Bank Marketing Campaign Project

Group Name: Yomna's Group

Name: Yomna Eisa

Email: yomnaabelrahmaneisa@gmail.com

Country: Saudi Arabia

College/Company: N/A (fresh graduate)

Specialization: Data Science

GitHub Link:
https://github.com/YomnaEisa/Data-Glacier-Projects-YomnaEisa/tree/main/week13

1. **Problem description**

   ABC Bank plans to launch a term deposit product and seek to build a predictive model to identify potential customers likely to purchase it. By utilizing machine learning (ML) models, the bank aims to optimize its marketing efforts, targeting customers with a higher probability of buying the product. This strategy, implemented through telemarketing, SMS, and marketing channels, aims to save resources and reduce costs associated with resource billing.

2. **Business understanding**

   These marketing campaigns will greatly benefit ABC Bank in several ways. Firstly, by utilizing a predictive model to identify potential customers who are more likely to purchase the term deposit product, the bank can optimize its marketing efforts. This targeted approach ensures that resources, such as time and budget, are allocated efficiently, focusing only on customers with a higher probability of buying the product. As a result, the bank can significantly reduce marketing costs and improve overall cost-effectiveness.

   Secondly, the campaigns allow ABC Bank to establish stronger connections with their target customers. By reaching out through various marketing channels like

telemarketing, SMS, and email marketing, the bank can engage with customers in a more personalized and direct manner. This personalized interaction can enhance customer satisfaction, increase brand loyalty, and potentially lead to long-term customer relationships.

Furthermore, the campaigns provide valuable insights into customer behavior and preferences. By analyzing the outcomes and responses from these marketing efforts, ABC Bank can gain a deeper understanding of its target market. This knowledge can help the bank refine its product offerings, tailor future marketing campaigns, and make data-driven business decisions.

Overall, these marketing campaigns enabled by the predictive model contribute to the bank's profitability and growth. By effectively targeting potential customers, optimizing resource allocation, and enhancing customer relationships, ABC Bank can maximize the success rate of selling its term deposit product and achieve its business objectives.

3. Data understanding

The dataset named ‘bank-additional-full’ is a CSV file that consists of 21 columns and 41188 rows. The file contains data from May 2008 to November 2010. The data covers information regarding the marketing campaign itself such as employment variation rate, number of employees, consumer confidence index, and Euribor 3-month rate. As well as all the basic client information, such as age, job, education, marital status…etc. Lastly, there’s the variable ‘y’ which is an answer to the question ‘Has the client subscribed to a term deposit? ’ The answer is a binary yes or no.

4. Type of data for analysis

Column Name	Data Type	No. null/unknown	No. of outliners

		values	
age	Integer	0	0
job	String	**330**	0
marital	String	**80**	0
education	String	**1731**	0
default	String	**8597**	0
housing	String	**990**	0
loan	String	**990**	0
contact	String	0	0
day	String	0	0
month	String	0	0
year	Integer	0	0
pdays	Integer	0	0
previous	Integer	0	0
poutcome	String	0	0
campaign	String	0	0
day_of_week	String	0	0
duration	Integer	0	**1446**
Emp.var.rate	Float	0	0
cons.price.idx	Float	0	0

Cons.conf.idx	Float	0	0
euribor3m	Float	0	0
nr.employed	Float	0	0
y	String	0	0

5. Problems in the data

5.1 Null values

No null values were found in the dataset.

5.2 'Unknown' values

The following 'unknown' values were found in the dataset:

Column 'job' has **330** 'unknown' values.

Column 'marital' has **80** 'unknown' values.

Column 'education' has **1731** 'unknown' values.

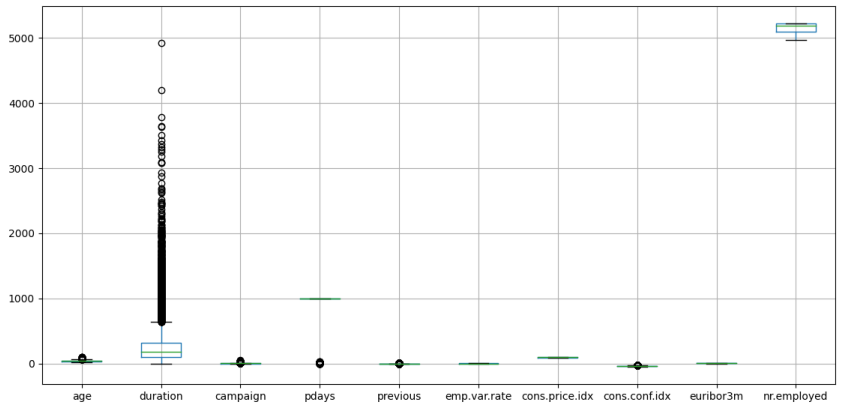
Column 'default' has **8597** 'unknown' values.

Column 'housing' has **990** 'unknown' values.

Column 'loan' has **990** 'unknown' values.

5.3 Outliner values

The column 'duration' contains outliner data based on the fact that it has a mean of 258.28 while the max value is 4918. The graph below confirms the presence of outliner data in the column 'duration':



5.4 Unbalanced data

When calculating the proportion of each class in the target variable ‘y’, these were the results:

no 0.887346

yes 0.112654

We can conclude based on the above that the dataset is unbalanced given the fact that the class ‘no’ is significantly larger than the class ‘yes’

6. Approaches to solving problems in the data

6.1 Solution for unknown values

Depending on the column itself and the data it holds, either we will drop the row or replace the ‘unknown’ value with the mode. For the housing and loan columns, we replace the missing values with the mode. Moreover, for the columns job and marital, we will drop the missing values. Lastly, for the ‘education’ and ‘default’ columns, we’ll use logistic regression.

6.2 Solution for Outliner Values

For removing the outliners, we’ll use the Z-Score Method and the IQR

6.3 Solution for unbalanced data

For the unbalanced data, we decided on undersampling: which involves reducing the number of instances in the majority class to balance it with the minority class.

7. Final Recommendations Exploratory Data Analysis

Based on the analysis of the provided data, several key business recommendations can be formulated to enhance the bank's marketing strategy for term purchases.

Firstly, it is strongly recommended that the bank directs its marketing efforts towards married couples. The data indicates that married individuals have demonstrated a notably higher likelihood of engaging in term purchases. This

target demographic should be a primary focus for the bank's promotional activities.

Secondly, the bank should consider tailoring its marketing campaigns towards university graduates. This demographic has shown a higher propensity for term purchases, suggesting that targeting this educated group may yield more favorable results.

Additionally, individuals without outstanding loans should be a priority in the bank's marketing strategy. This group exhibits a heightened inclination to engage in term purchases, making them a valuable segment to target.

Moreover, the bank is encouraged to concentrate its marketing endeavors on individuals in their late twenties to early forties. This age group has exhibited a greater inclination to participate in term purchases, and targeting them may lead to increased engagement.

Lastly, individuals with higher Consumer Price Index (CPI) values should be a focus of the bank's marketing efforts. The data suggests that those with elevated CPI values are more likely to engage in term purchases, making them a potentially lucrative target audience.

By implementing these recommended strategies, the bank can refine its marketing approach to better reach and engage potential customers for term purchases, ultimately contributing to increased business success.

8. Machine Learning Model:

After a thorough analysis of the available data, we arrived at the decision to employ a combination of three distinct machine learning models: K-Nearest Neighbors (KNN), Decision Tree, and Bernoulli Naive Bayes. This selection stems from a nuanced understanding of the dataset's characteristics and the

specific challenges it presents. KNN, with its instance-based learning approach, is well-suited for cases where proximity in feature space directly correlates with similarity in output. Decision Trees, on the other hand, are excellent for capturing non-linear relationships and complex decision boundaries within the data. Lastly, Bernoulli Naive Bayes is adept at handling categorical features and is particularly effective when dealing with binary classification tasks. By harnessing the strengths of these three models, we aim to construct a robust predictive framework that maximizes accuracy and generalization on our dataset.

Given the performance metrics of each model, including K-Nearest Neighbors (KNN) with an accuracy of 93%, Decision Tree with 91%, and Bernoulli Naive Bayes with 86%, we have opted to implement the KNN model for this project. The notably higher accuracy of the KNN model suggests that it is the most suitable choice for our specific dataset and predictive task. This decision aligns with our goal of maximizing accuracy and precision in our model's predictions.

9. Conclusion

After thorough analysis, we proposed targeted marketing strategies, prioritizing married couples, university graduates, loan-free individuals, specific age brackets, and those with higher Consumer Price Index values. We successfully addressed data challenges, including 'unknown' values and outliers, implementing effective solutions. Employing K-Nearest Neighbors, Decision Tree, and Bernoulli Naive Bayes models, we observed exceptional accuracy of 93% with KNN. This comprehensive project equips ABC Bank with actionable insights to refine its marketing approach and drive the success of its new product.