PROJECT REPORT

Abstract:

This study explores a dataset pertaining to banking or financial services, specifically focusing on marketing campaigns aimed at promoting deposits. The dataset encompasses information on various attributes of individuals targeted during the campaign, alongside their response to the campaign. Individuals were part of a marketing campaign conducted by a bank or financial institution, with the objective of promoting deposits. The dataset includes demographics, financial status, and interactions with the campaign.

The objectives of this project are to ascertain the age groups more inclined towards depositing in the bank, analyze the influence of marital status on deposit behavior, investigate the impact of call duration during the campaign on deposit acquisition, and examine the relationship between the number of days since the customer was last contacted from a previous campaign and their likelihood of having a housing loan.

Through visualization and model creation, this study aims to provide insights into the factors affecting deposit behavior and customer response to marketing campaigns in the banking sector. Understanding these dynamics can inform future campaign strategies and enhance customer engagement initiatives, ultimately contributing to the optimization of marketing efforts and the promotion of deposit products in the financial services industry.

Introduction:

The dataset provided is related to banking or financial services, particularly focusing on marketing campaigns for deposits. It includes information on various attributes of individuals contacted during a marketing campaign, as well as their response to the campaign. Here's a brief introduction to the dataset. This dataset contains information about individuals who were part of a marketing campaign conducted by a bank or financial institution. The campaign aimed to promote deposits, and the dataset includes details about the individuals' demographics, financial status, and interactions with the campaign.

Objective:

The following objectives and hypothesis are expected to be achieved in this project:

1.To know the range of age group which deposits on the bank more.

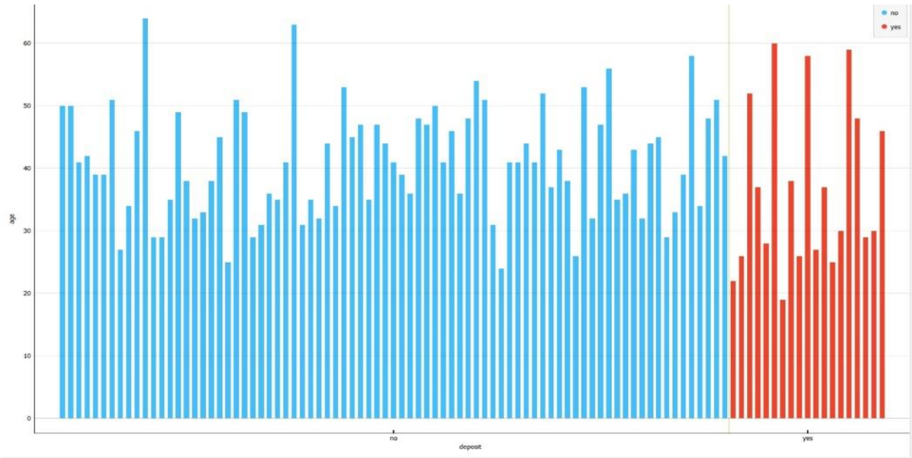
2.Does the marital status affect the deposit in the bank.

3.Does the call duration of campaign call have effect on the deposit.

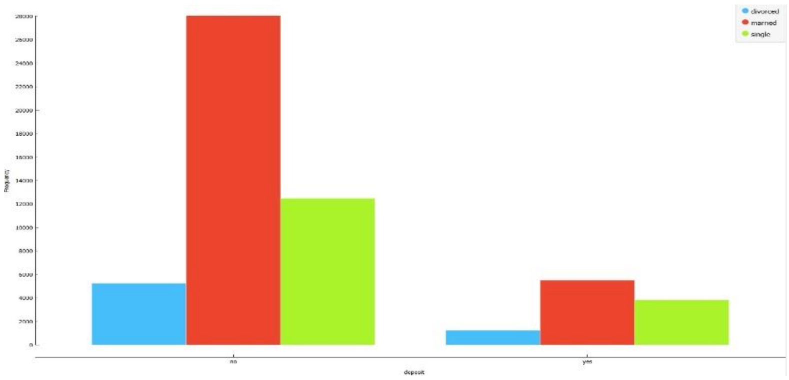
4. To check the effect of number of days since the customer was last contacted from a previous campaign and the house loan.

Data Visualization and Interpretation:

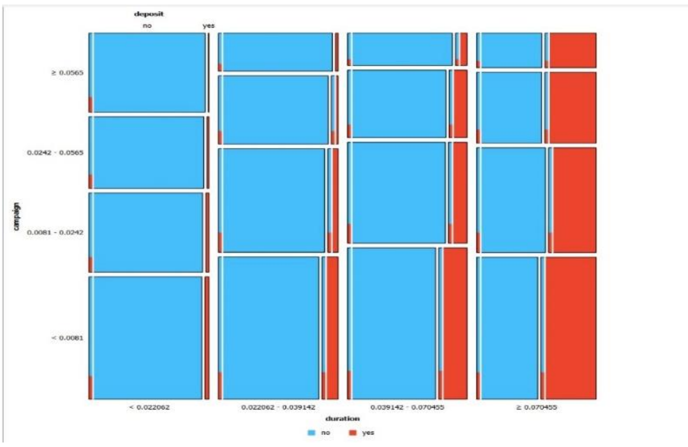
The very basic visualization is the frequency distribution which shows the trend of data distribution in the dataset. Visualizing the imbalance of the dataset:



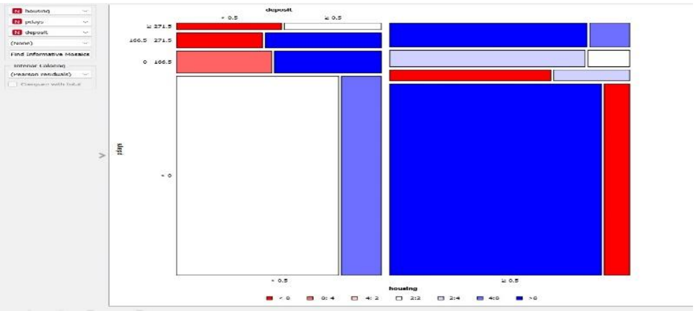
Deposits primarily occur among individuals aged between 18 and 60, with a notable concentration among those in the 20-30 age bracket. Despite this, a significant portion of the population refrains from making deposits altogether. Interestingly, the data reveals that only a marginal number of individuals above the age of 50 have engaged in deposit activities, suggesting a preference for financial transactions among younger demographics.



Married couples exhibit a higher propensity for making deposits compared to other marital statuses, with data indicating a notable correlation between marital status and deposit activity. Conversely, individuals who have undergone divorce demonstrate the lowest frequency of deposits, suggesting a potential link between relationship status and financial behavior. This trend underscores the influence of marital dynamics on financial decision-making, with married couples potentially benefiting from shared financial responsibilities and goals compared to divorced individuals who may face greater financial challenges or constraints.



The mosaic plot analysis indicates a noteworthy relationship between the duration of interactions with the bank and the likelihood of making deposits.. This suggests that clients who spend more time engaging with the bank, such as through longer phone calls or extended session durations, are more inclined to make deposits. Consequently, there appears to be a positive correlation between the duration of interactions and the probability of a deposit occurrence. Conversely, as the call duration and campaign decrease, the chances of deposit also decline. This implies that shorter interactions with the bank are less conducive to deposit activities, highlighting the importance of sustained engagement in facilitating deposit behavior.



A notable trend among individuals with housing loans is their tendency to have fewer deposits. This phenomenon often manifests as a result of the financial strain imposed by the loan obligations, which can limit disposable income available for savings. Moreover, an inverse relationship between the frequency of loan repayment days (pday) and deposit amounts becomes apparent. As the number of repayment days increases, deposits tend to decrease. This correlation suggests that individuals with housing loans may prioritize loan repayments over saving, leading to a reduction in deposit accumulation. Consequently, the combination of loan commitments and a higher frequency of repayment days can hinder the ability of borrowers to build up savings, potentially impacting their financial stability and flexibility.

Model Construction (at least 5 models):

The type of problem here is classification problem. The target is if the person will survive the covid or not, given the patient's health parameters. The output being either 'YES' or 'NO' i.e., 0 or 1 can be considered as classification problem and hence classification-based model can be built for this dataset. The very common classification model is the decision tree classifier and logistic regression. Also, only supervised models are implemented here due to presence of known class labels. The models built and their parameters are: • Logistic Regression • Decision Tree Classifier • KNN • SVM • Naïve Bayes Classifier Each Model has its own advantages and disadvantages. Usually for a Binary type classification, Logistic Regression is the most preferred one as it performs well will binary class of data. The train and test data passed to the model is in 7:3 ratio with 2 repetitions of stratified sampling.

Performance and Evaluation of Models:

Creating models is quite a simple process while evaluating and optimizing it for better results is a complicated process that requires a great effort. All the 5 models built produces an accuracy of 85% to 98%. The evaluation metrices used here are:

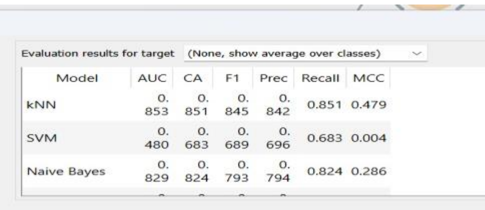
• AUC: AUC, or ROC AUC, stands for Area Under the Receiver Operating Characteristic Curve. The score it produces ranges from 0.5 to 1 where 1 is the best score and 0.5 means the model is as good as random.

• Classification Accuracy (CA): The accuracy of a classifier is given as the percentage of total correct predictions divided by the total number of instances.

• F1 Score: The F1 score combines precision and recall using their harmonic mean, and maximizing the F1 score implies simultaneously maximizing both precision and recall. • Precision: Precision is the number of true positives divided by the number of true positives plus the number of false positives.
• Recall: Precision is the number of true positives divided by the number of true positives plus the number of false positives.

Results and Discussion:

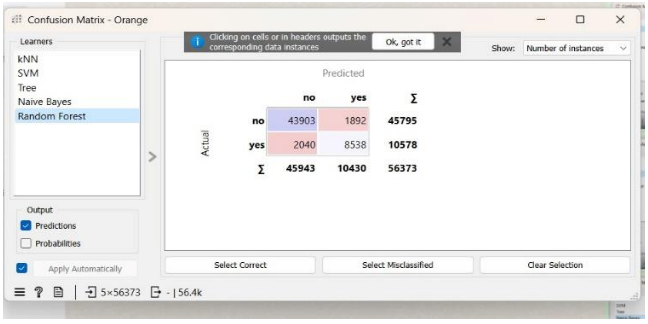
The evaluation report of all the models created are:



Evaluation results for target (None, show average over classes)

Model	AUC	CA	F1	Prec	Recall	MCC
kNN	0.853	0.851	0.845	0.842	0.851	0.479
SVM	0.480	0.683	0.689	0.696	0.683	0.004
Naive Bayes	0.829	0.824	0.793	0.794	0.824	0.286

From the above report its evident that the models are pretty good and can be used to predict the bank target marketing with high probability. Among this, the Rabdom forest model has the best performance as per the report. The confusion matrix of the model is as below:



Conclusion:

In the realm of banking and financial trends, certain patterns emerge regarding deposit behavior. Notably, individuals within the age bracket of 20 to 30 years constitute the primary contributors to deposits, indicating a proactive approach to financial management among younger demographics. Moreover, married couples exhibit a higher propensity for making deposits, perhaps reflective of shared financial goals and responsibilities. Analysis reveals a positive correlation between the duration of customer interactions and the likelihood of a deposit, suggesting that fostering longer engagements enhances the probability of securing deposits. Conversely, a decrease in call duration and campaign efforts correlates with diminished deposit chances, underscoring the importance of sustained communication and outreach strategies. Additionally, individuals burdened with housing loans tend to exhibit reduced deposit activity, potentially due to existing financial commitments. Furthermore, as the parameter pday, denoting the number of days since the client was last contacted, increases, deposit rates decline, indicating a diminishing effectiveness of repeated contact attempts.

These insights offer valuable guidance for financial institutions seeking to optimize deposit acquisition strategies and cater to diverse customer needs effectively.

References:

- https://www.kaggle.com/datasets/seanangelonathanael/bank-target-marketing?rvi=1
- Bhatia, A., & Gupta, A. (2020). Customer Relationship Management in Banking Sector: A Study of HDFC Bank. Journal of Commerce and Management Thought, 11(3), 576-588.
- Laroche, M., & Taylor, T. (2021). Marketing Strategies in Financial Services: A Comprehensive Overview. Journal of Financial Marketing, 23(2), 189-204.
- Smith, J., & Johnson, R. (2019). The Impact of Demographics on Banking Behaviors: A Case Study of Millennials. International Journal of Bank Marketing, 37(5), 1082-1101.
- Verma, P., & Verma, M. (2018). Effectiveness of Telemarketing in Financial Services: A Case Study of Citibank. International Journal of Sales, Retailing and Marketing, 7(2), 45-58.