ABSTRACT

Spam is a hot topic for decades. After all the tech advancements still all of us stumble upon spam messages then and now. Spam has become more realistic where a normal people could not distinguish whether it is real or not. Random job offers appearing on your WhatsApp messenger from unknown source, Spam messages on every comment section of almost all the social medias. Occasionally, individuals may receive spam messages from their friends. Tech giants are spending millions to keep these spammers from their application. The project titled “ONLINE CHATTING WEBSITE WITH SPAM MESSAGE DETECTION” is aimed to provide a spam free chatting system to users. To detect and avoid spam messages we can use natural language processing and machine learning. In this system we developed a machine learning model using TF-IDF Vectorizer algorithm and Decision Tree Classifier to predict a SMS is a spam or non-spam(ham) and it was discovered that the model out performs existing models. The TF-IDF algorithm is a common technique in natural language processing (NLP) for text data preprocessing and feature extraction. This algorithm is used for feature extraction in this system. A decision tree classifier is a type of supervised learning algorithm that is commonly used in machine learning for classification tasks. Decision trees can be trained quickly and can handle large datasets with many features. This is important in the context of SMS spam detection, where there may be a large number of messages to classify and many features to consider. This model is used to predict the spam messages in this online chatting website. In this website user can register and login if his request of registration is accepted by the admin. The user can send friend request to other users and can send SMS to their friends. The system evaluates messages and predicts whether they are spam or ham.

Frame work : Django

Database : MySQL