

Portafolio análisis

Luis Arturo Rendón

A01703572

SMA0102: Técnicas analíticas

Jupyter Notebook: https://colab.research.google.com/drive/1uCP2xKZgJEtoW6C-VM3TOoAVZy_z2qwT?usp=sharing

- Utiliza al menos 2 técnicas de preprocesamiento de acuerdo con el problema como escalamiento, detección de anomalías, imputación, etc.
 - Lo primero que hice fue dropear todas las rows que presentaban valores con NaN, esto más como costumbre. A partir de este primer intento todo empezó a explotar en mi modelo. Aprendí que no debes de hacer eso cuando tienes registros a través del tiempo, porque irrumpes en la línea temporal.
 - Los únicos dos manejos de NaN en el notebook, primero es en el momento de entrenar el modelo ya que este no acepta valores nulos ni infinitos y segundo cuando se hace álgebra matricial, porque también falla cuando se usan valores nulos.
 - Vale la pena mencionar que el primer primer primer manejo de datos fue mergear ambos datasets.
 - El primer procesamiento o cálculo fue sacar la variable dependiente “future_yr”, cabe mencionar que en las primeras entregas usé otra variable dependiente que no era la correcta jejeje... Esto se encuentra en “Obtenemos la variable dependiente” en el notebook.
 - El segundo paso de preprocesamiento de datos fue sacar los ratios financieros a utilizar durante todo el trabajo. Estos están explicados y obtenidos en la parte que dice “Ahora sacamos 4 financial ratios en base a los datos ya obtenidos” en el jupyter notebook.

- El tercer procesamiento serían la obtención de las variables de control en “Por último para la primera parte obtenemos las 2 variables de control”.
- Después se filtra la información mediante la industria seleccionada y se entrena a un modelo con esta información. Los datos outliers o extremos causan problemas porque mueven todos los datos y presentan promedios “ilusorios”.
- Por último, aplico winsorización a los datos para manejar estos datos extremos y delimitarlos. De esta manera no afectan a todo el data-set. Datos extremos como Apple, Microsoft, etc. Esto se encuentra en la sección que dice “Windsorizing” en el notebook.
- Explica claramente el uso de cada técnica de análisis utilizada y su relevancia en el set de datos.
 - Esto se encuentra en todo el jupyter notebook, en los análisis, datos obtenidos y las gráficas. No puedo decir exactamente en dónde debido a que los análisis y las interpretaciones se encuentran desde el principio hasta el final del jupyter notebook.

SMA0104: Análisis de información

Análisis del modelo: [analisis_modelo_bloque1_a01703572.docx](#)

- Evalúa el modelo con un conjunto de prueba y un conjunto de validación.
<https://github.com/00sen/JuliaLinearRegressionProject/blob/main/dam-julia.ipynb>
 - En la parte de “Comprobando la accuracy” se prueba la eficiencia del modelo con un data-set de validación con el 10% de los datos y con un data-set de test con el 10% de los datos. Se usó el otro 80% para el train.
 - Esto se puede ver en el punto las secciones del punto 6 del análisis del modelo “Uso y pruebas”.
- Detecta correctamente el grado de bias o sesgo: bajo medio alto.

- Se menciona el underfitt de mi modelo en el punto 6.1, 6.2, 6.3 y 8. y el que yo creía que era overfitt del modelo del framework en el punto 7.3 y 7.4.
- Detecta correctamente el grado de varianza: bajo medio alto.
 - Presenta un grado de varianza bajo, con resultados similares entre el train, validation y test. Pero gracias a las matrices de confusión se ve claramente qué es lo que sucede con la técnica que aprende mi modelo. Esto se puede ver en los puntos 6.1, 6.2, 6.3 y 6.4.
- Explica el nivel de ajuste del modelo: underfitt fitt overfitt.
 - Se explica el underfitting de mi modelo en los puntos 6.1, 6.2, 6.3 y 8. Y se explica el que creía que era el overfitting del framework en el punto 7.3 y 7.4.
- Utiliza técnicas de regularización para mejorar el desempeño del modelo.
 - Se utilizó la normalización en las dos columnas seleccionadas, esto porque el odómetro tiene valores que llegan al millón y la cantidad ronda lo 20-30 litros. Esto viene en la sección 3.7.

SEG0403: Compromiso ético y ciudadano

Repositorio del reto: https://github.com/salgue441/Al-concentracion/blob/main/kaggle_challenge/covid19-segmentation.ipynb

- Explica como la solución cumple leyes, normas y principios éticos de la industria o el ambiente del reto.
 - Se explicó en clase el reto y el propósito de este. De dónde sacaríamos la información relevante y para qué se usaría, siguiendo las condiciones y equipos puestos por el profesor.
- Explica en su repositorio cual es la normatividad correspondiente del reto o socio formador.
 - Para cubrir ambas competencias: Al principio del notebook se hace referencia al reto que es de donde sacamos todos los materiales necesarios, así como la guía de resolución del reto. Además, se

especifica un poco más abajo que se usan datos anónimos para mantener la confidencialidad.