Prediction model to identify potential customers for banking institution

A Project Work Synopsis

Submitted in the partial fulfilment for the award of the degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE WITH SPECIALIZATION IN

CLOUD COMPUTING

Submitted by:

Anubhav Jain

Hardik

Tanu

Rohit

Under the Supervision of:

Prof.(Dr.) Ranjan Walia

CHANDIGARH UNIVERSITY
Discover. Learn. Empower.

CHANDIGARH UNIVERSITY, GHARUAN, MOHALI - 140413, PUNJAB

April,2024

Abstract

This research delves into an in-depth analysis of the data derived from a comprehensive marketing campaign spearheaded by Banco de Portugal. The central objective of this initiative was to augment the subscription rates of customers towards fixed-term deposit products, prominently encompassing Certificates of Deposit (CDs). Drawing upon a rich reservoir of knowledge acquired from relevant coursework, an array of sophisticated machine learning algorithms were systematically deployed. These algorithms were meticulously applied to unravel a critical question: What actionable strategies can financial institutions adopt to effectively market fixed-term deposit products, thereby optimizing efficiency and amplifying success rates? Through a meticulous examination of the intricate layers of data, this study endeavors to furnish nuanced insights and actionable recommendations tailored to augment the efficacy of marketing endeavors within the banking domain, with a particular focus on enhancing the allure and uptake of fixed-term deposit offerings among discerning consumers.

Table of Contents

Title Page

Abstract

1. Introduction 1

1.1 Problem Definition 1

1.2 Problem Overview 2

2. Literature Survey 3

2.1 Existing System 5

2.2 Proposed System 6

3. Problem Formulation 7

4. Research Objective 8

5. Methodologies 9

6. Experimental Setup 11

7. Conclusion 13

8. Reference 14

1. INTRODUCTION

1.1 Problem Definition

In the contemporary landscape of advertising inundated with digital noise, businesses, especially those in highly competitive sectors like banking, struggle to effectively capture consumer attention and promote their products, such as fixed-term deposit offerings. The primary challenge lies in devising advertising strategies that cut through the clutter and resonate with the target audience, ultimately driving customer engagement and business growth.

Given this scenario, the problem formulation for banks aiming to optimize their marketing strategies for fixed-term deposit products can be articulated as follows:

1. Identifying Target Audience:
   - How can banks accurately identify and define their target audience for fixed-term deposit products amidst a diverse consumer base with varying financial needs and preferences?

2. Understanding Consumer Behavior:
   - What are the key drivers influencing consumer behavior and decision-making processes regarding fixed-term deposit investments?

3. Leveraging Data Insights:
   - How can banks effectively collect, analyze, and leverage data from past marketing campaigns, customer interactions, and market trends to inform their advertising strategies?

4. Personalization and Segmentation:
   - What methodologies and technologies can banks employ to segment customers based on demographic, psychographic, and behavioral attributes, enabling personalized advertising campaigns?

5. Measuring Campaign Effectiveness:
   - What metrics and KPIs should banks prioritize to evaluate the effectiveness and ROI of

their advertising initiatives for fixed-term deposit products?

6. Adapting to Market Dynamics:

- How can banks adapt their marketing strategies in response to changing market dynamics, competitor actions, and regulatory factors impacting the banking sector?

7. Integration of Technology:

- How can banks integrate advanced technologies such as machine learning, predictive analytics, and marketing automation tools to enhance the efficiency and efficacy of their advertising efforts?

8. Ensuring Compliance and Trust:

- How can banks ensure that their advertising strategies comply with regulatory standards and industry best practices while fostering trust and credibility among consumers?

By addressing these key questions and challenges, banks can develop comprehensive and data-driven advertising strategies tailored to promote fixed-term deposit products effectively, thereby attracting more customers and achieving business growth objectives in a competitive market environment.The crux of the problem in advertising fixed-term deposit products lies in effectively leveraging the plethora of available data sources. Banks possess a wealth of data, including customer demographics, transaction history, and market trends, yet harnessing this data to inform advertising strategies remains a challenge. One key issue is the sheer volume and diversity of data sources available to banks. Customer data may be scattered across various systems and databases, making it difficult to consolidate and analyze effectively. Transaction history provides valuable insights into individual customer behavior, but it must be supplemented with broader market trends to paint a comprehensive picture.

Furthermore, the challenge extends beyond data aggregation to data integration. Integrating diverse data streams, such as demographic information, transactional data, and external market data, requires sophisticated data management and analytics capabilities. Banks must invest in robust data infrastructure and analytical tools to extract actionable insights from

disparate data sources. Another challenge is ensuring data quality and accuracy. Inaccurate or incomplete data can lead to flawed insights and ineffective advertising strategies. Banks must implement rigorous data validation and cleansing processes to ensure the reliability of their data sources .

1.2 Problem Overview

Banks, positioned as the primary clients in this context, are motivated by the imperative to fortify their deposit business through refined advertising strategies for fixed-term deposit products. The core of understanding banks' needs lies in recognizing their multifaceted objectives, which encompass augmenting customer acquisition, fostering engagement, and ultimately, amplifying profitability. In essence, banks aspire to harness advertising endeavors as a means to heighten their competitive edge within the dynamic financial services sector. Their overarching goal is to position fixed-term deposit products prominently in the minds of consumers, thereby translating awareness into action through increased subscription rates. This involves crafting campaigns that not only capture attention but also resonate deeply with target audiences, compelling them to consider and ultimately opt for these deposit offerings. Moreover, banks aim to cultivate lasting relationships with customers, built on trust and value. By optimizing advertising efforts, they seek to nurture engagement and loyalty, fostering a sense of affinity that transcends mere transactions. This entails aligning messaging and communication channels with customer preferences, ensuring relevance and resonance at every touchpoint. Furthermore, the quest to enhance profitability underpins banks' endeavors to maximize the return on investment (ROI) from their advertising endeavors. By deploying resources judiciously and refining tactics based on data-driven insights, banks endeavor to optimize campaign performance and yield measurable results in terms of increased deposits and revenue generation. Overall, the identification of banks as the primary clients underscores 11 their pivotal role in driving the strategic direction of advertising initiatives for fixed-term deposit products. Understanding their needs involves recognizing their aspirations for growth, engagement, and profitability, and tailoring advertising strategies accordingly to meet these objectives effectively within a fiercely competitive market landscape.

2. LITERATURE SURVEY

Introduction:

Securing text transfer system using cloud is a crucial topic, especially as more businesses and individuals are moving towards cloud-based solutions. In this literature survey, we will review the existing research on the topic to identify the most effective methods for securing text transfer using cloud.

Literature Review:

[1] Athanassopoulos, Antreas D., Non-Parametric Frontier Models for Assessing the Market and Cost Efficiency of Large Scale Bank Branch Networks. Journal of Money, Credit, and Banking, Vol. 30 No. 2, May,1998

In this paper we propose models for assessing the efficiency inlarge networks of bank branches. We distinguish bank branch efficiency into market and cost components suitably modified to capture different tiers of bank-management. The paper proposes a methodology which includes the use of multivariate analysis in order to ensure the homogeneity of the branches assessed and then data envelopment analysis for assessing efficiency. The methodology is applied on a sample of 580 branches of a commercial bank in the UK. The results obtained reinforced previous claims regarding the presence of high technical inefficiencies and economics/diseconomies of scale at the branch level from a production and cost point of view. Furthermore, the decision to pre-cluster the network of branches into homogenous groups has had profound implications on the magnitude of the assessed efficiencies.

[2] Y. Freund, R. Shapire Computational Learning Theory, 1995, Volume 904 A decision-theoretic generalization of on-line learning and application to boosting

We consider the problem of dynamically apportioning resources among a set of options in a worst-case on-line framework. The model we study can be interpreted as a broad, abstract extension of the well-studied on-line prediction model to a general decision-theoretic setting.

We show that the multiplicative weight-update rule of Littlestone and Warmuth [10] can be adapted to this mode yielding bounds that are slightly weaker in some cases, but applicable to a considerably more general class of learning problems. We show how the resulting learning algorithm can be applied to a variety of problems, including gambling, multiple-outcome prediction, repeated games and prediction of points in $\mathbb{R}n$. We also show how the weight-update rule can be used to derive a new boosting algorithm which does not require prior knowledge about the performance of the weak learning algorithm.

[3] Jijo, Bahzad & Mohsin Abdulazeez, Adnan. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. Journal of Applied Science and Technology Trends. 2. 20-28

Decision tree classifiers are regarded to be a standout of the most well-known methods to data classification representation of classifiers. Different researchers from various fields and backgrounds have considered the problem of extending a decision tree from available data, such as machine study, pattern recognition, and statistics. In various fields such as medical disease analysis, text classification, user smartphone classification, images, and many more the employment of Decision tree classifiers has been proposed in many ways. This paper provides a detailed approach to the decision trees. Furthermore, paper specifics, such as algorithms/approaches used, datasets, and outcomes achieved, are evaluated and outlined comprehensively. In addition, all of the approaches analyzed were discussed to illustrate the themes of the authors and identify the most accurate classifiers. As a result, the uses of different types of datasets are discussed and their findings are analyzed.

[4] Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. New York: Chapman & Hall.research collaborations using link prediction and random forest classifiers .

We introduce a method to predict or recommend high-potential future (i.e., not yet realized) collaborations. The proposed method is based on a combination of link prediction and machine learning techniques. First, a weighted co-authorship network is constructed. We calculate scores for each node pair according to different measures called predictors. The resulting scores can be interpreted as indicative of the likelihood of future linkage for the given node pair. To determine the relative merit of each predictor, we train a random forest classifier on older data. The same classifier can then generate predictions for newer data. The top predictions are treated as recommendations for future collaboration. We apply the technique to research collaborations between cities in Africa, the Middle East and South-Asia, focusing on the topics of malaria and tuberculosis. Results show that the method yields accurate recommendations. Moreover, the method can be used to determine the relative strengths of each predictor.

Conclusion:

In conclusion, navigating contemporary challenges in advertising fixed-term deposit products requires banks to adopt agile, data-driven strategies tailored to the dynamic nature of the financial landscape. By staying abreast of economic market volatility, fluctuating interest rates, regulatory changes, evolving consumer preferences, and the proliferation of digital channels, banks can position themselves for success in promoting fixed-term deposit products. By leveraging predictive models and advanced analytics, banks can stay competitive and effectively reach their target audience amidst a rapidly evolving market environment.

2.1 Existing System

The existing systems for identifying potential customers for a banking institution typically involve a combination of traditional and digital methods, as well as data-driven approaches. Here's an overview of the existing systems:

1. Customer Relationship Management (CRM) Systems:
   - Banking institutions utilize CRM systems to manage interactions with current and potential customers. These systems store customer data, communication history, and transaction details, allowing banks to track customer behavior and preferences over time.

2. Transaction Data Analysis:
   - Banking institutions analyze transaction data to identify patterns and trends indicative of potential customers for specific products, including fixed-term deposits. This analysis may involve identifying customers with stable income, high savings balances, or frequent transactions indicative of financial stability and savings habits.

3. Market Segmentation:
   - Banks segment their customer base into different groups based on demographics, financial behaviors, and preferences. Market segmentation allows banks to tailor their marketing efforts and product offerings to specific customer segments, including those likely to be interested in fixed-term deposit products.

4. Lead Scoring:
   - Lead scoring models are used to prioritize potential customers based on their likelihood to purchase specific products or services. These models assign scores to individual leads based on various criteria, such as income level, credit score, banking history, and engagement with marketing campaigns.

5. Digital Marketing and Advertising:
   - Banking institutions leverage digital marketing channels, including social media, search engine advertising, email marketing, and targeted display ads, to reach potential customers online. Digital advertising campaigns are often targeted based on demographic and behavioral data, allowing banks to reach individuals with relevant financial needs and interests.

6. Data Mining and Predictive Analytics:

- Banks use data mining and predictive analytics techniques to analyze large datasets and identify patterns indicative of potential customers. Predictive models may be used to forecast customer behavior, such as the likelihood of opening a fixed-term deposit account based on historical data and customer attributes.

7. Customer Acquisition Strategies:

- Banking institutions implement customer acquisition strategies to attract potential customers to their products and services. These strategies may include offering promotional incentives, discounts, or special offers to encourage customers to open fixed-term deposit accounts.

8. Cross-Selling and Upselling:

- Banks cross-sell and upsell their products and services to existing customers based on their financial needs and preferences. Customer data and predictive analytics are used to identify opportunities for cross-selling fixed-term deposit products to customers who already have other banking relationships.

Overall, the existing systems for identifying potential customers for banking institutions combine data analysis, market segmentation, digital marketing, and customer relationship management to target individuals with the highest likelihood of being interested in fixed-term deposit products. These systems aim to optimize customer acquisition efforts and drive business growth while providing value to customers based on their financial goals and needs.

2.2 Proposed System

3. Problem Formulation

The integration of data science and machine learning techniques offers significant potential for enhancing bank marketing campaigns. By leveraging advanced analytical tools and methodologies, banks can gain valuable insights into customer behavior and develop targeted strategies to maximize campaign success. Decision Trees are considered to be one of the most popular approaches for representing classifiers. Researchers from various disciplines such as statistics, machine learning, pattern recognition, and Data Mining have dealt with the issue of growing a decision tree from available data.

This paper presents an updated survey of current methods for constructing decision tree classifiers in a top-down manner. The chapter suggests a unified algorithmic framework for presenting these algorithms and describes various splitting criteria and pruning methodologies.

Decision tree classifiers are regarded to be a standout of the most well-known methods to data classification representation of classifiers. Different researchers from various fields and backgrounds have considered the problem of extending a decision tree from available data, such as machine study, pattern recognition, and statistics. In various fields such as medical disease analysis, text classification, user smartphone classification, images, and many more the employment of Decision tree classifiers has been proposed in many ways. This paper provides a detailed approach to the decision trees.

Furthermore, paper specifics, such as algorithms/approaches used, datasets, and outcomes achieved, are evaluated and outlined comprehensively. In addition, all of the approaches analyzed were discussed to illustrate the themes of the authors and identify the most accurate classifiers.

As a result, the uses of different types of datasets are discussed and their findings are analyzed. Presently, Customer Churn Prediction (CCP) becomes a tedious task among decision-makers and machine learning (ML) communities. Since the Internet of Things (IoT) and Cloud Computing (CC) platform generates a massive amount of customer data, it is necessary to construct a CCP model using the customer data from IoT devices. In this view, this paper devises a new model using optimal meta-heuristic based feature selection with Gradient Boosting Tree 34 classification for CCP. The presented CCP model

involves four main processes, namely data acquisition, preprocessing, feature selection and classification.

At the earlier stage, the data collection process takes place utilizing IoT gadgets such as laptops, smartphones, wearable, etc. The IoT gadgets transmit the sensed data to the cloud data server (CDS). Then, gathered data is preprocessed and the missing values are imputed effectively. Next, the ant colony optimization (ACO) algorithm is applied as a feature selector to select an optimal set of features.

Finally, the GBT algorithm is employed as a classifier to classify the data into churn or non-churn. A comprehensive simulation was performed to indicate the betterment of the proposed model. The experimental results stated that the ACO GBT model has reached a maximum sensitivity of 95.82%, specificity of 74.59%, accuracy of 92.71%, Fscore of 95.73% and kappa value of 70.71%. In this paper, we propose models for assessing the efficiency in large networks of bank branches. We distinguish bank branch efficiency into market and cost components suitably modified to capture different tiers of bank-management. The paper proposes a methodology which includes the use of multivariate analysis in order to ensure the homogeneity of the branches assessed and then data envelopment analysis for assessing efficiency.

3. Problem Formulation

In the contemporary landscape of advertising inundated with digital noise, businesses, especially those in highly competitive sectors like banking, struggle to effectively capture consumer attention and promote their products, such as fixed-term deposit offerings. The primary challenge lies in devising advertising strategies that cut through the clutter and resonate with the target audience, ultimately driving customer engagement and business growth.

Given this scenario, the problem formulation for banks aiming to optimize their marketing strategies for fixed-term deposit products can be articulated as follows:

1. Identifying Target Audience:
   - How can banks accurately identify and define their target audience for fixed-term deposit products amidst a diverse consumer base with varying financial needs and preferences?

2. Understanding Consumer Behavior:
   - What are the key drivers influencing consumer behavior and decision-making processes regarding fixed-term deposit investments?

3. Leveraging Data Insights:
   - How can banks effectively collect, analyze, and leverage data from past marketing campaigns, customer interactions, and market trends to inform their advertising strategies?

4. Personalization and Segmentation:
   - What methodologies and technologies can banks employ to segment customers based on demographic, psychographic, and behavioral attributes, enabling personalized advertising campaigns?

5. Measuring Campaign Effectiveness:
   - What metrics and KPIs should banks prioritize to evaluate the effectiveness and ROI of their advertising initiatives for fixed-term deposit products?

6. Adapting to Market Dynamics:

- How can banks adapt their marketing strategies in response to changing market dynamics, competitor actions, and regulatory factors impacting the banking sector?

7. Integration of Technology:

- How can banks integrate advanced technologies such as machine learning, predictive analytics, and marketing automation tools to enhance the efficiency and efficacy of their advertising efforts?

8. Ensuring Compliance and Trust:

- How can banks ensure that their advertising strategies comply with regulatory standards and industry best practices while fostering trust and credibility among consumers?

By addressing these key questions and challenges, banks can develop comprehensive and data-driven advertising strategies tailored to promote fixed-term deposit products effectively, thereby attracting more customers and achieving business growth objectives in a competitive market environment.

4. RESEARCH OBJECTIVES

Research Objective:

To develop and evaluate machine learning models for predicting the effectiveness of bank marketing campaigns, specifically focusing on subscription rates for fixed-term deposit products, leveraging data-driven insights and advanced techniques.

1. Model Development:

- Develop and compare machine learning models, including logistic regression, decision trees, and ensemble methods, for predicting customer subscription rates to fixed-term deposit products based on demographic, campaign, and market data.

2. Feature Selection and Engineering:

- Identify relevant features and perform feature engineering to enhance the predictive power of the models. Explore the impact of different feature combinations, including customer demographics, campaign attributes, and market dynamics, on the prediction accuracy.

3. Data Analysis and Insights:

- Analyze historical marketing campaign data to uncover patterns and trends influencing customer behavior, such as response rates and conversion rates for fixed-term deposit subscriptions. Extract actionable insights to inform marketing strategies and campaign optimization.

4. Model Evaluation and Performance Metrics:

- Evaluate the performance of the developed models using appropriate evaluation metrics, such as accuracy, precision, recall, F1-score, and ROC-AUC. Compare the effectiveness of different models and identify the most suitable approach for predicting subscription rates.

5. Personalization and Segmentation:

- Investigate the effectiveness of personalized marketing campaigns tailored to individual customer segments based on demographic, behavioral, and financial attributes. Explore segmentation strategies to target specific customer groups more effectively.

6. Integration of Advanced Techniques:

- Explore the integration of advanced techniques such as ensemble learning, gradient boosting, and neural networks to improve the predictive accuracy and robustness of the models. Assess the scalability and computational efficiency of these techniques for real-world applications.

7. Practical Recommendations for Banks:

- Provide practical recommendations for banks to optimize their marketing strategies for promoting fixed-term deposit products based on the research findings. Suggest actionable insights for campaign design, audience targeting, messaging, and budget allocation.

8. Ethical Considerations and Regulatory Compliance:

- Consider ethical considerations and regulatory compliance requirements in the development and deployment of predictive models for marketing purposes. Ensure transparency, fairness, and privacy protection in data collection, analysis, and use for decision-making.

5. METHODOLOGY

The dataset was provided by the U. C. Irvine Machine

The first step is to load the dataset into a dataframe for easy manipulation and exploration using the *pandas* package. The 'duration' feature was dropped due to the risk of data leakage. This feature measures the length of the phone call between the bank's marketing representative and the customer. Since this time cannot be known until after the call has ended (when the outcome for that customer is already known), including it in a predictive model would not provide realistic results.

The next step was to explore and clean the categorical variables such as 'job type,' 'marital status,' 'education,' etc. Plots for each were produced that looked at their relative frequency as well as normalized relative frequency. In Python, these graphs are created using the *seaborn* package.

Many of these features contain unknown values so the next question is how to deal with this missing data. Simply discarding these rows would lead to a huge reduction in the amount of data and thus greatly interfere with the results. Instead, these missing values are imputed using other inde- pendent variables to infer the missing values. While this does not guarantee that all the missing data will be restored, a majority of it will be. For instance, cross-tabulation between 'job' and 'education' was used based on the hypothesis that a person's job will be influenced by their education. Thus, a person's job is used to predict their education level. The Python function *cross tab* was created for this cross-tabulation step. A similar cross-tabulation process was carried out for the 'house ownership' and 'loan status' features. It's important to note that in making these imputations, care was taken to ensure the correlations made sense in the real world. If not, the values were not replaced. Throughout this process, dataframes using the *pandas* package were invaluable. Python provides quickness, ease of modifiability and ease of replacement of values throughout the dataset thanks to this tool.

The next task is to deal with missing data among the numerical features. In this particular dataset, all missing values were encoded as '999.' It's quickly noted that while only the 'pdays' (number of days since that customer had been contacted from the previous campaign)

column contained suchvalues, they made up the majority of the data for this feature. In other words, this column was missing more data than it contained. Further exploration showed that this missingness was due to customers who had not been contacted previously at all. To deal with this, the numerical feature ‘pdays’ was replaced with a categorical feature based on whether the customer had never been contacted, contacted 5 or less days ago, 6-15 days ago, etc.

Finally, a heatmap was created to show us whether there is strong correlation between the target variable and any independent variables. The heatmap is created using Spear- man correlation, which measures the degree to which the rankings of each variable (as opposed to the actual values) align, thus minimizing the effect of outliers[2]. Once this is measured, those variables are expected to be significant during the modeling stage. This graphic was created using Python’s *seaborn* package and the specially written function *drawheatmap*, which takes a dataframe as an input. The code for this function can be seen in the Jupyter notebook for this project.

6. RESULT

Based on the feature importance plot, some recommendations can be made to the bank's marketing team:

Fig. Most important Features based on the AdaBoost model

- The marketing team should collaborate with economic experts so that as soon as they have some signals indicating the Libor going up (or the economic situation improving, i.e., consumer price index or consumer confidence index goes up), they can expect more customers to subscribe for the term deposit and should pro-actively reach out to them before the bank's competitors do.
- The marketing team should target relatively old age customers who would be looking for safe and profitable investment options. The marketers should ensure to con- vey the peace of mind and steady source of income these products provide as a value proposition to these customers.
- Although the 'duration' (length of marketing phone call) variable was not used in the prediction models for various reasons cited earlier, the correlation of the 'duration' variable with the target variable shows that the higher the duration, the more likely it is that the customer will subscribe to the term deposits (correlation = 0.405). This makes intuitive sense because longer duration shows that the customer is interested in the product. Hence, the marketers should try to make the call engaging and increase the duration of the call.
- The telephone seems to be the most preferred mode of communication.

- The marketing team should prioritize those customers to whom they previously reached out during previous campaigns. They are likely to subscribe for the term deposit.

Document the system: Document the system architecture, security controls, and testing results. This documentation will be useful for future reference, maintenance, and updates.

Deploy the system: Deploy the secure text transfer system in the cloud environment. Monitor the system for any issues or incidents and update the security controls as necessary.

Evaluate the system: Evaluate the effectiveness of the secure text transfer system by collecting feedback from users and monitoring performance metrics. Use this feedback to make improvements and enhancements to the system.

7. CONCLUSION

From this project, we learned how banks can improve their marketing campaigns by focusing their efforts on certain prime-grade clients and also how they can recognize market conditions which are favorable to increase client subscription for the fixed-term products they are offering. All of this was possible by implementing data science and machine learning methods in Python. Tools such as dataframes, arrays, for loops, etc. were all critical for the success of this project. A large number of other tools and techniques from the Python for Data Science course were used and these were invaluable for making our analyses and predictions. This project demonstrated how powerful Python can be for data science applications.

8. REFERENCE

[1] Rokach, Lior & Maimon, Oded. (2005). Decision Trees. 10.1007/0-387-25465-X_9.

[2] Jijo, Bahzad & Mohsin Abdulazeez, Adnan. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. Journal of Applied Science and Technology Trends. 2. 20-28

[3] S. Venkatesh Dept. of Computer Science Metaheuristic based Optimal Feature Subset Selection with Gradient Boosting Tree Model for IoT Assisted Customer Churn Prediction.

[4] Athanassopoulos, Antreas D., Non-Parametric Frontier Models for Assessing the Market and Cost Efficiency of Large Scale Bank Branch Networks. Journal of Money, Credit, and Banking, Vol. 30 No. 2, May,1998

[5] Y. Freund, R. Shapire Computational Learning Theory, 1995, Volume 904 A decision-theoretic generalization of on-line learning and application to boosting

[6] Fawagreh, K., Gaber, M. M., & Elyan, E. (2014). Random forests: from early developments to recent advancements. Systems Science & Control Engineering, 2(1), 602–609

[7] Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. New York: Chapman & Hall.research collaborations using link prediction and random forest classifiers .