



INSTITUTO SUPERIOR POLITÉCNICO DE CÓRDOBA

TECNICATURA SUPERIOR EN CIENCIA DE DATOS E INTELIGENCIA ARTIFICIAL

Módulo de Analista de Datos

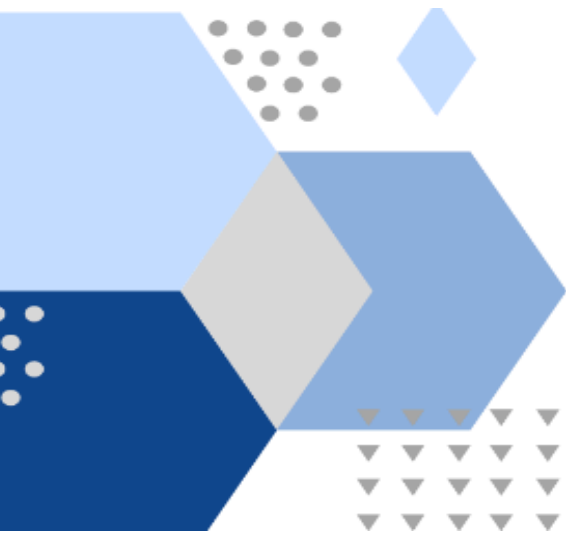
Informe Final y Conclusiones

Docentes:

- Pratta, Nahuel
- Ugarte, Marcos

Integrantes:

- López, Erick
- Nüesch, Christian
- Zurita Rojo, Débora



Índice

| | |
|---|----------|
| Índice..... | 2 |
| Presentación..... | 3 |
| Alcance de metas generales..... | 3 |
| Alcance de metas técnicas o específicas..... | 5 |

Presentación

En este documento se presenta un análisis final que reflexiona sobre los objetivos alcanzados en relación con la proyección inicial del proyecto. Se evaluará cuántas de las ideas originales se lograron cumplir y materializar, y hasta qué punto se alcanzaron estos objetivos. Dicho en otras palabras, este documento ofrece una reflexión final sobre los logros del proyecto en comparación con las metas planteadas al inicio.

Alcance de metas generales

El proyecto original tenía como objetivo desarrollar una solución innovadora para automatizar la generación de informes destinados a concesionarios y aseguradoras en la industria automotriz. La idea central era utilizar el análisis de datos para generar información relevante y personalizada, según las necesidades específicas de cada usuario. Este objetivo se cumplió parcialmente, ya que se logró trabajar con los datos base y generar predicciones. Sin embargo, no se llegó a implementar la interfaz final ni a aplicar el proyecto en un entorno real.

Este proyecto se enmarca en la innovación tecnológica, que, aunque prometedora, puede enfrentarse a demoras y extensiones de plazos durante su desarrollo. Y también muy importante, este proyecto formó parte de un espacio curricular, y su propósito principal era aprender sobre ciencia de datos, análisis de datos y sus aplicaciones en el aprendizaje automático, y esta parte sí que se cumplió.

Los principales ejes temáticos de este proyecto fueron la ciencia de datos y la inteligencia artificial, con un enfoque en la recolección, limpieza y modelado de datos. Esto incluyó el análisis exploratorio de datos y la aplicación de técnicas de *machine learning*. Estos objetivos se cumplieron satisfactoriamente.

Se logró alcanzar los objetivos relacionados con el desarrollo e implementación de Python, incluyendo el uso de bibliotecas especializadas y la aplicación de modelos predictivos, con un enfoque en la precisión y la interpretación de los resultados. También se cumplieron

las metas de visualización de datos a nivel de programación. Sin embargo, no se pudo desarrollar la generación de gráficos interactivos ni tablas que facilitaran la interpretación de grandes volúmenes de datos, debido a la limitación de tiempo.

De acuerdo con lo mencionado anteriormente, se lograron desarrollar de manera satisfactoria las competencias originales previstas, como el análisis avanzado de datos utilizando bibliotecas específicas y la habilidad de extraer *insights* a partir de grandes conjuntos de datos.

La decisión de realizar este proyecto surgió de la necesidad de enfrentar la falta de datos actualizados en la industria automotriz, lo cual dificultaba la creación de predicciones sobre clientes y otros aspectos. Teniendo en cuenta esta problemática, el proyecto fue exitoso, ya que permitió sentar las bases para ofrecer productos o servicios de análisis en este sector. Se desarrollaron capacidades importantes como la limpieza y el preprocesamiento de datos, además de generar las primeras proyecciones o predicciones mediante sistemas de aprendizaje automático.

Este avance constituyó un buen punto de partida para un futuro desarrollo completo orientado a la industria. Sin embargo, durante el transcurso de la cursada surgieron diversas cuestiones técnicas que requirieron tiempo para aprender y aplicar nuevos conocimientos. Esto absorbió la mayor parte del tiempo disponible, lo que pospuso la implementación práctica o real del proyecto para una etapa posterior.

Para concluir esta parte, este proyecto fue significativo y tuvo una relevancia personal para cada uno de los integrantes. Nos permitió enfrentarnos a diversas problemáticas del entorno real, especialmente en la solución de desafíos relacionados con la limpieza y preparación de los datos. Las mayores dificultades se presentaron al procesar los datos y seleccionar cuáles utilizar para generar predicciones que produjeran modelos consistentes y estables.

Estas experiencias y dificultades resultaron de gran valor, ya que nos prepararon para enfrentar retos similares en nuestra futura carrera profesional, fortaleciendo nuestras habilidades y conocimientos prácticos en ciencia de datos y aprendizaje automático.

Alcance de metas técnicas o específicas

En cuanto a las metas más específicas y los objetivos técnicos relacionados con las materias del espacio curricular de **Analista de Datos**, se llevaron a cabo numerosas actividades y acciones a lo largo del proyecto. Estas actividades nos permitieron adquirir conocimientos prácticos importantes, ya que implicaron un esfuerzo a lo largo de la cursada, ya que uno debía entender los conceptos para poder aplicarlos, pero también a la inversa a partir del desarrollo de aplicaciones logramos entender nuevos conceptos. En general, la comprensión de cada paso fue muy importante, lo que fortaleció nuestro aprendizaje.

Durante el desarrollo en Python, especialmente utilizando notebooks de Jupyter, se trabajó en la manipulación de columnas y el procesamiento de datos originales en bruto, sin ningún tipo de preparación previa. Esta etapa de preparación de datos fue crucial antes de iniciar el análisis. El conjunto de datos original incluía una gran variedad de tipos de datos, desde texto y números hasta valores flotantes y datos booleanos. Esta variedad nos permitió experimentar de manera amplia desde el inicio del proyecto.

En general, se practicó, entendió y aprendió de manera efectiva todo lo relacionado con el uso de bibliotecas externas en los lenguajes de programación y en los entornos de desarrollo. Esto incluyó cargar, leer y presentar los datos, así como realizar una primera exploración para identificar datos faltantes, determinar qué tipo de datos estaban ausentes y en qué cantidad.

También se llevaron a cabo análisis estadísticos preliminares sobre el conjunto de datos, evaluando aspectos como máximos, mínimos, promedios, medias, varianzas y desviaciones estándar. En resumen, se realizaron las descripciones básicas e iniciales necesarias para cualquier proceso de tratamiento de datos.

Dentro de las metas originales del proyecto, una parte importante fue el tratamiento de los datos o proceso de curado del conjunto de datos. Esto incluyó completar los valores ausentes, corregir los datos mal formateados, eliminar registros duplicados y realizar otras actividades propias del preprocesamiento, como la corrección de datos erróneos y la conversión de formatos incorrectos. Todas estas metas se cumplieron a la perfección.

En la parte final de la cursada, comenzamos a desarrollar y aplicar conceptos de *machine learning*, iniciando con una regresión lineal simple y luego avanzando a una regresión lineal múltiple.

Con respecto a la regresión lineal simple, los resultados no fueron muy satisfactorios en general, ya que predecir el precio de algo solo a partir de datos históricos no dio buenos resultados. Sin embargo, esta etapa fue valiosa para comprender bien la separación de los conjuntos de datos en entrenamiento y prueba. Aprendimos a realizar una doble separación, primero para el entrenamiento y luego para las predicciones, comparando los resultados obtenidos con los datos reales. Estos objetivos se cumplieron y los conceptos se entendieron correctamente.

En cuanto a la regresión lineal múltiple, la situación se complicó considerablemente. Aunque en un principio obtuvimos resultados prometedores y logramos una buena interpretación de los resultados en base al modelo, nos enfrentamos a problemas con los datos categóricos, que debían ser codificados para ser utilizados por el modelo de regresión. Además, surgieron serios problemas de colinealidad, lo que planteó un desafío al momento de elegir las variables a trabajar. Tuvimos que aprender desde cero cómo seleccionar las variables adecuadas y cómo determinar cuáles serían las mejores para el modelo.

Este proceso fue arduo y demandó mucho trabajo, ya que fue un constante ida y vuelta. A medida que avanzábamos, surgían nuevos problemas que nos obligaban a retroceder y ajustar el modelo o la idea general. En varias ocasiones, tuvimos que rediseñar la estrategia original para aplicar estos conceptos. Esta parte del proyecto fue, sin duda, la más desafiante, pero también la más enriquecedora, ya que nos permitió aprender mucho, incluso sobre temas que no estaban incluidos en los contenidos de la cursada, y para resolver estos problemas, tuvimos que investigar fuera del ámbito académico, lo que resultó en un añadido.

En conclusión, el proyecto fue completamente exitoso en cuanto a los objetivos y metas específicas originales. Se lograron cumplir todas las metas técnicas planteadas, desde el manejo de bibliotecas en lenguajes de programación hasta el análisis y tratamiento de los datos.