



Text Classifier

30.6.2021

Tarun.S

Class 11

Vivekam Sr. Sec. School

Machine Learning Project

Overview

Built to classify input text into categories such as Educational and Non-Educational text which can be later on used to categorize websites, YouTube videos into the former categories.

Goals

1. To convert the text given into Educational and Non-Educational content using Machine Learning.
2. This can be later on extended so that it can also classify data of websites, blogs or even classify the videos available on platforms such as YouTube, which makes it easier for students to study without distractions.

Specifications

1. The most important concept used is Machine learning which is done with 'nltk' library which is used in all kinds of text preprocessing steps.
2. Along with which libraries like 'pandas' are also used. Bag of words and vectorization is done to the data in order to use it to train the model.
3. Naive Bayes classifier is taken advantage of to train the model.
4. The dataset used to train the model has been scraped from various websites and also downloaded from legitimate open-source websites.

Milestones

1. If utilized and developed properly this could be a huge advantage for every student across the world who wants to study without any distractions.
2. With current dataset the model is 99% accurate.
3. This project would turn out to be a revolution in the Educational industry.