FAKE NEWS DETECTION USING PYTHON AND MACHINE LEARNING

Model:

Logistic Regression:

Logistic regression is a supervised learning algorithm used in fake news detection. It learns a binary classification model to separate real and fake news based on given features. The algorithm uses a logistic function to map the input features to probabilities, indicating the likelihood of an instance being fake news. A decision boundary is set to classify instances into real or fake news. Logistic regression is interpretable and can provide insights into feature importance.

Intel Extension for Scikit-Learn is used in Logistic Regression to enhance the performance.

```python
from sklearnex import patch_sklearn
patch_sklearn()
```

The efficiency of the scikit learn algorithms is optimized by the scikit-learn-Intelex addition, which is added to scikit-learn by executing 'patch_sklearn()'

```
'Intel Extension for scikit learn time: 1.90s'
```

The training time for Logistic Regression using the Intel Extension for Scikit-Learn is around 1. seconds.

```python
from sklearnex import unpatch_sklearn
unpatch_sklearn()
```

The 'unpatch_sklearn()' function is used to remove the Intel extension from Scikit-Learn and return to the normal implementation. The two figs show that the Intel extension for Scikit-Learn consumes less time than the original Scikit-Learn.

```
'Original scikit learn time: 2.69s'
```
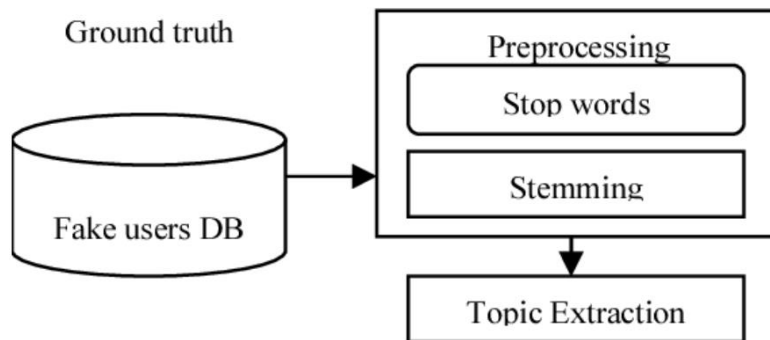
Random Forest Classifier:

Ensemble learning involves combining the predictions of multiple individual models to make final predictions. The Random Forest Classifier utilizes the concept of ensemble learning by combining multiple decision trees. A random forest is a collection of decision trees, where each tree is trained on a random subset of the trained data and features. The randomness introduces during training helps to reduce overfitting and imporve generalization. Intel Extension for Scikit-learn is also used in Random Forest Classifier like Logistic Regression. At the first step, Intel extension is used to patch the scikit learn for enhancement of performance.

```
'Intel Extension for scikit learn time: 12.93s '

'Original scikit learn time: 12.97s '
```
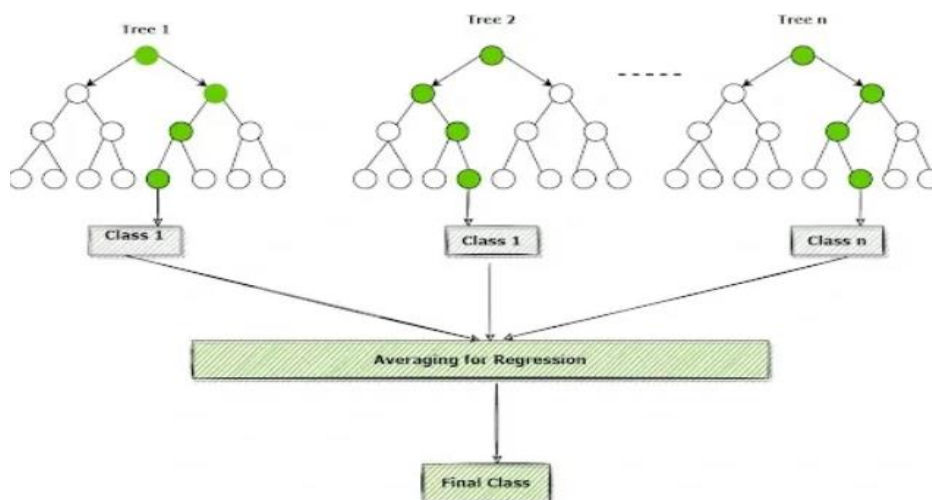
Decision Tree Classifier:

The decision tree algorithm will learn to make splits in the data based on the features and labels, creating a tree-like structure that represents decision rules for classifying news articles as fake or real.

Gradient Boosting Classifier:

This algorithm will sequentially build an ensemble of weak decision trees, where each subsequent tree corrects the mistakes made by previous tree. This process creates a strong predictive model.