A Machine Learning based Framework for Sentiment Classification: Indian Railways Case Study

D. Krishna Madhuri

Abstract*: *Machine learning in the field of computer science is the application of Artificial Intelligence (AI) that helps in making systems intelligent. It focuses on producing algorithms that may lead to AI applications in the real world. As enterprises are producing huge amount of data, it became indispensable to have machine learning techniques in place for discovering business intelligence from data for strategic decision making. However, in the contemporary era, the traditional data may be deemed inadequate for decision making. The rationale behind this is that people of all walks of life are able to exchange ideas and opinions/sentiments over social media like Facebook and Twitter. In other words, there is social feedback exists in Online Social Networks (OSNs). Collection of social media data related to business and using machine learning algorithms to extract useful knowhow from such data bestows competitive edge to enterprises. The existing literature on sentiment analysis has plenty of methods for discovering sentiments. However, it is still an open problem to have optimizations. In this paper we proposed a framework for discovering sentiments from tweets of Indian Railways. This is a domain specific framework which leverages business intelligence through different classifiers such as C4.5, Naive Bayes, SVM and Random Forest. An evaluation procedure with measures like precision, recall, F-Measure and accuracy is provided. The empirical study with Indian Railways case study revealed that the proposed framework is useful in sentiment analysis and can be tailored to suit other domains as well. By considering the atweets of Indian Railways as a case study evaluation is made in terms of precision, recall and F-Measure.

Keywords: Sentiment classification, machine learning, C4.5, Naive Bayes, SVM, Random Forest

I. INTRODUCTION

Knowledge required for accurate decision making is driving force in the research pertaining to AI. Through machine learning phenomenon, it has been around and became an established area for helping organizations in making accurate decisions. Information Technology (IT) sections of organizations should have model operandi to exploit AI procedures to be crucial in the growth of respective organizations. With the emergence of social media, focus is shifted (in addition to traditional approaches) to this wealth of knowledge where sentiments of people can help in making expert decisions for business growth. Social feedback is the term that refers to the opinions or sentiments of people exchanges over OSNs. In this context, the research in this paper is purely based on the data of social media Twitter. Before describing the research carried out in this paper, here is the essence of literature review made. Supervised learning methods for sentiment classification are explored in [1], [2], [3], [5], [12] and [18]. There are many specific method employed for sentiment analysis. They include sentiment-based word embedding [1], N-gram machine learning [3], deep learning [8], dependency based rules [13], and graph based approach [23], lexicon-based method [26] and Convolution Neural Networks (CNN). From the review of literature, it is understood that the methods employed are useful in sentiment classification. However, it is an open problem to have further optimizations and domain specific investigations to exploit sentiment classification in a better way. Towards this end, in this paper we proposed a framework for garnering sentiments for Indian Railways. The Indian Railways case study is considered as it helps in investigating domain specific dynamics of sentiment analysis. A framework is proposed to explore sentiment classification with supervised learning methods like C4.5, Naive Bayes, SVM and Random Forest. The framework guides the operations in terms of training and testing phases. The methodology includes evaluation procedure with metrics like accuracy, precision, recall and F-Measure based on confusion matrix made with True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN). Empirical study is made by using dataset collected using Twitter API from Twitter accounts like @RailMinIndia and @IR. A prototype application is built to demonstrate proof of the concept. Our contributions in this paper are as follows.

- We proposed a methodology for sentiment classification using machine learning algorithms like C4.5, Naive Bayes, SVM and Random Forest.
- We built a prototype application to evaluate the performance of the aforementioned algorithms with tweets of Indian Railways as case study. Evaluation is made in terms of precision, recall and F-Measure.

The reminder of the paper is structured as follows. In Section 2, existing literature on sentiment classification is reviewed to know the state of the art. In Section 3, we proposed a methodology based on machine learning algorithms for sentiment classification. Section 4 provides details of the supervised learning algorithms used in this paper.
Section 5 presents experimental results while section 6 covers conclusions and gives possible future scope of the research.

II. RELATED WORK

This section reviews literature related to sentiment classification methods including machine learning. Tang et al.

Revised Manuscript Received on 8 February 2019.
D. Krishna Madhuri, Department of Computer Science and Engineering, GRIET, Hyderabad , India.

[1] explored Twitter sentiment classification with the approach known as sentiment-specific word embedding. Recurrent Neural Network (RNN) is employed in [2] for sentiment classification. On the other hand, n-gram machine learning phenomenon is used in [3] for sentiment analysis. Convolution networks at character level are used for text classification in [4]. Merger of offline and online feedback is the main focus in [5] for analysing public sentiments. Lexical based approaches are followed in [26] while combination of learning based and lexicon based methodology is employed in [6]. The concept of lifelong learning is investigated in [7] for sentiment classification. Deep learning based analysis is described in [8], [16] and [17]. Text mining approaches like extraction, classification and clustering are the focus in [9] while emotions are included in the study of sentiment analysis in [10] and [15]. Genetic rank aggregation coupled with feature selection is used in [11]. Sentiment classification with context-sensitive approach is studied in [12]. Sentiment patterns are explored in [13] based on the dependency based rules. A multi-label classification approach is used in [14] for better classification of sentiments. Twitter corpus is used in [18] for opinion mining while sentient strength is detected in [19]. The concept of word vectors is employed in [20]. A good survey of sentiment analysis is made in [24] while in [28] digital pathology scans are used for classification.

Other sentiment based approaches found in the literature include dependency tree based method [21], target dependency based [22], graph based approach [23], ontology based [25]; social relations orientation [27], CNN based [29] and Twin Transfer Learning (TTL). From the review of literature, it is understood that machine learning has potential to mine sentiments associated with domain specific data sources. However, there is need for evaluation of performance of supervised learning methods for sentiment classification on tweets of Indian Railways so as to provide useful insights to make well informed decisions.

III. MACHINE LEARNING TECHNIQUES

Machine learning techniques are broadly classified into supervised and unsupervised categories. The supervised methods need training phase while the other category does not need training and uses some sort of similarity measure to learn and perform intended operation. In this paper four supervised learning methods are employed for sentiment classification of tweets of Indian Railways. The algorithms are known as C4.5, Naive Bayes, SVM and Random Forest. C4.5 is extension of well knows algorithm known as ID3. C4.5 generates decision trees that are used for classification of sentiments. Naive Bayes is a probabilistic classifier which is based on Bayes' theorem. SVM is the classification algorithm which is widely used in data mining research. It is a binary classifier but support kernels for multi-class classification as well. Random forest is another algorithm that can be used for classification in machine learning. All these supervised learning algorithms are used in this paper for sentiment classification. More details on the usage of them are provided in the Section 4.

IV. METHODOLOGY FOR SENTIMENT CLASSIFICATION

This section provides methodology used to have sentiment classification with the tweets dataset collection from Twitter accounts of Indian Railways. Twitter API is used to have live connectivity to twitter and collected tweets. Afterwards, the tweets are pre-processed to have training and testing sets. The following sub sections provide more details on the methodology employed.

A Problem Formulation

Social networking web sites like Facebook and Twitter, to mention few, have changed how information is disseminated or shared instantly. Moreover, they became a platform for social meetings and information exchange. In the process, there is possibility that people express their valuable opinions on products and services. Thus social feedback is made available readily. Any organization cannot afford to have a blind eye on this feedback. However, it is challenging to collect social media content and use it for discovering business intelligence. This is the problem addressed in this paper with Indian Railways case study.

B Proposed Framework

A framework is proposed to guide the research on sentiment classification. Since sentiments are opinions in social media, tweets of Indian Railways are collected from Twitter web site. The tweets thus collected have wealth of knowledge in the form of sentiments of people. Garnering such intelligence can boost decision making process of any organization. Therefore, the framework is aimed at helping in sentiment analysis with different machine learning algorithms like C4.5, Naive Bayes, SVM and Random Forest. The framework is illustrated in Figure 1. There are two phases like training and testing. It is required as the machine learning approach is known as supervised learning that needs training set for Building classifier

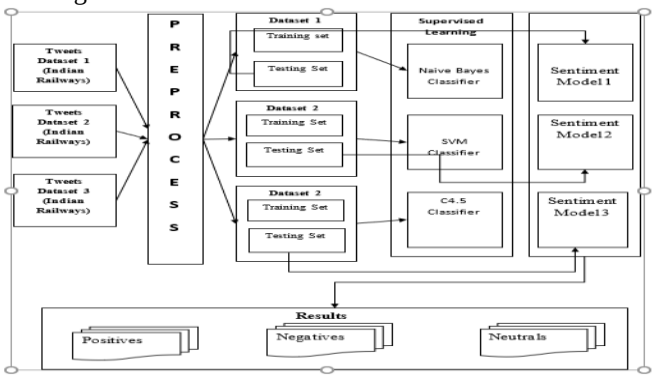

Figure 1: Framework for sentiment classification of tweets of Indian Railways.

The given tweets dataset is subjected to pre-processing. It generates training and testing datasets. The training set is used to learn a classifier. The Random Forest is not explicitly provided in the framework but the procedure remains same for it as well. The supervised learning (training) phase lets classification algorithms to train a classifier based on the class labels provided in the training set.

The testing set does not contain class labels as they are to be predicted by the classifiers built. The framework facilitates to build classifier for all algorithms. Once classifier is built, the training phase is said to be completed. The testing set is given input to testing phase in which the knowledge model built by the classifier is used to predict class labels. The results are provided in terms of positives, negatives and neutral. There is necessity to know how accurately classification is made by the algorithms. This is done with the help of evaluation methodology described in Section 4.3.

C Evaluation

This section provides methodology used for evaluation of machine learning techniques and their utility in mining sentiments from tweets of Indian Railways. Figure 2 shows the procedure employed for evaluation

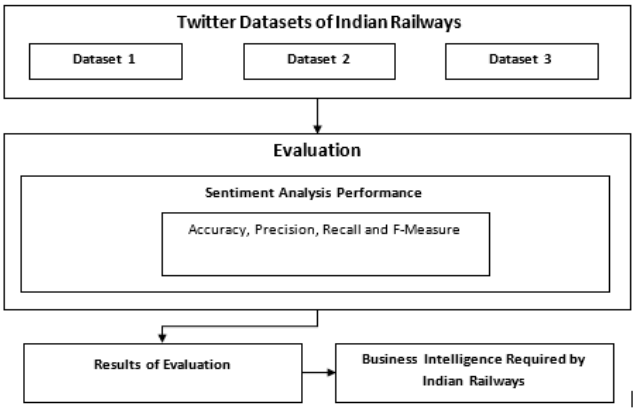

Figure 2: Evaluation methodology

As shown in Figure 2, evaluation procedure is illustrated. The metrics used for evaluation are accuracy, precision, recall and F-Measure. The evaluation results and the classification results provide required business intelligence to Indian Railways. This social knowhow can help the organization to make strategic decisions. Table 1 shows the confusion matrix based on which evaluation is realized.

	Ground Truth (Correct Prediction)	Ground Truth (In correct Prediction)
Result of algorithm (Correct Prediction)	True Positive (TP)	False Positive (FP)
Result of algorithm (Incorrect Prediction)	False Negative (FN)	True Negative (TN)

Table 1: Shows confusion matrix

As shown in Table 1, the confusion matrix provides the meaning of TP, FP, TN and FN. These values are used to have metric like precision, recall, accuracy and F-Measure. Accuracy is computed as in Eq. (1).

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

This measure shows the ability of a classification technique to differentiate a sentiment with non sentiment in given dataset. It uses TP, FP, TN and FN provided in the confusion matrix. Precision on the other hand is the measure used to know the ability of a classification technique to know the fraction of correctly classified instances among all the instances considered [9]. It is computed as in Eq. (2).

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

In the same fashion, recall is the measure to know fraction of correctly classified instances among all the instances that are supposed to be in correctly labelled instances. It is measured as in Eq. (3).

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

There is another measure which is based on the precision and recall. It is harmonic mean of these two measures. It is called F-Measure which is computed as in Eq. (4).

F-Measure = 2*((precision*recall)/(precision + recall)) (4)

V. EXPERIMENTAL RESULTS

Experiments are made with the machine learning methods such as C4.5, Naive Bayes, SVM and Random Forest against dataset containing tweets of Indian Railways collected from its Twitter accounts. The results are observed in terms of TP, FP, TN and FN and the measures of evaluation as discussed in Section 4.3 are used to observe performance dynamics of the supervised learning methods.

Machine Learning Algorithms	FP Rate
C.4.5	8.2
NaiveBayes	9
SVM	6.9
Random Forest	7.6

Table 2: Results in terms of false positives

As shown in Table 2, the FPR is presented for all algorithms. It, along with TP, TN and FN, is used to compute precision, recall and F-Measure.

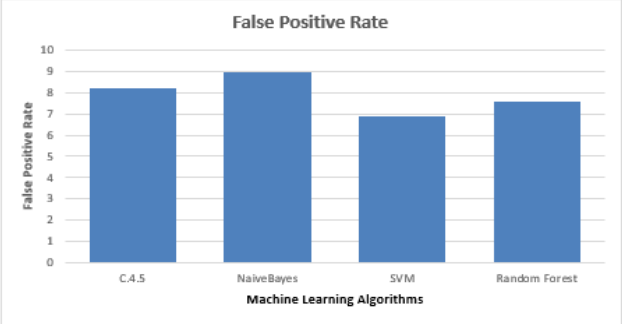

Figure 3: FPR performance of machine learning approaches

As presented in Figure 3, it is evident that SVM has better performance with least percentage of false positives. Naive Bayes is the algorithm that showed highest number of false positives. Higher FPR indicates less in performance.

Algorithms	Accuracy	Precision	F-Measure	Recall
C4.5	89.5	84.2	83	81.8
NaiveBayes	89	83.3	81.5	78
SVM	91.5	88.5	87.5	83
Random Forest	90.5	89.2	87	81.5

Table 3: Performance of machine learning techniques in sentiment classification

As shown in Table 3, the computed accuracy, precision, recall and F-Measure are presented against the machine learning techniques.

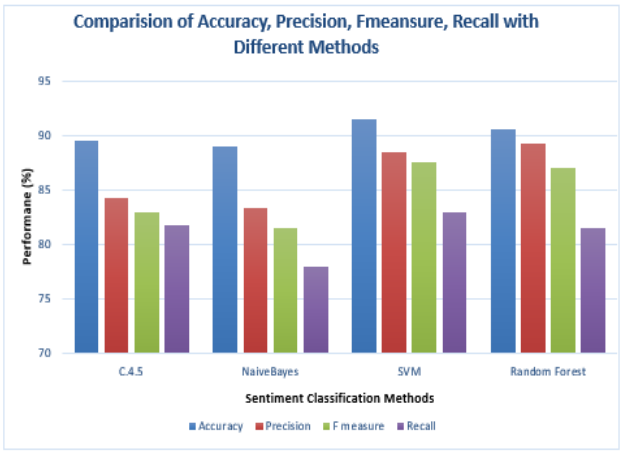

Figure 4: Performance of machine learning algorithms in sentiment classification

As presented in Figure 4, the supervised learning methods used in experiments are presented in horizontal axis. The vertical axis on the other hand shows the performance of the algorithms in terms of accuracy, precision, recall and F-Measure. Highest accuracy is exhibited by SVM. Random Forest and SVM showed highest precision and recall respectively. SVM shows highest F-Measure over all other algorithms. SVM has higher performance in sentiment classification.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, a methodology for sentiment classification has proposed. The case study considered is Indian Railways. It explores supervised learning methods like C4.5, Naive Bayes, SVM and Random Forest in sentiment classification of tweets of Indian Railways. Since tweets carry valuable social feedback and opinion on Indian Railways, this study provides useful insights on sentiment classification. It considers positive, negative and neutral sentiments. The proposed framework has two phases such as training and testing. In the training phase, the classifiers are built to have different knowledge models that can be used in sentiment classification. In the testing phase, the generate knowledge models are actually sussed to label unlabelled instances that arrive online. An evaluation methodology is defined with measures like accuracy, precision, recall and F-Measure. The experimental results revealed that SVM has shown higher performance when compared with C4.5, Naive Bayes and Random Forest in the proposed sentiment classification framework. In future we consider fuzzy logic for further refinement of sentiment classification with Highly Positive, Positive, Neutral, Negative and Highly Negative. In this proposed methodology, It will help in obtaining fine grained business intelligence for strategic decision making based on the feed back.

REFERENCES

1. Duyu Tang, Furu Wei, Nan Yang, Ming Zhou, Ting Liu and Bing Qin. (2014). Learning Sentiment-Specific Word Embedding for Twitter Sentiment Classification, p1555–1565.
2. Duyu Tang, Bing Qin and Ting Liu. (2015). Document Modeling with Gated Recurrent Neural Network for Sentiment Classification, p1422–1432.
3. Abinash Tripathy, Ankit Agrawal and Santanu Kumar Rath. (2016). Classification of sentiment reviews using n-gram machine learning approach. elsever, p117–126.
4. Xiang Zhang, Junbo Zhao and Yann LeCun. (2015). Character-level Convolutional Networks for Text Classification, p1-9.
5. Leona Yi-Fan Su, Michael A. Cacciatore, Xuan Liang, Dominique Brossard, Dietram A. Scheufele and Michael A. Xenos. (2016). Analyzing public sentiments online combining human- and computer-based content analysis. Information, Communication & Society, p1-24.
6. Lei Zhang, Riddhiman Ghosh, Mohamed Dekhil, Meichun Hsu and Bing Liu . (2011). Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis, p1-9.
7. Zhiyuan Chen, Nianzu Ma and Bing Liu. (2018). Lifelong Learning for Sentiment Classification, p1-8.
8. Navonil Majumder, Soujanya Poria, Alexander Gelbukh and Erik Cambria . (2017). Deep Learning-Based Document Modeling for Personality Detection from Text. iEEE iNTElliGENT SYSTEmS, p74-79.
9. Mehdi Allahyari. (2017). A Brief Survey of Text Mining Classification, Clustering and Extraction Techniques. KDD Bigdas, p1-13.
10. Maria Giatsoglou. (2017). Sentiment analysis leveraging emotions and word embeddings. elsever, p214–224.
11. Aytug˘ Onan and Serdar Korukog˘lu. (2017). A feature selection model based on genetic rank aggregation for text sentiment classification. Journal of Information Science. 43 (1), p25–38.
12. Yafeng Ren, Yue Zhang, Meishan Zhang and Donghong Ji. (2016). Context-Sensitive Twitter Sentiment Classification Using Neural Network, p1-7.
13. Soujanya Poria, Erik Cambria, Grégoire Winterstein and Guang-Bin Huang. (2014). Sentic patterns: Dependency-based rules for concept-level sentiment analysis. Elsever, 69, p45–63.
14. Shuhua Monica Liu and Jiun-Hung Chen. (2015). A multi-label classification based approach for sentiment classification. elsever . 42, p1083–1093.
15. Weiyuan Li and Hua Xu . (2013). Text-based emotion classification using emotion cause extraction. elsever, p1-8.
16. Xavier Glorot. (2011). Domain Adaptation for Large-Scale Sentiment Classification A Deep Learning Approach, p1-8.
17. Richard Socher, Alex Perelygin, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng and Christopher Potts. (2013). Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank, p1631–1642.
18. Alexander Pak and Patrick Paroubek. (2013). Twitter as a Corpus for Sentiment Analysis and Opinion Mining, p1320-1326.
19. Mike Thelwall, Kevan Buckley, Georgios Paltoglou and Di Cai . (2012). Sentiment Strength Detection in Short Informal Text. Journal of the American Society for Information Science and Technology. 61 (12), p2544–2558.
20. Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng and Christopher Potts. (2011). Learning Word Vectors for Sentiment Analysis, p142–150.
21. Tetsuji Nakagawa. (2010). Dependency Tree-based Sentiment Classification using CRFs with Hidden Variables, p786–794.
22. Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu and Tiejun Zhao. (2011). Target-dependent Twitter Sentiment Classification, p151–160.
23. Xiaolong Wang. (2011). Topic Sentiment Analysis in Twitter: A Graph-based Hashtag Sentiment Classification Approach. ACM, p1-10.
24. G.Vinodhini and RM.Chandrasekaran. (2012). Sentiment Analysis and Opinion Mining: A Survey. International Journal of Advanced Research in Computer Science and Software Engineering. 2 (6), p1-11.
25. Efstratios Kontopoulos . (2013). Ontology-based sentiment analysis of twitter posts. elsever, p4065–4074.

26. Yan Dang, Yulei Zhang, and Hsinchun Chen. (2010). A Lexicon-Enhanced Method for Sentiment Classification: An Experiment on Online Product Reviews. IEEE, p1-8.
27. Xia Hu, Lei Tang, Jiliang Tang and Huan Liu. (2013). Exploiting Social Relations for Sentiment Analysis in Microblogging. ACM, p1-10.
28. Morteza Babaie. (2011). Classification and Retrieval of Digital Pathology Scans: A New Dataset. IEEE, p1-10.
29. Soujanya Poriaa. (2017). Ensemble application of convolutional neural networks and multiple kernel learning for multimodal sentiment analysis. elsever, p217–230.

AUTHORS PROFILE

D. Krishna Madhuri Received her B.Tech degree in computer Science and engineering from Swarnandhra college of Engineering & Technology, Narsapur, Andhra Pradesh in 2009, and M.Tech Degree in Computer Science and Engineering from Sridevi Women's Engineering College Hyderabad, Telangana in 2012. She is currently pursuing her Ph.D degree from Sri Satyasai University, Sehore, Madhya Pradesh. Currently she is working as Assistant professor in the department of Computer Science and Engineering of Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, Telangana India. Her research interest includes Data Bases, Data warehouse and Data Mining and Big Data.