1. Overview of Problem:

This project is aimed at analyzing the effectiveness and performance of different machine learning algorithms in its ability to classify twitter accounts as bots or humans. In this project, we apply XGBoost, Logistic Regression, SVM, Random Forest, and LSTM neural networks to identify tweets across social media and compare their accuracies. Using these trained models, we build a classifier for Twitter content that predicts the likelihood a given tweet was created by a troll bot in real time.

1. Preprocessing Data:
2. Check shape, data type and Missing values.
3. Fill the Missing values.
4. We are using English language tweets so separate the dataset accordingly.
5. EDA
6. Check for total no of English tweets, plot for different categories like region,type,data,etc.
7. NLP:
8. Work on bag of words, tokenization.
9. Identify different categories of words like most and least frequently used, positive and negative words, words used by trolls and humans.
10. Training the Classification Model

Train using different algorithms & Test the model on test set

1. Logistic Regression
2. Decision Tree
3. Random Forest
4. XGBoost Classifier
5. SVM
6. LSTM Neural Networks
7. Closing thoughts / Room for Improvement
8. Deployment & Inferencing

Deployment using Flask / Django / heroku