Problem Statement:

Build a Decision Tree classifier to predict if the client will subscribe to a Term Deposit based on their demographic and behavioral data.

Dataset

This dataset is a Bank Marketing dataset from the UCI Machine Learning Repository (https://docs.google.com/spreadsheets/d/1oTqQ1nwwVfv49gWszShu4CfWVuCfYj9r/edit?usp=sharing&ouid=107093930789725076517&rtpof=true&sd=true)

The data is related with direct marketing campaigns (phone calls) of a Portuguese banking institution. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed.

Column Description:

age : Integer Age

job : Occupation

marital : Marital Status

education : Education Level

default : has credit in default?

balance : average yearly balance in euros

housing : has housing loan?

loan : has personal loan?

contact : contact communication type

day\_of\_month: last contact day of the month

duration : last contact duration, in seconds (numeric)

campaign : number of contacts performed during this campaign and for this client

pdays : number of days that passed by after the client was last contacted from a previous campaign(numeric, -1 means client was not previously contacted)

previous : number of contacts performed before this campaign and for this client

poutcome : outcome of the previous marketing campaign

y : has the client subscribed a term deposit? (outcome)

Conclusion and Key Insights

In this project, we embarked on a journey to build a predictive model for term deposit subscriptions in a banking context to help a Portuguese banking institution tailor its marketing efforts effectively. Let's recap the key takeaways and insights from this endeavor:

Data Pre-processing:

The initial data inspection revealed a well-structured dataset with no missing values. While we encountered some duplicate entries, they were not considered errors but rather different campaigns targeting the same clients.

Exploratory Data Analysis (EDA):

The age distribution of clients highlighted that the majority fell in the 30-40 age range, coinciding with the highest subscription rate for term deposits.

Occupation-wise, clients in administrative, blue-collar and technician roles formed the majority and were more likely to subscribe to term deposits.

Marital status leaned heavily toward the "married" category.

Clients with university degrees or high-level education were more prominent and had a higher subscription rate.

Credit defaults were rare, with the vast majority of clients having no credit defaults.

Clients with housing loans were more likely to subscribe to term deposits, while those with personal loans were less likely.

Model Building and Evaluation:

We implemented a Decision Tree classifier to predict term deposit subscriptions. T

The model exhibited strong performance, achieving an accuracy rate of approximately 91.52% on both training and testing data.

Cross-validation confirmed the robustness of our model, with an accuracy score of around 91.60%.

The classification report revealed a balanced model with reasonable precision, recall, and F1-scores for both subscribed and non-subscribed clients.

Insights:

The duration of calls was a significant predictor of term deposit subscriptions. Clients tended to subscribe after longer conversations, suggesting the importance of engaging clients effectively during calls.

Previous attempts at contact also played a role, indicating that repeated contact can be fruitful in convincing clients to subscribe.

Multicollinearity among some input features was observed, potentially affecting model performance. Consideration should be given to addressing this issue in future iterations.

Future Directions:

To further enhance predictive accuracy, exploring other machine learning algorithms or ensemble methods could be beneficial. Feature engineering and selection methods could be employed to mitigate multicollinearity. Incorporating external data sources, such as economic indicators or market trends, may provide additional insights and improve model performance. A real-time deployment of the model into the bank's operational systems can automate decision-making and streamline marketing efforts.

Thank you !