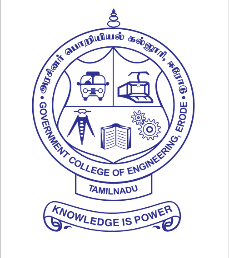
**GOVERNMENT COLLEGE OF ENGINEERING-ERODE**

**NAAN MUDHAVAN IBM – PROJECT**

**FAKE NEWS DEDECTION USING NLP**

**PROJECT IDEA**

**ELECTRONICS AND COMMUNICATION ENGINEERING**

**MENTOR**

**Dr.M.POONGOTHAI**

**IT DEPARTMENT**

**TEAM MEMBERS**

**NAME NM-ID**

**1)THANUJ KUMAR.R au731121106050**

**2)SARULATHA.M au731121106041**

**3)KARTHIKA.S au731121106022**

**4)PREETHIGA.A au731121106037**

**5)DIVYADHARSHINI .S au731121106012**

**Detecting fake news requires a critical approach and the application of various strategies to assess the credibility of the information**

**ALGORITHM**

1.Import the dataset

2. Explore the data to figure out what they look like

3. Pre-process the data

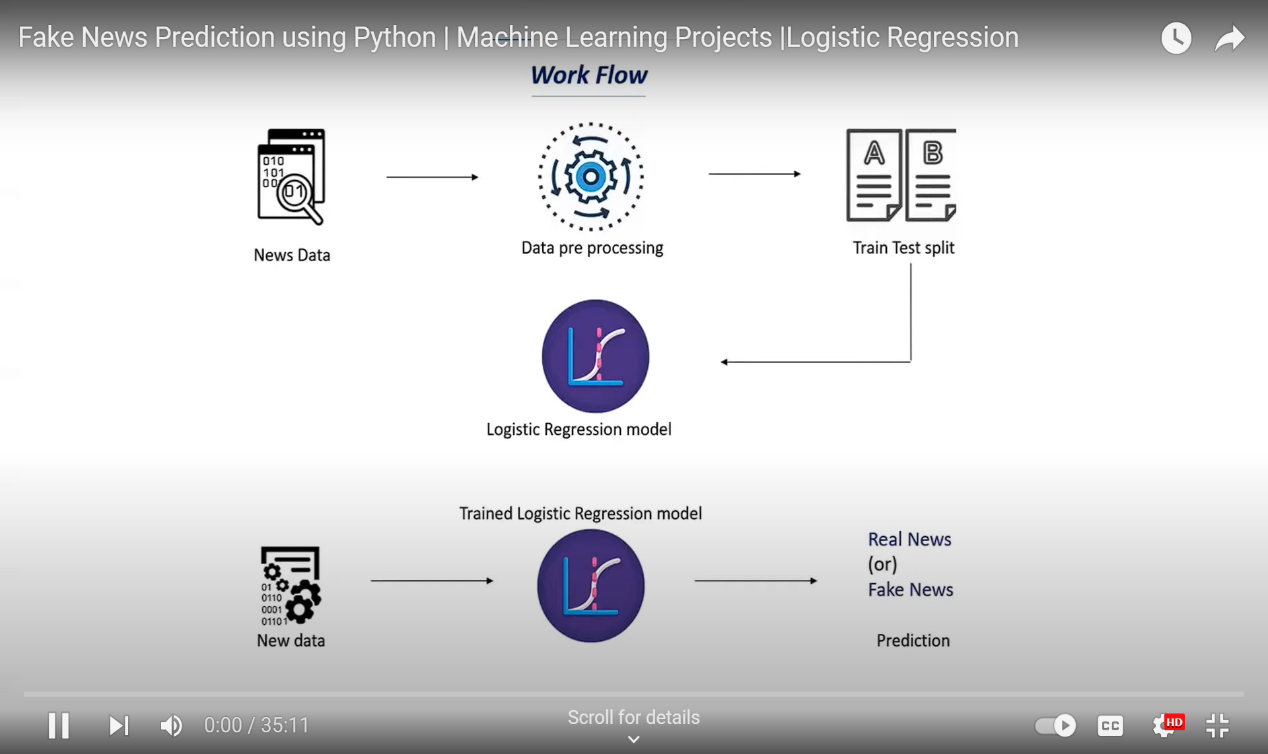
4. Split the data into attributes and labels

5. Divide the data into training and testing sets

6. Train the logistic regression

7. Make some predictions

8. Evaluate the results of the algorithm



**EXPLANATION**

**1.Data Preprocessing:**

Start by collecting and preprocessing your dataset of news articles. Clean the text data by removing stopwords, punctuation, and converting words to lowercase.

**2. Feature Extraction:**

Transform the cleaned text data into numerical features. One common approach is to use TF-IDF (Term Frequency- Inverse Document Frequency) to represent the text. This technique assigns a numerical value to each word based on its importance in the document and across the entire dataset.

**3.Label Encoding:**

Assign labels to your data, where "1" represents fake news and "O" represents real news.

**4.Splitting Data:**

Split your dataset into a training set and a testing set to train and evaluate your model.

**5. Logistic Regression Model:**

Build a Logistic Regression model. In Python, you can use libraries like Scikit-Learn for this purpose. Train the model on the training data.

**6. Evaluation:**

Evaluate the model's performance on the testing set using metrics like accuracy, precision, recall, and F1-score. These metrics will help you understand how well your model is performing in detecting fake news.

**7. Fine-Tuning:**

You can experiment with different hyperparameters of the Logistic Regression model, such as regularization strength, to optimize its performance.

**8.Deployment:**

Once satisfied with the model's performance, you can deploy it in a real-world application to classify news articles as real or fake.

**TEXT CLASSIFICATION MODEL ACCURACY CHART:**

**LOGISTIC REGRESSION**

* Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value.
* It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.
* Logistic Regression is much similar to the Linear Regression except that how they are used. Linear Regression is used for solving Regression problems, whereas Logistic regression is used for solving the classification problems.
* In Logistic regression, instead of fitting a regression line, we fit an “S” shaped logistic function, which predicts two maximum values (0 or 1).
* The curve from the logistic function indicates the likelihood of something such as whether the cells are cancerous or not, a mouse is obese or not based on its weight, etc.
* Logistic Regression is a significant machine learning algorithm because it has the ability to provide probabilities and classify new data using continuous and discrete datasets.
* Logistic Regression can be used to classify the observations using different types of data and can easily determine the most effective variables used for the classification.

**CONCLUSION**

This project works based on Logistics regression because it has high accuracy, In Logistic regression machine learning model was used and a fake news classification model was built taking the vector generator from the TF-IDF vectorizer as input for training and evaluation.

**“Remember, critical thinking is the best tools against fake news. Always be safe and verify information from multiple reliable sources before believing and sharing it.”**