A Recommender System for Indian Credit Cards using Text Analytics

Madhukeshwar R K
madhukeshwar.ba06@reva.edu.in
REVA Academy for Corporate Excellence, REVA University, Bengaluru, India
Ratnakar Pandey
ratnakarpandey@race.reva.edu.in
REVA Academy for Corporate Excellence, REVA University, Bengaluru, India
Shinu Abhi
shinuabhi@reva.edu.in
REVA Academy for Corporate Excellence, REVA University, Bengaluru, India

Abstract

With the increase in web and social media usage, views and experiences on products or services shared by online users have increased. Social media acts as the main source of data for text analytics on which the users' sentiments can be performed. Organizations can gain valuable insights into social marketing strategies by finding sentiments and emotions towards their products. Sentiment analysis is a process of computationally finding, classifying, and categorizing opinions expressed on the block of text, to decide whether sentiment towards a particular topic or a product, is positive, negative, or neutral and whether emotions are happy, sad, angry, etc. by combining Machine Learning and Natural Language Processing. In this paper, we perform sentiment analysis through NLP techniques on the reviews and tweets collected from websites and Twitter on prominent Indian credit cards. Predicted sentiment values are used to develop a recommendation model which recommends similar credit cards based on categories.

Keywords: ***Text Mining, Sentiment Analysis, Sentiment Classification, Natural Language Processing, Recommendation engine, Credit Cards.***
DOINumber:10.14704/nq.2022.20.8.NQ44922 **NeuroQuantology 2022; 20(8): 9021-9028**

1.INTRODUCTION

Credit cardholders make credit payments to buy products or to the required services using a payment card which is commonly known as a credit card. Usually, this is associated with the revolving account that gives credit to the cardholder to make payment for his / her needs and later pay back to the card issuer within the agreed time limit. Credit cards can be classified into two groups i.e. consumer cards and business cards (O'Sullivan et al., 2003). On high-value purchases, the cardholder can convert the total amount of purchases into low-cost EMIs to enable easy repayment over a period thus revolutionizing their shopping experience.

Banks issue credit cards with a credit limit, allowing the cardholder to make payments. The card limit is based on the cardholder's income, credit score, and bank account transaction history. Repayment of spent amount using the credit card need can be done without paying interest, by making the repayment within the predefined period (*Credit Card Industry Analysis - Overview, Market Dynamics, Costs*, 2021). Learning how to use a credit card and how to make use of the credit card period and repay the amount on time helps cardholders to boost their credit score and help them to get better eligibility for high credit without any difficulty.

In the last few years, custom-designed credit cards based on the type of usage have become a big selling point. Co-branded and

affinity cards have become more popular than ever (*History of the Credit Card*, 2017). Credit cards custom designed for travel at the airport or railway station gives the cardholders a unique experience through complimentary lounge access, priority check-in, discounts at the restaurants
and more. A few of the travel cards also cover comprehensive travel insurance. Discounts on credit
1
cards get extended on movie tickets, online shopping, health, and wellness outlets, and surcharge waivers at petrol/diesel pumps across the country.

Credit card customers use social media and online review portals to post their experiences and opinions openly with the world. The availability of this huge customer data has pulled business users and researchers from different fields to better understand their customer sentiments. Sentiment analysis' primary objective is to predict the polarity of the text i.e. its positive, negative, or neutral (*Sentiment Analysis & Machine Learning Techniques - Data Analytics*, 2021).

This paper intends to design a recommendation system to help the users to select the right card for the right occasion based on the collected users reviews from websites and predict their sentiment polarity using machine learning techniques (*Sentiment Analysis - Wikipedia*).

2. LITERATURE REVIEW

This section examines the extant literature available on the need for developing a recommendation system based on sentiment analysis with natural language techniques in the credit card market.

Given the fierce competition in the global market the credit card industry is highly vulnerable to customer power in the emerging e-payment and fintech industry. Innovation in financial services to provide a personalized experience to customers with a better offer on credit card transactions become inevitable (Umuhoza et al., 2020). Mining credit card reviews can help the banks to find interesting patterns among different variables that may be used in the future to design better products (Zaza & Al Emran, 2016).

The digital transformation in the banking sector has enforced e-payment customers to purchase products online by staying at home and due to the proliferation of online stores, online reviews have also increased as the source of information on product quality and durability. Opinion mining on these online reviews helps customers to make a better decision on the purchase (Mittal & Agrawal, 2022).

Online shopping has become increasingly popular due to the variety of products, lower prices, availability of different models/brands, and fast logistic systems. The explosion of offers has forced the shoppers to make use of debit/credit cards and to avail of this cashback and discounts offered by card issuers. Credit card holders typically aggregate reward points through various offers from multiple credit cards. Cashback or rewards acquired from the credit card transactions varies in percentage from the issuer of the credit cards, identifying the best reward or cashback for a given card is difficult (Javkar et al., 2016).

Post shopping, shoppers provide their ratings, review, and emotions on websites which becomes the main source for purchaser's sentiments data generation. Multiple tools and techniques are available in the market for automatically classifying the sentiments for user-generated data. Sentiment analysis helps users to make better purchases through their collective analysis of sentiments. In deep learning models, the network learns to extract the features while the learning/training process. Word2vec modeling technique uses CNN to get trained and to classify the sentiments on reviews collected (Shah, 2021).

User text data can also be fed to a stochastic learning algorithm that analyses and classifies the feedback as negative, positive, and neutral and provides recommendations to shoppers for their next purchases (P, 2020). Lexicon, machine learning, or a hybrid combination of both are the most commonly used approach (Ahmad et al., 2020).
2
Lexicon Based algorithms can calssify the user sentiment through polarity score or using a machine learning classifier to identify specific text into a sentiment class. Two problems to be solved here are subjectivity classification; a text is subjective or objective and polarity

classification; the text is a positive or negative or neutral (*Sentiment Analysis in Banking - Maveric Systems*, n.d.). The lexicon

based approach is easy to understand and implement (Chakrabarti et al., 2018). However, user-shared reviews raise challenges due to insufficient coverage of emotions expressed. With an unsupervised approach, accuracy is determined by the classifier that might need modifications or negations (Asghar et al., 2017).

Recommendation systems typically are classified into content, collaborative, and hybrid-based recommendations. When properties of targets are considered for the recommendation it is called content based. When the system recommends the targets based on the comparison measures between other targets and users it's called collaborative filtering. A hybrid recommendation is based on a combination of content based and collaborative (Shaikh et al., 2017).

Content-based algorithms come with limitations of lack of diversified reviewer's interests, so the content fusion of reviewer behavior is suggested. It is implemented by building the correlation between the popularity of the reviewer's interest and the text and then finding the user preferences along with time utility and finally fusing the potential and user preferences to provide a recommendation list (Li & Wang, 2020).

In the collaborative filtering technique, the number of users increases the amount of work required by the system. The technique should be able to provide quality recommendations for complex problems. For complex problems, the preferred technique is item-based collaborative filtering. The item-based technique uses indirect computing recommendations for the user from the relationship identified between different targets which is an output of the user-target matrix (Sarwar et al., 2001).

Xue and Zhang propose to calculate a new distance between the short and long text's similarity as a technique to identify the nearest neighbor set from the social network of the user and recommend the texts to the user's nearest neighbor set (Xue & Zhang, 2019).

A study on common recommendation techniques reveals that 55% of approaches are content based filtering, around 18% are collaborative filtering, and 16% are graph-based recommendations. Hybrid recommendations, stereotyping and item-centric recommendations are the other techniques that are applied (Beel et al., 2014).

3. METHODOLOGY

Natural Language Processing (NLP) falls under the branch of Artificial Intelligence (AI) which is a branch of Computer Science. It mainly provides an understanding ability of computers like the way human beings can text and speak words. NLP includes rule-based modeling, computational linguistics with Machine Learning (ML), Statistics, and Deep Learning Models as shown in Fig 3.1. With a combination of these technologies, text data produced out of human languages is understood by the computers to provide meaning to the full user's sentiments and intents.

3

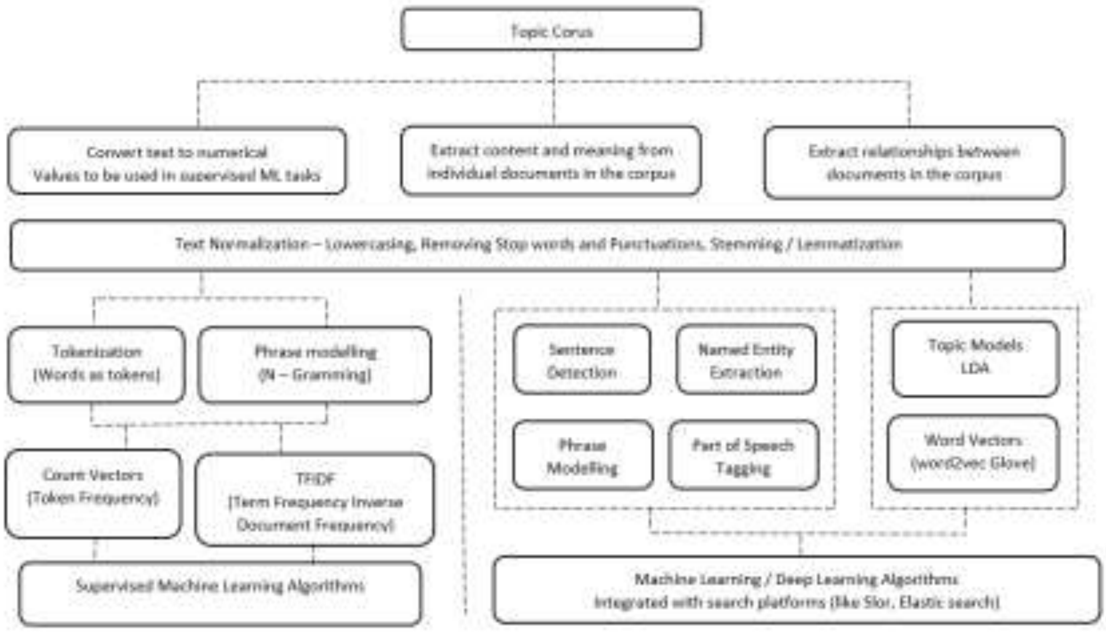


Figure 3.1 NLP Flow Diagram (*Breaking through Text Clutter with Natural Language Processing (NLP)*)

The data for this study, user reviews from the website, and tweets are collected using a trial version of the web scraping tool – Octoparse and using scraps and tweets from the Twitter API using R programming. Various text preprocessing activities were done on the text. We have used the lexicon method for scoring. Exploratory data analysis is done on the frequency of words and repeated words, followed by feature extraction. Here we transform the unstructured text data into machine-readable format and numbers using a bag of words, TF-IDF, and word embedding. We did sampling and used various classifiers and topic detection models to provide the sentiment of texts as positive, negative, and neutral on the reviews and tweets.

4. PROPOSED SYSTEM

Data collected from the websites and tweets are preprocessed as mentioned in the data pipeline diagram (Figure 4.1). Each text undergoes text normalization i.e., converting the text to lowercase, removing the stop words, punctuations, stemming, and lemmatization. EDA is performed on the texts for word frequency and word cloud followed by feature extraction. Finally, a separate set of datasets are created and stored in the datastore. This data is further processed to achieve the required dataset.

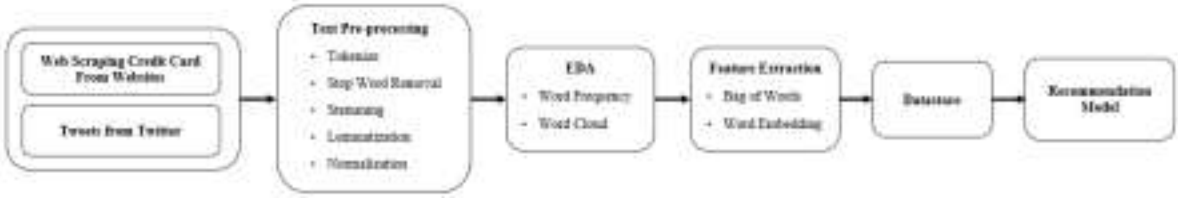


Figure 4.1 Data pipeline for data preprocessing

Dataset from the data store is further processed for building a recommendation engine pipeline (Figure 4.2). Data in the dataset is filtered which had only positive sentiments. Two features credit card and card category are combined to get a unique feature called *credit card category*. The final dataset is reduced to needed features followed by binning the polarity.

4

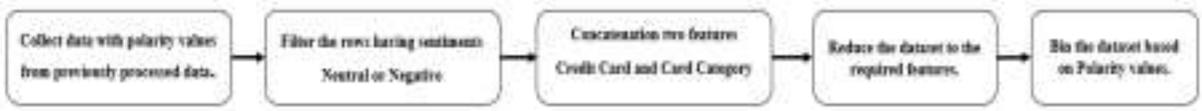


Figure 4.2 Data Preparation Steps for Recommendation Engine

5. MODELING

Preprocessed data was fed into multiple models to get the polarity value for the sentiments. The classification technique used was the Lexicon-Based approach, Text Blob Sentiment Analysis, VADER sentiment analysis, TF-IDF Vectorizer, and Fine Tuning with BERT.

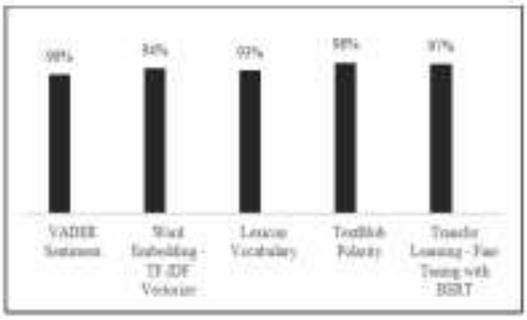
Text Blob is a python package that calculates sentiment scoring based on the polarity of the dataset from -1 to 1 (Hermansyah & Sarno, 2020). It consists of a large number of corpora sets and provides stemmers and algorithms to perform text analysis. KNN is a regression and classification machine learning technique, it looks at the labels of several data points near a targeted data point to make an educated guess regarding the data category. Even though it is straightforward, KNN is a powerful machine learning technique (Gafoor et al., 2022).

6. ANALYSIS AND RESULTS

Accuracy of Lexicon with AFFIN vocabulary, classifiers with Text Blob Polarity, VADER Sentiment, TF-IDF Vectorizer, and Transfer Learning using fine-tuned BERT Model are compared. It is important to have accurate labeling for all comments.

S.N.	Approach	Accuracy	Classifier with Best Result
1.	Text Blob Polarity	98%	Decision Tree Classifier
2.	Fine Tuning with BERT	97%	Transfer Learning
3.	Word Embedding TF-IDF Vectorizer	94%	Gradient Boosting Classifier
4.	Lexicon Vocabulary	93%	Gradient Boosting Classifier
5.	VADER Sentiment	90%	Random Forest Classifier

Table 5.1 Modeling Accuracy



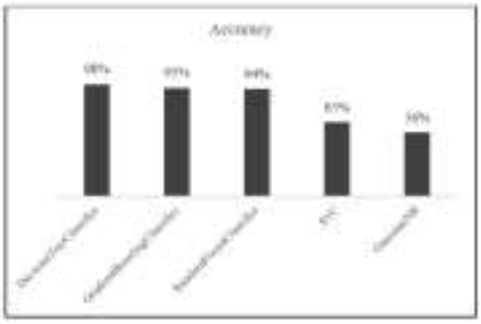


Figure 5.1 Comparison of model accuracies Figure 5.2 Comparison of accuracies Text Blob Polarity Model

Figure (5.1) shows that the classifier Text Blob Polarity classifier has the highest accuracy of 98% for the sentiment analysis. VADER Sentiment classifier with the lowest accuracy of 90%. Text Blob Polarity with Decision Tree Classifier has given the highest accuracy of 98% for the sentiment analysis.

5

There are numerous factors affecting the model performance of other approaches, especially for deep learning techniques for which we need to have more optimized data.

Decision Tree Classifier	Precision	Recall	F1-Score	Support
Negative	.98	.97	.97	828
Positive	.98	.96	.97	458
Accuracy			.98	1286

Macro Average	.98	.96	.97	1286
Weighted Average	.98	.97	.97	1286

Table 5.2 Decision Tree Classifier – Result.

Table 5.2 shows that the Decision Tree classifier for Text Blob Polarity has an accuracy of 98%, precision and recall for the positive classifier being 98% and 96% and F1-Score being 0.97.

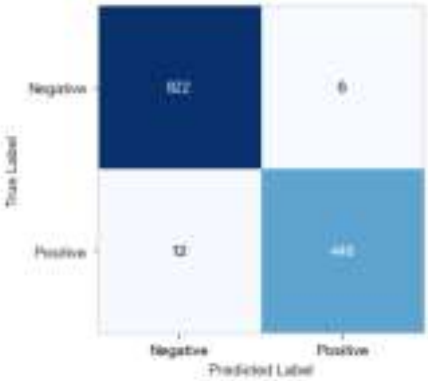


Figure 5.3 Decision Tree Confusion Matrix – Text Blob.

KNN-based collaborative filtering model relies on the similarity of item features and data distribution without making any assumptions. Distance is calculated between the target/label and every other item label/ target within the dataset. Top items are returned based on ranks calculated between the distances.

```
Recommendations for HDFC Bank Credit Card - fuel
1 : HDFC Bank Credit Card - lifestyle, with a distance of 0.86552552474476
2 : HDFC Bank Credit Card - general, with a distance of 0.8730984066413561
3 : HDFC Bank Credit Card - reward, with a distance of 0.8753430253253338
4 : ICICI Bank Credit Card - fuel, with a distance of 1.0
```

Figure 5.4 KNN Collaborative Filtering Models Recommendations.

The Decision Tree Classifier for Text Blob Polarity has 98% accuracy. There could be several factors affecting the performance of the other classifiers like deep learning and transfer learning with fine-tuned BERT models. The model performance can be improved with the following fine-tuning process.

- Having more datasets.
- Topic modeling with more data points.
- Handling class balance for the bigger dataset.
- Domain-specific support for accurate labeling by specific corpus.
- More efficient negation handling.

6

KNN Collaborative Filtering Recommendation framework is focused on the premise that it is possible to use items that are close to the item to guess how far distance from another item is shown in Figure 5.7.

```
Recommendations for State Bank of India Credit Card - general
1 : American Express Credit Card - shopping, with a distance of 0.8786329908905896
2 : American Express Credit Card - reward, with a distance of 0.9510487013222725
3 : American Express Credit Card - general, with a distance of 0.9786605149782409
4 : ICICI Bank Credit Card - general, with a distance of 1.0
```

Figure 5.7 Recommendation for general credit card

The model was able to predict other bank cards by distinct categories when one bank card's category was provided with the parameter distance.

7. CONCLUSION

The purpose of this paper is to perform text analysis and sentiment analysis on the reviews and tweets collected from websites and Twitter and to develop a tableau-based dashboard and a credit card recommendation engine for the users, to choose the right credit card for the right offer or discounts offered by the credit card issuer.

This study uses sentiment analysis using various text analytics approaches to find whether a text is negative or positive. Significant approaches were used like Lexicon with AFFIN vocabulary, classifiers with Text Blob Polarity, VADER Sentiment, TF-IDF Vectorizer, and Transfer Learning using fine tuned BERT Models. Text Blob Polarity with Decision Tree Classifier has given the highest accuracy and hence was used for creating the polarity values. Polarity values were used to feed the KNN Collaborative Filtering Model to recommend the credit cards based on the category to recommend the other credit cards category. Retrieval Credit Card was developed using Deep Learning.

8. REFERENCE

Ahmad, M., Aftab, S., Muhammad, S. S., & Ahmad, S. (2020). Machine Learning Techniques for Sentiment Analysis: A Review. *A Journal of Physical Sciences, Engineering and Technology*, *12*(02), 72–78. www.ijmse.org

Asghar, M. Z., Khan, A., Ahmad, S., Qasim, M., & Khan, I. A. (2017). Lexicon-enhanced sentiment analysis framework using rule-based classification scheme. *PLoS ONE*, *12*(2), 1–22. https://doi.org/10.1371/journal.pone.0171649

Beel, J., Gipp, B., Langer, S., & Breitinger, C. (2014). *Research Paper Recommender Systems : A Literature Survey Table of Content*. 1–68. https://www.scss.tcd.ie/joeran.beel/pubs/2016 IJDL -- - Research Paper Recommender Systems -- A Literature Survey (preprint).pdf

Breaking through text clutter with natural language processing (NLP). (n.d.). Retrieved June 27, 2022, from https://www.latentview.com/blog/breaking-text-clutter-natural-language-processing/

Chakrabarti, S., Trehan, D., & Makhija, M. (2018). Assessment of service quality using text mining – evidence from private sector banks in India. *International Journal of Bank Marketing*, *36*(4), 594–615. https://doi.org/10.1108/IJBM-04-2017-0070

Credit Card Industry Analysis - Overview, Market Dynamics, Costs. (2021, September 5). https://corporatefinanceinstitute.com/resources/knowledge/credit/credit-card-industry-analysis/

Gafoor, A., Srujana, A. L., Nagasri, A., Durgaprasad, G. S. S., & Dasari, L. S. K. (2022). KNN based Entertainment Enhancing System. *2022 6th International Conference on Trends in Electronics*
7
and Informatics, ICOEI 2022 - Proceedings, *Icoei*, 1056–1061. https://doi.org/10.1109/ICOEI53556.2022.9777225

Hermansyah, R., & Sarno, R. (2020). Sentiment Analysis about Product and Service Evaluation of PT Telekomunikasi Indonesia Tbk from Tweets Using TextBlob, Naive Bayes & K-NN Method. *International Sem Inar on Application for Technology of Information and Communication*, 511– 516.

History of the Credit Card. (2017). https://en.wikipedia.org/wiki/Credit_card

Javkar, K. G., Vora, S. H., Rodge, A. S., Bose, J., & Sharma, H. (2016). Best offer recommendation service. *2016 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2016*, 2430–2436. https://doi.org/10.1109/ICACCI.2016.7732421

Li, L., & Wang, L. (2020). News recommendation based on content fusion of user behavior. *Proceedings - 2020 13th International Symposium on Computational Intelligence and Design, ISCID 2020*, *1*, 217–220. https://doi.org/10.1109/ISCID51228.2020.00055

Mittal, D., & Agrawal, S. R. (2022). Determining banking service attributes from online reviews: text mining and sentiment analysis. *International Journal of Bank Marketing*, *40*(3), 558–577. https://doi.org/10.1108/IJBM-08-2021-0380

O'Sullivan, A., Sheffrin, S., & Perez, S. (2003). *Microeconomics: Principles, Applications and Tools, Student Value Edition (9th Edition)*. 29.

P, R. K. M. E. A. (2020). *Sentiment analysis in E-Commerce using Recommendation System*. *8*(12), 114–119.

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. *Proceedings of the 10th International Conference on World Wide Web, WWW 2001*, 285–295. https://doi.org/10.1145/371920.372071

Sentiment analysis - Wikipedia. (n.d.).

Retrieved July 24, 2022, from https://en.wikipedia.org/wiki/Sentiment_analysis#cite_note-SentiStrength2010-17

Sentiment Analysis & Machine Learning Techniques - Data Analytics. (2021). https://vitalflux.com/sentiment-analysis-machine-learning-techniques/

Sentiment Analysis in banking - Maveric Systems. (n.d.). Retrieved August 10, 2022, from https://maveric-systems.com/blog/sentiment-analysis-in-banking/

Shah, A. (2021). Sentiment analysis of product reviews using supervised learning. *Reliability: Theory and Applications*, *16*, 243–253. https://doi.org/10.1145/3447568.3448513

Shaikh, S., Rathi, S., & Janrao, P. (2017). Graph Based Approached. *2017 IEEE 7th International Advance Computing Conference*, 932–935. https://doi.org/10.1109/IACC.2017.180

Umuhoza, E., Ntirushwamaboko, D., Awuah, J., & Birir, B. (2020). Using unsupervised machine learning techniques for behavioral-based credit card users segmentation in Africa. *SAIEE Africa Research Journal*, *111*(3), 95–101. https://doi.org/10.23919/saiee.2020.9142602

Xue, H., & Zhang, D. (2019). *Based on Content and Social Network*. *Itaic*, 477–481.

Zaza, S., & Al-Emran, M. (2016). Mining and exploration of credit cards data in UAE. *Proceedings - 2015 5th International Conference on e-Learning, ECONF 2015*, 275–279. https://doi.org/10.1109/ECONF.2015.57
8