Project Title: Spam Detection

Summary:

This project aims to develop a spam detection system using natural language processing (NLP) to ensure the safety of online users. The dataset comes from a Kaggle dataset that combines 4 different datasets, including 5,574 unique text messages with 13% of messages being spam. The project team plans to apply LSTM, BERT, and RNN models to filter out spam messages from legitimate ones. The models will be able to account for the entire text to detect spam, consider the entire text with self-attention to learn about the context of words, and use inputs and hidden layers to provide an output. The project's outcomes have the potential to be beneficial to individuals and organisations alike by improving the accuracy of spam detection, increasing efficiency, reducing frustration, and providing user safety. The initial LSTM model has been successful in detecting spam with an accuracy of 99.89%, and the team is looking forward to see how the other models will perform.

Grading (Highly innovative/ Neutral/ Not innovative):

I would classify this project as highly innovative because it addresses a common problem faced by millions of users worldwide and uses advanced NLP techniques and machine learning models to develop a spam detection system. The project combines LSTM, BERT, and RNN models to provide a more comprehensive analysis of text data, which can lead to higher accuracy in spam detection. Moreover, the project team's use of a diverse dataset from various sources can help improve the generalisation of the model.

The project is also innovative because it takes a proactive approach to tackle the issue of spam, which has been a persistent problem for decades. By leveraging the power of advanced NLP models, the team is working to protect users from unwanted messages, increase efficiency, and reduce the risk of phishing and other threats. Overall, the project has the potential to have a significant impact on user safety and efficiency in online communication, making it a highly innovative project.

Suggestion/Feedback:

The project sounds very promising, as it addresses an important problem that affects millions of online users worldwide. The combination of advanced NLP techniques and machine learning models to develop a spam detection system is an exciting development. The use of LSTM, BERT, and RNN models can provide a more comprehensive analysis of text data, which can

lead to higher accuracy in spam detection. The team's decision to use a diverse dataset from various sources can help improve the generalisation of the model.

However, it would be helpful and beneficial to have more information on how the project team plans to validate the accuracy of the models and how they plan to implement the final system.

Overall, the project is a great initiative, and it has the potential to have a significant impact on user safety and efficiency in online communication.