Abstract

This report explores the development and comparison of different recommender systems for personalized financial product recommendations. We aimed to recommend products like credit cards, loans, savings accounts, and insurances based on customer data. We implemented three recommendation methods: rule-based, machine learning (ML) classification, and content-based filtering. Our analysis revealed that rule-based recommendations, driven by predefined business rules, achieved perfect accuracy in controlled conditions. However, this approach may miss complex real-world customer needs. Among the ML models tested, Random Forest and Support Vector Classification (SVC) performed best, showcasing their ability to handle diverse data patterns effectively. Content-based filtering also showed high accuracy in aligning products with individual customer profiles. Integrating these systems can improve customer engagement and satisfaction in the financial sector. Future work should focus on continuous adaptation to keep up with evolving customer expectations and financial landscapes.

Table of Contents

List of acronyms

EDA: Exploratory Data Analysis

ML: Machine Learning

SVM: Support Vector Machine

KNN: k-Nearest Neighbors

SVC: Support Vector Classification

AI: Artificial Intelligence

Introduction

Background

In the financial sector personalized product recommendations boost customer satisfaction. Financial institutions analyze customer data and tailor suggestions for credit cards, loans, savings accounts and insurances. This improves retention and cross-selling opportunities. The issue addressed in this project is the development and comparison of different recommender system approaches to provide personalized financial product recommendations. Specifically, we investigated how rule-based and content-based recommender systems can be utilized to suggest suitable financial products based on customer data, and which method proves more effective in terms of accuracy.

Objectives

The primary objectives of this report are:

1. To develop a recommender system that provides personalized recommendations for financial products such as credit cards, loans, savings accounts, and insurance options.
2. To implement and compare rule-based, ML classifications and content-based recommendation approaches.
3. To evaluate the effectiveness of each approach based on set evaluation metrics based on methods.

Project Description

This project focuses on recommending personalized financial products such as credit cards, loans, savings accounts, investment options, and insurance products based on user data. By utilizing both rule-based and content-based recommendation systems, we aim to compare their effectiveness and insights.

Scope

Our project is confined to the available customer data and predefined financial products. External factors influencing customer choices and the integration of advanced deep learning algorithms are beyond the current explored scope.

Methodology

Business Understanding

Many financial institutions have prioritized personalization as a majority of customers have already moved to primarily digital relationships, with 45% using mobile and 27% managing their everyday banking needs from the web. When we dive deep into how the banking sector does the current recommendations, we can see the different rules associated with each feature a user has. Below are the rules we considered in this project(Whitestone, n.d.).

1. **Credit Cards:**
    - **Upgrade/Downgrade Cards:** Based on spending patterns, credit limit, and usage (e.g., Avg_Utilization_Ratio), users can be recommended to upgrade to a premium card or downgrade to a more suitable card.
    - **Cashback/Rewards Cards:** For users with high total transaction amount and total transaction credit, cards offering better rewards or cashback are suggested.
2. **Loans:**
    - **Personal Loans:** Recommended for users with a high credit limit and consistent credit usage for big purchases or debt consolidation.
    - **Auto Loans:** Suggested for users aged 30-50 with stable incomes.
    - **Home Loans:** Suitable for married users with higher income categories.
3. **Savings and Investment Products:**
    - **High-Yield Savings Accounts:** For users with high Avg_Open_To_Buy, these accounts offer better interest rates.
    - **Investment Accounts:** Users with stable income and high education levels might be interested in mutual funds, stocks, or retirement accounts.
4. **Insurance Products:**
    - **Life Insurance:** Recommended for users with dependents (Dependent_count).
    - **Health Insurance:** Suitable for users in higher age brackets or with more dependents.

Data Collection

The dataset used for the project consists of 10,000 customers and contains 18 columns. The features include demographic information and financial status of the customers. Some of the columns are coverage, salary, marital status, credit card limits, credit card categories, total spending amount, credit card usage frequency, debt status, past payment habits. The scope of the dataset provides sufficient detail to understand various financial behaviours and profiles of the customers. This dataset is found in kaggle and it is attached at the references section for reference (Goyal, n.d.).

Data analysis and understanding

The initial steps taken were just to explore the high level details in the dataset. The dataset has no missing values therefore no mitigations were needed there. The numerical columns' descriptive analysis was the next step we took.

- **Credit_Limit**: Indicates the maximum amount a customer can charge to their credit card, useful for understanding financial flexibility. The maximum limit for our data is 34516 while the mean is 8631.95.
- **Total_Trans_Amt**: The total transaction amount, reflecting spending behavior of our users. The mean amount being 4404.08 and standard deviation of 3397.
- **Avg_Utilization_Ratio**: Shows how much of the credit limit is being used on average, important for assessing credit risk. The mean average utilizing ratio is 0.27.
- **Total_Revolving_Bal**: Represents the unpaid balance carried month to month, relevant for understanding customer debt levels which will affect the loan recommendations later on.
- **Total_Trans_Ct**: The total number of transactions, indicating usage frequency.

Given our project's focus, we outline a streamlined EDA approach that targets the most relevant aspects of your dataset. This will include an analysis of key variables that are likely to influence product recommendations such as customer age, credit limit, transaction behavior, and categorical variables like income and card categories. Then we moved on to plotting graphs to see the distributions and correlations between features.

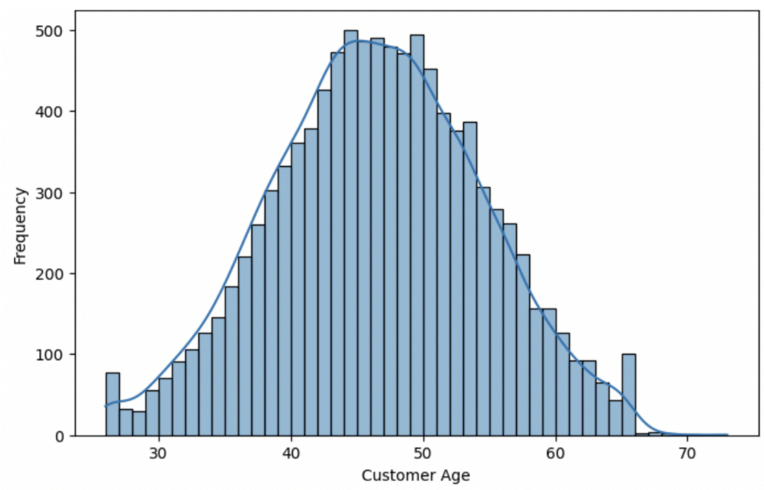

Figure 1. Distribution of customer age in the dataset

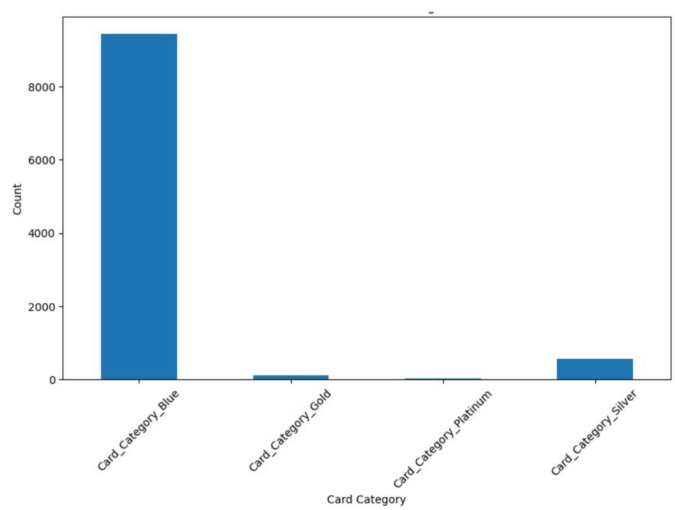

Figure 2. Distribution of credit limit to card category

The above figure shows that there are four card types, these show us that the majority of our users use the Blue card which is the ordinary credit card and only a very small number of them have platinum which makes sense in a real world scenario. This is not a scenario that needs to be mitigated as an outlier because the rareness actually helps our engine understand the real world scenario.

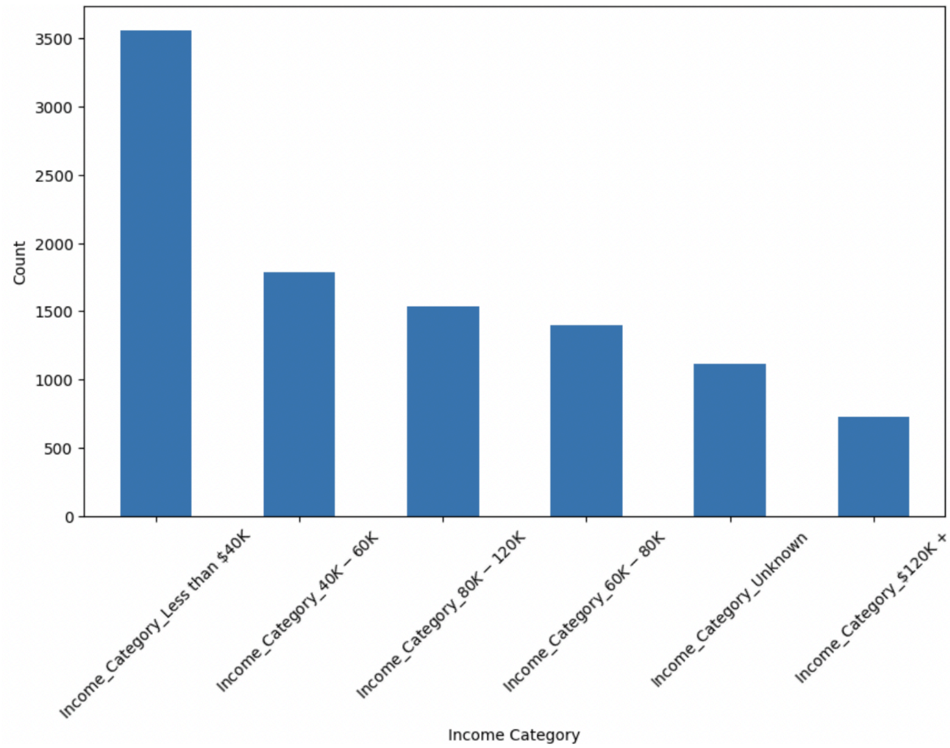

Figure 3. Distribution of the income categories by range

The above figure shows the distribution of income in our dataset categorized into ranges. The distribution is skewed to the left but due to the labels being stored in a random order, normalization of this data is not feasible at this stage. Therefore, we kept this feature as it is.

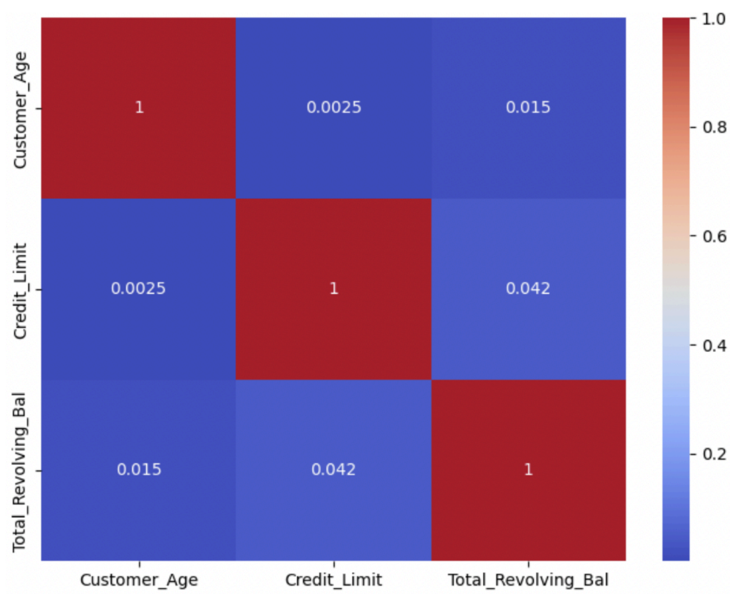

Figure 4. Correlation matrix between the most important numerical features

Recommendation Methods

1. Rule-based recommendation

A rule-based recommender system utilizes predefined rules to suggest items to users based on specific criteria and conditions. These rules are often crafted by domain experts and are based on logical inferences derived from user data and behavior patterns. The system matches user profiles or actions with these rules to provide relevant recommendations. This approach is particularly effective in scenarios where clear, well-understood relationships exist between user actions and suitable recommendations, allowing for straightforward and transparent suggestion mechanisms (Apolloni et al., 2020).

The Rule-Based model is utilized in this project to provide quick and easily understandable recommendations based on predefined business logic. Given the structured nature of the financial dataset, with clear attributes and a reasonably clear business understanding, rule-based recommendation can effectively capture straightforward eligibility criteria for different financial products.

We implemented a function 'determine_product' based on our understanding and details we found that makes a user eligible for certain financial products. The function takes in one row of input data and returns the recommended product for the user represented. Customer segmentation is done in the rule-based model by defining conditions under which specific financial products are recommended based on customer attributes.

By Financial Behavior:

Credit Card Recommendations:

Customers are segmented based on their credit limit and average utilization ratio. Transaction amounts and counts are used to identify customers who could benefit from cashback or rewards cards, segmenting them by their spending patterns.

By Demographic Factors:

Loan Products:

Age and income are used to segment customers for auto loans, while marital status and income category are criteria for home loan recommendations, targeting married customers within higher income brackets.

By Financial Needs and Goals:

Savings and Investment Products:

Customers with a high average open to buy are considered for high-yield savings accounts. Investment accounts are suggested for those with stable incomes and presumably higher financial literacy or investment interests, indicating a different segment of customers likely focused on building wealth.

By Protection Needs:

Insurance Products:

Life insurance products are recommended based on the presence of dependents, segmenting customers by their family responsibilities. Health insurance recommendations consider age and dependent count.

2. **ML models/ Classification**

In the paper "A Survey of Recommendation Systems: Recommendation Models, Techniques, and Application Fields" by Ko et al. (2022), classification models are highlighted as effective methods for recommendation systems. These models are used to categorize user preferences and predict the most relevant items for recommendation.

We used 5 models to compare the effectiveness. We used Random Forest, Logistic Regression, SVM, GradientBoosting, KNN. Before we applied the models on our dataset, we made a pipeline that scales the data using standard scaler and a dimensionality reduction level so we consider features in a more manageable way computing wise. For splitting of the data into train and test, we considered 75% of the data to train the model and the other 25% to test it.

3. **Content-based recommendation**

Content-based recommendation systems utilize the inherent attributes of items to make personalized suggestions by analyzing item properties and features instead of relying on user interactions alone.

Content-Based Filtering is implemented to personalize financial product recommendations by analyzing the attributes of both the customers and the products. Using features like income categories, marital status, and spending patterns, this model matches products to users based on attribute similarities.

Formula:

Cosine Similarity between two vectors A and B:

$$\text{cosine\_similarity}(A, B) = \frac{A \cdot B}{\|A\| \|B\|}$$

Where:

- $A \cdot B$ is the dot product of vectors A and B.
- $\|A\|$ is the magnitude of vector A.
- $\|B\|$ is the magnitude of vector B.

The first step is to identify the relevant features of items. This includes both quantitative attributes, such as Credit_Limit and Customer_Age, and categorical features like Income_Category. Afterwards, cosine similarity is used to measure how closely each item's features align with the user's profile. This metric quantifies the similarity between vectors, providing a score that reflects how well each item matches the user’s preferences. Finally, the similarity scores are added to the dataset, and the items are sorted based on these scores. The top N items with the highest similarity scores are selected as recommendations. In this case, we used the top 5 items.

Findings and Discussion

1. Rule-based recommendation

The rule-based model demonstrated perfect accuracy with a score of 1.0 in our evaluation. This indicates that the model was able to correctly classify all instances in the dataset without any errors. This is because we defined the rules explicitly for the system. Rule-based models often achieve high accuracy in controlled environments where rules are explicitly defined to handle all cases correctly. However, they have to always be taken with a grain of salt because there is no guarantee that in a complex and real-world scenario, this engine would make the best decisions. There is no real way of evaluating this. For instance, if a loan wasn't recommended for a particular customer, there is no absolute certainty that that user wouldn't have taken the loan and returned it at the right time with interest. There could also be cases where complex case scenarios are not taken into consideration.

2. ML models/Classification

The evaluation of our models reveals that the Random Forest classifier and Support Vector Classification (SVC) both achieved the highest average cross-validation scores of approximately 0.888. The Random Forest slightly outperformed SVC, which is why it is going to be used to compare the classification method with the rule-based and content-based methods in later stages. Gradient Boosting followed with an average score of about 0.873, while Logistic Regression achieved a solid score of approximately 0.843. On the other hand, k-Nearest Neighbors (k-NN) obtained the lowest average score of around 0.830 and was less effective compared to the other algorithms. These findings highlight that while ensemble methods like Random Forest and Gradient Boosting are generally more accurate in this case.

Category	Precision	Recall	F1-Score	Support
Auto Loan	0.89	0.92	0.90	214
Cashback/Rewards Card	0.97	0.99	0.98	168
Downgrade to Basic Card	0.76	0.77	0.77	238
Flexi Balance Transfer Card	0.81	0.57	0.67	30
Health Insurance	0.91	0.79	0.85	87
Home Loan	0.93	0.98	0.95	92
Investment Account	0.89	0.88	0.88	88
Life Insurance	0.89	0.93	0.91	1138
Personal Loan	0.89	0.83	0.86	444
Platinum Card	1.00	0.00	0.00	1
Standard Card	0.75	0.38	0.50	32
Accuracy			**0.89**	2532
Macro Avg	**0.88**	**0.73**	**0.75**	2532
Weighted Avg	**0.88**	**0.89**	**0.88**	2532

Figure 5. Cross validation scores for each class in random forest model

3. Content-based recommendation

Content-based filtering showed high similarity scores, with the top recommended products being 'Personal Loan' and 'Life Insurance'. The similarity scores for these recommendations were extremely close to 1. This indicates a high relevance to the user's profile. For instance, 'Personal Loan' appeared with similarity scores of 0.9999, while 'Life Insurance' scored 0.9999 as well. This method effectively identifies products that align closely with individual customer preferences, providing highly personalized suggestions.

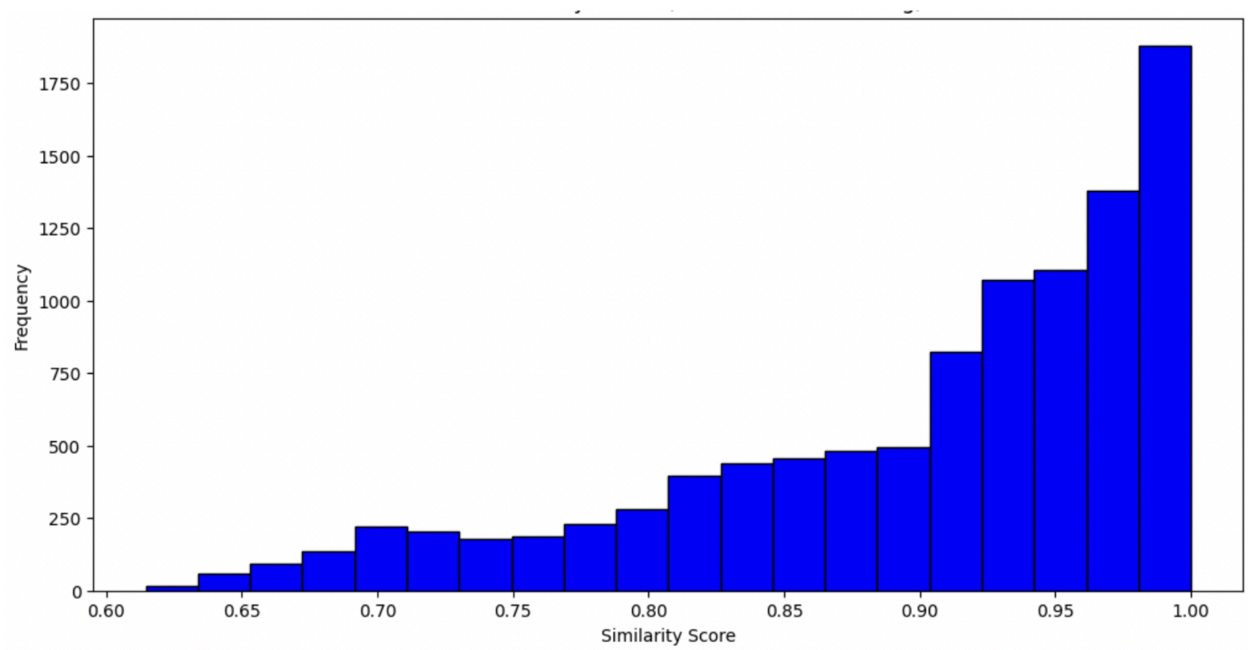

Figure 6. Distribution of similarity scores

The histogram shown above illustrates the distribution of cosine similarity scores calculated between a user profile and a set of items in our dataset. The distribution is right-skewed, indicating that a significant number of items have high similarity scores relative to the user profile. This skew towards higher similarity scores is typically positive in a recommendation system, suggesting that many items are relevant to the user's interests or needs as encapsulated by the user profile.

Conclusion

This project analyzes the effectiveness of different recommender systems—rule-based, machine learning, and content-based filtering—for personalized financial product recommendations. The rule-based method, although precise in controlled environments, may not completely grasp intricate, real-life customer preferences. Machine learning algorithms, such as Random Forest and SVM, have shown resilience in efficiently managing various data patterns, providing more customized suggestions. Content-based filtering stood out in matching recommendations closely with unique user profiles, showcasing its value in providing personalized financial product recommendations.

The findings underscore the significant potential of integrating these systems to enhance customer engagement and satisfaction in the financial services sector. Each method brings unique strengths, suggesting a combined approach could leverage the precision of rule-based systems with the adaptability and personalization of machine learning and content-based models.

Looking at the bigger picture, this project emphasizes how advanced recommender systems are changing the financial industry. With the increasing complexity of customer data, these systems offer vital resources for offering better product suggestions, which in turn helps build stronger customer connections and boosts business expansion. Future enhancements and ongoing adaptations to these systems are crucial to keep pace with the evolving financial landscape and customer expectations.

References

- https://www.dynamicyield.com/article/recommendations-in-financial-services/
- Apolloni, B., Karakatic, S., Imran, H., 2020. A Rule-Based Recommender System. [online] Available at: https://www.researchgate.net/publication/304108078_A_Rule_Based_Recommender_System
- Ko, H., Lee, S., Park, Y., & Choi, A. (2022). A Survey of Recommendation Systems: Recommendation Models, Techniques, and Application Fields. Electronics, 11(1), 141. Available at: https://www.mdpi.com/2079-9292/11/1/141
- Whitestone, E. (n.d.). The Power of Recommendations in Financial Services. Dynamic Yield. Available at: https://www.dynamicyield.com/article/recommendations-in-financial-services/
- Goyal, S. (n.d.). Credit Card Customers. Kaggle. Available at: https://www.kaggle.com/datasets/sakshigoyal7/credit-card-customers/data