# Sentiment Classification of Indian Banks' Customer Complaints


Gutha Jaya Krishna
Centre of Excellence in Analytics
Institute for Development and Research in Banking Technology
Hyderabad-500057, India
krishna.gutha@gmail.com

Vadlamani Ravi
Centre of Excellence in Analytics
Institute for Development and Research in Banking Technology
Hyderabad-500057, India
vravi@idrbt.ac.in

Bheemidi Vikram Reddy
Centre of Excellence in Analytics
Institute for Development and Research in Banking Technology
Hyderabad-500057, India
vreddy27193@gmail.com

Mohammad Zaheeruddin
Centre of Excellence in Analytics
Institute for Development and Research in Banking Technology
Hyderabad-500057, India
mohdzaheer544@gmail.com

Harshal Jaiswal
Centre of Excellence in Analytics
Institute for Development and Research in Banking Technology
Hyderabad-500057, India
harshaljaiswal1234@gmail.com

P. Sai Ravi Teja
Centre of Excellence in Analytics
Institute for Development and Research in Banking Technology
Hyderabad-500057, India
pulipaka.raviteja@gmail.com

Rohit Gavval
Centre of Excellence in Analytics
Institute for Development and Research in Banking Technology
Hyderabad-500057, India
rohitg6790@gmail.com



***Abstract*— Sentiment Analysis of the customer complaints of Indian banks is an interesting task and critical for appropriate business functionality and improvement. Applying machine learning (ML) techniques on these raw textual data is increasingly gaining traction. Towards pre-processing the raw textual data, we employed techniques like document term matrix (DTM) driven by Term Frequency - Inverse Document Frequency (TF-IDF), embedding model likeWord2Vec and psycho-linguistic method like Linguistic Inquiry and Word Count (LIWC). For the purpose of classification, the raw textual data of complaints is labeled as "moderate" or "extreme" by the three human annotators. Results indicate that the LIWC in combination with Random Forest and Naïve Bayes techniques performed the best in three banks datasets. The results were statistically corroborated with a t-test.**



***Keywords—Banks, Classification, Complaints, Machine Learning, Sentiment Analysis.***


## I. Introduction

In the current decade of automation, organizations make future insights and make changes to their business models based on the analytical techniques used for customer relationship management (CRM) to serve the customers better. To fulfil their business needs organizations are focusing on the analytical part of the CRM and to keep up with the aggressive pace of other organizations in the market. CRM is essential for dealing with customers in different forms of communications such as accepting feedbacks through telephonic calls, e-mails, SMS, online feedback and complaints portals. Analytics plays an essential job in analyzing these feedback channels. Therefore, sentiment analysis the current topic of study in this paper is one of the critical methods to analyze the unstructured raw feedback data of the customers in these channels.

Sentiment analysis is valuable in different applications such as formulating market strategies, stock value prediction, forex rate prediction. Sentiment Analysis is most useful in the examination of individuals feelings. For instance, in the 2016 USA election campaign, parties started using sentiment analysis on social media information to get the general voter sentiment on the policies made by governments. This kind of analysis is called Swing Analytics.

In this paper, we concentrated on classifying the sentiment of the customers of banks from the online complaints and feedback portals of respective banks. Therefore, we engaged our study on four banks complaints, i.e. on SBI, ICICI, AXIS, HDFC and gathered the customer complaints information from the "complaintsboard.com". We gathered 674 grievances of SBI bank, 440 grievances of ICICI bank, 513 grievances of AXIS bank, 637 grievances of HDFC bank. We then annotated the complaints in moderate or severe with the help of three professional human annotators and pre-processed the complaints using DTM, Word2Vec and LIWC methods for the extraction of the features from the raw textual data. The features extracted from the above pre-processing methods is given to various ML techniques such as Support Vector Machines (SVM), Naive Bayes (NB), Logistic Regression (LR), Decision Tree (DT), K-Nearest Neighbor (KNN), Random Forest (RF), XGBoost and Multi-layer Perceptron (MLP).

The paper is organized in the following manner with the literature review in section II. Sections III present the current study. In section IV, the proposed methodology is described. Section V describes the experimental setup and the description of the datasets. Results and discussion are presented in section VI, and finally, we conclude with some future directions in section VII.

## II. Literature Review

Various machine learning techniques which are applied for sentiment classification are reviewed here. Pang et al. [1] in their the seminal work applied machine learning techniques viz. NB, Maximum Entropy (ME), and SVM for binary sentiment classification of movie reviews. They explored different feature engineering methods, where SVM yielded the most remarkable accuracy of 82.9% with the unigram features. McDonald et al. [2] developed models for characterizing the sentiment at sentence and document level and employed the Viterbi algorithm for the inference modelling. Dang et al. [3]

performed sentiment classification using SVM by utilizing variety feature selection techniques. Saleh et al. [4] performed sentiment classification utilizing SVM with different feature selection strategies. Bai [5] developed a Tabu Search-Markov-Blanket Classifier (TS-MBC). Zhang et al. [6] performed sentiment classification by utilizing NB and SVM on restaurant reviews in Cantonese. Tan et al. [7] proposed an automatic methodology to recognize sentiment at the phrase level. Wang et al. [8] compared the three ensemble techniques viz. bagging, boosting, and random subspace methods on five base learners NB, ME, DT, kNN, and SVM for classification of sentiment. Moraes et al. [9] applied SVM and NB with Artificial Neural Network (ANN) -based methodology for the assessment of sentiments. Basari et al. [10] developed a hybrid strategy for sentiment classification with Particle Swarm Optimization (PSO) and SVM on movie reviews data. Ghiassi et al. [11] used a supervised feature elimination method by utilizing n-grams and a statistical methodology to make a Twitter-based dictionary for sentiment analysis. Dynamic Artificial Neural Network (DAN2) and SVM are utilized as multi-class classifiers in [11]. Ravi et al. in [12], collected the above mentioned four banks datasets and analyzed using fuzzy formal concept analysis for the first time.

## III. CONTRIBUTION

Many articles in sentiment analysis of customer reviews involve preprocessing methods like DTM are Doc2Vec. However, these methods fail to capture the psycho-linguistic features from the customer reviews which we believe to yield very crucial information about the reviews. In [13], psycho-linguistic features were used to detect irony in the reviews. Therefore, in this study we employed LIWC to extract psycho-linguistic features for sentiment classification.

## IV. PROPOSED METHODOLOGY

The proposed methodology is as follows:

1. Gather the bank customer complaints data,
2. Preprocess the raw textual data include tokenization of the textual data, removal of stop words, punctuations and numerals and stemming.
3. Convert the preprocessed data in DTM and Word2Vec formats or converting the raw textual data to LIWC features and normalizing the LIWC features.
4. Split the data generated by DTM, Word2VEC and LIWC into a stratified 80:20 percent split. Keeping the 80 percent as train data and remaining 20 percent as test data. Performing 10-fold cross validation (10-FCV) on 80 percent train data.
5. Apply the machine leaning techniques mentioned in section IV on train data generated of DTM, Word2Vec and LIWC and testing the model on the remaining 20 percent test data.
6. Perform pairwise statistical t-test on the best performing models in each of the three feature extraction techniques.

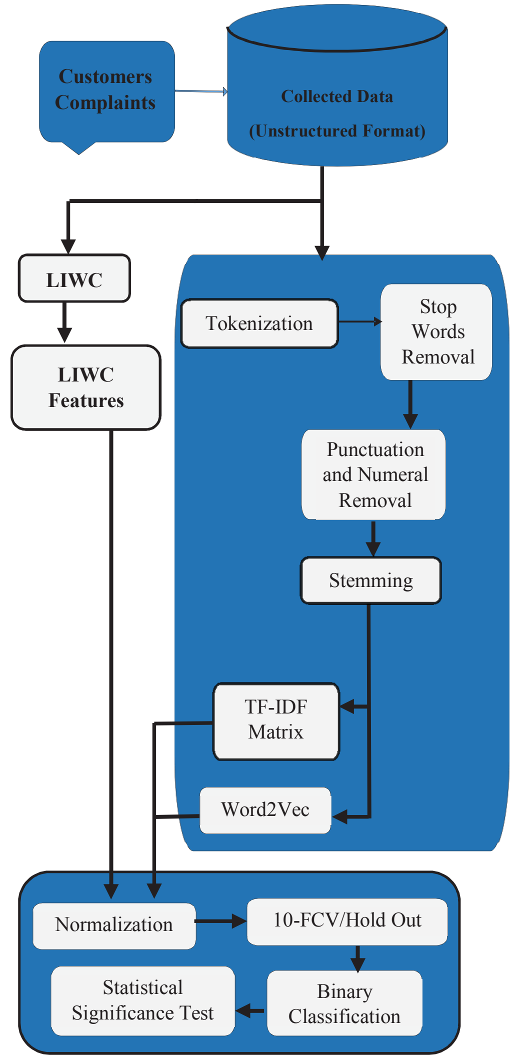


Fig. 1. Schematic Diagram

### *A. Feature Extraction Methods*

Most of the machine learning models cannot process raw textual information; instead, they can be applied only if the data is numeric. DTM, Word2Vec, LIWC are few of the many approaches by which we can change the raw textual data to a numeric format.

*Document Term Matrix*

TF-IDF is one of the methods of building a document term matrix. Term Frequency (TF) is the word frequency in each review in the corpus. The proportion of the number of times the word present in the review of all words in the review. Inverse Document Frequency (IDF) is the find the weight of uncommon words in all the reviews in the corpus. The words that infrequently appeared in the corpus have a high IDF score. TF-IDF [14] gives each word in a review, a weight dependent on its term frequency and inverse term frequency. The terms with high weight scores can be considered as essential words or features.

*Word2Vec*

In a bag of words, each of the unique words represented by discrete identifier prompts for information sparsity and implies that we may require more information to prepare machine learning models effectively. Utilizing vector representations can in a sense overcome some of the difficulties as mentioned above.

In Word2vec [15], the words are represented in a vector format Words which are comparatively similar are grouped, and different words are placed far away. By this method, a kind semantic relationship is found among the words. The Word2Vec can also be applied to a review or a document, by considering the vectors of the individual words in the review or document and averaging them.

*Linguistic Inquiry and Word Count (LIWC)*

LIWC [16] produces the summary of the raw textual language, i.e. it produces grammatical and linguistic features along with psychological features. LIWC utilizes LIWC lexicon of 6400 words at the back end, which gets as input corpus of raw text and creates word counts of pre-indicated psychologically significant features as the output. It is seen that a word is bucketed into numerous features. LIWC gives the psychometrics of the word usage, content, and style of words and so on. The content words include nouns, verbs, adjectives, and modifiers. Style or function words include articles, pronouns, conjunctions, auxiliary verbs, prepositional words, and so on. From a psychological perspective, function words pass on the style of correspondence, while content words demonstrate the communicated truth. Consequently, we can say that style words show the proportions of individuals' social and mental traits. In general, LIWC gives three classifications of raw textual data, i.e. the summary of the raw text as well as the linguistic and mental classification of the raw text. In [13], LIWC features are extracted for irony detection from the customer reviews.

### B. Machine Learning Techniques

*Support Vector Machines*

Support vector machines (SVM) [17] locates the maximum separating hyperplanes for the target classes. Therefore, SVM utilizes a form of constrained optimization. Solving the primal equations of SVM are hard, so the problem is changed over to dual with linear constraints. For non-linear separation problems kernel trick is employed, where linear, polynomial, sigmoid and radial basis kernels are utilized to explain the non-linear separation of classes.

*Naive Bayes*

Naive Bayes (NB) [18] computes the conditional a-posterior probabilities of a categorical class variable given independent predictor variables using the Bayes rule. The algorithm makes a dominant assumption about the data having features independent of each other. Naive Bayes has the advantage of being built on a robust probabilistic foundation and is very robust.

*Logistic Regression*

Logistic Regression (LR) [19] is typically applied to a two-class problem. It estimates the parameters 'β' for the specific inputs using regression. More formally, this model is one where the log-odds of the probability of occurrence of an event is a linear combination of independent input variables. The two classes of the output variable are changed over as 0 and 1, which for instance are two target classes.

*Decision Tree*

Decision tree [20] models which take a discrete set of class names are called classification trees; in these tree structures, leaves are the class names and branches are the conjunctions of input variables that lead to those class names. They utilize a variety of impurity measures for branching such as entropy, Gini index.

*K-Nearest Neighbor*

In k-Nearest Neighbor (kNN) [21] classification, the samples are grouped by a majority vote of its neighbours, with the sample being given the class of the most common among its k closest neighbours (k is a positive whole number, usually small). On the off chance that $k = 1$, the sample is appointed to the class of its single closest neighbour.

*Random Forest*

Random Forest [22] is a machine learning algorithm which can perform both regression and classification tasks. It creates a strong ensemble learner by adding several week learners. It builds multiple classification trees and combines the predictions outcome from all trees. It can also be used for multi-class problems

*XGBoost*

XGBoost is a boosting tree ensemble method. XGBoost [23] is an extension of gradient boosting trees. The algorithmic implementation of XGBoost has an addition of regularisation function to the loss function which enables the algorithm to avoid overfitting. In XGBoost [23], besides the regularised objective, shrinkage and column (feature) subsampling are two additional techniques used to reduce overfitting further. Also, an approximate algorithm for splitting allows XGBoost to handle big data.

*Multi-layer Perceptron*

Multilayer Perceptron (MLP) [24] is employed for the two-class classification problem in the current study with an input layer, hidden layers and an output layer. The MLP employs backpropagation learning.

## V. Experimental Setup and Description of Datasets

Training and testing were done on an Intel (R) Core (TM) i7-6700 processor with 32GB RAM. The two-class models for the sentiment classification problem are built in Python. Our emphasis is not on improving the training time. Until the system can run the code, system configuration has no role in the study.

### *A. Dataset Description*

The customer complaints information of SBI, ICICI, AXIS and HDFC banks are gathered from the "complaintsboard.com". We gathered 674 grievances of SBI bank, 440 grievances of ICICI bank, 513 grievances of AXIS bank, 637 grievances of HDFC bank. We then annotated the complaints as moderate or severe with the help of three professional human annotators and then pre-processed the complaints using DTM, Word2Vec and LIWC methods for the extraction of the features from the raw textual data. The maximum features by DTM for each bank in the corpus crossed 3000. Therefore, for DTM, we took the top 2000 features in all the four banks. For Word2Vec, we considered the maximum vector dimension as 50. For LIWC, 77 features were considered out of 93 features. From the 93 features generated by LIWC, word count, three general descriptor categories and twelve punctuation categories are excluded.

## VI. Results and Discussion

Area under ROC curve (AUC) is employed as the measure to judge the performance of the models. We observe that LIWC with RF and NB as the machine learning techniques yielded the best results in three banks datasets and this is corroborated statistically by a pairwise t-test. Results of the study are presented in Tables I to XII

TABLE I. Results of Axis Bank with DTM

| | AUC | AUC (SD) |
|---|---|---|
| **LR** | 0.8176 | 0.0174 |
| **KNN** | 0.7288 | 0.0214 |
| **Decision Tree** | 0.6753 | 0.0295 |
| **SVM (Linear)** | 0.7749 | 0.0243 |
| **SVM (Sigmoid)** | 0.7967 | 0.0128 |
| **SVM (RDF)** | 0.77 | 0.0144 |
| **SVM (Poly.)** | 0.7956 | 0.0171 |
| **Naive Bayes** | 0.6792 | 0.0275 |
| **Random Forest** | **0.8268** | 0.0159 |
| **XGBoost** | 0.8008 | 0.0129 |
| **MLP** | 0.8055 | 0.0161 |

TABLE II. Results of Axis Bank with Word2Vec

| | AUC | AUC (SD) |
|---|---|---|
| **LR** | 0.729 | 0.0209 |
| **KNN** | 0.6472 | 0.0198 |
| **Decision Tree** | 0.6033 | 0.0206 |
| **SVM (Linear)** | 0.7198 | 0.0242 |
| **SVM (Sigmoid)** | 0.7344 | 0.0172 |
| **SVM (RDF)** | 0.5609 | 0.0083 |
| **SVM (Poly.)** | 0.7386 | 0.0205 |
| **Naive Bayes** | 0.549 | 0.0058 |
| **Random Forest** | 0.629 | 0.0422 |
| **XGBoost** | 0.8008 | 0.0129 |
| **MLP** | **0.8055** | 0.0161 |

TABLE III. Results of Axis Bank with LIWC

| | AUC | AUC (SD) |
|---|---|---|
| **Logistic** | 0.753 | 0.0119 |
| **KNN** | 0.6927 | 0.0121 |
| **Decision Tree** | 0.6434 | 0.039 |
| **SVM (Linear)** | 0.7309 | 0.0294 |
| **SVM (Sigmoid)** | 0.7555 | 0.0144 |
| **SVM (RDF)** | 0.7498 | 0.0163 |
| **SVM (Poly.)** | 0.7583 | 0.0163 |
| **Naive Bayes** | 0.566 | 0.0252 |
| **Random Forest** | **0.7926** | 0.0171 |
| **XGBoost** | 0.7633 | 0.0239 |
| **MLP** | 0.7471 | 0.0175 |

TABLE IV. Results of SBI with DTM

| | AUC | AUC (SD) |
|---|---|---|
| **LR** | 0.7567 | 0.0111 |
| **KNN** | 0.7312 | 0.014 |
| **Decision Tree** | 0.6797 | 0.0335 |
| **SVM (Linear)** | 0.7649 | 0.0128 |
| **SVM (Sigmoid)** | 0.769 | 0.0149 |
| **SVM (RDF)** | 0.735 | 0.0156 |
| **SVM (Poly.)** | 0.7649 | 0.0128 |
| **Naive Bayes** | 0.6869 | 0.0095 |
| **Random Forest** | **0.7931** | 0.0099 |
| **XGBoost** | 0.7819 | 0.0215 |
| **MLP** | 0.7653 | 0.0185 |

TABLE V. Results of SBI with Word2Vec

| | AUC | AUC (SD) |
|---|---|---|
| **LR** | 0.7189 | 0.0071 |
| **KNN** | 0.7186 | 0.0188 |
| **Decision Tree** | 0.604 | 0.0367 |
| **SVM (Linear)** | 0.7294 | 0.0198 |
| **SVM (Sigmoid)** | 0.6774 | 0.0116 |
| **SVM (RDF)** | 0.6859 | 0.0126 |
| **SVM (Poly.)** | 0.708 | 0.018 |
| **Naive Bayes** | 0.5856 | 0.0101 |
| **Random Forest** | 0.6425 | 0.0126 |
| **XGBoost** | 0.6642 | 0.0341 |
| **MLP** | **0.73** | 0.0213 |

TABLE VI. Results of SBI with LIWC

| | AUC | AUC (SD) |
|---|---|---|
| **Logistic** | 0.7156 | 0.0089 |
| **KNN** | 0.7073 | 0.0141 |
| **Decision Tree** | 0.6773 | 0.037 |
| **SVM (Linear)** | 0.7238 | 0.0224 |
| **SVM (Sigmoid)** | 0.7164 | 0.0125 |
| **SVM (RDF)** | 0.7267 | 0.0218 |
| **SVM (Poly.)** | 0.7366 | 0.0192 |
| **Naive Bayes** | 0.686 | 0.0407 |
| **Random Forest** | **0.7794** | 0.0138 |
| **XGBoost** | 0.7697 | 0.0173 |
| **MLP** | 0.7359 | 0.015 |

TABLE VII. RESULTS OF ICICI BANK WITH DTM

| | AUC | AUC (SD) |
|---|---|---|
| **LR** | **0.7713** | 0.0149 |
| **KNN** | 0.7221 | 0.0168 |
| **Decision Tree** | 0.6967 | 0.0556 |
| **SVM (Linear)** | 0.7554 | 0.0178 |
| **SVM (Sigmoid)** | 0.7577 | 0.0182 |
| **SVM (RDF)** | 0.724 | 0.0163 |
| **SVM (Poly.)** | 0.7554 | 0.0178 |
| **Naive Bayes** | 0.6395 | 0.0162 |
| **Random Forest** | 0.7534 | 0.0198 |
| **XGBoost** | 0.7474 | 0.016 |
| **MLP** | 0.7269 | 0.0272 |

TABLE VIII. ICICI BANK WORD2VEC

| | AUC | AUC (SD) |
|---|---|---|
| **LR** | 0.6759 | 0.0144 |
| **KNN** | 0.6854 | 0.0245 |
| **Decision Tree** | 0.6061 | 0.0531 |
| **SVM (Linear)** | 0.7263 | 0.0265 |
| **SVM (Sigmoid)** | 0.6592 | 0.0103 |
| **SVM (RDF)** | 0.7244 | 0.0253 |
| **SVM (Poly.)** | **0.7477** | 0.0272 |
| **Naive Bayes** | 0.6339 | 0.0115 |
| **Random Forest** | 0.6584 | 0.0323 |
| **XGBoost** | 0.6579 | 0.032 |
| **MLP** | 0.7293 | 0.0195 |

TABLE IX. RESULTS OF ICICI BANK WITH LIWC

| | AUC | AUC (SD) |
|---|---|---|
| **Logistic** | 0.6676 | 0.0199 |
| **KNN** | 0.5872 | 0.0175 |
| **Decision Tree** | 0.7019 | 0.0364 |
| **SVM (Linear)** | 0.7083 | 0.0157 |
| **SVM (Sigmoid)** | 0.6713 | 0.0254 |
| **SVM (RDF)** | 0.7071 | 0.0155 |
| **SVM (Poly.)** | 0.683 | 0.0207 |
| **Naive Bayes** | 0.5397 | 0.0116 |
| **Random Forest** | **0.7744** | 0.0136 |
| **XGBoost** | 0.7277 | 0.024 |
| **MLP** | 0.6472 | 0.0122 |

TABLE X. RESULTS OF HDBF BANK WITH DTM

| | AUC | AUC (SD) |
|---|---|---|
| **LR** | 0.6159 | 0.0119 |
| **KNN** | **0.6776** | 0.0149 |
| **Decision Tree** | 0.613 | 0.043 |
| **SVM (Linear)** | 0.6434 | 0.0112 |
| **SVM (Sigmoid)** | 0.6419 | 0.0096 |
| **SVM (RDF)** | 0.63 | 0.0185 |
| **SVM (Poly.)** | 0.6434 | 0.0112 |
| **Naive Bayes** | 0.6432 | 0.0236 |
| **Random Forest** | 0.6268 | 0.0166 |
| **XGBoost** | 0.677 | 0.0248 |
| **MLP** | 0.6079 | 0.0136 |

TABLE XI. RESULTS OF HDFC BANK WITH WORD2VEC

| | AUC | AUC (SD) |
|---|---|---|
| **LR** | 0.5956 | 0.0157 |
| **KNN** | **0.6884** | 0.0275 |
| **Decision Tree** | 0.5985 | 0.0484 |
| **SVM (Linear)** | 0.5864 | 0.0138 |
| **SVM (Sigmoid)** | 0.5 | 0 |
| **SVM (RDF)** | 0.5 | 0 |
| **SVM (Poly.)** | 0.5959 | 0.0163 |
| **Naive Bayes** | 0.583 | 0.0113 |
| **Random Forest** | 0.6692 | 0.0285 |
| **XGBoost** | 0.641 | 0.0345 |
| **MLP** | 0.6751 | 0.0212 |

TABLE XII. RESULTS OF HDFC BANK WITH LIWC

| | AUC | AUC (SD) |
|---|---|---|
| **Logistic** | 0.6644 | 0.0202 |
| **KNN** | 0.6694 | 0.018 |
| **Decision Tree** | 0.5931 | 0.0482 |
| **SVM (Linear)** | 0.6613 | 0.0213 |
| **SVM (Sigmoid)** | 0.6512 | 0.0255 |
| **SVM (RDF)** | 0.6613 | 0.0213 |
| **SVM (Poly.)** | 0.6897 | 0.0123 |
| **Naive Bayes** | **0.7451** | 0.0267 |
| **Random Forest** | 0.6665 | 0.0264 |
| **XGBoost** | 0.7114 | 0.0151 |
| **MLP** | 0.6002 | 0.0203 |

TABLE XIII. STATISTICAL SIGNIFICANCE TEST RESULTS

| Bank | Compared Methods | t-value | p-value | Sig. (T/F) |
|---|---|---|---|---|
| **Axis Bank** | DTM (RF) vs Word2Vec (MLP) | 8.062 | 2.20E-07 | T |
| | DTM (RF) vs LIWC (RF) | 4.631 | 0.000207 | T |
| **SBI** | DTM (RF) vs Word2Vec (MLP) | 8.495 | 1.00E-07 | T |
| | DTM (RF) vs LIWC (RF) | 2.550 | 0.020064 | F |
| **ICICI** | LIWC (RF) vs Word2Vec (SVM-Poly) | 2.776 | 0.01244 | F |
| | LIWC (RF) vs DTM (LR) | 0.485 | 0.63288 | F |
| **HDFC** | LIWC (NB) vs Word2Vec (kNN) | 4.677 | 0.00018 | T |
| | LIWC (NB) vs DTM (kNN) | 6.981 | 1.61E-06 | T |

The pairwise t-test is performed on the average AUC over 10 folds at 1% significance level and 18 degrees of freedom. We compared the best performing models in each of the three feature extraction techniques. We observe from Table XIII that LIWC is statistically significant in the case of HDFC bank and statistically the same in the cases of DTM of SBI and DTM as well as Word2Vec of ICICI banks. Therefore, we prefer LIWC with RF and NB over most of the other models that we built as it is has less features and as it also captures psycho-linguistic features of the review.

If the p-value is under 0.01, then the null hypothesis is dismissed thereby stating that the pair of models are different statistically. Else, if the p-value is more than 0.01, at this point the null hypothesis is not rejected, and the pair of models compared are the same statistically. t-Test results are presented in table XIII, with a significance value of 'T' as true, when the p-value is below 0.01 and significance value of 'F' as false, when the p-value is higher than 0.01. It is observed that overall LIWC played a stellar role in picking up important features about the reviews that could not be obtained by the other two methods.

It is noteworthy that despite capturing the psycho-linguistic features, LIWC could only yield an AUC of 0.7926. One idea could be to consider the union of feature sets obtained by DTM, Word2Vec and LIWC before invoking any classifier. This strategy could potentially improve the results because of the fact that syntactic, semantic and psycho-linguistic features are considered together, thereby enabling us to study complaints comprehensively.

## VII. Conclusion and Future Directions

In this study, we applied machine learning classifiers for the sentiment classification on the reviews of customers from SBI, ICICI, AXIS and HDFC banks information gathered from the online complaints and feedback portals of respective banks. We employed three feature extraction methods and concluded that LIWC with RF and NB is statistically significant with a pairwise t-test on the results of AUC obtained from various classifiers. In future, we can employ feature selection methods on the results obtained from the feature extraction methods and use deep learning methods as classifiers. Also, we include more banks in the study and get feedbacks from the respective banks to improve our study.

## References


[1] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 conference on Empirical methods in natural language processing* , 2002, vol. 10, pp. 79–86.

[2] R. McDonald, K. Hannan, T. Neylon, M. Wells, and J. Reynar, "Structured Models for Fine-to-Coarse Sentiment Analysis," in *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, 2007, pp. 432–439.

[3] Y. Dang, Y. Zhang, and H. Chen, "A Lexicon-Enhanced Method for Sentiment Classification: An Experiment on Online Product Reviews," *IEEE Intell. Syst.*, vol. 25, no. 4, pp. 46–53, Jul. 2010.

[4] M. Rushdi Saleh, M. T. Martín-Valdivia, A. Montejo-Ráez, and L. A. Ureña-López, "Experiments with SVM to classify opinions in different domains," *Expert Syst. Appl.*, vol. 38, no. 12, pp. 14799–14804, Nov. 2011.

[5] X. Bai, "Predicting consumer sentiments from online text," *Decis. Support Syst.*, vol. 50, no. 4, pp. 732–742, Mar. 2011.

[6] Z. Zhang, Q. Ye, Z. Zhang, and Y. Li, "Sentiment classification of Internet restaurant reviews written in Cantonese," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 7674–7682, Jun. 2011.

[7] L. K.-W. Tan, J.-C. Na, Y.-L. Theng, and K. Chang, "Phrase-Level Sentiment Polarity Classification Using Rule-Based Typed Dependencies and Additional Complex Phrases Consideration," *J. Comput. Sci. Technol.*, vol. 27, no. 3, pp. 650–666, Jan. 2012.

[8] G. Wang, J. Sun, J. Ma, K. Xu, and J. Gu, "Sentiment classification: The contribution of ensemble learning," *Decis. Support Syst.*, vol. 57, pp. 77–93, Jan. 2014.

[9] R. Moraes, J. F. Valiati, and W. P. Gavião Neto, "Document-level sentiment classification: An empirical comparison between SVM and ANN," *Expert Syst. Appl.*, vol. 40, no. 2, pp. 621–633, Feb. 2013.

[10] A. S. H. Basari, B. Hussin, I. G. P. Ananta, and J. Zeniarja, "Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization," *Procedia Eng.*, vol. 53, pp. 453–462, Jan. 2013.

[11] M. Ghiassi, J. Skinner, and D. Zimbra, "Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network," *Expert Syst. Appl.*, vol. 40, no. 16, pp. 6266–6282, Nov. 2013.

[12] K. Ravi, V. Ravi, and P. S. R. K. Prasad, "Fuzzy formal concept analysis based opinion mining for CRM in financial services," *Appl. Soft Comput.*, vol. 60, pp. 786–807, Nov. 2017.

[13] K. Ravi and V. Ravi, "Irony Detection Using Neural Network Language Model, Psycholinguistic Features and Text Mining," in *2018 IEEE 17th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC)*, 2018, pp. 254–260.

[14] T. Roelleke and J. Wang, "TF-IDF uncovered: a study of theories and probabilities," in *31st Annual International ACM SIGIR conference on research and development in information retrieval*, 2008, pp. 435–442.

[15] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," in *Proceedings of the International Conference on Learning Representations*, 2013.

[16] Y. Tausczik and J. Pennebaker, "The psychological meaning of words: LIWC and computerized text analysis methods," *J. Lang. Soc. Psychol.*, vol. 29, no. 1, pp. 24–54, 2010.

[17] B. Schölkopf and A. J. Smola, *Learning with kernels : support vector machines, regularization, optimization, and beyond*. Cambridge, MA, USA: MIT Press, 2002.

[18] I. Rish, "An empirical study of the naive Bayes classifier," in *IJCAI-workshop on empirical methods in artificial intelligence*, 2001, pp. 41–46.

[19] S. H. Walker and D. B. Duncan, "Estimation of the probability of an event as a function of several independent variables.," *Biometrika*, vol. 54, no. 1, pp. 167–79, Jun. 1967.

[20] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, Mar. 1986.

[21] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967.

[22] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[23] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, 2016, pp. 785–794.

[24] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986.


.