Group : 18

CS 513 - B KDD Project Proposal : Kaggle Bot Account

Team Members:

- Shrey Shah
- Brinda Nathvani
- Namra Patel
- Pratiksha Naik
- Saumil Trivedi

Problem Statement:

The objective of this project is to predict the Kaggle Bot Accounts dataset to predict that the user is bot or human.

Dataset:

The "Kaggle Bot Account Detection" dataset is a collection of information on user accounts on the Kaggle platform. Kaggle is a popular data science community where users can participate in data science challenges, share their work, and collaborate with other data scientists.

The dataset contains 1,321,188 user accounts, labeled as either "bot" or "human." The data includes features such as the user's name, profile picture, number of followers, number of notebooks created, and number of scripts created. The dataset also includes text features such as the user's bio and description.

The goal of the dataset is to train a machine learning model to accurately predict whether a given account is a bot or a human based on the available features. The dataset can be used for research in detecting and preventing bot activity on online platforms, which is a common problem in many online communities.

Source: https://www.kaggle.com/datasets/shriyashjagtap/kaggle-bot-account-detection

Implementation Strategy and Algorithms Used:

1. EDA on the dataset
2. Data Normalisation or Standardisation
3. Hot encoding
4. K- nearest neighbor
5. Naive Bayes
6. Xgboost
7. SVM
8. Decision Trees
9. Random Forest
10. NN (tentative)

Model Metrics and Evaluation:

1. AUC-ROC
2. Confusion Matrix
3. F1 Score
4. Recall
5. Precision
6. Log Loss