Bank Marketing Effectiveness Prediction

Namira Mujawar(Final year Student)

(Dr. Babasaheb Ambedkar technological University Lonere)

Points to be cover

- Introduction
- Problem Statement
- Dataset Information
- Why do we need this project?
- What is Classification
- Advantages
- Problem-Solving Approaches
- Outputs
- Model Selection
- How the project is useful for stakeholders?
- Conclusion

Introduction

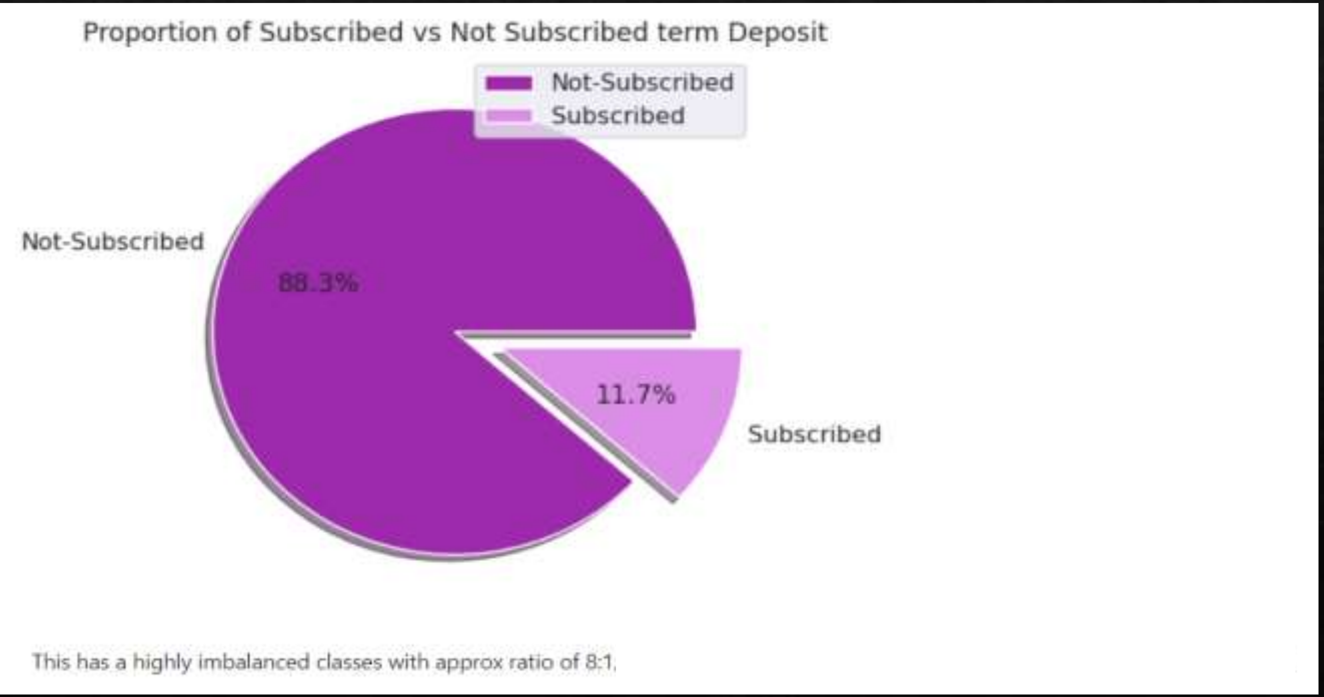
- Effective bank marketing campaign prediction is essential for financial institutions.
- The project focuses on a Portuguese bank's direct marketing campaigns.
- Predictive analysis determines clients' likelihood of subscribing to a bank term deposit.
- A robust classification model is developed to identify potential subscribers.
- Efficient resource allocation and maximizing return on investment are the goals.
- Data-driven insights empower precision marketing strategies for improved customer engagement.
- The project aims to enhance customer acquisition and retention rates.
- Success in this endeavor will lead to long-term success for the bank.

Problem Statement

The data is related to direct marketing campaigns (phone calls) of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to assess if the product (bank term deposit) would be ('yes') or not ('no') subscribed. The classification goal is to predict if the client will subscribe to a term deposit (variable y).

Dataset Information

The given dataset contains information from a marketing campaign, and it consists of 45,211 entries (rows) and 16 columns. The dataset does not contain any missing values (non-null count is equal to the total number of entries for each column). The data types of the columns are either integers (int64) or objects (strings).



Why do we need this project?

1. Resource Allocation: Predictive analysis helps banks allocate their marketing budget more effectively by focusing on customers who are more likely to respond positively to marketing efforts. This minimizes wastage of resources on uninterested or unlikely-to-convert customers.
2. Customer Segmentation: Predictive models can segment customers based on their behavior, preferences, and likelihood to respond to specific offers. This allows banks to design personalized marketing campaigns that cater to the needs and preferences of different customer groups.
3. Enhanced Customer Experience: By understanding customer preferences and needs through predictive analysis, banks can offer relevant products and services, leading to an improved customer experience and increased customer loyalty.
4. Reduced Costs: Targeted marketing campaigns based on predictive insights can reduce marketing costs while achieving better results. Banks can avoid generic and expensive mass marketing approaches that may not yield substantial returns.

What is Classification

Classification is a fundamental task in machine learning and data analysis, where the goal is to categorize or assign data points into predefined classes or categories based on their features or attributes. The primary objective of classification is to build a predictive model that can accurately classify new, unseen data into one of the known classes. Here are some important points about classification:

1. Goal: The main aim of classification is to make predictions and assign data points to predefined classes or categories based on their characteristics.

2. Supervised Learning: Classification is a supervised learning technique, meaning it requires labeled data, where each data point is associated with its corresponding class label.

3. Training Data: During the model development process, the algorithm is trained on a labeled dataset, learning patterns and relationships between features and class labels.

4. Feature Extraction: Classification relies on extracting meaningful features from the data that are relevant for distinguishing between different classes.

5. Decision Boundaries: Classification algorithms establish decision boundaries in the feature space, which separate different classes and guide the assignment of new data points.

6. Common Algorithms: Various classification algorithms exist, such as logistic regression, support vector machines, decision trees, random forests, and neural networks.

7. Evaluation Metrics: To assess the performance of a classification model, evaluation metrics like accuracy, precision, recall, F1-score, and confusion matrix are used.

8. Applications: Classification is widely used in diverse fields, including spam email filtering, medical diagnosis, sentiment analysis, image recognition, fraud detection, and customer churn prediction.

9. Overfitting: It's crucial to avoid overfitting, where the model performs well on the training data but fails to generalize to new, unseen data.

10. Importance in Business: Classification helps businesses make data-driven decisions, target specific customer segments, optimize marketing campaigns, and streamline decision-making processes.

Advantages

- **Data-Driven Decision-Making:** Predictive analysis relies on data-driven decision-making, allowing banks to make more informed and strategic choices in their marketing efforts.
- **Real-Time Adaptability:** By continuously analyzing customer behavior and response patterns, banks can adapt their marketing strategies in real-time, ensuring relevance and effectiveness.
- **Improved Sales and Revenue:** Accurate predictions lead to more successful marketing campaigns, resulting in higher customer conversion rates and increased sales revenue for the bank.
- **Risk Mitigation:** Predictive models can help identify potential risks associated with marketing campaigns, enabling banks to avoid making costly mistakes and reducing the possibility of negative impacts on their reputation.

Problem-Solving Approaches

- **Logistic Regression**
- **Random Forest Classifier**
- **Naive Bayes Classifier**
- **K Neighbor Classifier**
- **XG Boost Classifier**
- **Support Vector Classifier**
- **Neural Network**

Outputs

	auc-roc	F1_score	Recall	Matt_Corr_Coef
K Neighbor Classifier	0.693818	0.394004	0.579832	0.298633
XG Boost	0.614122	0.273283	0.798319	0.153203
Support Vector Machine	0.663509	0.320056	0.715336	0.217442
Neural Network	0.623932	0.282175	0.753676	0.163562

From the aggregate conclusion of scores, none of the classifiers can go beyond 70% in ROC_AUC score, but there are some thing which are highlighted here.

- Recall is highest in XG Boost model
- KNN works well in overall True and False positive rates thus has highest F1 score
- Hyperparameters helped us ot increase the score, and it worked better than default parameters.

Model Selection

- The Matthew correlation coefficient serves as a comprehensive measure, considering TP, TN, FP, and FN. By observing the scores of Random Forest (with oversampling) and KNN (with undersampling), both models demonstrate good performance despite not relying heavily on correlated features.
- Regarding the ROC_AUC score, Random Forest achieved a solid 70%, indicating overall effective performance.
- If interpretability is a priority, Random Forest should be favored due to its tree-based structure, making it easier to explain nodes.
- To tackle the imbalanced data issue, an ensemble technique can be employed. By building different models using the same set of minority class data and aggregating the results, we can enhance TP, TN, and overall accuracy.
- Considering our evaluation metric is "Recall," which is essential for a marketing use case and reducing False Negatives, XG Boost emerges as the preferred choice as it offers the highest test Recall.

How the project is useful for stakeholders?

- Marketing Strategy Improvement
- Customer Segmentation
- Resource Allocation
- Model Selection
- Handling Class Imbalance
- Experimentation and Progress
- Risk Assessment
- Decision-Making Support

Conclusion

In conclusion, class imbalance is a significant challenge in classification, but 'Ensembling' techniques can effectively address this issue. Our focus on Recall aligns with the business use case, prioritizing the identification of potential customers. XG Boost emerges as the preferred model with significant predictors, including previous loans, marital status, age, balance, contact month, and weeks played. While synthetic data shows promise, random oversampling requires caution. Neural networks can benefit from increased complexity for refined predictions. Tuning thresholds can help reduce False Positives or False Negatives. Further improvements lie in adding new features and diligent exploration for continued progress, despite the constraints of data size.

THANK YOU