

MSc. in Computing Practicum Approval Form

Section 1: Student Details

Project Title:	Fake review prediction using Neural Network with different activation functions
Student ID:	20210415 21263045
Student name:	Kashish Kishinchandani Aditi Bhat
Student email	kashish.kishinchandani2@mail.dcu.ie aditi.bhat2@mail.dcu.ie
Chosen major:	Data Analytics
Supervisor	Prof. Manoj Kesavulu
Date of Submission	28/01/2022

Section 2: About your Practicum

Please answer all questions below. Please pay special attention to the word counts in all cases.

What is the topic of your proposed practicum? (100 words)

Reviews help the consumers choose between products. However, sometimes vendors use bots to post reviews(bogus/fake) to increase the ratings and reach of the products. This undermines the product's actual reviews, thereby negatively affecting consumer satisfaction. Our research focuses on understanding how activation functions and word embeddings complement each other in detecting fake reviews in Neural Networks. After considering several combinations of word embeddings and activations, the one with the highest accuracy will be picked as the best. The focus is on the activation functions as they play a vital role in specifying how a node or nodes in a network layer change the weighted sum of the input into an output.

Please provide details of the papers you have read on this topic (details of 5 papers expected).

1. Khan, Hanif, et al. "Fake review classification using supervised machine learning." *International Conference on Pattern Recognition*. Springer, Cham, 2021.
2. P. Shetgaonkar, J. T. Rodrigues, S. Aswale, V. L. K. Gonsalves, J. C. Rodrigues and A. Naik, "Fake Review Detection Using Sentiment Analysis and Deep Learning," 2021 International

Conference on Technological Advancements and Innovations (ICTAI), 2021, pp. 140-145, doi: 10.1109/ICTAI53825.2021.9673375.

3. S. M. Anas and S. Kumari, "Opinion Mining based Fake Product review Monitoring and Removal System," 2021 6th International Conference on Inventive Computation Technologies (ICICT), 2021, pp. 985-988, doi: 10.1109/ICICT50816.2021.9358716.

4. I. Amin and M. Kumar Dubey, "An overview of soft computing techniques on Review Spam Detection," 2021 2nd International Conference on Intelligent Engineering and Management (ICIEM), 2021, pp. 91-96, doi: 10.1109/ICIEM51511.2021.9445280.

5. G. Shahariar, Swapnil Biswas, Faiza Omar, Faisal Shah and Samiha Hassan, "Spam Review Detection Using Deep Learning", 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)

How does your proposal relate to existing work on this topic described in these papers? (200 words)

The proposal focuses on predicting the legitimacy of the reviews from the dataset. The supervised learning technique is employed for classifying spam and genuine reviews starts by entering the input review, pre-processing the review, and finally classifying it as fake and genuine[1]. Different techniques are used to detect fake reviews such as LSTM, Bi-LSTM, GRU and more[2][3][5]. Using the research done in [4], we intend to use it as a framework for our research, we will be using Neural Networks to predict the legitimacy of the review. After taking into account several combinations of word embeddings and activation, the one with the highest accuracy will be presented as the proposed solution.

What are the research questions that you will attempt to answer? (200 words)

- How do individual activation functions combined with different word embedding techniques work in fake review prediction?
- What makes the difference in a specific combination of a word embedding and an activation function that increases accuracy compared to other combinations?

How will you explore these questions? (Please address the following points. Note that three or four sentences on each will suffice.)

- What software and programming environment will you use?

- Programming Language: Python
- Environment: Jupyter Notebooks, Jupyter Labs, AWS/GCP

- What coding/development will you do?

- The dataset will be acquired from an already existing dataset. The dataset will then be preprocessed using NLP techniques.
- Combinations of word embeddings and activation functions are the primary focus.
- Feature engineering techniques will be employed on this dataset, and Neural Networks will be employed to train the model.
- Lastly, the best combination will be evaluated to compare the accuracies of all combinations of word embeddings and activation functions.

- What data will be used for your investigations?

- The dataset will be acquired from:
 - Yelp reviews dataset - <https://www.yelp.com/dataset>
 - Kaggle e-commerce reviews dataset - <https://www.kaggle.com/furkangozukara/turkish-product-reviews>

- Is this data currently available, if not, where will it come from?

- The Kaggle datasets are readily available.
- The Yelp dataset has been requested and acquired.

- What experiments do you expect to run?

- We will perform preprocessing on the dataset using tokenization, stop word removal and other NLP processes. Additionally, we will employ techniques to reduce noise in the data to avoid faulty training.
- We will use word embeddings to convert the text data into numeric data as Neural Networks need a numeric input.
- Lastly, we will use Neural Network to train the data and evaluate the accuracies of different activation functions in complement with the word embeddings.

- What output do you expect to gather?

- Activation function and word embedding combinations are employed, and the results are predicted. The predicted results are verified against the actual results, and accuracy is calculated for each. The combination with the highest accuracy will be selected as the best combination for our chosen Neural Network model.