Sci. Educ. Innov. Context Mod. Probl. P-ISSN: 2790-0169 E-ISSN: 2790-0177
Issue 3, Vol. 8, 2025, IMCRA

Article

Recommender System for Banking Industry with Collaborative Filtering and XGBoost Classifier

Anee Pritam
anee_pritam@siu.edu.in, ORCID: 0000-0001-7304-4934
Symbiosis Centre for Information Technology, Symbiosis International (Deemed University), Pune, India.

Dhanya Pramod
dhanya@scit.edu, ORCID: 0000-0003-3451-9794
Symbiosis Centre for Information Technology, Symbiosis International (Deemed University), Pune, India.

Received date: 14.01.2025; Accepted date: 16.03.2025;
Publication date: 05.05.2025, doi: 10.56334/sei/8.3.50

Abstract

As netizens, we interact with recommendation systems daily - while using movie or music streaming services, dating apps, shopping online, browsing social media, or while usual Google searches. Recommendation systems are one of the most popular and heavily used AI applications. Amazon and Netflix use machine learning and AI to power up their recommendation systems. These systems can effectively increase sales, revenue, click-through rates, and customer conversions. Customizing product or content recommendations based on the preferences of a particular user creates a positive effect on the user experience. This FinTech system can be a revolution in the banking industry by assisting banks in growing revenue, containing costs, providing customer satisfaction, and bringing brand loyalty. This study aims to improve customer engagement in the banking sector by providing suitable product recommendations to the right customer using a combination of collaborative filtering and classification approaches. In this work,

Citation. Anee P, Dhanya P. (2025). Recommender System for Banking Industry with Collaborative Filtering and XGBoost Classifier. *Science, Education and Innovations in the Context of Modern Problems*, 8(3), 790-801. doi: 10.56352/sei/8.3.50. https://imcra-az.org/archive/358-science-education-and-innovations-in-the-context-of-modern-problems-issue-3-volviii-2025.html

a hybrid recommendation model is proposed to provide recommendations to banks' banking clients. The model is built on a Python environment to address various challenges related to recommendation systems in the banking industry.

Keywords: Netizens, Recommendation Systems, Click-through Rates, FinTech, Banking Industry, Customer Satisfaction, Brand Loyalty, AI, Collaborative Filtering, Classification

Introduction

Recommendation systems enable companies to change the way customers interact with websites and other online platforms. This, in turn, helps the companies to increase their ROIsby increasing customer conversions and user click-through rates. Amazon and Netflix use these tools to power up their business and provide personalized customer experiences. Customer demographic details are collected and analyzed to garner information regarding past purchases, product reviews, and customer behavior. Such information is used to assess how customers rate the relevant product sets or the possibility of purchasing an additional product.There are different types of recommendation systems. Other approaches are associated with each recommendation system.

Collaborative Filtering: It collects customer ratings for items and computes the similarity between them for recommending products. For instance, if person A likes iPhone and person B likes iPhone and OnePlus, person A will even like OnePlus.

Content-based Filtering: It recommends products to a user which have similar features to those purchased before. User interaction with products is not necessary as the recommendation is made purely based on product-related features.

Demographic-based Recommendation System: Demographic variables classify users into different segments, and products are recommended based on segments. Rating history of users is not required in this system.

Utility-based Recommendation System: Each customer's utility is calculated for recommending products. Utility function depends on the industry. Its advantage over a content-based recommendation system is that it is independent of product features.

Knowledge-based Recommendation System: Recommendations are made based on the demands and choices of customers. A user's requirements are matched with well-suited products using knowledge-based functions.

Hybrid Recommendation System: A conglomeration of any two recommendation systems described above can be called a hybrid system. It has added advantage of eliminating limitations of one recommendation system [3]. Companies in many industries have begun implementing recommendation systems to retain customers, improve the online shopping experience and increase sales. Business owners have recognized the ability to gather large amounts of information related to transactions and user buying patterns. Future interactions can be fruitful if systematic storage ofsuch information is done continuously.

In addition to improving the customer experience, the information collected can also be used as an advertising targeting tool. Advertising exchanges can have the ability to target other customers on the company website depending on user interactions by including a referral system.

Basic strategies like linking product recommendations related to verified purchases, gathering information about abandoned shopping carts, sharing trending products and purchase history of other customers, and making personalized recommendations can increase revenue to a great extent.

Stimulating e-mails based on online interactions is another way to make the most of information. For example, the company may e-mail a customer who views five phone pages with a coupon code to persuade them to purchase the product. Businesses can use reverse triggers via e-mails to target products that customers do not yet see.

The demand for recommendation engines is growing with the availability of more and more products online. They help increase sales and customer interactions and ensure customer loyalty by recommending personalized products in real-time. The recommendation engine is not that widely used in the banking industry. This study can be a game-changer in the fintech world.AI based hybrid approaches combining collaborative filtering and classification techniques to recommend appropriate products and services to customers. Personalization and customization can be taken to the subsequent level with a suitable recommendation approach for banks' customers. Research work aims to highlight the importance of recommendation engines for the banking domain. It can help increase the revenue of banks and improve customer satisfaction by recommending products and services of interest. Customer-centric or Product-centric offerings can be made based on past transactions and profiles of the customers. Highly appealing offers can increase retention rates and attract more customers. Such systems can improve with time and the use of other advanced machine learning algorithms.

Literature Review

FinTech is "disruptive," "revolutionary," and armed with "digital weapons" that will "tear down" barriers and traditional financial institutions. [4]It is changing economies around the world. Information technology plays a crucial role in transforming banking. Technological innovations like blockchain decentralized ledger technology, smartphones, artificial intelligence, cloud computing, chatbots, and recommendation systems are entering the financial services sector, including peer-to-peer lending. Data privacy is a significant concern for FinTech. As more and more financial systems go digital, cyber-attacks become a more substantial risk. According to PwC’s Global FinTech Survey 2016, almost 56% of the respondents identified information security and privacy as threats to the rise of fintech. Yet, it is an ample business opportunity for companies and consumers. Security specialists need to redesign the architecture for fintech and other digital sectors. Security is paramount for customers. Therefore, it is the provider's responsibility to ensure data security and privacy to strengthen customers' confidence.

Indian FinTech is one of the top five markets by the value of capital funding and investments in the sector, with nearly $270 million of funding in 2016 [8]. It can be ranged into six broad areas: credit, personal finance management, payments, investment management, bank-tech, and insurance tech. The financial services landscape is reshaping in India with FinTech start-ups, innovative models including big data and machine learning, and improved capabilities and culture. India is now the world's third-largest FinTech center with $3.7 billion investments in the sector in 2019 [10].

Recommendation System is a significant area of research for the FinTech world. It is a machine learning algorithm, which provides tailored services and enhanced personalization to customers based on profile and purchase pattern[13]. Bayesian Personalized Ranking, Alternating Least Squares, and Word2Vec are different algorithms for building implicit feedback recommender. Amazon, Netflix, Spotify, Best Buy, and YouTube use recommenders that become more efficient with each use. Recommendation engines help them improve cart value, customer engagement and create customer delight. YouTube implements video recommendations feature on its homepage to recommend personalized sets of videos. Recommendations account for close to 60% of video clicks from the home page[7].Proposals are compared with modules like most viewed, top favorites, and top-rated to avoid presentation bias. Amazon's recommendation engine uses item-to-item collaborative filtering. This technique recommends products based on purchase history, product ratings, items added to cart but abandoned, and page views. Items that are similar to purchased items are added to the recommendation list. Click-through rates, conversion, and revenue are at their peaks for Amazon because of such real-time recommendations. Its recommender generates 35% of this multi-billion-dollar titan's revenue. Banks analyze vast chunks of data to meet the needs and demands of their customers, from personal banking to investment banking. The success rate of product recommendations has increased tenfold by using Next Product to Buy (NPtB) [6] recommendation engines for small and medium-scale enterprises (SMEs) and medium-scale enterprises.

Banking institutions are very willing to introduce decision support systems. There is excellent support for peer-to-peer loan customization than conventional loan services. Although the insurance industry is small, many applications recommend both insurance policies and endorsements. Some articles deal with real estate recommendations; a good part of them is just an empirical study. Methods for forecasting stock prices and providing buy/sell signals exist, but they are not customized. Several papers feature an interactive user interface for inventory management, but only a few articles offer machine learning methods for a personalized inventory recommendation. Based on modern portfolio theory, several ways are introduced to find efficient portfolios for different levels of risk aversion; however, customization is achieved by selecting only the level of risk. Some jobs apply machine learning methods to compose custom portfolios based on individual attributes. Promising applications of recommendation systems for venture capital funds, capital funds, and the business plan questionnaire also exist.

In [1], the K-nearest neighbors (KNN) algorithm was implemented to provide real-time recommendations by identifying visitor clickstream data and matching it to a particular user group. The results of their experiment show that a real-time automated recommendation engine powered by a K-NN classifier [15] implemented with the Euclidean distance metric can produce accurate and valuable client-specific classifications and recommendations at any time. Immediate requirements are given more preference over previous interactions with the site. Some proposed a new distributed recommendation system for big data based on Apache Spark using matrix factorization (Jamali & Ester, 2010) and random forest[2]. The experiment was conducted on three real-world data sets. The results outperformed existing recommendation methods for Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and computational time. In [11] explored concept similarity and spatial mapping to build an Expert Knowledge Recommendation System (EKRS) used open data sets to verify the algorithm. The model manifested stable performance with better convergence and robustness for small-scale knowledge base datasets with different sparsity. Some discussed the use of communication networks and multivariate analysis to recommend customized items from a set of commercially available items. Performance and preference predictions were based on consumer problems and product response patterns.

Their work justified using the deep recommendation model by combining models built on the user's auxiliary information and item's textual information. A hybrid probabilistic matrix factorization algorithm was proposed. Experimental results confirmed the upper hand of their proposed approach compared to reference models, and advanced deep recommendation approaches. According, the Chameleon-based Recommender system performs better compared to the K-means-based Recommender System. This recommender uses the Chameleon Hierarchical clustering algorithm to group the users or items into a set of clusters. Ratings of items are based on the voting system. Private banking is also leveraging techniques developed by retailers like Amazon to tailor product recommendations. Collaborative filtering facilitates the selection of the following best product for private banking clients. ACF [5], a boosted collaborative filtering algorithm, predicts the unknown ratings and recommends the following best buy.

Methodology

3.1. Data

The purpose of the study is to build a recommendation system for the banking industry. Primary data was collected from personal banking customers (sample participants) across India for this purpose. Convenience sampling was employed to collect responses using a structured questionnaire. User-related and product-related variables are required to build an efficient and valuable recommendation model that was kept in mind while designing the questionnaire. The data in the questionnaire included the following parameters. Table 1 shows the user-related variables.

Table 1: User-related variables.

Sr.No.	Variable	Meaning/Context
1.	Name	Name of the customer
2.	E-mail	Email-Id of the customer
3.	Age	Age of the customer
4.	Gender	Gender of the customer
5.	Marital Status	Married/Single
6.	Location	Current city
7.	Education	Not-Graduate/Graduate/Post-Graduate
8.	Employment	Unemployed/Student/Salaried/Self-employed
9.	Yearly Earnings	Annual earnings of the customer

Table 2: Item-related variables.

Sr. No.	Variable	Meaning/Context
1.	*Savings Account*	(Whether using the product or not) & (Rating for the product)
2.	*Current Account*	(Whether using the product or not) & (Rating for the product)
3.	*Home Loan*	(Whether using the product or not) & (Rating for the product)
4.	*Education Loan*	(Whether using the product or not) & (Rating for the product)
5.	*Car Loan*	(Whether using the product or not) & (Rating for the product)
6.	*Fixed Deposit*	(Whether using the product or not) &(Rating for the product)
7.	*Recurring Deposit*	(Whether using the product or not) & (Rating for the product)
8.	*Mutual Funds*	(Whether using the product or not) & (Rating for the product)

A hybrid approach was implemented to build the recommendation system using Gradient Boosted (XGBoost) Multilabel Classifierand Collaborative Filtering technique. This approach helped remove the biases of one single technique to a great extent. Banks can use this model to recommend products to their banking clients. The model's efficiency can be significantly improved by adding more user-related and item-related features to the data, as shown in Table 2.

XGBoost for Multilabel Classification

The data were pre-processed for building the recommendation model. Train-Test splitting [14] was done to check the performance later. User-related variables were fed as independent variables to the Gradient Boosted Multilabel Classifier. The target variables were whether the customer has opted for the products or not (Yes/No). The model was built on training observations, and the performance was checked on test observations. Predictions related to opting for products can be made for old and new customers using this model. A classification report was prepared for evaluation purposes.

Collaborative Filtering (using KNN algorithm)

The ratings for different products used by customers were collected via survey. Item-to-item collaborative filtering technique was implemented on the sparse matrix (consisting of ratings). The matrix is sparse because all the customers do not use all the products. Therefore, they only provide ratings for the products they use. K-Nearest Neighbors (KNN)[12] algorithm was adopted for recommending products. Cosine similarity (Li & Han, 2013) was used as a distance metric for finding K-Nearest Neighbors. The formula for cosine similarity is given below:

$$\text{Cosine Similarity} = \cos(\theta) = \frac{A.B}{||A||||B||},$$

Where A.B = dot product of vectors A and B,

||A||, ||B|| = magnitudes of vectors A and B respectively.

In KNN [16], the distance metric for the unclassified observation is calculated from all other classified comments. The 'k's similar words for the unclassified observation are recorded. The count of each class for these observations is noted. The class with the highest frequency is the class of the new unclassified observation.If a customer uses a particular product, other products can be recommended to the customer using this technique. The class of the customer can be found using the KNN algorithm, and products that customers of his class are already using can be offered to him.

Results

Data Analysis

Most of the respondents belonged to the age groups of 24-28 and 45-55. The age distribution graph is shown in Figure 1.

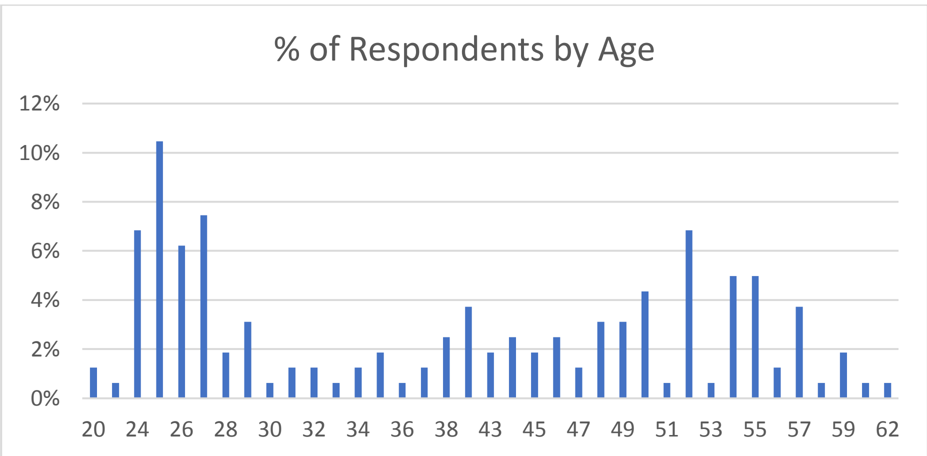

Figure 1: Age distribution graph.

Among the survey participants, around 60% were male, and the rest were female. 61% of them married. 61% were salaried employees, 14% were self-employed, 18% were students, and the rest were unemployed. Equal participation from graduates and post-graduates was observed. Yearly earnings of around 34.2% of them were greater than 10 Lakhs.

Model evaluation: XGBoost Classifier

The classification report for the Gradient Boosted Multilabel Classifier is shown in Figure 2.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	53
1	0.50	0.55	0.52	20
2	0.69	0.78	0.73	23
3	0.60	0.43	0.50	21
4	0.62	0.45	0.53	22
5	0.67	0.46	0.55	26
6	0.85	0.70	0.77	33
7	0.61	0.91	0.73	22
micro avg	0.74	0.71	0.73	220
macro avg	0.69	0.66	0.67	220
weighted avg	0.75	0.71	0.72	220
samples avg	0.74	0.69	0.66	220

Figure 2: Classification report for XGBoost Classifier.

F1-score (F-measure) [9] for the classifier was found to be 0.73. As this value is more significant than 0.5, the model can reasonably predict whether new or existing customers will opt for specific banking products or not. We have chosen f1-score over precision and recall because it is a better measure of model evaluation and a combination of both precision and recall. The area under the curve for ROC (Receiver Operating Characteristics) curve was 0.76 (76%).

User-Clusters based on Ratings

K-Means clustering and PCA (Principal Components Analysis) were utilized to find clusters with product rating data collected from respondents. Ratings for different products were converted into two principal components or eigenvectors. Clusters were formed using K-Means to get an overall idea about the likeness of specific customers to a particular cluster towards particular products, as shown in Figure 3.

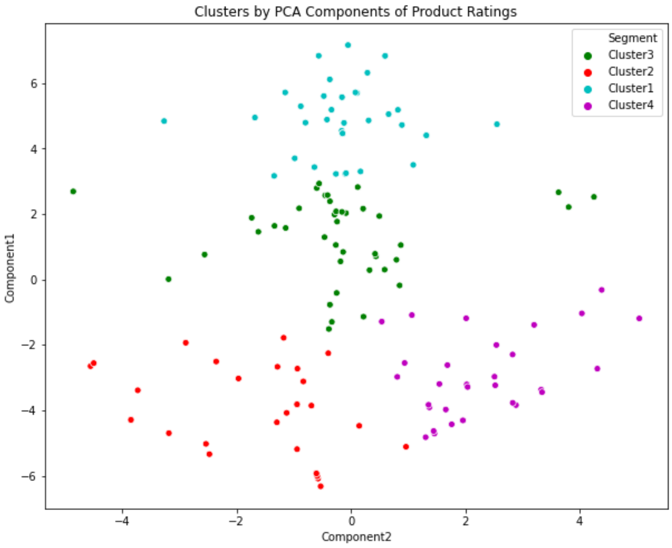

Figure 3: Clusters formed from product ratings' data of customers.

Recommendation with Collaborative Filtering

Item-to-item Collaborative Filtering was implemented to predict which other products a customer will choose depending on the effect he has decided on earlier. The nearest neighbors were chosen as 20 for applying the KNN algorithm. Three recommendations can be offered to the customer using this technique. The recommendations can be increased depending on the number of personal banking products. A sample recommendation for a customer who has a "Fixed Deposit" account in a bank is shown in Figure 4.

```
You have input item: Fixed Deposit
Found possible matches in our database: ['Fixed Deposit']

The following products can be offered to the customers...
......

Recommendations for Fixed Deposit:
1: Mutual Funds, with distance of 0.20084022853821593
2: Savings Account, with distance of 0.18595029214973646
3: Recurring Deposit, with distance of 0.16986196784836605
```

Figure 4: Sample output for recommender system.

The above output implies that a customer already has an active "Fixed Deposit" account can be provided a recommendation to opt for "Mutual Funds," "Savings Account," and "Recurring Deposit."Similar proposals can be made for a customer who owns other products. These recommendations are based on product ratings provided by customers of the same class. The customers are grouped into categories using the KNN algorithm.

Findings

The hybrid recommendation approach implemented in our research work can recommend the following:

(i) Will an existing or new personal banking client opt for a particular product of a bank or not? Multilabel XGBoostClassifier can predict the choice of each customer based on user-specific parameters.

(ii) If a personal banking customer is already opting for a product from the bank, what other products can be recommended to them? KNN-based Collaborative Filtering technique can recommend products to customers based on product ratings.

Conclusion

In this work, a hybrid recommendation model is proposed to provide recommendations to banks' banking clients. The model is built on a Python environment to address various challenges related to recommendation systems in the banking industry. The key point of this proposal is to address the user cold-start problem that arises if only a collaborative filtering technique is implemented. If a new user is added, he can be recommended products using the XGBoost classifier even if the collaborative filtering fails. Second, this approach can provide better efficiency than existing models if trained and evaluated on a vast customer dataset of banks. Third, the scalability issue can also be managed if implemented on an HDFS platform. However, our proposed work will still exist in popularity bias, sparsity of rating matrix, and item cold-start problems. These issues can be addressed in the future. This recommendation system can also be extended to other financial domains, including corporate banking, insurance, and stock markets. Ultimate personalization and customization can attract more customers and provide a competitive advantage over others.

Our hybrid recommendation approach is a proof of concept for providing recommendations to personal banking customers. It combines both the benefits of the Gradient Boosted Multilabel Classifier and the KNN-based Collaborative Filtering technique. Banks can use this approach to recommend products to their customers, boost sales and increase customer satisfaction. There are several ways to improve it in future work. Continuous research on other classification techniques and recommendation methods can take the study much further. These methods should be implemented on massive customer data to improve the computational cost and efficiency of the model. The dataset used for our research work was limited to people of metro-cities and few urban cities. For future work, data consisting of details of customers all around India should be used. The model should be implemented on a big data platform to reduce scalability issues to a large extent. Partnering and collaboration with an Indian bank are also essential to address the problems while implementing recommendation systems. More research is also needed on many other knowledge-based recommendation techniques. The results can be compared with this model. The most effective model to deal with a problem of this nature should be adopted shortly.

References

Adeniyi, D. A., Wei, Z., & Yongquan, Y. (2016). Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method. *Applied Computing and Informatics*.

Ajesh, A., Nair, J., & Jijin, P. S. (2016). A random forest approach for rating-based recommender system. *International conference on advances in computing, communications, and informatics (ICACCI).*

Al-Hassan, M., Lu, H., & Lu, J. (2015). A semantic enhanced hybrid recommendation approach: A case study of e-Government tourism service recommendation system. *Decision Support Systems 72: 97-109*.

Beyond Fintech: A Pragmatic Assessment Of Disruptive Potential In Financial Services. (2017). *World Economic Forum*.

Chirkina, A., & Rankov, B. (2018). *A Recommender System for Private Banking*. (InCube) Retrieved from https://www.incubegroup.com/blog/recommender-system-for-private-banking/.

Crespo, I., Naveira, C. F., & Kwon, H. (2018). *Using data to unlock the potential of an SME and mid-corporate franchise.* McKinsey Insights.

Davidson, J., Liebald, B., Liu, J., Nandy, P., Vleet, T. V., Gargi, U., & Gupta, S. (2010). The YouTube video recommendation system. *Fourth ACM conference on Recommender systems.*

Deloitte. (2017). *FinTech in India.*

Dembczynski, K., Jachnik, A., Kotlowski, W., Waegeman, W., & Hüllermeier, E. (2013). Optimizing the F-measure in multi-label classification: Plug-in rule approach versus structured loss minimization. *International conference on machine learning, pp. 1130-1138*.

Fintech Investments In India Double To $3.7 Billion In 2019: Accenture Analysis. (2020). (Bloomberg) Retrieved from https://www.bloombergquint.com/business/fintech-investments-in-india-double-to-37-billion-in-2019-accenture-analysis

Gao, L., Dai, K., Gao, L., & Jin, T. (2019). Expert knowledge recommendation systems based on conceptual similarity and space mapping. *Expert Systems with Applications*.

García-Laencina, P. J., Sancho-Gómez, J.-L., Figueiras-Vidal, A. R., & Verleysen, M. (2009). K nearest neighbors with mutual information for simultaneous classification and missing data imputation. *Neurocomputing 72, no. 7-9: 1483-1493*.

Gigli, A., Lillo, F., & Regoli, D. (2017). Recommender Systems for Banking and Financial Services.

Guggisberg, S. (n.d.). *How to Split a Dataframe into Train and Test Set with Python*. Retrieved from https://towardsdatascience.com/how-to-split-a-dataframe-into-train-and-test-set-with-python-eaa1630ca7b3

Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003). KNN model-based approach in classification. *OTM Confederated International Conferences.*

Murugan, S., Ganesh Babu TR, and C. Srinivasan. "Underwater Object Recognition Using KNN Classifier." *International Journal of MC Square Scientific Research* 9, no. 3 (2017): 48-52.

IMCRA - International Meetings and Conferences Research Association

www.imcra-az.org; E-mail (Submission & Contact): editor@imcra-az.org
"Science, Education and Innovations in the context of modern problems" Journal
ISSN (pe): 2790-0169 / 2790-0177

DOI- 10.56334/sei/8.3.50