

Reportes Individuales por Fase

Apartado de Reportes Individuales por Fase

@Kevin Ramirez Luna

- **Comprensión del Negocio**

Mi principal contribución en esta fase fue participar hacerle preguntas a los especialistas del CAETEC, apoyar en la validación de que los objetivos de negocio estuvieran conectados con necesidades reales del rancho y no solo con intuiciones del equipo técnico. Además, contribuí en la elaboración de las matrices de riesgos, entendimiento de las vacas y de sus curvas de producción. Ayude en elaboración de documentos que satisfagan al socio formador como el One page.

- **Comprensión de los Datos**

En esta fase estuve ayudando a comprender como opera el CAETEC para la extracción de sus datos en el robot de ordeña. Comprendí que sin un entendimiento profundo del contexto de los datos, cualquier análisis queda incompleto y no se podría idear un primer acercamiento a las necesidades existentes en el CAETEC, que siempre se necesita realizar un ETL/EDA

Una de mis contribuciones fue revisar la consistencia entre distintas fuentes, identificar qué datasets podían integrarse y cuáles no, ver que acercamientos podrían ser buenos buscando papers en sitios web para agarrar inspiración y ver si las decisiones que tomábamos ya estaban estudiadas o implementadas. Ayude en la comprensión de los datos mediante graficas, para conocer como estaban distribuidos y que tan completa era la información que teníamos y documente todos mis hallazgos en los reportes de entendimiento de los datos.

- **Preparación de los Datos**

En esta fase implementé un merge de los datos iniciales para analizarlos en volumen y poder usarlos en nuestros modelos durante la primera y parte de la segunda iteración de modelado. Me di cuenta de que preparar datos implica tomar decisiones sobre qué conservar, qué descartar y qué transformar.

Mi participación se centró en limpiar estos datos y en documentar la justificación detrás de cada decisión de eliminación o remplazo de atributos. También ayudé a estandarizar formatos de fecha, corregir duplicados y preparar datasets que alimentaron las iteraciones realizadas.

Finalmente, ayude a mantener una comunicación efectiva con el CAETEC para asistir al rancho con el fin de asistir por primera y segunda vez viendo que los

resultados de las primeras iteraciones no eran validas

- **Modelado**

Viendo que los datos necesarios no eran completos y no se alineaban con el nuevo enfoque que planteamos en la 3ra iteración, después de que compañeros prepararan nuevos datos para la 3ra iteración, apoye en la implementación de modelos de Deep Learning (MLP) en función de cumplir los objetivos de negocio y crear el benchmark final con la comparativa de todos los modelos.

- **Evaluación**

En esta fase como tal no apoye mucho, solamente en la redacción de la documentación asegurándose que el modelo seleccionado cumpliera con los objetivos establecidos y nuestro modelo cumpliera los objetivos de minería de datos para realizar un benchmark

- **Deployment**

Mi aporte en esta fase fue pequeña. ayudando en la redacción del reporte final para sintetizar nuestro proyecto y algunas de las subfases de CRISP-DM, como también, junto con mi compañero Cristian realice el manual de usuario para que las personas que usen nuestro proyecto tengan una visión clara y útil de como usar la herramienta.

@Cristian Chávez Guía

- **Comprensión del Negocio**

En esta primera parte de la metodología, apliqué mis conocimientos de ingeniería de software aprendidos en anteriores semestres, ya que al inicio el socio formador llegó sin una necesidad, solo llegó con datos y nos explicó el proceso que hace el CAETEC referente a las vacas. A partir de eso identificamos áreas de oportunidad para el proyecto, con las cuales podríamos trabajar y, aplicando los principios de mejora continua, nuestro enfoque fue reducir el tiempo de los ciclos de retroalimentación. Apoyé en la definición de criterios de éxito, que eventualmente nos servirían para evaluar el proyecto; también realicé junto con Kevin el one page, apoyé en la definición de la matriz de stakeholders, ayudé a definir riesgos, planes de contingencia y de mitigación, investigué y pregunté a la doctora Lupita ciertas cuestiones sobre cómo se manejaban las cosas en el rancho del CAETEC, definí ciertas partes del background del proyecto con todas mis notas que tomé de las visitas al CAETEC y en conjunto con mis compañeros definimos criterios de éxito y objetivos de negocios.

- **Comprensión de los Datos**

Como tal, no realicé directamente la exploración de los datos en los Google Colab;

sin embargo, mi trabajo en esta fase fue documentar ciertos notebooks que realizaron mis compañeros. Mi aprendizaje fue sobre todo con las clases y workshops del profesor Doctor Dorantes, en donde, por medio del uso de gráficos, exploración de distribuciones, matrices de correlación, pude comprender el gran valor de esta fase y toda la información que podemos extraer de nuestros datos.

- **Preparación de los Datos**

Con la idea en mente de realizar predicciones de la producción de leche a futuro, realicé junto con mis compañeros Ángel y Kevin el data preparation para un modelo de series de tiempo que fue nuestra segunda iteración de esta fase. También documenté toda esta iteración del data preparation y con los resultados hicimos la segunda iteración de modelado.

- **Modelado**

Para la etapa de modelado, realice un modelo de series de tiempo de LSTM, el cual, por temas de viabilidad, no fue posible utilizarlo para pasar a la siguiente etapa, puesto que el enfoque era difícil de realizar, no era viable y no se podía determinar que realmente el modelo nos pudiera ayudar a cumplir con los objetivos del proyecto. Esto se debió a la complejidad de intentar predecir la producción de una vaca, ya que estas pueden llegar a ser muy aleatorias en su producción; no encontramos patrones que ayudaran a realizar una predicción acertada. Por ende, este modelo se deprecó y ya no se utilizó más adelante.

Para la tercera iteración, realicé un XGBoost como parte del benchmark, ya que nos interesaba comparar entre modelos. Esas 2 fueron mis contribuciones en la etapa del modelado; el modelo XGBoost fue una mejor aproximación, ya que la problemática se podía resolver con un random forest así. que sabíamos que modelos más robustos la podían resolver igual.

- **Evaluación**

En evaluación, lo único que hice fue checar si realmente el resultado cumplía con los criterios establecidos y detalles menores, así que mi contribución fue más de revisión y no una tan activa como las de mis otros compañeros.

- **Deployment**

Por último, en esta fase de deployment tampoco tuve un rol tan importante; lo único que hice fue realizar 2 diagramas de despliegue para la aplicación y hacer el reporte de cómo manejábamos los datos y su ciclo de vida. Y si al caso también realice alguna que otra documentación.

- **Comprendión del Negocio**

Una de las cosas principales que aprendí fue el uso e implementación de la metodología CRISP-DM, donde en la primera fase aprendí que un proyecto de ciencia de datos no inicia con códigos o modelos, sino con el entendimiento del negocio, donde lo que se hizo fue convertir un problema ganadero (identificar el momento óptimo para el secado de las vacas) a objetivos analíticos medible. También aprendí lo importante que es entender como es que los expertos trabajan en el CAETEC (la Doctora Lupita) para poder entender su manera de trabajar y buscar cuales pudieran ser problemas o limitantes que tienen en el rancho. En esta fase contribuí a la investigación de como es el funcionamiento del CAETEC, durante la visita, ayude a recopilar información relevante sobre como trabajan con las vacas, como funciona la maquina de ordeño, que datos proporciona la maquina a los especialistas y más importante que datos no tienen ellos. De igual forma contribuí en la generación de los objetivos del negocio donde durante toda la implementación de la metodología ayude a actualizar los objetivos y compararlos con los avances que se fueron desarrollando del proyecto. De igual forma parte contribuí con documentación de los objetivos, los beneficios, las terminologías, etc, para un registro para futuras consultas.

- **Comprendión de los Datos**

Para la fase de la comprensión de los datos aprendí a trabajar con datos reales que implicaron inconsistencias, valores nulos, formatos mixtos y muchas veces ambigüedades en las definiciones. Aprendí hacer todo el preprocesamiento de los datos para entender que tipos de datos se nos habían proporcionado, como eran sus tendencias, si es que existían relaciones significativas entre los tributos, entender la importancia de que atributos fueron relevantes para los objetivos y como impactan en la problemática que queremos resolver. Para esta fase una de mis contribuciones fue el análisis y documentación de algunos de los datos que se nos proporcionaron además contribuir con la investigación y creación del diccionario de datos debido a que al ser terminologías desconocidas para el equipo fue necesario hacer investigaciones de los atributos además de entender determinar si serían relevantes para las siguientes fases del proyecto. De igual forma contribuí con la documentación de los datos que explore para tener un registro sobre lo que se tenía y lo que no.

- **Preparación de los Datos**

Para la fase de la preparación de los datos aprendí a trabajar con diferentes fuentes de información para generar nuevos datasets complementarios que cumplieran con los objetivos. A pesar de que no todas las preparaciones de los datos fueron utilizadas para la iteración final, contribuí en preparación de los datos de las iteraciones anteriores para poder llevar a cabo las diferentes maneras en las que se decidió abordar el problema, que aunque no quedó como solución final, contribuyo para poder cambiar de iteración y entender si era o no óptima la forma en la que se decidía abordar las propuestas de solución.

- **Modelado**

Durante la fase de modelado enfrentamos los problemas más complicados debido a que terminamos haciendo 3 iteraciones diferentes debido que la manera en la que abordamos las primeras 2 iteraciones no fue la mejor, de igual forma gracias a eso pudimos llegar a la iteración final donde logramos cumplir el objetivo. Para la tercera iteración aprendí la importancia de no quedarnos con un solo modelo debido a que en un principio pensamos que el único modelo que serviría bien sin necesidad de implementar una arquitectura demasiado robusta para un problema no tan complicado era el Random Forest pero al final hicimos pruebas como 5 modelos donde se determinó que el mejor fue un modelo que estuviera en un punto intermedio entre los modelos sencillos y robustos que probamos (XGBoost). Durante esta fase mis contribuciones fueron ayudar a buscar cuales serían las arquitecturas a implementar, hacer uso de una de ellas para observar su rendimiento y ayudar a hacer las comparaciones de los distintos resultados de los modelos para determinar cual sería la más adecuada como modelo final. De igual forma contribuí con la documentación de los resultados y comparaciones de los modelos para hacer la comparación final que cual de ellos cumplía con los objetivos, cumplía con los scores de las métricas que definimos y cual rendimiento era el más óptimo.

- **Evaluación**

Para la fase de evaluación aprendí sobre como evaluar y hacer el análisis de las métricas no solo de manera numérica sino también analizar sobre los sesgos, sobreajustes, validez biológica, etc. Aprendí también a conectar y comparar los resultados técnicos con los criterios de éxito del negocio y a evaluar que tan bien las predicciones ayudan a tomar decisiones sobre el estado de las vacas. Para esta fase una de las contribuciones que hice fue la documentación de las evaluaciones finales con respecto a los resultados obtenidos y los objetivos iniciales que se tuvieron en fases anteriores.

- **Deployment**

Para esta fase final entendí y aprendí que un modelo no es suficiente dentro de un notebook por que con la documentación, guías, interpretación y un mecanismo práctico necesita ser adaptado para su implementación y uso. Para esta etapa mis contribuciones fueron además del diseño de como queríamos desplegar la información de manera que fuera más clara para quien implemente el modelo también ayude a la documentación del reporte final que resume todo el proyecto y la revisión final de proyecto para saber que se hizo bien, que se hizo mal, que se pudo hacer mejor, etc.

@Fidel Alexander Bonilla Montalvo

- **Comprensión del Negocio**

En esta fase me ayudo a conocer cómo funciona el CAETEC en general pues era un

lugar que no conocía pero sobre todo cuando lo visitamos fui entendiendo cómo se maneja un rancho como el CAETEC, comencé a ver que detrás de todo ese sistema hay muchísima tecnología, decisiones muy delicadas sobre cada vaca, tanto cuánto cuesta alimentar, por qué es tan importante detectar el secado, y cómo funciona el robot DeLaval, me ayudó a ver que el proyecto tenía un impacto directo en decisiones diarias del rancho.

Y mientras visitábamos el CAETEC y la maestra Lupita nos enseñaba, de lo que nos dijo lo que más me llamó la atención fue que una vaca con baja producción no es solo un dato, significa más tiempo del robot ocupado, más consumo de alimento sin retorno, más riesgo de enfermedades y retrasos en las demás vacas.

Esta fase me enseñó que un buen modelo empieza con un buen entendimiento del negocio y que si el objetivo está claro, las demás fases se vuelven mucho más sencillas.

- **Comprensión de los Datos**

Esta fase fue donde aprendí del proyecto pues al principio con los datos parecía que ya teníamos todo lo necesario para empezar a modelar, pero conforme fuimos explorando los datasets nos dimos cuenta de que estaban llenos de inconsistencias, de columnas sin información, archivos duplicados, cambios de formato según la vaca o el día, y muchas variables que no aportaban nada, y de allí empezó el trabajo interesante pues se tenía que analizar a fondo esos datos para saber como seguir con el objetivo de negocio y así fue pues conforme fuimos avanzando se encontraba la relación entre los registros de ordeño, el inventario y las fichas técnicas, gracias a esta fase pudimos definir qué datos valía la pena preparar y cuáles no, y ya con eso supimos que más datos solicitar para la continuación del proyecto.

- **Preparación de los Datos**

Lo que más costó en la fase de preparación de los datos fue unir los datasets, pues cada archivo tenía su propio formato, diferentes estructuras, distintas fechas y columnas que no coincidían exactamente, pero también, cada vaca tenía su propio archivo, así que era complicado manejar entre tantos datos de información, en esta fase se tuvieron varias partes pero contribuyendo en las primeras pudimos ver que archivos serían buenos juntas, cuales no se podrían y las columnas importantes de estas mismas. Y esto fue lo que estuve haciendo en esta parte, analizar columna por columna de algunos data set para ver si tenían información útil y las observaciones notorias documentarlas y compartirlas con el equipo.

- **Modelado**

Esta fase resultó especialmente complicada porque al inicio no estaba claro qué tipo de modelo sería el adecuado para resolver la necesidad. Por esta razón se probaron varias alternativas, comenzando por enfoques básicos y avanzando gradualmente hacia modelos más complejos. Este proceso me enseñó que es importante iniciar con lo más simple y solo escalar cuando sea necesario, evaluando en cada paso si

el modelo realmente aporta valor. Probar distintas opciones permitió identificar qué técnicas eran útiles y cuáles no, reforzando la idea de que el modelado es un proceso iterativo que requiere análisis y adaptación.

- **Evaluación**

Cuando los modelos estuvieron listos, la evaluación era tan importante porque no bastaba con ver cuál tenía mayor precisión, porque eso no siempre significa que sea el mejor. Fue necesario analizar varias métricas, revisar las matrices de confusión y verificar si alguno presentaba sobreajuste, ya que un modelo que funciona muy bien con los datos actuales podría fallar cuando lleguen datos nuevos.

Comparar los modelos con este enfoque permitió identificar que XGBoost no solo tenía buen desempeño, sino que también era el más estable y menos propenso al sobreajuste. Esto fue clave, porque un modelo útil debe mantenerse confiable en escenarios reales, no solo en el conjunto de entrenamiento. Y esto fue en lo que ayude en esta parte, apoye con la decisión del modelo

- **Deployment**

En esta fase vimos que un modelo solo tiene valor cuando puede utilizarse en la práctica. No basta con entregar el archivo del modelo es necesario acompañarlo con un dashboard, instrucciones claras y ejemplos que faciliten su uso, y eso fu lo que continuamos haciendo pues decidir como seria el diseño del dashboard y su diseño fue algo que costo en un principio y requirió varias versiones para llegar a un diseño final. Y en esta parte lo que apoye fue con el diseño con un estilo del dashboard, con las mismas funciones que las establecidas pero con un diseño sencillo para decidir cual de los diseños era mejor.

@Diego Antonio García Padilla

- **Comprendión del Negocio**

En esta fase entendí que entender el marco teórico del negocio es lo más importante para saber qué hacer: para un trabajo de calidad no sólo es importante hacer la solución de la manera correcta, si no que sea la solución correcta en primer lugar. Además, aprendí que las decisiones productivas en el CAETEC tienen un componente biológico y económico que deben estar reflejados en los objetivos analíticos.

En esta fase contribuí al entendimiento del negocio y a definir los objetivos y los criterios de los objetivos de minería de datos, siguiendo el primer criterio de la calidad: “hacer el producto correcto”. Igualmente, apoyé en la planificación y definición del plan.

Además, definí requisitos funcionales para el desarrollo de la plataforma de

visualización y predicción de datos.

- **Comprendión de los Datos**

Pude aprender a interpretar datos fisiológicos reales, entendiendo que no se comportan como datasets típicos, sino que tienen más factores involucrados, teniendo variabilidad natural, y que su análisis requiere sensibilidad biológica y conocimiento en el campo.

En esta fase aporté en analizar los nuevos datos recibidos en la tercera iteración, creando dashboards en Tableau para analizar cómo se comportaban y cómo podríamos usarlos.

- **Preparación de los Datos**

Aprendí que preparar datos para series de tiempo requiere técnicas diferentes, como la reconstrucción de secuencias, ventanas temporales y estandarización por individuo. Para esta fase colaboré en la creación del pipeline de limpieza.

- **Modelado**

Aprendí que los modelos complejos no siempre son la respuesta, especialmente cuando los datos no son suficientes para soportarlos. Para ello, es necesario incrementar de complejidad progresivamente, no es necesario resolver algo con un modelo muy complejo cuando una regresión logística puede hacer lo mismo.

Durante esta fase, mi contribución fue en la tercera iteración. Al ver que los modelos implementados previamente no cumplían con los objetivos de minería de datos, colaboré en la investigación y elección de otros modelos que se podrían aplicar, basados en el nuevo entendimiento de los datos.

- **Evaluación**

En esta penúltima etapa, aprendí que evaluar modelos en contextos reales requiere considerar limitaciones operativas y biológicas, no solo estadísticas. Por ejemplo, hay situaciones donde menos accuracy es aceptable, si un modelo tiene menos falsos negativos.

Mi contribución fue revisar los resultados con base en los criterios del negocio y proponer ajustes para mejorar la estabilidad del modelo. También ayudé a documentar escenarios donde el modelo podría fallar y cómo mitigarlos en el uso real.

- **Deployment**

Aprendí que la fase de deployment no solo consiste en entregar un modelo o una plataforma, sino dejar las bases para que este proyecto pueda ser objeto de otros, buscando mejorarlo y entregar mejores resultados.

Por último, contribuí en el desarrollo de la presentación final, en la revisión del repositorio final, asegurándome de que los archivos estuvieran organizados, reproducibles y con instrucciones claras. También apoyé a documentar las limitaciones para que futuros equipos puedan iterar sobre la solución.

Igualmente, desarrollé la plataforma web Saturno, en dónde los colaboradores del CAETEC pueden acceder, subir los archivos de [Global Hato](#) y visualizarlos de manera interactiva, así como clasificar las vacas en su ciclo de lactancia con el modelo generado.

@Ángel Mauricio Ramírez Herrera

- **Comprendión del Negocio**

Durante esta fase comprendí que entender el problema del CAETEC no solo era analizar los datos, sino entender que se toman decisiones dentro de los hatos que afectan la productividad, bienestar animal y costos.

Aprendí que una solución para esto, era “generar conocimiento” que puedan ayudar a al flujo de trabajo de las personas como las veterinarias que cuidan a las vacas para tomar decisiones informadas.

- **Comprendión de los Datos**

En esta fase me quedó claro que los datos crudos nunca están listos para modelarse. Me enfrenté a información fragmentada por vaca, columnas inconsistentes, valores nulos masivos y formatos que cambiaban entre arhivos.

Tuve que analizar cada columna, entender su significado con apoyo de los stakeholders y mis compañeros y distinguir qué variables tenían valor y cuáles no. Esta fase fue fundamental para descubrir que muchas de las variables críticas como el estado reproductivo ni siquiera existían explícitamente en los datos, lo que determinó el rumbo del proyecto

- **Preparación de los Datos**

Esta fue la fase donde más participé y donde más aprendí. Me encargué del procesamiento de datos, y confirmé que esta fase determina directamente la calidad del modelo.

Reconstruí variables que no existían, unifiqué archivos dispersos, limpié columnas ruidosas, normalicé datos y consolidé un dataset que pudiera realmente modelarse. También trabajé en la agregación diaria de ordeña, algo crítico para reducir ruido y representar el comportamiento productivo real.

En el proceso descubrí algo clave: **un buen modelo no se construye con**

algoritmos, se construye con buenos datos.

Si la preparación fallaba, el modelo fallaría sin importar qué tan avanzado fuera. Por eso considero que esta fase fue la más importante del proyecto y donde pude aportar más valor al equipo.

- **Modelado**

Durante el modelado aprendí que la parte técnica no se trata solo de probar modelos, sino de elegir el enfoque correcto. Pasamos por regresión, LSTM y finalmente clasificación multiclas.

Ver cómo el desempeño cambiaba dependiendo de la representación de los datos reforzó algo que viví en la fase anterior, el modelado refleja la calidad del procesamiento.

Me di cuenta de que el algoritmo es solo la parte final del pipeline, ya que sin datos estables, ninguna técnica hubiera funcionado. También entendí la importancia de evaluar objetivamente sin enamorarse de un modelo solo porque tiene buenas métricas aparentes, especialmente cuando existe riesgo de leakage.

- **Evaluación**

La fase de evaluación me enseñó a cuestionar los resultados, incluso cuando parecen buenos. Aprendí a revisar si las métricas realmente reflejaban aprendizaje y no overfitting, y eso solo fue posible gracias al pipeline de datos que desarrollamos.

Comprendí la importancia de validar con expertos del dominio y que no basta que un modelo prediga bien sino debe ser coherente. Esta fase cerró el círculo entre lo técnico y lo operativo, y me hizo valorar aún más la metodología CRISP y su énfasis en la retroalimentación continua.

- **Deployment**

En esta fase entendí que un proyecto de análisis no se completa hasta que entrega valor en la práctica. Ver cómo el modelo, el dataset y el pipeline se traducían en un dashboard utilizable por veterinarias cerró todo el trabajo previo.

Me quedó claro que el procesamiento de datos no solo sirve para entrenar modelos, también determina la facilidad con la que los usuarios pueden interactuar con la información en un entorno real. Esta fase me enseñó que la ingeniería de datos debe pensarse con visión de producto, no solo con visión técnica.

Además, en esta fase comprendí que muchas veces existen errores a lo largo del proyecto que pueden ser críticos en el futuro y que a veces no se pueden tomar decisiones que sean menos costosas.

@José Eduardo Viveros Escamilla

- **Comprendión del Negocio**

En esta fase aprendí la importancia de definir con claridad qué problema se quiere resolver y por qué. Comprendí que un proyecto de ciencia de datos sin una comprensión profunda del negocio corre el riesgo de construir soluciones irrelevantes. Mi contribución en esta fase fue poder generar los objetivos de negocio basado en la información recabada del CAETEC, esto a si mismo me dio la noción de toda la terminología que teníamos que tener para que la documentación tuviera coherencia, de igual forma pude apoyar a mis compañeros inventario de tecnología que usaremos en el proyecto, así como los riesgos y contingencias, esto me ayudo a tener una noción de como iba a arrancar el proyecto pero a pesar de eso todavía me sentía un tanto perdido en cuanto a la cuestión de los datos, sin embargo eso fue algo que durante las siguientes fases mejoró progresivamente.

- **Comprendión de los Datos**

Aprendí que los datos provenientes del robot de ordeña requieren interpretación contextual y que no basta con analizarlos de manera aislada. Mi contribución fue detectar inconsistencias entre los datos del corral 6 y los del hato completo, identificar columnas con semántica dudosa y validar con el equipo qué atributos podían usarse o debían descartarse. También apoyé en la elaboración de visualizaciones interpretativas para comprender cómo se comportan las distintas etapas del ciclo reproductivo.

Todo este proceso me ayudo muchísimo durante la primera iteración, la realidad es que un inicio tenía miedo de meterle la mano a los datos, al ver que a pesar de que entendía la teoría no sabía como extraer ese conocimiento al notebook de Google Colab, sin embargo mi compañero Kevin se encargo de esa parte y me estuvo explicando como poder ir moviéndolo y los pasos requeridos en un buen AED (Análisis de Exploración de Datos), y esto mismo me ayudo para que eventualmente en la tercera iteración pudiera hacer el análisis del .csv mas importante que es “global hato”, comprender los datos fue la parte mas importante para eventualmente saber que hacer con en ellos en la preparación.

- **Preparación de los Datos**

Aprendí que la preparación de los datos requiere un equilibrio entre automatización y decisiones manuales basadas en criterio experto. Mi contribución fue generar las transformaciones para los modelos de la primera y tercera iteración, la comprensión de datos donde originalmente estuve como apoyo, me ayudo mucho a poder saber que transformaciones realizar. En la primera iteración previamente mi compañero Kevin había realizado el AED (Análisis de Exploración de Datos), después de eso yo tome I a rienda de la iteración y realice una prueba de valores nulos donde estos mismo se imputaron para no afectar el rendimiento del modelo, a si mismo también genere atributos derivados que ayudaron a poder darle mas correlación. Después de esto durante la tercera iteración mi compañero Ángel realizo las transformaciones

más importantes para los 5 modelos que hicimos en esa iteración, sin embargo para el modelo que realice que fue el TabNet, necesitaba realizar otros atributos derivados. Considero que desde el AED de la primera iteración a este punto donde realice el Modelo TabNet, tuve una mucho mejor comprensión de lo que estaba sucediendo con todo lo que queríamos predecir

- **Modelado**

Aprendí que modelar implica explorar alternativas, fallar rápido y aprender de esos errores. Mi contribución fue implementar un primero modelo durante la primera iteración, el cual fue un benchmark de varios algoritmos que nos pudiera ayudar para iteraciones futuras que nos tocárían hacer, este notebook fue basicamente spliterar los datos en una distribución train, val y test, realice una Regresion lineal, Random forest, Lasso y Ridge teniendo con un mejor accuracy el Random forest con 0.7109.

Durante la tercera iteración sabíamos que al cambiar el enfoque después del resultado del segundo modelo LST, sabíamos que teníamos que volver a algo que ya conocíamos y fue que propuse la idea de volver a un Random Forest que fue el modelo que mejor se ajustó a nuestros datos desde un inicio, fue por eso mismo que también consideramos tener una versión mejorada de este que es XGBoost que fue finalmente el que terminamos eligiendo, durante esta iteración fui un apoyo para mi compañero Cristian quien se encargo de este modelo, a si mismo yo fui el cargado de un modelo con datos tabulares, me encargue de todo el diseño de la arquitectura, y del entrenamiento de esta, a si mismo esta vez haríamos un benchmark, con TabNet, Random Forest, XGBoost, Regresión Logistic y una MLP. Hablando de una forma muy sincera en este punto del proyecto yo sentía tenía el conocimiento suficiente para poder comprender todo lo que estaba sucediendo con los datos, lo cual me dio mucho gusto por que a como inicie durante la parte de Data Preparation, justo en esta etapa ya estaba mucho mas preparado. A si mismo apoye con la etapa final de la decisión del modelo que finalmente fue el XGBoost y el Random Forest como finalistas

- **Evaluación**

En esta fase aprendí que evaluar significa también cuestionar la validez externa del modelo. Comprendí que es importante pensar en qué podría pasar cuando se apliquen datos futuros o situaciones no vistas, realizamos pruebas y apoye en la documentación de la fase, de los modelos aprobados. Mi contribución fue revisar la estabilidad de los modelos aprobados y finalmente con mi equipo poder tomar un desicion en conjunto de que modelo es que el usuarios, que en este caso fue el XGBoost.

- **Deployment**

Aprendí que un proyecto debe ser accesible para quien lo use. Contribuí en la creación de instrucciones de uso simplificadas para el manual de usuario apoyando a mi compañero de Kevin, comprendiendo más el trabajo de la aplicación web que

hizo mi compañero Diego, a si mismo estuve realizando la presentación final que presentaremos el dia de mañana.