



Código	FPI-002
Objeto	Protocolo de presentación de proyectos de investigación SIGEVA UNLaM
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	4
Vigencia	12/11/2021

Unidad Ejecutora:
Departamento de Ingeniería e Investigaciones Tecnológicas

Programa de acreditación:
CyTMA2

Título del Programa de Investigación¹:
Director del Programa:

Título del proyecto de investigación:
Automatización de servicios de telecomunicaciones con Machine Learning

PIDC:

Elija un elemento.

PII

Elija un elemento.

Director del proyecto:
Caiafa, Marcelo Dante

Co-Director del proyecto:
Aurelio, Ariel Rodrigo

Integrantes del equipo:
Bevilacqua, Alejandro
Baner, Deborah Yanina (alumna)

Fecha de inicio:

01/01/2022

Fecha de finalización:

31/12/2023

¹ Completar solo en caso de que el presente proyecto se encuadre en el marco de un Programa de Investigación

1-Cuadro resumen de horas semanales dedicadas al proyecto por parte de director e integrantes del equipo de investigación:²

Rol del integrante	Nombre y Apellido	Cantidad de horas semanales dedicadas al proyecto
Director	Caiafa, Marcelo Dante	4hs
Co-director	Aurelio, Ariel Rodrigo	4hs
Director de Programa		
Docente-investigador UNLaM	Bevilacqua, Alejandro	4hs
Investigador externo ³		
Asesor-Especialista externo ⁴		
Graduado de la UNLaM ⁵		
Estudiante de carreras de posgrado (UNLaM) ⁶		
Alumno de carreras de grado (UNLaM) ⁷	Baner, Deborah Yanina	4hs
Personal de apoyo técnico administrativo		

2-Plan de investigación

2. Tipo de actividad I+D: Aplicada

2.1. Resumen del Proyecto:

La transformación digital viene generando en los últimas décadas profundos cambios en la forma de producir y comercializar bienes y servicios. En este contexto, los procesos de digitalización suponen modificaciones en la cultura y en los modelos de negocio de las organizaciones. Con esto surgen nuevos canales, productos y servicios, enmarcados en estrategias digitales de comercialización. Como motor de esta transformación se destacan las tecnologías de Big Data para el almacenamiento de datos masivos provenientes de múltiples fuentes, y Machine Learning que habilitan a los sistemas de información a la interpretación de la información en grandes cantidades con diversos fines aplicables a cualquier empresa, por ejemplo: mejorar la experiencia del cliente, incrementar la eficiencia en procesos productivos, encontrar nuevas oportunidades de negocio, etc. La generación de información a partir del gran volumen de datos representa un gran valor para las organizaciones. Su capacidad para generar información relevante a partir de la explotación de datos en forma masiva es posible mediante la utilización de herramientas de Inteligencia artificial. El aprendizaje automático es un método de análisis de datos que automatiza la construcción de modelos analíticos. Es una rama de la inteligencia artificial basada en la idea de que los sistemas pueden aprender de los datos, identificar patrones y tomar decisiones con la mínima intervención humana. El objetivo del trabajo es utilizar una plataforma de Machine Learning de código abierto para automatizar procesos directamente relacionados con servicios soportados por sistemas de comunicaciones.

Este trabajo pretende aprovechar la disponibilidad de la gran cantidad de datos existente en distintas plataformas de telecomunicaciones. Su adecuado procesamiento permite generar

² Incluir todos los integrantes del equipo de investigación, agregando tantas filas para cada rol de integrante del equipo de investigación como sea necesario.

³ Deberá adjuntar FPI 28, 29 y 30 debidamente firmados.

⁴ Idem nota 2.

⁵ Idem nota 2

⁶ Adjuntar certificado de materias aprobadas de estudiantes de carrera de posgrado.

⁷ Adjuntar certificado de materias aprobadas de estudiantes de carrera de grado.

información relevante para mejorar la eficiencia y el riesgo operativo de la infraestructura de comunicaciones.

Se trata de la continuación de un trabajo de investigación anterior que permitió la generación de un tablero de comando a partir de la generación de una serie de indicadores. En este caso ahora se trata de otorgar capacidades de predicción y toma de decisiones programadas a partir de los indicadores de las variables consignadas en los tableros.

Mediante la utilización de herramientas de aprendizaje automático se pretende la automatización de procesos vinculados a la tecnología de la información. Se pretende determinar patrones de tráfico en forma autónoma para detectar comportamientos anómalos, monitoreo de la calidad de los servicios de comunicaciones que posibilitan su restablecimiento sin intervención de un operador humano.

El estudio pretende aportar valor al proceso de toma de decisiones y gestión de activos como infraestructura y privacidad de datos para mejorar la eficiencia de distintos procesos operativos en las tecnologías de la información.

2.2. Palabras clave: Tecnologías de la información, aprendizaje automático, transformación digital.

2.3 Resumen del Proyecto (inglés):

The digital transformation has been generating deep changes in the way of producing and marketing goods and services in recent decades. In this context, digitization processes entail changes in the culture and business models of organizations. As the engine of this transformation, Big Data technologies stand out for the storage of massive data from multiple sources, and Machine Learning enable information systems to interpret information in large quantities for many purposes applicable to any company, for example: improve the customer experience, increase efficiency in production processes, find new business opportunities, etc.

The generation of information from the large volume of data represents great value for organizations. Its ability to generate relevant information from the exploitation of data in a massive way is possible through the use of artificial intelligence tools. Machine learning is a data analysis method that automates the building of analytical models. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention. The objective of the work is to use an open source Machine Learning platform to automate processes directly related to services supported by communication systems.

This work aims to take advantage of the availability of the large amount of data existing in different telecommunications platforms. Its proper processing allows the generation of relevant information to improve the efficiency and operational risk of the communications infrastructure.

This proposal is the continuation of a previous research work that allowed the generation of a dashboard from the generation of a series of indicators. In this case, it is now a matter of granting programmed prediction and decision-making capabilities based on variables' indicators recorded on the boards.

Through the use of automatic learning tools, the automation of processes linked to information technology is intended. It is intended to determine traffic patterns autonomously to detect abnormal behavior, monitoring the quality of communications services that enable their restoration without intervention of a human operator.

The study aims to add value to the decision-making process and management of assets such as infrastructure and data privacy to improve the efficiency of different operating processes in information technology.

2.4 Palabras clave (inglés): machine learning, digital transformation, ICT.

2.5 Disciplina desagregada: 2108, Ingeniería de las comunicaciones

2.6 Campo de aplicación: 651, Comunicaciones- Telecomunicaciones

2.7 Especialidad: Aprendizaje automático

2.8 Estado actual del conocimiento:

Los cambios que la transformación digital viene generando en las últimas décadas, han producido un cambio de paradigma en la forma de producir y comercializar bienes y servicios. Internet está

transformando los hábitos de consumo, de compra, de entretenimiento, de comunicación y relacionamiento entre las comunidades.

La convergencia y los avances en tecnologías como el Big Data, el Cloud Computing o el Internet de las Cosas (IoT) son los motores de cambio que están siendo incorporados para dar respuesta a estas nuevas necesidades de los clientes, y que a su vez están impulsando el crecimiento y la innovación en la Inteligencia Artificial (IA en adelante). Estas tecnologías representan una oportunidad para la banca tradicional privada ya que permitirían, tanto gestionar la gran cantidad de información a la que se tiene acceso hoy en día tomando decisiones estratégicas en las compañías a partir de los datos,

Hasta hace unos pocos años la inteligencia artificial, con 70 años de historia, comprendía un terreno del mundo tecnológico que no había alcanzado las expectativas esperadas en su nacimiento. Pero hoy existen diversos factores, con origen fundamentalmente en la tecnología, que impulsan su uso: el incremento sin precedentes del volumen y tipología de datos disponibles, la conectividad y el acceso al dato, la mejora de los algoritmos utilizados y el aumento de la capacidad computacional de los sistemas. En las últimas 2 décadas la disponibilidad de datos digitales se ha incrementado casi 1000 veces, mientras que su relación con la algoritmia, se han registrado mejoras de 10 veces en su eficiencia, al tiempo que la velocidad de computación ha aumentado 100 veces su desempeño (Brynjolfsson & McAfee, 2017).

Estamos viviendo la revolución de los datos. Las compañías recolectan datos extremadamente detallados y proveen de información y conocimiento a sus clientes, proveedores, aliados e incluso competidores (Brynjolfsson, Hitt, & Kim, 2011) y se evidencia mayor dependencia de este tipo de extracción de conocimiento (Oussous, Benjelloun, Ait Lahcen, & Belfkih, 2018). Esta revolución se materializa en el uso de tecnologías de información en el ámbito corporativo, donde sistemas de gestión de recursos (ERP), suministros (SCM) y clientes (CRM), han probado tener efecto positivo directo las operaciones (McAfee, 2002) y productividad (Aral, Brynjolfsson, & Wu, 2006).

Machine Learning es una herramienta actual que ayuda a las empresas a impulsarse en la transformación digital ya que las deja en una mejor situación que garantiza su continuidad y rentabilidad a largo plazo. En el centro de esta transformación se encuentra la capacidad de utilizar datos para proporcionar información útil y predicciones mediante el aprendizaje automático y la inteligencia artificial.

El Machine Learning es una rama de la inteligencia artificial que, mediante algoritmos, conjuntos de instrucciones definidas, ordenadas y acotadas para resolver un problema o realizar una tarea, provee a las computadoras de capacidad de identificar modelos de datos masivos y realizar un análisis predictivo. Por lo que las computadoras realizan una tarea específica automáticamente sin programarse.

Por otro lado tenemos la oportunidad de mejora en la eficiencia de los procesos, en la productividad y en la gestión del riesgo operativo a partir de la incorporación de herramientas de analítica predictiva. Según los autores, el impacto económico de la IA en las empresas sería impulsado por 3 pilares: las ganancias en la productividad debido a la automatización, la mayor demanda de los consumidores y la difusión de la innovación. La primera promesa consiste en obtener ganancias de productividad masivas para las empresas que automatizan sus procesos. La IA puede aportar automatización inteligente, que se diferencia de manera crucial de todas las formas anteriores de automatización debido a que las máquinas de IA pueden aprender. (Katz, 2017)

En el evento Mobile World Congress del 2019 el CEO de Telefónica, José María Álvarez-Pallete, y el CEO de Microsoft, Satya Nadella presentaron en una de las alianzas más comentadas del primer día del evento. No es porque la alianza entre ambas compañías sea nueva —Microsoft fue una de las primeras en colaborar con Telefónica en su cuarta plataforma cognitiva— sino por el hecho de que los dos CEOs se sentaran juntos en una conferencia de prensa —todo un síntoma de época—. El acuerdo anunciado viene a profundizar en el trabajo ya realizado por ambas compañías sobre Aura al permitir combinar las redes de Telefónica y la plataforma Azure AI para diseñar los servicios que las empresas de telecomunicaciones ofrecerán en el futuro. Como decíamos, no es un tema nuevo para Telefónica, que ya ha mostrado cómo está intentando empujar la inteligencia artificial dentro de su ecosistema, apalancado en acuerdos con socios.

En concreto, Microsoft y Telefónica buscarán extender el uso de Azure Cognitive Services tanto en la relación con los clientes —en dónde ya están trabajando desde Aura— como en los procesos de

negocio, creando nuevas experiencias en el hogar impulsadas por la inteligencia artificial, además de explorar el uso de tecnologías inteligentes para transformar las redes. Las compañías crearán un grupo de trabajo para estudiar la aplicación de IA y Machine Learning en optimización de redes, reducción de costos y transformación del sector. Es decir, según lo que podemos entender de este anuncio, Microsoft y Telefónica se proponen utilizar IA en cualquier espacio dónde pueda ser útil para mejorar procesos de la organización.

No son las únicas compañías que están trabajando en cómo IA y las tecnologías de aprendizaje automático pueden ayudar a las empresas de telecomunicaciones a mejorar procesos y generar nuevos negocios. Este año, MWC fue sede de una buena cantidad de lanzamientos de productos que cubren casi todos los ámbitos del negocio de un operador de telecomunicaciones. Nokia anunció que, junto con China Telecommunications Corporation innovaciones en machine learning para la optimización de 5G mMIMO. Se trata de los primeros resultados de la cooperación firmada entre ambas compañías para desarrollar aplicaciones de aprendizaje automático. Nokia desarrolló el algoritmo base y trabajó junto con CTC para analizar el potencial de mejora de performance en las futuras redes 5G y crear una primera demostración y prueba de concepto.

En concreto, se creó un programador basado en machine learning que permite mejorar la eficiencia espectral y la experiencia de usuario. Además, se desarrolló un algoritmo de AI/ML usando la huella de RF para mejorar la precisión en locaciones indoor en redes celulares.

Del lado de los dispositivos, Qualcomm Technologies anunció la cuarta versión de su motor de IA que permite experiencias de voz, cámara y hasta realidad aumentada en la mayoría de los smartphones que utilizan plataforma Snapdragon 855, de la marca Sony y Xiaomi.

Por otro lado, Calix añadió capacidades de aprendizaje automático a su producto de marketing en la nube que, asegura, podría ayudar a los operadores de servicios de telecomunicaciones a reducir la tasa de abandono de clientes. Este algoritmo analiza cerca de 30 atributos de comportamiento y experiencia de servicio de los usuarios, permite segmentar a clientes de alto riesgo y entregar esa información en tablas diseñadas especialmente. (Pautasio, 2019)

2.9. Problemática a investigar:

El presente proyecto es una continuación del trabajo anterior donde se ejecutó un tablero de comando con Big Data a partir de datos ingresados de las centrales telefónicas de comunicaciones para un monitoreo de infraestructura y calidad de servicios. Los datos se obtuvieron a partir de los CDR (call detail records) y la tecnología utilizada fue una herramienta de Big Data. Particularmente la plataforma utilizada para su tratamiento fue ELK (Elastic-Logstash-Kibana). Una de las finalidades fue disponer de un tablero de comando que permita la visualización de la información, generada a partir del gran volumen de datos, para la ejecución del mantenimiento correctivo.

Este presente proyecto pretende ser la continuación del anterior. El objetivo es agregar funcionalidades de aprendizaje automático para que el sistema logre el desempeño de mantenimiento preventivo. Es decir que se pretende automatizar los diferentes procesos involucrados con tecnología Machine Learning para lograr reducir los tiempos de respuesta y los riesgos operativos en la gestión de incidentes.

La capacidad de aprendizaje automático, conocida como Machine Learning (ML), es un área de la Inteligencia Artificial (IA) cuyas técnicas aprenden a partir de datos. Big Data y ML se complementan mutuamente. Big Data proporciona grandes volúmenes de datos; las técnicas de ML son capaces de aprender a partir de los datos.

Estos datos pueden ser utilizados por organizaciones pequeñas, medianas o grandes. Las empresas proveedoras de los servicios de comunicaciones los utilizan en general para los procesos de carga, liquidación, facturación, eficiencia de la red, detección de fraude, aseguramiento de ingresos, detección de abandono, servicios de valor agregado, inteligencia de negocios, etc. Adicionalmente pueden ayudar a mejorar servicios y procesos existentes en distintas áreas como marketing, auditorías y resolución técnica de incidentes. (Agrawal, 2011)

Basándonos en las técnicas de ML podemos diseñar y entrenar modelos descriptivos, predictivos y prescriptivos. Los primeros nos permiten obtener una explicación a fenómenos observados y activar alarmas de manera temprana. Los segundos nos permiten anticipar situaciones predefinidas

o predecir valores de ciertos. Por último, los modelos prescriptivos evalúan la relación riesgo-beneficio asociada a cada una de las posibles acciones a realizar.

- ¿Puede Machine Learning mejorar la precisión de resultados a la par de reducir los tiempos de resolución de incidentes en los servicios de IT?
- ¿Se encuentra en riesgo, a corto y mediano plazo, la continuidad laboral de especialistas que ejecutan tareas de soporte y administración en los servicios de IT a partir de la implementación de Machine Learning?
- ¿Qué características tiene la industria que impulsado soluciones de Machine Learning en monitoreo y gestión de servicios IT?

2.10. Objetivos:⁸

Las organizaciones tienen muchos datos de los clientes: antigüedad, planes contratados, consumo diario, llamadas mensuales al servicio de atención al cliente, últimos cambios de planes contratados... pero seguramente los usa solo para facturar y para hacer estadísticas. ¿Qué más puede hacer con esos datos? Se pueden usar para predecir cuándo un cliente se va a dar de baja y gestionar la mejor acción que lo evite. En pocas palabras, con Machine Learning se puede pasar de ser reactivos a ser proactivos. Los datos históricos del conjunto de los clientes, debidamente organizados y tratados en bloque, generan una base de datos que se puede explotar para predecir futuros comportamientos, favorecer aquellos que mejoran los objetivos de negocio y evitar aquellos que son perjudiciales.

Esa cantidad ingente de datos son imposibles de analizar por una persona para sacar conclusiones y menos todavía para hacer predicciones. Los algoritmos en cambio sí pueden detectar patrones de comportamiento contando con las variables que le proporcionamos y descubrir cuáles son las que han llevado, en este caso, a darse de baja como cliente

Tradicionalmente la operación y administración de las distintas plataformas tecnológicas que proveen la infraestructura necesaria a los distintos servicios de comunicaciones se gestionaba de forma completamente aislada. Respondía a una metodología que se conoce generalmente como el tratamiento en silos. Es decir que, cada uno de los proveedores ofrece sus herramientas propietarias para la gestión de dichos recursos, esto resulta con una baja capacidad de integración. Actualmente existe un nivel de integración entre las plataformas de comunicaciones que no detiene su crecimiento. En este contexto es posible disponer de herramientas de gestión que sean capaces de consolidar la información en un único tablero de comando que ofrezca información relevante.

Objetivo General

Incorporar funcionalidades de aprendizaje automático para la automatización de procesos de servicio de telecomunicaciones. Se pretende automatizar diferentes procesos mediante lam utilización de herramientas de Machine Learning para reducir tiempo de respuesta y riesgos operativos en la gestión de incidentes en servicios de tecnología.

Objetivos Específicos

- Construcción de perfiles de tráfico. Utilizando herramientas de aprendizaje automático para conocer el comportamiento regular o estándar del sistema. Es decir que, a partir de ese comportamiento esperable, se pueda detectar automáticamente cualquier tráfico anómalo que permita de manera inmediata ejecutar procesos para minimizar ilícitos. Mediante la identificación de los **factores de ocupación** OTR (Outcoming Traffic Rate) de enlaces de conectividad se automatiza el monitoreo de los flujos de tráfico.

⁸ Detallar objetivo general y objetivos específicos.

- Mejorar la disponibilidad de la infraestructura tecnológica. A partir de la medición de los valores relativos de ocupación de enlaces e indicadores de disponibilidad del servicio, **niveles de tráfico entrante/saliente de servicio**, como el AHT (Average Holding Time). Se busca que la toma de decisiones sea autónoma para reducir la velocidad de respuesta, mejorando así la continuidad de servicios de comunicaciones ante eventuales incidentes.
- Automatización del proceso de monitoreo de **calidad de servicio**. Como resultante de la representación de los parámetros de QM (quality monitoring) se pretende que la herramienta ejecute automáticamente procesos de medición de calidad de servicio de comunicaciones que actualmente se ejecutan con la intervención de un operador humano.

2.11. Marco teórico:

El Machine Learning, traducido al español como aprendizaje automático, es una rama de la inteligencia artificial que permite que las máquinas aprendan sin ser expresamente programadas para ello. Es el campo de investigación que enfoca en el estudio de la teoría, performance, y propiedades de los algoritmos y sistemas de aprendizaje automatizado (J. Qiu, Wu, Ding, Xu, & Feng, 2016). Su objetivo es el descubrimiento de información que pueda ser transformada en conocimiento y permita mayor inteligencia en la toma de decisiones (Oussous et al., 2018).

Un estudio de la consultora internacional McKinsey que examinó diferentes aspectos relacionados con la IA, analizó la distribución de la inversión total, incluyendo tanto la inversión de grandes corporaciones como la financiación de capital de riesgo y fondos de capital privado. Encuestaron a más de 3.000 ejecutivos de alto nivel, procedentes de 10 países y 14 sectores diferentes, acerca del uso de tecnologías de IA en sus empresas, las perspectivas de un mayor despliegue de estas tecnologías en dichas empresas y el impacto de la IA en los mercados, los gobiernos y los individuos. Para eso clasificaron la utilización de las herramientas de IA en cinco áreas: robótica y vehículos autónomos, visión computarizada, Lenguaje natural, agentes virtuales y aprendizaje automático (machine learning). De acuerdo con sus resultados: liderada, especialmente, por gigantes tecnológicos como Google o Baidu. Estima que la inversión global de estos gigantes tecnológicos en 2016 fue de entre 20.000 y 30.000 millones de dólares, de los cuales el 90% se dedicaron a I+D y despliegue y el 10% restante a adquisiciones. La financiación de capital privado y capital riesgo ha alcanzado un total de entre 6.000 y 9.000 millones de dólares. El aprendizaje automático es la tecnología que recibe el mayor porcentaje de ese dinero, tanto en inversiones externas como internas. Seguida de la robótica y el reconocimiento de voz. (McKinsey, 2017)

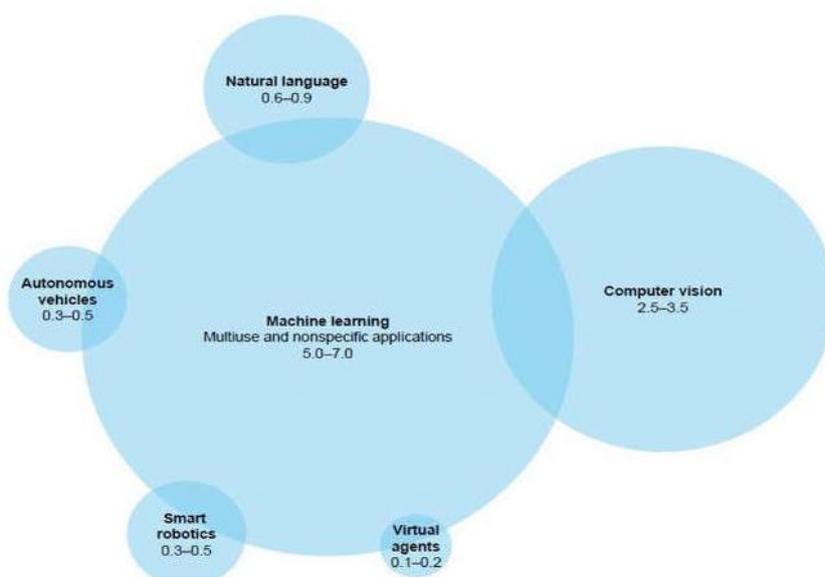


Gráfico 1: Distribución de la inversión en IA a partir de sus diferentes tecnologías.
Fuente: *External investment in AI focused companies by technology category, McKinsey*

ML es un tipo de IA que permite a las máquinas aprender en forma directa de ejemplos y de experiencia, la cual es adquirida a partir de un conjunto de datos (conocido como dataset), a través de un entrenamiento. La programación tradicional está basada en reglas estáticas, las cuales establecen cómo resolver un problema, paso a paso. Por lo contrario, con ML se dispone una gran cantidad de datos para utilizar como ejemplo de cómo la tarea puede ser realizada, o para detectar patrones (The Royal Society, 2017).

De acuerdo a algunos autores (Broussard, 2018) los tipos de aprendizaje se clasifican en tres subdominios en función de su metodología: aprendizaje supervisado, no supervisado y por refuerzo.

El aprendizaje supervisado corresponde al entrenamiento por medio de datos previamente etiquetados y resultados esperados, el objetivo es que las máquinas aprendan a hacer el trabajo ya hecho. En contraste con el no supervisado, se proveen grandes cantidades de datos de entrada que conforman características de un mismo objeto y la finalidad es que aprendan a determinar qué es a partir de la información recopilada.

El tercer subdominio es el de los algoritmos por refuerzo, donde la máquina aprende en base a prueba y error, identificando los patrones de éxito y perfeccionando el método de detección (J. Qiu et al., 2016).

El caso del **aprendizaje por refuerzo** se produce cuando una máquina aprende por medio de prueba y error hasta alcanzar la mejor manera de completar una tarea dada. Por ejemplo, el sistema aprende a modificar su conducta en base de “recompensas” para que resuelva la tarea asignada, sin programarlo específicamente ni para que lo realice de una forma determinada.

El **aprendizaje supervisado** se produce cuando se entrena a las máquinas con datos etiquetados. Por ejemplo, fotos con descripciones de los elementos que aparecen en ellas. El algoritmo que utiliza la máquina es capaz de seleccionar esas etiquetas en otras bases de datos. Así, la máquina puede identificar imágenes similares.

Finalmente, en el caso del **aprendizaje no supervisado**, las máquinas no identifican patrones en bases de datos etiquetadas, sino que buscan similitudes. En este caso, los algoritmos no están programados para detectar un tipo específico de datos, sino que buscan ejemplos que se parezcan y puedan agrupar. Es lo que ocurre, por ejemplo, en el caso del reconocimiento facial, en el que el algoritmo no busca unos rasgos concretos, sino una serie de patrones comunes que indicarían que se trata del mismo rostro.

Según la consultora internacional Gartner, en su informe de 2018, denominado Ciclo de expectativas para tecnologías emergentes, conocido en inglés como Hype Cycle for Emerging Technologies machine learning se encuentra en un momento de grandes expectativas, esperando que llegue a una etapa de madurez y adopción masiva en un lapso que va entre dos y cinco años. (Gartner 2018)

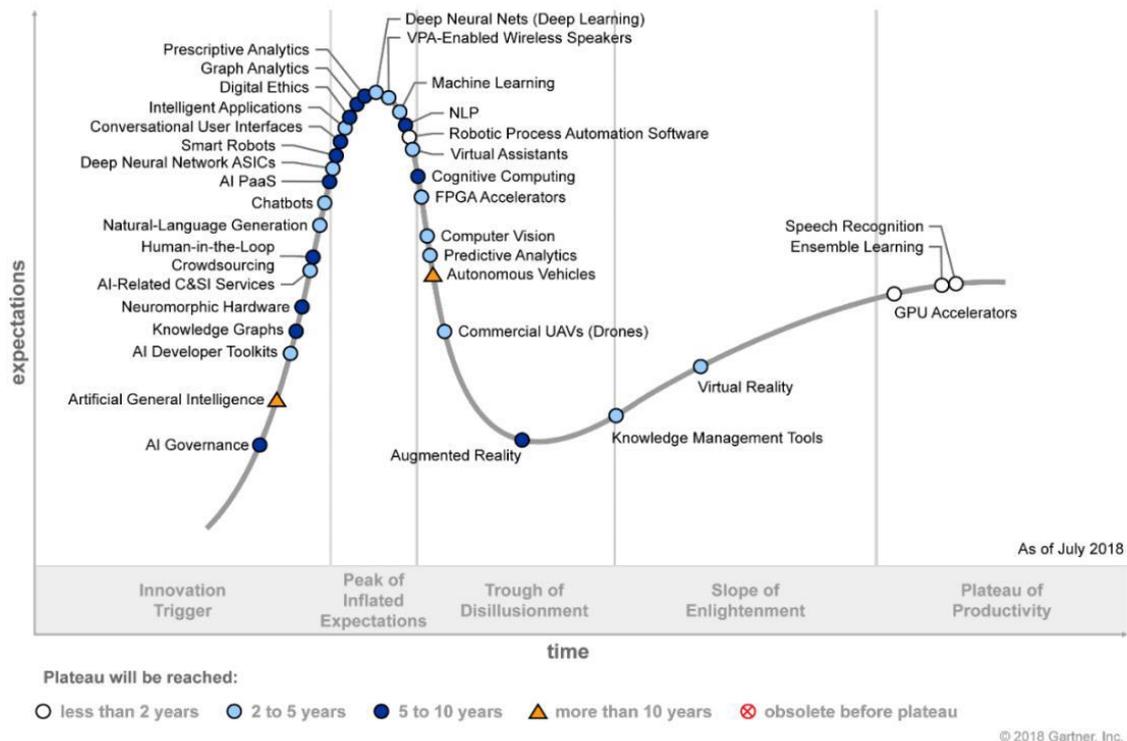


Gráfico 2: Ciclo de evolución de expectativas de las tecnologías emergentes
Fuente: *The Hype Cycle of Emerging Technologies*, Gartner.

Un factor que puede ayudar a entender la importancia y crecimiento que está teniendo esta tecnología más allá de la adopción, es la inversión y el gasto que las empresas incurren en esta tecnología. En el primer caso, se identifica un crecimiento importante y sobre todo el foco de esa inversión se encuentra en aprendizaje automático (McKinsey, 2017a). También los analistas de la industria de TI coinciden en que el tamaño del mercado para la tecnología de IA experimentará un fuerte crecimiento en los próximos tres años.

La IA puede contribuir a la economía mundial, para 2030, con \$15,7 trillones de dólares. Un número que supera el aporte a la economía actual de China e India. Dicho número puede descomponerse en \$6,6 trillones provenientes de un crecimiento en la productividad y \$9,1 trillones provenientes del consumo. Dentro de dicho aporte, China y Norte América serían los máximos beneficiados, haciéndose de un casi 70% del crecimiento (Gillham, 2018).

Región	Impacto económico de la IA	% del Total
China	\$7.0 trillion	44.6%
América del Norte	\$3.7 trillion	23.6%
Europa del Norte	\$1.8 trillion	11.5%
Asia desarrollada	\$0.9 trillion	5.7%
Europa del Sur	\$0.7 trillion	4.5%
América Latina	\$0.5 trillion	3.2%
Resto del mundo	\$1.2 trillion	7.6%
Total	\$15.7 trillion	100.0%

Tabla 1: Contribución estimada de la IA a la macroeconomía mundial para 2030.
Fuente: *The macroeconomic impact of the artificial intelligence* (Gillham, 2018).

Para caracterizar la revolución que IA significa, el fondo de inversión Woodside Capital Partners detalla, en su informe “*The Top Acquirers in Artificial Intelligence*” publicado en el 2019, una serie de conceptos que ayudan a comprender el fenómeno.

Ya para el 2018, Google poseía miles de proyectos basados en IA, en desarrollo. Esto se suma a que también es la compañía perteneciente a las “4 Grandes” (Google, Facebook, Amazon y Microsoft) más activa en lo que respecta a adquisiciones e inversiones de aplicaciones y talentos vinculados a la IA, siendo la compañía que más inversión está realizando. Desde el 2015, la misma ha pasado de ser “*mobile first*” a “*AI-first*”, invirtiendo anualmente más de \$10 billones de dólares en infraestructura para la creación de negocios basados en IA. Con el objetivo de construir el mejor laboratorio de IA del mundo, Facebook contrata en 2013 a Yann LeCun, uno de los pioneros en el diseño y uso de RNC, como director del área de Investigación en IA (FAIR por su nombre en inglés, *Facebook Artificial Intelligence Research*). Como ya hemos mencionado, las oportunidades en IA crecen mientras se tenga grandes cantidades de datos a analizar. Alineado a esto, Facebook cuenta, al año 2019, con 2,2 billones de usuarios mensuales activos, de los cuales apenas le está sacando el provecho que tamaña muestra puede generar. Actualmente su ARPU (promedio de ganancia por usuario) mundial es de \$5,97 dólares. Por otro lado, el ARPU en Estados Unidos y Canadá es de \$25,91 dólares. En ambos casos hay grandes oportunidades para hacer crecer dichas cifras (Porter, 2019).

A través de la IA, Facebook tiene la oportunidad de potenciar sus productos insignia (Facebook, Instagram, Whatsapp, entre otros) y expandirse a otras verticales, como ser salud, *retail* o servicios para empresas. Prácticamente todo lo que el usuario puede visualizar en Facebook está atravesado por IA y ML. Algunas de las tecnologías comúnmente utilizadas hoy son las tecnologías del lenguaje (traducciones, reconocimiento de voz, comprensión de lenguaje natural), filtro de contenidos (reducción de artículos no deseados, noticias falsas, discursos con incitación al odio), visión (su algoritmo de reconocimiento facial posee una precisión del 97% (contra una precisión del 85%, que posee el FBI)) y publicidad (dónde y cuándo mostrarla, maximizando la cantidad de clicks) (Porter, 2019).

En épocas de constantes cambios en el comportamiento de los consumidores, para mantener una organización competitiva la clave será tener la capacidad de entender esos comportamientos y reaccionar rápido.

Las principales tendencias en 2021 fue la utilización de inteligencia artificial y machine learning. En esa línea los autores destacan dos impactos en particular:

- Uso de machine learning para predecir la demanda: Cuando se usan correctamente las variables de contexto y los datos históricos por ejemplo las ventas de un producto, se pueden generar predicciones con una precisión mucho más alta que con los métodos tradicionales, capturando características de los negocios como la estacionalidad o correlaciones no intuitivas. Esto, entre otras cosas, impacta directamente en la gestión más eficiente del stock evitando faltante de productos y la reducción de costos optimizando la capacidad de atención.

- Los perfiles de analistas de todas las organizaciones deberán convertirse en analistas de datos: Las empresas o instituciones que busquen tomar decisiones basadas en datos, no podrán depender de una sola área centralizada de datos, eso será insuficiente por el volumen de trabajo que tendrán estos equipos.

Se propone que las áreas expertas en datos se encarguen de diseñar la estrategia de incorporación de IA/ML y los algoritmos más complejos, pero el resto de la compañía deberá incorporar las habilidades de datos necesarias para impulsar estas iniciativas. Habilidades como estadística básica, creación de tableros visuales, contar historias con datos y hacerle consultas a los datos mediante lenguajes como SQL. (Vivas, 2020)

2.12. Hipótesis de trabajo o los supuestos implícitos (según corresponda al diseño metodológico) :⁹
Como se mencionó anteriormente, el presente trabajo pretende dar continuidad a partir de los resultados de un trabajo de investigación precedente. En el trabajo anterior de logró la elaboración

⁹ En proyectos de desarrollo tecnológico puede ser reemplazada una hipótesis de trabajo por la propuesta de solución al problema de investigación mediante el diseño de un prototipo o elemento equivalente.

de tableros mediante herramienta de Big Data que procesaron grandes volúmenes de datos. La herramienta de código abierto utilizada fue ELK (Elastic-Logstash-Kibana) y se utilizaron los datos provenientes de los CDR (call detail records) de las centrales de comunicaciones de distintas plataformas tecnológicas. Esto permitió disponer de un tablero de tráfico y métricas de calidad de servicio e infraestructura. Con esto se implementó la ingesta de datos en módulo Logstash, para su adecuada indexación procesamiento con el motor de búsqueda de Elastic y una posterior visualización y representación de resultados a partir de los filtros del módulo de Kibana.

Se pretende en este caso avanzar a partir de ese trabajo y dotarlo al sistema de funcionalidades de aprendizaje automático mediante el módulo de Machine Learning.

Este trabajo pretende ampliar el alcance del anterior en cuanto a funcionalidades y plataformas tecnológicas, siempre dentro del encuadre dentro de servicios de telecomunicaciones.

Se pretende comprobar la efectividad de la herramienta y su alcance en estos procesos particulares de análisis de tráfico, monitoreo de calidad y disponibilidad de infraestructura, como se detalló en los objetivos.

En cuanto a los participantes se pretende identificar las habilidades técnicas específicas y competencias genéricas transversales que requieren dominar y son plausibles de desarrollar en un proyecto de estas características.

2.13. Metodología:

Este trabajo se basa en una metodología del tipo sintética, es decir que a partir de varios elementos por separado se pretende llegar a un resultado concreto. (Ruiz Limon, 2006)

El método sintético es un proceso de razonamiento que tiende a reconstruir un todo, a partir de los elementos distinguidos por el análisis; se trata de reconstruir el todo en sus aspectos y relaciones esenciales para una mayor comprensión de los elementos constituyentes. Cuando se dice que va de lo abstracto a lo concreto significa que los elementos aislados se reúnen y se obtiene un todo concreto real. (Mialaret, 2001).

Como primera instancia se trabajará en la identificación de procesos susceptibles de automatización. Se preparará la ingesta de datos a partir de los archivos de gran volumen de información para su procesamiento con tecnología de Big Data, a partir del trabajo anterior. Luego se conformarán la representación de indicadores y sobre eso se aplicará la herramienta de aprendizaje automático para que los procesos que requieren intervención manual puedan se ejecutados de forma desatendida con la consecuente mejora en la eficiencia y reducción de riesgos y tiempo de respuesta.

Se indica en el siguiente gráfico las distintas etapas de trabajo

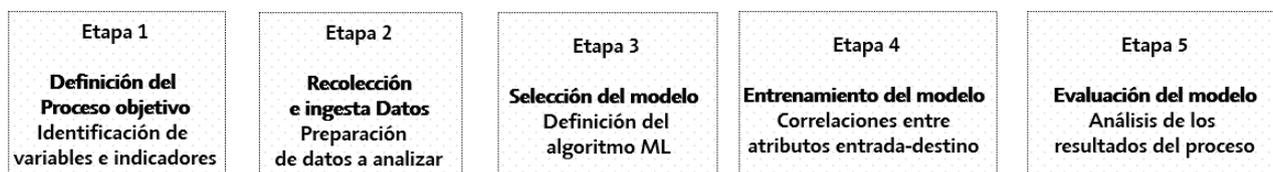


Gráfico 3: Etapas de ejecución del proceso de automatización

Fuente: Elaboración propia

Etapas del proceso proyecto

1. Definición del proceso objetivo.

El primer paso para el aprendizaje automático consiste en definir un objetivo, plantear un problema que requiera solución, es decir especificar el proceso que se automatizará.

Existen diferentes formas de estructurar el objetivo de un proyecto en el que se va a implementar tecnología Machine Learning. Eso sí, siempre hay que establecerlo en base a las necesidades empresariales y a las posibilidades que tiene la compañía en función de los datos de los que disponemos.

Para definir el proceso objetivo a automatizar, algunas de las preguntas a plantear son:

- ¿Cuál es el proceso objetivo específico a automatizar?
- ¿Qué se pretende enseñar a la máquina?
- ¿Es posible alcanzar el objetivo con los datos disponibles?
- ¿Cuál es el estado actual de la variable objetivo?
- ¿Cómo se van a medir los resultados?

2. Recolección e ingesta de datos.

Para que un software que utiliza Machine Learning funcione correctamente, debemos nutrirlo de una gran cantidad de datos. Estos servirán como punto de partida. Después, la máquina continuará su aprendizaje en base a los nuevos datos que vaya extrayendo y procesando.

Es más importante la calidad que la cantidad. No obstante, lo idóneo es plantear un equilibrio, es decir, cuantos más y mejores datos tengamos para empezar, mejor será el rendimiento del modelo. Para recolectar los datos, podemos recurrir al data scraping. Se trata de un método muy utilizado en IA para recopilar datos de fuentes diversas, como por ejemplo una página web, un blog o una hoja de cálculo o un web scraping.

La preparación de los datos para un proyecto de aprendizaje automático es un proceso largo y tedioso en el que el Data Analyst se enfrenta a grandes retos. Los datos tienen que ser analizados y procesados correctamente para evitar resultados engañosos.

3. Selección del modelo de ML

Esta es la etapa para elegir un modelo/algoritmo de Machine Learning que encuadre con el objetivo.

Existen diversos algoritmos de aprendizaje automático: predictivos, de clasificación, regresión lineal, clustering, Deep Learning, entre otras muchas variantes:

- Clasificación binaria: predice un resultado binario, es decir, experimentos estadísticos con solo dos resultados posibles. Ejemplo: ¿este email es spam o no?
- Modelo de clasificación multiclase: proporciona predicciones para varias clases, es decir, es capaz de predecir uno de más de dos resultados. Ejemplo: ¿qué categoría de productos recibe mayor cantidad de consultas?
- Modelo de regresión: predice un valor numérico. Ejemplo: ¿cuál será la temperatura en Bs As en el día de mañana?. En este sentido, algunos de los modelos más usados son los siguientes: random forest, regresión logística, regresión lineal, etc.

4. Entrenamiento del modelo.

El entrenamiento de los modelos de aprendizaje automático consiste en suministrar la información que permita que el algoritmo de Machine Learning haga su aprendizaje inicial.

En esta fase, los datos deben estar totalmente contrastados y albergar las respuestas correctas, también conocidas como atributos de destino. De esta forma, el algoritmo de aprendizaje es capaz de plantear correlaciones en los datos de entrenamiento que han sido asignados en los atributos de entrada y se proporciona un modelo que almacena dichas correlaciones.

5. Evaluación del modelo.

En esta etapa el proceso de evaluación determinará el funcionamiento del modelo. Sin ella, no se podrá dilucidar la validez del proyecto.

Para poder evaluar un modelo correctamente se debe separar en dos partes bien diferenciadas. Por un lado, la muestra de datos de prueba y por otro, la fuente de datos de entrenamiento.

Hay que adjuntar los datos al modelo elegido y comparar las predicciones devueltas con el valor de destino (objetivo de aprendizaje).

Además, se debe diseñar una métrica que indique la efectividad de la predicción y la coincidencia de valores. Si la exactitud es menor o igual a 50% la veracidad no será válida.

Tras la puesta en marcha de un modelo de Machine Learning pueden darse múltiples situaciones que den lugar a equívocos. Por ello, el análisis de errores es la última fase, aquella que permite modelar y cambiar los aspectos no relevantes para el proyecto.

Este análisis permitirá mejorar el rendimiento del modelo de aprendizaje automático y profundizar aún más en las fases previas a la implementación y entrenamiento.

Además, con el análisis de errores, se obtendrán las causas de los fallos durante el proceso y las nuevas conclusiones sobre oportunidad de mejora.

2.14. Bibliografía:

Agrawal, D, P. Bernstein, E. Bertino, S. Davidson and U. Dayal, "Challenges and Opportunities with Big Data". USA, CYBER CENTER TECHNICAL REPORTS, 2011, <http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1000&context=cctech>

Aral, S., Brynjolfsson, E., & Wu, D. J. (2006). Which came first, IT or Productivity? The virtuous cycle of investment and use in Enterprise Systems. International Conference on Information Systems, 1–22.

Broussard, M. (2018). *Artificial Unintelligence How Computers Misunderstand the World*. MIT Press.

Brynjolfsson, E. & McAfee, A. (2017). *Machine, platform, crowd: Harnessing our digital future*. WW Norton & Company

Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance? (SSRN Scholarly Paper Núm. ID 1819486). Recuperado de Social Science Research Network website: <https://papers.ssrn.com/abstract=1819486>

Gartner. (2018). Hype Cycle for Emerging Technologies, 2018. Disponible en https://www.gartner.com/technology/mediaproducts/reprints/ntt_com/340159.html?mkt_tok=eyJpIjojWIRFM05tSTVNeIE1WVRVMSIsInQiOiJQaWZlZ1dEVUVkNWpmWHBtRjNYb3h4bWFqOFhRUnI2WkZjY3JYZ0hcLzI4MXlTM1lOdmFEQ0ZobkdtdkUrmkU3R1FrTkV3aEZBTncwVkQwNzNhSkIcaXlGNdCrbnpaVTJC

Jiang, H., Wang, K., Wang, Y., Gao, M., & Zhang, Y. (2016). Energy big data: A survey. IEEE Access, 4, 3844–3861. <https://doi.org/10.1109/ACCESS.2016.2580581>

Katz, R. L. (2017). Social and economic impact of Digital Transformation on the economy. Disponible en: https://www.itu.int/en/ITU-D/Conferences/GSR/Documents/GSR2017/Soc_Eco_impact_Digital_transformation_finalGSR.pdf

Mckinsey Global Institute. (2017). Artificial Intelligence. The Next Digital frontier? Disponible en: <https://www.mckinsey.com/~media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx>

McAfee, A. (2002). The impact of enterprise information technology adoption on operational performance: An empirical investigation. *Production and Operations Management*, 11(1), 33–53. <https://doi.org/10.1111/j.1937-5956.2002.tb00183.x>

Mialaret, Gastón. *Psicología de la Educación*. 1ed. Paris: Siglo XXI Editores, 2001. pág.147

Oussous, A., Benjelloun, F.-Z., Ait Lahcen, A., & Belfkih, S. (2018). Big Data technologies: A survey. *Journal of King Saud University - Computer and Information Sciences*, 30(4), 431–448. <https://doi.org/10.1016/j.jksuci.2017.06.001>

Pautasio, Leticia (2019). IA comienza a mostrar aplicaciones en telecomunicaciones. MWC. Mobile World Congress 2019.

Porter, K. (2019). The Top Acquires in Artificial Intelligence And Select Emerging Growth Companies. *Woodside Capital Partners*, (February).

Gillham Jonathan (2018). The macroeconomic impact of AI, (February), 1–75. Publicado para PricewaterhouseCoopers. Disponible en https://www.researchgate.net/publication/326424203_The_macro-economic_impact_of_artificial_intelligence

Ruiz Limon, Ramón, “Historia y evolución del pensamiento científico”, México 2006, pág 130

The Royal Society. (2017). *Machine learning: the power and promise of computers that learn by example. Report by the Royal Society* (Vol. 66). <https://doi.org/10.1126/scitranslmed.3002564>

Vivas, Fredi (2020) *Data Science: El 2020 demostró que será esencial para nuestras vidas*. Artículo publicado en FiloNews en diciembre 2020. Disponible en <https://www.filo.news/ciencia/Data-Science-El-2020-demostro-que-sera-esencial-para-nuestras-vidas-20201224-0022.html>

2.15. Programación de actividades (Gantt):¹⁰

El proyecto tiene un tiempo planificado en 2 años. Se han establecido 5 etapas para el primer año de trabajo, las cuales tienen un total de actividades que se detallan en el cronograma respectivo. Para el segundo año se indican las actividades en el mismo cronograma. En la planilla siguiente se indican las tareas a lo largo de los dos años de ejecución. Cada tarea está acompañada de las iniciales de las personas que participarán más activamente en las mismas.

¹⁰ Definir la programación de actividades para cada objetivo específico, y las personas responsables de su ejecución.

Actividades/Responsables 1er año. 2022	Mes 1	Mes 2	Mes 3	Mes 4	Mes 5	Mes 6	Mes 7	Mes 8	Mes 9	Mes 10	Mes 11	Mes 12
Etapa 1: Identificación del proceso objetivo. Validación del cronograma de trabajo. Definición de alcance por etapas MC,AB,AB,BD	x	x	x									
Selección de plataforma de ML Definición de variables y distribución de tareas. MC,AA			x	x								
Configuración de herramienta ML Descripción de indicadores MC,AA,					x							
Etapa 2: Recolección e Ingesta de Datos. Ajustes y carga de datos MC, AB,					x	x						
Etapa 3: Selección del modelo. Especificación del algoritmos y ajustes del formato de cálculo. MC, DB,						x						
Etapa 4: Configuración del Elastic Indexación de datos Tratamiento de datos. MC,DB, AB						x	x	x	x			
Etapa 5: Análisis de datos en Kibana Elaboración de tablero preliminar. MC, AB,									x	x	x	x
Actividades/ Responsables 2do año. 2023	Mes 1	Mes 2	Mes 3	Mes 4	Mes 5	Mes 6	Mes 7	Mes 8	Mes 9	Mes 10	Mes 11	Mes 12
Etapa 5: Evaluación de los objetivos. Revisión de la alcances. MC, AB, AA	x	x	x									
Determinación de mejoras viables. Análisis de distintos procesos para automatizar. MC, AA,			x	x	x							
Implementación de mejoras. Exposiciones. MC, AA, AB,					x	x	x	X				
Ajustes finales y modelizaci MC, AB, BD								X	x	x		
Análisis final de resultados y elaboración de informes. MC, AB, AA, BD										x	x	x

Donde la descripción de las referencias en la asignación de tareas es:

MC: Marcelo Caiafa

AA: Ariel Aurelio

AB: Alejandro Bevilacqua

DB: Deborah Baner

Todas las tareas que detalladas oportunamente en el Gantt se concretarán en los laboratorios de informática del DIIT, que cuenta con computadoras con conexión a Internet, las cuales son adecuadas para la cantidad de personas que estarán asignadas al proyecto en distintas franjas horarias.

2.16. Resultados en cuanto a la producción de conocimiento:

Uno de los resultados del presente trabajo será la identificación de las mejores prácticas que validen el valor y la oportunidad que representa la plataforma de Machine Learning de código abierto conforme a su capacidad para generar información importante para los gestores de los sistemas de telecomunicaciones.

Se estima que lo anterior sea escalable a otras plataformas tecnológicas de forma tal que el servicio pueda ser ofrecido por los alumnos de la universidad en el mercado local.

2.17. Resultados en cuanto a la formación de recursos humanos:

Dentro de la formación del equipo de trabajo se destacan particularmente dos integrantes que tendrán su inicio en la investigación académica. Por un lado está la participación de una estudiante de la carrera de ingeniería Débora Baner, que dispone ya de más del 70% de las materias aprobadas. Se brindará particular atención a las tareas de asesoramiento, consultoría y mentoreo para el desarrollo de competencias a partir de las tareas que ejecute a lo largo del proyecto.

Por otro lado también será el inicio en trabajos de investigación de Alejandro Bevilacqua, docente del departamento de ingeniería hace más de 10 años en nuestra universidad.

Es importante aclarar que todas las tareas estarán supervisadas por el director del proyecto. Todos los integrantes del equipo compartirá las reuniones de planificación y seguimiento.

2.18. Resultados en cuanto a la difusión de resultados:

Está planificada la asistencia a congresos como CONAIISI, TEyET, CADI que además de permitir la difusión de los resultados del presente trabajo, posibilita el intercambio con pares sobre diversos tópicos relacionados con las plataformas tecnológicas utilizadas, los procedimientos seleccionados y las conclusiones alcanzadas.

2.19. Resultados en cuanto a transferencia hacia las actividades de docencia y extensión:

Actualmente en la currícula de la Universidad existen diversas asignaturas que dan tratamiento a las tecnologías de la información y las comunicaciones. Por ejemplo, en las asignaturas de ingeniería electrónica, Servicios avanzados de comunicaciones, Redes de computadoras y Desarrollo de Mercados Tecnológicos, entre otras. La particularidad que comparten entre ellas es el abordaje sobre distintas temáticas de la gestión de los sistemas de comunicaciones. Tanto desde un enfoque estrictamente técnico como de gestión de recursos para lo cual esta propuesta espera representar un agregado de valor.

2.20. Resultados en cuanto a la transferencia de resultados a organismos externos a la UNLaM:

Está prevista la transferencia de conocimientos a organismos externos a la UNLAM como el IEEE mediante la publicación de los resultados en su librería digital IEEE Xplore.

2.21. Vinculación del proyecto con otros grupos de investigación del país y del exterior:

No aplica

2.22. Destinatarios:

Tipo de destinatario		Subtipo de destinatario ¹¹	¿Cuál? Especificar	Demandante ¹²	Adoptante ¹³
Sector Gubernamental	Gobiernos	Del Poder Ejecutivo nacional			
		Del Poder Ejecutivo provincial			
		Del Poder Ejecutivo municipal			
	Otras Instituciones gubernamentales	Poder Legislativo en sus distintas jurisdicciones			
		Poder Judicial en sus distintas jurisdicciones			
Sector Salud		Hospitales, centros comunitarios de salud y otras entidades del sistema de atención			
Sector Educativo	Sistema universitario		unlam		X
	Sistema de educación básica y secundaria				
	Sistema de educación terciaria				
Sector Productivo	Empresas				
	Cooperativas de trabajo y producción				
	Asociaciones del Sector				
Sociedad Civil	ONG's y otras organizaciones sin fines de lucro				
	Comunidades locales y particulares				

¹¹ Marcar con una X

¹² Demandante: entidad administrativa de gobierno nacional, provincial o municipal constituida como demandante externo de las tecnologías desarrolladas, que determina la necesidad del proyecto por su importancia social. Marcar con una X

¹³ Adoptante: beneficiario o usuario en capacidad de aplicar los resultados desarrollados (organismos gubernamentales de ciencia y tecnología nacionales o provinciales; universidades e institutos universitarios de gestión pública o privada; empresas públicas o privadas; entidades administrativas de gobierno nacionales, provinciales o municipales; entidades sin fines de lucro; hospitales públicos o privados; instituciones educativas no universitarias; y organismos multilaterales. Marcar con una X

3-Recursos Existentes¹⁴

Descripción/ concepto	Cantidad	Observaciones

4-Recursos financieros¹⁵

	Rubro	Año 1	Año 2	Total
Gastos de capital (equipamiento)	a) Equipamiento (1)			
	b) Licencias (2)			
	c) Bibliografía (3)	\$5500,00	\$5500,00	\$11000,00
	Total Gastos de Capital	\$ 5500,00	\$ 5500,00	\$ 11000,00
Gastos corrientes (funcionamiento)	d) Bienes de consumo	\$ 5600,00	\$5600,00	\$11200,00
	e) Viajes y viáticos (4)	\$28600,00	\$28600,00	\$57200,00
	f) Difusión y/o protección de resultados (5)	\$5000,00	\$5000,00	\$10000,00
	g) Servicios de terceros (6)			
	h) Otros gastos (7)			
	Total Gastos Corrientes	\$ 39200,00	\$ 39200,00	\$ 78400,00
Total Gastos (Capital + Corrientes)		\$44700,00	\$44700,00	\$89400,00

Aclaraciones sobre rubros del presupuesto

1 Equipamiento: Equipamiento, repuestos o accesorios de equipos, etc.

2 Licencias: Adquisición de licencias de tecnología (software, o cualquier otro insumo que implique un contrato de licencia con el proveedor).

3 Bibliografía: En el caso de compra de bibliografía, ésta no debe estar accesible como suscripción en la Biblioteca Electrónica.

Bibliografía: se estima inversión de \$5.500,00 en cada año.

¹⁴ Antes de confeccionar el presupuesto del proyecto, será necesario que el Director incluya en esta tabla si dispone de recursos adquiridos con fondos de proyectos anteriores (equipamiento, bibliografía, bienes de consumo, etc.) a ser utilizados en el proyecto a presentar, y además se recomienda consultar en la Unidad académica la disponibilidad de recursos existentes factibles de ser utilizados en el presente proyecto.

¹⁵ Justificar presupuesto detallado. Para compras de un importe superior a \$15000.- se requieren tres presupuestos. (Resolución Rectoral N°177/2021.)

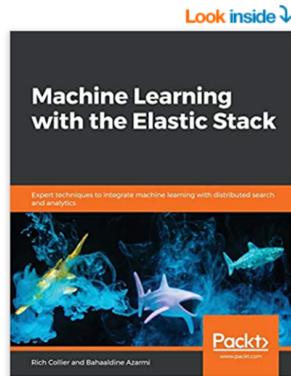
Comprar por las categorías principales que se envían a otros países

Libros > Computadoras y Tecnología > Bases de Datos

Machine Learning with the Elastic Stack: Expert techniques to integrate machine learning with distributed search and analytics

de Rich Collier (Author), Bahaaldine Azarmi (Author)

★★★★★ 10 calificaciones



ISBN-13: 978-1788477543

Kindle **Pasta blanda** **US\$ 39.30** Todos Ver todas las 2 versiones

Nuevo: **US\$ 39.30**
Disponible.
Vendido y enviado por Amazon.com.
Disponible a un precio menor de otros vendedores que podrían no ofrecer envío Prime gratis.

Entrega el **Viernes, Marzo 11**.
Realiza el pedido en 15 hrs 39 mins
[Enviar a Argentina](#)
Cantidad: 1
[Agregar al Carrito](#)
[Comprar ahora](#)

4 Viajes y viáticos: Viajes y viáticos en el país: Gastos de viajes, viáticos de campaña y pasantías en otros centros de investigación estrictamente listados en el proyecto. Gastos de viaje en el exterior: (no deberán superar el 20% del monto del proyecto).

Costo estimado de 1 viajes y viáticos asociados: \$28600,00

Desayuno

Huentala Hotel

Mendoza. A 587 m del centro [Ver en mapa](#)

8.1 ★★★★★

[Reserva flexible](#) Incluye traslado ida y vuelta

Últimas 3 habitaciones

Ahorrás \$4.616

Precio final por persona \$30.874

\$28.566

Final 2 personas \$57.132 Incluye impuestos, tasas y cargos

[Siguiente](#)

[Pasaporte Despegar](#) Sumarías 1.061 puntos

Hasta en 12 cuotas sin interés [Ver medios de pago](#)

Vuelo directo BUE ↔ MDZ

5 Difusión y/o protección de resultados: Ej.: (Gastos para publicación de artículos, edición de libros inscripción a congresos y/o reuniones científicas).

Gastos de Participación en Eventos científicos

Costo de inscripción \$5000,00 (considerando un mínimo de 1 evento anual)

ORIGINAL							
UBATEC SA		B COD. 006		FACTURA			
				Punto de Venta: 00002 Comp. Nro: 00004735		Fecha de Emisión: 01/10/2021	
Razón Social: UBATEC SA		Domicilio Comercial: Saenz Peña R. Pte. Av. 938 Piso:6 - Ciudad de Buenos Aires		CUIT: 30644259486		Ingresos Brutos: 01506571-05	
Condición frente al IVA: IVA Responsable Inscripto				Fecha de Inicio de Actividades: 01/06/1991			
Período Facturado Desde: 01/10/2021		Hasta: 01/10/2021		Fecha de Vto. para el pago: 01/10/2021			
CUIT: 30646228685		Apellido y Nombre / Razón Social: UNIVERSIDAD NACIONAL DE LA MATANZA					
Condición frente al IVA: IVA Sujeto Exento		Domicilio: Florencio Varela 1903 - San Justo, Buenos Aires					
Condición de venta: Contado							
Código	Producto / Servicio	Cantidad	U. Medida	Precio Unit.	% Bonif	Imp. Bonif.	Subtotal
1	Pago Inscripción Congreso CADI/ CLADI/CAEDI 2021 - Marcelo Caiafa	1,00	unidades	3500,00	0,00	0,00	3500,00

Gastos de Publicación

Costo de inscripción \$10000,00 (considerando un mínimo de 2 eventos a lo largo de todo el proyecto)

6 Servicios de terceros: Servicios de terceros no personales (reparaciones, análisis, fotografía, etc.).

7 Otros gastos: Incluir, si es necesario, gastos a realizar que no fueron incluidos en los otros rubros.

Se adjunta el detalle de la justificación de la estimación de gastos:

a) Bienes de consumo:

1. Cartucho de toner \$3400,00

articulo.mercadolibre.com.ar/MLA-622613882-toner-hp-c9700a-original-121a-negro-1500-2500-..._JM?searchVariation=48824631985#searchVariati...

También puede interesarte: toner 85a - toner brother hl1212w - toner samsung m2020w - tn 1060 - toner xerox 3020 - toner hp p1102w

Volver al listado | Computación > Impresión > Insumos de Impresión > Tóneres

Compartir | Vender uno igual

Nuevo

Toner Hp C9700a Original 121a Negro 1500 2500

\$ 3.400
en 12x \$ 482¹⁵

Ver los medios de pago

Envío gratis a todo el país
Conocé los tiempos y las formas de envío.
Calcular cuándo llega

Devolución gratis
Tenés 30 días desde que lo recibís.
Conocer más

Color De La Tinta: Negro

2. Resmas de papel \$2200,00 (3 unidades)



Nuevo | 5 vendidos

Resmas Tempo Naturaleza A4 80 Gr.



★★★★★ 1750 opiniones

\$ 740

en 12x de \$ 136⁰⁹ pagando con Mercado Crédito

[Ver los medios de pago](#)

Envío a todo el país
Conocé los tiempos y las formas de envío.
[Calcular cuándo llega](#)

Devolución gratis
Tenés 30 días desde que lo recibís.
[Conocer más](#)

Color: **Bianco**

4.1 Origen de los fondos solicitados

Institución	% Financiamiento
UNLaM	100%
Otros (indicar cuál)	



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

Departamento: Ingeniería e Investigaciones Tecnológicas

**Programa de acreditación:
CyTMA2**

Programa de Investigación¹:

Código del Proyecto: C2-ING-092

**Título del proyecto
Automatización de servicios de telecomunicaciones con Machine Learning**

PIDC:

Elija un elemento.

PII:

Elija un elemento.

Director: Marcelo Dante Caiafa

Director externo:

Codirector: Ariel Aurelio

Integrantes: Alejandro Bevilacqua

Investigador Externo, Asesor- Especialista, Graduado UNLaM:

Alumnos de grado: Deborah Baner (Beca UNLaM)

Resolución Rectoral de acreditación: N°760/22

Fecha de inicio: 01/01/22

Fecha de finalización: 31/12/23

¹ Los Programas de Investigación de la UNLaM están acreditados con resolución rectoral, según lo indica la Resolución HCS N° 014/15 sobre **Lineamientos generales para el establecimiento, desarrollo y gestión de Programas de Investigación a desarrollarse en la Universidad Nacional de La Matanza**. Consultar en el departamento académico correspondiente la inscripción del proyecto en un Programa acreditado.



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

A. Desarrollo del proyecto

A.1. Grado de ejecución de los objetivos inicialmente planteados, modificaciones o ampliaciones u obstáculos encontrados para su realización (desarrolle en no más de dos (2) páginas).

En lo que respecta a modificaciones sucedidas durante el desarrollo del proyecto, es importante destacar que la alumna y becaria Daborah Baner completó exitosamente la carrera de ingeniería electrónica y alcanzó su graduación.

A su vez es importante destacar que para el ing Alejandro Bevilacqua, exalumno y docente de la UNLAM, la participación en este trabajo represento para él su primera experiencia en la investigación académica.

Referido a la planificación inicialmente propuesta para las diferentes tareas, se puede comunicar en este informe final que, los puntos completos respecto a la planificación prevista alcanzaron un nivel satisfactorio de completitud.

Si bien las tareas propuestas en la planificación se alcanzaron a completar, se presentaron algunas cuestiones que alteraron en parte los tiempos inicialmente previstos para cada una de ellas.

Oportunamente se planificó que la ejecución de las tareas de implementación, instalación y configuración de la plataforma Python se realizaria a partir de los recursos informáticos disponibles en los laboratorios de ingeniería de la sede de la Unlam. Sin embargo resultó mucho más práctico que cada integrante utilizara como herramientas de desarrollo sus propias computadoras agilizando considerablemente el avance del proyecto.

Otra tarea que demandó mayor tiempo al previsto fue la adquisición de los datos relacionados con el proceso a automatizar. Inicialmente se supuso que dispondríamos de planillas de datos en formato txt o similar. En la práctica fue necesario ejecutar un proceso de scraping sobre un servidor web para la obtención de los datos reales. Para ello se utilizó la librería BeautifulSoup de Python.

Si bien lo anterior generó demoras respecto de la planificación propuesta posibilitó desarrollar técnicas que, si bien no estaban previstas inicialmente, contribuyen al ejercicio en un ambiente profesional de cualquier proyecto de similares características.



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

La estructura de diccionario de datos fue algo que no se pudo adecuar, es decir que fue preciso desarrollar habilidades de análisis a partir de los datos disponibles. De hecho una de las propuestas fue enriquecer el dataset para nutrir con mayor cantidad de variables al conjunto.

B. Principales resultados de la investigación

B.1. Publicaciones en revistas (informar cada producción por separado)

Artículo 1:	
Autores	<i>Caiafa Marcelo Dante Aurelio Ariel Bevilacqua Alejandro</i>
Título del artículo	<i>Tecnología Big Data para monitorear datos de servicios TIC</i>
N° de fascículo	10
N° de Volumen	3
Revista	<i>International Journal of Human Sciences Research</i>
Año	2023
Institución editora de la revista	<i>Atena</i>
País de procedencia de institución editora	<i>Brasil</i>
Arbitraje	SI
ISSN:	2764-0558
URL de descarga del artículo	https://www.atenaedi-tora.com.br/catalogo/revista/international-journal-of-human-sciences-research-v-3-n-10-issn-2764-0558/295
N° DOI	10.22533

B.2. Libros

Libro 1	
Autores	
Título del Libro	



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

Año	
Editorial	
Lugar de impresión	
Arbitraje	Elija un elemento.
ISBN:	
URL de descarga del libro	
N° DOI	

B.3. Capítulos de libros

Autores	<i>Caiafa Marcelo Dante Aurelio Ariel Bevilacqua Alejandro</i>
Título del Capítulo	<i>Análisis de Datos sobre la tasa de abandono en telecomunicaciones. Un proyecto de vinculación profesional evaluado con rúbricas analíticas.</i>
Título del Libro	<i>Ingeniería: Innovación, ciencia y tecnología</i>
Año	2024
Editores del libro/Compiladores	<i>Denis Medina Guedes Leonardo Franco da Silva Victor Crespo de Oliveira</i>
Lugar de impresión	<i>Ponta Grossa. Brasil</i>
Arbitraje	SI
ISBN:	978-65-258-2273-0
URL de descarga del capítulo	www.atenaeditora.com.br/catalogo/post/analisis-de-datos-sobre-la-tasa-de-abandono-en-telecomunicaciones-un-proyecto-de-vinculacion-profesional-evaluado-con-rubricas-analiticas
N° DOI	10.22533

B.4. 1 Trabajos presentados a congresos y/o seminarios

--



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

Autores	<i>Caiafa Marcelo Dante Aurelio Ariel Bevilacqua Alejandro Baner Deborah Yanina</i>
Título	<i>Proyecto de análisis de datos en servicios de telecomunicaciones evaluado con rúbricas analíticas</i>
Año	<i>2023</i>
Evento	<i>XVIII Congreso de Tecnología en Educación y Educación en Tecnología 2023</i>
Lugar de realización	<i>Universidad Nacional de Hurlingham</i>
Fecha de presentación de la ponencia	<i>15 de junio 2023</i>
Entidad que organiza	<i>RedUNCI</i>
URL de descarga del trabajo (especificar solo si es la descarga del trabajo; formatos pdf, e-pub, etc.)	<i>https://teyet2023.unahur.edu.ar/</i>

B.4. 2. Trabajos presentados a congresos y/o seminarios

Autores	<i>Caiafa Marcelo Dante Aurelio Ariel Bevilacqua Alejandro Baner Deborah Yanina</i>
Título	<i>Data analysis for telecommunication services</i>
Año	<i>2023</i>
Evento	<i>XI jornadas de cloud computing, Big Data and Emerging Topics</i>
Lugar de realización	<i>Universidad Nacional de La Plata</i>
Fecha de presentación de la ponencia	<i>25 de junio 2023</i>
Entidad que organiza	<i>LIDI</i>
URL de descarga del trabajo (especificar solo si es la descarga del trabajo; formatos pdf, e-pub, etc.)	<i>http://se-dici.unlp.edu.ar/handle/10915/155281</i>

B.4. 3. Trabajos presentados a congresos y/o seminarios



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

Autores	<i>Caiafa Marcelo Dante Aurelio Ariel Bevilacqua Alejandro Baner Deborah Yarina</i>
Título	<i>Data analysis for telecommunication services</i>
Año	<i>2023</i>
Evento	<i>52 jornadas argentinas de Ingeniería Informática</i>
Lugar de realización	<i>Universidad Nacional de Tres de Febrero</i>
Fecha de presentación de la ponencia	<i>6 de septiembre 2023</i>
Entidad que organiza	<i>SADIO</i>
URL de descarga del trabajo (especificar solo si es la descarga del trabajo; formatos pdf, e-pub, etc.)	<i>https://publicaciones.sadio.org.ar/index.php/JAIIO/issue/archive</i>

B.4. 4. Trabajos presentados a congresos y/o seminarios

Autores	<i>Caiafa Marcelo Dante Aurelio Ariel Bevilacqua Alejandro Baner Deborah Yarina</i>
Título	<i>Implementación de análisis de datos en servicios de telecomunicaciones</i>
Año	<i>2023</i>
Evento	<i>VII Jornadas de la RedVITEC</i>



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

Lugar de realización	<i>UNL Universidad Nacional del Litoral</i>
Fecha de presentación de la ponencia	<i>6 de diciembre 2023</i>
Entidad que organiza	<i>RedVITEC Red de Vinculación Tecnológica de las universidades Nacionales de Argentina</i>
URL de descarga del trabajo (especificar solo si es la descarga del trabajo; formatos pdf, e-pub, etc.)	<i>https://redvitec.cin.edu.ar/jornadas2023.html</i>

B.5. Otras publicaciones

Autores	
Año	
Título	
Medio de Publicación	

C. Otros resultados. Indicar aquellos resultados pasibles de ser protegidos a través de instrumentos de propiedad intelectual, como patentes, derechos de autor, derechos de obtentor, etc. y desarrollos que no pueden ser protegidos por instrumentos de propiedad intelectual, como las tecnologías organizacionales y otros. Complete un cuadro por cada uno de estos dos tipos de productos.

C.1. Títulos de propiedad intelectual. Indicar: Tipo (marcas, patentes, modelos y diseños, la transferencia tecnológica) de desarrollo o producto, Titular, Fecha de solicitud, Fecha de otorgamiento

Tipo	Titular	Fecha de Solicitud	Fecha de Emisión

C.2. Otros desarrollos no pasibles de ser protegidos por títulos de propiedad intelectual. Indicar: Producto y Descripción.

Producto	Descripción



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

D. Formación de recursos humanos. Trabajos finales de graduación, tesis de grado y posgrado. Completar un cuadro por cada uno de los trabajos generados en el marco del proyecto.

D.1. Tesis de grado

Director (apellido y nombre)	y Autor (apellido y nombre)	Institución	Calificación	Fecha /En curso	Título de la tesis

D.2 Trabajo Final de Especialización

Director (apellido y nombre)	y Autor (apellido y nombre)	Institución	Calificación	Fecha /En curso	Título del Trabajo Final

D.2. Tesis de posgrado: Maestría

Director (apellido y nombre)	y Tesista (apellido y nombre)	Institución	Calificación	Fecha /En curso	Título de la tesis

D.3. Tesis de posgrado: Doctorado

Director (apellido y nombre)	y Tesista (apellido y nombre)	Institución	Calificación	Fecha /En curso	Título de la tesis

D.4. Trabajos de Posdoctorado

Director (apellido y nombre)	y Posdoctorando (apellido y nombre)	Institución	Calificación	Fecha /En curso	Título del trabajo	Publicación

E. Otros recursos humanos en formación: estudiantes/ investigadores (grado/posgrado/ posdoctorado)



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

Apellido y nombre del Recurso Humano	Tipo	Institución	Período (desde/hasta)	Actividad asignada ²
Deborah Yanina Bamer	Estudiante de grado	UNLAM	01/01/2022 31/12/2023	
Alejandro Bevilacqua	investigador	UNLAM	01/01/2022 31/12/2023	

F. Vinculación³: Indicar conformación de redes, intercambio científico, etc. con otros grupos de investigación; con el ámbito productivo o con entidades públicas. Desarrolle en no más de dos (2) páginas.

Se ha desarrollado una estrecha vinculación entre este grupo de investigación académica y el ámbito productivo con una institución corporativa dedicada al área de servicios dentro del mercado masivo local conformada por más de 7000 clientes distribuidos dentro del territorio de la provincia de Bs As.

De esta manera se logró el acceso a registros de información de plataformas productivas de telecomunicaciones y sus registros relacionados. Estos son fuente de datos de una infraestructura productiva en Argentina para su análisis en el presente trabajo.

Es importante destacar que se logró completar la participación en los congresos

TEyET 2023

XVIII Congreso de Tecnología en Educación y Educación en Tecnología 2023 realizadas en la Universidad Nacional de Hurlingham

JCC-BD&ET 2023

XI Jornadas de Cloud Computing-Big Data and Emerging Topics realizadas en la Universidad nacional de La Plata

JAIIO2023

52 Jornadas Argentinas de ingeniería informática realizadas en la Universidad de Tres de Febrero

RedVITEC2023

VII Jornadas académicas de la RedVITEC Vinculación tecnológica en las universidades

² Descripción de la/s actividad/es a cargo (máximo 30 palabras)

³ Entendemos por acciones de “vinculación” aquellas que tienen por objetivo dar respuesta a problemas, generando la creación de productos o servicios innovadores y confeccionados “a medida” de sus contrapartes.



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

G. Otra información. Incluir toda otra información que se considere pertinente.

H. Cuerpo de anexos:

- Anexo I: Copia de cada uno de los trabajos mencionados en los puntos B, C y D, y certificaciones cuando corresponda.⁴
- Anexo II:
 - FPI-013: Evaluación de alumnos integrantes. (si corresponde)
 - FPI-014: Comprobante de liquidación y rendición de viáticos. (si corresponde)
 - FPI-015: Rendición de gastos del proyecto de investigación acompañado de las hojas foliadas con los comprobantes de gastos.
 - FPI-035: Formulario de reasignación de fondos en Presupuesto.
- Nota justificando baja de integrantes del equipo de investigación.

_____ Caiafa Marcelo Dante _____
Firma y aclaración
del director del proyecto.

Lugar y fecha: San Justo, 29 de Febrero 2024.

- Cargar este formulario junto con los documentos correspondientes **exclusivamente** al Anexo I en SIGEVA UNLaM.

ANEXO I

⁴ En caso de libros, podrá presentarse una fotocopia de la primera hoja significativa o su equivalente y el índice.



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

Presentación a TEyET 2023

Proyecto de análisis de datos de servicios de telecomunicaciones evaluado con rúbricas analíticas

Marcelo Dante Caiafa¹, Ariel Rodrigo Aurelio¹, Alejandro Bevilacqua¹, Debora Baner¹
*Departamento de Ingeniería e Investigaciones Tecnológicas,
Universidad Nacional de La Matanza, Argentina.*

mcaiafa@unlam.edu.ar, aaurelio@unlam.edu.ar, abevilacqua@unlam.edu.ar, dbaner@alumno.unlam.edu.ar

- ORIGINAL ARTICLE -

Resumen

La vinculación profesional entre la universidad y la industria es objetivo de nuestra labor académica. Es interés del trabajo registrar un proyecto de articulación basado en el análisis de datos de servicios de telecomunicaciones. El desempeño de los estudiantes de ingeniería que lo implementan se evalúa a partir de rúbricas analíticas elaboradas para este caso.

Los procesos de transformación digital originan profundos cambios, entre otros en la generación de bienes y servicios. En este contexto la obtención de información relevante a partir de datos disponibles resulta un aporte de valor para las organizaciones.

El proyecto se basa en un estudio exploratorio de datos reales, que utiliza como herramienta la plataforma Python y sus diferentes librerías, para construir un perfil de cliente con alto potencial de abandono en los servicios de telecomunicaciones.

Se pretende contribuir al fomento de actividades que integren la academia con la industria reflejando el valor que la aplicación de nuevas herramientas de evaluación pueden aportar a estudiantes y docentes.

Palabras Clave: Tecnología de la Información, análisis de datos, rúbricas analíticas, competencias profesionales.

El valor del análisis de datos

La transformación digital está generando en las últimas décadas profundos cambios en la forma de producir y comercializar bienes y servicios [1]. En este contexto la gestión de la información permite que las organizaciones mejoren sus procesos, “los datos son la nueva moneda que sustenta cambios fundamentales en la cuarta revolución industrial” [2]. Una cultura organizacional orientada en datos dispone de ventajas en la toma de decisiones porque se basa en evidencias [3].

Algunos estudios [4] indican que las organizaciones con enfoque basado en datos mejoran su productividad y rentabilidad. Según consultoras internacionales [5] el tamaño del mercado global de analítica de datos representó USD 31,800 millones para 2021 y estiman alcance los USD 329,800 millones en 2030. Eso registraría una tasa de crecimiento anual compuesta (CAGR) del 29,9% en ese período.

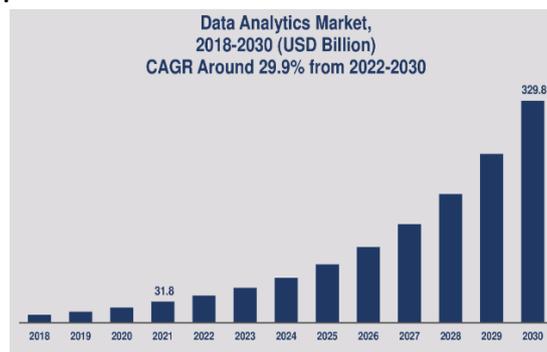


Figura 1: Mercado de Analítica de Datos

Fuente: Acumen R&C, 2021

Se puede definir el análisis de datos como un proceso multidisciplinario que gestiona datos con el propósito de generar información para mejorar la toma de decisiones [6].

Tukey definió el análisis de datos como "los procedimientos y técnicas para interpretar resultados, formas de planificar la recopilación de datos para hacer su análisis más sencillo y preciso" [7]. Actualmente se clasifican diferentes tipos de analítica:

- a) Analítica descriptiva: ilustra los datos de los resultados recopilados durante un intervalo de tiempo.
- b) Analítica de diagnóstico: busca la causa raíz de un problema.
- c) Analítica predictiva: utiliza datos pasados para realizar pronósticos.
- d) Analítica prescriptiva: está dedicada a encontrar la solución más adecuada.

Si bien existen diferentes modelos de madurez analítica, en el siguiente gráfico se ordenan los distintos tipos de análisis. Estos se pueden considerar como etapas correlativas del mismo proyecto, que a medida que avanza nivel de complejidad incrementa el valor del aporte [8].



Figura 2: Modelo de ascendencia analítica
Fuente: Gartner, 2012.

Aunque no existe completo consenso sobre el modelo, resulta útil asociarlo a diferentes tipos de preguntas para ayudar a centrar los esfuerzos analíticos. [9]

Alcance del proyecto

El alcance del presente proyecto queda enmarcado dentro de la instancia inicial del modelo de madurez analítico, es decir que se trata de un análisis descriptivo que genera información a partir del estudio de un marco de datos históricos reales y concretos.

Roles y perfiles en un equipo

Para identificar las habilidades a desarrollar en el proyecto se parte de las definiciones genéricas de responsabilidades de los distintos roles desempeñados en un equipo de trabajo para este tipo de proyectos. Algunos autores [10] indican que entre los perfiles requeridos por el mercado laboral están:

- Analista de datos, ejecuta análisis estadísticos de diversas fuentes de información.
- Científico de datos, aplica matemáticas, estadística y programación para explorar, examinar y graficar datos.
- Gerente de datos (Chief Data Officer), es el ejecutivo responsable final de toda la gestión de datos y sus políticas de privacidad.
- Ingeniero de datos, es responsable de probar, implementar y mantener la infraestructura de datos y sus repositorios.
- Arquitecto de datos, es quien define, diseña y supervisa la implementación de los sistemas subyacentes que se utilizan.
- Traductor de datos, es un rol emergente que enlaza con las unidades de negocio relevando necesidades, validando requerimientos y comunicando resultados.

Etapas del proyecto

El estudio está basado en la articulación del mundo académico-profesional. Se enfoca en un proyecto de análisis exploratorio de datos de servicios de telecomunicaciones. En la planificación del proyecto se toman como referencia las recomendaciones de diferentes autores [11], [12]. Si bien un autor citado [13] avanza en detalle para lenguaje R, el presente trabajo fue realizado en Python y sus librerías (Pandas, Numpy y Matplotlib).

Para facilitar la valoración y el seguimiento del desarrollo de competencias profesionales el análisis de datos se agrupó en tres etapas [14] que están identificadas como A, B, C respectivamente en las rúbricas.

Etapas de Relevamiento de datos: interpretación del contexto, especificación del objetivo y validación de requerimientos.

Etapas de Tratamiento de datos: colección de datos, limpieza de datos, identificación de variables, construcción de diccionario de datos, análisis de relaciones entre variables, descripción estadística de datos.

Etapas de Presentación: visualización de la información, sumarización de observaciones y exposición de resultados.

Una característica del EDA (Exploratory Data Analysis) es que se centra más en el proceso que en la teoría. Se puede aprender un puñado de técnicas de análisis de nivel de entrada (estadísticas descriptivas, correlaciones y visualizaciones básicas) en una cantidad de tiempo de algunas semanas [15]. Esto combinado con la curva de aprendizaje de un lenguaje interpretativo de alto nivel como Python resultó indicado para este proyecto.

Problemática para investigar

Un relevamiento del CIECTI (Centro Interdisciplinario de Estudios en Ciencia, Tecnología e Innovación) revela cierta desconexión entre la producción académica y los desarrollos de empresas locales [16].

Algunos autores [17] dentro de su propuesta de enseñanza-aprendizaje incorporan como estrategia y métodos de evaluación en la enseñanza de ingeniería de software, la ejecución de proyectos universidad-empresa. La definición de ciencia multidisciplinaria, la diversidad de perfiles y los distintos roles de los integrantes de un equipo de análisis de datos supone que será necesario el desarrollo de distintos tipos de habilidades.

Como caso de estudio concreto se consideran los datos históricos de un conjunto de clientes de servicios de telecomunicaciones, debidamente organizados y tratados en bloque, para definir un perfil de cliente con alta probabilidad de solicitud de baja de servicio.

De allí que las preguntas de investigación que guían el trabajo son:

- ¿Cómo desarrollar un proyecto de análisis de datos de servicios de telecomunicaciones con estudiantes de ingeniería que los vincule con la industria?
- ¿Cuáles son las competencias profesionales, habilidades técnicas y genéricas, que requiere un ingeniero que se desempeña en proyectos de análisis de datos y cómo pueden ser evaluadas?

Objetivos del trabajo

El propósito de la investigación apunta a evaluar el desarrollo de las habilidades y competencias profesionales puestas en práctica por estudiantes de ingeniería en la ejecución de un proyecto de análisis de datos.

En este marco, los objetivos del estudio son:

1. Registrar el proceso para identificar un perfil típico de cliente con mayor tasa de abandono, a partir de las variables más representativas detectadas en un conjunto de datos de servicios de telecomunicaciones.

2. Construir propuestas metodológicas de evaluación de desempeño de competencias profesionales acorde a necesidades actuales.

Se pretende aportar valor a la formación del perfil del profesional que trabaja en el sector TIC (tecnología informática y comunicaciones) destacando las habilidades necesarias para lograr un eficiente desempeño en la exploración de datos. Este tipo de proyectos, no trata sólo de implementación de infraestructura tecnológica sino que, por su naturaleza y potencial impacto son estratégicos para sustentar la toma de decisiones basada en evidencia [18].

Rúbricas analíticas

El resultado final pretende aportar un caso concreto de vinculación académica evaluado a partir de la utilización de rúbricas analíticas. Una rúbrica es un instrumento cuya principal finalidad es compartir los criterios de realización de las tareas de aprendizaje y de evaluación con los estudiantes y entre el profesorado. La rúbrica, como guía u hoja de ruta de las tareas, muestra las expectativas que estudiantes y docentes comparten sobre distintas actividades, organizadas en diferentes niveles de desempeño [19].

La evaluación del proceso de enseñanza-aprendizaje requiere, además de la asimilación de conocimientos, el desarrollo de habilidades necesarias para resolver problemas en contextos reales. Según algunos autores [20], una habilidad es la capacidad de utilizar recursos necesarios para responder efectivamente a una situación compleja en un contexto específico.

Si bien existen distintos enfoques, se siguen las definiciones de algunos autores que clasifican las habilidades en dos enfoques: tipo técnico (conocimientos) o interpersonales (actitudes sociales) [21], así el perfil del ingeniero está conformado por competencias técnicas y competencias genéricas [22].

Competencias técnicas específicas

Suelen ser referenciadas como habilidades duras, son las requeridas para el desempeño de una ocupación en concreto. Están relacionadas con funciones o puestos de trabajo en particular. Aportan al estudiante los conocimientos, actitudes, habilidades y valores propios de cada profesión y actividad laboral.

Habilidades genéricas transversales

Las habilidades genéricas, también llamadas habilidades blandas, en ingeniería de software se pueden identificar a partir de un modelo semántico que especifica el marco de trabajo para su evaluación.

Este entorno de tareas se basa en el hecho que el estudiante tiene un perfil de habilidades por desarrollar y para cada una define un conjunto de hitos a alcanzar [23].

Desarrollo del trabajo

El desarrollo del proyecto de investigación se estructura en cinco instancias ejecutadas en forma secuencial.

En la primera instancia se realizaron las tareas de selección de la herramienta de software y la elaboración de las rúbricas.

En la segunda instancia, que corresponde a la etapa de relevamiento de datos, se estudia el contexto, se validan los requerimientos, y se definen los indicadores.

La tercera instancia corresponde al tratamiento de los datos. Con los datos obtenidos se identifican las variables, se realiza un diccionario de datos, se limpian, se ordenan y se ejecuta el análisis estadístico.

En la instancia número cuatro se ejecutan las operaciones vinculadas con la presentación de resultados. Se construye la matriz de correlación para identificar las variables relevantes y son

validadas con el mapa de calor. También se elaboran los gráficos para la presentación de los informes.

La última instancia es para el análisis final del proyecto la elaboración de conclusiones y propuestas de futuros trabajos.

Selección de la herramienta de software

Inicialmente se analizaron diversas opciones como R, ELK, powerBI. Se definió la utilización de Python como lenguaje de programación por ser la más adecuada por sus características. Entre otras tenemos un lenguaje interpretado, de alto nivel, de tipado dinámico, multiplataforma, de código abierto, orientado a objeto, su gran comunidad y amplias y diversas librerías, entre otras [24].

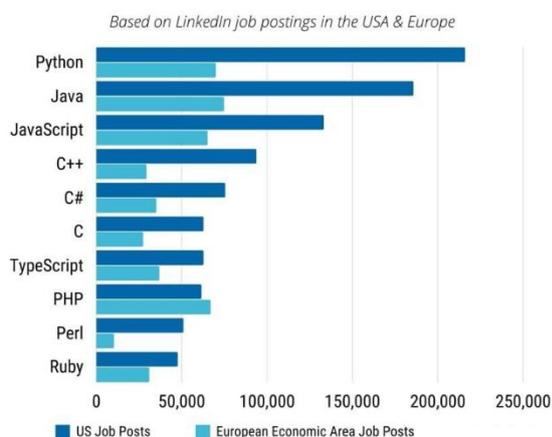


Figura 3: Lenguajes de programación más requeridos

Fuente: LinkedIn, 2022.

Otra de las razones que motivó la decisión de utilizar Python como lenguaje es su gran demanda en las ofertas de empleo en USA y EU que registra LinkedIn y se muestra en el gráfico anterior [25].

En esta primera instancia además, a partir de las recomendaciones del CONFEDI (Consejo Federal de Decanos de Ingeniería) [26] y otras, se confeccionaron las rúbricas para el seguimiento y evaluación del desarrollo de catorce competencias de parte de los estudiantes durante el proceso. Se adjuntan en hoja separada.

Etapas de Relevamiento de Datos

En esta instancia se estudió el contexto de la problemática a resolver y su impacto. Es decir que se indagó sobre el valor que aportaría la solución buscada. Esto significó estudiar la relación entre el ciclo de vida del cliente y la tasa de abandono. Además se profundizó sobre los costos relacionados con la retención de un cliente existente vs los costos de captación de uno nuevo, distintas estrategias de fidelización y venta cruzada, etc.

Se colectaron los datos y se procedió a un primer acercamiento en un análisis preliminar según las mejores prácticas [27].

Etapas de Tratamiento de Datos

Aquí se realizan distintas manipulaciones del conjunto de datos. Primero se consolida el tipo de dato por cada variable. Se detectan datos faltantes y outliers, se ordenan y consolidan para un posterior análisis estadístico y se mapean variables categóricas a variables numéricas para facilitar su procesamiento.

customerID	object	DeviceProtection	object
gender	object	TechSupport	object
SeniorCitizen	int64	StreamingTV	object
Partner	object	StreamingMovies	object
Dependents	object	Contract	object
tenure	int64	PaperlessBilling	object
PhoneService	object	PaymentMethod	object
MultipleLines	object	MonthlyCharges	float64
InternetService	object	TotalCharges	object
OnlineSecurity	object	Churn	object
OnlineBackup	object		dtype: object

Figura 4: Identificación de tipo de dato x variable

Fuente: Elaboración propia

Luego se eliminan datos nulos e inconsistentes. Estas tareas se ejecutan con librería Pandas y Numpy, según las buenas prácticas [28].

customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	DeviceProtection	Tech
0	Female	0	Yes	No	1	No	No phone service	DSL	No	...	No	No
1	Male	0	No	No	34	Yes	No	DSL	Yes	...	Yes	Yes
2	Male	0	No	No	2	Yes	No	DSL	Yes	...	No	No
3	Male	0	No	No	45	No	No phone service	DSL	Yes	...	Yes	Yes
4	Female	0	No	No	2	Yes	No	Fiber optic	No	...	No	No

Figura 5: Extracto representativo de datos

Fuente: Elaboración propia

Se ordenan las variables, se renombran algunas y otras se redefinen a partir de los indicadores. Se elabora el diccionario de datos.

Variable	Descripción
a)Customer ID:	Identificación interna del cliente
b)Gender:	Tipo de género
c)Senior Citizen:	Si el cliente es un adulto
d)Partner:	El cliente tiene pareja o no
e)Dependents:	El cliente tiene dependientes o no
f)Tenure:	Número de meses de antigüedad del cliente en la compañía
g)Phone Service:	El cliente tiene contratado el servicio de telefonía o no
h)Multiple lines:	El cliente tiene contratadas múltiples líneas o no
i)Internet Service:	Tipo de tecnología del servicio de internet (DSL, fibra óptica, etc)
j)Online Security:	El cliente tiene contratado el servicio de seguridad online o no
k)Online Backup:	El cliente tiene contratado el servicio de backup online o no
l)Device Protection:	Si el cliente tiene contratado el servicio de protección de dispositivos
m)Tech Support:	El cliente tiene contratado el servicio de soporte técnico o no
n)Streaming TV:	El cliente tiene contratado el servicio de TV online o no
o)Streaming movies:	El cliente tiene contratado el servicio de películas online o no
p)Contract:	Tipo de contrato seleccionado por el cliente
q)Paperless Billing:	El cliente solicitó factura electrónica o no
r)Payment Method:	Método de pago seleccionado por el cliente
s)Monthly Charges:	Cargos mensuales del cliente
t)Total Charges:	Cargos totales del cliente
u)Churn:	El cliente solicitó la baja de los servicios o no

Figura 6: Diccionario de Datos

Fuente: Elaboración propia

Luego se realiza el análisis estadístico para determinar las variables más representativas y se construye la matriz de correlación.

Figura 7: Matriz de correlación

Fuente: Elaboración propia

	SeniorCitizen	tenure	MonthlyCharges	TotalCharges	Churn	gender_Female	gender_Male	Partner_No	Partner_Yes	Dependents
SeniorCitizen	1.000000	0.015683	0.219874	0.102411	0.150541	0.001819	-0.001819	-0.016957	0.016957	0.2108
tenure	0.015683	1.000000	0.246862	0.825880	-0.354040	-0.005285	0.005285	-0.381912	0.381912	-0.1633
MonthlyCharges	0.219874	0.246862	1.000000	0.051905	-0.190853	0.043270	-0.043270	0.007305	-0.007305	0.1422

```

Index(['Churn', 'Contract_Month-to-month', 'OnlineSecurity_No',
      'TechSupport_No', 'InternetService_Fiber optic',
      'PaymentMethod_Electronic check', 'OnlineBackup_No',
      'DeviceProtection_No', 'MonthlyCharges', 'PaperlessBilling_Yes'],
      dtype='object')

```

Por último se obtienen las variables más representativas para la tasa de abandono.

Figura 8: Listado de variables c/mayor incidencia

Fuente: Elaboración propia

Etapa de Presentación de resultados

En esta etapa se realizan gráficos para la visualización de los resultados del análisis estadístico. Se elabora el mapa de calor, histogramas de cada variable, su distribución estadística y dispersión. Estas tareas se apoyan en librerías Matplotlib y Seaborn [29]. Como ejemplo se muestra la variable de cargos mensuales vs tasa de abandono (churn).

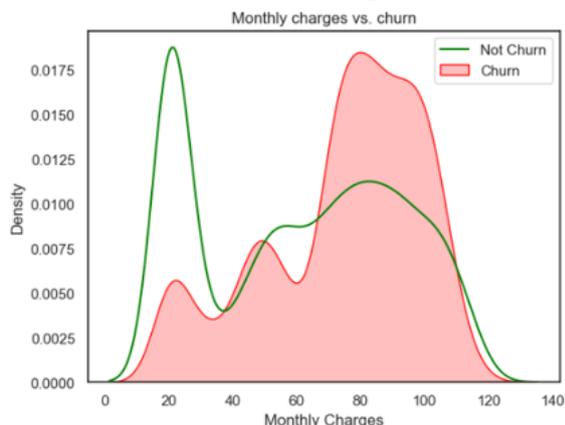


Figura 9: Matriz de correlación

Fuente: Elaboración propia

En el gráfico anterior se observa que los clientes con mayor influencia en tasa de abandono son aquellos que tienen cargos mensuales entre usd 70 y usd 100 mensuales, mientras que los clientes con menor tasa de baja de servicio son quienes tienen una facturación media mensual de usd 20 dólares.

Conclusiones

El proyecto de investigación permitió valorar la experiencia del proceso enseñanza-aprendizaje a través de la participación en un proyecto concreto de articulación academia-industria, que ofreció a los estudiantes una aproximación al contexto productivo.

El trabajo tipificar las variables más representativas en la tasa de abandono de clientes de servicios de telecomunicaciones: tipo de contrato (mensual), soporte técnico (no contratado), método de pago (cheque electrónico) y seguridad en línea (no contratado).

Las tareas de la primera y última instancia del proyecto demandaron del dominio de habilidades blandas. Esto se reflejó en la interpretación del contexto, adecuada integración del equipo de trabajo y la comunicación de resultados. Mientras que el dominio de competencias técnicas en manejo de programación Python y las librerías citadas se evidenció en las etapas intermedias del análisis de datos estrictamente.

Se comprobó además que la utilización de rúbricas analíticas en la evaluación facilita el seguimiento del desarrollo de habilidades y apalanca un mayor compromiso del estudiante.

El lenguaje de programación Python, con una curva de aprendizaje adecuada permitió ajustarse a los tiempos previstos del proyecto. A su vez le dio escalabilidad y dinámica al análisis descriptivo de datos, y define un primer paso en el estudio de ciencia de datos.

Para futuros trabajos se propone continuar con el desarrollo del modelo de madurez analítico, basándose en este análisis descriptivo, avanzando al predictivo con machine learning, mediante librerías Scipy y Scikit-Learn. Adicionalmente se podría aplicar este modelo en otras carreras, distintas a ingeniería, que utilizaran conjuntos de datos vinculados con sus disciplinas específicas, para integrar conocimientos y habilidades particulares de sus ambientes productivos.

Bibliografía

- [1] K. Schwab. *La cuarta revolución industrial*. World Economic Forum. Ed Debate, 2016.
- [2] R. Privdeville. *Prepare for data Revolution*. Data-driven world. Armanino, 2019.
- [3] M. Schwartz. *War & Peace & IT: Business Leadership, Technology and Success in the digital age*. Portland. Ed IT Revolution Press, 2019.
- [4] A. McAfee, & E. Brynjolfsson. *Big Data: The Management Revolution*. Harvard Business Review. 2012.
- [5] Acumen, Research & Consulting. *Global Data Analytic Market*. 2022. Disponible en <https://www.acumenresearchandconsulting.com/data-analytics-market>
- [6] W. Cleveland. *Data science: An action plan for expanding the technical areas of the field of statistics*. The ASA Data Science Journal, 7(6), 414-417. 2014
- [7] J. Tukey. *The future of Data Analysis*. The Collected Works of John Tukey, ed. L. V. Jones, Monterey, CA: Wadsworth & Brooks/Cole, pp. 391–484 1962
- [8] T. Elliot. Gartner BI: *Analytics moves to the core*. Retrieved from Digital Business & Business Analytics: <https://timoelliott.com/blog/2013/02/gartnerbi-emea-2013-part-1-analytics-moves-to-the-core.html>
- [9] W. Jensen. *Statistics=Analytics?*. 2021. *Quality Engineering*. W.L. Gore & Associates, Inc., Flagstaff, Arizona, pag 7
- [10] M. Pratt. *How to structure and manage a data science team*. (2021). Retrieved from Tech-Target. <https://www.techtarget.com/searchbusinessanalytics/feature/How-to-structure-and-manage-a-data-science-team>
- [11] C. O’Neil & R. Schutt. *Doing Data Science: Straight Talk from the Frontline*. O’Reilly Media Inc, California (2013)
- [12] J. Saltz & Shamshurin, I.: Exploring the process of doing data science via an ethnographic study of a media advertising company. In: 2015 IEEE International Conference on Big Data (Big Data), pp. 2098–2105. IEEE (2015)
- [13] Wickham, H., Grolemund, G.: *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O’Reilly Media Inc, California (2016)
- [14] S. Van Daele & G. Jansseswillen. *Identifying the Steps in an Exploratory Data Analysis*. ICPM(international Conference on process Mining (2022) p 526. Ed. Springer
- [15] M. Courtney. *Exploratory Data Analysis in Schools: A Logic Model to Guide Implementation*. (2021) IJEPL (International Journal Education Policy & Leadership. Volume17(4) doi:10.22230/ijepl.2021v17n4a1041

- [16] F. Malvicino y G. Yoguel (2016), "Big Data. Avances recientes a nivel internacional y perspectivas para el desarrollo local. CIECTI (Centro interdisciplinario de estudios en ciencia, tecnología e innovación)
<http://www.ciecti.org.ar/dt3-big-data/>
- [17] V. Garousi. (2016). *Incorporating real-world industrial testing projects in software testing courses: opportunities, challenges, and lessons learned*. En Software Engineering Education and Training (CSEE&T 2011), 24th IEEE-CS Conference on, pp. 396-400. IEEE F.
- [18] P. Rodríguez, N. Palomino y J.Mondaca. *El uso de datos masivos y sus técnicas analíticas para el diseño e implementación de políticas públicas en Latinoamérica y el Caribe*. BID. Sector de Conocimiento y Aprendizaje. 2017
- [19] J.A. Del Pozo Flores. *Competencias profesionales. Herramientas de evaluación: el portafolios, la rúbrica y las pruebas situacionales*. Ed. Narcea. (2017).
- [20] P. Perrenoud. *Diez nuevas competencias para enseñar. Invitación al viaje*. Editorial Graó. 2004
- [21] C. Lévy-Leboyer, 2003, *Gestión de las competencias. Cómo analizarlas, cómo evaluarlas, cómo desarrollarlas*. Ed Gestión
- [22] Career Space. *Perfiles de capacidades profesionales genéricas de TIC*. Capacidades profesionales futuras para el mundo del mañana. Publicaciones Oficiales de las Comunidades Europeas. 2001. www.carrer-space.com
- [23] M. Ilahi, Cheniti-Belcadhi, L., & Braham, R. *Formal competence-based assessment: on closing the gap between academia and industry*. Proceedings of the 2nd International Conference on Technological Ecosystems for Enhancing Multiculturality, pp. 581-587. ACM. 2014
- [24] W. Bel, *Algoritmos y estructuras de datos en Python*. Un enfoque ágil. Facultad Ciencia y Tec. 2020. pg17. Ed Uader.
- [25] Coding Nomads. 2022. *Most In-Demand Programming Languages* Disponible en: <https://codingnomads.co/blog/the-best-programming-languages-to-learn/>
- [26] C. Neil. *Aportes para la implementación de un modelo de formación orientado a competencias*. RIISIC. (Red de Ingeniería en Informática Sistemas de Información del CONFEDI). Ed UAI. 2021
- [27] S. Mukhiya & U. Ahmed. *Hands-On Exploratory Data Analysis with Python: Perform EDA techniques to understand your data*. Ed. Packt 2020.
- [28] W. McKinney. *Python for Data Analysis: Data Wrangling with Pandas & Numpy*. Third Edition. Ed. O'Reilly. 2022
- [29] A. Pajankar. *Hands-on Matplotlib*. Learn Plotting and Visualizations with Python 3. Ed Apress 2021.

TITULO: RUBRICAS ANALITICAS PARA EVALUACION DE COMPETENCIAS TECNICAS ESPECIFICAS (PYTHON Y LIBRERIAS UTILIZADAS EN EL PROYECTO)

CATEGORIA (Aspectos a Evaluar)	VALORACION x etapa del proyecto			NIVEL DE DESEMPEÑO			
	A	B	C	EXCELENTE	SATISFACTORIO	MEJORABLE	INSUFICIENTE
Manejo de tipos de datos en Python				Logra dominar completamente los tipos de datos y sus transformaciones	El manejo de los diferentes tipos de datos y sus transformaciones es aceptable	Necesita ejercitar en el manejo de tipos de datos y sus transformaciones	No logra dominar adecuadamente los distintos tipos de datos
Manejo de métodos y funciones en Python				Logra dominar completamente los distintos tipos de métodos y funciones empleados en el proyecto	El manejo de métodos y funciones de variables resulta aceptable	Necesita ejercitar el manejo de métodos y funciones de variables	Los métodos y funciones se utilizan con criterio aleatorio sin aportar significado
Manejo de estructuras de control en Python				Domina completamente los flujos de control if, while, for, elif y sus andadores	El manejo de estructuras de bucles y loops es aceptable	El manejo de estructuras de control es básico, necesita ejercitarlo	El manejo de las estructuras de control es insuficiente
Manejo de librería NUMPY, análisis numérico de datos en Python				Domina completamente la ejecución de manejo numérico del conjunto de datos	El dominio de librería Numpy para el análisis numérico es aceptable	El manejo de estrategias de indexación, slicing y transformación es básico necesita ser mejorarlo	El manejo de las funciones de la librería es insuficiente
Manejo de librería PANDAS, manipulación y limpieza de datos en Python				Domina completamente las ejecución de tareas de limpieza, ordenamiento y consolidación de datos	La manipulación de la librería Pandas para limpieza de datos es aceptable	El manejo de tareas de limpieza y ordenamiento de datos requiere más ejercitación	No se registra evidencia de la utilización de la librería en el proyecto
Manejo de librería MATPLOTLIB, para visualización estática de datos				Domina completamente la ejecución de comandos para la visualización adecuada de datos y resultados	El dominio de la librería Matplotlib es aceptable	La ejecución de histogramas y boxplots, scatter requiere de asistencia, necesita ejercitar	El análisis resulta insuficiente y/o presenta dificultad de comprensión, no se alcanza ningún objetivo
Ejecución del EDA				El análisis exploratorio realizado es excelente y aporta información relevante.	La ejecución del EDA es aceptable alcanza los objetivos esperados	El análisis exploratorio de datos es mejorable en su ejecución y sus resultados	
OBSERVACIONES				AUTONOMÍA, DECISION, MANEJO 100%	AUTONOMÍA, MANEJO SATISFACTORIO	RESUELVE CON COLABORACION.	REQUIERE ASISTENCIA

Figura 10: Rúbricas analíticas para la evaluación de competencias técnicas específicas

TITULO : RÚBRICAS ANALITICAS PARA EVALUACION DE COMPETENCIAS GENERICAS TRANSVERSALES

CATEGORIA	VALORACION en función de etapa del proyecto			NIVEL DE DESEMPEÑO			
	A	B	C	EXCELENTE	SATISFACTORIO	MEJORABLE	INSUFICIENTE
Gestión de la demanda y manejo de expectativas				Propone y documenta opciones en todos los encuentros. Flexible en la búsqueda del mejor resultado final	Acepta las sugerencias de cambio y maneja alternativas de solución	Interviene en algunas reuniones, no propone opciones ni acepta cambios	No acepta cambios ni interactúa con el solicitante del requerimiento
Pensamiento crítico				Propone respuestas y soluciones adecuadamente fundamentadas	Genera temas e investiga y luego realiza algún análisis con aporte de valor	Sin demasiado cuestionamiento sobre las decisiones, capacidad de análisis mejorable	No se cuestiona ni analiza durante los encuentros
Trabajo en equipo				Siempre participa, se propone a colaborar. Assume sus responsabilidades y da asistencia a otros	Se une a todos los encuentros. La mayoría de las veces hace aportes. A veces se lleva trabajo	Se une en ocasiones. No realiza propuestas.	No participa de los encuentros. No logra el trabajo conjunto
Comunicación de los resultados				Se reúne habitualmente y elabora informes de seguimiento que expone sin dificultad	Genera reuniones de seguimiento, elabora informes y busca consensos	Espacios para seguimiento deficientes, búsqueda de consensos insuficiente	No se define un seguimiento de avance de proyecto
Toma de decisiones				Propone plan de trabajo, asigna tareas y participa en todas las decisiones	Detecta información útil y acepta el nivel de avance del proyecto	Requiere sugerencias y asistencia para definir acciones a seguir	No logra definir acciones en base a la información disponible
Compromiso				Assume la responsabilidad durante la ejecución completa del proyecto	Está alineado con la estrategia y los tiempos del equipo. Entrega sus aportes	Acepta las tareas asignadas	No asume responsabilidad.
Gestión en entornos de incertidumbre				Es capaz de establecer diferentes escenarios y analizar alternativas de solución frente a imponderables	Con la información disponible toma decisiones exponiendo riesgo controlado	Requiere la totalidad de información crítica para definir el camino	No asume riesgos. No toma decisiones sin la información necesaria
OBSERVACIONES				AUTONOMÍA, DECISION, COMPROMISO, COMUNICACIÓN, FLEXIBILIDAD 100%	AUTONOMÍA, MANEJO	REQUIERE APOYO, NECESITA MEJORAR RELACIONAMIENTO Y CONFIANZA	NO ALCANZA EL OBJETIVO

Figura 11: Rúbricas analíticas para la evaluación de competencias genéricas transversales

Data analysis in telecommunication services

Process automation through machine learning techniques

Caiafa Marcelo Dante¹ , Aurelio Ariel¹ , Bevilacqua Alejandro¹ 

¹ *Engineering and Technological Research Department, Universidad Nacional de La Matanza, Argentina*
mcaiafa@unlam.edu.ar; aurelio@unlam.edu.ar; abevilacqua@unlam.edu.ar

Abstract

The data analysis has become a main source of knowledge for organizations. Its adequate treatment allows to obtain valuable information.

The professional link between university and industry is the goal of our academic work. It is intended to register an articulation project based on data analysis.

The work is based on the use of Python and its libraries, for the data processing of a telecommunication services institution.

The research addresses the need to consolidate results that add value to decision-making processes.

The project is based on an exploratory, correlation and segmentation study to build a customer profile with high dropout potential.

It is intended to contribute to the promotion of activities that integrate the academy with the industry, reflecting the value that the application of new tools is capable of contributing to students and teachers in the ICT (information and communication technology) sector.

Keywords: TIC, Data analysis, telecommunications, churn.

Resumen

El análisis de datos se ha convertido en una fuente principal de conocimiento para las organizaciones. Su adecuado tratamiento permite obtener valiosa información.

El vínculo profesional entre la universidad y la industria es el objetivo de nuestro trabajo académico. Se pretende registrar un proyecto de articulación basado en el análisis de datos.

El trabajo se basa en la utilización de Python y sus librerías, para el procesamiento de datos de una institución de servicios de telecomunicaciones.

La investigación aborda la necesidad de consolidar resultados que agreguen valor a los procesos de toma de decisiones.

El proyecto se basa en un estudio exploratorio, de correlación y segmentación para construir un perfil de cliente con alto potencial de abandono.

Se pretende contribuir a la promoción de actividades que integren la academia con la industria, reflejando el valor que la aplicación de nuevas herramientas es capaz de aportar a estudiantes y docentes en el sector TIC (tecnologías de la información y la comunicación).

Palabras claves: TIC, análisis de datos, telecomunicaciones, tasa de abandono de servicios.

1. The value of data analysis

The digital transformation is generating in the last decades profound changes in the way of producing and marketing goods and services [1]. In this context, information management allows organizations to improve their processes, "data is the new currency that sustains fundamental changes in the fourth industrial revolution" [2].

A data-oriented organizational culture has advantages in decision-making because it is based on evidence [3]. Some studies [4] indicate that organizations with a data-driven approach improve their productivity and profitability. According to international consultants [5], the size of the global data analytics market represented USD 31.8 billion by 2021 and they estimate it will reach USD 329.8 billion in 2030. That would register a

compound annual growth rate (CAGR) of 29.9% in that period.

Data analysis can be defined as a multi-disciplinary process that manages data with the purpose of generating information to improve decision making [6].

Tukey defined data analysis as "the procedures and techniques for interpreting results, ways of planning data collection to make data analysis easier and more accurate" [7].



Figure 1: Data Analytics Market
Source: Acumen R&C, 2021

Different types of analytics are currently classified:

- a) Descriptive analytics: It illustrates outcome data collected over a time interval.
- b) Diagnostic analytics: it looks for the root cause of a problem.
- c) Predictive analytics: it uses past data to make forecasts.
- d) Prescriptive analytics: it involves machine learning techniques to provide recommendations for specific problems.

Although there are different analytical maturity models, the following chart lists the different types of analysis. These can be considered as correlative stages of the same project, which as the level of complexity progresses increases the value of the contribution [8].

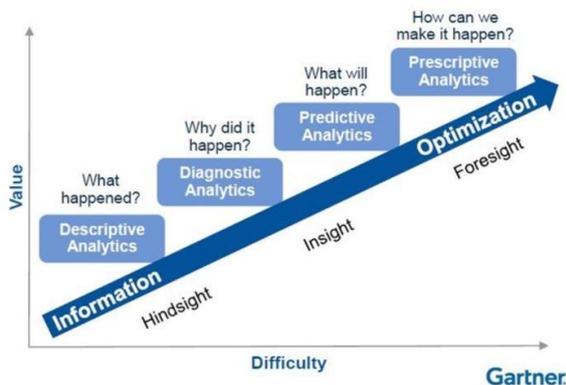


Figure 2: Analytical Maturity Model

Source: Gartner, 2012.

Although there is not complete consensus on the model, it is useful associate it with different types of questions to help focus analytical efforts. [9]

2. Scope of Analysis

The scope of this project is involved within the different stages of the analytical maturity model. It is a descriptive and prospective analysis that generates information from the study of a framework of real and specific historical data.

In order to identify the skills to be developed in the project, we start from the generic definitions of responsibilities of the different roles performed in a work team for this type of project. Some authors [10] indicate that among the profiles required by the labor market are:

- Data Analyst, runs analysis statistics from various sources of information.
- Data scientist, apply mathematics, statistics, and programming to explore, examine, and graph data.
- Data manager (Chief Data Officer), is the ultimate responsible executive for all data management and its privacy policies.
- Data engineer, is responsible to test, implement and maintain the data infrastructure and its repositories.
- Data architect, is the one who defines, designs and oversees the implementation of the underlying systems used.
- Data translator, is an emerging role that links with the business units, assessing needs, requirements validation and communicating results.

The aim is to add value to the training of professionals working in the ICT sector, highlighting the required skills to achieve a successful and systematic process of data exploration.

3. Project Design

The study is based on the articulation of the academic-professional world. It focuses on a project of exploratory analysis of data from telecommunications services. In the planning of the project, the recommendations of different authors [11], [12] are taken as a reference. Although a cited author [13] advances in detail for the R language, the present work was carried out in Python's libraries.

To facilitate the project's development, the data analysis was grouped into three main stages [14].

Data Collection Stage: the ADC (automated data capture) eliminates the manual efforts required for data entry. It was implemented by scraping process on CRM internal web page using BeautifulSoup.

Data Treatment Stage: In this instance the Pandas and Numpy libraries were used for data cleaning, identification of variables, construction of a data dictionary, analysis of relationships between variables, statistical description of data.

Presentation Stage: The information was visualized by Matplotlib and Seaborn libraries. They were so useful for summary of observations and results presentation.

Finally Scikit-learn library was used for modeling of prediction with machine learning techniques.

A typical characteristic of EDA (Exploratory Data Analysis) is that it focuses more on the process than on the theory. A handful of entry-level analysis techniques (descriptive statistics, correlations, and basic visualizations) can be learned [15]. This combined with the learning curve of a high level interpretive language like Python was appropriated for this project.

4. Problem to investigate

As a specific case study, the historical data of a set of telecommunications service customers, duly organized and treated as a block, is considered to define a customer

profile with a high probability of service cancellation request.

The research question that guides the work is how to develop a data analysis project that links engineer students to the telecommunications services sector ?. In this framework, the objectives of the study are:

1. Record the process to identify a typical customer profile with the highest dropout rate, based on the most representative variables detected in a data set of telecommunications services.
2. Build methodological proposal for EDA project with forecasting capabilities.

It is intended to add value to the professional who works in the ICT sector (computer technology and communications). These types of projects are not only about the implementation of technological infrastructure but due to their nature and potential impact, they are strategic to support decision-making based on evidence [16].

5. Development

This research project is structured in five instances executed sequentially. In the first instance, the tasks of selecting the software tool and the preparation of the rubrics were carried out. In the second instance, which corresponds to the data collection stage, the context is studied, the requirements are validated, and the indicators are defined. The third instance corresponds to data processing. With the data obtained, the variables are identified, a data dictionary is made, they are cleaned, ordered and the statistical analysis is executed.

In the stage number four, the operations related to the presentation of results are executed. The correlation matrix is built to identify the relevant variables and they are validated with the heat map. The graphs for the presentation of the reports are also elaborated.

The last instance is for the final analysis of the project, the elaboration of conclusions and proposals for future work.

Selection of the software tool had initially different options. R, ELK, powerBI

were analyzed. The use of Python 3.9.12 release as a programming language was defined as being the most appropriate due to its characteristics. Among others we have an interpreted, high-level, dynamically typed, multiplatform, open source, object-oriented language, its large community and extensive libraries, among others [17].

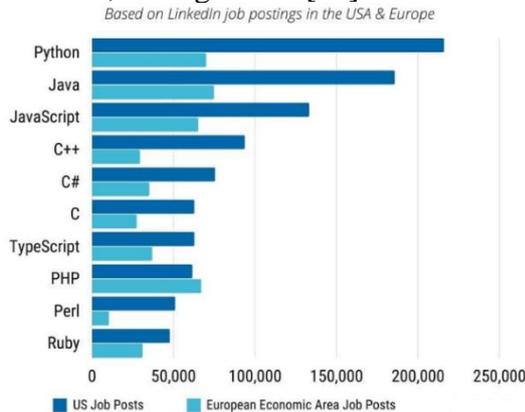


Figure 3: Most required programming languages
Source: LinkedIn, 2022.

Another of the reasons that led to the decision to use Python as a language is its great demand in the job offers in the USA and the EU [18].

5.1. Data collection

In this instance, the context of the problem to be solved and its impact were studied. In other words, they inquired about the value that the sought solution would provide. This meant studying the relationship between customer lifecycle and churn rate. The costs related to retaining an existing client vs. the costs of attracting a new one and the different loyalty and cross-selling strategies were also studied in depth.

```
import requests
import pandas as pd
from bs4 import BeautifulSoup
from urllib3.exceptions import InsecureRequestWarning
from urllib3 import disable_warnings #soluciona problemas con max ssl certifi
disable_warnings(InsecureRequestWarning)
dias = ['01feb', '02feb', '03feb', '04feb', '05feb', '06feb', .....]
def obtener_data(dia):
    web = f"https://intranet/wiki/{dia}CustomerData" #se enmascara nombre real
    response = requests.get(web, verify=False)
    contenido = response.text
    soup = BeautifulSoup(contenido, "lxml") # lxml será el parser a utilizar
    datas = soup.find_all("div", class_ = "detailsbox")
    MonthlyCharges = []
    Churn = []
    .....
    for data in datas:
        MonthlyCharges.append(data.find("th", class_ = "MonthlyCharges").get_text())
        Churn.append(data.find("th", class_ = "Churn").get_text())
    .....
    diccionario_data = {"MonthlyCharges": MonthlyCharges, "Churn": Churn, .....}
    df_data = pd.DataFrame(diccionario_data)
    df_data["dia"] = dia
    return df_data
#genero una lista en base a un bucle for para coleccionar la data historica
telco = [obtener_data(year) for dia in dias]
#genero el df concatenando todos los registros de la lista
df_telco = pd.concat (telco, ignore_index=True)
#genero un csv con el df llamado dataframe
ruta_csv = "\\2023\\UNLAM\\Investig_2023\\dataframe.csv"
df_telco.to_csv(path_or_buf=ruta_csv, index=False)
```

Figure 4: web scraping process

Data were collected by web scraping process on the intranet web page which is fed by CRM (customer relationship manager) [20]. This case study is based in historical data with churn rate 26%, a significantly high number as it exceeds the industry average.

5.2. Data Processing

Various manipulations of the data set are performed here. First, the data type is consolidated for each variable. Missing data and outliers are detected, ordered and consolidated for subsequent statistical analysis, and categorical variables are mapped to numerical variables to facilitate their processing. Null and inconsistent data are removed. These tasks are executed with Pandas and Numpy libraries, according to good practices [21].

Variable	Descripción
a)Customer ID:	Identificación interna del cliente
b)Gender:	Tipo de género
c)Senior Citizen:	Si el cliente es un adulto
d)Partner:	El cliente tiene pareja o no
e)Dependents:	El cliente tiene dependientes o no
f)Tenure:	Número de meses de antigüedad del cliente en la compañía
g)Phone Service:	El cliente tiene contratado el servicio de telefonía o no
h)Multiple lines:	El cliente tiene contratadas múltiples líneas o no
i)Internet Service:	Tipo de tecnología del servicio de internet (DSL, fibra óptica, etc)
j)Online Security:	El cliente tiene contratado el servicio de seguridad online o no
k)Online Backup:	El cliente tiene contratado el servicio de backup online o no
l)Device Protection:	Si el cliente tiene contratado el servicio de protección de dispositivos
m)Tech Support:	El cliente tiene contratado el servicio de soporte técnico o no
n)Streaming TV:	El cliente tiene contratado el servicio de TV online o no
o)Streaming movies:	El cliente tiene contratado el servicio de películas online o no
p)Contract:	Tipo de contrato seleccionado por el cliente
q)Paperless Billing:	El cliente solicitó factura electrónica o no
r)Payment Method:	Método de pago seleccionado por el cliente
s)Monthly Charges:	Cargos mensuales del cliente
t)Total Charges:	Cargos totales del cliente
u)Churn:	El cliente solicitó la baja de los servicios o no

Figura 6: Data dictionary

The variables are ordered, some are renamed and others are redefined based on the indicators. The data dictionary is created.

5.3. Data Exploration

The statistical analysis is carried out to determine the most representative variables and the correlation matrix is built.

	SeniorCitizen	tenure	MonthlyCharges	TotalCharges	Churn	gender_Female	gender_Male	Partner_No	Partner_Yes	Dependents
SeniorCitizen	1.00000	0.015683	0.219874	0.102411	0.150541	0.001819	-0.001819	-0.016957	0.016957	0.2102
tenure	0.015683	1.00000	0.246892	0.825880	-0.354049	-0.005285	0.005285	-0.381912	0.381912	-0.1633
MonthlyCharges	0.219874	0.246892	1.00000	0.651065	0.192858	0.013779	-0.013779	-0.097825	0.097825	0.1122
TotalCharges	0.102411	0.825880	0.651065	1.00000	-0.199484	-0.000048	0.000048	-0.319072	0.319072	-0.064
Churn	0.150541	-0.354049	0.192858	-0.199484	1.00000	0.008545	-0.008545	0.149892	-0.149892	0.1633
gender_Female	0.001819	-0.005285	0.013779	-0.000048	0.008545	1.00000	-1.00000	-0.001379	0.001379	0.0102
gender_Male	-0.001819	0.005285	-0.013779	0.000048	-0.008545	-1.00000	1.00000	0.001379	-0.001379	-0.0102
Partner_No	-0.016957	-0.381912	-0.097825	-0.319072	0.149892	-0.001379	0.001379	1.00000	-1.00000	0.4522

Fig. 7: Correlation Matrix

Finally, the most representative variables for the dropout rate are obtained. A partial vision of heat map is shown below:

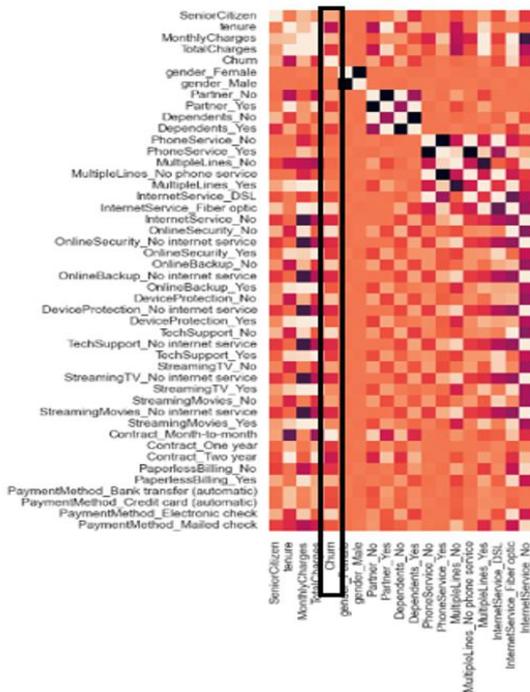


Fig. 8: Heat map

5.4. Data Visualization

In this stage, the graphs are made to visualize the results of the statistical analysis. The heat map, histograms of each variable, its statistical distribution and dispersion are elaborated. These tasks are supported by Matplotlib and Seaborn libraries [22]. As an example, the variable of monthly charges vs. abandonment rate (churn) is shown.

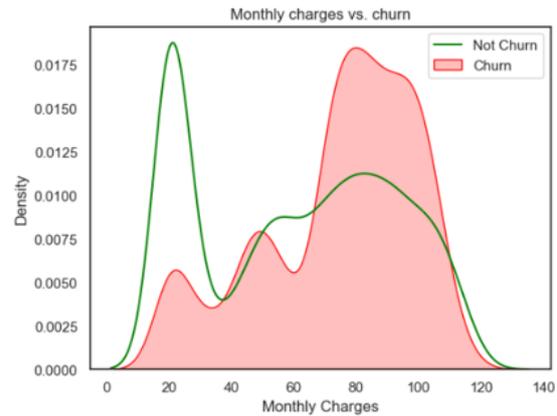


Fig. 9: MonthlyCharges vs Churn

5.5. Data Modelling

The ET

Three instances are identified. The first is the input. It consists of indicating the file path to extract the data. The next step is the filter where all fields are listed in comma separated format. In the particular case of dates, the data received in UNIX format can be converted to a data type. Finally, the output indicates the name of the index where the data will be loaded.

6. Results

The final parameter.

Fig. 2 Outgoing telephone traffic distribution

The filter was appli

Fig. 3 Customer care level dashboard

Two grap

7. Conclusions

The measurements reflected in levels.

Comparing the calls

.

Competing interests

The authors have declared that no competing interests exist.

Authors' contribution

"CM conceived the idea and conducted the experiments; CM, AA and BA analyzed the results and revised the manuscript. All authors read and approved the final manuscript."

8. References

- [1] K. Schwab. *La cuarta revolución industrial*. World Economic Forum. Ed Debate, 2016.
- [2] R. Privdeville. *Prepare for data Revolution*. Data-driven world. Armanino, 2019.
- [3] M. Schwartz. *War & Peace & IT: Business Leadership, Technology and Success in the digital age*. Portland. Ed IT Revolution Press, 2019.
- [4] A. McAfee, & E. Brynjolfsson. *Big Data: The Management Revolution*. Harvard Business Review. 2012.
- [5] Acumen, Research & Consulting. *Global Data Analytic Market*. 2022. Disponible en <https://www.acumenresearchandconsulting.com/data-analytics-market>
- [6] W. Cleveland. *Data science: An action plan for expanding the technical areas of the field of statistics*. The ASA Data Science Journal, 7(6), 414-417. 2014
- [7] J. Tukey. *The future of Data Analysis*. The Collected Works of John Tukey, ed. L. V. Jones, Monterey, CA: Wadsworth & Brooks/Cole, pp. 391-484 1962
- [8] T. Elliot. Gartner BI: *Analytics moves to the core*. Retrieved from Digital Business & Business Analytics: <https://timoelliott.com/blog/2013/02/gartnerbi-emea-2013-part-1-analytics-moves-to-the-core.html>
- [9] W. Jensen. *Statistics=Analytics?*. 2021. *Quality Engineering*. W.L. Gore & Associates, Inc., Flagstaff, Arizona, pag 7
- [10] M. Pratt. *How to structure and manage a data science team*. (2021). Retrieved from TechTarget. <https://www.techtarget.com/searchbusinessanalytics/feature/How-to-structure-and-manage-a-data-science-team>
- [11] C. O'Neil & R. Schutt. *Doing Data Science: Straight Talk from the Frontline*. O'Reilly Media Inc, California (2013)
- [12] J. Saltz & Shamshurin, I.: Exploring the process of doing data science via an ethnographic study of a media advertising company. In: 2015 IEEE International Conference on Big Data (Big Data), pp. 2098-2105. IEEE (2015)
- [13] Wickham, H., Golemund, G.: *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly Media Inc, California (2016)
- [14] S. Van Daele & G. Jansseswillen. *Identifying the Steps in an Exploratory Data Analysis*. ICPM(international Conference on process Mining (2022) p 526. Ed. Springer
- [15] M. Courtney. *Exploratory Data Analysis in Schools: A Logic Model to Guide Implementation*. (2021) IJEPL (International Journal Education Policy & Leadership. Volume17(4) doi:10.22230/ijepl.2021v17n4a1041
- [16] P. Rodríguez, N. Palomino y J.Mondaca. *El uso de datos masivos y sus técnicas analíticas para el diseño e implementación de políticas públicas en Latinoamérica y el Caribe*. BID. Sector de Conocimiento y Aprendizaje. 2017
- [17] W. Bel, *Algoritmos y estructuras de datos en Python*. Un enfoque ágil. Facultad Ciencia y Tec. 2020. pg17. Ed Uader
- [17] Coding Nomads. 2022. *Most In-Demand Programming Languages* Disponible en: <https://codingnomads.co/blog/the-best-programming-languages-to-learn/>
- [18] P. Rodríguez, N. Palomino y J.Mondaca. *El uso de datos masivos y sus técnicas analíticas para el diseño e implementación de políticas públicas en Latinoamérica y el Caribe*. BID. Sector de Conocimiento y Aprendizaje. 2017
- [19] S. Mukhiya & U. Ahmed. *Hands-On Exploratory Data Analysis with Python: Perform EDA techniques to understand your data*. Ed. Packt 2020.
- [20] W. McKinney. *Python for Data Analysis: Data Wrangling with Pandas & Numpy*. Third Edition. Ed. O'Reilly. 2022
- [21] A. Pajankar. *Hands-on Matplotlib*. Learn Plotting and Visualizations with Python 3.Ed Apress 2021.



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

Presentación a JAIIO

Data analysis for telecommunication services

Marcelo Dante Caiafa¹ , Ariel Aurelio¹ , Alejandro Bevilacqua¹ 

¹ Universidad Nacional de La Matanza, Florencio Varela 1903, Buenos Aires, Argentina
{macaiafa, aaurelio, abevilacqua}@unlam.edu.ar

Abstract. The data has become a main source of knowledge for organizations. Its adequate treatment allows to obtain valuable information. The professional link between university and industry is the goal of our academic work. It is intended to register an articulation project based on data analysis. It is an initiative based on a collaboration between a company and the academy. The aim is using the knowledge and skills of the students to improve the efficiency of the company by automating a process and at the same time bring computer engineering students closer to the productive environment.

The work is based on Python and its libraries for data processing of a telecommunication services. The project is based on an exploratory, correlation and segmentation study of data processing for telecommunication services. The goal is to build a customer profile with high dropout potential.

The work's intention is to add value to decision makers in a data-based organization and verify the contribution that new tools can offer to students and teachers in the ICT sector (information and communication technologies).

Keywords: Data analysis, ICT, telecommunications, churn.

1 Introduction

The digital transformation is generating in the last decades big changes in the way of producing and consuming goods and services [1]. In this context, the information management allows organizations to improve their processes, "data is the new currency that sustains fundamental changes in the fourth industrial revolution" [2].

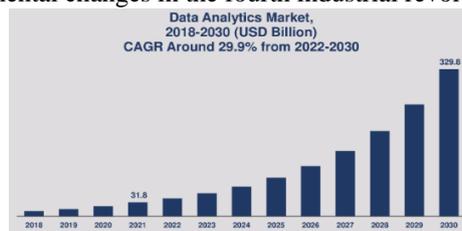


Fig. 1. Data Analytics Market (Acumen R&C 2021)

A data-oriented organizational culture has advantages in decision-making because it is based on evidence [3][16]. Some studies [4] indicate organizations with a data-driven approach improve their productivity and profitability. According to international consultants [5], the size of the global data analytics market represented USD 31.8 billion by 2021 and they estimate it will reach USD 329.8 billion in 2030. It would register a compound annual growth rate (CAGR) of 29.9% in that period.

Different types of analytics are classified in:

- Descriptive analytics: It illustrates outcome data collected over a time interval.
- Diagnostic analytics: it looks for the root cause of a problem.
- Predictive analytics: it uses past data to make forecasts.
- Prescriptive analytics: it involves machine learning techniques to forecasting.

Although there are different analytical maturity models, the following chart lists different types of analysis. These can be considered as correlative stages of the same project, which as the level of complexity progresses increases the contribution value [6].



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

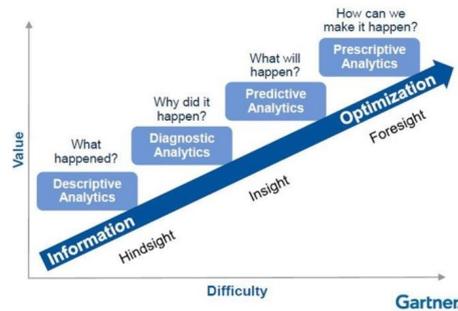


Fig. 2. Analytical Maturity Model (Gartner 2012)

The study is based on the articulation of the academic-professional world. It focuses on a project of exploratory analysis of data from telecommunications services. In the planning of the project, the recommendations of different authors [7], [8] are taken as a reference. Although a cited author [9] advances in detail for the R language, the present work was carried out in Python's libraries. To facilitate the project's development, the data analysis was grouped into three main stages [10].

Data Collection Stage: the automated data capture eliminates the manual efforts required for data entry. It was implemented by scraping process on CRM internal web site using BeautifulSoup.

Data Treatment Stage: In this instance the Pandas and Numpy libraries were used for data cleaning, identification of variables, construction of a data dictionary, analysis of relationships between variables, statistical description of data.

Data Presentation Stage: The information was visualized by Matplotlib and Seaborn libraries. They were so useful for summary of observations and results presentation.

A typical characteristic of EDA (Exploratory Data Analysis) is that it focuses more on the process than on the theory. A handful of entry-level analysis techniques (descriptive statistics, correlations, and basic visualizations) can be learned [11]. This kind of projects is not only about the implementation of technological infrastructure but due to their potential impact, they are strategic to support decision-making based on evidence.

The research question that guides the work is how to develop data analysis project that links engineer students to the telecommunications service sector. In this framework, the objectives are:

1. To document the process to identify a typical customer profile with the highest dropout rate, based on the most representative variables detected.
2. To build methodological proposal for EDA project with forecasting capabilities.

2 Work Development

This research project is structured in five instances executed sequentially. In the first instance, the tasks of selecting the software tool were carried out. In the second instance, the context and requirements are validated. The data collection is processed by scraping and the indicators are defined. The third instance is about data processing. They are cleaned, ordered and the statistical analysis is executed. In the stage number four, the operations related to presentation of results are executed. The correlation matrix is built. The relevant variables are identified and validated with heat map. The graphs for presentation reports are elaborated. The last instance is the elaboration of conclusions and new proposals for future work.

Software selection had initially many options. R, ELK, powerBI were analyzed. Python 3.9.12 release as a programming language was defined as being the most appropriate due to its features: high-level language, dynamically typed, multi-platform, open source, object-oriented language, its big community and extensive libraries, among others [12].

2.1 Data Collection

In this instance, the context of the problem to be solved and its impact were studied. In other words, they inquired about the value that the sought solution would provide. This meant studying the relationship between customer



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

lifecycle and churn rate. The costs related to retaining an existing client vs. the costs of attracting a new one and the different loyalty and cross-selling strategies were also studied deeply.

```
import requests
import pandas as pd
from bs4 import BeautifulSoup
from urllib3.exceptions import InsecureRequestWarning
from urllib3 import disable_warnings #soluciona problemas con max ssl certif
disable_warnings(InsecureRequestWarning)
dias = [03012023,03022023,03032023, 03042023, 03052023, 03062023, .....]
def obtener_data(dia):
    web = f"https://intranet/wiki/{dia}CustomerData" #se enmascara nombre real
    response = requests.get(web, verify=False)
    contenido = response.text
    soup = BeautifulSoup(contenido, "lxml") # lxml será el parser a utilizar
    datas = soup.find_all("div", class_ = "detailsbox")
    MonthlyCharges = []
    Churn = []
    .....
    for data in datas:
        MonthlyCharges.append(data.find("th", class_ = "MonthlyCharges").get_text())
        Churn.append(data.find("th", class_ = "Churn").get_text())
    .....
    diccionario_data = {"MonthlyCharges": MonthlyCharges, "Churn": Churn, .....}
    df_data = pd.DataFrame(diccionario_data)
    df_data["dia"] = dia
    return df_data
#genero una lista en base a un bucle for para coleccionar la data historica
telco = [obtener_data(year) for dia in dias]
#genero el df concatenando todos los registros de la lista
df_telco = pd.concat (telco, ignore_index=True)
#genero un csv con el df llamado dataframe
ruta_csv = "\\2023\\UNLAM\\Investig_2023\\dataframe.csv"
df_telco.to_csv(path_or_buf=ruta_csv, index=False)
```

Fig. 3. Scrapping Process

Data were collected by web scraping process on the intranet web page which is fed daily by CRM server (customer relationship manager) [13].

2.2 Data Processing

Various manipulations of the data set are performed here. First, the data type is consolidated for each variable. Missing data and outliers are detected, ordered and consolidated for subsequent statistical analysis, and categorical variables are mapped to numerical variables to facilitate their processing. Null and inconsistent data are removed. These tasks are executed with Pandas and Numpy libraries, according to good practices [14]. The variables are ordered, some of them are renamed and others are redefined based on the indicators. This case study is based in historical data with churn rate 26%.

2.3 Data Exploration

Statistical analysis is carried out to determine the most representative variables, correlation matrix is built and most relevant variables for churn prediction are identified.

	Churn	CargosMensuales	Permanencia	CargosTotales	Adultos	TipodeContrato_Anual	TipodeContrato_Bianual	TipodeContrato_Mensi
Churn	1.000000	0.192858	-0.354049	-0.199484	0.150541	-0.178225	-0.301552	0.4045
CargosMensuales	0.192858	1.000000	0.246862	0.651065	0.219874	0.004810	-0.073256	0.0589
Permanencia	-0.354049	0.246862	1.000000	0.825880	0.015683	0.202338	0.563801	-0.6493
CargosTotales	-0.199484	0.651065	0.825880	1.000000	0.102411	0.170569	0.358036	-0.4467
Adultos	0.150541	0.219874	0.015683	0.102411	1.000000	-0.046491	-0.116205	0.1377
TipodeContrato_Anual	-0.178225	0.004810	0.202338	0.170569	-0.046491	1.000000	-0.288843	-0.5700
TipodeContrato_Bianual	-0.301552	-0.073256	0.563801	0.358036	-0.116205	-0.288843	1.000000	-0.6219
TipodeContrato_Mensual	0.404565	0.058933	-0.649346	-0.446776	0.137752	-0.570053	-0.621933	1.0000

Fig. 4. Correlation Matrix

Finally, the most representative variables for the dropout rate are validated by heat map which is shown.



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

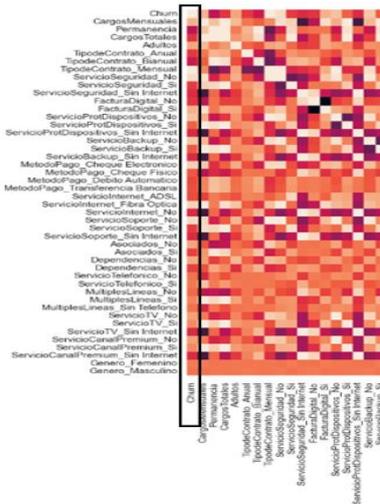


Fig. 5. Partial view of Heat map

3 Results

It is observed the clients with highest unsubscribe requests are those with highest monthly charges and lower than average tenure.

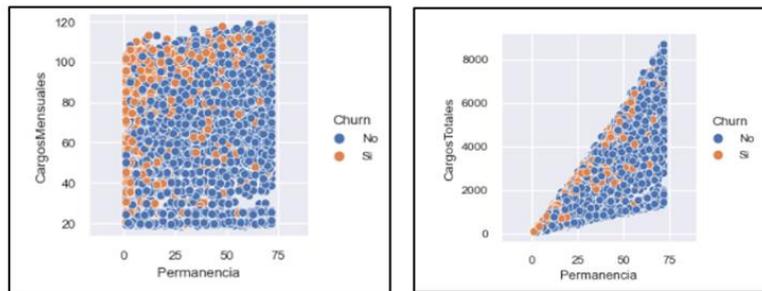


Fig.6. Interrelación de variables (cargos mensuales y totales vs permanencia)

4 Conclusions

The most representative variables in the telecommunications service customer abandonment rate were typified: type of contract (monthly), technical support (not contracted), payment method (electronic check) and online security (not contracted).

It was possible to detect and ponder the most relevant causes that give early signs of a possible change contractual conditions. This allows preventive, constant and continuous analysis to avoid loss of assets such as the client portfolio. It gave scalability and dynamics to descriptive data analysis, defines it a first step in the data science study.

For future work, it is proposed to continue with the development of the analytical maturity model, based on this descriptive analysis, advancing to the predictive one with machine learning, through SciPy libraries. Additionally, this model could be applied in other careers, other than engineering, that use data sets linked to their specific disciplines, to integrate particular knowledge and skills from their productive environments.

Competing interests

The authors have declared that no competing interests exist.

Authors' contribution

"MC conceived the idea and conducted the experiments; MC, AA and AB analyzed the results and revised the manuscript.



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

All authors read and approved the final manuscript."

5 References

- [1] S Schwab. La cuarta revolución industrial. World Economic Forum. Ed Debate, 2016.
- [2] R. Privdeville. Prepare for data Revolution. Data-driven world. Armanino, 2019.
- [3] M. Schwartz. War & Peace & IT: Business Leadership, Technology and Success in the digital age. Portland. Ed IT Revolution Press, 2019.
- [4] A. McAfee, & Brynjolfsson. The Management Revolution. Harvard Business Review. 2012.
- [5] Acumen, Research & Consulting. Global Data Analytic Market. 2022. Disponible en <https://www.acumenresearchandconsulting.com/data-analytics-market>
- [6] W. Jensen. Statistics=Analytics?. 2021. Quality Engineering, Inc., Flagstaff, Arizona, p 7
- [7] C, O'Neil & R. Schutt. Doing Data Science: Straight Talk from the Frontline. O'Reilly Media Inc, California (2013)
- [8] J. Saltz & Shamshurin, I.: Exploring the process of doing data science via an ethnographic study of a media advertising company. In: 2015 IEEE International Conference on Big Data pp. 2098–2105. IEEE (2015)
- [9] Wickham, H., Golemund, G.: R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. O'Reilly Media Inc, California (2016)
- [10] S. Van Daele & G. Jansseswillen. Identifying the Steps in an Exploratory Data Analysis. ICPM(international Conference on process Mining (2022) p 526. Ed. Springer
- [11] M. Courtney. Exploratory Data Analysis in Schools: A Logic Model to Guide Implementation. (2021) IJEPL (International Journal Education Policy & Leadership. Volume17(4)
- [12] W. Bel. Algoritmos y estructuras de datos en Python. Facultad C y Tec. 2020. p17. Ed Uader
- [13] S. Mukhiya & U. Ahmed. Hands-On Exploratory Data Analysis with Python: Perform EDA techniques to understand your data. Ed. Packt 2020.
- [14] W. McKinney. Python for Data Analysis: Data Wrangling with Pandas & Numpy. Third Edition. Ed. O'Reilly. 2022
- [15] A. Pajankar. Hands-on Matplotlib. Learn Plotting and Visualizations with Python 3. Ed Apress 2021.

Presentación a RedVITEC

Convocatoria para la presentación de trabajos

VII Jornadas académicas de la RedVITEC

"Vinculación tecnológica en las universidades: desarrollo y transformación territorial"

6 y 7 de diciembre de 2023

Universidad Nacional del Litoral. Santa Fe, Argentina.

Consejo Interuniversitario Nacional, diciembre de 2023

Anexo formato presentación de las actividades de vinculación tecnológica

(Completar uno por cada caso presentado)

UNIVERSIDAD/ES	Universidad Nacional de La Matanza
Nombre	Marcelo Dante Caiafa
mail de contacto	mcaiafa@unlam.edu.ar
sitio web	www.unlam.edu.ar



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

Redes sociales	https://www.linkedin.com/in/marcecaiafa/
----------------	---

AUTORES:	
Autor 1	Marcelo Dante Caiafa
Autor 2	Ariel Rodrigo Aurelio
Autor 3	Alejandro Bevilacqua
Correo electrónico (completar por cada autor):	mcaiafa@unlam.edu.ar aaurelio@unlam.edu.ar abevilacqua@unlam.edu.ar

PRESENTACIÓN NÚ- MERO:	
-----------------------------------	--

(Hasta 3 presentaciones por universidad)

TÍ- TULO:	Implementación de Análisis de Datos en Servicios de Telecomunicaciones
----------------------	--

(70 caracteres con espacios):

SUBTÍ- TULO:	Proyecto evaluado con rúbricas analíticas
-------------------------	---

(100 caracteres con espacios):

ÁREA TEMÁTICA	
----------------------	--

SECTOR VINCULADO (marcar la opción que corresponda):

Sector privado	Nombre - mail de contacto - web
Empresas	
Pymes	X
Emprendimien- tos	
Cooperativas	
Otras	

Instituciones Privadas sin Fines de Lucro	Nombre- mail de contacto - web



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

Administración Pública	Nombre- mail de contacto - web
Nacional	
Provincial	
Municipal o comunal	

DESCRIPCIÓN DEL CASO (máximo 3500 caracteres con espacios en total)

Descripción del trabajo:

Objetivo del trabajo: El objetivo general de la investigación apunta a la ejecución de un análisis de datos que permita identificar los clientes propensos al abandono del servicio de telecomunicaciones a partir de la recopilación de datos históricos y el desarrollo de las competencias profesionales puestas en práctica por estudiantes de ingeniería en su implementación. Los objetivos específicos del estudio son:

1. Colectar, limpiar e interpretar datos históricos de clientes de servicios de telecomunicaciones.
2. Caracterizar los clientes que abandonan y los que permanecen con el servicio.
3. Determinar las variables de mayor incidencia en el grupo de clientes que cancela el servicio
4. Elaborar las rúbricas analíticas para la evaluación de competencias profesionales en análisis de datos.

Se pretende aportar valor a la formación del perfil del profesional que trabaja en el sector TIC (tecnología informática y comunicaciones) destacando las habilidades necesarias para lograr un eficiente desempeño en la exploración de datos. Este tipo de proyectos no tratan sólo de implementación de infraestructura tecnológica sino que, por su naturaleza e impacto son estratégicos para soportar la toma de decisiones basada en evidencia.

Contextualización del trabajo: Como caso de estudio concreto se consideran los datos históricos de un conjunto de clientes de servicios de telecomunicaciones, debidamente organizados y tratados en bloque, para definir un perfil de cliente con alta probabilidad de solicitud de baja de servicio.

De allí que las preguntas de investigación que guían el trabajo son:

- ¿Cómo desarrollar un proyecto de análisis de datos de servicios de telecomunicaciones con estudiantes de ingeniería que los vincule con la industria?
- ¿Cuáles son las competencias profesionales, habilidades técnicas y genéricas, que requiere un ingeniero que se desempeña en proyectos de análisis de datos y cómo pueden ser evaluadas?

La tasa de abandono de servicio, también conocida por la palabra en inglés churn, se utiliza en la industria como un indicador para validar el modelo de negocio. Un valor inferior a la media representa un servicio que responde favorablemente a las necesidades del cliente, mientras que un valor mayor al promedio indica insatisfacción.



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

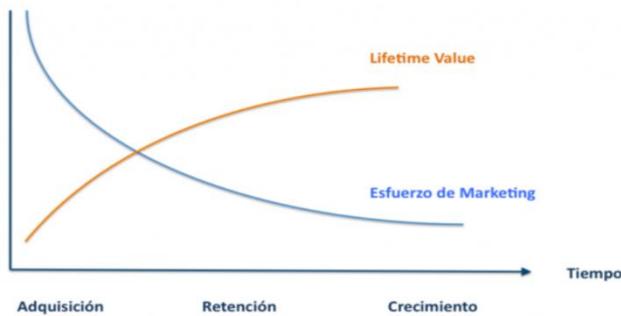


Figura. Valor de vida del cliente

Fuente: Bejou, D., (2006).

La figura anterior representa la evolución del valor de la vida del cliente y la inversión de marketing en función del tiempo. Esto permite distinguir tres etapas de adquisición, retención y crecimiento de clientes. El valor de la vida del cliente crece a medida que el tiempo transcurre y se logra evitar el abandono. La tasa de cancelación de servicio tiene impacto directo en la disminución de la facturación, la reducción de la rentabilidad y la pérdida de ingresos por ventas cruzadas, entre otros.

Los esfuerzos de marketing en la reducción de la tasa de cancelación resultan en inversiones en programas de fidelización, en mejoras en la calidad del servicio, en la atención al cliente, etc.

Según algunos autores [24], el costo de adquirir un nuevo cliente es como mínimo cinco veces más alto que el de evitar su cancelación en la etapa de adquisición, aunque esto varía en función del tipo de industria, el tipo de servicio, los precios relativos de mercado, etc.

Tipo de vinculación: En los casos en que corresponda mencionar brevemente el modo en que se estableció el contacto entre la universidad y los demás actores (*iniciativa de la universidad, de la contraparte o de un tercero; conocimiento previo a nivel de unidad ejecutora, o institucional con la unidad académica o con la universidad, etc.; vinculación y/o antecedentes de trabajo preexistente*)

Descripción del proyecto:

El desarrollo del proyecto de investigación se estructura en cinco instancias ejecutadas en forma secuencial.

En la primera instancia se realizaron las tareas de selección de la herramienta de software y la elaboración de las rúbricas.

En la segunda instancia, que corresponde a la etapa de relevamiento de datos, se estudia el contexto, se validan los requerimientos, y se definen los indicadores.

La tercera instancia corresponde al tratamiento de los datos. Con los datos obtenidos se identifican las variables, se realiza un diccionario de datos, se limpian, se ordenan y se ejecuta el análisis estadístico.

En la instancia cuatro se ejecutan las operaciones vinculadas con la presentación de resultados. Se construye la matriz de correlación para identificar las variables relevantes y son validadas con el mapa de calor. También se elaboran los gráficos para la presentación de los informes.



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

La última instancia es para el análisis final del proyecto la elaboración de conclusiones y propuestas de futuros trabajos.

Selección de la herramienta de software

Inicialmente se analizaron diversas opciones como R, ELK, PowerBI. Se definió la utilización de Python como lenguaje de programación por ser la más adecuada por sus características. Entre otras tenemos un lenguaje interpretado, de alto nivel, de tipado dinámico, multi-plataforma, de código abierto, orientado a objeto, su gran comunidad, diversidad de librerías, entre otras [25].:

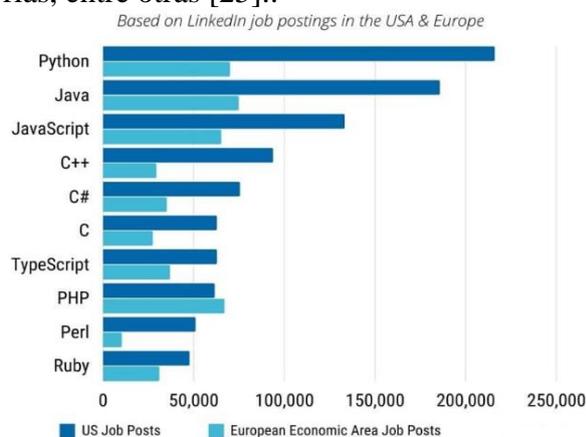


Figura 3. Lenguajes de programación más requeridos
Fuente: LinkedIn, 2022.

Otra de las razones que motivó la decisión de utilizar Python como lenguaje es su gran demanda en las ofertas de empleo en USA y EU que registra LinkedIn y se muestra en el gráfico anterior [26].

En esta primera instancia, a partir de las recomendaciones del CONFEDI (Consejo Federal de Decanos de Ingeniería) [27] y otras, se confeccionaron las rúbricas para el seguimiento y evaluación del desarrollo de catorce competencias de parte de los estudiantes durante el proceso. Se adjuntan en hoja separada.

Etapa de relevamiento de datos

En esta instancia se estudió el contexto de la problemática a resolver y su impacto. Es decir que se indagó sobre el valor que aportaría la solución buscada. Esto significó estudiar la relación entre el ciclo de vida del cliente y la tasa de abandono. Además se profundizó sobre los costos relacionados con la retención de un cliente existente vs los costos de captación de uno nuevo, distintas estrategias de fidelización y venta cruzada, etc.

Se colectaron los datos y se procedió a un primer acercamiento en un análisis preliminar según las mejores prácticas [28].

Etapa de tratamiento de datos

Aquí se realizan distintas manipulaciones del conjunto de datos. Primero se consolida el tipo de dato por cada variable. Se detectan datos faltantes y outliers, se ordenan y consolidan para un posterior análisis estadístico y se mapean variables categóricas a variables numéricas para facilitar su procesamiento.



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

Luego se eliminan datos nulos e inconsistentes. Estas tareas se ejecutan con librería Pandas y Numpy, según las buenas prácticas [29].

La base de los datos históricos sobre los que se trabajó responde a la siguiente estructura: donde los datos del usuario quedan representados a partir de las variables listadas

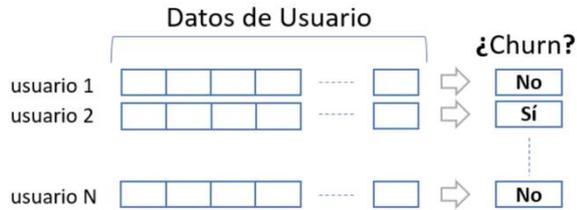


Figura. Estructura del conjunto de datos

Se ordenan las variables, se renombran algunas y otras se redefinen a partir de los indicadores. Se elabora el diccionario de datos del usuario que quedan representados a partir de diecinueve variables.

Luego se realiza distinto tratamiento a las variables categóricas respecto de las numéricas. Para determinar las variables más representativas y se construye la matriz de correlación graficada como un mapa de calor.

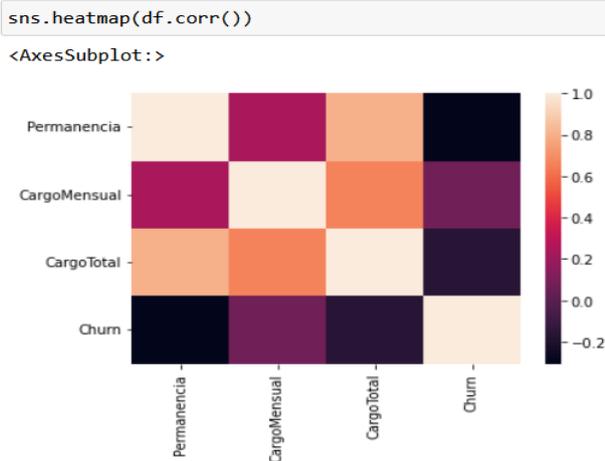


Figura. Matrix de correlación

Para facilitar su interpretación se grafica a continuación la tabla de correlación donde se distingue la relación negativa respecto de la permanencia.

	Permanencia	CargoMensual	CargoTotal	Churn
Permanencia	1.000	0.239	0.804	-0.306
CargoMensual	0.239	1.000	0.660	0.067
CargoTotal	0.804	0.660	1.000	-0.164
Churn	-0.306	0.067	-0.164	1.000

Figura. Tabla de correlación
Fuente: Elaboración propia

A las variables categóricas se les da un tratamiento de gráfico de puntos para analizar los datos en forma descriptiva con la intención de determinar relaciones notorias entre ellas. Para ello se toman las variables de cargomensual y permanencia.



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

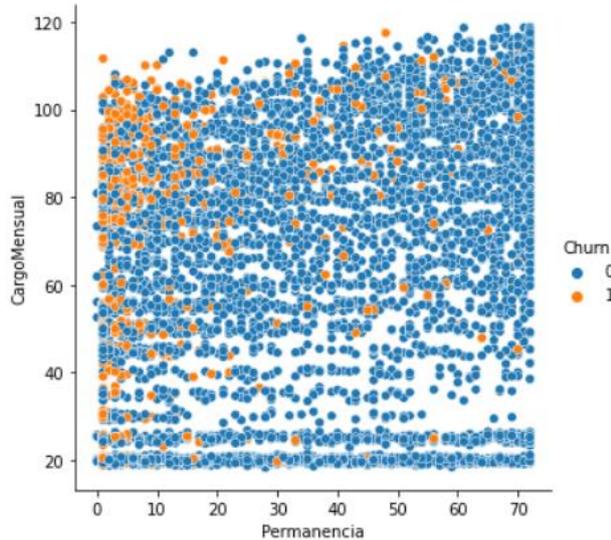


Figura. Lenguajes de programación más requeridos
Fuente: Elaboración propia.

En la figura anterior se observa una tendencia clara de concentración de bajas de servicio a menor permanencia del cliente y para mayores cargos mensuales.

Luego se las compara a partir de las características por el tipo de cliente clasificado por género, por edad, por tipo de servicio contratado (telefonía, internet y soporte), por tipo de pago y tipo de contrato. Todos estos se adjuntan en hoja separada al final del trabajo.

Resultados e impacto:

En esta etapa se realizan gráficos para la visualización de los resultados del análisis estadístico. Se elabora el mapa de calor, histogramas de cada variable, su distribución estadística y dispersión. Estas tareas se apoyan en



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

librerías Matplotlib y Seaborn [30]. Como ejemplo se muestra la variable de cargos mensuales vs tasa de abandono (churn).

Distribución de CHURN

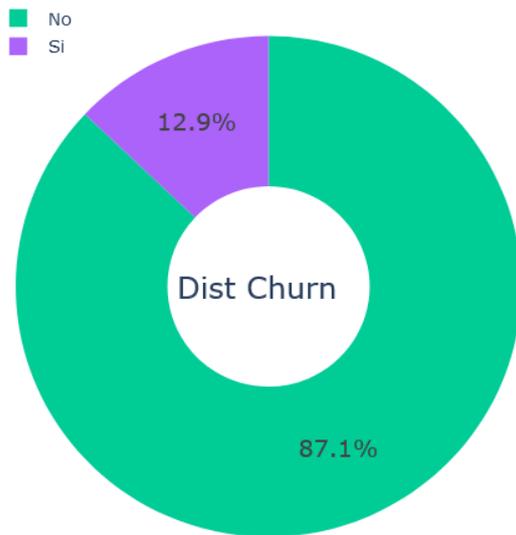


Figura. Cálculo de la tasa de abandono
Fuente: Elaboración propia.

La tasa de cancelación se calcula como el cociente entre el valor total de clientes que solicitaron la cancelación de servicio para un determinado período de tiempo respecto del total de clientes suscriptos al inicio del período especificado. En este caso se trató del primer trimestre del 2023.

El proyecto de investigación permitió valorar la experiencia del proceso enseñanza-aprendizaje a través de la participación en un proyecto concreto de articulación academia-industria, que ofreció a los estudiantes una aproximación al contexto productivo.

El valor de la tasa de abandono del caso es considerablemente superior a la media de la industria por lo que se recomienda avanzar en el estudio en la instancia de predicción utilizando tecnología de aprendizaje supervisado de máquina.

El trabajo tipificar las variables más representativas en la tasa de abandono de clientes de servicios de telecomunicaciones: permanencia, los cargos mensuales, el tipo de contrato (mensual) y el soporte técnico (no contratado).

Las tareas de la primera y última instancia del proyecto demandaron del dominio de habilidades blandas. Esto se reflejó en la interpretación del contexto, adecuada integración del equipo de trabajo y la comunicación de resultados. Mientras que el dominio de competencias técnicas en manejo de programación Python y las librerías citadas se evidenció en las etapas intermedias del análisis de datos estrictamente.

Se comprobó además que la utilización de rúbricas analíticas en la evaluación facilita el seguimiento del desarrollo de habilidades y apalanca un mayor compromiso del estudiante.

El lenguaje de programación Python, con una curva de aprendizaje adecuada permitió ajustarse a los tiempos previstos del proyecto. A su vez le dio escalabilidad y dinámica al análisis descriptivo de datos, y define un primer paso en el estudio de ciencia de datos.

Palabras clave: análisis de datos, TIC, rúbricas



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLAM
Versión	5
Vigencia	03/9/2019

Ejes transversales:

El resultado final pretende aportar un caso concreto de vinculación académica evaluado a partir de la utilización de rúbricas analíticas. Una rúbrica es un instrumento cuya principal finalidad es compartir los criterios de realización de las tareas de aprendizaje y de evaluación con los estudiantes y entre el profesorado. La rúbrica, como guía u hoja de ruta de las tareas, muestra las expectativas que estudiantes y docentes comparten sobre distintas actividades, organizadas en diferentes niveles de desempeño [19].

La evaluación del proceso de enseñanza-aprendizaje requiere, además de la asimilación de conocimientos, el desarrollo de habilidades necesarias para resolver problemas en contextos reales. Según algunos autores [20], una habilidad es la capacidad de utilizar recursos necesarios para responder efectivamente a una situación compleja en un contexto específico. Si bien existen distintos enfoques, se siguen las definiciones de algunos autores que clasifican las habilidades en dos enfoques: tipo técnico (conocimientos) o interpersonales (actitudes sociales) [21], así el perfil del ingeniero está conformado por competencias técnicas y competencias genéricas [22].

Las competencias técnicas específicas suelen ser referenciadas como habilidades duras. Son las requeridas para el desempeño de una ocupación en concreto. Están relacionadas con funciones o puestos de trabajo en particular. Aportan al estudiante los conocimientos, actitudes, habilidades y valores propios de cada profesión y actividad laboral.

Por otro lado, están las habilidades genéricas también llamadas habilidades blandas. En ingeniería de software se pueden identificar a partir de un modelo semántico que especifica el marco de trabajo para su evaluación. Este entorno de tareas se basa en el hecho que el estudiante tiene un perfil de habilidades por desarrollar y para cada una define un conjunto de hitos a alcanzar [23].

DATOS DE LA UNIDAD EJECUTORA O EL EQUIPO DE TRABAJO:

Denominación:	DIIT (Departamento de Ingeniería e Investigaciones Tecnológicas)
Responsable institucional:	Marcelo Dante Caiafa
Teléfono:	1140295936
Web:	www.unlam.edu.ar
Correo electrónico:	mcaiafa@unlam.edu.ar

A fin de facilitar y el trabajo de producción general y con el objetivo de reducir los plazos para la publicación, se solicita respetar las siguientes especificaciones técnicas:

Función	Cantidad	Modo de presentación
----------------	-----------------	-----------------------------



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

Título	Máximo 70 caracteres con espacios Puede ser un título de fantasía que haga alusión simbólicamente al proyecto. <i>Ej: Alimentando los sueños</i>	Textos en Word. Calibri. Cuerpo 11, interlineado simple. Alineados a izquierda.
Subtítulos	Máximo 100 caracteres. Describe objetivamente el proyecto, sobre todo si se utilizó un título simbólico/poético.	
Descripción del caso	Máximo 3500 caracteres con espacios en total	
Imágenes	Hasta 6 imágenes (entre fotos y gráficos) y hasta 2 imágenes Las imágenes deberán ser de libre disponibilidad o autorizadas por sus titulares. (las imágenes que ilustren la publicación serán seleccionadas de entre las enviadas en función de su calidad fotográfica y comunicacional)	Fotografías o ilustraciones: JPG. PNG. EPS. Incluidas en el documento y además en archivos separados. Definición 300 dpi en tamaño 15 x 20 cm Tablas y gráficos en Excel o en programa vectorial (Corel Draw, Adobe Illustrator, PDF editable). Incluidas en el documento y además en archivos separados.

Referencias

Se enumeran a continuación las referencias bibliográficas.

- [16][1] K. Schwab. La cuarta revolución industrial. World Economic Forum. Ed Debate, 2016.
- [17][2] R. Privdeville. Prepare for data Revolution. Data-driven world. Armanino, 2019.
- [18][3] M. Schwartz. War & Peace & IT: Business Leadership, Technology and Success in the digital age. Portland. Ed IT Revolution Press, 2019.
- [19][4] A. McAfee, & E. Brynjolfsson. Big Data: The Management Revolution. Harvard Business Review. 2012.
- [20][5] Acumen, Research & Consulting. Global Data Analytic Market. 2022. Disponible en <https://www.acumenresearchandconsulting.com/data-analytics-market>
- [21][6] W.Cleveland. Data science: An action plan for expanding the technical areas of the field of statistics. The ASA Data Science Journal, 7(6), 414-417. 2014
- [22][7] J. Tukey. The future of Data Analysis. The Collected Works of John Tukey, ed. L. V. Jones, Monterey, CA: Wadsworth & Brooks/Cole, pp. 391-484 1962
- [23][8] T. Elliot. Gartner BI: Analytics moves to the core. Retrieved from Digital Business & Business Analytics: <https://timoelliott.com/blog/2013/02/gartnerbi-emea-2013-part-1-analytics-moves-to-the-core.html>
- [24][9] W. Jensen. Statistics=Analytics?. 2021.Quality Engineering. W.L. Gore & Associates, Inc., Flagstaff, Arizona, pag 7
- [25][10] M. Pratt. How to structure and manage a data science team. (2021). Retrieved from TechTarget. <https://www.techtarget.com/searchbusinessanalytics/feature/How-to-structure-and-manage-a-data-science-team>
- [26][11] C, O'Neil & R. Schutt. Doing Data Science: Straight Talk from the Frontline. O'Reilly Media Inc, California (2013)
- [27][12] J. Saltz & Shamshurin, I.: Exploring the process of doing data science via an ethnographic study of a media advertising company. In: 2015 IEEE International Conference on Big Data (Big Data), pp. 2098-2105. IEEE (2015)
- [28][13] Wickham, H., Golemund, G.: R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. O'Reilly Media Inc, California (2016)
- [29][14] S. Van Daele & G. Jansseswillen. Identifying the Steps in an Exploratory Data Analysis. ICPM(international Conference on process Mining (2022) p 526. Ed. Springer
- [30][15] M. Courtney. Exploratory Data Analysis in Schools: A Logic Model to Guide Implementation. (2021) IJEPL (International Journal Education Policy & Leadership. Volume17(4) doi:10.22230/ijepl.2021v17n4a1041
- [31][16] F. Malvicino y G. Yoguel (2016), "Big Data. Avances recientes a nivel internacional y perspectivas para el desarrollo local. CIECTI (Centro interdisciplinario de estudios en ciencia, tecnología e innovación). <http://www.ciecti.org.ar/dt3-big-data/>
- [32][17] V. Garousi. (2016). Incorporating real-world industrial testing projects in software testing courses: opportunities, challenges, and lessons learned. En Software Engineering Education and Training (CSEE&T 2011), 24th IEEE-CS Conference on, pp. 396-400. IEEE F.



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

- [33][18] P. Rodríguez, N. Palomino y J.Mondaca. El uso de datos masivos y sus técnicas analíticas para el diseño e implementación de políticas públicas en Latinoamérica y el Caribe. BID. Sector de Conocimiento y Aprendizaje. 2017
- [34][19] J.A. Del Pozo Flores. Competencias profesionales. Herramientas de evaluación: el portafolios, la rúbrica y las pruebas situacionales. Ed. Narcea. (2017).
- [35][20] P. Perrenoud. Diez nuevas competencias para enseñar. Invitación al viaje. Editorial Graó. 2004
- [36][21] C. Lévy-Leboyer, 2003, Gestión de las competencias. Cómo analizarlas, cómo evaluarlas, cómo desarrollarlas. Ed Gestión
- [37][22] Career Space. Perfiles de capacidades profesionales genéricas de TIC. Capacidades profesionales futuras para el mundo del mañana. Publicaciones Oficiales de las Comunidades Europeas. 2001. www.carrer-space.com
- [38][23] M. Ilahi, Cheniti-Belcadhi, L., & Braham, R. Formal competence-based assessment: on closing the gap between academia and industry. Proceedings of the 2nd International Conference on Technological Ecosystems for Enhancing Multiculturality, pp. 581-587. ACM. 2014
- [39][24] Frederick Reichheld, "Winning on purpose", HBR, 2021
- [40][25] W. Bel, Algoritmos y estructuras de datos en Python. Un enfoque ágil. Facultad Ciencia y Tec. 2020. pg17. Ed Uader.
- [41][26] Coding Nomads. 2022. Most In-Demand Programming Languages Disponible en: <https://codingnomads.co/blog/the-best-programming-languages-to-learn/>
- [42][27] C. Neil. Aportes para la implementación de un modelo de formación orientado a competencias. RIISIC. (Red de Ingeniería en Informática Sistemas de Información del CONFEDI). Ed UAI. 2021
- [43][28] S. Mukhiya & U. Ahmed. Hands-On Exploratory Data Analysis with Python: Perform EDA techniques to understand your data. Ed. Packt 2020.
- [44][29] W. McKinney. Python for Data Analysis: Data Wrangling with Pandas & Numpy. Third Edition. Ed. O'Reilly. 2022
- [45][30] A. Pajankar. Hands-on Matplotlib. Learn Plotting and Visualizations with Python 3.Ed Apress 2021.



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

CATEGORÍA (Aspectos a Evaluar)	VALORACION x etapa del proyecto			NIVEL DE DESEMPEÑO			
	A	B	C	EXCELENTE	SATISFACTORIO	MEJORABLE	INSUFICIENTE
Manejo de tipos de datos en Python				Logra dominar completamente los tipos de datos y sus transformaciones	El manejo de los diferentes tipos de datos y sus transformaciones es aceptable	Necesita ejercitar en el manejo de tipos de datos y sus transformaciones	No logra dominar adecuadamente los distintos tipos de datos
Manejo de métodos y funciones en Python				Logra dominar completamente los distintos tipos de métodos y funciones empleados en el proyecto	El manejo de métodos y funciones de variables resulta aceptable	Necesita ejercitar el manejo de métodos y funciones de variables	Los métodos y funciones se utilizan con criterio aleatorio sin aportar significado
Manejo de estructuras de control en Python				Domina completamente los flujos de control if, while, for, elif y sus anidaciones	El manejo de estructuras de bucles y loops es aceptable	El manejo de estructuras de control es básico, necesita ejercitarlo	El manejo de las estructuras de control es insuficiente
Manejo de librería NUMPY, análisis numérico de datos en Python				Domina completamente la ejecución de manejo numérico del conjunto de datos	El dominio de librería Numpy para el análisis numérico es aceptable	El manejo de estrategias de indexación, slicing y transformación es básico necesita ser mejorado	El manejo de las funciones de la librería es insuficiente
Manejo de librería PANDAS, manipulación y limpieza de datos en Python				Domina completamente las ejecución de tareas de limpieza, ordenamiento y consolidación de datos	La manipulación de la librería Pandas para limpieza de datos es aceptable	El manejo de tareas de limpieza y ordenamiento de datos requiere más ejercitación	No demuestra ejercicio de la librería Pandas
Manejo de librería MATPLOTLIB, para visualización estática de datos				Domina completamente la ejecución de comandos para la visualización adecuada de datos y resultados	El dominio de la librería Matplotlib es aceptable	La ejecución de histogramas y boxplots, scatter requiere de asistencia, necesita ejercitar	No se registra evidencia de la utilización de la librería en el proyecto
Ejecución del EDA				El análisis exploratorio realizado es excelente y aporta información relevante.	La ejecución del EDA es aceptable alcanza los objetivos esperados	El análisis exploratorio de datos es mejorable en su ejecución y sus resultados	El análisis resulta insuficiente y/o presenta dificultad de comprensión, no se alcanza ningún objetivo
OBSERVACIONES				AUTONOMÍA, DECISION, MANEJO 100%	AUTONOMÍA, MANEJO SATISFACTORIO	RESUELVE CON COLABORACION.	REQUIERE ASISTENCIA

TÍTULO: RUBRICAS ANALITICAS PARA EVALUACION DE COMPETENCIAS TECNICAS ESPECIFICAS (PYTHON Y LIBRERIAS UTILIZADAS EN EL PROYECTO)

Figura 10: Rúbricas analíticas para la evaluación de competencias técnicas específicas



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

TITULO: RÚBRICAS ANALÍTICAS PARA EVALUACION DE COMPETENCIAS GENERICAS TRANSVERSALES

CATEGORÍA	VALORACION en función de etapa del proyecto			NIVEL DE DESEMPEÑO			
	A	B	C	EXCELENTE	SATISFACTORIO	MEJORABLE	INSUFICIENTE
Gestión de la demanda y manejo de expectativas				Propone y documenta opciones en todos los encuentros. Flexible en la búsqueda del mejor resultado final	Acepta las sugerencias de cambio y maneja alternativas de solución	Interviene en algunas reuniones, no propone opciones ni acepta cambios	No acepta cambios ni interactúa con el solicitante del requerimiento
Pensamiento crítico				Propone respuestas y soluciones adecuadamente fundamentadas	Genera temas e investiga y luego realiza algún análisis con aporte de valor	Sin demasiado cuestionamiento sobre las decisiones, capacidad de análisis mejorable	No se cuestiona ni analiza durante los encuentros
Trabajo en equipo				Siempre participa, se propone a colaborar. Asume sus responsabilidades y da asistencia a otros	Se une a todos los encuentros. La mayoría de las veces hace aportes. A veces se lleva trabajo	Se une en ocasiones. No realiza propuestas.	No participa de los encuentros. No logra el trabajo conjunto
Comunicación de los resultados				Se reúne habitualmente y elabora informes de seguimiento que expone sin dificultad	Genera reuniones de seguimiento, elabora informes y busca consensos	Espacios para seguimiento deficientes, búsqueda de consensos insuficiente	No se define un seguimiento de avance de proyecto
Toma de decisiones				Propone plan de trabajo, asigna tareas y participa en todas las decisiones	Detecta información útil y acepta el nivel de avance del proyecto	Requiere sugerencias y asistencia para definir acciones a seguir	No logra definir acciones en base a la información disponible
Compromiso				Asume la responsabilidad durante la ejecución completa del proyecto	Está alineado con la estrategia y los tiempos del equipo. Entrega sus aportes	Acepta las tareas asignadas	No asume responsabilidad.
Gestión en entornos de incertidumbre				Es capaz de establecer diferentes escenarios y analizar alternativas de solución frente a imponderables	Con la información disponible toma decisiones exponiendo riesgo controlado	Requiere la totalidad de información crítica para definir el camino	No asume riesgos. No toma decisiones sin la información necesaria
OBSERVACIONES				AUTONOMIA, DECISION, COMPROMISO, COMUNICACION, FLEXIBILIDAD 100%	AUTONOMIA, MANEJO	REQUIERE APOYO, NECESITA MEJORAR RELACIONAMIENTO Y CONFIANZA	NO ALCANZA EL OBJETIVO

Figura 11: Rúbricas analíticas para evaluación de competencias genéricas transversales



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

ANEXO II

Unidad Académica: DIIT

Código: C2-ING-092

Título del Proyecto: Automatización de servicios de telecomunicaciones con Machine Learning

Director del Proyecto: Caiafa Marcelo Dante

Programa de acreditación: PROINCE.... CyTMA2: X..

Fecha de inicio:01/01/2022

Fecha de finalización: 31/12/2023

1. Datos del alumno

Apellido y Nombre: Baner, Deborah Yanina

DNI:34148315

Unidad Académica: DIIT

Carrera que cursa:ingeniería electrónica

Período evaluado:2023

2. Dictamen de evaluación de desempeño del alumno:

Colocar una cruz donde corresponda

2.1 Satisfactorio: X

2.1 No satisfactorio:

Fundamentos del dictamen:

El desempeño de la becaria resulta satisfactorio dado su compromiso puesto de manifiesto en la esmerada realización de las tareas asignadas a lo largo de los dos años de duración del proyecto. Se destaca además su participación activa en todas las reuniones de equipo para el seguimiento del proyecto, y su iniciativa en la proposición de distintas alternativas para la ejecución de los scripts de automatización en la captura de datos.

Su valora además su contribución en la elaboración de las presentaciones a los diferentes congresos en los que se presentó el trabajo de investigación a saber,

XVIII Congreso TEyET 2023 realizado en la Universidad de Hurlingham

XI jornadas del congreso JCC&DB-ET 2023 realizado en la Universidad de La Plata

XI Jornadas de Cloud Computing-Big Data and Emerging Topics realizado en la Universidad nacional de La Plata

52 Jornadas Argentinas de ingeniería informática realizadas en la Universidad de Tres de Febrero

VII Jornadas académicas de la RedVITEC Vinculación tecnológica en las universidades



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

Se adjuntan los certificados correspondientes



Se certifica que *Marcelo Caiafa* ha participado en calidad de expositor (modalidad virtual) del artículo *Data Analysis for Telecommunications services* en las XI Jornadas de Cloud Computing, Big Data & Emerging Topics desarrolladas del 27 al 29 de junio de 2023 en la ciudad de La Plata (Argentina).


Dr. Marcelo Naiouf
Coordinador Comité Científico



CiN RedVITEC
Red de Vinculación Tecnológica
de las Universidades
Nacionales de Argentina

Por cuanto se certifica que *Caiafa, Marcelo Dante*, ha participado en calidad de **expositor/a**, en las **VII Jornadas Académicas de la RedVITEC**, con el trabajo *"Implementación de análisis de datos en servicios de Telecomunicaciones"*.

Las mismas se desarrollaron del 6 al 7 de diciembre de 2023 y estuvieron organizadas por la RedVITEC y la Universidad Nacional del Litoral.

Santa Fe, diciembre de dos mil veintitrés.

Dr. Javier Lottersberger
Secretario de Vinculación Tecnológica e Innovación - UNL
Coordinador ejecutivo de la RedVITEC

Dr. Enrique José Mammarella
Rector Universidad Nacional del Litoral
Presidente de la Comisión de Vinculación Tecnológica del CiN





Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019



XVIII CONGRESO DE
**TECNOLOGÍA EN EDUCACIÓN &
EDUCACIÓN EN TECNOLOGÍA**

**15 y 16 de junio
de 2023**

Universidad Nacional de Hurlingham
Tte. Orígone 151, Villa Tesei,
provincia de Buenos Aires



Educación
pública, gratuita
y de calidad



Por cuanto

• **Marcelo Dante Caiafa (UNLaM)**

ha participado como autor del trabajo
"Proyecto de análisis de datos de servicios de telecomunicaciones evaluado
con rúbricas analíticas."
en el XVIII CONGRESO DE TECNOLOGÍA EN EDUCACIÓN & EDUCACIÓN EN
TECNOLOGÍA, organizado por la Universidad Nacional
de Hurlingham, los días 15 y 16 de junio de 2023,
se le otorga el presente certificado.

Hurlingham, 16 de junio de 2023

Mg. Walter Wallach
Vicerrector (rector en ejercicio)
UNAHUR

Lic. Patricia Pesado
Coordinadora
Red UNCI



Código	FPI-009
Objeto	Guía de elaboración de Informe de avance de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019



DECLARACIÓN DE PUBLICACIÓN

Atena Editora, especializada en la publicación de libros, revistas internacionales y colecciones de artículos científicos en todas las áreas del conocimiento, con sede en la ciudad de Ponta Grossa-PR, declara que después de una evaluación por pares de doble ciego, por miembros de nuestro Consejo Editorial, el capítulo titulado "Big Data Technology for monitoring ICT service data", de autoría de "Marcelo Dante Caiafa, Aurelio Ariel, Busto Adrian Marcelo," fue aprobado y publicado en la revista "International Journal of Human Sciences Research (ISSN 2764-0558)", sob ISSN 2764-0558 y DOI 10.22533/at.ed.5583102331034.

Agradezco su elección por Atena Editora como medio de difusión del trabajo al público científico y académico, y felicitamos a los autores por aceptar la publicación.

Reitero votos de la más alta estima y consideración.

Ponta Grossa, 12 de mayo de 2023.

Prof.ª Antonella Carvalho de Oliveira
Doutora em ensino de ciência e tecnologia
Editora chefe
ATENA EDITORA
PREFIXO EDITORIAL ISBN 97865
PREFIXO EDITORIAL DOI 10.22533