

**MAESTRÍA**

**EN**

**EXPLOTACIÓN**

**DE**

**DATOS**

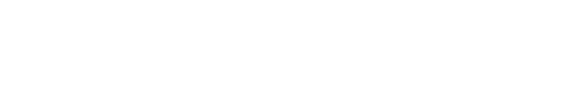
**Y**

**GESTIÓN**

**DEL**

**CONOCIMIENTO**

**Regresión Avanzada**



**Autor de contenidos:**

Lic. Hernán G. Ifran

Informe

El siguiente informe trata de dar una hoja de ruta sobre cómo se fue trabajando cada ejercicio.

En el ejercicio N°1, había que realizar una regresión logística para predecir en función de 3 variables (X1, X2, X3) el binomio (nafta-diesel).

En principio, observé que las clases están desbalanceadas y quise evaluar cuanto mejoraba el modelo al balancearlas. Por eso, lo siguiente que realice fue hacer la regresión logística con un modelo en el que las clases estaban desbalanceadas y otro en el que estaban balanceadas. El resultado fue favorable para el balanceado ya que obtuve un menor AIC (agregar).

Luego continue todo el análisis posterior con el mejor modelo, realizando el test de verosimilitud, test Hosmer-Lemeshow, test Roc y matriz de confusión.

Finalmente trabaje con otras clasificaciones como SVM, LDA y QDA, para concluir finalmente que la que mejor precisión se obtuvo de igual manera en los modelos de SVM y QDA.

En cuanto a las clasificaciones, me resultaron un desafío ya que en las clases prácticas no encontré una guía de cómo realizarlas. De todas maneras, dedique varias horas en sacar esto adelante y me llevo una mochila de aprendizajes.

Entiendo que la mejor manera de graficar los modelos podría hacer sido en 3D pero decidí realizarlo en 2D para que se visualicen más claro los conceptos.

En el ejercicio N°2, se trabajó con ANOVA. Para lo cual se evaluó si lo supuesto se cumplía. Se observó que a pesar de que los test de Shapiro y Anderson-Darling daban evidencia de normalidad, en el gráfico de qqplot los residuos no seguían una distribución normal, por tal motivo, se realizó la transformación de Box-Cox que no obtuvo mejores resultados.

Finalmente, para complementar es TP se realizó un test de Tukey. Para mayores definiciones se complementa la entrega junto con un Rmarkdown que contiene comentarios mas detallados del proceso y los resultados.

El objetivo de este informe ha sido reflejar los esfuerzos y aprendizajes aplicados en la resolución de los ejercicios y destacar la importancia de abordar las discrepancias entre las pruebas de normalidad y las observaciones visuales de los datos. Ha sido una gran oportunidad para seguir aprendiendo. Gracias.