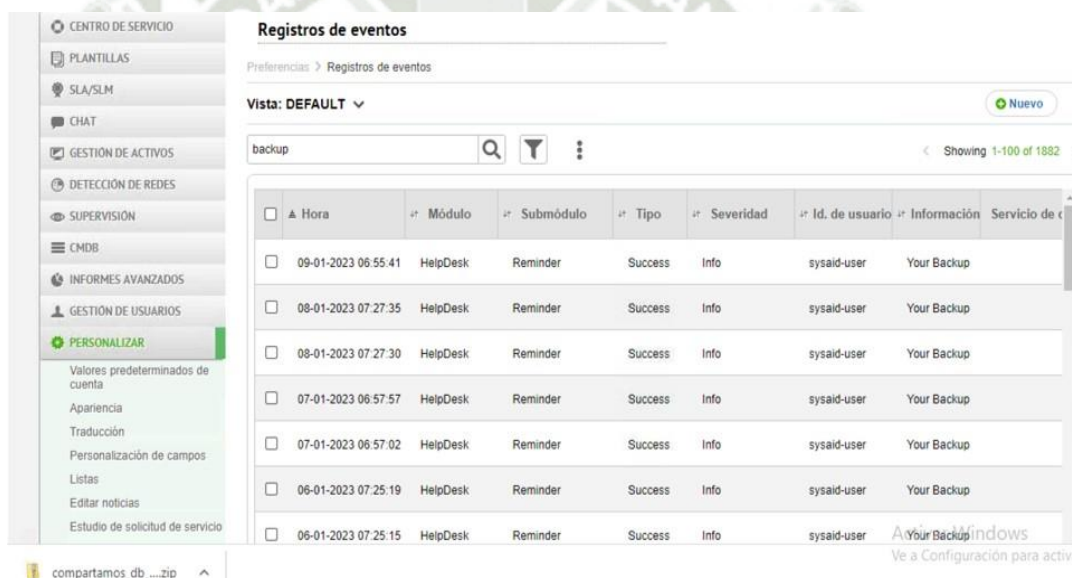
Nota: Elaboración propia.

2.4.1. Insumos

El presente trabajo se basó en extraer la información de la plataforma sysaid sobre las incidencias registradas para poder analizar la información como se muestra en las siguientes imágenes.

En la figura 9 se muestra cómo será extraída la data de la plataforma sysaid ingresando como administrador, luego pondremos a buscar el archivo backup para poder descargar toda la información almacenada que registraron el personal cuando presentaron una incidencia.

Figura 9. *Extracción de datos*



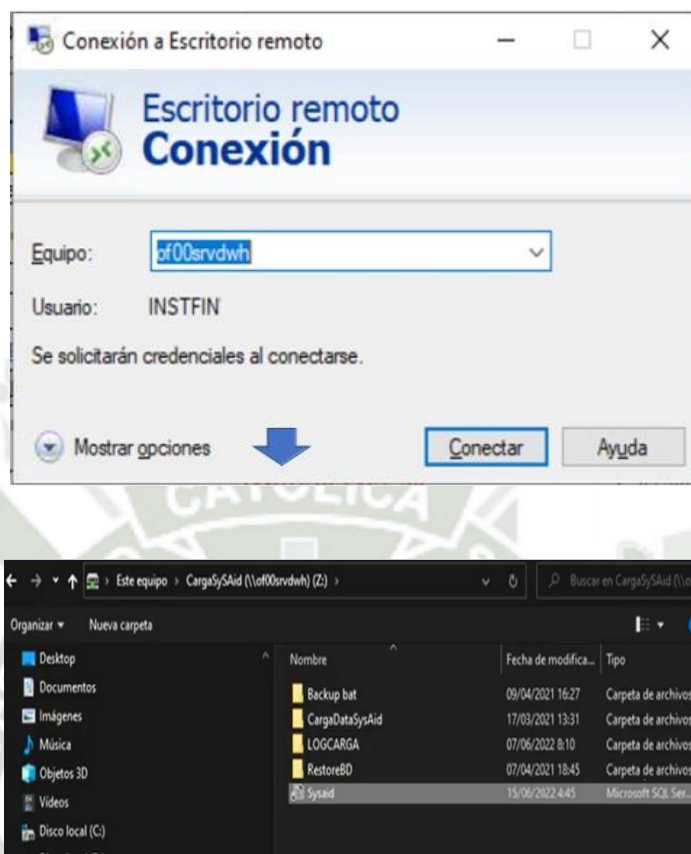
The screenshot shows the Sysaid 'Registros de eventos' (Event Records) page. The search bar contains the text 'backup'. The results table displays the following data:

<input type="checkbox"/>	▲ Hora	↕ Módulo	↕ Submódulo	↕ Tipo	↕ Severidad	↕ Id. de usuario	↕ Información	Servicio de c
<input type="checkbox"/>	09-01-2023 06:55:41	HelpDesk	Reminder	Success	Info	sysaid-user	Your Backup	
<input type="checkbox"/>	08-01-2023 07:27:35	HelpDesk	Reminder	Success	Info	sysaid-user	Your Backup	
<input type="checkbox"/>	08-01-2023 07:27:30	HelpDesk	Reminder	Success	Info	sysaid-user	Your Backup	
<input type="checkbox"/>	07-01-2023 06:57:57	HelpDesk	Reminder	Success	Info	sysaid-user	Your Backup	
<input type="checkbox"/>	07-01-2023 06:57:02	HelpDesk	Reminder	Success	Info	sysaid-user	Your Backup	
<input type="checkbox"/>	06-01-2023 07:25:19	HelpDesk	Reminder	Success	Info	sysaid-user	Your Backup	
<input type="checkbox"/>	06-01-2023 07:25:15	HelpDesk	Reminder	Success	Info	sysaid-user	Your Backup	

Nota: Elaboración propia.

En la figura 10 se procederá a copiar la información descargada con nombre de sysaid al servidor de la empresa con nombre of00srvdwh donde se almacena la información de gran cantidad de datos para mantenerlos de forma segura; luego se procederá a ingresar al servidor donde se colocó la información, en el cual se extraerá dicha información comprimida.

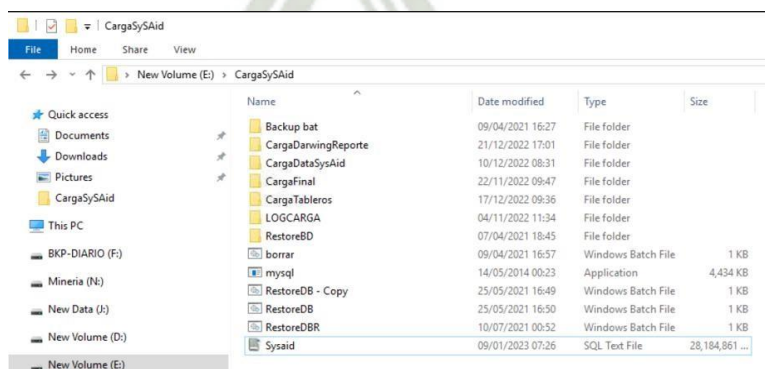
Figura 10. Descarga de datos



Nota: Elaboración propia.

En la figura 11 tenemos la información ya descargada y lista para poder ser procesada y visualizada en data.

Figura 11. Base de datos sysaid



Nota: Elaboración propia

2.4.2. Procesamiento

Los datos obtenidos usando ETL lo que hará es extraer, transformar y cargar una nueva base de datos donde se podrá ya realizar el análisis o la extracción de información procesada. Luego de la extracción de datos previamente mencionado se procederá a realizar los filtros necesarios para que la información sea mostrada en el Power BI como en Excel.

2.4.3. Almacenamiento

El almacenamiento diario, que se extrae de la información del sysaid, se copia a un servidor de una datawarehouse para tener el alcance de un análisis diario al área de soporte y tecnología de la información para poder medir las metas y objetivos planteados.

2.4.4. Destino

El destino de la información se podrá tener listo en un reporte de Power BI, el cual se podrá extraer dicha de Excel o en gráficos, que permitirá hacer el análisis y poder procesar los datos en Python para el presente proyecto.

2.5. Revisión y depuración de datos

Después de extraer la base de datos del sysaid se procederá a hacer la extracción de datos y la filtración de información por el SQL server como se muestra en la figura 12 para hacer el proceso de filtración para procesar dicha información.

Figura 12. *Filtración de información*

```
SELECT
A.[Tipo Servicio], A.[Grupo Administradores],a.Motivo, A.[Gestor Proceso],
A.Oficina, A.Prioridad, A.Año, A.Mes,
A.[Fecha Solicitud], A.[Fecha Vencimiento], A.[Cumplimiento SLA], A.Status,
A.[Oficina User], A.[Perfil User], a.[Area Atencion]
FROM SysAidAna A
WHERE Origen='Email Integration'
and A.[Fecha Solicitud]>='01/01/2016'
AND [Area Atencion] in ('TECNOLOGIA DE LA INFORMACIÓN')
AND [Tipo Servicio] ='Incidente'
and [Grupo Administradores] ='MESA DE SERVICIO'
```

Nota: Elaboración propia.

En las siguientes figuras se mostrarán los resultados obtenidos por el Power BI de cada año. Encontraremos el total de tickets ingresados por el personal, por motivo de un incidente si fue por error, capacitación y proceso en una financiera, donde el mismo personal registra su ticket por cada inconveniente ocurrido en el transcurso del día y puedan ser atendidos en su debido momento. Después de extraída la información se procederá a utilizar la información para poder determinar el porcentaje del motivo de ingreso de un incidente y determinar el nivel de rendimiento del personal.

En la figura 13 podremos ver el total de tickets ingresados en el año 2017 por el Power BI, en el cual veremos que la mayor parte de incidentes fue por el motivo de proceso con 6113 ingresados.

Figura 13. *Ticket por motivo*



Nota: Elaboración propia.

En la figura 14 podremos ver el total de tickets ingresados en el año 2018 por el Power BI, en el cual veremos que la mayor parte de incidentes fueron por el motivo del proceso de igual forma que el año anterior, con una cantidad de tickets de 6991.

Figura 14. *Ticket por motivo*



Nota: Elaboración propia.

En la figura 15 podremos ver el total de tickets ingresados en el año 2019 por el Power BI, el cual veremos que la mayor parte de incidentes fueron por el motivo del proceso con 13392 tickets.

Figura 15. *Ticket por motivo*



Nota: Elaboración propia.

En la figura 16 podremos ver el total de tickets ingresados en el año 2020 por el Power BI, el cual veremos que la mayor parte de incidentes fueron por el motivo del error de sistema con 17739 registros.

Figura 16. *Ticket por motivo*



Nota: Elaboración propia

En la figura 17 podremos ver el total de tickets ingresados en el año 2021 por el Power BI, el cual veremos que la mayor parte de incidentes fue por el motivo del error de sistema con 32477 ticket.

Figura 17. *Ticket por motivo*



Nota: Elaboración propia.

CAPITULO III: RESULTADO DE LA INVESTIGACIÓN

3. RESULTADO DE LA INVESTIGACIÓN

En este punto se detallará el procedimiento que se utilizará aplicando data science para poder definir el nivel de rendimiento del personal y así determinarlas.

3.1. Pre procesamiento de la data

Para la elaboración de resultados se ha contado con información real de una financiera; la información obtenida se consiguió cumpliendo los parámetros de privacidad, salvaguardando datos sensibles del personal como el de la financiera.

En esta etapa utilizaremos la base de datos del sysaid ya extraída y limpia, como se especifica en el capítulo II, para poder ser utilizada en Python en Jupyter Notebooks, obteniendo los resultados y puedan ser analizados para determinar el nivel de rendimiento. Se procederá a utilizar la data del 2017 al 2021 y se procederá a elaborar en un archivo CSV.

Por otra parte, se procederá a utilizar el programa Anaconda, el cual ayudará para poder programar en Python utilizando Jupyter Notebooks para cargar la data sysaid.csv.

En la figura 18 podremos ver la data que se subió, en este caso el archivo sysaid.csv en Python para que pueda ser procesada, en este archivo juntaremos la información de todos los años del 2017 al 2021, en la que podemos encontrar un total de 134628 registros. Luego se procederá a importar las librerías para poder hacer la lectura del archivo CSV.

Figura 18. *Librerías y procesamiento de data*

```
In [1]: import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import pandas as pd

In [5]: pd.read_csv('sysaid.csv')

Out[5]:
```

	Ticket	Oficina	Prioridad	Año	Motivo	Cumplimiento SLA	Unnamed: 6	Unnamed: 7
0	58491	024 Agencia San Camilo	Bajo	2017	Capacitacion	Cumple SLA	NaN	NaN
1	58526	024 Agencia San Camilo	Bajo	2017	Capacitacion	Cumple SLA	NaN	NaN
2	59954	024 Agencia San Camilo	Planificada	2017	Proceso	Cumple SLA	NaN	NaN
3	59315	052 Agencia Moquegua	Bajo	2017	Capacitacion	Cumple SLA	NaN	NaN
4	60715	059 Agencia Zapallal	Bajo	2017	Error Sistema	Cumple SLA	NaN	NaN
...
134623	493111	005 Agencia Miraflores	Medio	2021	Error Sistema	Cumple SLA	NaN	NaN
134624	499094	999 CEAS LIMA - LEURO	Bajo	2021	NaN	Cumple SLA	NaN	NaN
134625	437081	024 Agencia San Camilo	Medio	2021	Error Sistema	Cumple SLA	NaN	NaN
134626	464881	068 Agencia La Union	Bajo	2021	Error Sistema	Cumple SLA	NaN	NaN
134627	464942	079 Agencia Villa el Salvador	Medio	2021	Error Sistema	Cumple SLA	NaN	NaN

134628 rows x 8 columns

Nota. Elaboración propia.

3.2. Instancias de prueba

Primero se procedió a validar los datos que se utilizarán de los 5 años de la financiera, como para no tener inconvenientes. Se procedió a verificar y realizar una última limpieza en Python para no tener valores nulos y si los hubiera eliminar y tener una data limpia. En este caso se puede ver que en la figura 19 se encontró data nula, pudiendo tener error en los resultados, por lo cual se procede a eliminar para así tener una data limpia.

Figura 19. *Procesamiento de data*

```
In [49]: s1.isnull().sum()

Out[49]: Ticket      0
Oficina      0
Prioridad     0
Año          0
Motivo      7503
Cumplimiento SLA  0
Unnamed: 6    134628
Unnamed: 7    134626
dtype: int64

In [57]: s1.isnull().sum()

Out[57]: Ticket      0
Oficina      0
Prioridad     0
Año          0
Motivo      0
Cumplimiento SLA  0
dtype: int64
```

Nota: Elaboración propia.

En la figura 20 tendremos la data limpia con 127125 registros en total de los años 2017 al 2021.

Figura 20. *Procesamiento*

Out[58]:

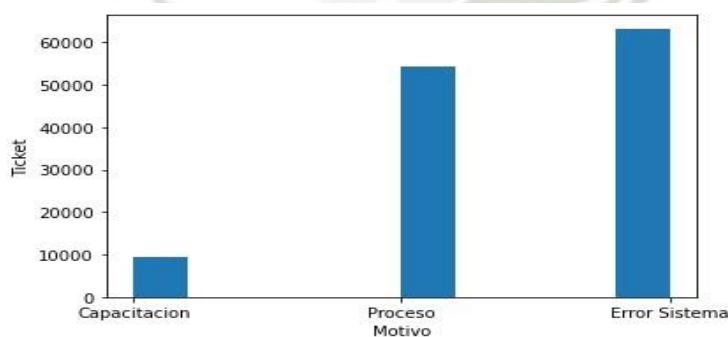
	Ticket	Oficina	Prioridad	Año	Motivo	Cumplimiento SLA
0	58491	024 Agencia San Camilo	Bajo	2017	Capacitacion	Cumple SLA
1	58526	024 Agencia San Camilo	Bajo	2017	Capacitacion	Cumple SLA
2	59954	024 Agencia San Camilo	Planificada	2017	Proceso	Cumple SLA
3	59315	052 Agencia Moquegua	Bajo	2017	Capacitacion	Cumple SLA
4	60715	059 Agencia Zapallal	Bajo	2017	Error Sistema	Cumple SLA
...
127120	463424	017 Agencia Ceres	Medio	2021	Error Sistema	Cumple SLA
127121	493111	005 Agencia Miraflores	Medio	2021	Error Sistema	Cumple SLA
127122	437081	024 Agencia San Camilo	Medio	2021	Error Sistema	Cumple SLA
127123	464881	068 Agencia La Union	Bajo	2021	Error Sistema	Cumple SLA
127124	464942	079 Agencia Villa el Salvador	Medio	2021	Error Sistema	Cumple SLA

127125 rows x 6 columns

Nota. Elaboración propia.

Después de tener una data limpia y comprobarla en Python, se procederá a realizar el análisis de visualización con la data cargada. En la figura 21 se ve un diagrama de los 5 años del total de los tickets registrados por motivo, donde la mayor parte de ingreso de incidentes fue por error de sistema; seguido de ello está el de proceso que no está muy alejado del registro antes mencionado, el cual muestra una gran cantidad de incidentes.

Figura 21. *Cantidad de ticket registrados*



Nota: Elaboración propia.

3.3. Visualización de datos

En las siguientes figuras se mostrará las estadísticas para poder determinar el nivel de rendimiento de los usuarios.

En la figura 22 se muestra la cantidad de tickets ingresados por años. En donde podemos ver que por cada año el ingreso de tickets fue incrementando, como se dijo anteriormente, mientras va creciendo, una financiera también incrementará el tema de ingresos de tickets y el ingreso del personal.

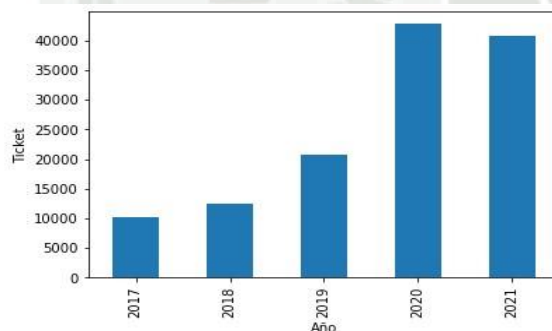
Figura 22. *Resumen de ticket por año*

Año	Ticket
2017	10257
2018	12515
2019	20782
2020	42799
2021	40772

Nota: Elaboración propia.

Por esto en la figura 23 se realiza la representación gráfica de tickets por años obtenidos en la figura anterior.

Figura 23. *Gráfico de ticket por año*



Nota: Elaboración propia.

En la figura 24 se procede a calcular la cantidad de tickets ingresados dentro de los 5 años por el motivo, en donde podemos observar que la cantidad de tickets por error son de 63283 y seguido está el de proceso con 54423 obteniendo una diferencia solo de 8860 tickets.

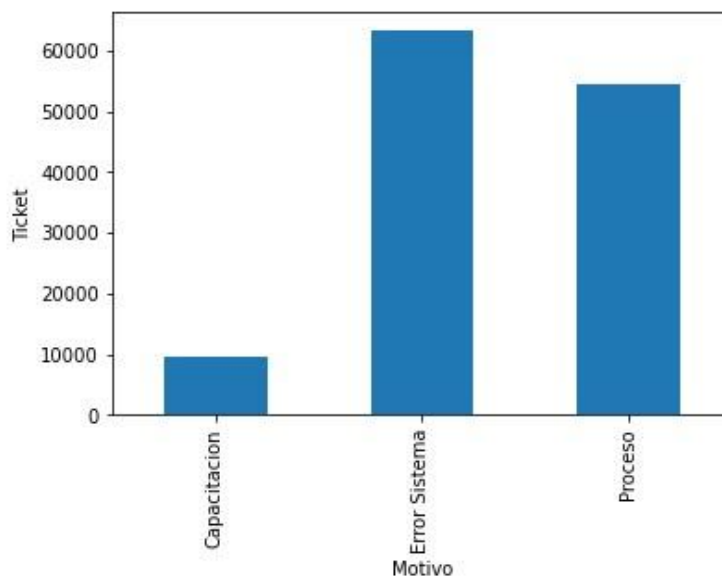
Figura 24. *Resumen por motivo*

Motivo	Ticket
Capacitacion	9419
Error Sistema	63283
Proceso	54423

Nota: Elaboración propia.

A continuación, en la figura 25 mostraremos la gráfica de la cantidad de tickets ingresados por motivo de la figura anterior.

Figura 25. *Gráfico de motivo*



Nota: Elaboración propia.

Se procede a realizar un análisis para poder comparar, si la gran cantidad de tickets ingresados cumplen con las SLA lo que significa si cumplieron con el servicio acordado o pactado con otras áreas de la financiera. En la figura 26 tendremos los resultados del cumplimiento del total de tickets ingresados por los usuarios dentro de los 5 años si se lograron atender según la SLA.

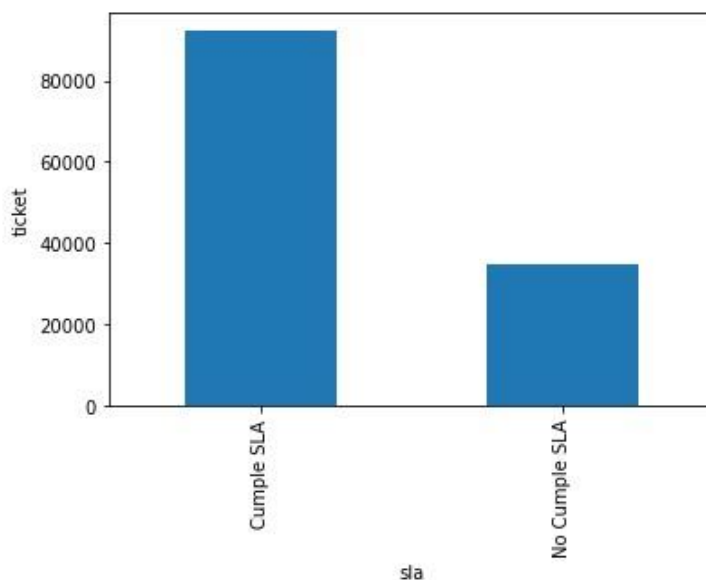
Figura 26. *Resumen de Sla*

Cumplimiento SLA	
Cumple SLA	92280
No Cumple SLA	34845

Nota: Elaboración propia

Asimismo, en la figura 27 se obtiene la gráfica del resultado anterior del cumplimiento de la SLA.

Figura 27. *Gráfico de Sla*



Nota: Elaboración propia.

En este punto se realizó el promedio de tickets ingresados por motivo y por año, donde vemos que el promedio de tickets fue obtenido por la cantidad de tickets de cada año, en el año 2018 casi a finales de año, el promedio de tickets fue por capacitación, también vemos que por promedio en el año 2020 a comienzos hubo más errores de sistema y por último en el año 2019 a comienzos fue por proceso.

En la figura 28 se muestra el promedio de motivos de ticket ingresados por año.

Figura 28. *Promedio por motivo*

Motivo	Promedio
Capacitacion	2018.981845
Error Sistema	2020.356920
Proceso	2019.103173

Nota: Elaboración propia.

3.4. Aplicación del método de Machine Learning

3.4.1. Aplicación del método de Regresión Lineal

Luego se procederá a definir el modelo de ejecución aplicando Machine Learning lo que significa identificar patrones para poder generar predicciones con la inteligencia artificial, La figura 29 se muestra la fórmula para aplicar el modelo de regresión.

Figura 29. *Modelo Regresión Lineal*

$$Y = mX + b$$

Nota. Elaboración propia.

Para poder trabajar con la información del sysaid extraída, se procederá a convertir los datos que son caracteres a numéricos como se muestra en la siguiente figura 30.

Figura 30. *Datos*

	Ticket	Oficina	Prioridad	Año	Motivo	Cumplimiento SLA
0	58491	21	1	2017	0	0
1	58526	21	1	2017	0	0
2	59954	21	4	2017	2	0
3	59315	48	1	2017	0	0
4	60715	53	1	2017	1	0
...
127120	463424	16	3	2021	1	0
127121	493111	4	3	2021	1	0
127122	437081	21	3	2021	1	0
127123	464881	62	1	2021	1	0
127124	464942	73	3	2021	1	0

Nota. Elaboración propia.

Para poder aplicar maching learning se procederá a convertir los datos en números en la tabla 3 podemos encontrar como se clasifica cada motivo en número.

Tabla 3. *Conversión de datos Motivo*

Etiqueta	Valor numérico
Capacitación	0
Error de sistema	1
Proceso	2

Nota. Elaboración propia.

En la tabla 4 podemos encontrar como se clasifica el campo cumplimiento en número.

Tabla 4. *Conversión de datos Cumplimiento*

Etiqueta	Valor numérico
Cumple	0
No cumple	1

Nota. Elaboración propia.

En la figura 31, se procederá a aplicar el modelo de regresión de la data que se tiene del sysaid, en este caso de procederá a separar X y Y para reestructurarlos en datos columnas para que estén en una matriz para poder aplicar maching learning, al definir las variables en lista, para luego hacer un entrenamiento y un test con la fórmula que nos brinda el data science, en este caso para hacer entrenamiento para que con el restante pueda hacer test por ello se usará el 20% de test para poder hacer la pruebas aleatorias con la información, donde X son los años, Y es el motivo en el cual se obtiene un conjunto de datos, por último se procede a aplicar regresión lineal obteniendo así los coeficientes y la intersección lineal del modelo aplicado.

Podemos ver que según los resultados obtenidos, si el coeficiente es negativo y la variable de intercepción es inversa, podemos decir que sí hay una mejora y que el modelo aplicado de regresión lineal es efectivo para este proyecto.

Figura 31. *Regresión lineal*

```
In [10]: X=np.array(a).reshape(-1,1)
        Y=np.array(b).reshape(-1,1)

In [16]: X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=0.2, random_state=80)

In [17]: y_test.shape
Out[17]: (25425, 1)

In [18]: modelo = LinearRegression()

In [19]: modelo.fit(X_train,y_train)
Out[19]: LinearRegression
         LinearRegression()

In [20]: modelo.coef_
Out[20]: array([[ -0.13713642]])

In [21]: modelo.intercept_
Out[21]: array([278.3310045])
```

Nota: Elaboración propia.

En la figura 32 luego se hará las predicciones sobre X con los resultados obtenidos en la regresión lineal, en donde concluimos que se van a acercando a ticket por proceso.

Figura 32. *Regresión lineal*

```
r=modelo.predict(X_test)

array([[1.17829756],
       [1.72684325],
       [1.31543398],
       ...,
       [1.31543398],
       [1.31543398],
       [1.17829756]])
```

Nota: Elaboración propia.

Posterior comparemos respuestas con Y con las predicciones que se obtuvo en X, lo cual se procederá a comparar con métricas la cual utilizaremos será el MAE (Mean Absolute Error) donde obtendremos el promedio entre los datos que tenemos de los registros de ticket y el de predicción como se muestra en la figura 33.

Figura 33. *Mean Absolute Error*

$$MAE(y, \hat{y}) = \frac{1}{m} \sum_{i=1}^n |y_i - \hat{y}_i|$$

```
print('Mean Absolute Error:', metrics.mean_absolute_error(y_test, r))
```

Mean Absolute Error: 0.48107344105723304

Nota: Elaboración propia.

Posterior en la figura 34 tendremos el MSE (Mean Squared Error) donde obtendremos el promedio de los valores estimados y el valor real.

Figura 34. *Mean Squared Error*

$$MSE = \left(\frac{1}{n}\right) \sum_{i=1}^n (y_i - x_i)^2$$

```
print('Mean Squared Error:', metrics.mean_squared_error(y_test, r))
```

Mean Absolute Error: 0.3487315114489374

Nota: Elaboración propia.

Luego de tener el promedio en la figura 35 procederemos a sacar el RMSE(Root Mean Squared Error) la raíz cuadrada del resultado obtenido anteriormente.

Figura 35. *Root Mean Squared Error*

$$RMSE = \sqrt{\left(\frac{1}{n}\right) \sum_{i=1}^n (y_i - x_i)^2}$$

```
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(y_test, r)))
```

Mean Absolute Error: 0.5905349366878622

Nota: Elaboración propia.

Al tener una data limpia y aplicar el modelo de regresión lineal haciendo una separación de variables de X y Y posterior a ello un entrenamiento usando el 80% y para hacer el test con el 20% para luego verificar los resultados sobre las variables, concluimos que al usar regresión lineal y validar el resultado obtenido con la métrica de MSE (Mean Squared Error) donde obtuvimos como resultado de 0.3487315114489374, donde nuestro error de promedio es pequeño donde comprobamos que es efectivo mientras el resultado es más pequeño es efectivo el modelo aplicado, lo que afirma que el margen de error es muy bajo, lo que confirma que los usuarios no están bien capacitados en su flujo.

3.4.2. Aplicación del método de Clasificación

A continuación, aplicaremos el siguiente modelo de clasificación donde podremos definir por etiquetas, aquí también confirmamos la falta de proceso en el personal. Para este método se usarán los mismos datos que se usó en regresión lineal. En este caso, para hacer el test, se usará el 20% para poder hacer las pruebas aleatorias con la información donde X son todos los campos como se muestra en la Figura 36 y nuestra Y será el campo de solo motivo. Por último, se procede a aplicar regresión logística obteniendo así los resultados para luego hacer un entrenamiento sobre el modelo aplicado. En este caso, sobre el resultado se predicen los tickets que ingresarán por etiquetas, luego se hace una comparación con Y sobre los resultados que confirman la falta de proceso del usuario.

Figura 36. *Regresión logística*

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=80)
```

```
X_train.shape
```

```
(101700, 5)
```

```
modelo=LogisticRegression()
```

```
modelo.fit(X_train,y_train)
```

```
▼ LogisticRegression
```

```
LogisticRegression()
```

```
res=modelo.predict(X_test)
```

```
res
```

```
array(['Error Sistema', 'Proceso', 'Error Sistema', ..., 'Proceso',  
      'Error Sistema', 'Error Sistema'], dtype=object)
```

```
y_test
```

```
array(['Proceso', 'Proceso', 'Proceso', ..., 'Proceso', 'Error Sistema',  
      'Error Sistema'], dtype=object)
```

Nota: Elaboración propia.

Después de los resultados obtenidos se procederá aplicar métrica para poder medir la efectividad del clasificador.

En la figura 37 podremos obtener el reporte de clasificación donde tenemos una precisión de error de sistema es del 74% y de proceso de 72 % donde el total de datos procesados la exactitud de respuesta es del 73 por ciento sobre 25425.

Figura 37. Reporte de clasificación

```
print (metrics.classification_report(y_test,res))
```

	precision	recall	f1-score	support
Capacitacion	0.00	0.00	0.00	1885
Error Sistema	0.74	0.85	0.79	12667
Proceso	0.72	0.72	0.72	10873
accuracy			0.73	25425
macro avg	0.49	0.52	0.50	25425
weighted avg	0.67	0.73	0.70	25425

```
print('accuracy is',accuracy_score(y_test,res))
```

```
accuracy is 0.7297541789577188
```

Nota: Elaboración propia

En la siguiente figura 38 usamos el modelo clasificador Naive Bayes. Es un modelo donde podemos tener más exactitud.

También encontramos que tenemos una precisión de error de sistema del 98% y de proceso de 95 %. En el total de datos procesados, la exactitud de respuesta es del 97 por ciento sobre 25425.

Figura 38. *Reporte de clasificación*

```

: classifier =GaussianNB()
  classifier.fit(X_train, y_train)
  y_test=classifier.predict(X_test)

```

```

: metrics.confusion_matrix(y_test,res)
  print (metrics.classification_report(y_test,res))
  print('accuracy is',accuracy_score(y_test,res))

```

	precision	recall	f1-score	support
Error Sistema	0.98	0.97	0.97	14759
Proceso	0.95	0.97	0.96	10666
accuracy			0.97	25425
macro avg	0.96	0.97	0.97	25425
weighted avg	0.97	0.97	0.97	25425

accuracy is 0.9663716814159292

Nota: Elaboración propia.

Aplicar el modelo de clasificación haciendo una separación de variables de X y Y posterior a ello un entrenamiento usando el 80% y para hacer el test con el 20% para luego verificar los resultados sobre las variables, concluimos que al usar clasificación y validar el resultado obtenido usando la métrica de efectividad de gaussiannb donde obtuvimos como resultado del 97% de exactitud, el cual es super alta. En este caso, si el resultado es al 100%, es muy efectiva donde comprobamos nuevamente que es efectivo y donde concluimos que hay una falta de capacitación.

3.4.3. Comparación de Modelos

En la siguiente tabla 5 podemos ver los resultados obtenidos en ambos modelos. En el modelo de regresión lineal vemos que aplicando las tres métricas la más efectiva es el mean squared error, obteniendo 0.38 si nuestro valor es más pequeño, es más efectivo y confirmando que el modelo aplicado es efectivo. En el modelo de clasificación se aplicaron 2 métricas, de las cuales la más efectiva fue la de clasificador Naive Bayes donde obtenemos un 97 %. En este modelo, si está cerca al 100 es efectivo. En este caso, sobre los 2 modelos aplicados vemos que el mejor fue el de regresión logística aplicando Naive Bayes puesto que estamos confirmando que hay un bajo nivel de rendimiento.

Tabla 5. Comparativo de modelos

Modelo	Concepto	Métricas	Resultado
Regresión lineal	Funcionan únicamente con datos numéricos. Para comprobar el modelo, la métrica debe ser pequeño el resultado obtenido	Mean Absolute Error	Promedio de los datos registrados con la predicción 0.48
		Mean Squared Error	Promedio de los valores estimados y el valor real 0.38
		Root Mean Squared Error	La raíz cuadra del resultado obtenido 0.59
Regresión Logística	Predice una categoría/clase. Para comprobar que el modelo sea 100% efectivo el resultado debe tener el 100%	Métricas de clasificación	Precisión de exactitud 0.73%
		Clasificador Naive Bayes	Precisión exactitud del 97 %

Nota: En la siguiente comparación se valida los resultados obtenidos con los dos modelos utilizados, el objetivo es ver los resultados obtenidos y cuál es el más eficiente “Comparativo de modelos”, elaboración propia.

3.5. Validación de resultados

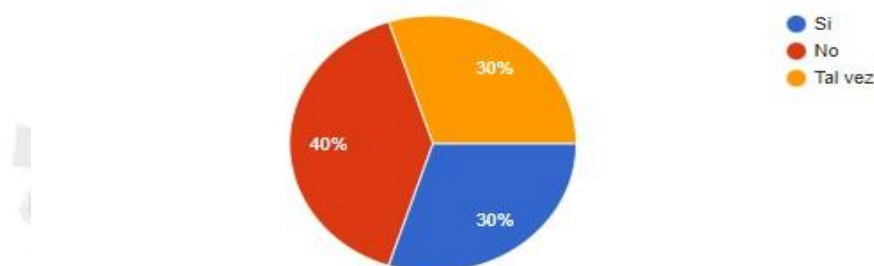
Después de realizar las estadísticas de la base de datos del sysaid se realizó una encuesta al personal, en la cual se obtuvieron los siguientes resultados.

En la figura 39 se puede ver que el 30% del personal piensa que son errores del sistema y el otro 30% duda de su respuesta, por lo que el 60% del personal ingresa un ticket para que

pueda ser ayudado sin saber que realmente es un error o un tema de proceso. Por lo tanto, vemos que, si un usuario conoce bien el proceso, no registrará un ticket, lo que generará que haya menos tickets. La intersección bajará y con ello podemos indicar que, si hay una buena capacitación, habrá menos tickets de proceso.

Figura 39. Encuesta 1

2. Todos los mensajes que te muestra el sistema piensas que son errores.

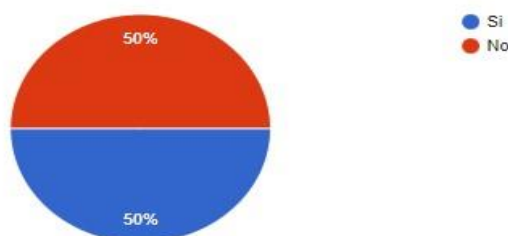


Nota: Elaboración propia.

En la figura 40 se logra ver una de las causas por las que el usuario tiene un bajo rendimiento, puesto que la mitad de los usuarios no revisa los manuales o normativas de la empresa. Lo cual comprueba que, si un usuario no revisa la normativa primero, también genere ticket por proceso.

Figura 40. Encuesta 2

3. Antes de reportar un ticket sobre el mensaje que te muestra el sistema, revisas tu manual.



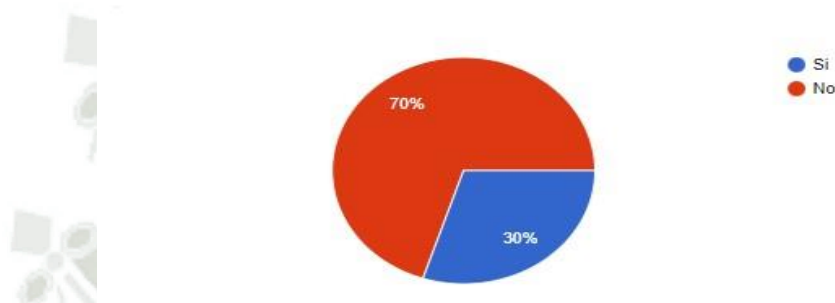
Nota: Elaboración propia

En la figura 41 Se logra ver que una de las causas más importantes es que el usuario no se siente capacitado para manejar todos los flujos del sistema, lo que hace que el personal

no sea polifuncional en su trabajo y haya un bajo rendimiento de los usuarios. En este caso, según la encuesta, se valida que el 70% del personal sostiene que no está capacitado en todos los flujos del sistema y ese porcentaje ocasiona también que se registren tickets por no saber el proceso o el flujo del sistema.

Figura 41. Encuesta 3

5. Sientes que estas capacitado en todos los flujos del sistema.

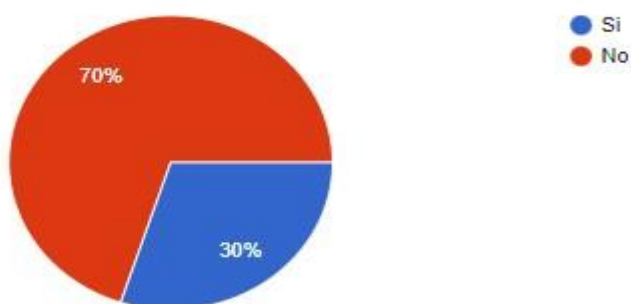


Nota: Elaboración propia

También se muestra que en la figura 42 podemos ver que no solo es culpa del usuario, sino que los mensajes mostrados por el sistema no son claros para ellos, lo que ocasiona que ese porcentaje genere ticket al no quedarle claro el mensaje mostrado por el sistema. En este caso el 70 % de colaboradores indica que los mensajes no son claros.

Figura 42. Encuesta 4

7. Las alertas que muestran en el sistema son claras.



Nota: Elaboración propia.

CONCLUSIONES

1. Se logró aplicar técnicas de data science y se confirmó que después de aplicar machine learning y ver los resultados obtenidos al aplicar la técnica de regresión lineal; se identificó un margen de error mínimo del 0.3 %. Así mismo, al aplicar la técnica de clasificación, confirmamos que el 97 % de efectividad en el pronóstico al predecir la cantidad de tickets ingresados por cada categoría.
2. Se recolectó toda la información que posteriormente se empleó en todo el desarrollo, donde se puede concluir que todos los datos recolectados de la financiera durante 5 años, dentro de los periodos de los años 2017 al 2021 se encontraron 134628 tickets; posterior a la aplicación de las técnicas de data science se obtuvo un valor de 127125, la cual representa una data limpia. Esta acción de limpiar los datos nos da una mayor certeza de los resultados; esto nos permite tomar acciones óptimas sobre la data recolectada.
3. Se logró validar dos modelos para hacer un análisis y asegurar los resultados. Se procede a aplicar regresión lineal y regresión logística; para ello se realizó una verificación con métricas de efectividad. Al tener ambos resultados eficaces, se comprobó el bajo nivel de rendimiento del uso del sistema de créditos y cuando un parámetro se eleve, se pueden determinar las causas y así tomar las acciones en ese momento.
4. Se comparó el uso de dos modelos aplicando machine learning donde concluimos que, al usar regresión lineal, la métrica de MSE (Mean Squared Error) obtuvimos un 0.3487315114489374 y en clasificación tenemos un

accuracy(exactitud) del 97 % comparando con las encuestas realizadas a los usuarios, determinamos que el personal tiene un bajo nivel de rendimiento.

5. Se determino que existe bajo rendimiento en el personal de acuerdo con los modelos aplicados, se concluye, que ambos modelos son efectivos; comparando con la encuesta realizada al personal, se valida que el 70% no se sintió capacitado en todos los flujos del sistema, donde el 40% también pensó que todas las alertas que le salían eran errores sin antes revisar sus manuales de capacitación obteniendo ahí que el 50% no lee sus manuales para sus procesos, confirmando así que el sistema presentaba complicaciones donde el usuario no se adaptaba fácilmente, confirmando que no se encontraba completamente integrado el sistema y con ello provocando que el usuario no tenga la confianza necesaria para ser utilizado, con estas falencias que se presentan y con los resultados obtenidos; se confirma que el personal no está completamente capacitado en todos los flujos de la empresa y que el sistema no era muy asequible para los usuarios, por lo cual dicho sistema fue cambiado para que pueda ser más factible con el usuario.

RECOMENDACIONES

1. Se recomienda analizar la información que almacena la plataforma sysaid sobre todo el campo de motivo de incidente y no solo los otros parámetros de gestión que tiene determinado la financiera, para así poder determinar el motivo de las causas del ticket de capacitación o proceso.
2. Hacer un reporte de la cantidad de tickets por motivo ayudará a poder analizar la información y poder evitar tickets por proceso y tomar las correcciones respectivas sobre los indicadores que ya están preestablecidos por la financiera.
3. Tener capacitados a los colaboradores de una financiera permitirá que no haya pérdida de tiempo en la atención de incidentes importantes y generar una atención rápida al cliente.
4. Utilizar herramientas adecuadas para una adecuada capacitación para cada cargo.
5. Tener un personal siempre capacitado para todas las funciones para que pueda ser polifuncional, para que cuando un personal salga de vacaciones o licencia cumpla con la función y tenga el conocimiento en el proceso.

REFERENCIAS

- Arcila, C. (2016). *TÉCNICAS BIG DATA: ANÁLISIS DE TEXTOS A GRAN ESCALA PARA LA INVESTIGACIÓN CIENTÍFICA Y PERIODÍSTICA*. España. Obtenido de http://eprints.rclis.org/34193/1/EPI_2016_25_4.pdf.pdf
- Bern, J. (2020). Obtenido de fti.unai.edu: <https://fti.unai.edu/wp-content/uploads/2020/02/Introduction-to-Data-Science.pdf> blog mdcloud. (s.f.). *mdcloud*. Recuperado el 11 de Mayo de 2022, de <https://blog.mdcloud.es/procesos-etl-definicion-desarrollo-aplicaciones/>
- Bohórquez, D. (2018). *Funcionamiento del Sistema Bancario*. Obtenido de https://www.researchgate.net/publication/324896570_Notas_para_entender_el_funcionamiento_del_Sistema_Financiero_colombiano/citation/download
- Constain, S., Agudelo, M., & Pico, C. (noviembre de 2019). G.SIS.04 Guía de Arquitectura de Soluciones Tecnológicas. *El futuro digital es de todos*, 2-61. Obtenido de https://www.mintic.gov.co/arquitecturati/630/articles117954_recurso_pdf.pdf
- Delgado, M. (2021). *ANÁLISIS DE LAS CAPACIDADES DE PLANIFICACIÓN*. Guayaquil. Obtenido de <http://201.159.223.180/bitstream/3317/17658/1/TUCSG-POS-MFEE-256.pdf>
- Díaz, N., Maestre, M., & Romero, F. (2018). Generación de valor: factor clave en la toma de decisiones de las pymes. *Económicas CUC*, 9-24. doi: <http://dx.doi.org/10.17981/econcuc.39.2.2018.01>
- Economipedia. (2022). *Economipedia*. Recuperado el 27 de Mayo de 2022, de <https://economipedia.com/definiciones/sistema-bancario.html>
- GDM. (13 de Mayo de 2022). *GDM*. Obtenido de <https://blog.gdm.com.mx/blog/tiposde-indicadores-para-evaluar-desempe%C3%B1o-del-personal>
- Gómez, O. (2017). *La importancia de la capacitacion en los asesores de confiar para el buen funcionamiento del proceso de analisis de los creditos*. Medellin.
- Gomez, O. A. (2017). *LA IMPORTANCIA DE LA CAPACITACIÓN EN LOS ASESORES DE CONFIAR PARA EL BUEN FUNCIONAMIENTO DEL PROCESO DE ANALISIS DE LOS CREDITOS*. Medellin.
- Frnka-Davis, L. (20 de Abril de 2020). Obtenido de www.utphysicians.com: <https://www.utphysicians.com/las-ventajas-de-la-tecnologia-en-el-manejo-de-la-salud-mental/>

Galicia, R. (2 de junio de 2021). Obtenido de www.airavirtual.com:

<https://www.airavirtual.com/blog/herramientas-disenador-ux>

Kroese, D., Botev, Z., & Taimre, T. (2020). *Data Science and Machine Learning. Mathematical and Statistical Methods*. New York: Chapman and Hall/CRC.
doi:<https://doi.org/10.1201/9780367816971>

KYOCERA Document Solutions. (2021). *KYOCERA*. Recuperado el 9 de Mayo de 2022, de <https://www.kyoceradocumentsolutions.es/es/smarter-workspaces/businesschallenges/the-cloud/los-6-principales-tipos-sistemas-informacion.html>

Martinez , J. (02 de Marzo de 2020). Obtenido de www.eleconomista.es:
<https://www.eleconomista.es/opinion-blogs/noticias/10390466/03/20/El-proceso-de-toma-de-decisiones-en-el-sector-financiero.html>

Novoseltseva, E. (09 de Marzo de 2021). Obtenido de apiumhub.com:
<https://apiumhub.com/es/tech-blog-barcelona/caracteristicas-principiosarquitectura-de-datos/>

Oracle. (2022). *Oracle*. Recuperado el 27 de MAYO de 2022, de <https://www.oracle.com/mx/data-science/what-is-data-science/>

Oracle. (2022). *Oracle*. Recuperado el 11 de Mayo de 2022, de <https://www.oracle.com/mx/data-science/what-is-data-science/>

power data. (2017). *PowerData*. Recuperado el 09 de mayo de 2022, de <https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/10-tecnicas-de-analisis-de-datos-para-estadisticas-de-big-data>

QuestionPro. (13 de mayo de 2022). *QuestionPro*. Obtenido de <https://www.questionpro.com/blog/es/desempeno-laboral/>

SBS. (2022). Recuperado el 2022, de SBS: <https://www.sbs.gob.pe/inclusion-financiera/Inclusion-Financiera/Evolucion>

Rollins, J. (2015). *Metodología Fundamental para la Ciencia de Datos*. Estados Unidos de América: IBM Corporation. Obtenido de <https://www.ibm.com/downloads/cas/6RZMKDN8>

Rubiños, A. (2020). *Aplicación de Data Science en la empresa Partners Technology S. A. C. lima*. Obtenido de

https://repositorioacademico.upc.edu.pe/bitstream/handle/10757/655788/AngelsR_G.pdf?sequence=3&isAllowed=y

Sala, V. (14 de Enero de 2020). Obtenido de www.cyberclick.es:

<https://www.cyberclick.es/que-es/data-science>

Sanchez, C., & Prada, D. (2018). La co-creación y los nuevos retos de generación de valor que enfrentan las organizaciones. Bogotá, Colombia.

SBS. (2022). Recuperado el 2022, de SBS:

<https://www.sbs.gob.pe/inclusionfinanciera/Inclusion-Financiera/Evolucion>

Simón, H. (2013). *Administrative Behavior. A study of decision-making processes in administrative organization*. New York: The Free Press.

Soto, C. (2017). *Gestión financiera Empresarial*. Machala, Ecuador: Editorial UTMACH. Obtenido de

<http://repositorio.utmachala.edu.ec/bitstream/48000/14354/1/Cap.2%20Fundamentos%20de%20la%20gesti%C3%B3n%20financiera.pdf>

Sysaid. (18 de Mayo de 2022). *Sysaid*. Obtenido de <https://www.sysaid.com/customers>

Amazon Web Services . (2023). *AWS*. Obtenido de <https://aws.amazon.com/es/what-is/datascience/>

Terrazas, R. (2009). *MODELO DE GESTIÓN FINANCIERA PARA UNA ORGANIZACIÓN*. Bolivia. Obtenido de <https://www.redalyc.org/pdf/4259/425942159005.pdf>

ANEXOS

Anexo 1 *Glosario de términos*

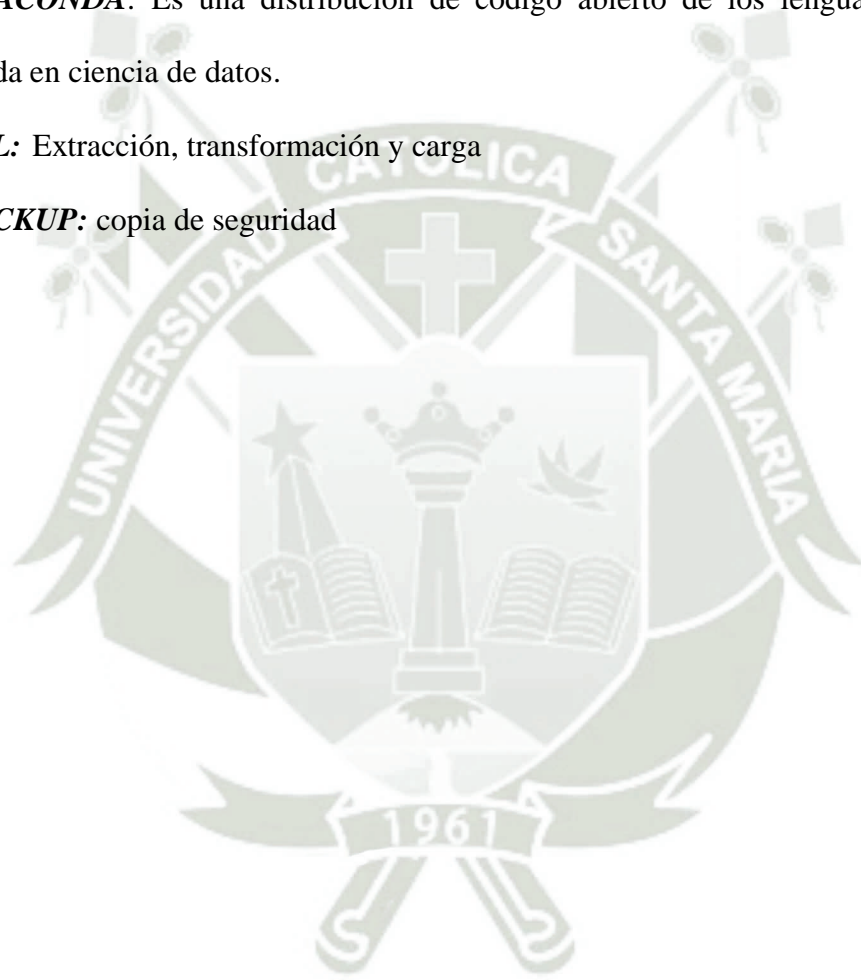
DATA SCIENCE: En español podemos encontrar como ciencia de datos, análisis de gran cantidad de datos para extraer información.

SLA: Service Level Agreement En español acuerdo de nivel de servicio

ANACONDA: Es una distribución de código abierto de los lenguajes Python y R, utilizada en ciencia de datos.

ETL: Extracción, transformación y carga

BACKUP: copia de seguridad

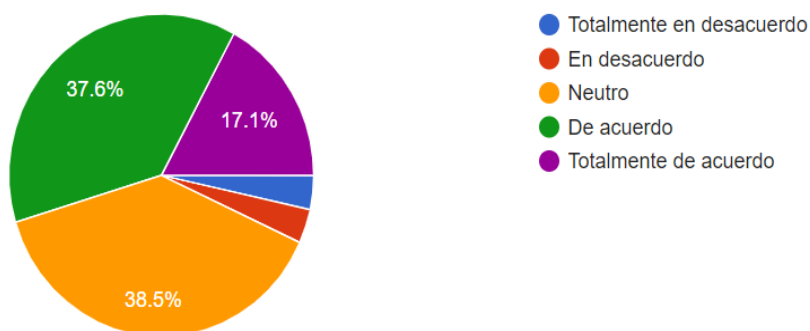


Anexo 2 Encuesta

A continuación, se aplicó el sistema de escalas de usabilidad SUS para poder así evaluar la usabilidad de un nuevo sistema después de usar otro, obtendremos los siguientes resultados de los usuarios sobre el nuevo sistema que consideran que es bueno.

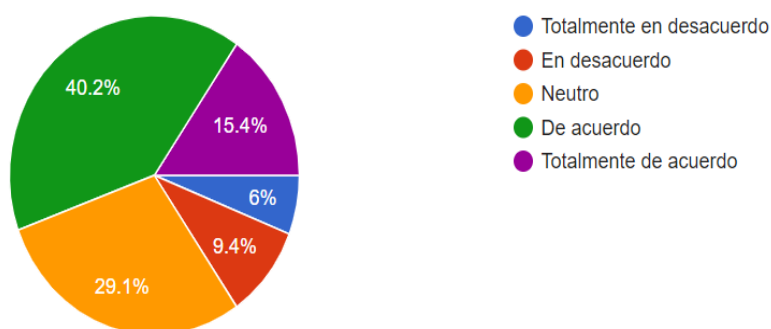
En la siguiente pregunta podemos ver que mas del 50% de usuarios les gusta usar el nuevo sistema confirmando que el sistema es bueno al anterior.

1. Creo que me gustaría utilizar este sistema con frecuencia



En la siguiente pregunta podemos ver que el 55% de usuarios ratifica que el sistema nuevo es facil usar a comparacion del antiguo sistema.

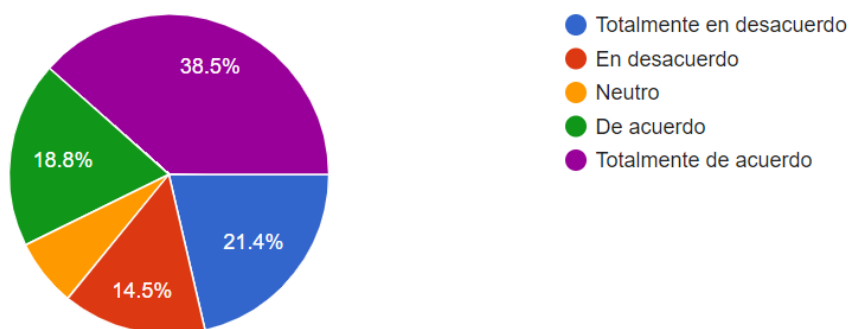
3. Pensé que el sistema nuevo era fácil de usar



Validamos que 57% de usuarios confirma que siempre necesita apoyo para el uso del sistema y donde el 43% confirma puede hacer uso del sistema sin ayuda de un tecnico reafirmando que el sistema es facil de usar para el usuario.

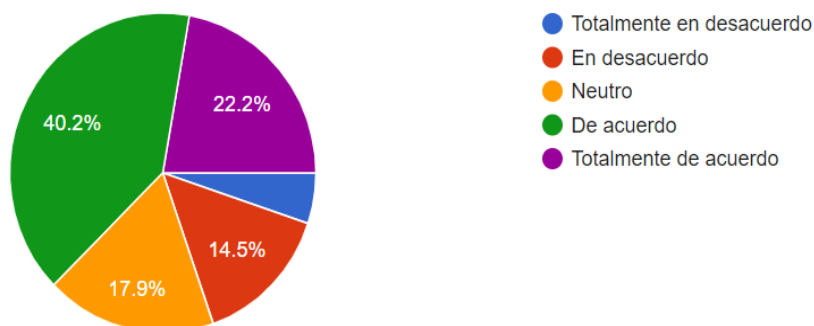
4. Pensé que necesitaría el apoyo de un técnico para poder utilizar el sistema nuevo

Figura 4



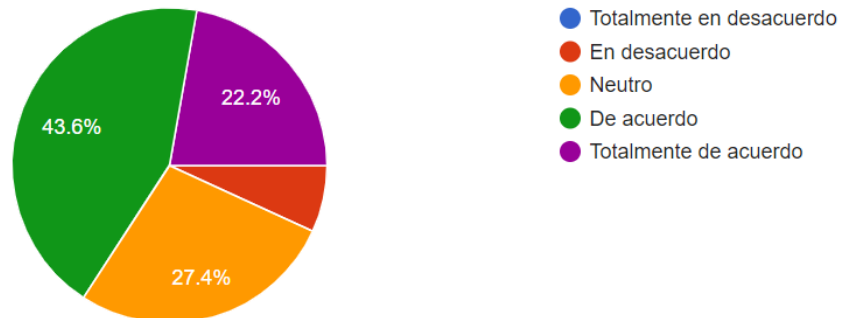
Aseguramos que el 62% de usuarios opina que el nuevo sistema en su software esta mejor integrado en sus funciones a diferencia del anterior aseverando que el cambio a este nuevo sistema tiene mejoras en el procesode los flujos.

5. Encontré que las diversas funciones del sistema nuevo estaban bien integradas a diferencia del anterior sistema



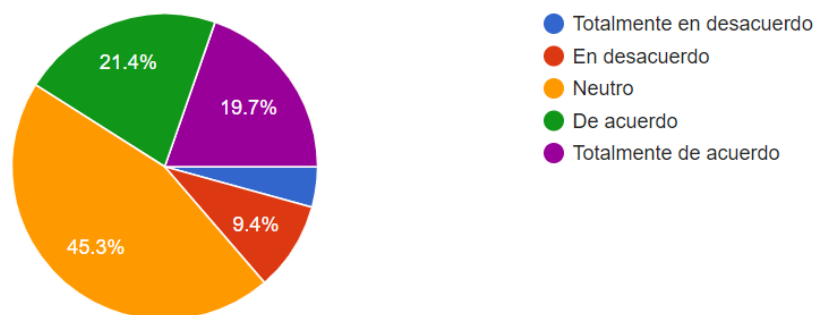
En la siguiente pregunta podemos ver que el 66% de usuarios esta seguros de usar este nuevo sistema ratificando que tiene mucha confianza al usar el sistema sintiendose seguro del mismo.

9. Me sentí muy seguro usando el sistema



En la siguiente pregunta podemos ver que el 41% de usuarios confirman que deben aprender muchas cosas confirmando los nuevos cambios con las nuevas funciones integradas al nuevo software que no estaba completamente entendible en su uso para su flujo.

10. Necesitaba aprender muchas cosas antes de empezar con este sistema nuevo



Anexo 2 Aporte de la investigación

APORTE DE LA INVESTIGACIÓN

En el siguiente proyecto realizado y como participe de este proceso por más de 5 años, se pudo evaluar que los usuarios presentan bajo rendimiento en su capacitación sobre el uso del sistema, donde podemos afirmar que el motivo de los tickets ingresados por los usuarios no está muy alejado por el motivo de error puesto que seguido de ellos está el motivo de proceso.

Analizar la información que almacena la plataforma sysaid sobre todo el campo de motivo de incidente ayudará a identificar si hay un bajo rendimiento y poder gestionarlo en su debido momento en la financiera, el cual se podrá determinar el motivo de las causas del ticket de capacitación o proceso.

También podemos afirmar que la tercera parte de tickets ingresados no cumple con SLA de la empresa lo que ocasiona que algunos tickets vencidos sean atendidos fuera de tiempo. Por otro lado, podemos ver que aplicando ambos modelos confirman la falta de capacitación del personal con ello confirmamos con las encuestas realizadas que afirma que el 70% del personal no se encuentra capacitados lo cual confirma con los resultados obtenidos en ambos modelos.

A continuación, también se realiza la consulta a dos usuarios de la misma área que realizan la atención de incidentes en la financiera, para poder corroborar la afirmación de los modelos aplicado en el siguiente proyecto.

El usuario responde que es correcto la validación que se obtiene, que los tickets ingresados por error están muy cerca al del proceso, puesto que todos los del personal entienden diferente ya sea por una enseñanza de forma distinta o de la misma forma, lo que ocasiona que la enseñanza para un grupo sea diferente y la captación para otro grupo sea tener un aprendizaje más rápido, también se podrá encontrar algún personal con falta de interés de aprender en cada capacitación dada o que falten a sus reuniones de capacitación.

El último usuario afirma que los resultados que muestran los modelos aplicados con la encuesta al personal son correctos ya que reafirma que la mayor parte del personal falta capacitación y esto se debe al poco entusiasmo por aprender, lo que ocasiona que tengan una falta de comprensión con el personal y a su vez pensando que no existe compañerismo o empatía.

Utilizar diferentes métodos de análisis de datos y validar los datos como se está haciendo en este proyecto, es importante para tener un parámetro que nos alerte cuando éste elevado y así se pueda determinar las causas que está generando para poder saber el motivo y así tomar acciones pertinentes en ese momento.

De acuerdo con el análisis realizado al personal, indica generar encuestas al personal cuando un parámetro se encuentre alto, para que le permita determinar así las causas y resolver los inconvenientes y no se genere como una problemática.

Anexo 3 script `datasysaid.ipynb` `sysaid.csv`

```
Import numpy as np import matplotlib.pyplot as plt %matplotlib inline import seaborn as
sns import pandas as pd from sklearn import datasets #ML para splits from
sklearn.model_selection import train_test_split
```

```
#ML para LinearRegression from
sklearn.linear_model import LinearRegression
import pandas from sklearn.preprocessing import
LabelEncoder
pd.read_csv('sysaid.csv')
s1=pd.read_csv('sysaid.csv',encoding='utf-
8') s1.describe() s1.isnull().sum()
plt.hist(s1['Motivo'], label='ticket') #plt.gca().legend(('ticket','motivo'))
x=(s1['Motivo']) y=(s1['Ticket'])
plt.xlabel('Motivo')
plt.ylabel('Ticket')
plt.show()
df2=s1[['Año', 'Motivo','Cumplimiento SLA']]
df2.groupby(df2['Año']).size()
df2.groupby(df2['Motivo']).size()
df2.groupby(df2['Cumplimiento SLA']).size()
s1['Año'].groupby(s1['Motivo']).mean()
a=s1['Año'].groupby(s1['Motivo']).mean()
s1.groupby(s1['Motivo']).mean()
df2.groupby(df2['Año']).size().plot(kind='bar')
plt.xlabel('Año') plt.ylabel('Ticket') plt.show()
df2.groupby(df2['Motivo']).size().plot(kind='bar') plt.xlabel('Motivo')
plt.ylabel('Ticket') plt.show()
df2.groupby(df2['Cumplimiento SLA']).size().plot(kind='bar') plt.xlabel('Sla')
plt.ylabel('Ticket') plt.show()
```

```
m=l.fit_transform(s1['Motivo'])  
G1=s1['Año']  
G2=s1['Motivo']  
X=np.array(G1).reshape (-1,1)  
Y=np.array(m).reshape (-1,1)  
X_train, X_test, y_train, y_test = train_test_split(X, Y,  
test_size=0.2, random_state=80) modelo = LinearRegression()  
modelo.fit(X_train,y_train) modelo.coef_ modelo.intercept_
```

