**Approach to Solve Fake News Detector**

**Introduction**

​Develop a model to detect if a given news article is a fake one or legitimate(real) one.

**Approach**

Following are the steps taken

1. Exploring the data
2. Cleaning the data
3. Modelling with two ML models

**Exploring the data**

Things done in this

1. Found the number of null values. While doing so, it was observed that there are few null values and it can be ignored.
2. After removing the null values there was slight imbalance in the classes and was ignored as of now.
3. Detected count of frequency of words in both real and fake new with the word cloud.

**Cleaning the data**

Things done in this

1. Removed the punctuations from the text data.
2. Removed the stop words from the text.

**Modelling**

Things done in this

1. Tried 3 different ways to develop the model.
   1. Used only text data for modelling
   2. Used only title data for modelling
   3. Used concatenation of text and title data
2. Found that concatenation was giving better result
3. Developed 2 models using Logistic Regression and Decision tree.
4. Logistic Regression gave an accuracy of 95.57% with Precision as 0.96 and Recall as 0.93
5. Decision Tree gave an accuracy of 94.5% with Precision as 0.92 and Recall as 0.95

**Conclusion**

Concatenation gave the better result so used that as the final data for Modelling

**Future Scope**

Following things can be done

1. Use ngrams with 2- or 3-word phrases
2. Can try Deep Learning Embedding Models