ISSN: 2366-1313

Detection of fake online reviews using semi supervised and supervised learning

[1]B. TRIVENI, [2]DASARI RANJITH, [3]NANNURU SHIVA KRISHNA REDDY, [4]VUPPALA NITHISH KUMAR

[1]Associate Professor, Dept.of CSE, Teegala Krishna Reddy Engineering College, Meerpet, Hyderabad,

[2]BTech student, Dept.of CSE, Teegala Krishna Reddy Engineering College, Meerpet, Hyderabad, dasariranjith33@gmail.com

[3]BTech student, Dept.of CSE, Teegala Krishna Reddy Engineering College, Meerpet, Hyderabad, Shivakrishnareddyn@gmail.com

[4]BTech student, Dept.of CSE, Teegala Krishna Reddy Engineering College, Meerpet, Hyderabad, nithishkumarvuppala222@gmail.com

Abstract: *Online reviews have a great impact on business and commerce today. The decision to buy products online is generally based on the reviews given by the users. Therefore, opportunistic individuals or companies attempt to manipulate product reviews for their own interests. This paper presents some models of supervised and supervised text mining models to find fake online reviews and compares the performance of both methods on datasets containing hotel reviews.*

Keywords: *Supervised, online user reviews, products, detection of fake user identification.*

I. INTRODUCTION

Technologies are developing rapidly. Old technologies are constantly being replaced by new and developing ones. These new technologies allow humans to do their jobs efficiently. Such a technological development is the online market. We may purchase and reserve the use of the Websites online. Almost everyone tries to do reviews before buying some products or services. Therefore, online reviews have become a huge source of group popularity. In addition, they have a significant impact on the marketing and promotion of services and products. As the online market has expanded, so have fake reviews online. It is surprisingly dependent on the problem. People can run fake reviews promoting their own personal products that harm real customers. In addition, aggressive groups may attempt to harm the recognition of others by providing false negative evaluations.

ISSN: 2366-1313

Researchers have been studying roughly several methods to spot such fake reviews online. Some strategies are primarily based on review content, and others are based entirely on user behavior post reviews. The content-based analysis makes specialization in what is written in the review. This is the review text where the focus on user behavior is US, IP address, number of reviewers' posts, etc. Most of the proposed operations are models for supervised classes. Furthermore, a few researchers have worked with semi-supervised models. Semi-moderated methods for losing trusted marks are offered for reviews [1].

Nowadays, social media is more popular as mobile devices can easily access social media from anywhere. Therefore, social networks have become a vital research topic in many fields. As the diversity of humans using the social community grows daily, they talk to their peers so they can share their own feelings daily, and wide-ranging insights are generated. Monitoring or censoring social media is the most important issue in the contemporary situation these days. These days, many organizations have been using social media marketing to promote their products or brands, so it will be important for them to calculate each product's achievement and value [2]. Several tools with extensions were required: one to assess how many customers in your logo are attracted by your promotion, and the second to find out what people think about your unique brand to build social media monitoring. Comparing user reviews is not as clean as it seems to all customers. It may be necessary to perform a sentiment analysis, which is defined as recognizing the polarities of buyer behaviour, personality, and feelings in a single file or sentence to assess your position. To address this, we need machine learning and natural language processing techniques, where most developers run into trouble while looking to build their own teams. In recent years, there has been increasing interest in the help of social media analytics for advertising, sentiment assessment, and knowledge network coherence. Social media data is modelled according to the various classifications attributed to 'aggregate statistics', i.e., Size, speed, and versatility. The analysis of social media desires on large volumes of records must be done in an efficient and timely manner. Media content analysis was centralized in the social sciences, given the important role social networks play in shaping public opinion. This type of test usually relies on the initial coding of the text being tested, a step that involves parsing and annotating the text and which limits the amount of data that can be

ISSN: 2366-1313

analysed. As the web improves, more and more people are going online and becoming record makers rather than more effective agents of information in the afterlife, leading to intense alarm and data overload. There is a lot of special information in the online craft reviews, which plays a big role in the selection techniques. For example, a customer will decide what to buy if they see valuable reviews[3].

In this paper, we do some kinds of processes to detect fake reviews online, some of which can be semi-monitored, and some of them can be censored. For semi-moderated knowledge, we use a set of expectation maximization rules. Naive Bayes statistical classifier and support vector machines (SVMs) are used as classifiers in our study paints to improve classification performance. We have particularly focused on the content of assessment-based technologies. as our advantage. Remember the frequency of the phrase used, the polarity of feelings, and the length of the overall description posted by others, especially users trusted friend

II. LITERATURE SURVEY

Researchers have been studying about many approaches for detection of these fake online reviews. Some approaches are reviewing content based and some are based on behaviour of the user who is posting reviews. Content based study focuses on what is written on the review that is the text of the review where user behaviour-based method focuses on country, ip-address, number of posts of the reviewer etc. Most of the proposed approaches are supervised classification models. Few researchers, also have worked with semi-supervised models. Semi-supervised methods are being introduced for lack of reliable labelling of the reviews

This paper chose primarily three methods for text classification because of their relative popularity and success in prediction of sentiments:

Naive Bayes: This works on the assumption of conditional independence and despite this oversimplified assumption, Naive Bayes performs well in many complex real-world problems. Naive Bayes classifier is superior in terms of CPU and memory consumption. **Support Vector Machines:** SVM also provides a robust approach to build text classifiers and was picked because of its ability to handle High dimensional input space. When learning text classifiers, many

ISSN: 2366-1313

(morethan10000) features can been countered. Since SVMs use over fitting protection, which does not necessarily depend on the number of features, they have the potential to handle these large feature spaces.

Maximum Entropy: MaxEnt Naïve Bayes is based on conditional independence assumption, hence to ensure that this paper covers an alternative, it uses Maximum Entropy that does not assume conditional independence. It is based on the Principle of Maximum Entropy and from all the models that fit the training data, selects the one which has the largest entropy. Although it takes more time than Naïve Bayes to train the model, this method has proven to be useful in cases where we do not know anything about the prior distribution (Hening-Thurau etal., 2003) state that customer comments articulated via the Internet are available to a large number of other customers, and therefore can be expected to have a significant impact on the success of goods and services. This on consumer buying and communication behavior are tested in a large-scale empirical study. The results illustrate that consumers read online articulations mainly to save decision-making time and make better buying decisions. Structural equation modeling shows that their motives for retrieving online articulations strongly influence their behavior (Duan et al., 2008) showed that both a movie's box office revenue and WOM valence significantly influence WOM volume. WOM volume in turn leads to higher retrieve other customer's online articulations from webbased consumer opinion platforms. The relevance of these motives and their impact box office performance. This positive feedback mechanism highlights the importance of WOM in generating and sustaining retail revenue. (Chevalier & Mayzlin, 2006) hypothesized that buyers suspect that many reviewers are authors or other biased parties. They found marginal (negative) impact of 1-star reviews is greater than the (positive) impact of 5-star reviews. The results suggest that new forms of customer communication on the Internet have an important impact on customer behavior. Work on sentiment analysis found using a formal approach is the work by (Simancík and Lee, 2009). The paper presents a method to detect sentiment of newspaper headlines, in fact partially using the same grammar formalism that later will be presented and used in this work, however without the combinatorial logic approach. The paper focus on some specific problems arising with analysing newspaper headlines, e.g., such as headline texts often do not constitute a complete sentence, etc.

ISSN: 2366-1313

J. K. Rout et al. [4] With more consumers using online opinion reviews to inform their service decision making, opinion reviews have an economic impact on the bottom line of businesses. Unsurprisingly, opportunistic individuals or groups have attempted to abuse or manipulate online opinion reviews (e.g., spam reviews) to make profits and so on, and that detecting deceptive and fake opinion reviews is a topic of ongoing research interest. In this paper, we explain how semi-supervised learning methods can be used to detect spam reviews, prior to demonstrating its utility using a data set of hotel reviews.

E. P. Lim, et al. [5] This job seeks to detect customers creating spam reviews or reviewing spammers. We detected several distinct behaviours for testing spammers and changed these behaviours if you want to find spammers. Specifically, we seek to model subsequent behaviours. First, spammers may also target specific goods or product organizations to maximize their impact. Second, they tend to avoid alternative reviewers of their merchandise results. We then select a subset of relatively suspicious reviewers for further examination by our consumer testers with the help of an online spammer testing program developed primarily for test subjects. Our results show that our proposed methods and supervised classification effectively detect spammers and outperform other standard methods that rely solely on help votes. In the end, we showed that detected spammers impact ratings more than useless reviewers.

Cardie et al. [6] Consumer buying decisions are increasingly influenced by the reviews that consumers generate online. As a result, there was a situation developing about being able to submit unsolicited misleading reviews - fake reviews that were intentionally written to appear true, to lie to the reader. In this paper, we explore widespread tactics for identifying deceptive online review spam based entirely on a new gold-preferred dataset, which consists of statistics from 3 unique domain names (e.g., hotel, restaurant, medical), each of which includes three patterns of opinions, ie. The customer gave honest reviews, Turker gave misleading reviews, and staff (professionals) gave misleading reviews. Furthermore, our technology attempts to take advantage of the general difference in language usage between misleading and honest reviews, which we hope will help customers make purchasing decisions and review portal operators, such as TripAdvisor or Yelp, to screen for potentially fraudulent practices on their websites.

III. PROPOSED WORK

ISSN: 2366-1313

We have applied both semi-supervised and supervised classifications. We used the Expectation-Maximization (EM) algorithm for the semi-moderated record set category. The expectation-maximization algorithm, first proposed by Karimpour et al., is designed to label unlabeled records for training. The rules work as follows: the classifier is first derived from the set of classified data. This workbook is then used to label the unlabeled dataset. Let this expected set of stickers be PU. Now any other classifier is derived from the mixed units of the labeled and unlabeled datasets and used to classify the unlabeled dataset again. This method is repeated until the PU group has stabilized. After producing a robust PU set, we taught the typing algorithm with the combined learning set for each labeled and unlabeled dataset and set it up to predict the validation dataset. The algorithm is given below.

Expectation-Maximization (EM) algorithm

Algorithm 2 EM Algorithm

INPUT: Labeled instance set L, and unlabeled instance set U.

OUTPUT: Deployable classifier, C.

```
1: C ← train(L);
2: PU = ∅;
3: while true do
4:    PU = predict(C,U);
5:    if PU same as in previous iteration then
6:       return C;
7:    end if
8:    C ← train(L ∪ PU);
9: end while
```

As classifier, we have used Support Vector machines (SVM) and Naive Bayes(NB) classifier with EM algorithm. Scikit Learn package of Python programming language provides sophisticated library of these classifiers. Hence for our research work, we have used Python with scikit-learn and numpy packages. We have tuned the parameters of the SVM for better results. For supervised classification, we have used Naive Bayes and SVM classifiers. We know, Naive Bayes classifier can be implemented where conditional independence property is maintained. As, text comes randomly from user mind, we can't know what the next line and word is going to be. Hence, Naive Bayes classifier is popularly used in text mining. It is probabilistic method hence it can be used

ISSN: 2366-1313

both for classification and regression. It is also very fast to calculate

To detect fake reviews online, we start with raw text data. We used a dataset that previous researchers had already compiled. We move unnecessary texts like articles and prepositions into facts. These textual statistics are then converted into numeric facts to make them fit for the workbook. Important and important abilities are drawn, after which the separation technique took place. We also used the 'Popular Gold' dataset curated with the help of Ott et al. [7], we didn't need ladders such as handling missing values, inconsistency removal, redundancy removal, etc. Instead, we needed to merge texts, create a dictionary, and assign texts to numeric rate due to pre-processing obligations. We used recall of phrase frequency, sentiment polarity, and evaluation duration as our characteristics. We have taken 2000 sentences as functions. Therefore, the dimensions of our function vector are 160_2002. We did not take n-grams or parts of speech as features because these are functions derived from a bag of words and may be intended for overfitting. The feature extraction system is summarized in Figure 1.

SYSTEM ARCHITECTURE

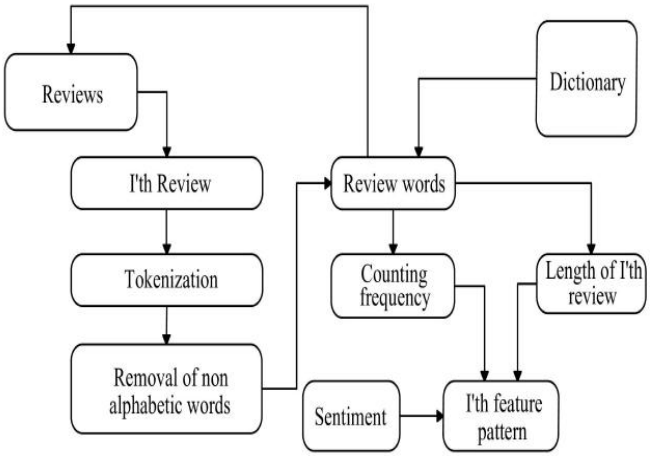


Fig. Stages of proposed feature extraction process

From the figure 1, we can see that, when we are working with i'th review, it's corresponding features are generated in the following procedure.

1) Each review goes through tokenization process first. Then, unnecessary words are removed and candidate feature words are generated.

2) Each candidate feature words are checked against the dictionary and if it's entry is available in the dictionary then it's frequency is counted and added to the column in the feature vector that corresponds the numeric map of the word.

3) Alongside with counting frequency, the length of the review is measured and added to the feature vector.

4) Finally, sentiment score which is available in the data set is added in the feature vector. We have assigned negative

ISSN: 2366-1313

sentiment as zero valued and positive sentiment as some positive valued in the feature vector.

We have applied both semi-supervised and supervised ratings. We used the maximization and expectation (EM) rule set for the semi-supervised classification of the record set.

The set of expectation maximization (EM) algorithm proposed by Karimpour et al. [8] is designed to label unlabeled information for use in education. The algorithm works as follows: the classifier is first derived from the labeled data set. This workbook is then used to label the unlabeled dataset. Let this expected set of stickers be PU. Now any other classifier is derived from the combined units of the labeled and unlabeled datasets and used to classify the unlabeled dataset again. This procedure is repeated until the hard PU is installed. After producing a robust PU set, we test the class rules with the combined training set of labeled and unlabeled datasets and initialize it to predict the test dataset

.

IV. RESULTS AND PERFORMANCE ANALYSIS

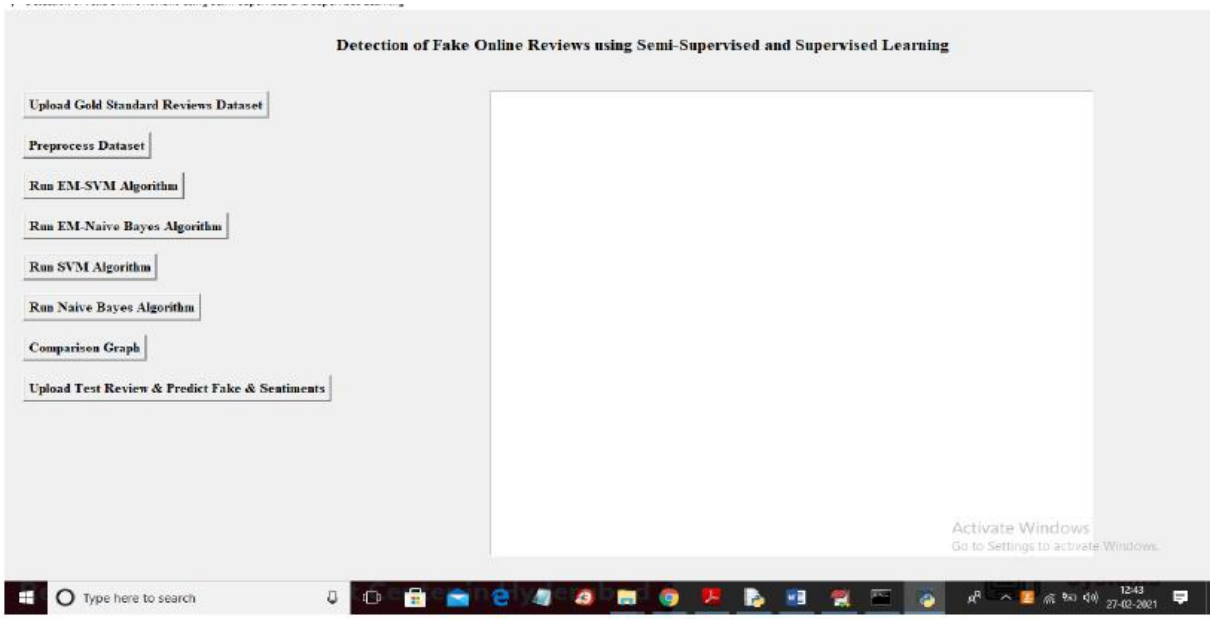


Fig.2 Application interface

In above screen click on 'Upload Gold Standard Reviews Dataset' button to upload dataset

ISSN: 2366-1313

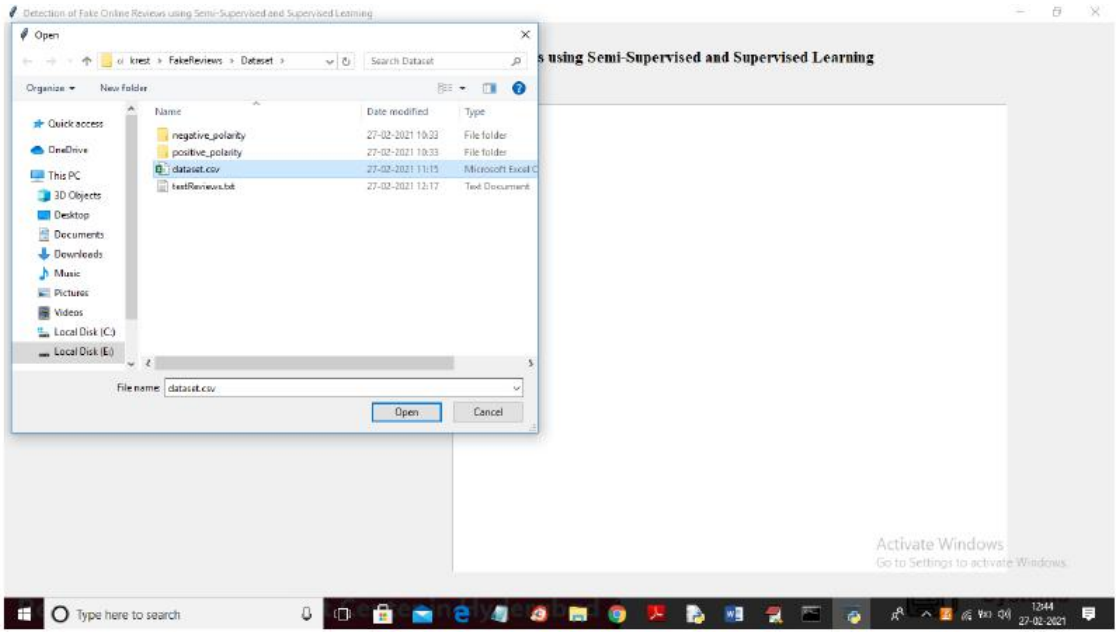


Fig.3 Uploading dataset

In above screen selecting and uploading ‘dataset.csv’ file and then click on ‘Open’ button to load dataset and to get below screen

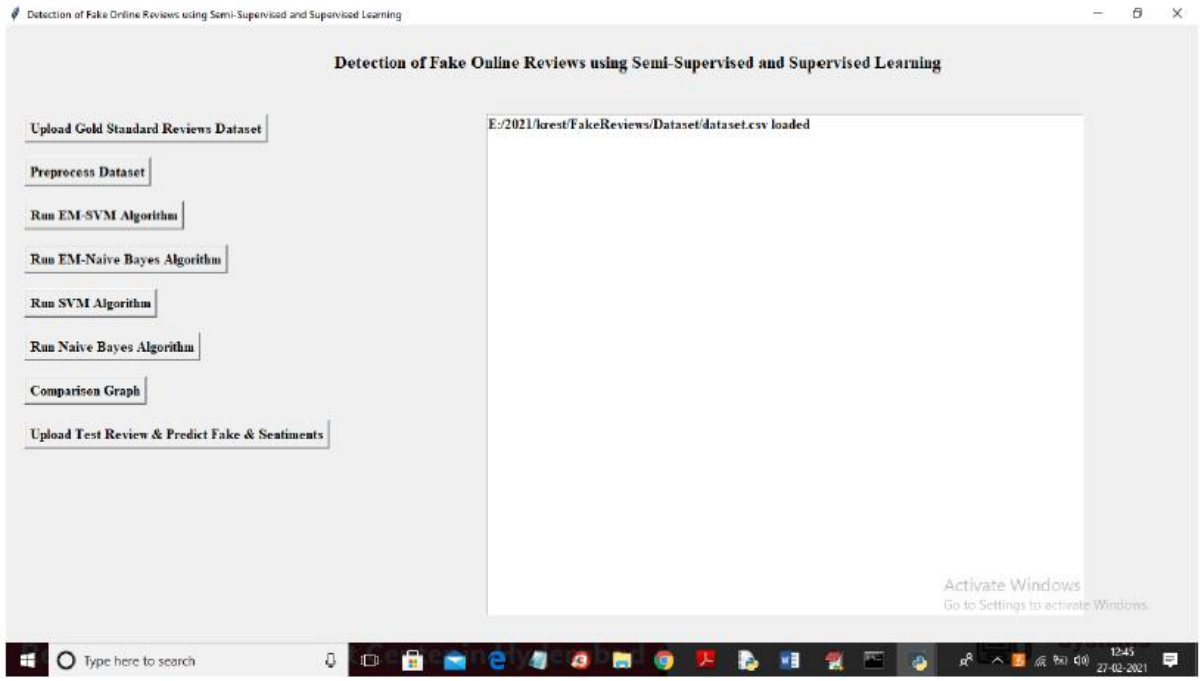


Fig.4 Pre-processing Dataset

In above screen dataset loaded and now click on ‘Preprocess Dataset’ button to read and process dataset and to get below screen

ISSN: 2366-1313

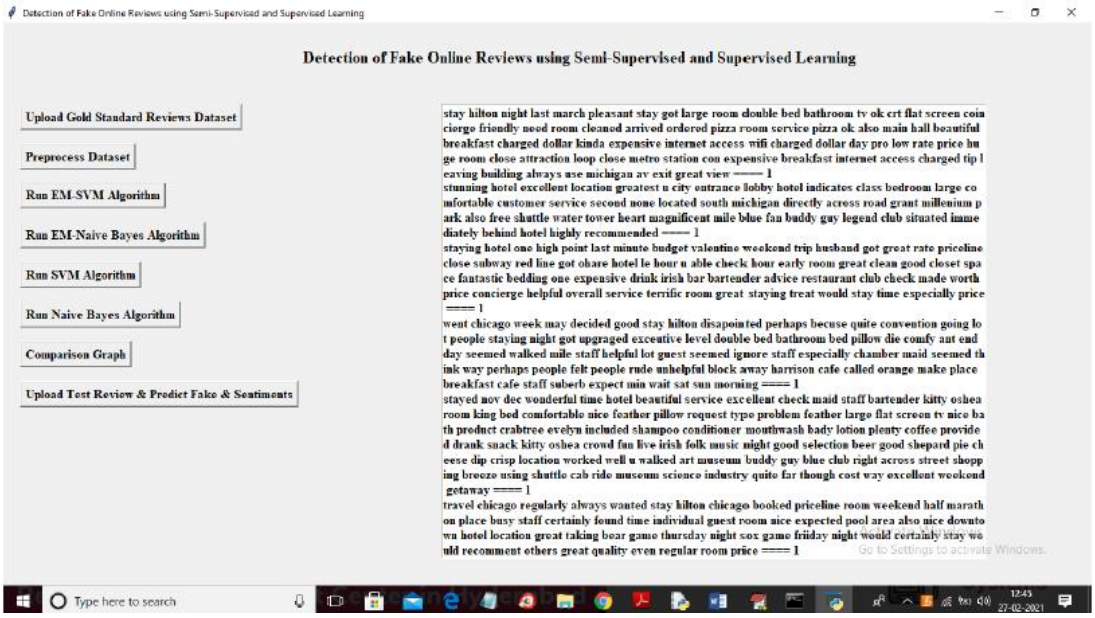


Fig.5 Preprocessed Dataset

In above screen from all reviews we removed stop words and after === symbol we can see it label as 0 or 1 and now scroll down above screen to bottom to see TF-IDF vector. You can see below screen

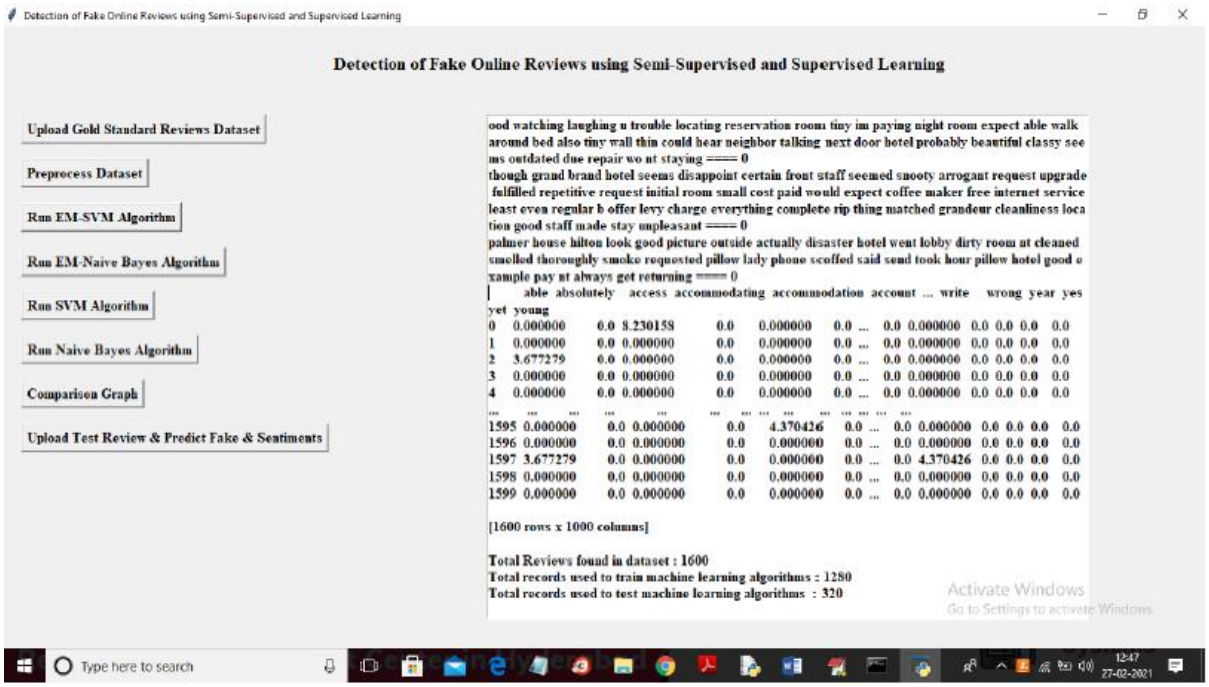


Fig.6 Run EM-SVM Algorithm

In above screen in top row we can see all words separated by TAB and in below top row we can see its numeric value calculated using TF-IDF. In above screen in bottom we can see dataset contains total 1600 reviews and then application using 1280 reviews for training and

ISSN: 2366-1313

320 reviews for testing. Now train and test data is ready and now click on 'Run EM-SVM Algorithm' button to train it.

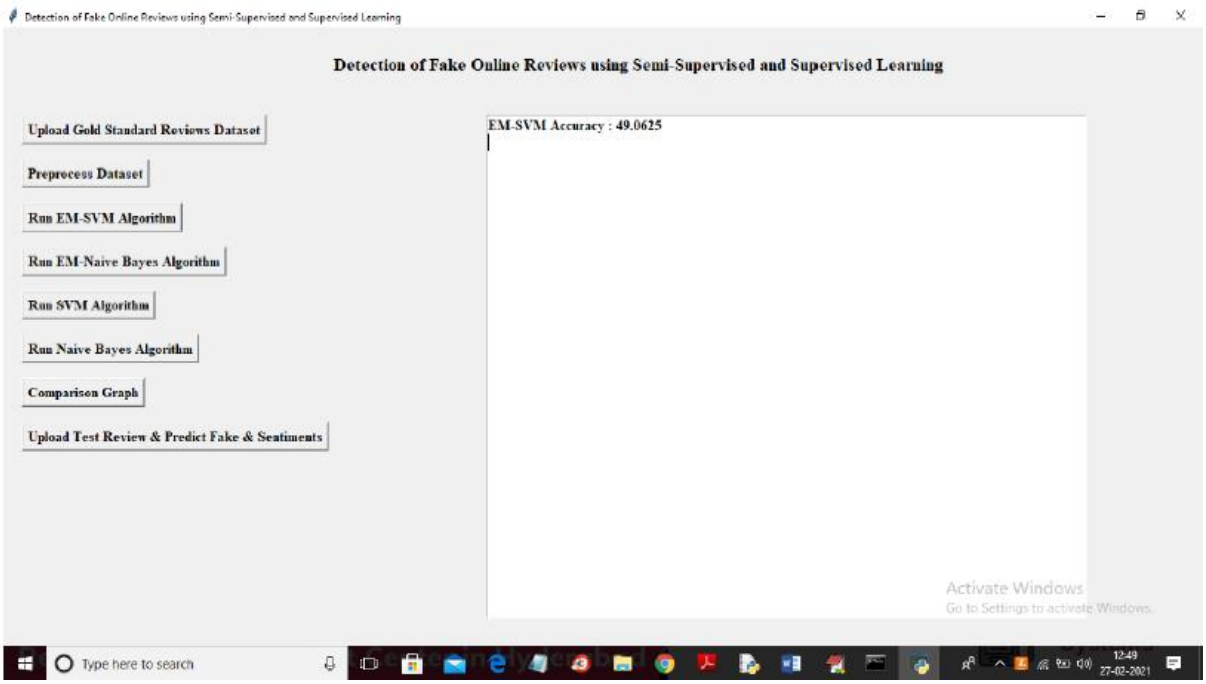


Fig.7 EM-SVM Accuracy

In above screen EM-SVM got 49% accuracy and similarly click next 3 buttons to train all algorithms

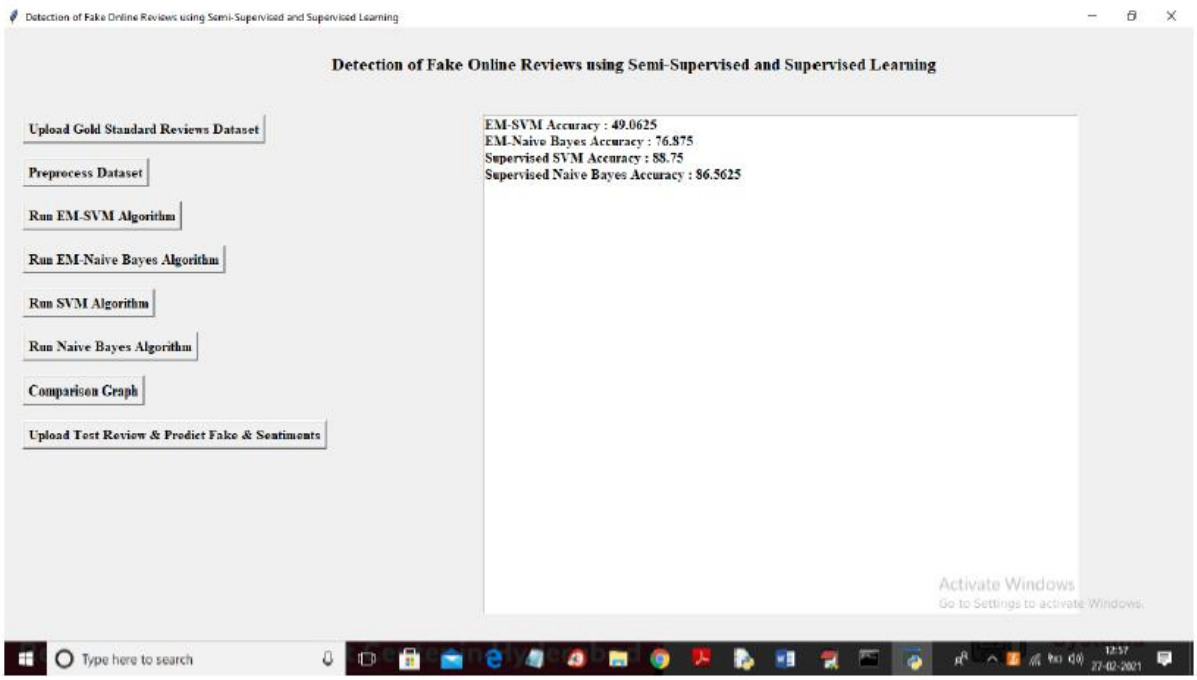


Fig.8 Accuracy of used algorithm

ISSN: 2366-1313

In above screen we can see EM algorithms are not working well but supervise algorithms are giving better accuracy and now click on 'Comparison Graph' button to get below graph.

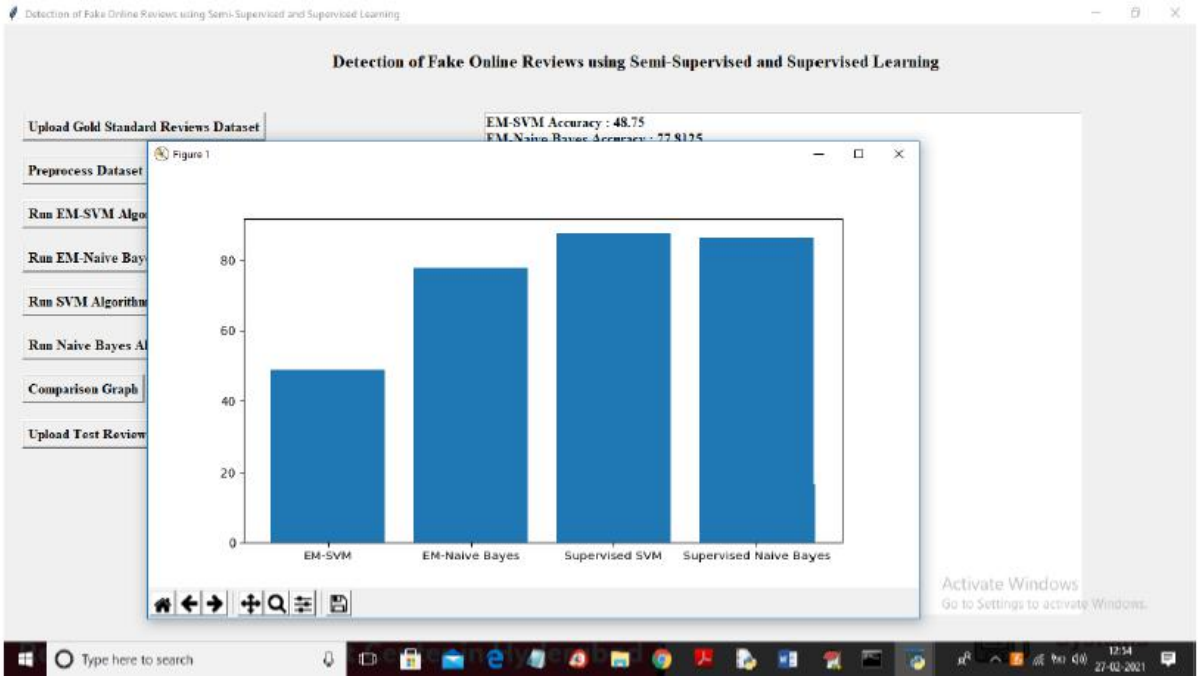
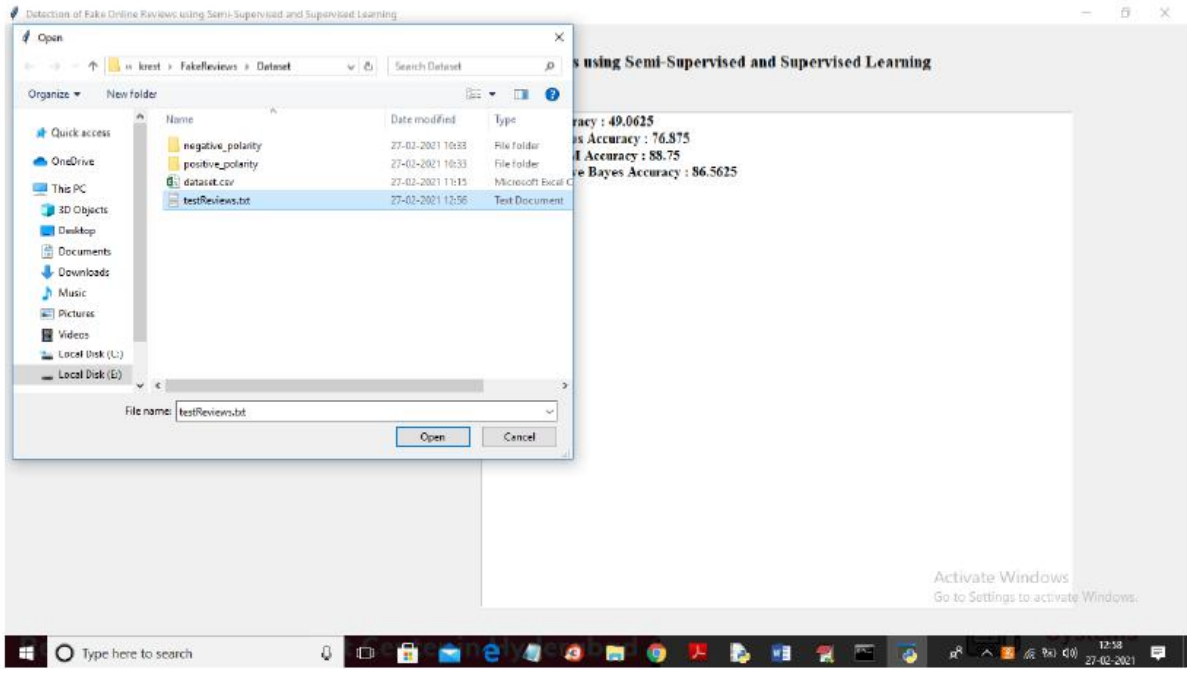


Fig.9 Bar Graph Representing Accuracy

In above screen x-axis represents algorithm name and y-axis represents accuracy of those algorithms and from above graph we can say supervised algorithms are better than EM. Now click on 'Upload Test Review & Predict Fake & Sentiments' button to upload test review and to get below output for each review.



ISSN: 2366-1313

Fig.10 Uploading Test Reviews

In above screen selecting and uploading 'testReviews' and then click on 'Open' button to get below result

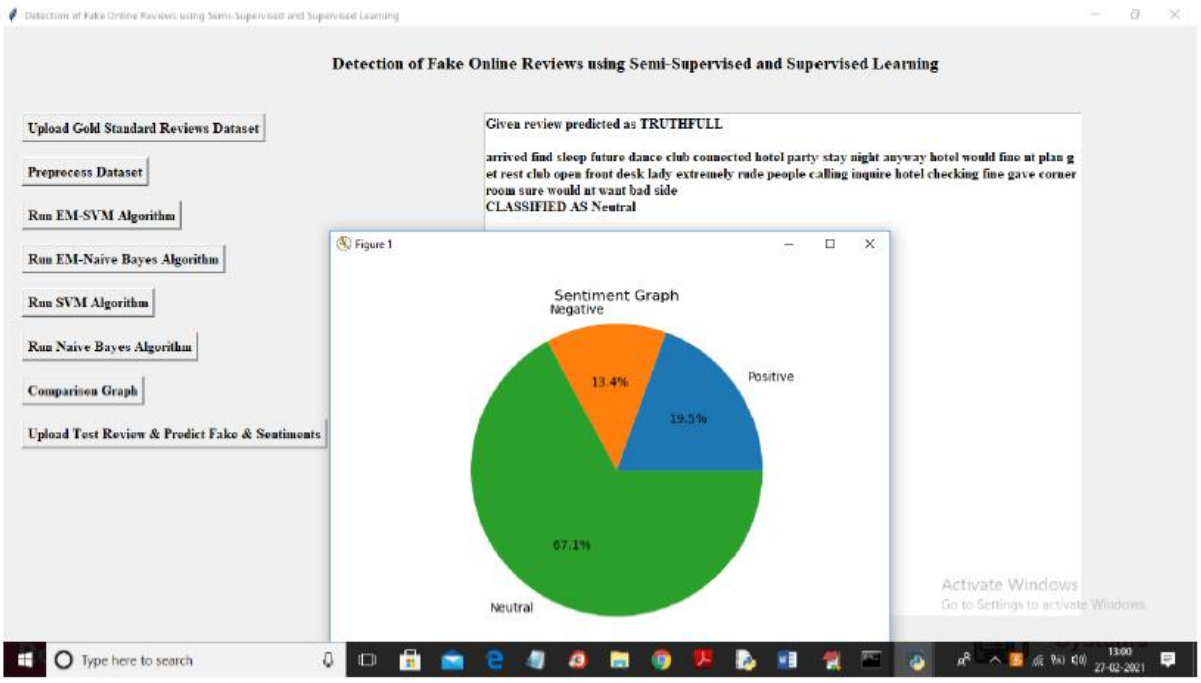


Fig.11 Pie Chart Representation of Test

In above screen we can see review detected as TRUTHFULL and its sentiment predicted as NEUTRAL.

V. CONCLUSION

We have shown several semi-supervised and supervised text mining techniques for detecting fake online reviews in this research. We have combined features from several research works to create a better feature set. Also, we have tried some other classifier that were not used on the previous work. Thus, we have been able to increase the accuracy of previous semi supervised techniques done by Jiten et al.. We have also found out that supervised Naive Bayes classifier gives the highest accuracy. This ensures that our dataset is labeled well as we know semi-supervised model works well when reliable labeling is not available. In our research work we have worked on just user reviews.

REFERENCES

[1] Chengai Sun, Qiaolin Du and Gang Tian, "Exploiting Product Related Review Features for Fake Review Detection,"

ISSN: 2366-1313

Mathematical Problems in Engineering, 2016.

[2] A. Heydari, M. A. Tavakoli, N. Salim, and Z. Heydari, ”Detection of review spam: a survey”, Expert Systems with Applications, vol. 42, no. 7, pp. 3634–3642, 2015.

[3] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, “Finding deceptive opinion spam by any stretch of the imagination,” in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT), vol. 1, pp. 309–319, Association for Computational Linguistics, Portland, Ore, USA, June 2011.

[4] J. W. Pennebaker, M. E. Francis, and R. J. Booth, ”Linguistic Inquiry and Word Count: Liwc,” vol. 71, 2001.

[5] S. Feng, R. Banerjee, and Y. Choi, “Syntactic stylometry for deception detection,” in Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers, Vol. 2, 2012.

[6] J. Li, M. Ott, C. Cardie, and E. Hovy, “Towards a general rule for identifying deceptive opinion spam,” in Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL), 2014.

[7] E. P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, “Detecting product review spammers using rating behaviors,” in Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM), 2010.

[8] J. K. Rout, A. Dalmia, and K.-K. R. Choo, “Revisiting semi-supervised learning for online deceptive review detection,” IEEE Access, Vol. 5, pp. 1319–1327, 2017.

[9] J. Karimpour, A. A. Noroozi, and S. Alizadeh, “Web spam detection by learning from small labeled samples,” International Journal of Computer Applications, vol. 50, no. 21, pp. 1–5, July 2012.

[10] W. Luo, F. Zhuang, X. Cheng, Q. H, Z. Shi, “Ratable aspects over sentiments: predicting ratings for unrated reviews,” IEEE International Conference on Data Mining (ICDM), 2014, pp. 380-389