Phase 2 - Report

Arnav Gupta, 200968030, Batch 3

Literature Review:

S.No	Title	Author	Date of Publishing	Publisher	Models implemented	Pros/Cons
1	FAKEDETECTOR: Effective Fake News Detection with Deep Diffusive Neural Network	Jiawei Zhang, Bowen Dong, Philip S. Yu	10 Aug 2019	https://arxiv.org/	1) Deep Diffusive Unit Model 2) RNN 3) SVM	Pros: To learn the latent representations of the textual input To model the correlation among news articles, creators and subjects Cons: Less effective on noisier datasets with overlapping classes
2	Fake news detection in social media	Kelly Stahl	15 May 2018	https://www.csustan.edu/	1) Naïve Bayes Classifier 2) SVM	Pros: To compute the likelihood of a certain outcome by using past knowledge of it Fast and a highly accessible technique Cons: Training time with SVMs can be high Less effective on noisier datasets with overlapping classes

3	Fake News Detection Using Machine Learning approaches: A systematic Review	Syed Ishfaq Manzoor, Dr Jimmy Singla, Nikita	23-25 April 2019	IEEE Xplore Part Number: CFP19J32-ART	1) Naïve Bayes 2) Decision trees 3) SVM 4) Neural Networks 5) Random Forest 6) XG Boost.	Pros: Content cues which include lexical and semantic level of analysis were implemented by the authors Classification of falsified and fabricated news items Cons: Everchanging characteristics and features of fake news in social media networks is posing a challenge in categorization of fake news
4	Fake News detection Using Machine Learning	Jasmine Shaikh, Rupali Patil	19 March 2021	https://ieeexplore.ieee.org/xpl/conhome/9378700/proceeding	1) SVM 2) Passive Aggressive Classifier (PAC)	Pros: SVM chooses best hyperplane (straight line) considering maximum margin between support vectors for nonlinear data Passive Aggressive Classifier is easy to use and work fast Cons: PAC does not provide high accuracy The algorithm remains passive for a correct classification outcome, and it turns aggressive if

						it is an incorrect classification, updating and adjusting
5	Fake News Detection Using Machine Learning Ensemble Methods	M. Irfan Uddin	17 Oct 2020	https://www.hindawi.com/journals/complexity/2020/8885861/	1) Logistic Regression 2) KNN 3) Ensemble Learners	Pros: Logistic Regression provides the intuitive equation to classify problems into binary or multiple classes With Ensemble Learners, a number of models can be trained on different set of parameters to create multiple decision boundaries on randomly chosen data points as training data Cons: In KNN, accuracy depends on the quality of the data. Amount of data should also be large.
6	Fake news detection based on news content and social contexts: a transformer-based approach	Shaina Raza, Chen Ding	30 January 2022	https://link.springer.com/article/10.1007/s41060-021-00302-z	Bidirectional and AutoRegressive Transformer (BART) model	Pros: BART model combines the unique features (bidirectional and autoregressive) of both text generation and temporal modelling, which we require to meet our goals

						Cons: Limitation on the input sequence length
7	Fake News Detection on the Web: An LSTM-based Approach	Piyush Vyas, Jun Liu, Omar El-Gayar	August 2021	https://aisel.aisnet.org/	Long Short Term Memory (LSTM)	Pros: Overcomes the long-term dependency problem Cons: Prone to overfitting
8	Fake News Detection: A long way to go	Sunidhi Sharma, Prof. Dilip Kumar Sharma	November 2019	https://www.researchgate.net/publication/339975451_Fake_News_Detection_A_long_way_to_go	Recurrent Neural Networks	Pros: RNN remembers each and every information through time Cons: Computation is slow
9	Analysis of Classifiers for Fake News Detection	Vasu Agarwala, H.Parveen Sultanaa, SrijanMal Hotraa, AmitrajitS Arkar	2019	https://pdf.sciencedirectassets.com/ www.sciencedirect.com	1) Logistic Regression 2) Linear SVM	Pros: Easier and faster binary classifications Cons: Data is erratic and this means that any type of prediction model can have anomalies and can make mistakes
10	Localization of Fake News Detection via Multitask Transfer Learning	Jan Christian Blaise Cruz,	16 May 2020	https://aclanthology.org/2020.lrec-1.316.pdf	1) Siamese Neural Network 2) Deep Bidirectional Representations (BERT)	Pros: Siamese network could then be trained to differentiate between classes in order to

		Julianne Agatha Tan, Charibeth Cheng				perform classification BERT, allows the model to compute weighted importance for each token in a sequence, effectively pinpointing context reference Generative Pre-Trained Transformers (GPT), uses encoders decoders

Basic Conclusions Derived:

1) SVM is commonly used in fake news detection because it's a good text classifier, combined with other python libraries would give us good accuracy and output.

2) Logistic Regression and SVM give good accuracy for the fake/real classification task.

3) LSTM, long-term memory makes it powerful to do forecasting or detecting.