Journal of the Kerala Statistical Association
Vol.31, December 2020, 68-84 *ISSN 2249-4553*

Comparison of Machine Learning Techniques for Recommender Systems for Financial Data

Divya G Nair*and K. Muralidharan†
Department of Statistics, Faculty of Science,
The Maharaja Sayajirao University of Baroda, Vadodara 390002, India

ABSTRACT

Recommender Systems are one of the most successful and widespread application of machine learning technologies in business. These are the software tools used to give suggestions to users on the basis of their requirements. Increase in number of options: be it number of online websites or number of products, it has become difficult for the customer to choose from a wide range of products. Today there is no system available for banks to provide financial advisory services to the customers and offer them relevant products as per their preference before they approach the bank. Like any other industries, financial service rarely has any like, feedback and browsing history to record ratings of services. So it becomes a challenge to build recommender systems for financial services. In this research paper, authors propose a collaborative filtering technique to recommend various products to the customer in order to increase the product per customer (PPC) ratio of bank. The advantage of these recommender systems is that it provides better suggestion to the customer based on his needs/requirements for his/her savings, expenditure and investments.

Key words and Phrases:*Financial Analysis, Recommender systems, Collaborative Filtering, Cosine Similarity, Singular value decomposition.*

*divya.nair.ap@gmail.com
†lmv_murali@yahoo.com

1 Introduction

In today's competitive world, growing customer base and satisfying them is considered the most challenging task. Traditional retail banks with physical branches and high headcounts continue to offer valuable services to consumer but with the increase in digitalisation it becomes a challenge to retain the loyal customers and attract new customers by coping with the new technologies. With digitalisation in financial sector, people expect banks to provide service at their doorstep without being called to the branches. Also with the younger generation i.e. millennial starting with their banking needs, it becomes essential for the financial sector to keep pace with modernisation.

Big data in the banking and financial services have been responsible for helping to create better customer experiences and also help protect businesses. The financial sector and banking institutions can benefit from big data by using that information to customize audience sets by demographic, behaviour, etc. and offer them personalized products (Gigli et al., 2017). Big data facilitates banks and financial institutions to be more specific about product offerings, likely increasing the chance that the right product will be offered to the right person (Amakobe, 2015).

Recommender systems (Carlos et. al 2015, Kumar et. al 2017) are beneficial both to the customer as well as the service providers. These were introduced with the aim of offering products which seems to be more suitable for the customer based on their past behaviour, purchase patterns, financial status and so on. Big Industries like Amazon, Netflix and Facebook uses recommender systems for recommending products, movies and friends to their customers based on the buying behaviour, ratings and browsing history. The existing recommender systems are generally based on ratings, likes, feedback or browsing data. Banking industry has also started embracing digitalisation and has initiated steps to attract customers. Though these techniques are quite common in retail segment like Amazon and across online movie and music industry, this has yet to come strongly into the financial industry. Recommender Systems help stop attrition of the customers by providing quick financial advisory services and help them to find the relevant products at a quick glance be it

mortgage, savings account, stocks and bonds, investments or loans. Today there are no such recommendation systems available for financial products to help the customers find their suitable products. Unlike the retail industry, for banking industry, ratings, likes, feedback are not available and so a different methodology had to be derived to find out a single value which represents the ratings of the customers.

Hence, an attempt has been made to propose a recommendation system for service institutions like bank in this paper. The proposed system makes use of the customer demography, income, credit–debit transactions, product holdings etc. to identify customer behaviour and similarity among other customers to provide a very efficient and effective product offering. This will result in improved customer service and customer satisfaction thereby increasing the conversion ratio of the leads and resulting in increased Product per customer (PPC) of the bank.

The organization of the research paper is as follows. The Section 2 provides a detailed literature survey. The proposed framework for Recommender system along with the experimental setup for financial analysis is studied in Section 3. The detailed analysis and the result are presented in Section 4. The last section presents the conclusion and the recommendations.

2 Literature Survey

A recommender system is a technology that is deployed in the environment where items (products, movies, events, articles) are to be recommended to users (customers, visitors, app users, readers) or the opposite (https:/medium.com/recombee-blog/recommender-systems-explained-d98e8221f468). Typically, large number of users and large number of products make it difficult and expensive to know/study every customer's preference and offer the right product or to identify the right customers for each product. Efficient and effective recommender systems are a solution to such situations. Recommender systems provide ratings/preference order to the unrated products based on their past ratings for other products.

Recommender systems (RS) filtering can be categorized into three main approaches. According to Su and Khoshgoftaar (2009) they are

1. Content based filtering: Recommends items by matching attributes with other items that a given user have rated. In content-based recommender systems, a recommendation is based on the relation between the attributes of items that a given user has previously rated and items which the user have not yet rated.

2. Collaborative filtering: Collaborative fultering (CF) based systems propose items based on an analysis of user feedback along with the preferences of similar users. This additional robustness makes CF the most widely used and successful RS method. Recommends items by comparing a given user with a set of users that have rated other items similarly.

3. Hybrid filtering: Recommends items by combining different type of approaches together. This technique overcomes the drawbacks of content based and collaborative filtering technique and improves prediction performance. These have increased complexity and expense for implementation. Also require additional data like unstructured data which is not easily available.

The proposed framework makes use of structured data and user-item similarity based on collaborative filtering technique. Note that, Collaborative filtering (CF) systems have two main approaches for filtering, namely memory-based and model-based. Memory-based collaborative filtering techniques also called neighbourhood-based collaborative filtering includes clustering, user-user and item-item similarity based CF. The methods are based on the correlations or similarity metrics like cosine, Jaccard (Bag et. al 2019) between users and items to produce a preference score that predicts the likelihood of a user acquiring an item in the future and provide corresponding recommendations. User and item-based algorithms are the most common types of memory-based recommendation methods. User-based methods generate recommendations according to the similarities between users, whereas item-based methods compute similarities within a space of items to find strong relationships with items that have already been rated by an active user. These techniques are simple and easy to implement. These techniques use the entire dataset to classify or identify the similarity among customers. The Model-based approach devel-

ops a model based on the existing user-item ratings thereby taking a probabilistic approach to calculate the expected value of a user for a particular item. Model based collaborative filtering approach applies statistical method and machine learning technique like Neural Network, Singular Value Decomposition, principal component analysis etc. for mining the rating matrix. In many cases, the ratings matrix is sparse as ratings for every user to item may not exist. These techniques work better with sparse matrix by dimensionality reduction and thereby improve performance. (Su and Khoshgoftaar (2009) and Vijaya Kumar et al. (2014)).

The study presented here compares the performance of two collaborative filtering approaches i.e. memory-based and model-based, using banking industry data. Sarwar et al. (2001) and many others have discussed about both memory-based and model-based techniques in different areas. In this paper, we have implemented both memory based and model based approach for banking dataset to study the performance. The performance of each approach was evaluated using offline testing and user-based testing.

3 Proposed data framework

The proposed framework is to develop recommender systems to offer retail segment products like loans/deposits/ investments to the banking customers as per their life cycle or behavioural requirements. The uses of big data in banking industry are discussed by Amakobe (2015) in the areas of fraud detection, marketing and credit risk management. In India, banks are offering different products to the customers based on their eligibility. But for customer satisfaction, offering right products to the right customers at the right time is more essential. This improves the customer relationship wherein the customer feels that the bank understands their requirements and offers the products well in advance even before the customer reaches out to the bank for his requirements.

In practice, many commercial recommender systems are used to evaluate very large product sets. The user-item matrix used for collaborative filtering will thus be extremely sparse and the performances of the predictions or recommendations

of the CF systems are challenged. The data sparsity challenge appears in several situations, specifically, the cold start problem occurs when a new user or item has just entered the system, it is difficult to find similar ones because there is not enough information (in some literature, the cold start problem is also called the new user problem or new item problem). New items cannot be recommended until some users rate it, and new users are unlikely given good recommendations because of the lack of their rating or purchase history (Su and Khoshgoftaar, 2009). In our study also, we face data sparsity issue as more than 70% of the customers hold only Savings bank account with the bank and so the scores for remaining products are not available.

Banking data is very rich and confidential in the sense that it contains all the financial details of the customer like income, loans that he has already availed (for house, education, vehicle etc.), deposits (which shows the liquidity he is holding), investments, spending pattern etc. The data becomes even richer if we make use of unstructured data like transaction mining, browsing history of customers, social media data etc. In this framework we are going to make use of only structured data i.e. the demographics, product holdings, geography, transaction, external bureau data etc. in our study.

Experimental setup: The data of specific set of customers from a bank ABC is considered for the study. Customers getting regular income/salaried in the last six months were included for the study. Data preparation includes missing value imputation and outlier detection. For instance: the birth date of a customer may be a default value or incorrect which gives some unrealistic figure as age or in some cases the birth date may be missing so in these cases, the missing value imputation is done either by using the average value or some other technique depending on the variable. For outlier detection, extreme values are discarded and coerced at 3sigma values. The Data mining process is done using SQL and SPSS Modeller. After data preparation, the biggest challenge is to prepare a user-item rating matrix which will be used as an input for the model. This input matrix has a rating for each customer – product combination.

Ratings/score are the heart of recommendation engines. There are two types of ratings viz. explicit and implicit ratings. Recommender systems that are developed for movie reviews or for e-commerce sites are based on the ratings /feedbacks/likes received from the customers known as explicit ratings. Explicit data is one-action feedback: a single click tells us that a user liked a video or rated a product positively. Explicit data has its own advantages and disadvantages. These are always more valuable to businesses as it is given by the user himself and is clear, unambiguous, and gives a definite picture of the user. But explicit data may also be shallow. Like if a user provides ratings haphazardly only because it was mandatory to provide ratings then the ratings become meaningless (Aggrawal, 2016). For example: Many social media platforms like Face book, twitter etc. have the feature to like a content displayed but the unlike option is not available. Similarly, it is observed that people generally do not provide any positive feedback for any services or applications used but always approach the page for negative feedbacks/complaints. This would result in biased opinion about the product.

One of the challenges of recommender systems in the wider commercial world is that one rarely has explicit ratings data. For example in banking sector there is no concept of providing feedback or rating to a particular product. However, there is often nontrivial information about the interactions, e.g. clicks, purchases, spending pattern etc. Such indirect "ratings" information about user-item interactions is known as implicit feedback. Modelling implicit feedback is a difficult but important problem. The main challenge here is to derive the ratings. The accuracy of the model wholly depends on the ratings and hence at most care has to be given while deriving the implicit ratings. Here we will be dealing with implicit ratings as explicit ratings are not available for banking dataset.

Deriving Score: Implicit ratings have to be derived based on the user behaviour/pattern available. These can be derived either based on some existing document/score card/parameters for a particular item or based on the significant parameters identified through feature selection method. We have performed feature selection for each product separately to identify the significant parameters contributing to each

Figure 1: Word cloud for housing loan parameters

product. Since the behaviour and requirement of each product is different from each other, the weightage of parameters and significance may also differ. For example the parameters found significant for housing loan is shown below in the form of word cloud (Figure 1). Based on the weightage and significance of parameters, a score is derived for each product–customer combination. For instance: If a Customer (C1) has availed a housing loan and his annual income is 16 lakhs whose occupation is State Government employee, EMI is 30% of his net monthly income, then the rating for customer C1 for housing loan will be 35 (= weight assigned to income bracket "¿15 lakhs") +30 (= weight assigned to state government employees) + 10 (weight assigned to EMI bracket 20-30%). So the score for pair C1–housing loan will be 75. **Data Set:** The input matrix used for experiment consists of 1.4 crore rows and 8 columns i.e. approximately 1.4 crores customers and eight products viz. home loan, auto loan, personal loan, pension loan, deposit products like recurring deposit, fixed deposit and investment products like PPF and Mutual Fund. So the input matrix with implicit rankings should ideally be like the table as mentioned below: But since in our case, only 20% of the total customers had availed any other product other than the basic savings bank account, it resulted in a sparse matrix as shown in Table 2.

4 Methodology

In real life scenario, we may be more interested in identifying the top k preferences of a customer rather than estimating the rating that he will give to a particular

Table 1: Input table

User	Product							
	Savings Account	Home Loan	Auto Loan	Personal Loan	Pension Loan	Mutual Fund	PPF	Fixed Deposit
X1	85	56	43	12	32	25	35	43
X2	72	34	23	55	21	44	55	34
—	—	—	—	—	—	—	—	—

Table 2: Sparse matrix table

User	Product							
	Savings Account	Home Loan	Auto Loan	Personal Loan	Pension Loan	Mutual Fund	PPF	Fixed Deposit
X1	32	0	0	43	0	0	0	0
X2	43	0	0	0	0	0	58	43
X3	67	0	45	0	0	0	0	23
—	—	—	—	—	—	—	—	—

product. For instance: the rating a customer may give to a home loan product based on its features like interest rates etc. may be high but it may not be his top preference. In this paper we will be discussing the method to identify the top preferences of a customer based on customer similarity which is also known as the top-k recommendation problem (Aggarwal, 2016). In this paper we will discuss about the collaborative filtering techniques for recommending the top-k products to a customer. Two types of CF techniques viz. memory-based and model-based CF methods are included in this paper.

Memory based CF further includes: k- means clustering technique to segregate the heterogeneous set of customers into homogeneous set thereby identifying the similar set of customers. This method is performed using SPSS Modeller on a system of 64 bit of 48GB RAM and 1 TB storage capacity. This clustering technique gave a silhouette score of 0.6. However, the clusters obtained by this technique were not much differentiable which could be due to sparsity in the data. Thus, clustering technique does not seem to work well with sparse dataset, which is a drawback of this method. The next technique within memory-based CF used here is user-user similarity using Python. This method uses the entire dataset to find similarity among the customers. But due to huge volume of data, the data could not be processed and hence data scalability seems to be a drawback for user-user similarity technique.

Thus major challenges in memory based recommender systems are data sparsity, scalability, diversity etc. Data sparsity leads to the cold start problem i.e. new customers with no purchase history. Also, when number of existing users and items grow tremendously, traditional CF algorithms will suffer serious scalability problems. These problems can degrade the performance of recommendation process.

To achieve better prediction performance and overcome shortcomings of memory-based CF algorithms, model-based CF approaches have been investigated. Model based CF techniques use the rating data to estimate and predicts the top-k preferences (Su and Khoshgoftaar, 2009). Model based CF algorithms include methods such as Bayesian belief nets, Markov Decision Process-based CF, Dimensionality

reduction techniques like Singular Value Decomposition (SVD) that are capable of handling problems like data sparsity and scalability.

To overcome the drawbacks of memory-based CF i.e. data sparsity and scalability, Simon Funk's Singular Value Decomposition (SVD) technique which is a model based CF technique is studied in this paper. SVD techniques are dimensionality reducing techniques and hence are known for predicting the ratings for large scale data. After its Success in Netflix Price, Simon Funk's SVD has become a common approach in dealing with huge sparse matrices. Here, the actual rating matrix (A) is decomposed into three matrices as below:

$$A = USI^T,$$

where U is the latent factors matrix of the users, S explains the relationship between the latent factors of the user and item, I is the latent factors matrix of the items. In this case by latent factors we mean characteristics of the user/item.
The resulting dot product calculates the ratings for all user-item pair by minimising the squared error. That is for each rating, error is calculated as

$$E_{ij} = R_{ij} - \hat{R}_{ij}.$$

Using the error values, new values in the User matrix and Item matrix are updated as below. A regularisation parameter is added to the equation to avoid over fitting of the generalised model.

$$U_i = 2\alpha * (A - P) * U_i; \;\; U_i = \text{ith Value in User matrix}$$

$$I_j = 2\alpha * (A - P) * I_j; \;\; I_j = \text{jth value in Item matrix},$$

where α is the learning rate. By multiple iterations, the error is minimized using the gradient descent function. Thus, the nearest approximation is arrived at and the item/service with highest rating is offered as the next best product for the customer.

Our dataset is divided geography-wise into 18 data sets. Each data set consists of approximately 7-8 lakhs data. Only 20% of the customers have availed any other product other than SB. Therefore, the ratings matrix obtained is sparse and hence to

deal with data sparsity, Stochastic Gradient Descent (SGD) is used along with SVD popularly as mentioned above to optimize the ratings and thereby minimizing the error. SVD predicts the user-item ratings and hence the error between the actual and the predicted rating value can be calculated. Using the stochastic gradient descent we try to minimize the error through multiple iterations and obtaining the local minima value of error giving the nearest predicted value and increasing the accuracy.

The quality of a recommender system can be decided based on the result of evaluation and interpretation based on business logic. According to Herlocker et al. (2004), metrics evaluating recommendation systems can be broadly classified into the following categories: predictive accuracy metrics, such as Mean Absolute Error (MAE) and its variations; classification accuracy metrics, such as precision, recall, F1-measure, and ROC sensitivity; rank accuracy metrics, such as Pearson's product-moment correlation, Kendall's Tau, Mean Average Precision (MAP), half-life utility, and normalized distance-based performance metric (NDPM). We only introduce the commonly-used CF metrics Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) here. MAE computes the average of the absolute difference between the predictions and true ratings. The MAE and RMSE are given by

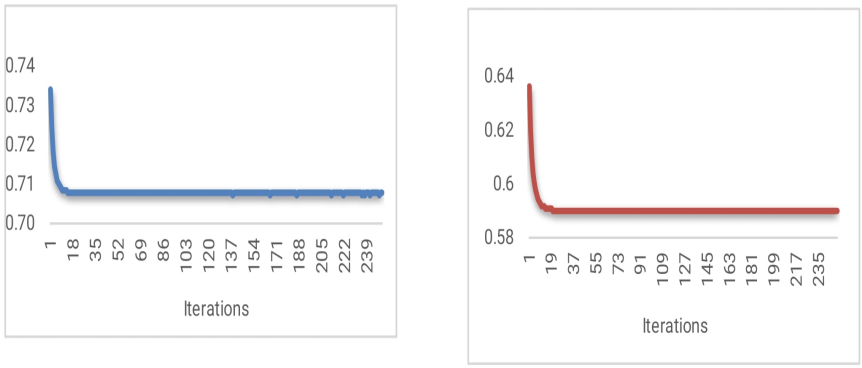
$$MAE = \frac{\sum_{ij} |p_{ij} - r_{ij}|}{n} \tag{4.1}$$

and

$$RMSE = \sqrt{\frac{1}{n} \sum_{ij} (p_{ij} - r_{ij})^2}, \tag{4.2}$$

where n is the total number of ratings over all users, p_{ij} is the predicted rating for user i on item j, and r_{ij} is the actual rating. The lower the MAE betters the prediction. The RMSE shown in (4.2) amplifies the contributions of the absolute errors between the predictions and the true values. Both MAE and RMSE do not have any upper bounds or lower bounds that justify whether the predictive power or accuracy is good or not. They are only comparable with respect to previous trail or with any other dataset.

Figure 2: Plot of RMSE (a) set-1, (b) set-2



(a) (b)

Here Funk's SVD is performed with 250 iterations and the learning rate and regularization parameters (Aggrawal, 2016) were finalised based on trial and error methods. The RMSE values become stable after some iteration as shown in Figures 2 (a) and (b) below. The results are also validated as per business logic and it is observed that customers with single product i.e. the cold start problem are also handled well by the algorithm as per their ratings and the suggestions were efficient enough. For example: A person having a tendency of saving is offered investment products like FD, Mutual Funds etc., while a person with existing liability product like home loan may be offered a personal loan. Thus, deriving the score/rating plays a vital role in recommender systems. Therefore, model based CF using Simon Funk's SVD has proved to be providing efficient results.

Table 3 presents the recommendation of the CF algorithm for some select cases. The table may be interpreted as follows: For user-0, the first preference is item number 5 (ie. Recurring deposit), the second preference is item number 7 (mutual fund), and so on. The preference table is shown in Table 4.

Table 3: Recommendation in terms of preferences

User	Item0	Item1	Item2	Item3	Item4	Item5	Item6	Item7	Item8
0	54	75	33	12	53	87	10	66	47
1	42	30	13	53	77	10	64	23	87
2	68	74	21	13	10	34	87	58	49
3	41	81	11	63	30	21	72	35	56
4	29	62	74	20	83	30	12	54	41
5	57	60	31	12	23	20	81	73	46
6	40	79	28	10	84	18	57	31	60
7	22	72	‘62	10	82	17	58	34	40
8	43	55	81	32	14	76	20	27	69
9	41	23	75	20	31	15	86	60	54
10	35	12	26	13	82	66	77	50	45

Table 4: Item preference table

Preference	
Item 0	SB
Item 1	Home Loan
Item 2	Car Loan
Item 3	Personal Loan
Item 4	Pension loan
Item 5	Recurring Deposit
Item 6	Fixed Deposit
Item 7	Mutual Fund
Item 8	PPF

5 Conclusion

The traditional method of offering products to customers without considering their requirement/preference causes irritation among the customers thereby leading to dissatisfaction. Collaborative filtering based recommender systems are the latest techniques that can be used to identify a customer's preference among certain set of products. In this paper, we have discussed memory- based and model-based approaches. In memory-based approach, k-means clustering and user-user similarity were studied. But because of data sparsity and scalability, these approaches did not work for the bank dataset. Whereas, the model-based approach helped in dealing with the data sparsity and scalability and provided efficient results. The proposed framework will help to provide the right product to the right customer. Thereby, improving the customer relationship and satisfaction with the bank. This also helps in reducing the customer attrition rate and increasing the product per customer index. The proposed system though better than the traditional system may have many shortcomings in case of data gaps like unknown salary, inadequate transaction data etc. This can be overcome with the help of Collaborative filtering techniques using big data tools. Capturing the digital footprints of the customer and implementing Hybrid collaborative filtering may also help in improving the prediction power of the models. The newer generation customers expect doorstep banking by ways of digitisation and with increasing levels of expectation they have the tendency of high attrition rate on getting good offers by the competitors. Therefore it is important to foresee the customers' preference and offer the products accordingly to reduce the attrition rate. Thus Recommender systems can be of huge benefit for the financial industry by using it wisely to attract customers.

Acknowledgements: Both the authors thank the referee and Editor for their valuable comments.

References

Aggarwal, C. C. (2016). Recommender Systems: The Textbook, DOI 10.1007/978-3- 319-29659-3 1, *Springer International Publishing Switzerland.*

Andrea Gigli, Fabrizio Lillo, and Daniele Regoli. (2017). Recommender Systems for Banking and Financial Services. In *Proceedings of RecSys 2017 Posters, Como, Italy.*

Francesco Ricci and Lior Rokach and Bracha Shapira (2011). *Introduction to Recommender Systems Handbook*, Springer, pages 1-35.

Amakobe, Moody. (2015). *The Impact of Big Data Analytics on the Banking Industry.* DOI : 10.13140/RG.2.1.1138.4163.

Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl (2001). Item-based collaborative filtering recommendation algorithms. *10th Int. Conf. on World Wide Web* pp. 285–295

Carlos, A., Gomez-Uribe, Neil Hunt. (2015). The Netflix Recommender System: Algorithms, Business Value, and Innovation, Netflix, Inc. ACM Trans. Manage. Inf. Syst. 6, 4, Article 13, http://dx.doi.org/10.1145/2843948.

Minh-Phung Thi Do, Dung Van Nguyen, Loc Nguyen (2010). Model-based Approach for Collaborative Filtering - *The 6th International Conference on Information Technology for Education.*

Aditya, P. H., Budi, I., and Munajat, Q. (2016). A comparative analysis of memory-based and model-based collaborative filtering on the implementation of recommender system for E-commerce in Indonesia: A case study, https://ieeexplore.ieee.org/document/7872755 Published in: 2016 *International Conference on Advanced Computer Science and Information Systems (ICACSIS).*

Kumar, M. K. P., Manjusha, M., and Sidharth, S. R. (2017). A Framework for Development of Recommender System for Financial Data Analysis,*I.J. Information Engineering and Electronic Business*,5, 18-27

P. N. Vijaya Kumar, Dr. V. Raghunatha Reddy (2014) - A Survey on Recommender Systems (RSS) and Its Applications, *International Journal of Innovative Research in Computer and Communication Engineering*

Safir Najafi & Ziad Salam (2016) - Evaluating Prediction Accuracy for Collaborative Filtering Algorithms in Recommender Systems,KTH Royal Institute of Technology: Stockholm, Sweden.

Su, X., and Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques, *Advances in Artificial Intelligence*.

Herlocker, J. L., Konstan, J. A., Terveen, L. G., and Riedl, J. T. (2004). "Evaluating collaborative filtering recommender systems," *ACM Transactions on Information Systems*, vol. 22, no. 1, pp. 5–53.

Bag, S., Kumar, K., Tiwari, M. K. (2019). An efficient recommendation generation using relevant Jaccard similarity, *Information Sciences* Volume 483,Pages 53-64.

Ivchenko, G.I., Honov, S.A. (1998). On the jaccard similarity test. *Journal of Mathematical Sciences* 88(6), 789–794

Zhengzheng Xian, Qiliang Li, Gai Li, and Lei Li. (2017). New Collaborative Filtering Algorithms Based on SVD++ and Differential Privacy, *Mathematical Problems in Engineering*, Vol. 2017, Article ID 1975719.

https://www.techfunnel.com/information-technology/how-the-financial-sector-will- benefit-from-big-data/

https://medium.com/recombee-blog/machine-learning-for-recommender-systems-part-1-algorithms-evaluation-and-cold-start-6f696683d0ed

https://mapr.com/blog/big-data-in-banking-many-challenges-more-opportunities/

https://medium.com/recombee-blog/recommender-systems-explained-d98e8221f468

https://nikhilwins.wordpress.com/2015/09/18/movie-recommendations-how-does-netflix-do-it-a-9-step-coding-intuitive-guide-into-collaborative-filtering/

https://towardsdatascience.com/paper-summary-matrix-factorization-techniques-for-Recommander-Systems-82d1a7ace74

https://blog.mirumee.com/the-difference-between-implicit-and-explicit-data-for-business-351f70ff3fbf