**Introduction:**

This report presents a comprehensive analysis conducted using the two datasets one is an independent dataset and another is a guided dataset. The purpose of this analysis is to explore and understand the various characteristics of the data, utilizing a range of classification techniques. The notebook provided for this report suggests a focus on data exploration, visualization, and preliminary analysis.

**Methodology:**

**EDA & Data Visualisation:**

In the initial phase, the dataset is loaded and examined to gain a fundamental understanding of its structure, contents, and type of data we have. This includes assessing the dataset's shape, size, variable types, and initial statistical properties. Preliminary exploration is crucial to guide any subsequent, more detailed analysis and to ensure appropriate handling of the data. Following the initial exploration, the notebook delves into more specific data visualization techniques. These visualizations are primarily focused on understanding the distribution of various categories within the dataset and their characteristics.

In both of the datasets, we did count plots, bar plots, and correlation matrices. Which helps us understand data clearly and see the trend and pattern beforehand.

**Preprocessing:**

We used preprocessing techniques such as null data handling, Handling categorical values if there, Feature selection, Train, and test split before doing the modeling on the data.

We applied the same for both of the datasets to further do modeling.

**Data Modelling:**

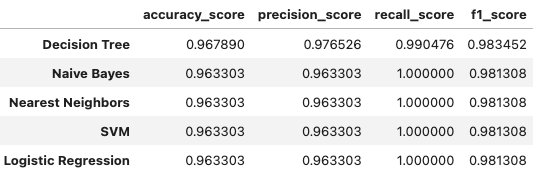
The groundwork laid by the initial data exploration and visualization provides a solid foundation for such future analyses. We applied 5 different models Logistic Regression, SVM, Linear SVM, KNN, Decision Tree, and Naive Bayes model for both datasets.

For our dataset (Animal dataset) we the dataset had very less negative labels. For which it was biased towards positive class. This Issue was handled in the guided dataset(churn dataset) when we applied the SMOTE algorithm. We checked that the accuracy decreases but overall performance of the model increases.

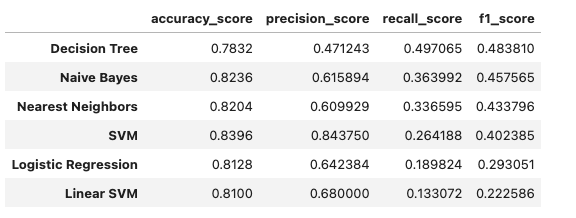
**Conclusion**

Based on the analysis conducted in the Jupyter Notebook, we have gained valuable insights into both of the datasets. The focus was primarily on understanding the dataset's structure, exploring its various attributes, and visualizing key aspects. These steps have laid a solid foundation for any further detailed analysis. These insights from the visualization are crucial for anyone looking to understand the dataset's scope and for planning further analysis.

At the end we show all the metrics like accuracy, precision, recall, f1 score in a tabular format, and the confusion metrics for each algorithm, We found the best algorithm for each dataset.

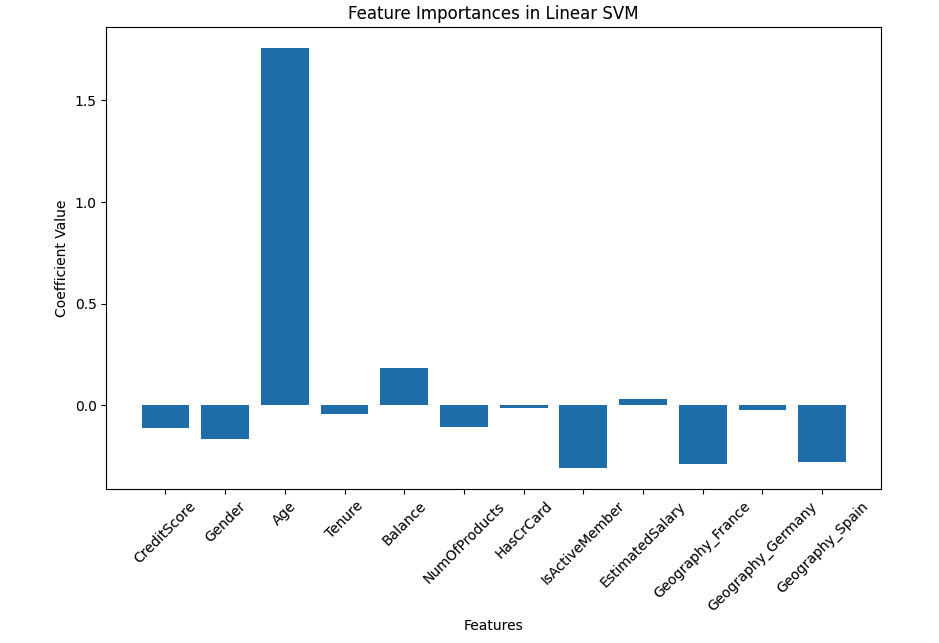


[Own dataset: Animal dataset]



[Guided dataset: Churn dataset]

For the churn dataset, we visualise feature importance for the SVM model which indicates which feature has having positive or negative impact on a user being churned.



**Recommendations for Future Work**

**In-Depth Statistical Analysis:**

Dive deeper into the dataset through hypothesis testing and correlation studies. This can help uncover hidden relationships between different variables and provide a better understanding of the data.

**Data Preprocessing and Feature Engineering:**

Before modeling, ensuring the dataset is clean and well-preprocessed is essential. This may include handling missing values, normalizing data, or creating new features that could enhance the model's performance.

**Apply Ensemble Models:**

Models like random forests, XGboost, and Catboost could be explored based on the dataset's characteristics. If the dataset has more categorical variables then a model like catboost can be used. The imbalance dataset can be handled better in XGBoost.