Opinion Spam Detection using Ensemble Models

Project ID: 22MPA005

Final Review

Group Members

RA1811003030002 Dhawal Sarin

RA1811003030008 Ekansh Sharma

RA1811003030010 Simran Kathuria

Supervised By:

Mr. Nishant Anand

Assistant Professor

Department of Computer Science & Engineering

Faculty of Engineering & Technology

SRM Institute of Science & Technology

May 10, 2022

Review Comments

Comment

1) In our review-I, we were asked to add serial number to all the mentioned papers in the reference slides.

Comment

2) In our review-II, we were asked to add comparative table.

Table of Contents

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY (Deemed to be University u/s 3 of UGC Act, 1956) DELHI-NCR CAMPUS, GHAZIABAD (U.P)

Objective

In the continuous period of prospering web-based business and other web-based platforms, individuals prefer online purchasing items and services to save time. These internet-based buy choices are for the most part impacted by the audits/assessments of other people who already have experienced them. Malicious users utilize this experience sharing to advance or corrupt items/services for their unjust money related advantages, known as review spam. For the same problem statement, many different solutions have been proposed but we will ensemble different learning methods so that we could achieve the highest accuracy, to predict that the given review is deceptive or truthful.

Literature Survey

Author	Title	Summary
Yan Zhu,Zhuang Xu,Mushtaq Ahmad,Muqeet Ahmad	Enactment of Ensemble Learning for Review Spam Detection on Selected Features (2019) [6]	In this study, they investigated the performance and effectiveness of ensemble learning with feature selection techniques for review spam detection on real and semi-real-life datasets. They employ ChiSquared feature selection technique with ELM and getting precision 0.851, 0.820, and 0.774 for Yelp, M.Ott Pos-Pol, and M.Ott Neg-Pol datasets, respectively.
Ashish Salunkhe	Attention-based Bidirectional LSTM for Deceptive Opinion Spam Classification (2021) [9]	In this paper, they have used different machine learning models that can be used to train the final model. Different classifiers are implemented namely Naive Bayes Classifier, Linear Classifier: Logistic Regression, Support Vector Machine, Deep Neural Networks and compare their results with performance metrics and choose the final model for the classification.
Siyuan Zhao,Zhiwei Xu,Limin Liu,Mengjie Guo,Jing Yun	Towards Accurate Deceptive Opinions Detection based on Word Order-preserving CNN (2018) [10]	In this paper, the CNN in the deep learning model is used to identify the deceptive opinions. And getting accuracy of 84%.
Michela Fazzolari,Francesco Buccafurri,Gianluca Lax,Marinella Petrocchi	Improving Opinion Spam Detection by Cumulative Relative Frequency Distribution (2020) [1]	In this paper, they used several supervised machine-learning algorithms, namely Logistic Regression, Support Vector Machine, Decision Tree, Naive Bayes, K-Nearest Neighbors and to ensure higher reliability of results, they have applied a Stratified k-Fold Cross Validation approach.
Nitin Jindal and Bing Liu	Opinion Spam and Analysis (2008) [2]	In this paper, they are using logistic regression and getting accuracy of 78%.

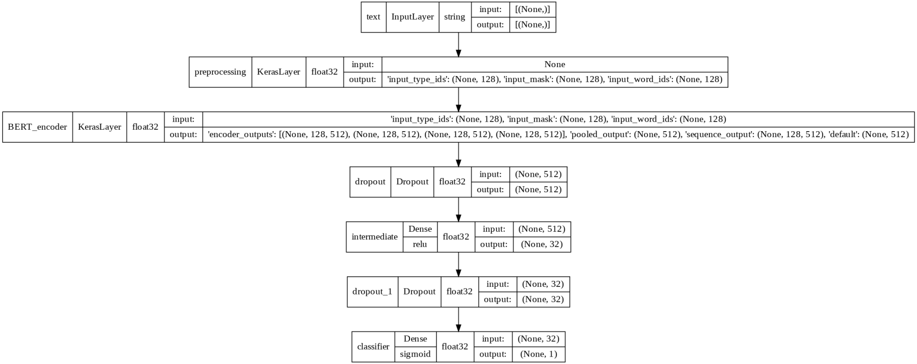
SRM INSTITUTE OF SCIENCE AND TECHNOLOGY (Deemed to be University u/s 3 of UGC Act, 1956) DELHI-NCR CAMPUS, GHAZIABAD (U.P)

Litrature Survey

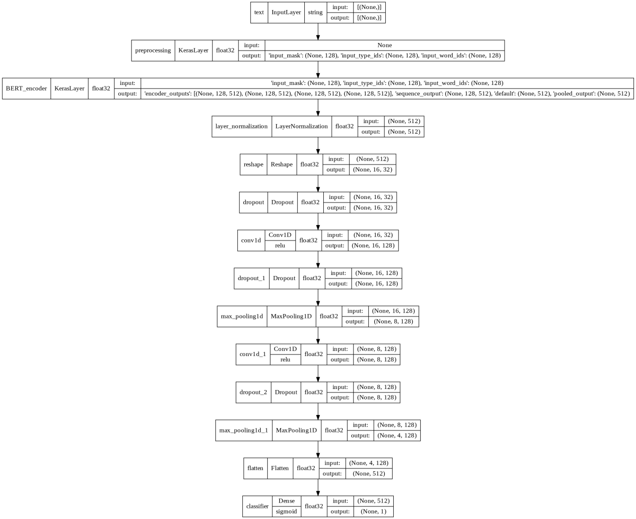
Nitin Jindal and Bing Liu	Review Spam Detection (2007) [4]	In this paper, they have used the statistical package R to perform logistic regression that gives the average AUC value of 78% by appling 10-fold cross validation on the data.
Fangtao Li, Minlie Huang, Yi Yang and Xiaoyan Zhu	Learning to Identify Review Spam [3]	In this paper, they first employ supervised learning methods and then a two-view semi-supervised methods to exploit the large amount of unlabeled data. The experiment results show that the two-view co-training algorithms can achieve better results than the single-view algorithm and getting accuracy of 63.1%.
Arjun Mukherjee, Vivek Venkataraman, Bing Liu, Natalie Glance	What Yelp Fake Review Filter Might Be Doing? (2013) [7]	This paper performed an in-depth investigation of the nature fake reviews in the commercial setting of Yelp.com. This study shows that although linguistic methods in (Ott et al., 2011; Feng et al., 2012b) reported very high (90%) detection accuracy on crowdsourced fake reviews,but it do not work well on real-life fake reviews. Behavioral features yielded a respectable 86% accuracy.
Naveed Hussain, Hamid Turab Mirza, Ghulam Rasool, Ibrar Hussain and Mohammad Kaleem	Spam Review Detection Techniques: A Systematic Literature Review (2019) [8]	This literature review presents an overall discussion about different feature extraction approaches from review datasets. Research gaps and future directions in the area of spam review detection are also presented.
Jagmeet Kaur and Dr. Munish Sabharwal	Spam Detection in Online Social Networks Using Feed Forward Neural Network (2018) [5]	In this paper they compares various algorithms with FFNN+ICA such as Random Tree, Random Forest, Naive Bayes. Their proposed model reaches an accuracy of 90%

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY (Deemed to be University u/s 3 of UGC Act, 1956) DELHI-NCR CAMPUS, GHAZIABAD (U.P)

Architectural Design of the Proposed System



Architectural Design of the Proposed System



SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
(Deemed to be University u/s 3 of UGC Act, 1956)
DELHI-NCR CAMPUS, GHAZIABAD (U.P)

Data-set Specification

Our corpus consists of truthful and deceptive hotel reviews of 20 Chicago hotels.

- 400 truthful positive reviews from TripAdvisor.
- 400 deceptive positive reviews from Mechanical Turk.
- 400 truthful negative reviews from Expedia, Hotels.com, Orbitz, Priceline, TripAdvisor and Yelp.
- 400 deceptive negative reviews from Mechanical Turk.

Acknowledgement
If you use any of this data in your work, Please direct questions to Myle Ott (myleott@cs.cornell.edu).

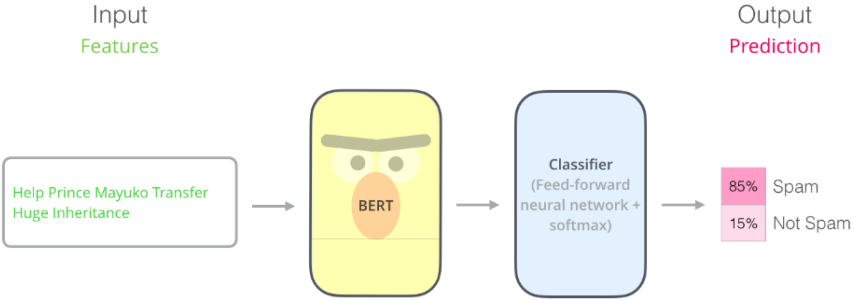


Figure: Workflow of our model.

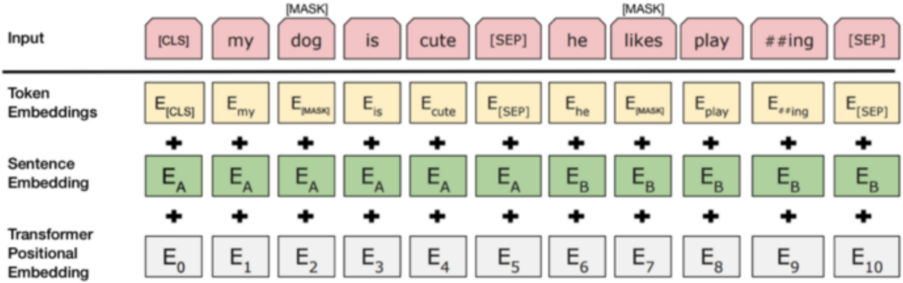
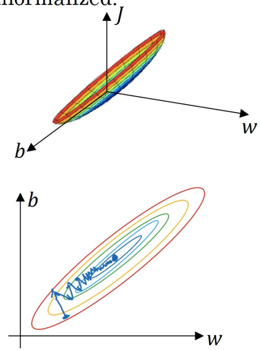


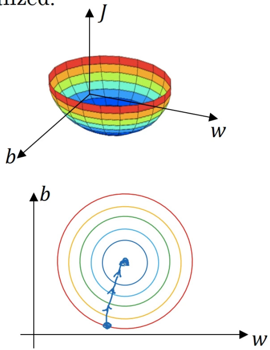
Figure: **Working of BERT.**

Why normalize inputs?

Unnormalized:



Normalized:



SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
(Deemed to be University u/s 3 of UGC Act, 1956)
DELHI-NCR CAMPUS, GHAZIABAD (U.P)

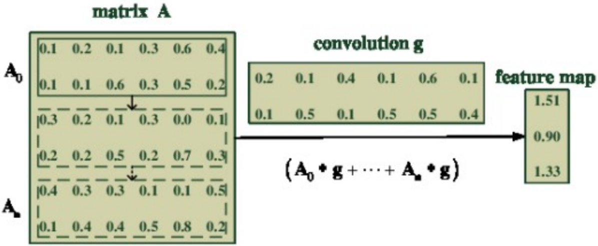


Figure: **Working of 1D Convolution**

The process of convolution in the CNN model deals with the data. The matrix A is multiplied by the corresponding elements of the convolution matrix g and summed up. During the process of convolution, we will get the feature map.

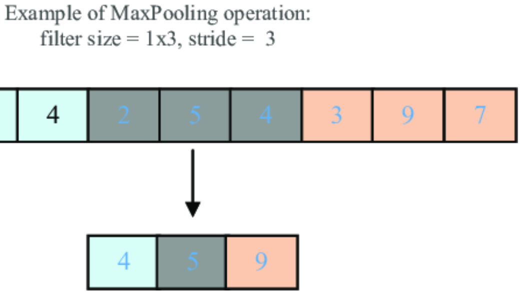


Figure: **Working of 1-D Maxpooling.**

Model Outcomes

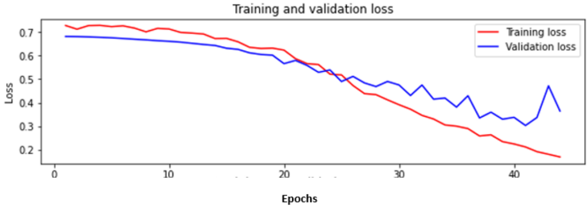


Figure: Loss Graph

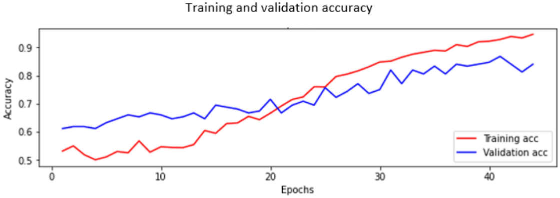


Figure: Accuracy Graph

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
(Deemed to be University u/s 3 of UGC Act, 1956)
DELHI-NCR CAMPUS, GHAZIABAD (U.P)

Model Outcomes

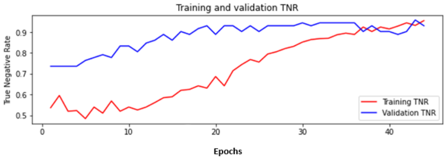


Figure: Graph of TNR

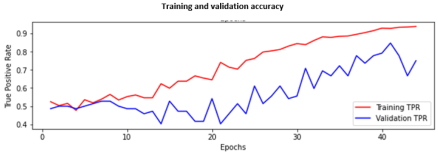


Figure: Graph of TPR.

- Deceptive Comment: 1
- Not Deceptive Comment: 0
- True Positive Rate: 0.94
- True Negative Rate: 0.88

Model Outcomes

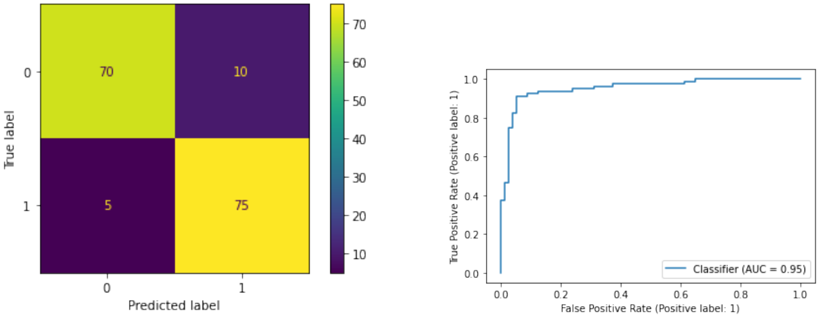
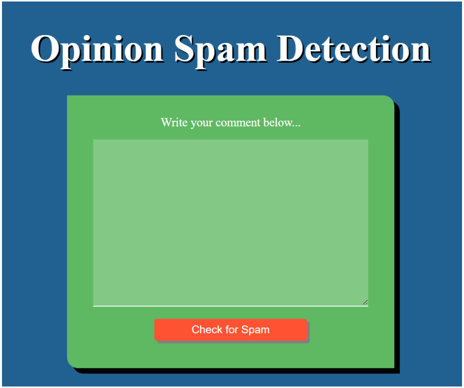


Figure: Confusion Metrics and ROC Curve

We have validation dataset with **160** entries. Out of which, **80 are deceptive (i.e. true label/output = 1) and 80 are not deceptive (i.e. true label/output = 0).**

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY (Deemed to be University u/s 3 of UGC Act, 1956) DELHI-NCR CAMPUS, GHAZIABAD (U.P)

GUI Results

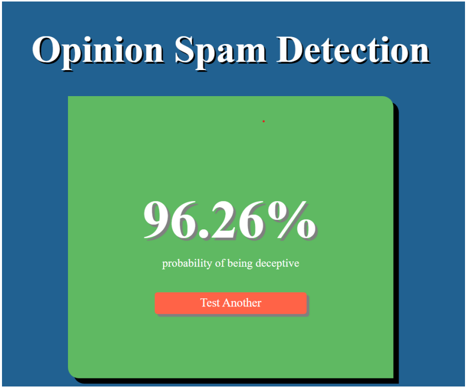


Opinion Spam Detection
Write your comment below...
Check for Spam

GUI Results

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
(Deemed to be University u/s 3 of UGC Act, 1956)
DELHI-NCR CAMPUS, GHAZIABAD (U.P)

GUI Results


Opinion Spam Detection
96.26%
probability of being deceptive
Test Another

Comparison Table

Author	Title	Accuracy	Our Accuracy
Yan Zhu,Zhuang Xu,Mushtaq Ahmad,Muqeet Ahmad	Enactment of Ensemble Learning for Review Spam Detection on Selected Features (2019) [6]	**Precision Score** 85.1% (Yelp dataset), 82.0%(M.Ott Pos-Pol dataset) and 77.4%(M.Ott Neg-Pol datasets)	**Precision Score** 88.2%.
Ashish Salunkhe	Attention-based Bidirectional LSTM for Deceptive Opinion Spam Classification (2021) [9]	**F1 Score** SVM:68.35%, . Logistic Regression: 86.45% Attention based Bidirectional Long Short Term :90.25% and Recurrent	**F1 Score** 91%
Siyuan Zhao,Zhiwei Xu,Limin Liu,Mengjie Guo,Jing Yun	Towards Accurate Deceptive Opinions Detection based on Word Order-preserving CNN (2018) [10]	**F1 Score** order-preserving-CNN (OPCNN): 84%.	**F1 Score** 91%.
Nitin Jindal and Bing Liu	Opinion Spam and Analysis (2008) [2]	**AUC Score** Logistic Regression: 78%	**AUC Score** 95%

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
(Deemed to be University u/s 3 of UGC Act, 1956)
DELHI-NCR CAMPUS, GHAZIABAD (U.P)

Comparison Table

Author	Title	Accuracy	Our Accuracy
Nitin Jindal and Bing Liu	Review Spam Detection (2007) [4]	**AUC Score** logistic regression that gives the average of 78% by appling 10-fold cross validation on the data.	**AUC Score** 95%.
Fangtao Li, Minlie Huang, Yi Yang and Xiaoyan Zhu	Learning to Identify Review Spam [3]	**F1 Score** The experiment results show that the two-view co-training algorithms can achieve better results than the single-view algorithm and getting accuracy of 63.1%.	**F1 Score** 91%
Arjun Mukherjee, Vivek Venkataraman, Bing Liu, Natalie Glance	What Yelp Fake Review Filter Might Be Doing? (2013) [7]	Behavioral features yielded a respectable 86% accuracy indicating that Yelp's filter might be using a behavioral based approach.	**Accuracy** 91%

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
(Deemed to be University u/s 3 of UGC Act, 1956)
DELHI-NCR CAMPUS, GHAZIABAD (U.P)

References I

M. Fazzolari, F. Buccafurri, G. Lax, and M. Petrocchi.
Improving opinion spam detection by cumulative relative frequency distribution.
arXiv preprint arXiv:2012.13905, 2020.

N. Hussain, H. Turab Mirza, G. Rasool, I. Hussain, and M. Kaleem.
Spam review detection techniques: A systematic literature review.
Applied Sciences, 9(5):987, 2019.

N. Jindal and B. Liu.
Review spam detection.
In *Proceedings of the 16th international conference on World Wide Web*, pages 1189–1190, 2007.

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
(Deemed to be University u/s 3 of UGC Act, 1956)
DELHI-NCR CAMPUS, GHAZIABAD (U.P)

N. Jindal and B. Liu.
Opinion spam and analysis.
In *Proceedings of the 2008 international conference on web search and data mining*, pages 219–230, 2008.

J. Kaur and M. Sabharwal.
Spam detection in online social networks using feed forward neural network.
In *RSRI conference on recent trends in science and engineering*, volume 2, pages 69–78, 2018.

F. Khurshid, Y. Zhu, X. Zhuang, M. Ahmad, and M. Ahmad.
Enactment of ensemble learning for review spam detection on selected features.
International Journal of Computational Intelligence Systems, 12(1):387, 2018.

References III

F. H. Li, M. Huang, Y. Yang, and X. Zhu.
Learning to identify review spam.
In *Twenty-second international joint conference on artificial intelligence*, 2011.

A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance.
What yelp fake review filter might be doing?
In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 7, 2013.

A. Salunkhe.
Attention-based bidirectional lstm for deceptive opinion spam classification.
arXiv preprint arXiv:2112.14789, 2021.

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY (Deemed to be University u/s 3 of UGC Act, 1956) DELHI-NCR CAMPUS, GHAZIABAD (U.P)

References IV

S. Zhao, Z. Xu, L. Liu, M. Guo, and J. Yun.
Towards accurate deceptive opinions detection based on word order-preserving cnn.
Mathematical Problems in Engineering, 2018, 2018.

SRM
INSTITUTE OF SCIENCE AND TECHNOLOGY
(Deemed to be University u/s 3 of UGC Act, 1956)
DELHI-NCR CAMPUS, GHAZIABAD (U.P)