

Proyecto final

Cristian Azocar

Universidad Tecnica Federico Santa Maria

4 de diciembre de 2023



- Se guardaron los csv en dataframes usando pandas.
- Se separaron los atributos continuos y categoricos que se iban a utilizar.
- Se conto los datos faltantes para cada columna.



- Este es uno de los boxplot de los atributos continuos.

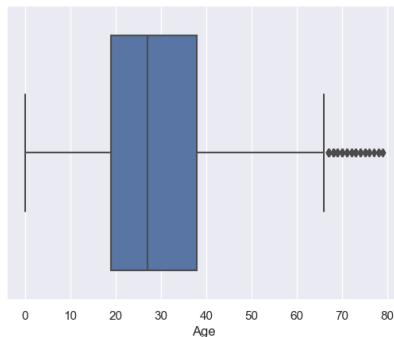


Figura: boxplot del atributo ".Age"



- Este es uno de los countplot de los atributos categoricos.

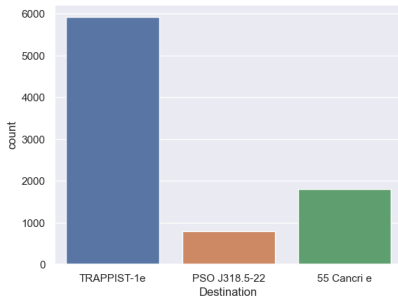


Figura: countplot del atributo "Destination"



Preprocesamiento

- Se hizo dropna() al dataframe con datos de entrenamiento.
- Para normalizar los atributos categoricos se utilizo RobustScaler.
- Para pasar los atributos categoricos a numero entero se utilizo LabelEncoder.



Seleccion de modelo

- Perceptron.
- RandomForestClassifier.
- SVC (support vector classification).
- KNeighborsClassifier.



- El modelo con la peor prediccion fue el Perceptron.
- El modelo con la mejor prediccion fue RandomForestClassifier.
- Las metricas para estos modelos se vieron utilizando `classification_report`.



Conclusion

- Al subir los resultados de la prediccion al kaggle de la competencia se obtuvo un score de 0,77.
- Algunos aspectos para considerar y con los que se podria obtener un mejor score:
- Utilizar otro metodo para optimizar los parametros de los modelos
- Utilizar otro modelo (distinto a los seleccionados para este proyecto).
- Analizar de manera distinta los datos (Preprocesamiento)
- Utilizar un conjunto de entrenamiento mas grande.

