**Project Report**

**Fake News Detection**

**Introduction:**

Fake news is a phenomenon that has been prevalent in recent times, and it is a significant issue as it has the potential to sway people's opinions and beliefs. Therefore, detecting fake news is crucial to mitigate its negative impact. In this project, we aim to detect fake news using machine learning algorithms.

**Data:**

The dataset used for this project is obtained from a CSV file that contains news articles with their labels (fake or real). The dataset has a total of 6,335 articles, each having a unique ID, title, text, and label. After loading the data into a pandas dataframe, we dropped the unnecessary columns (ID and title) and removed any missing values.

**Data Cleaning:**

To clean the data, we removed any URLs, mentions, and hashtags in the text column using regular expression patterns. We also verified that there were no null values in the dataset.

**Data Preprocessing:**

Next, we used the TfidfVectorizer from sklearn to convert the text into a matrix of features that can be used for training our model. We split the data into training and testing sets using the train\_test\_split method.

**Model:**

We used two models for fake news detection: the RandomForestClassifier and the PassiveAggressiveClassifier. The RandomForestClassifier is an ensemble learning method that fits multiple decision tree classifiers on different sub-samples of the dataset and uses the average of the predictions as the final output. The PassiveAggressiveClassifier is a linear model for large-scale learning that works by updating the weights of the model whenever it makes a misclassification. We trained both models and evaluated their performance using various metrics such as accuracy, confusion matrix, and classification report.

**Results:**

The RandomForestClassifier achieved an accuracy of 91.62%, while the PassiveAggressiveClassifier achieved an accuracy of 92.41%. Therefore, both models showed promising results in detecting fake news. The confusion matrix and classification report also revealed that the models performed well in terms of precision, recall, and f1-score.

**Conclusion:**

In this project, we successfully implemented two machine learning models to detect fake news with high accuracy. As fake news continues to be a significant issue in today's society, these models can be utilized to identify and filter out fake news articles, thereby reducing its negative impact.