CS550/DSL501: Machine Learning (2023–24–M)
Project Report - Phase 2

Project Name: Fake News Detector

Group Members:

1) **Anil Kumar** **12240150**
2) **Harshit Kumawat** **12240660**
3) **Nikhil Babu** **12241280**
4) **Govind More** **12240610**
5) **Rajana Venkata Mohit** **12241440**

GitHub Repository:
https://github.com/anilkrmandal/FakeNewsDetecctor.git

1 Individual Contributions

- **Anil Kumar:** Contributed to embedding techniques and model development. Initially, he worked on embedding the Wikipedia data using Word2Vec, but later shifted to FastText for improved performance. Along with Harshit, Anil was also responsible for implementing the CNN-LSTM model, which became the final chosen architecture for the Fake News Detector.
- **Harshit Kumawat:** Collaborated with Govind on training an unsupervised FastText word embedding model and vectorizing the entire WelFake dataset. Additionally, Harshit worked with Anil to implement the CNN-LSTM model, ensuring the model could effectively classify news articles as fake or real.
- **Nikhil Babu:** Initially worked on developing a UI interface for the project, but the team later decided to drop the UI, as they determined that it was unnecessary for the scope of the project.
- **Govind More:** Partnered with Harshit in training the FastText word embedding model and managing the vectorization of the WelFake dataset.Contibuted in CNN-LSTM in improving accuracy.
- **Rajana Venkata Mohit:** Initially implemented logistic regression, random forest classifiers and SVM. However, as the team shifted to using CNN-LSTM, his earlier work was replaced to align with the new direction of the project.

2 Tasks and Milestones Achieved

- Data preprocessing and extraction through unsupervised FastText completed.
- CNN-LSTM model applied successfully.
- Achieved an accuracy of 89
- Frontend was initially designed but later dropped.
- GitHub repository set up with contribution guidelines.
- Regular team meetings held to discuss progress and challenges.

3 Dataset Link

The dataset used for this project is the WelFake dataset, which can be accessed at the following link:
https://zenodo.org/records/4561253

4 Pending Tasks for Final Phase

- Implement supervised FastText embedding techniques for further improvement.
- Incorporate approaches related to Explainable AI to provide insights into the model's decision-making process.
- Focus on improving accuracy through additional feature engineering and fine-tuning of the CNN-LSTM model.
- Finalize documentation and prepare for the project presentation.
- Test the model on real-world data to evaluate its effectiveness.