IMDB Dataset of 50k Movie Reviews

Introduction

- Over the last two decades, social media's popularity has increased tremendously. Everybody uses social media in their everyday life to express their thoughts and opinions about a range of subjects. Many companies process the data and utilize it to generate pertinent information that they may later use for business goals. There is a considerable quantity of text-based data on movie reviews that the public provides on websites like IMDB, Amazon, Rotten Tomatoes, and others, and manually analyzing that data is a challenging job. To tackle this problem and make the process simple and automated, sentiment analysis is used. The Sentimental analysis is a subfield of Natural Language Processing and AI, used for extracting meaningful information from the textual data which we can utilize further to improve business in all way.
- In this project, I have taken dataset of IMDB movie reviews to predict how the audience have rated the movies and predict the movies that have a positive or negative review. I will create a model which I will used for extracting information using Sentiment Analysis from the data and predict which classifier is best suitable for this dataset by looking at the accuracy. The dataset is collected from Kaggle and it has 50k movie reviews.

Dataset

- I find this dataset **"IMDB Dataset of 50k Movie Reviews"** from Kaggle dataset.
- **Dataset link:** https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews?select=IMDB+Dataset.csv

Problem Statement

- In our scenario, to extract the positive and negative aspects of a certain product, I'll perform **sentiment analysis** and identifies best classifier for this dataset by looking at the accuracy.

Framework: -

In this project, I have used Sentiment Analyses technique. The proposed framework for models includes following steps:

- Data Cleaning, Text Stemming, Removing Stop words and special characters, Feature Extraction, Modelling, Accuracy of the model

Result

- The major goal of this project is to develop a sentiment analysis model that will benefit in a better comprehension of movie reviews and enable for comparison of the results of different classifiers depending on the level of accuracy in each model.