

Tipo de documento: Tesis de maestría

Escuela de Negocios. Master in Management + Analytics

Modelo Predictivo de machine learning para la Anticipación y Optimización de Contrataciones en el Departamento de Postventa de una Comercializadora de Analizadores de Laboratorio Clínico

Autoría: González Nuñez, Eduardo A.

Año: 2024

¿Cómo citar este trabajo?

González Nuñez, E. (2024). "Modelo Predictivo de machine learning para la Anticipación y Optimización de Contrataciones en el Departamento de Postventa de una Comercializadora de Analizadores de Laboratorio Clínico". [Tesis de maestría. Universidad Torcuato Di Tella]. Repositorio Digital Universidad Torcuato Di Tella. <https://repositorio.utdt.edu/handle/20.500.13098/12937>

El presente documento se encuentra alojado en el Repositorio Digital de la Universidad Torcuato Di Tella bajo una licencia Creative Commons Atribución-No comercial-Sin derivadas 4.0 Internacional
Dirección: <https://repositorio.utdt.edu>



Master in Management + Analytics

“Modelo Predictivo de machine learning para la Anticipación y Optimización de Contrataciones en el Departamento de Postventa de una Comercializadora de Analizadores de Laboratorio Clínico”

TESIS

Eduardo A. González Núñez

Mayo 2024

Tutor: Javier Marengo

"A mi querida Sofía, gracias por ser la musa y la calma en el caos de este viaje académico. Tu presencia ha sido mi refugio y este logro es tan tuyo como mío."

"A Nancy y Benjamín, mis queridos padres, que me enseñaron que el conocimiento es la llave de la vida y cuyo amor y sacrificio me han permitido perseguir mis sueños."

Resumen

En el entorno competitivo de comercialización de analizadores de laboratorio clínico, GEMATEC SRL se enfrenta a retos significativos en su departamento de postventa, derivados del crecimiento constante de su base instalada. Esta expansión resulta en una demanda creciente de servicios de soporte técnico, un desafío exacerbado por la monopolización del conocimiento técnico necesario por los fabricantes de estos equipos. Dada la dificultad para encontrar técnicos con experiencia, un nuevo empleado en GEMATEC SRL requiere aproximadamente ocho meses de capacitación para alcanzar la autonomía en sus funciones.

Esta tesis propone abordar estas dificultades mediante el desarrollo de un modelo de predicción de demanda utilizando técnicas de machine learning, que permitirá anticipar las necesidades futuras de servicio de manera precisa. Paralelamente, se implementará un modelo de programación lineal entera para optimizar los tiempos de contratación de nuevos técnicos, asegurando que el departamento de postventa pueda responder de manera eficiente y efectiva a los requerimientos de servicio.

Con estos modelos, se espera no solo mejorar la planificación y gestión de recursos humanos en el departamento de postventa, sino también incrementar la satisfacción del cliente mediante respuestas más rápidas y efectivas. Esta integración de machine learning y optimización matemática proporcionará a GEMATEC SRL una herramienta estratégica para enfrentar los desafíos del mercado actual y mejorar continuamente su rendimiento operativo.

Abstract

In the competitive market of clinical laboratory analyzers, GEMATEC SRL faces significant challenges in its after-sales department due to the continuous growth of its installed base. This increase, a direct result of the company's successful business activities, has led to an increasing demand for technical support services. The demand for service represents a challenge for GEMATEC SRL because the technical knowledge required to support clinical analyzers is largely monopolized by the manufacturers of these technologies, making it very difficult to find experienced technicians in the labor market. For GEMATEC SRL in particular, a technician entering the company takes about eight months to be sufficiently trained to work autonomously.

This thesis proposes to address these difficulties by developing a predictive demand model using machine learning techniques, which will allow for precise anticipation of future service needs. Concurrently, an integer linear programming model will be implemented to optimize the timing of new technical staff hires, ensuring that the after-sales department can efficiently and effectively respond to service requirements.

With these models, it is expected not only to improve planning and human resource management in the after-sales department but also to enhance customer satisfaction through quicker and more effective responses. The integration of machine learning and mathematical optimization will provide GEMATEC SRL with a strategic tool to face the challenges of the current market and continually improve its operational performance.

Índice de Contenidos

CAPITULO I – INTRODUCCION	6
1.1. Contexto.....	6
1.2. Comprensión y predicción de la demanda futura mediante machine learning 8	8
1.3. Planificación Estratégica de Contratación Basada en la Predicción de la Demanda.....	8
1.4. Impacto de la tesis.....	8
CAPITULO II – ANALISIS EXPLORATORIO DE DATOS	10
2.1. Datos Disponibles	10
2.2. Análisis Exploratorio vía simulación de demanda	11
2.3. Comportamiento de analizadores nuevos	18
2.4. Proyección de instalaciones para el año 2024.....	25
2.5. Tiempos de contratación	28
2.6. Tiempos de capacitación.	29
CAPITULO III – METODOLOGÍA	31
3.1. Predicción de demanda de servicio futura modelado por machine learning a través de datos proporcionados por GEMATEC SRL.....	31
3.2. Modelo de programación Lineal entera	46
CAPITULO IV – CONCLUSIONES GENERALES	64
4.1. Conocimientos adquiridos y aplicados	64
4.2. Limitaciones del estudio	64
4.3. Futuras mejoras y dirección de la investigación.....	65
4.4. Aplicabilidad del trabajo	65
4.5. Posibles líneas de trabajo emergentes	65
4.6. Conclusión final	66
Bibliografía	67

CAPITULO I – INTRODUCCION

1.1. Contexto

GEMATEC SRL, una empresa especializada en la comercialización de analizadores de laboratorio clínico, enfrenta desafíos particulares en su departamento de postventa debido al constante crecimiento de su base instalada. Este incremento es resultado directo de su exitosa actividad comercial, que a su vez ha impulsado una demanda creciente de servicios de soporte técnico. La complejidad en la demanda de servicio se agrava debido a que el conocimiento técnico necesario para mantener estos dispositivos está en gran medida en manos de los fabricantes, lo que dificulta la contratación de técnicos con experiencia en el mercado laboral. En particular, un técnico que se integra a GEMATEC SRL necesita aproximadamente ocho meses de formación antes de poder operar de manera autónoma. La especialización requerida para manejar estos equipos sofisticados hace indispensable que la empresa cuente con un sistema predictivo de demanda eficiente. Este sistema sería crucial para anticipar futuras necesidades de servicio, permitiendo al departamento de postventa no solo ajustarse proactivamente a la demanda creciente, sino también gestionarla efectivamente para asegurar una respuesta oportuna y de alta calidad a las necesidades de sus clientes.

La bibliografía citada al final de este documento respalda la relevancia de utilizar técnicas de machine learning para la predicción de la demanda de servicios. Güçlükol (2020) en su estudio sobre la predicción de la demanda de servicios postventa, utiliza datos históricos de una empresa turca de electrodomésticos para comparar métodos estadísticos tradicionales con algoritmos de machine learning. Sus hallazgos muestran que las redes neuronales y la regresión de soporte vectorial superan en precisión a los métodos basados en promedios móviles, permitiendo una planificación más precisa de las necesidades de servicio y recursos humanos. Zohdi et al. (2022) aplican un enfoque similar en la predicción de demanda basada en la información del cliente, utilizando algoritmos de aprendizaje automático que consideran múltiples variables para

mejorar la exactitud de las predicciones, lo cual es fundamental para la gestión eficiente de recursos en empresas de servicios.

Además, la integración de inteligencia artificial en los procesos de selección de personal ha mostrado ser efectiva. Will, Krpan, y Lordan (2022) desarrollan el marco HIRE, que evalúa cómo la inteligencia artificial puede mejorar la toma de decisiones en la contratación. Este marco se enfoca en balancear la eficiencia y precisión de las máquinas con la experiencia humana, destacando que la combinación de ambos puede optimizar significativamente los procesos de selección de personal. Sin embargo, también resaltan los desafíos relacionados con la equidad y la transparencia, aspectos que deben ser cuidadosamente gestionados para evitar sesgos. Goretzko e Israel (2022) profundizan en estos problemas, analizando cómo la falta de transparencia y la baja calidad de datos pueden comprometer la equidad en la selección de personal basada en machine learning. Su investigación subraya la necesidad de garantizar la calidad de los datos y la transparencia de los algoritmos para mantener la confianza en estos sistemas.

Pessach et al. (2022) combinan técnicas de aprendizaje automático con programación matemática para optimizar los procesos de reclutamiento y capacitación en su estudio. Utilizan datos históricos y modelos predictivos para determinar el momento óptimo para iniciar el proceso de contratación, considerando el tiempo necesario para la formación y la demanda futura de servicios. Este enfoque no solo mejora la eficiencia en la contratación, sino que también asegura que los colaboradores estén disponibles cuando se necesiten. Smelyakov et al. (2023) complementan esta visión al analizar la efectividad de diversos algoritmos de machine learning en la toma de decisiones de contratación, demostrando que estos modelos pueden mejorar la precisión y reducir los tiempos de selección, contribuyendo a una mejor gestión de recursos humanos.

1.2. Comprensión y predicción de la demanda futura mediante machine learning

La tesis presentada en este documento ha abordado exitosamente la comprensión y predicción precisa de la demanda futura de servicios postventa en GEMATEC SRL mediante la implementación de modelos de machine learning. La especialización y la alta tecnicidad del equipo de GEMATEC SRL, que requiere un período considerable para la formación técnica, han hecho esencial una planificación efectiva de recursos. Se han evaluado y seleccionado algoritmos adecuados mediante pruebas exhaustivas, utilizando series de tiempo y datos históricos. Esto ha mejorado significativamente la capacidad de la empresa para adaptarse a las fluctuaciones del mercado y responder eficientemente a las necesidades del cliente.

1.3. Planificación Estratégica de Contratación Basada en la Predicción de la Demanda

El desarrollo de esta tesis permitió la elaboración de un modelo de optimización utilizando Programación Lineal Entera (PLE), implementado a través del software CPLEX. Este modelo ha sido crucial en la optimización de la contratación de personal técnico en GEMATEC SRL. El modelo determina el momento óptimo para iniciar la contratación de nuevos empleados, considerando el período de formación de ocho meses y los tiempos de contratación habituales. Al integrar las predicciones de demanda del modelo de machine learning, GEMATEC SRL ha logrado planificar estratégicamente el proceso de contratación y formación, optimizando los recursos, minimizando costos y manteniendo un alto nivel de servicio al cliente.

1.4. Impacto de la tesis

La descripción detallada de cómo se logró este éxito se explorará en los capítulos siguientes de esta tesis.

El éxito de la investigación ha sido evidenciado por la eficacia del modelo de machine learning implementado para predecir con precisión la demanda futura de servicios postventa y la efectividad del modelo de optimización PLE en determinar el momento óptimo para iniciar procesos de contratación y formación.

La implementación de estos modelos ha permitido a GEMATEC SRL contar con herramientas de analytics que le permitan ser más eficientes con sus procesos internos.

La eficiencia operativa se ha mejorado al reducir el tiempo de decisión a la hora de contratar nuevo personal, permitiendo esto tener al mismo capacitado y disponible al momento de experimentar un pico de demanda futuro, logrando así un mejor servicio a su cartera de clientes. Además, la planificación estratégica basada en las predicciones de demanda ha permitido a la empresa explorar nuevas alternativas comerciales que se adapten a las necesidades de los clientes.

La calidad del servicio de la empresa se ha medido mediante encuestas de satisfacción del cliente y análisis de tiempos de respuesta, mostrando una mejora en la rapidez y efectividad de la resolución de problemas técnicos. La satisfacción del cliente ha aumentado, reflejando una percepción positiva del servicio postventa.

Este logro aborda de manera efectiva los desafíos específicos de GEMATEC SRL en la gestión de su creciente demanda de servicios postventa, mediante la integración de tecnologías avanzadas de análisis de datos y optimización de recursos. El éxito alcanzado con estas soluciones marca un avance significativo en la eficiencia operativa y la calidad del servicio de la empresa, alineándose con sus objetivos estratégicos y comerciales.

CAPITULO II – ANALISIS EXPLORATORIO DE DATOS

2.1. Datos Disponibles

La empresa GEMATEC SRL, especializada en la comercialización de analizadores de química clínica, dispone de un valioso sistema de gestión de relaciones con el cliente (CRM). Este CRM contiene datos esenciales que incluyen números de serie de más de 3500 analizadores, información sobre más de 2726 clientes y sus transacciones en términos de servicios, equipamiento y reactivos. A pesar de la riqueza de esta base de datos, su potencial actual está subutilizado, ya que los informes generados son predominantemente retrospectivos.

Para esta tesis, el enfoque se centra en los datos relacionados con la demanda de postventa, que incluyen:

- Número de orden de servicio: Identificador único para cada incidencia.
- Información del cliente.
- Técnico/bioquímico asignado.
- Fechas de creación y resolución.
- Tipo de atención.
- Registros de tickets anteriores.
- Estado de la resolución y motivo si no fue resuelto.
- Tipo de analizador involucrado.
- Categoría de venta
- Equipo de distribución
- Número de Lote por analizador
- Equipo técnico de soporte

El CRM permite realizar un análisis sobre estos tickets, lo que nos posibilita discernir la labor y tiempo invertido por cada profesional en función de los diferentes tipos de analizadores y tareas atendidas. La base de datos, que cuenta con registros desde 2017, muestra un incremento significativo en las

incidencias, pasando de menos de 300 por mes en 2017 a más de 600 incidencias mensuales en 2023. Esta tendencia creciente subraya la importancia de anticipar la demanda futura, lo que fundamenta la propuesta de esta tesis.

La base de datos está estructurada en formato SQL, pero para mayor versatilidad en su manipulación y análisis, se ofrece la posibilidad de exportar los datos en formatos Excel o CSV. Además, el CRM cuenta con una API integrada con una plataforma de Power BI, lo que permite realizar análisis avanzados y visualizaciones detalladas.

Por otro lado, está a disposición el presupuesto de ventas por línea de equipamiento para el año 2024 de parte del departamento de ventas, esta información será crucial para entender en qué proporción va a crecer la base instalada y como estaría configurada su distribución en cuanto a tipo de equipamiento se refiere.

Además de la información contenida en el CRM, se cuenta con informes de una consultora externa de recursos humanos. Estos informes ofrecen métricas como el "Time to Hire" promedio y la tasa de contratación en relación con la cantidad de entrevistas realizadas para un puesto.

2.2. Análisis Exploratorio vía simulación de demanda

Para abordar y justificar de manera concreta el problema planteado en esta tesis, se desarrolló un código de simulación detallado que modela la demanda de servicios postventa en GEMATEC SRL. Este código simula una distribución normal basada en datos históricos de la demanda, proporcionando un marco cuantitativo para comprender y anticipar las necesidades futuras de servicios.

El proceso de simulación se llevó a cabo utilizando un conjunto de datos representativo que refleja el comportamiento histórico de la demanda. Para simular esta distribución, se calcularon los parámetros estadísticos clave (media y desviación estándar de los datos históricos). Estos parámetros se utilizaron

para configurar la distribución normal que modela la variabilidad esperada en la demanda de servicios.

El código de simulación implementado permitió realizar múltiples corridas, generando escenarios variados de demanda futura. Esta capacidad de simulación ofreció dos ventajas principales: primero, facilitó una comprensión más profunda del problema al permitir visualizar cómo las fluctuaciones en la demanda podrían impactar la gestión de recursos de la empresa. Segundo, proveyó una base sólida para la justificación del problema ante la gerencia general, demostrando mediante datos simulados la posible evolución de la demanda y la consiguiente necesidad de ajustes en el personal técnico.

Así, la simulación no solo respaldó el planteamiento del problema en este documento, sino que también sirvió como herramienta de validación para justificar ante la gerencia general de GEMATEC SRL. la necesidad de una planificación estratégica de recursos humanos. Este enfoque analítico y basado en datos facilitó la toma de decisiones informadas, permitiendo a la gerencia general comprender y apoyar las recomendaciones para optimizar la asignación de recursos y mejorar la capacidad de respuesta a la demanda de servicios postventa.



Figura 1: Evolución de cantidad de órdenes de servicio por trimestre desde el T1 2017 hasta el parcial T4 de 2023

Se puede observar claramente como en la figura 1 la demanda va incrementando con el pasar de los años, lo cual responde a la actividad comercial de la empresa que incrementa la base instalada de manera constante. (Al momento de la

extracción de estos datos en noviembre del 2023 no se contaba con los datos completos del trimestre lo cual explica la caída en la última data point)

A partir de estos datos crudos, se procedió a estructurar una planilla que resume la cantidad de incidencias por línea de producto para cada día del período analizado (01 de Julio del 2022 a 01 de Julio del 2023). Este resumen nos permitió calcular el desvío estándar y la varianza de las incidencias, elementos clave para nuestra posterior simulación de la demanda.

Técnico	Q. Clínica	Hemato	Gases	Inmuno	Ionos	Q. Clínica (T)	Hemato (T)	Gases (T)	Inmuno (T)	Ionos (T)	HORAS	DIAS HABILES	HORAS x DIA	TIEMPO OCIOSO	UTILIZACION
T1	2	1	174	20	2	5	3	117	30	0	1432,62	228	6,28	1,72	79%
T2	6	8	150	13	5	12	11	280	30	6	1642,44	235	6,99	1,01	87%
T3	14	6	96	4	25	0	0	95	5	7	1028,8	241	4,27	3,73	53%
T4	64	99	42	3	0	104	120	26	7	2	1732,64	228	7,6	0,4	95%
T1	64	74	17	7	4	44	40	34	5	0	1264,49	235	5,38	2,62	67%
T2	66	88	49	4	2	64	43	15	0	1	1530,05	244	6,27	1,73	78%
T3	73	69	200	21	8	36	27	46	11	5	2481,37	244	10,17	-2,17	127%
T4	24	79	115	3	1	7	12	14	2	0	1413,75	237	5,97	2,03	75%
TOTAL	313	424	843	75	47	272	256	627	90	21	12526,2	1892	52,93	11,07	83%

Figura 2: Datos de distribución de demanda por responsable y horas promedio de resolución por línea de producto.

DISTRIBUCION DE TIEMPO	QUIMICA CLINICA	HEMATOLOGIA	GASES	INMUNO	IONOS
VISITA IN SITU	3,7	2,95	2,66	2,37	1,87
ATENCION TELEFONICA	1,92	1,26	1,68	1,72	1,18
VIAJE A CLIENTE IDA Y VUELTA	3,25				
HORA DE ALMUERZO	1				
JORNADA LABORAL	9				
DIAS HABILES	251				

Figura 3: Promedios de tiempo requerido para realización de actividades técnicas.

La figura 2 es el resultante de un desglose de las órdenes de servicios asociadas al período previamente mencionado, en el cual se ordenaron dichas órdenes de servicio según el responsable y el tipo de analizador asociado, a su vez, se realizó una encuesta informal entre técnicos tanto de GEMATEC SRL como de otras empresas competidoras para calcular los tiempos promedios de resolución para cada tipo de analizador lo cual resultó en la creación de la figura 3. Con los nuevos datos asociados a esta encuesta se pudo construir el resto de la figura 2 la cual contiene cálculos de la utilización por responsable asociada al periodo y la cantidad de incidencias que ocurrieron por línea de analizador. A partir de esta información se realizaron cálculos de varianza y desvío estándar con la finalidad de simular una distribución normal que emulase el comportamiento de la demanda en un periodo determinado.

Producto	Fecha de finalización ²	Teléfono	Visita	Total, general	DESVIÓ TELEFONO	DESVIÓ VISITA	DESVIÓ CUADRADO TELEFONO	DESVIÓ CUADRADO VISITA	PROMEDIO TELEFONO	VARIANZA TELEFONO	DESV ESTANDAR TELEFONO	PROMEDIO VISITA	VARIANZA VISITA	DESV ESTANDAR VISITA
BS	1-ene	1	0	1	-0,12	-1,49	0,01	2,21	1,12	4,47	2,11	1,49	3,12	1,77
BS	4-ene	2	3	5	0,88	1,51	0,78	2,29						
BS	6-ene	1	1	2	-0,12	-0,49	0,01	0,24						
BS	9-ene	0	1	1	-1,12	-0,49	1,25	0,24						
BS	10-ene	1	1	2	-0,12	-0,49	0,01	0,24						
BS	11-ene	1	1	2	-0,12	-0,49	0,01	0,24						
BS	12-ene	0	1	1	-1,12	-0,49	1,25	0,24						
BS	13-ene	1	0	1	-0,12	-1,49	0,01	2,21						
BS	16-ene	0	1	1	-1,12	-0,49	1,25	0,24						

Figura 4: Extracto de planilla de cálculo de desvío estándar y varianza para cada día de trabajo del año 2023

Una vez realizados los cálculos de la figura 4 se procede a unificarlos todos en la figura 5, la cual se utilizó como input para el script de Python en donde se creó el entorno de simulación.

		TELEFONO					
		QUIMICA	HEMATO	GASES	INMUNO	IONO	OTROS
MEDIA		1,12	1,03	2,41	0,84	0,36	0,62
DESV.ESTANDAR		2,11	2,07	3,88	0,82	1,02	0,74
		VISITA					
		QUIMICA	HEMATO	GASES	INMUNO	IONO	OTROS
MEDIA		1,49	1,69	2,95	0,62	0,91	1,13
DESV.ESTANDAR		1,77	1,82	5,20	0,76	0,71	1,05

Figura 5: Tabla final de cálculos para utilización de input en entorno de simulación de Python

Posterior a la obtención de los datos de la figura 5, se procedió a generar la simulación en Python el script realiza una simulación para evaluar la utilización del servicio técnico de GEMATEC SRL. Utiliza datos como el tiempo promedio dedicado a visitas y llamadas telefónicas, el tiempo de viaje a los clientes, y la capacidad laboral de los técnicos, para generar demandas de servicio en distintas categorías. La simulación calcula la cantidad total de horas de trabajo requeridas por día y las compara con las horas disponibles de los técnicos, para determinar la utilización diaria del servicio técnico. El script luego calcula y visualiza la utilización promedio a lo largo de todas las simulaciones, comparándola con un umbral ideal de utilización (80%). Este análisis ayuda a entender si los recursos técnicos están siendo utilizados eficientemente o si hay exceso o falta de capacidad en relación con la demanda. A través de esta simulación, observamos que el sistema de GEMATEC SRL está casi saturado.

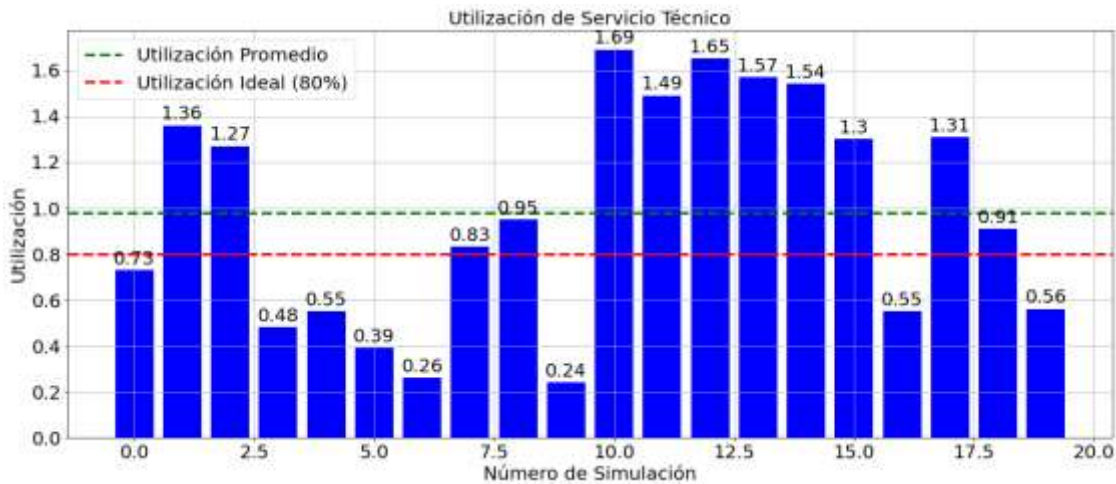


Figura 6: Gráfico de utilización para cada día simulado con una cantidad de 8 técnicos de campo (cantidad actual dentro de GEMATEC). Escala de eje vertical de 0 a 1.8 representando porcentajes de utilización, escala de eje horizontal de 0 a 20 representando número de días simulados.

Podemos observar en la figura 6 cómo la capacidad promedio del servicio (línea verde) actualmente está muy cerca de llegar a un punto de saturación que terminaría generando una acumulación considerable de demanda no atendida. Aunque en algunos días se presentan picos que exceden la capacidad, estos son compensados por días de baja demanda, lo que equilibra la utilización promedio. Sin embargo, en el contexto del mercado de la salud, este equilibrio es indicativo de una mala calidad de servicio, ya que puede llevar a la suspensión del diagnóstico que se brinda a los pacientes. A continuación, en la figura 7, veremos cómo cambiaría esta situación si GEMATEC SRL contara hoy en día con dos técnicos entrenados adicionales.

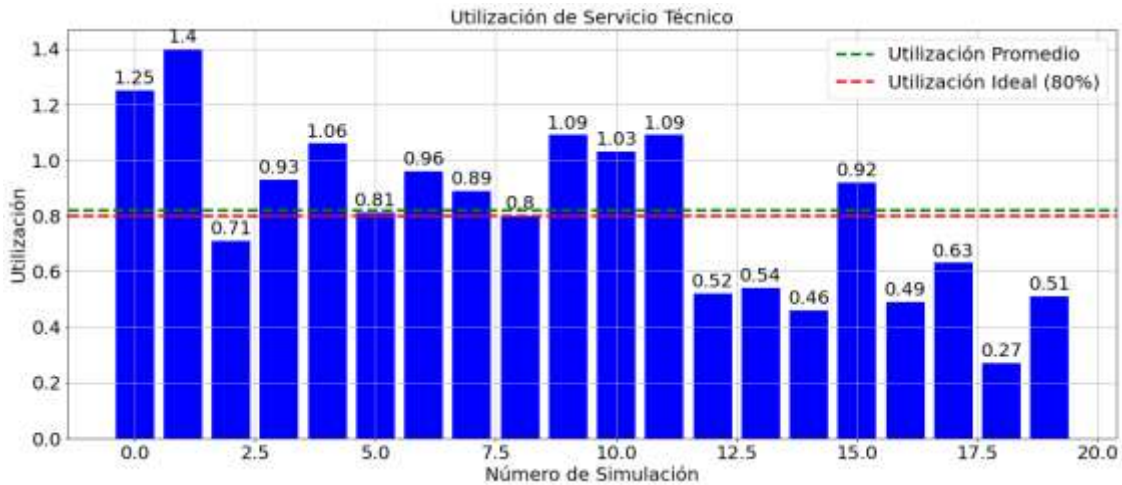


Figura 7: Mejora sustancial en la utilización al incrementar la cantidad de técnicos a 10.

La simulación y análisis realizados nos llevan a una conclusión crucial que se alinea directamente con el objetivo principal de esta tesis: la predicción precisa de la demanda futura es fundamental para la planificación efectiva de la fuerza técnica necesaria en GEMATEC SRL. Este entendimiento detallado de la demanda futura es esencial no solo para mantener, sino para mejorar la calidad del servicio al cliente.

Al identificar la proximidad del sistema a su capacidad máxima y reconocer el potencial de mejorar la respuesta al incrementar el número de técnicos, esta investigación resalta la importancia de anticipar la demanda de servicio. Esta anticipación permite planificar con suficiente antelación las contrataciones necesarias, asegurando que el personal esté adecuadamente capacitado y listo para enfrentar los desafíos futuros.

La capacidad de prever con precisión la demanda de servicios no solo optimiza la asignación de recursos humanos, sino que también garantiza una gestión más eficiente y una mayor satisfacción del cliente. Esta visión proactiva y basada en datos de la gestión de recursos es crucial para una empresa como GEMATEC SRL, donde la calidad del servicio técnico es directamente proporcional a la experiencia y habilidades del equipo técnico.

Por lo tanto, los hallazgos de este análisis respaldan el argumento central de la tesis: la necesidad de una herramienta predictiva avanzada para la proyección de demanda en servicios técnicos. Esta herramienta no solo abordará la problemática actual de la saturación del servicio técnico, sino que también establecerá un marco para una planificación estratégica más robusta y orientada al futuro. Esta planificación incluye la contratación y capacitación de técnicos, asegurando así que GEMATEC SRL continúe brindando un servicio de alta calidad a medida que su base de clientes y operaciones se expande.

2.3. Comportamiento de analizadores nuevos

Con la finalidad de comprender como sería el comportamiento de los nuevos analizadores que se sumen a la base instalada de GEMATEC SRL se realizó un análisis en profundidad de las tendencias de servicio resultantes de los lotes únicos.

En el presente apartado se hace mención a los hallazgos derivados del análisis de la base de datos de GEMATEC SRL, se considera pertinente resaltar que para esta exploración se utiliza una base de datos distinta que la base con la que se entrena el modelo final dado que contiene mucha información que correlaciona al 100% con la variable que se busca predecir. El periodo de análisis en esta base es exclusivamente lotes con órdenes de servicio generadas desde el 01/01/2020 al 31/03/2024. A continuación, las columnas de dicho archivo:

- Fecha: Fecha en la que se registró el evento o la acción.
- Equipo: Nombre o tipo de equipo involucrado.
- Categoría padre: Categoría padre de equipamiento. Por ejemplo, Hematología.
- Categoría interna: Subcategoría más específica de equipamiento. Por ejemplo, contador hematológico BS10
- Cliente: Nombre o identificador del cliente.
- Distribuidor: Nombre o identificador del distribuidor.
- Responsable: Persona o departamento responsable de la resolución.

- Equipo de trabajo: Cuadrillas a la que pertenece cada responsable relacionado a la atención de una categoría padre de producto.
- Tipo de resolución: Método o tipo de resolución aplicado.
- Etiqueta: Etiquetas adicionales relacionadas con el registro.
- Lote: Lote de producción del equipo o componente.
- Estado del equipo: Estado actual del equipo al momento del registro.
- Cantidad RMA: Indicador de órdenes de servicio (RMA), indica si un conjunto de lote/fecha tuvo o no una orden de servicio asociada, cabe a resaltar, que esta es la característica que se buscó eventualmente predecir en el modelo de machine learning.

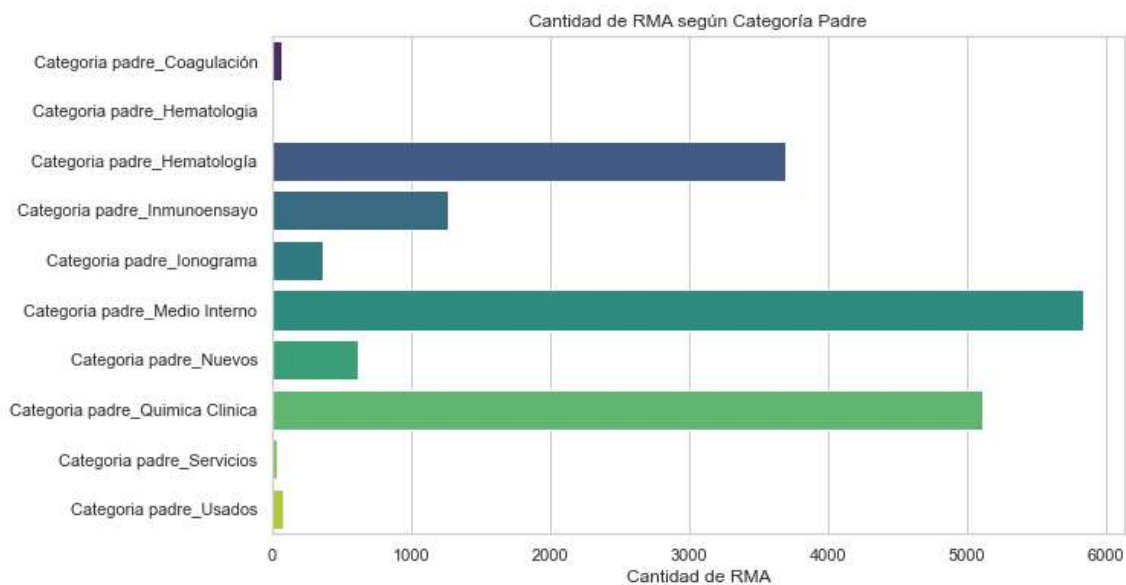


Figura 8: Cantidad de RMA según Categoría padre entre 01/01/2020 y 31/04/2024

En la figura 8 se pueden apreciar la distribución de las RMA asociadas al período, existe una polarización lógica en las líneas de Hematología, Química clínica y Medio interno dado que son las tres líneas principales de comercialización de GEMATEC SRL.

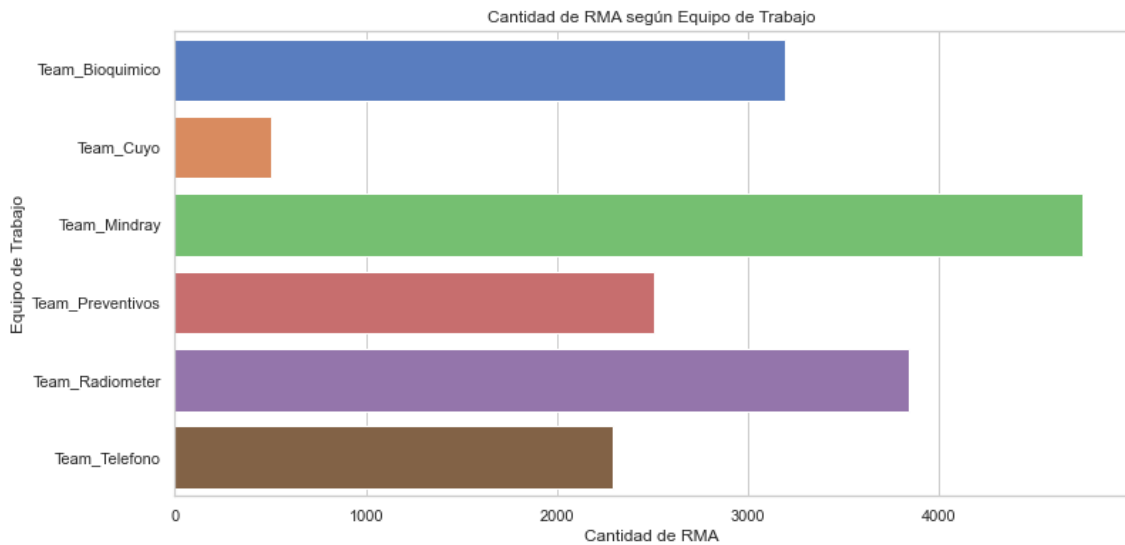


Figura 9: Distribución de RMA según equipo de atención asistencial.

A partir de los datos que figuran en la figura 9 se puede apreciar cómo está dada la distribución del trabajo según los equipos de atención asistencial, considerando que los equipos bioquímicos, radiometer y mindray poseen tres integrantes cada uno y entendiendo que las unidades de mindray son las que tienen tiempos de resolución más largos (sumatoria de unidades de química Clínica + hematología, referir a figura 5), se puede notar un desbalance en la cantidad de trabajo, sin embargo, al ser este un trabajo especializado no se pueden realizar cambios de equipo sin tener una caída marcada de la cantidad de servicios que se pueden resolver hasta que un profesional pueda ser autónomo posterior a un cambio de equipo.

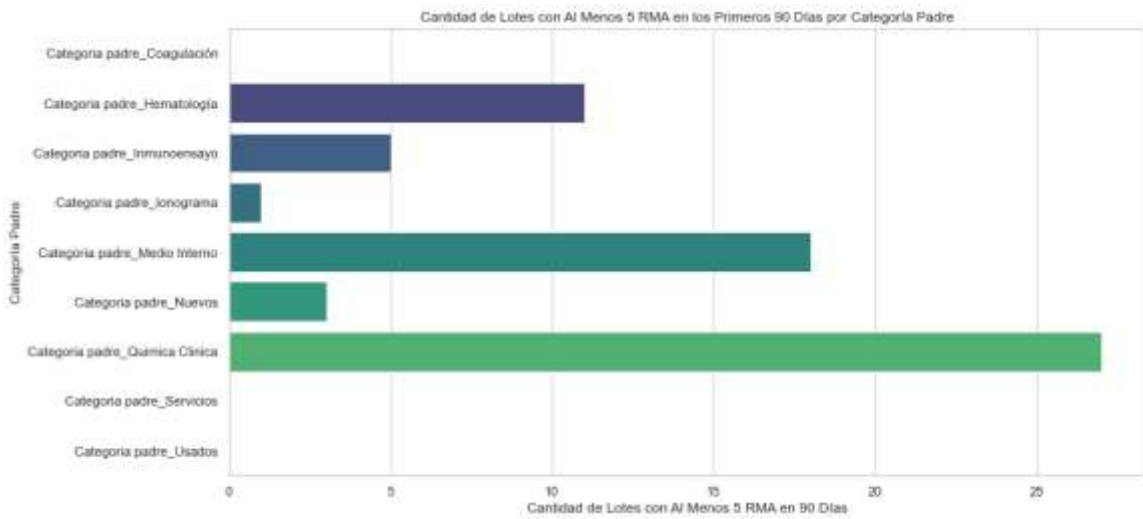


Figura 10: Lotes con 5RMA+ en sus primeros 90 días de funcionamiento

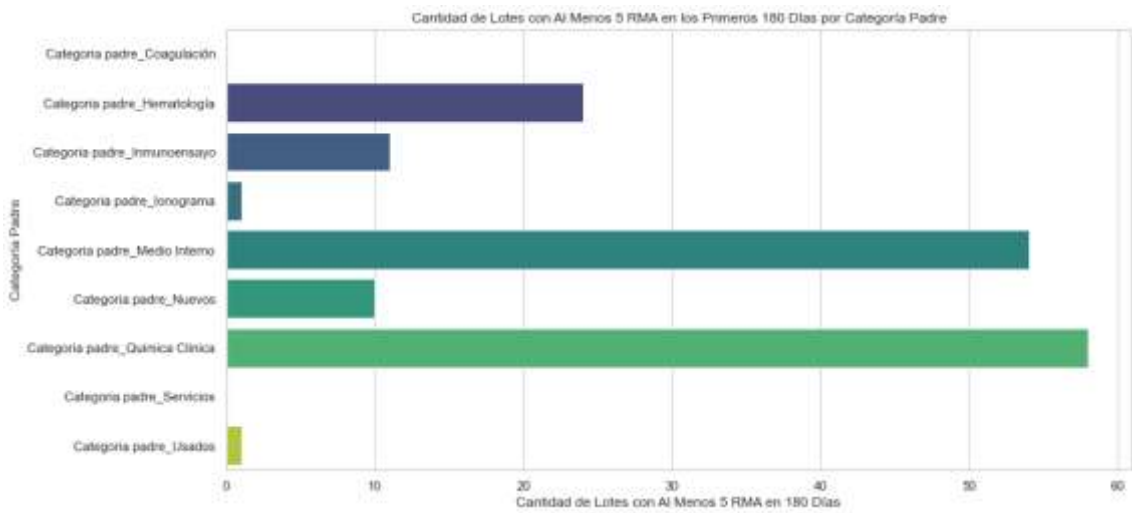


Figura 11: Lotes con 5RMA+ en sus primeros 180 días de funcionamiento

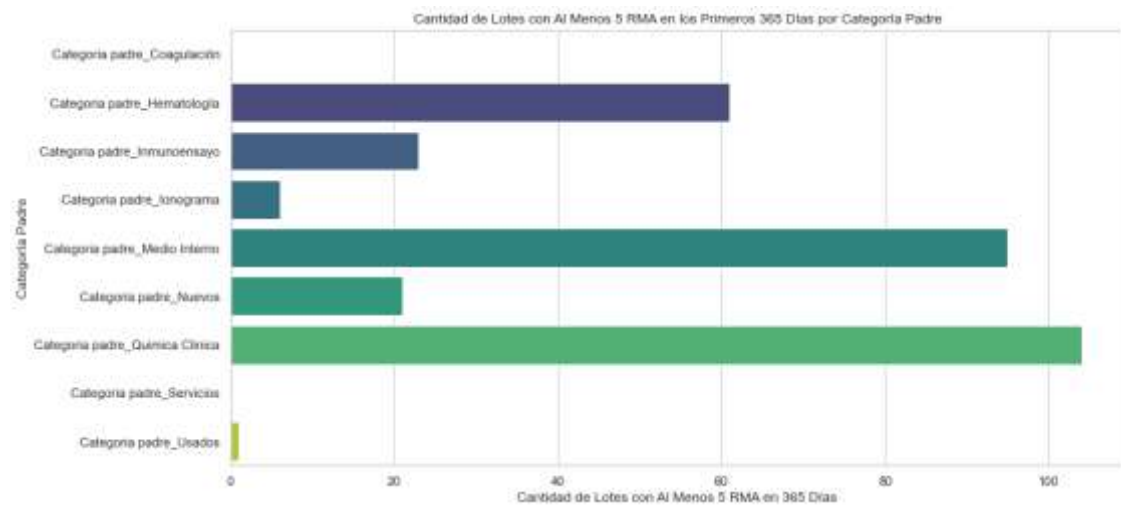


Figura 12: Lotes con 5RMA+ en sus primeros 365 días de funcionamiento

Al analizar los patrones de órdenes de servicio (RMA) para lotes de equipos en los intervalos de 90, 180 y 365 días, se observa un incremento progresivo en el número de órdenes a medida que el periodo de observación se extiende. Este patrón sugiere una correlación entre el tiempo en operación de los equipos y la frecuencia de mantenimiento requerido.

En particular, la categoría 'Química Clínica' presenta una frecuencia constante y elevada de órdenes de servicio en todos los intervalos de tiempo considerados, lo que podría indicar un uso más intensivo de estos equipos o una susceptibilidad a fallas que demanda atención técnica frecuente. Asimismo, la categoría 'Medio Interno' experimenta un incremento significativo en las órdenes de servicio a medida que se amplía el marco temporal, lo cual podría apuntar a problemas de desgaste o a la aparición de complicaciones técnicas con el uso prolongado.

Se debe destacar que categorías como 'Hematología' y 'Coagulación' muestran un aumento en la cantidad de RMA cuando se extiende la observación de 90 a 365 días. Este comportamiento podría interpretarse como un indicativo de periodos de operación inicialmente estables que, con el tiempo, evolucionan hacia una mayor necesidad de intervenciones técnicas.

Estos datos enfatizan la importancia de implementar estrategias de mantenimiento preventivo y de ajustar la planificación de recursos técnicos para abordar de manera proactiva las órdenes de servicio, especialmente en aquellas categorías con altas tasas de RMA. Un enfoque predictivo en la planificación podría mejorar significativamente la respuesta a las demandas de mantenimiento, optimizando la distribución de los recursos técnicos y, en consecuencia, mejorando la satisfacción del cliente y la eficiencia operativa a largo plazo.

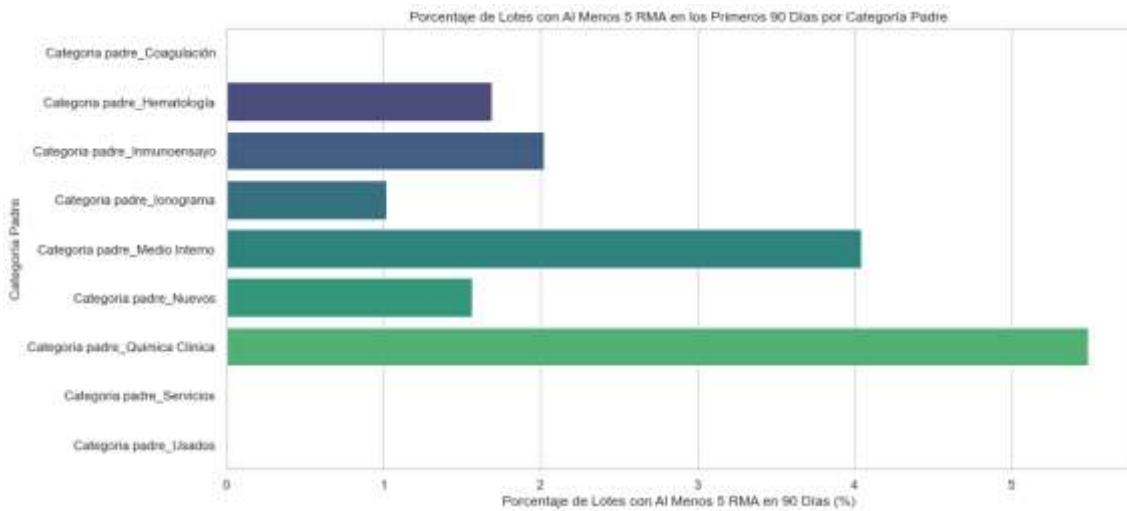


Figura 13: Porcentaje de Lotes sobre el total de lotes con más de 5RMA en los primeros 90 días de funcionamiento.

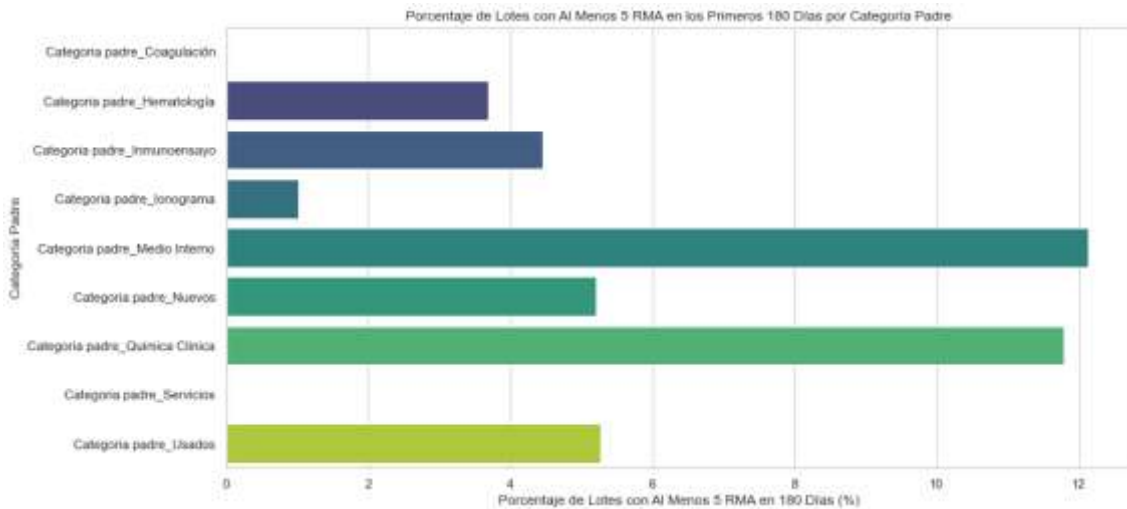


Figura 14: Porcentaje de Lotes sobre el total de lotes con más de 5RMA en los primeros 180 días de funcionamiento.

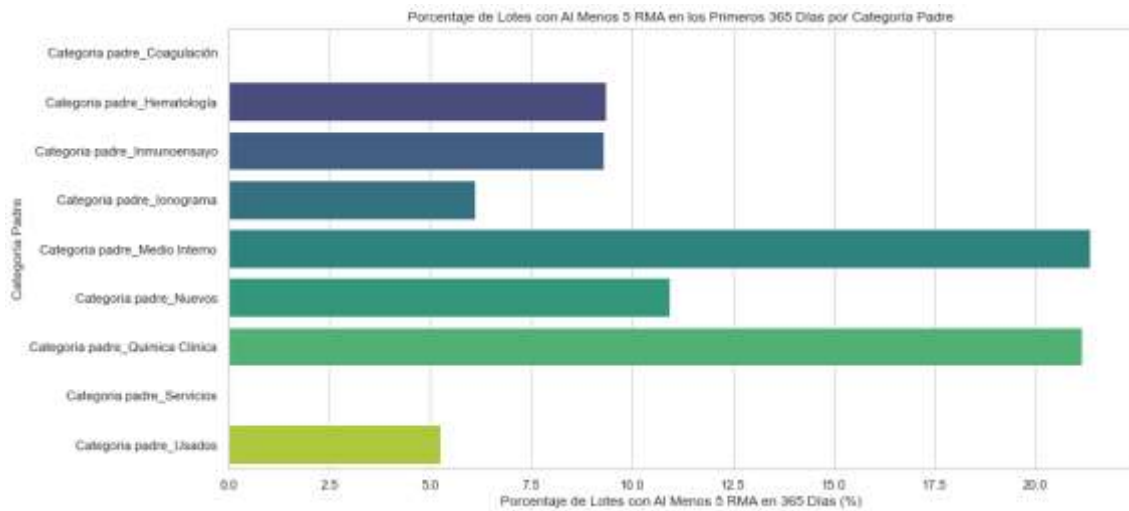


Figura 15: Porcentaje de Lotes sobre el total de lotes con más de 5RMA en los primeros 365 días de funcionamiento.

Observando los porcentajes de lotes que han requerido al menos 5 órdenes de servicio (RMA) en los periodos de 90, 180 y 365 días, se puede realizar una evaluación conjunta de la durabilidad y la fiabilidad de los equipos distribuidos por GEMATEC SRL.

En cada uno de los intervalos de tiempo evaluados, se aprecia que la categoría 'Química Clínica' mantiene la proporción más alta de lotes que requieren atención técnica. Esto sugiere que, en términos relativos, esta categoría podría estar más expuesta a fallas o que su uso frecuente conduce a una mayor incidencia de problemas que necesitan resolución técnica.

Al expandir el lapso de tiempo de 90 a 365 días, se evidencia un incremento consistente en el porcentaje de lotes con RMA en todas las categorías, lo que es coherente con el desgaste natural y la probabilidad acumulada de fallos técnicos con el uso extendido. Es notable que la categoría 'Usados' (Equipos refurbish) muestra un incremento sustancial en el porcentaje de RMA cuando se considera el período completo de un año, lo que plantea interrogantes sobre la eficacia de los procedimientos de reacondicionamiento y la idoneidad de estos equipos para un uso prolongado.

Estas observaciones señalan la relevancia de llevar a cabo un análisis detallado de las políticas de mantenimiento y la estrategia de gestión de equipos. La

implementación de mejoras en estas áreas no solo podría reducir la frecuencia de las órdenes de servicio y aumentar la longevidad de los equipos, sino que también proporcionaría una base para una mejor asignación de recursos técnicos y un incremento en la satisfacción global del cliente.

2.4. Proyección de instalaciones para el año 2024

Se incluirá en el análisis un cuadro que refleja las proyecciones de instalaciones futuras de analizadores, cortesía del departamento de ventas de GEMATEC SRL. La integración de estas proyecciones con la demanda existente de la base instalada es crucial para reforzar la precisión de nuestro modelo de forecasting de demanda basado en machine learning. Esta fusión de datos proporciona una estimación más holística del volumen de demanda que se puede esperar, lo cual es crucial para afinar el modelo y mejorar la exactitud de las predicciones. La valoración de esta información es primordial, ya que permite una asignación de recursos más efectiva y un mejor planeamiento estratégico en el departamento de soporte técnico, asegurando una infraestructura de servicio que esté en consonancia con las necesidades futuras de la empresa y sus clientes.

MODELO	LINEA	ene-24	feb-24	mar-24	abr-24	may-24	jun-24	jul-24	ago-24	sep-24	oct-24	nov-24	dic-24	TOTAL 2024
BC-10	Hematología 3P	1	1	3	3	3	3	3	3	4	4	4	3	35
BS-240	Química Clínica				2	2	3	1	1	2	1	1	2	15
393-899	Medio Interno		1	1	2	1	1	1	1	1	1	1	1	12
BC-5000	Hematología 5P			2	1	2	1	1	1	3	1	1	1	14
BS-230	Química Clínica			1		1		1						3
BS-360E	Química Clínica	1	1	1	1	1	1	1	1	1	1	1	2	13
BC-30s	Hematología 3P			1	1		1	1		1		1		6
BC-5390	Hematología 5P				1	1	1	2	2	2	1	1	1	12
23020003	CLIA				1	1	1	1	1	1	1	1	1	9
BC-3000 U	Hematología 3P	1	1				1			1		1		5
BC-780[R]	Hematología 6P			1		1		1		1		1		5
010101003301	CLIA				1		1		1		1		1	5
BC-5150	Hematología 5P			1			1							2
393-090	Medio Interno				1		1			1			1	4
BA-88A	Química Clínica			1			1			1				3
BS-480 U	Química Clínica						1			1			1	3
BT24[BC]	Microbiología			1				1				1		3
BT48[BC]	Microbiología				1				1				1	3
BC-6000	Hematología 5P						1			1				2
BS-430	Química Clínica						1				1			2
BS-380 U	Química Clínica						1			1				2
BC-6200	Hematología 5P									1				1
393-844	Medio Interno						1							1
Maldi Toff	Microbiología	1												1

Figura 16. Proyección de Ventas GEMATEC SRL 2024 por línea de producto.

El análisis de los datos proporcionados revela que las proyecciones de instalación de analizadores de GEMATEC S.R.L para el año 2024 varían significativamente entre modelos y líneas de producto. La línea de Hematología presenta la mayor cantidad de instalaciones planificadas, lo que puede indicar una concentración de esfuerzos de ventas o una demanda más alta en esta área.

Por otro lado, se anticipa una menor cantidad de instalaciones para los modelos dentro de las líneas de Medio Interno y Microbiología.

Los datos reflejan una estacionalidad en las proyecciones de instalaciones para ciertos modelos, como un aumento en los meses de verano y hacia el final del año, lo que podría estar alineado con los ciclos presupuestarios de los clientes o con lanzamientos estratégicos de productos. La distribución no uniforme de las instalaciones a lo largo del año sugiere que la demanda de servicio técnico podría experimentar picos correspondientes, lo cual debe ser considerado en el modelo de forecasting de demanda.

Los totales anuales por modelo brindan una perspectiva de las necesidades de recursos para mantenimiento y soporte, esencial para la planificación de la carga de trabajo del departamento de servicio técnico.

Esta información, sumada a la demanda existente de la base instalada, proporcionará una base sólida para estimar el nivel de volumen de la demanda de servicios técnicos. Con estos datos, GEMATEC S.R.L podrá optimizar la asignación de recursos, asegurando que el personal técnico esté disponible y bien distribuido a lo largo del año para cumplir con las expectativas de servicio y mantener la satisfacción del cliente.

2.5. Tiempos de contratación

Apellido	Fecha de Inicio de Búsqueda	Fecha de Ingreso	Tardanza en meses días	
Técnico 1	16/10/2016	19/12/2016	64	
Técnico 2	11/12/2016	11/1/2017	31	
Técnico 3	22/5/2017	3/7/2017	42	
Técnico 4	10/5/2017	1/8/2017	83	
Técnico 5	21/2/2018	26/3/2018	33	
Técnico 6	1/2/2019	20/3/2019	47	
Técnico 7	10/12/2019	2/1/2020	23	
Técnico 8	13/5/2020	16/6/2020	34	
Técnico 9	4/8/2020	15/9/2020	42	
Técnico 10	4/8/2020	15/9/2020	42	
Técnico 11	14/10/2020	1/2/2021	110	
Técnico 12	6/11/2020	1/2/2021	87	
Técnico 13	6/11/2020	1/2/2021	87	
Técnico 14	27/7/2021	1/9/2021	36	
Técnico 15	1/12/2021	10/1/2022	40	
Técnico 16	1/12/2021	10/1/2022	40	
Técnico 17	5/4/2022	1/5/2022	26	
Técnico 18	5/4/2022	1/5/2022	26	
Técnico 19	20/11/2023	1/2/2024	73	
Técnico 20	20/11/2023	1/2/2024	73	
Técnico 21	20/11/2023	1/2/2024	73	
			52,95	PROMEDIO

Figura 17. Tiempos promedio de contratación para servicio postventa.

El análisis de la tabla proporciona información detallada sobre el tiempo de contratación de los empleados de GEMATEC SRL para el servicio postventa, que es un dato fundamental para alimentar y precisar el modelo de optimización desarrollado por esta tesis.

Importancia de los Tiempos de Contratación:

- **Alimentación del Modelo:** Los tiempos de contratación son un parámetro crítico en el modelo de optimización, ya que determinan la disponibilidad efectiva de los recursos humanos. La exactitud en la estimación de este

tiempo es crucial para sincronizar las contrataciones con la demanda proyectada de servicios.

- **Variabilidad de los Tiempos:** La tabla muestra una variabilidad considerable en los tiempos de contratación, desde tan solo 23 días hasta 110 días. Esta variabilidad debe ser considerada en el modelo para manejar la incertidumbre y permitir la planificación de contrataciones con suficiente antelación.
- **Promedio de Tiempos de Contratación:** El promedio de tiempo de contratación, que es aproximadamente de 53 días, puede ser utilizado como un valor de referencia para el parámetro TiempoContratación en el modelo. No obstante, es importante tener en cuenta la distribución completa y no solo el promedio, ya que puede haber una amplia dispersión de los tiempos individuales.
- **Impacto en la Planificación:** Si el tiempo de contratación es subestimado, existe el riesgo de no tener suficiente personal cuando se necesita, mientras que una sobreestimación puede llevar a un sobrestaffing innecesario. Por lo tanto, el conocimiento preciso de estos tiempos permite afinar el modelo para que sea más efectivo y eficiente en términos de costos y operatividad.

Considerando el conocimiento interno del funcionamiento de la organización, se ha establecido un plazo de 60 días para las contrataciones, ya que, en casos urgentes, la empresa puede priorizar la búsqueda de perfiles aptos para el servicio de atención postventa, con el objetivo de siempre cubrir la demanda en el sector de la salud.

2.6. Tiempos de capacitación.

Los tiempos de capacitación son esenciales para garantizar que los técnicos recién contratados alcancen el nivel de competencia y autonomía necesarios para cumplir con las expectativas de rendimiento de la empresa. A continuación, se describe cómo este proceso de capacitación alimenta el modelo de optimización en el contexto de esta tesis:

- **Etapa Inicial de Capacitación:** Los recién contratados comienzan su formación con un conjunto de certificaciones a través de cursos tanto online como presenciales, que duran un total de 4 meses. Este período está dirigido por el Supervisor Técnico General y está diseñado para establecer una base sólida de conocimiento teórico y práctico.
- **Aplicación Práctica con Supervisión:** Tras completar la formación inicial, los técnicos entran en una fase práctica de 2 meses donde realizan mantenimientos preventivos bajo la supervisión de un Líder de Equipo. Este período de mentoría no solo refuerza la formación previa, sino que también permite evaluar la habilidad del técnico para aplicar su conocimiento en un entorno real de trabajo.
- **Evaluación Final y Autonomía:** Al finalizar los 2 meses de práctica supervisada, el Supervisor Técnico General realiza una evaluación compuesta por una prueba escrita y una serie de pruebas prácticas en el campo. La aprobación de esta evaluación indica que el técnico está listo para trabajar de manera autónoma dentro del equipo de preventivos, marcando así el final del proceso de capacitación de 6 meses.
- **Importancia para el Modelo de Optimización:** La inclusión de los tiempos de capacitación en el modelo de optimización es fundamental. No solo determinan cuándo un técnico recién contratado será productivo y podrá contribuir a la capacidad operativa total, sino que también influyen en el momento óptimo para iniciar el proceso de contratación. Al considerar un período de capacitación de 6 meses, el modelo puede anticipar y planificar las contrataciones de manera que se alineen con las fluctuaciones proyectadas en la demanda de servicios.

Integrar el feedback del Líder de Equipo y del Supervisor Técnico General sobre el desempeño de los técnicos en capacitación puede proporcionar datos valiosos para mejorar continuamente el contenido y la efectividad de los programas de formación. El conocimiento preciso de los tiempos de capacitación permite al modelo evitar tanto la escasez de personal técnico como el exceso de personal no capacitado. Esta optimización de recursos garantiza que los técnicos estén disponibles y sean plenamente productivos cuando se necesiten.

CAPITULO III – METODOLOGÍA

A partir de la presente se procederá a detallar exhaustivamente los procedimientos técnicos adoptados para el desarrollo de esta tesis. Se explica cada paso del proceso de investigación, desde la recopilación y análisis de datos hasta la implementación de los modelos de predicción y optimización. Este capítulo proporciona una visión completa y sistemática de las técnicas y herramientas utilizadas, asegurando la replicabilidad y la validez técnica de los resultados obtenidos.

3.1. Predicción de demanda de servicio futura modelado por machine learning a través de datos proporcionados por GEMATEC SRL

El conjunto de datos analizado para este estudio se obtuvo del sistema de gestión de relaciones con el cliente (CRM) de GEMATEC SRL mediante la implementación de herramientas de extracción de datos SQL. Con el acceso proporcionado al CRM, fue posible realizar consultas estructuradas para recopilar la información relevante necesaria para la investigación. Los datos extraídos se compilaron en un archivo CSV que sirve como fuente principal para el análisis detallado.

El dataset en cuestión comprende una serie de registros detallando las operaciones y movimientos de los lotes de equipos de GEMATEC SRL. Cada entrada es una combinación única de lote y fecha, abarcando un periodo de análisis desde el 01 de enero de 2022 hasta el 15 de abril de 2024. Los elementos clave del dataset incluyen información del lote, la fecha específica de la entrada, el equipo involucrado, categorías que describen la línea de productos y el estado de venta, así como la designación del último cliente y si se ha generado una orden de servicio (RMA).

La repetición de todos los lotes para cada fecha sugiere un registro cronológico de seguimiento, resultando en un total de 3,381,433 entradas. Esta gran cantidad de datos refleja una operación extensa y un control detallado sobre la gestión de los equipos. La información contenida en este dataset es esencial para analizar

patrones de demanda de servicio, la efectividad de las estrategias de ventas y distribución de productos, así como para el mantenimiento predictivo y la gestión de las órdenes de servicio a lo largo del tiempo.

La consistencia en la estructura de los datos facilita la extracción de información relevante para el modelado predictivo y el análisis estadístico, elementos críticos para la toma de decisiones estratégicas en el soporte y la postventa de equipos.

3.1.1. Análisis y Pre procesamiento de Datos:

El pre procesamiento de los datos constituye una etapa crucial en proyectos de aprendizaje automático. La tarea se inició cargando el conjunto de datos desde un archivo CSV y aplicando diversas transformaciones para preparar los datos para su análisis y modelado. Como se muestra en la figura 18, se convirtió la columna de fechas al formato adecuado para garantizar su procesamiento correcto por parte del algoritmo. Luego, se ordenó el DataFrame por 'Lote' y 'Fecha' con el propósito de facilitar las operaciones futuras. Además, se filtraron las filas para incluir únicamente aquellas correspondientes a fechas posteriores al 1 de febrero de 2022, limitando así el período de análisis.

```
# Convertimos la columna 'Fecha' a datetime  
df['Fecha'] = pd.to_datetime(df['Fecha'])  
  
# Ordenamos el DataFrame por 'Lote' y 'Fecha'  
df.sort_values(by=['Lote', 'Fecha'], inplace=True)  
  
# Filtrar el DataFrame para incluir solo las filas a partir del 1 de  
febrero de 2022  
df = df[df['Fecha'] >= '2022-01-01']
```

Figura 18. Filtrado y ordenamiento de fechas.

Se eliminaron columnas que podrían generar correlación del 100% con la variable a predecir (Véase figura 19).

```
# Eliminar La columna 'Responsable' del DataFrame
if 'Responsable' in df.columns:
    df = df.drop(['Responsable'], axis=1)

# Eliminar La columna 'Etiqueta' del DataFrame
if 'Etiqueta' in df.columns:
    df = df.drop(['Etiqueta'], axis=1)

# Eliminar La columna 'Tipo de resolucion' del DataFrame
if 'Tipo de resolucion' in df.columns:
    df = df.drop(['Tipo de resolucion'], axis=1)
```

Figura 19. Eliminación de columnas con correlación 100% con variable a predecir.

Se utilizaron técnicas de codificación one-hot en aquellas columnas con atributos categóricos para transformar la información en un formato numérico comprensible por el modelo. También se generaron nuevas características, como 'Fecha_Instalacion', 'Fecha_Ultima_RMA', 'Num_RMA_Previas', 'Dias_Desde_Instalacion' y 'Dias_Desde_Ultima_RMA', consideradas importantes para el análisis (detalle del código en figura 20). Finalmente, se desglosó la columna de fechas en componentes adicionales, como día de la semana, mes, etc., para capturar posibles patrones temporales en los datos, y se convirtieron las columnas booleanas generadas por esta operación a valores binarios para facilitar su procesamiento.

```

# Realizar one-hot encoding en la columna 'Lote'
df_lotes_dummies = pd.get_dummies(df['Lote'], prefix='Lote')

# Ahora, une estas nuevas columnas de 'Lote' al DataFrame original 'df'
df_with_dummies = pd.concat([df, df_lotes_dummies], axis=1)

del df, df_lotes_dummies
gc.collect()

# Ahora que hemos realizado el one-hot encoding para 'Lote', podemos calcular
las siguientes columnas:
df_with_dummies['Fecha_Instalacion'] =
df_with_dummies.groupby('Lote')['Fecha'].transform('min')
df_with_dummies['Fecha_Ultima_RMA'] = df_with_dummies[df_with_dummies['Hubo
RMA ?'] == 1].groupby('Lote')['Fecha'].shift()
df_with_dummies['Num_RMA_Previas'] = df_with_dummies.groupby('Lote')['Hubo RMA
?'].cumsum() - df_with_dummies['Hubo RMA ?']

# Desglosamos la columna 'Fecha' en componentes adicionales
add_datepart(df_with_dummies, 'Fecha', drop=False)

# Calculamos los días desde la instalación y desde la última RMA
df_with_dummies['Dias_Desde_Instalacion'] = (df_with_dummies['Fecha'] -
df_with_dummies['Fecha_Instalacion']).dt.days
df_with_dummies['Dias_Desde_Ultima_RMA'] = (df_with_dummies['Fecha'] -
df_with_dummies['Fecha_Ultima_RMA']).dt.days
df_with_dummies['Dias_Desde_Ultima_RMA'].fillna(9999, inplace=True)

# Eliminamos las columnas que ya no se necesitan
df_with_dummies.drop(['Fecha_Ultima_RMA', 'Fecha_Instalacion'], axis=1,
inplace=True)

# Convertimos las columnas booleanas generadas por add_datepart a valores
binarios
bool_columns = [col for col in df_with_dummies.columns if
df_with_dummies[col].dtype == 'bool']
for col in bool_columns:
    df_with_dummies[col] = df_with_dummies[col].astype(int)
    
```

Figura 20. One Hot Encoding y creación de nuevas features.

Además de desarrollar un modelo predictivo, se realizó un análisis de características para identificar las variables más influyentes en la predicción de si hubo o no una orden de servicio asociada a una observación. Este análisis fue crucial para el presente trabajo de tesis, ya que permito entender mejor los factores que impactan la variable objetivo y proporciono una base sólida para la toma de decisiones operativas. El análisis de características ayudo a destacar las variables que más contribuyen al comportamiento del sistema, lo que puede guiar futuras investigaciones y mejoras en los procesos.

Para esta finalidad fue el modelo Ridge el que fue seleccionado debido a su capacidad para manejar eficientemente la multicolinealidad entre las variables predictoras y su habilidad para regularizar los coeficientes de regresión, mejorando así la generalización del modelo. La regresión Ridge es especialmente útil cuando se espera que todas las características tengan alguna relevancia en la predicción y se desea evitar la eliminación completa de características, como ocurre con la regresión Lasso.

Además, el uso de la regresión Ridge ha sido una herramienta valiosa aprendida en la materia "Métodos Estadísticos Aplicados a Negocios" y ha sido utilizada a lo largo de la maestría por su efectividad comprobada en diversos análisis predictivos y de características.

En este análisis, la variable objetivo es binaria y denota si hubo (1) o no (0) una orden de servicio asociada a la observación. La predicción precisa de esta variable es crucial para anticipar y planificar la carga de trabajo, optimizando la asignación de recursos y mejorando la eficiencia operativa.

El análisis de importancia de características realizado con el modelo Ridge identificó las 10 características más influyentes (véase figura 21):

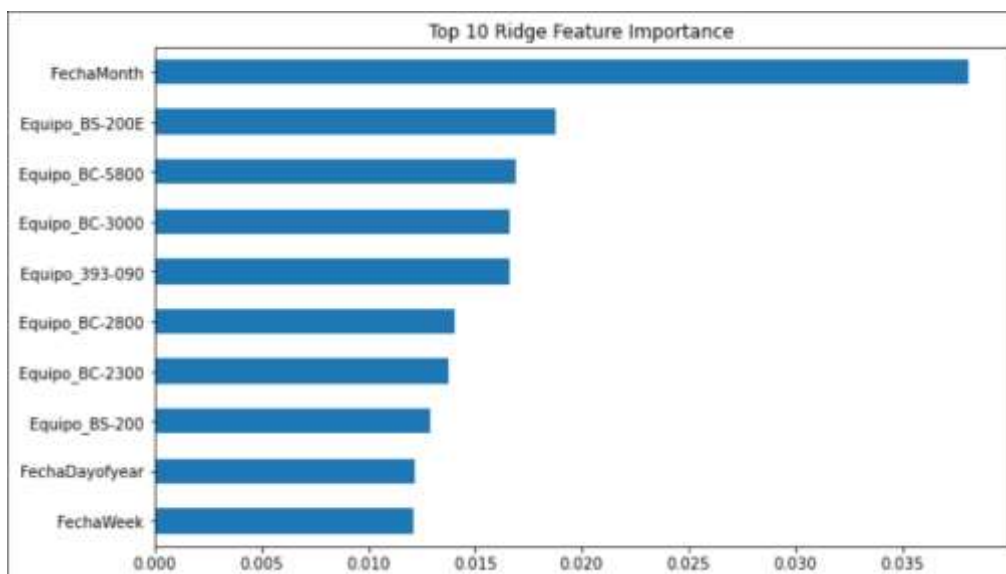


Figura 21. Análisis de características importantes con modelo Ridge.

El gráfico de importancia de características muestra los coeficientes de las 10 características más influyentes. Los coeficientes más grandes en valor absoluto indican una mayor influencia en la predicción del modelo, ayudando a identificar las características más relevantes.

Los resultados obtenidos muestran una lógica clara en la importancia de las características identificadas. Todos los analizadores clínicos mencionados en el ranking, desde la posición 2 hasta la 8, forman parte de los modelos más antiguos comercializados por la empresa. Esto tiene sentido, ya que los equipos más antiguos tienden a fallar con mayor frecuencia, lo que los convierte en características altamente predictoras en el análisis. Esta tendencia refuerza la validez del modelo y su capacidad para identificar correctamente los factores más relevantes que afectan la variable objetivo. Además, las características relacionadas con fechas (1, 9 y 10) resaltan la importancia de los efectos estacionales. La variable FechaMonth destaca patrones estacionales o cambios mensuales, mientras que FechaDayofyear y FechaWeek capturan efectos específicos del día del año y la semana, respectivamente. Estos resultados subrayan la influencia de la estacionalidad en la ocurrencia de órdenes de servicio, lo que es esencial para la planificación y optimización operativa.

3.1.2. Técnica de Separación de Conjuntos:

El conjunto de datos se dividió en tres partes principales: entrenamiento, validación y prueba. Esta división realizada en partes 80%, 10% y 10% respectivamente es efectiva para realizar una correcta prueba sobre el performance del modelo. Debido al enorme tamaño del dataset, y con la intención de procesar los datos en formato disperso, además de la separación de conjuntos se guarda un archivo por separado para cada conjunto con el fin de hacer el procesado adecuado de memoria, todo este proceso se detalla en la figura 22.

```

# Aplicamos one-hot encoding a las columnas categóricas restantes, excluyendo
'Lote' ya que ya fue codificado
categorical_columns
= df_with_dummies.select_dtypes(include=['object']).columns.tolist()
df_with_dummies = pd.get_dummies(df_with_dummies, columns=categorical_columns)

# Calcular Los índices para cortar los conjuntos de entrenamiento, validación
y prueba
train_end_idx = int(len(df_with_dummies) * 0.8)
val_end_idx = int(len(df_with_dummies) * 0.9)

# Dividir los datos
df_train = df_with_dummies.iloc[:train_end_idx]
df_val = df_with_dummies.iloc[train_end_idx:val_end_idx]
df_test = df_with_dummies.iloc[val_end_idx:]

del df_with_dummies
gc.collect()

# Guardar Los DataFrames
pickle.dump(df_train, open(output_path + 'df_train.pkl', 'wb'))
pickle.dump(df_val, open(output_path + 'df_val.pkl', 'wb'))
pickle.dump(df_test, open(output_path + 'df_test.pkl', 'wb'))
    
```

Figura 22. Separación de conjuntos y guardado de archivos separados.

3.1.3. Manejo de Memoria y Conversión a Matrices Dispersas:

El manejo eficiente de la memoria fue uno de los desafíos principales, especialmente al trabajar con archivos de datos grandes. Para mitigar este problema, se recurrió a la técnica de matrices dispersas para representar los datos de manera eficiente. Los DataFrame de pandas se convirtieron a matrices dispersas utilizando la biblioteca `scipy.sparse`, lo que permitió trabajar con conjuntos de datos más grandes sin agotar la memoria disponible.

Además, y como se muestra en la figura 23, se necesitó para fines prácticos apartar la columna a predecir de cada uno de los conjuntos y guardarla en archivos separados que luego se pudiesen concatenar a las matrices dispersas que se iban a generar.

```

# Extraer La columna de etiquetas para entrenamiento
if 'Hubo RMA ?' in df_train_loaded.columns:
    y_train = df_train_loaded['Hubo RMA ?'].copy()
    df_train_loaded.drop(['Hubo RMA ?'], axis=1, inplace=True)

# Guardar Las etiquetas de entrenamiento en un archivo pickle
with open(output_path + 'y_train.pkl', 'wb') as f:
    pickle.dump(y_train, f)

# Liberar memoria si ya no se necesita y_train
del y_train
gc.collect()

# Asegurarse de que todos Los datos en el DataFrame sean numéricos
for col in df_train_loaded.columns:
    if df_train_loaded[col].dtype == 'object'
    or df_train_loaded[col].dtype == 'bool':
        df_train_loaded[col] = pd.to_numeric(df_train_loaded[col],
errors='coerce')
df_train_loaded.fillna(0, inplace=True) # Llenar Los NaN con cero

# Convertir a tipo más pequeño para ahorrar memoria
df_train_loaded = df_train_loaded.astype('float32')
  
```

Figura 23. Separación de Etiqueta a predecir de DataFrame original.

Para el caso puntual del conjunto de entrenamiento se tuvo que recurrir a una conversión dispersa por batches, el ejemplo mostrado en la figura 24 muestra la aplicación de esta estrategia.

```
batch_size = Len(df_train_Loaded) // 100 # Dividir en 100 batches
aproximadamente iguales

# Inicializar una lista para almacenar las matrices sparse de cada batch
sparse_batches = []

for i in range(0, Len(df_train_Loaded), batch_size):
    # Seleccionar el batch actual
    batch = df_train_Loaded.iloc[i:i + batch_size]
    # Convertir el batch a una matriz CSR sparse
    sparse_batch = csr_matrix(batch.values)
    # Añadir la matriz sparse del batch a la lista
    sparse_batches.append(sparse_batch)

    # Eliminar la variable batch para liberar memoria
    del batch
    gc.collect()

del df_train_Loaded

# Concatenar todas las matrices sparse de los batches en una sola matriz
sparse
X_train_sparse = vstack(sparse_batches, format='csr')

# Guardar la matriz sparse resultante
with open(output_path + 'X_train.pkl', 'wb') as file:
    pickle.dump(X_train_sparse, file)
```

Figura 24. Creación de matriz dispersa para procesamiento de archivo voluminoso.

3.1.4. Selección de modelo, exploración de Hiperparámetros, Implementación del Modelo y Selección de métrica para evaluación de performance

Se realizó un análisis comparativo de varios algoritmos de aprendizaje automático para identificar la técnica más adecuada para el problema de clasificación binaria en cuestión. Inicialmente, se evaluaron modelos como Regresión Logística, Árboles de Decisión y Máquinas de Soporte Vectorial (SVM), teniendo en cuenta su capacidad para manejar datos de alta dimensionalidad y su eficacia en problemas similares. Sin embargo, se obtuvieron mejores resultados con métodos de gradient boosting, lo que llevó a

una exploración más profunda de algoritmos especializados en esta técnica, como XGBoost, LightGBM y CatBoost.

Finalmente, se seleccionó XGBoost debido a su robustez y eficiencia en el manejo de grandes volúmenes de datos, así como por su capacidad para optimizar múltiples funciones de pérdida y proporcionar modelos altamente precisos. La elección de XGBoost se justificó adicionalmente por sus mejores resultados en comparación con LightGBM y CatBoost. En pruebas de rendimiento, XGBoost obtuvo un AUC de 98,1 en un tiempo de 3 minutos con 34 segundos, superando a LightGBM (97,7) con un tiempo de 1 minuto y 40 segundos y CatBoost (97,8) con un tiempo de 7 minutos y 52 segundos. Esta superioridad en la métrica clave del área bajo la curva ROC (AUC) fue crucial para la optimización del rendimiento del modelo en términos de clasificación binaria. A pesar de no obtener el tiempo más rápido de los tres modelos, se optó por ir con el modelo más efectivo en cuanto a métrica teniendo en consideración que el volumen de datos a nivel computacional y la complejidad del problema en cuestión no amerita recortar tiempos en periodos tan cortos.

Algoritmo	AUC	Tiempo de Entrenamiento
XGBoost	98.1	3 minutos 34 segundos
LightGBM	97.7	1 minuto 40 segundos
CatBoost	97.8	7 minutos 52 segundos

Figura 25. Cuadro comparativo de performance de algoritmos de gradient boosting.

En la metodología aplicada, se llevó a cabo una exploración de hiperparámetros utilizando el enfoque de Búsqueda Aleatoria (Random Search), incorporando la librería scikit-learn y su clase RandomizedSearchCV. Este proceso implica la selección aleatoria de combinaciones de parámetros dentro de un espacio

predefinido, con el objetivo de identificar la configuración que maximiza la métrica de evaluación, en este caso, el área bajo la curva ROC.

Inicialmente, se definió un espacio de hiperparámetros que incluye la profundidad máxima del árbol (`max_depth`), el peso mínimo de los hijos (`min_child_weight`), la fracción de muestras para entrenar cada árbol (`subsample`), la fracción de columnas para entrenar cada árbol (`colsample_bytree`), y la tasa de aprendizaje (`eta`). Para cada parámetro, se especificó un rango de valores utilizando distribuciones estadísticas, permitiendo que `RandomizedSearchCV` seleccione valores al azar dentro de estos rangos.

Se configuró un modelo de XGBoost con la función objetivo `binary:logistic` y la métrica de evaluación `auc`, adecuado para problemas de clasificación binaria. El `RandomizedSearchCV` se estableció con 100 iteraciones, utilizando la métrica `roc_auc` para la selección del mejor modelo y validación cruzada de tres pliegues. Se ajustó esta configuración al conjunto de entrenamiento para evaluar diversas combinaciones de hiperparámetros (véase figura 26).

Finalmente, se extrajo el mejor conjunto de parámetros y la mejor puntuación AUC obtenida. El modelo resultante se evaluó sobre un conjunto de validación independiente para verificar su rendimiento general. Esta metodología asegura una búsqueda exhaustiva y aleatoria dentro del espacio de parámetros definido, favoreciendo la identificación de una configuración óptima para el modelo predictivo bajo estudio.

```

# Suponiendo que X_train_sparse y y_train ya están definidos y cargados
X_train, X_val, y_train, y_val = train_test_split(X_train_sparse,
y_train, test_size=0.2, random_state=42)

# Definir el espacio de hiperparámetros a explorar
param_dist = {
    'max_depth': stats.randint(3, 10),
    'min_child_weight': stats.randint(1, 6),
    'subsample': stats.uniform(0.5, 0.5),
    'colsample_bytree': stats.uniform(0.5, 0.5),
    'eta': stats.uniform(0.01, 0.3),
    'n_estimators': stats.randint(10, 50)
}

# Configurar el modelo XGBoost para usar con RandomizedSearchCV
clf = xgb.XGBClassifier(objective='binary:logistic', eval_metric='auc')

# Configurar RandomizedSearchCV
rs_clf = RandomizedSearchCV(clf, param_dist, n_iter=30,
scoring='roc_auc', cv=3, verbose=2, random_state=42, n_jobs=-1)

# Ejecutar RandomizedSearchCV
rs_clf.fit(X_train, y_train)

# Mejor modelo y parámetros
print("Mejor puntuación AUC: {:.3f}".format(rs_clf.best_score_))
print("Mejores parámetros: ", rs_clf.best_params_)

```

Figura 26. Implementación de algoritmo de random search para exploración de hiperparámetros.

En cuanto a la métrica utilizada para evaluar el rendimiento del modelo, se optó por el área bajo la curva ROC. Esta métrica es comúnmente utilizada en problemas de clasificación binaria y proporciona una medida del rendimiento del modelo al evaluar su capacidad para distinguir entre las clases positivas y negativas. Un valor de AUC cercano a 1 indica un modelo con un buen rendimiento, mientras que un valor cercano a 0.5 indica un rendimiento similar al azar.

La elección de AUC como métrica de evaluación se basó en su capacidad para proporcionar una evaluación integral del modelo en términos de sensibilidad y especificidad, lo que es importante para problemas de clasificación desbalanceada.

En la exploración de hiperparámetros realizada mediante el método de búsqueda aleatoria (Random Search), se obtuvo una configuración óptima de parámetros que maximiza la métrica de evaluación área bajo la curva (AUC) para el modelo de clasificación binaria basado en XGBoost. Los resultados de esta búsqueda mostraron que la mejor combinación de parámetros alcanzó una puntuación AUC de 0.981, indicando un excelente rendimiento del modelo en términos de su capacidad para distinguir entre las clases objetivo.

La configuración de parámetros que logró este rendimiento incluye:

- Una profundidad máxima (max_depth) de 6, lo que sugiere una complejidad moderada del modelo que equilibra bien el ajuste y la generalización.
- Un peso mínimo de los hijos (min_child_weight) de 1, indicando que el algoritmo fue sensible a las observaciones individuales sin causar sobreajuste.
- Una tasa de submuestreo (subsample) de aproximadamente 0.953, lo que permite que casi todas las observaciones participen en la construcción de cada árbol, contribuyendo así a la robustez del modelo.
- Una tasa de submuestreo por árbol (colsample_bytree) de aproximadamente 0.770, que proporciona una buena diversidad en las características utilizadas, ayudando a evitar el sobreajuste a las peculiaridades del conjunto de entrenamiento.
- Una tasa de aprendizaje (eta) de aproximadamente 0.252, que es relativamente alta, facilitando una convergencia rápida del modelo durante el entrenamiento.
- Un número de estimadores (n_estimators) de 48, optimizando el número de rondas de impulso para evitar el sobreajuste mientras se asegura una adecuada capacidad de aprendizaje.

Este conjunto de hiperparámetros proporciona un equilibrio entre la capacidad del modelo para capturar relaciones complejas en los datos y su habilidad para generalizar a datos no vistos, demostrado por la alta puntuación AUC obtenida en el conjunto de validación.

3.1.5. Interpretación de resultados

El modelo ha alcanzado un área bajo la curva ROC de aproximadamente 0.981 en el conjunto de prueba. Estas son métricas excepcionalmente altas que indican un excelente rendimiento del modelo en términos de su capacidad para distinguir entre las situaciones en las que se generará o no una orden de servicio. Un AUC cercano a 1.0 sugiere que el modelo tiene una gran precisión al clasificar los datos positivos y negativos correctamente.

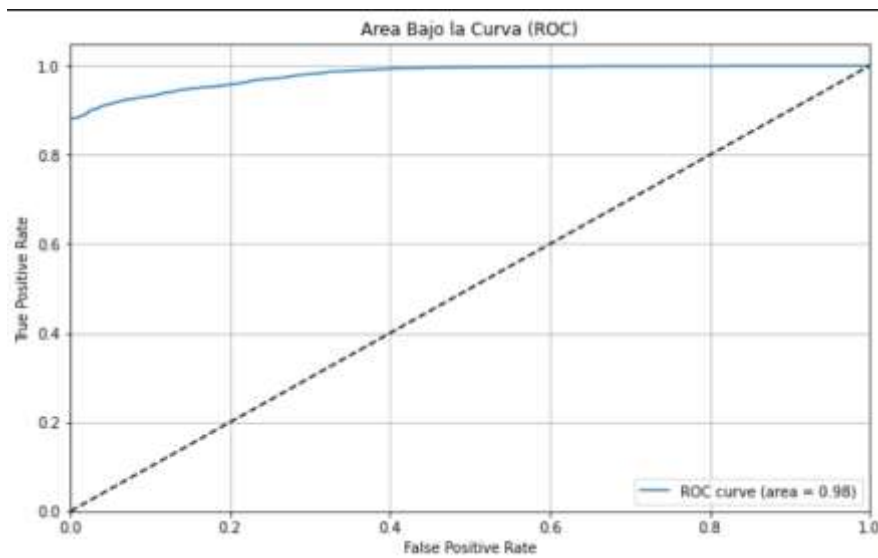


Figura 27. Área bajo la curva ROC resultante del modelo.

```
[30] train-auc:0.98354 eval-auc:0.96537
[31] train-auc:0.98388 eval-auc:0.96578
[32] train-auc:0.98419 eval-auc:0.96568
[33] train-auc:0.98441 eval-auc:0.96794
[34] train-auc:0.98470 eval-auc:0.96793
[35] train-auc:0.98499 eval-auc:0.96799
[36] train-auc:0.98517 eval-auc:0.96802
[37] train-auc:0.98538 eval-auc:0.97028
[38] train-auc:0.98563 eval-auc:0.97012
[39] train-auc:0.98571 eval-auc:0.96995
[40] train-auc:0.98586 eval-auc:0.97174
[41] train-auc:0.98600 eval-auc:0.97163
[42] train-auc:0.98613 eval-auc:0.97158
[43] train-auc:0.98620 eval-auc:0.97283
[44] train-auc:0.98635 eval-auc:0.97294
[45] train-auc:0.98644 eval-auc:0.97295
[46] train-auc:0.98652 eval-auc:0.97290
[47] train-auc:0.98667 eval-auc:0.97294
AUC en el conjunto de evaluación: 0.9806708526099397
Conteo de predicciones: {0: 335425, 1: 2719}
```

Figura 28. Resultados de AUC para entrenamiento de XGBOOST con hiperparametros específicos.

Interpretación en el Contexto Operativo:

- **Confianza en la Predicción:**

La alta puntuación AUC implica que el modelo es muy confiable para predecir si una combinación específica de lote y fecha resultará en una orden de servicio. Esto es crucial para la planificación y gestión de recursos, permitiendo a la empresa anticipar necesidades de mano de obra, materiales y otros recursos.

- **Planificación Basada en Predicciones:**

Con predicciones confiables, la gestión puede tomar decisiones informadas sobre la asignación de personal técnico y la programación de mantenimientos o revisiones, minimizando así los tiempos de inactividad y mejorando la eficiencia operativa.

- **Optimización de Recursos:**

Al predecir con precisión las órdenes de servicio, la empresa puede optimizar la utilización de sus recursos, asegurándose de que los técnicos estén disponibles cuando y donde se les necesite, sin incurrir en sobre-staffing o falta de personal en momentos críticos.

- **Intervenciones Proactivas:**

Este modelo permite realizar intervenciones proactivas al identificar anticipadamente las fechas y lotes que probablemente requieran atención. Esto puede ayudar a prevenir fallos o problemas antes de que ocurran, mejorando la satisfacción del cliente y la reputación de la empresa.

- **Feedback para Mejoras Continuas:**

La implementación de este modelo también ofrece una oportunidad para recopilar datos sobre su precisión y eficacia en condiciones reales de operación. Estos datos pueden ser utilizados para refinar y mejorar continuamente el modelo, adaptándolo a cambios en las condiciones de operación o en los patrones de demanda de servicio.

En conclusión, el modelo proporciona una herramienta valiosa para predecir órdenes de servicio con alta precisión, lo cual facilita una planificación eficaz y una mejor gestión de los recursos. La capacidad de anticipar la demanda de servicios no solo mejora la operativa diaria, sino que también contribuye a la estrategia general de la empresa en términos de mantenimiento preventivo y gestión de la carga de trabajo. Esto demuestra el potencial del modelado predictivo para transformar operaciones en entornos complejos, donde decisiones rápidas y basadas en datos pueden resultar en ahorros significativos y mejoras operativas.

3.2. Modelo de programación Lineal entera

En el modelo desarrollado, se busca minimizar el número de contrataciones a lo largo del horizonte de planificación. Esto se logra optimizando tanto el número como el momento de las contrataciones de técnicos, alineando estos elementos con la demanda proyectada y considerando los tiempos necesarios para la contratación y la capacitación.

Aunque un algoritmo goloso podría resolver el problema de optimizar la cantidad de contrataciones en el período, la elección de utilizar un modelo de programación lineal entera se fundamenta principalmente en la familiaridad con esta herramienta y su flexibilidad. La programación lineal entera no solo permite un manejo robusto y estructurado del problema, sino que también facilita la incorporación de nuevas restricciones en el futuro. Esta capacidad de adaptación es crucial para el modelo, permitiendo ajustes continuos que podrían surgir de cambios en las políticas de contratación o en las necesidades de capacitación. Al utilizar este enfoque, se asegura que el modelo pueda evolucionar y adaptarse a las necesidades cambiantes de la organización, manteniendo su relevancia y efectividad a largo plazo.

Para la realización del modelo, se considera un conjunto t que corresponde a cada período de planificación. El conjunto de todos los períodos t se denota como

$t \in \{1, 2, \dots, T\}$, donde T es el número total de meses en el horizonte de planificación.

3.2.1. Identificación de variables de decisión

VARIABLES DE DECISIÓN:

- X_t : Variable entera que indica el número de procesos de contratación que se inician en el período t . Cada unidad representa el inicio de un proceso de contratación que contribuirá técnicos operativos después de los períodos requeridos para su contratación y capacitación.
- Y_t : Variable entera que representa el número total de técnicos operativos disponibles en el período t , capaces de atender las órdenes de servicio. Esta cantidad se calcula tomando en cuenta los técnicos ya existentes y aquellos que se vuelven operativos tras completar los períodos de contratación y capacitación iniciados en períodos anteriores.

3.2.2. Parámetros del modelo

- ***Demanda_t***: Este parámetro representa el número estimado de órdenes de servicio que se deben atender en el período t . Es un valor crucial que define la carga de trabajo para la cual se requiere suficiente capacidad operativa.
- **TiempoContratación**: Define el número de períodos necesarios para completar el proceso de contratación de nuevos técnicos. Durante este tiempo, los técnicos contratados no están disponibles para trabajar.
- **TiempoCapacitación**: Este parámetro especifica el número de períodos que un técnico recién contratado requiere para estar completamente capacitado y listo para operar de manera efectiva. Solo después de este período, el técnico puede empezar a atender órdenes de servicio.
- **CapacidadTécnico**: Indica el número de órdenes de servicio que un técnico puede manejar por período. En este modelo, se especifica que un

técnico puede resolver 2 RMAs por día, lo que se traduce en una capacidad total de 44 RMAs por mes (considerando 22 días laborales).

- **Técnicos Existentes:** Número inicial de técnicos disponibles antes de iniciar cualquier contratación.

3.2.3. Función Objetivo

La función objetivo del modelo de programación lineal es minimizar el total de contrataciones (X_t) a lo largo del horizonte de planificación. Esta minimización refleja el esfuerzo por reducir los costos asociados a la contratación y capacitación de nuevo personal, teniendo en cuenta los tiempos necesarios para la contratación (*TiempoContratacion*) y capacitación (*TiempoCapacitacion*). La reducción en el número de contrataciones no solo disminuye los costos directos relacionados con estos procesos, sino también los costos indirectos asociados con la gestión de recursos humanos y la integración de nuevos empleados al flujo de trabajo existente.

La expresión matemática de la función objetivo se presenta de la siguiente manera:

$$\mathbf{Min} \mathbf{Z} = \sum_{t=1}^T \mathbf{X}_t$$

Donde:

- Z es el valor total de las contrataciones que se busca minimizar.
- X_t representa el número de técnicos contratados en el período t, tal como se define en la variable de decisión X_t
- T es el número total de períodos en el horizonte de planificación, relacionado con la longitud del vector de demanda ($Demanda_t$).

Esta función objetivo asegura que cada decisión de contratación se tome con el objetivo de satisfacer la demanda proyectada ($Demanda_t$), mientras se minimiza

el impacto económico a largo plazo, alineando las necesidades operativas con la disponibilidad efectiva de técnicos (Y_t).

3.2.4. Restricciones

- **Restricción de Satisfacción de la Demanda:** Cada mes, el número total de órdenes de servicio que pueden ser atendidas por los técnicos disponibles debe ser al menos igual a la demanda estimada para ese período. Matemáticamente, esto se expresa como:

$$Y_t \times \text{CapacidadTecnico} \geq \text{Demanda}_t, \forall_t$$

Donde Y_t representa el número de técnicos operativos disponibles en el período t (variable Y_t) y CapacidadTecnico es el número de órdenes de servicio que cada técnico puede manejar por período.

- **Restricción de Disponibilidad de Técnicos:** el número de técnicos operativos disponibles en cualquier mes se calcula considerando los nuevos técnicos que se han contratado y capacitado hasta ese momento. Los técnicos contratados en un mes t se vuelven operativos después de completar los períodos de contratación y capacitación. Matemáticamente, esto se expresa como:

$$Y_t = \text{TecnicosExistentes} \sum_{i=0}^{t-t_c-t_p} X_i, \forall_t \geq t_c + t_p$$

Para los primeros meses, antes de que se completen los periodos de contratación y capacitación:

$$Y_t = \text{TecnicosExistentes}, \forall_t < t_c + t_p$$

Donde:

- Y_t es el número de técnicos operativos disponibles en el período t .

- X_i es el número de técnicos contratados en el período i .
- t_c es el tiempo de contratación, el número de períodos que toma completar el proceso de contratación.
- t_p es el tiempo de capacitación, el número de períodos necesarios para capacitar a los técnicos antes de que estén listos para trabajar.

3.2.5. Implementación en Python

Como se puede observar en la figura 29, se inicia importando la biblioteca PuLP y se define el problema de optimización como un problema de minimización. Esto se realiza mediante la creación de una instancia de LpProblem donde se especifica el nombre del problema y el objetivo de minimización.

```
import pulp as lp
# Creación del modelo de optimizacion
problema = pulp.LpProblem("Minimizacion_Contrataciones", lp.LpMinimize)
```

Figura 29. Creación de modelo de programación lineal via biblioteca Pulp.

Las variables de decisión se definen utilizando LpVariable.dicts, que crea un diccionario de variables. Las contrataciones (X_t) son variables enteras que indican el número de técnicos contratados en cada periodo, mientras que los técnicos disponibles (y_t) son también variables enteras que representan el número de técnicos operativos en cada periodo (Véase figura 30).

```
# Variables de decisión
X = pulp.LpVariable.dicts("X", range(meses), lowBound=0,
cat='Integer') # Contrataciones por mes
Y = pulp.LpVariable.dicts("Y", range(meses), lowBound=0,
cat='Integer') # Técnicos operativos por mes
```

Figura 30. Creación de variables de decisión.

La función objetivo se define para minimizar el total de contrataciones a lo largo del horizonte de planificación. Como se muestra en la figura 31, esto se logra sumando todas las variables de contrataciones y utilizando el operador de adición de PuLP para establecer la función objetivo del modelo.

```
# Función objetivo
problema += pulp.lpSum([X[t] for t in range(meses)]),
"Minimizacion_Total_Contrataciones"
```

Figura 31. Generación de función objetivo.

Las restricciones se añaden al modelo para asegurar que se satisfagan las condiciones de demanda, capacidad y disponibilidad de técnicos (Véase figura 32).

```
# Restricciones

# Restricción de disponibilidad de Técnicos
for t in range(meses):
    if t < tiempo_contratacion + tiempo_capacitacion:
        problema += Y[t] == tecnicos_existentes,
f"Disponibilidad_Tecnicos_{t}"
    else:
        problema += Y[t] == tecnicos_existentes + pulp.lpSum([X[i] for
i in range(t - tiempo_contratacion - tiempo_capacitacion + 1)]),
f"Disponibilidad_Tecnicos_{t}"

# Restricción de satisfacción de la demanda
for t in range(meses):
    problema += Y[t] * capacidad_tecnico >= demandas[t],
f"Demanda_Satisfecha_{t}"
```

Figura 32. Ensamble de restricciones para modelo de programación lineal.

Finalmente, El modelo se resuelve utilizando el solucionador disponible en PuLP (Figura 33). Los resultados se obtienen accediendo al valor de las variables luego de la resolución.

```
#Resolver el problema
problema.solve()

# Verificar el estado de la solución
if pulp.LpStatus[problema.status] == 'Optimal':
    print("Estado: Óptimo")

    print("Contrataciones por mes:")
    for t in range(meses):
        print(f"Mes {t + 1}: {pulp.value(X[t])}")

    print("Técnicos operativos por mes:")
    for t in range(meses):
        print(f"Mes {t + 1}: {pulp.value(Y[t])}")
else:
    print("No se encontró solución óptima.")
```

Figura 33. Aplicación de función de resolución de modelo de programación lineal.

3.2.6. Solución y análisis del problema:

Durante la elaboración de esta tesis, específicamente en la fase de planteamiento del problema inicial, se realizó una simulación detallada para evaluar el impacto potencial de la contratación de personal adicional en la mejora de la atención al cliente en la empresa GEMATEC SRL. Esta simulación, cuyos detalles se encuentran expuestos en el Capítulo II, demostró que un incremento en el número de técnicos podría resultar en una mejora significativa en la eficiencia del servicio postventa para ese momento en particular.

A raíz de los resultados obtenidos de la simulación y tras presentar el planteamiento a la gerencia general de GEMATEC SRL, la empresa decidió incrementar el staff disponible para el servicio postventa. Esta decisión resultó en un sobrestaffing en el equipo, situación que persiste hasta el día de hoy. Dicho sobrestaffing proporciona un contexto desfavorable para la correcta evaluación del modelo de optimización desarrollado en esta tesis.

Con la finalidad de mostrar la efectividad del modelo construido, se optó por utilizar valores ficticios para los parámetros en el modelo de programación lineal

presentado. Esta decisión permite ilustrar cómo el modelo puede adaptarse y optimizar la asignación de recursos bajo diferentes escenarios de demanda y capacidad. Sin embargo, es importante recalcar que el modelo tiene la capacidad de incorporar valores reales resultantes del modelo de machine learning también descrito en este documento, para construir una optimización más ajustada a la realidad operativa de la empresa.

Esta estrategia no solo demuestra la flexibilidad y aplicabilidad del modelo en distintos contextos, sino que también subraya la importancia de una toma de decisiones basada en datos y simulaciones cuidadosamente analizadas. Al emplear los parámetros con valores ficticios, se asegura que el enfoque y las conclusiones derivadas de esta investigación sean claras y transferibles a otras situaciones similares, proporcionando un marco robusto para la toma de decisiones estratégicas en la gestión de recursos humanos en el ámbito del servicio postventa.

A continuación, en la figura 34 se detallan los parámetros simulados y utilizados, además, se aclaran los métodos de obtención de los mismos.

Parámetro	Valor
demandas	660, 660, 660, 660, 660, 660, 800, 900, 1000, 1100, 1200, 1500
tiempo_contratacion	2
tiempo_capacitacion	4
capacidad_tecnico	44
tecnicos_existentes	15
meses	12 (Longitud del arreglo de demandas)

Figura 34. Parámetros simulados para prueba de modelo de programación lineal.

- Demandas: Los datos de demanda mensual fueron simulados para mostrar un incremento substancial mes a mes que no es representativo del comportamiento de la base instalada real de GEMATEC SRL pero que sí pone a prueba el rendimiento real del modelo de programación lineal entera. De cualquier manera, basándose en un histórico y proyecciones del modelo de machine learning. Este modelo puede analizar las órdenes de servicio pasadas y, en combinación con el forecast de ventas de la

empresa, estima la cantidad de órdenes de servicio que se esperan para los próximos 15 meses. Este enfoque asegura que el modelo de optimización se alinea con las expectativas reales de la carga de trabajo futura.

- **Tiempo de Contratación:** Determinado por el departamento de Recursos Humanos, este parámetro refleja el número promedio de meses necesarios para completar el proceso de contratación de un nuevo técnico. Incluye el tiempo desde la publicación de la vacante hasta la incorporación efectiva del personal.
- **Tiempo de Capacitación:** Basado en un análisis detallado de las intervenciones de servicio por técnico y las certificaciones obtenidas, este tiempo representa los meses requeridos para que un nuevo técnico esté completamente capacitado y operativo. Este análisis permite asegurar que los técnicos no solo sean contratados, sino que alcancen un nivel de rendimiento que cumpla con los estándares de calidad de GEMATEC SRL.
- **Técnicos Existentes:** Este valor proviene del número actual de técnicos disponibles en el staff de servicio postventa. Es un dato a poner a mano cada vez que se corra el modelo dado que refleja la capacidad de servicio inmediata antes de cualquier nueva contratación.
- **Capacidad Técnico:** Definida por la gerencia general, esta cifra indica el número máximo de órdenes de servicio que un técnico debería gestionar eficientemente en un mes. Este parámetro se establece considerando la productividad deseada y las normas de calidad del servicio.

Cada uno de estos parámetros no solo es esencial para la construcción y validación del modelo de optimización, sino que también refleja un enfoque basado en datos y un compromiso con la mejora continua y la eficiencia operativa en GEMATEC SRL. Al simular estos datos en el contexto académico de esta tesis, se proporciona un marco robusto para la evaluación del modelo, asegurando que los resultados sean transferibles y aplicables en un escenario real de operación.

De esta manera y con los datos que fueron simulados, en la figura 35 se muestra la solución óptima proporcionada por el modelo:

Mes	Contrataciones por mes	Técnicos disponibles por mes
Enero	4.0	15.0
Febrero	2.0	15.0
Marzo	2.0	15.0
Abril	2.0	15.0
Mayo	3.0	15.0
Junio	7.0	15.0
Julio	0.0	19.0
Agosto	0.0	21.0
Septiembre	0.0	23.0
Octubre	0.0	25.0
Noviembre	0.0	28.0
Diciembre	0.0	35.0

Figura 35. Solución óptima del problema.

Los resultados del modelo de programación lineal revelan detalles importantes sobre la estrategia de contratación y disponibilidad de técnicos en la empresa GEMATEC SRL, a lo largo de un horizonte de planificación de 15 meses. Analicemos estos resultados en detalle:

Análisis de Contrataciones por Mes

- **Meses 1 a 6:** Se observa una serie de contrataciones escalonadas que aumentan gradualmente hasta el mes 6, lo cual indica una anticipación estratégica para enfrentar un aumento en la demanda que comienza a crecer significativamente a partir del mes 7. El pico de contrataciones en el mes 6 (7 técnicos) parece estar diseñado para prepararse para este incremento.
- **Meses 7 a 9:** Aunque la demanda sigue aumentando, el número de contrataciones disminuye a 2. Esto sugiere que las contrataciones realizadas en los meses anteriores están comenzando a tener efecto una vez que los técnicos están plenamente capacitados y disponibles para trabajar.
- **Meses 10 a 12:** No se realizan nuevas contrataciones en estos meses, lo que indica que el modelo anticipa que el número de técnicos disponibles

será suficiente para cubrir la demanda proyectada hasta el final del período de planificación.

Análisis de Técnicos Disponibles por Mes

- **Meses 1 a 6:** La cantidad de técnicos se mantiene constante en 15, lo cual coincide con el número de técnicos existentes. Esto muestra que, durante los primeros meses, no hay nuevos técnicos disponibles más allá del personal existente, debido a los tiempos de contratación y capacitación.
- **Mes 7 en adelante:** A partir de este mes, se observa un incremento gradual en el número de técnicos disponibles, lo cual es consecuencia directa de las contrataciones realizadas en los meses anteriores. Este incremento es continuo y sustancial, lo que refleja tanto la acumulación de efectos de las contrataciones realizadas como el tiempo necesario para que los técnicos contratados estén operativos.

Conclusiones del Análisis

- **Planificación Eficiente:** El modelo ha demostrado una planificación efectiva, donde las contrataciones se realizan de manera anticipada y en respuesta a la demanda proyectada, teniendo en cuenta los tiempos de contratación y capacitación.
- **Cobertura de Demanda:** El aumento en el número de técnicos disponibles está alineado con los incrementos en la demanda, lo que sugiere que el modelo es capaz de ajustar adecuadamente los recursos humanos para evitar tanto el sobrestaffing como la falta de personal.
- **Sostenibilidad:** La ausencia de nuevas contrataciones en los últimos meses del horizonte de planificación puede indicar una sostenibilidad en la operación, asumiendo que no hay cambios significativos en las condiciones de demanda o en la capacidad técnica requerida.

Este análisis muestra que el modelo de optimización utilizado es robusto y sensible a las necesidades operativas de la empresa, permitiendo una gestión

eficiente de los recursos técnicos en respuesta a las dinámicas de demanda y capacidad.

Para evaluar la robustez y flexibilidad del modelo de Programación Lineal Entera (PLE) desarrollado, se realizaron simulaciones bajo diferentes escenarios de demanda. Específicamente, se consideraron tres escenarios: el escenario base con la demanda original, un escenario con una demanda incrementada en un 15% a partir del mes 7 y un escenario con una demanda reducida en un 15% a partir del mismo mes. Este análisis permite demostrar cómo el modelo se adapta a variaciones en la demanda sin comprometer la capacidad operativa.

Escenario Base

El escenario base utiliza la demanda proyectada original sin modificaciones. En este caso, se planificaron las contrataciones necesarias para asegurar que se cubra la demanda mensual de órdenes de servicio. Inicialmente, con 15 técnicos operativos, el modelo ajusta el número de contrataciones de manera progresiva para alcanzar 35 técnicos operativos en el mes 12, asegurando la satisfacción de la demanda sin sobrepasar la capacidad de los técnicos disponibles.

Escenario con Demanda Incrementada

En este escenario, la demanda mensual se incrementa en un 15% a partir del mes 7. El modelo de PLE se ajusta para satisfacer esta demanda aumentada, lo que resulta en un mayor número de contrataciones adicionales. Como se observa en el gráfico, el número de técnicos operativos necesarios incrementa más rápidamente a partir del mes 7 en comparación con el escenario base, alcanzando 40 técnicos en el mes 12. Este aumento demuestra que el modelo puede responder eficazmente a incrementos en la demanda, garantizando la disponibilidad de recursos suficientes.

Escenario con Demanda Reducida

Para el escenario de demanda reducida, la demanda mensual se disminuye en un 15% a partir del mes 7. En este caso, el modelo ajusta las contrataciones de manera que se requieran menos técnicos operativos en comparación con el escenario base. En el gráfico, se observa que el número de técnicos operativos se mantiene más bajo, alcanzando aproximadamente 30 técnicos en el mes 12. Esto demuestra la capacidad del modelo para reducir eficientemente el número de contrataciones cuando la demanda es menor, evitando así una sobreasignación de recursos.

En la siguiente figura (numero 36) se ilustran los resultados obtenidos:

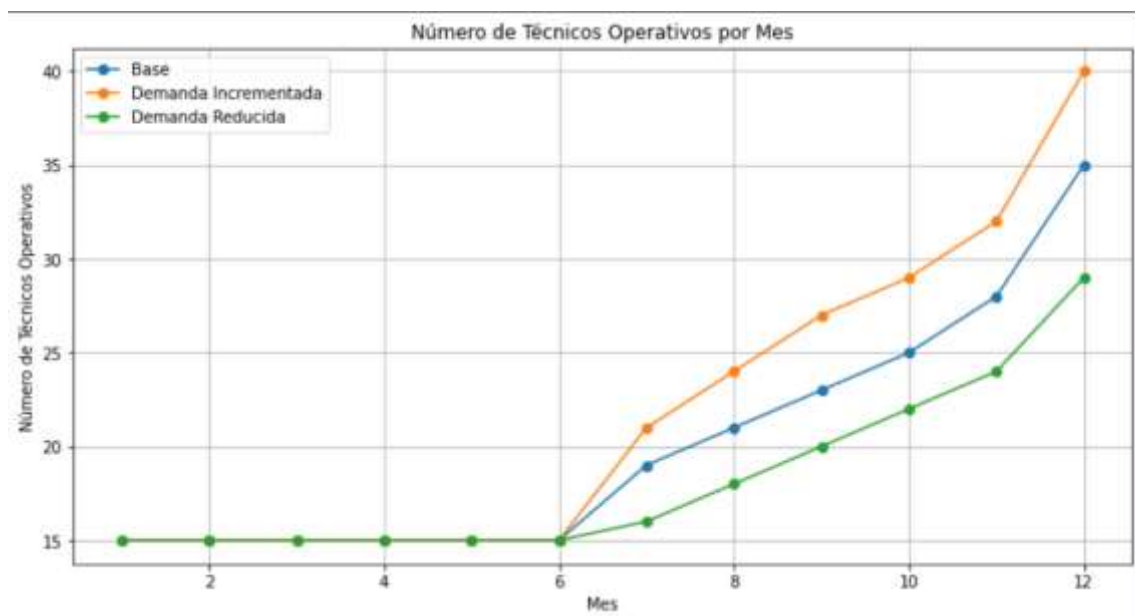


Figura 36. Gráfico de resultados de modelo de PLE para diferentes escenarios de demanda.

El análisis de los tres escenarios demuestra la flexibilidad y robustez del modelo de PLE. El modelo se adapta eficazmente a variaciones en la demanda, incrementando o disminuyendo el número de contrataciones de técnicos según sea necesario para asegurar que la demanda mensual se cumpla de manera eficiente. Esta capacidad de adaptación es crucial para la planificación operativa,

ya que permite responder rápidamente a cambios en las condiciones del mercado, asegurando siempre la disponibilidad adecuada de recursos técnicos para satisfacer la demanda de órdenes de servicio.

Este análisis proporciona evidencia sólida de que el modelo de PLE es una herramienta eficaz para la gestión y planificación de recursos en un entorno dinámico y variable, asegurando que la organización pueda mantener su capacidad operativa y cumplir con la demanda bajo diferentes escenarios.

3.2.7. Detalle de funcionamiento:

Para ilustrar el funcionamiento del modelo de programación lineal entera desarrollado por este trabajo de tesis, consideremos un ejemplo práctico que sigue los pasos del modelo en un período de planificación.

Supuestos Iniciales:

- Demandas mensuales ($Demanda_t$): [660, 660, 660, 660, 660, 660, 800, 900, 1000, 1100, 1200, 1500]
- Tiempo de contratación (TiempoContratacion): 2 meses
- Tiempo de capacitación (TiempoCapacitacion): 4 meses
- Capacidad de técnico (CapacidadTecnico): 44 órdenes de servicio por mes
- Técnicos existentes (TecnicosExistentes): 15 técnicos
- Horizonte de planificación: 12 meses

Paso 1: Inicialización

- En el mes 1, comenzamos con 15 técnicos.

Paso 2: Evaluación de la Demanda y Capacidad Disponible

- Para el mes 1, la demanda es de 660 órdenes de servicio.
- Con 15 técnicos, la capacidad disponible es $15 \times 44 = 660$ órdenes de servicio, lo cual es suficiente para satisfacer la demanda.

Paso 3: Planificación de Contrataciones Futuras

- Aunque la capacidad en el mes 1 es suficiente, debemos considerar los períodos futuros debido al tiempo de contratación y capacitación.
- Mirando hacia adelante 6 meses (2 meses de contratación + 4 meses de capacitación), para el mes 7 la demanda es de 800 órdenes de servicio.
- Con 15 técnicos, la capacidad sería $15 \times 44 = 660$ órdenes de servicio, lo que no es suficiente para cubrir la demanda futura.
- Por lo tanto, necesitamos planificar la contratación de más técnicos para asegurarnos de que estén operativos en el mes 7.

Paso 4: Determinación del Número de Contrataciones Necesarias

- Para cubrir la demanda del mes 7 (800 órdenes) con una capacidad de 44 órdenes por técnico, se necesitan al menos $\lceil 800/44 \rceil = 19$ técnicos.
- Actualmente tenemos 15 técnicos, por lo que necesitamos contratar $19 - 15 = 4$ técnicos en el mes 1 para que estén operativos en el mes 7.

Paso 5: Iteración Mensual

- Repetimos este proceso para cada mes del horizonte de planificación.
- En el mes 2, evaluamos la demanda para el mes 8, considerando que ya habremos contratado 4 técnicos adicionales en el mes 1.
- Continuamos este proceso para asegurar que cada mes se satisfaga la demanda proyectada con el mínimo número de contrataciones.

Resultados del Modelo

- El modelo determina el número óptimo de contrataciones mensuales (X_t) y calcula la disponibilidad de técnicos operativos (Y_t) en cada mes, asegurando que se cumplan todas las demandas de servicio proyectadas.
- Por ejemplo, los resultados pueden mostrar que se deben contratar 4 técnicos en el mes 1, 2 técnicos en el mes 2, y así sucesivamente, para

mantener la capacidad operativa adecuada a lo largo del horizonte de planificación.

Este enfoque garantiza que las decisiones de contratación se tomen de manera optimizada, considerando tanto las demandas futuras como los tiempos necesarios para la contratación y capacitación de nuevos técnicos.

3.2.8. Escalabilidad y posibles mejoras

El modelo de programación lineal entera desarrollado para la planificación de contrataciones es escalable y puede adaptarse a diferentes horizontes de planificación, estructuras de demanda y tamaños de equipo. Algunas características clave que facilitan la escalabilidad incluyen:

1. **Horizonte de Planificación Variable:** El modelo puede ajustarse fácilmente para diferentes números de períodos (meses) en el horizonte de planificación, simplemente modificando el parámetro T .
2. **Capacidad de Adaptación a Cambios en la Demanda:** El modelo puede manejar variaciones en la demanda mensual ($Demanda_t$) sin necesidad de reestructurarlo, lo que permite su uso en entornos con fluctuaciones en la carga de trabajo.
3. **Tiempos de Contratación y Capacitación Flexibles:** Los parámetros TiempoContratacion y TiempoCapacitacion pueden ajustarse según las necesidades específicas de la organización, permitiendo que el modelo sea útil en diferentes contextos operativos.
4. **Tamaño del Equipo Inicial:** El parámetro TecnicosExistentes se puede modificar para reflejar el tamaño inicial del equipo, lo que permite que el modelo se adapte a organizaciones con diferentes capacidades iniciales.

Posibles Mejoras del Modelo

Aunque el modelo actual es efectivo en optimizar la planificación de contrataciones, existen varias mejoras y ajustes que podrían hacerlo más robusto y adaptable:

1. Costos Variables de Contratación y Capacitación:

- Incorporar costos variables asociados a la contratación y capacitación en diferentes períodos permitiría evaluar no solo la cantidad, sino también el costo asociado a las decisiones de contratación.
- Esto podría incluir costos adicionales en períodos de alta demanda o costos reducidos en períodos de baja demanda.

2. Fluctuaciones en la Demanda:

- Incorporar escenarios de variación en la demanda (picos y valles) y desarrollar un modelo robusto que optimice las contrataciones bajo diferentes escenarios de demanda, mejoraría la flexibilidad y capacidad de respuesta.
- Podrían utilizarse técnicas de análisis de sensibilidad para entender mejor cómo afectan las fluctuaciones de demanda a las decisiones de contratación.

3. Limitaciones Presupuestarias:

- Incluir restricciones presupuestarias que limiten el número total de contrataciones en un período dado obligaría al modelo a encontrar soluciones que optimicen el uso de recursos financieros.
- Esto podría involucrar un presupuesto mensual o anual que no debe ser excedido.

4. Parámetros de Desempeño de Técnicos:

- Incluir variaciones en la capacidad de técnicos basadas en su experiencia y habilidades permitiría una planificación más detallada y precisa.

- Técnicos con diferentes niveles de experiencia podrían tener diferentes capacidades de manejo de órdenes de servicio, lo cual se puede modelar mediante diferentes parámetros de capacidad.

5. Incorporación de Desempeño Histórico:

- Utilizar datos históricos de desempeño y demanda para predecir mejor las necesidades futuras y ajustar el modelo en consecuencia.
- Modelos predictivos pueden integrarse para mejorar la precisión de las estimaciones de demanda.

Estas mejoras no solo harían el modelo más robusto y adaptable a diferentes escenarios y restricciones operativas, sino que también permitirían una optimización más detallada y efectiva de las decisiones de contratación, contribuyendo a una mejor gestión de recursos y a la eficiencia operativa general.

CAPITULO IV – CONCLUSIONES GENERALES

En la culminación de este trabajo de tesis, se han congregado una serie de conocimientos adquiridos a través del Máster en Management y Analytics. Esta confluencia de disciplinas ha permitido abordar con rigurosidad y perspectiva analítica los desafíos presentados por GEMATEC SRL en la optimización de su equipo de servicio postventa. A continuación, se desarrolla una reflexión comprensiva sobre la experiencia de investigación, las limitaciones encontradas, las potenciales mejoras y el impacto del programa académico en la realización de este estudio.

4.1. Conocimientos adquiridos y aplicados

A través del Máster en Management y Analytics, se adquirieron competencias fundamentales en análisis cuantitativo y cualitativo, permitiendo abordar problemas complejos con herramientas avanzadas de gestión de datos y optimización. El dominio de la programación lineal y el machine learning se convirtió en el eje central de esta investigación, proporcionando el soporte necesario para el desarrollo de un modelo de optimización robusto y sensible a las variables que inciden en la operativa de GEMATEC SRL.

La capacidad de integrar conocimientos de gestión empresarial y analítica avanzada ha sido crucial para traducir las necesidades operativas de la empresa en parámetros cuantificables y modelables. Esto se refleja en la habilidad para formular hipótesis, recoger y procesar datos de múltiples fuentes, y diseñar estrategias que respondan a los desafíos identificados.

4.2. Limitaciones del estudio

Si bien el modelo de optimización desarrollado representa un avance significativo, existen limitaciones inherentes a este tipo de investigación. La dependencia de datos precisos y actualizados puede ser un obstáculo, especialmente en escenarios volátiles donde la demanda y las condiciones de mercado cambian rápidamente. Asimismo, la simulación de parámetros y la utilización de valores ficticios, aunque necesarios para demostrar la flexibilidad y

aplicabilidad del modelo, pueden distanciarse de la realidad operacional compleja de la empresa.

4.3. Futuras mejoras y dirección de la investigación

Mirando hacia el futuro, se identifican oportunidades para la expansión de la investigación actual. La integración de modelos predictivos más sofisticados para la demanda, el uso de algoritmos de optimización en tiempo real y la aplicación de técnicas de machine learning para refinar continuamente las estimaciones de parámetros podrían enriquecer significativamente el modelo. Además, la exploración de modelos de simulación estocástica puede ofrecer una mayor comprensión de la variabilidad y el riesgo asociado a las decisiones de contratación y capacitación.

4.4. Aplicabilidad del trabajo

La aplicación de la herramienta desarrollada en el presente trabajo de tesis conlleva al día de hoy a una revisión trimestral de parte del responsable del departamento de postventa tanto del presupuesto de ventas como de los nuevos negocios propuesto por el área comercial de la empresa, pudiendo de esta manera adaptar la disponibilidad de su staff a las necesidades operativas que van a surgir de la actividad económica de la empresa. Con esta responsabilidad, el responsable de postventa puede acceder al sistema de gestión de la empresa, realizar un export de la base de datos, la cual, debe ser cargada al script que realiza el entrenamiento del algoritmo de predicción de demanda, una vez con el algoritmo entrenado con los nuevos datos, se estimara nuevamente el AUC (ROC) y se extraerá la tabla de resultados la cual será utilizada como dato de entrada en conjunto con los tiempos promedio de capacitación y contratación para poder correr el modelo de programación lineal entera que devengara en la toma de decisiones de staffing del área.

4.5. Posibles líneas de trabajo emergentes

Actualmente, se está desarrollando un proyecto de venta proactiva de mantenimientos preventivos por parte del departamento de postventa, basándose en la información resultante del modelo de XGBoost para predicción de demanda. En este proyecto se han podido apreciar comportamientos

característicos de lotes únicos pertenecientes a clientes con equipamiento específico. Con esta herramienta, es posible presentar a los clientes resúmenes sobre el rendimiento técnico de su equipo durante un período particular, mostrando comparativas entre una simulación del comportamiento de su lote único y un lote similar que recibe el mantenimiento adecuado. Además, se puede realizar una simulación de cómo se comportaría su propio lote si recibiera un servicio de mantenimiento adecuado, permitiendo mostrar al cliente la cantidad de tiempo de inactividad que se podría haber evitado en el año y el beneficio económico que tendría para su propia operación.

4.6. Conclusión final

El Máster en Management y Analytics ha sido instrumental en la realización de esta tesis, proporcionando no solo la base teórica sino también la confianza práctica para abordar los problemas de GEMATEC SRL. Las habilidades en gestión y analítica se han traducido en la capacidad de realizar un análisis crítico y exhaustivo, manejar grandes volúmenes de datos y aplicar técnicas estadísticas avanzadas, todas fundamentales para la comprensión y resolución del desafío presentado.

El desafío de GEMATEC SRL no solo era operativo sino también estratégico, donde se buscaba una solución que balanceara eficiencia y eficacia. La confluencia de conocimientos en negocios y analytics permitió abordar este desafío de manera holística, considerando no solo los números sino también el capital humano, la estrategia empresarial y la visión a largo plazo.

El viaje a través de esta tesis ha sido un reflejo de la trayectoria en el Máster, una que ha estado repleta de aprendizaje, descubrimiento y aplicación práctica. A medida que esta etapa académica llega a su fin, se deja abierta la puerta a futuras investigaciones y al continuo desarrollo profesional, en el cual los conocimientos y habilidades adquiridos servirán como sólidos cimientos para los retos venideros.

Bibliografía

1. Güçlükol, S. (2020). *Forecasting Customer Service Demand by Machine Learning with Real Life Implementation*. [Tesis de Maestría, Sabanci University]. Sabanci University Research Database. https://research.sabanciuniv.edu/id/eprint/41185/1/10346266_Guclukol_Simge.pdf
2. Will, P., Krpan, D., & Lordan, G. (2022). *People versus machines: Introducing the HIRE framework*. *Artificial Intelligence Review*, 56, 1071-1100. <https://doi.org/10.1007/s10462-022-10193-6>
3. Zohdi, M., Rafiee, M., Kayvanfar, V., & Salamiraad, A. (2022). *Demand forecasting based machine learning algorithms on customer information: An applied approach*. *International Journal of Information Technology*. <https://doi.org/10.1007/s41870-022-00875-3>
4. Goretzko, D., & Israel, L. S. F. (2022). *Pitfalls of Machine Learning-Based Personnel Selection: Fairness, Transparency, and Data Quality*. *Journal of Personnel Psychology*, 21(1), 37-47. Department of Psychology, Ludwig-Maximilians-Universität München, Munich, Germany. <https://doi.org/10.1027/1866-5888/a000287>
5. Pessach, D., Singer, G., Avrahami, D., Ben-Gal, H. C., & Shmueli, E. (2022). *Employees recruitment: A prescriptive analytics approach via machine learning and mathematical programming*. *International Journal of Information Technology*. <https://www.sciencedirect.com/science/article/pii/S0167923620300452>
6. Smelyakov, K., Hurova, Y., & Osieivskyi, S. (2023). *Analysis of the Effectiveness of Using Machine Learning Algorithms to Make Hiring Decisions*. In *Proceedings of the 7th International Conference on Computational Linguistics and Intelligent Systems (COLINS-2023)*. CEUR Workshop Proceedings. <https://ceur-ws.org/Vol-3387/paper7.pdf>