**Sentiment Analysis of IMDB Movie Reviews**

**Brief on the project :-**

The project is about identifying sentiment of the audience based on the movie reviews.The project is categorised of type Natural language processing is a subfield of artificial intelligence that deals with understanding and processing human language. In light of new advancements in machine learning.In this project we aim to use Sentiment Analysis on a set of movie reviews given by reviewers and try to understand what their overall reaction to the movie was,i.e.if they liked the movie or they hated it.We aim to utilise the relationships of the words in the review to predict the overall polarity of the review.

**Deliverables of the project :-**

* **Data Preprocessing :-** In this section we aim to do some operations on the dataset before training the model on it,processes like :
  1. Loading the dataset
  2. Data cleaning : Remove HTML tags
  3. Data cleaning : Remove special characters
  4. Data cleaning : Convert everything to lowercase
  5. Data cleaning : Remove stopwords
  6. Data cleaning : Stemming.
  7. Data cleaning : Removing emails
  8. Data cleaning : Remove url
  9. Data cleaning : Remove special characters.
* **Converting features to TFIDF vector form.**
* **Model Creation :-** The dataset is ready for training so we create Linear SVM model using scikit learn and then fit it to the data.
* **Model Evaluation :-** Finally we evaluate the model by getting accuracy, classification report

**Resources:-**

1. Data set source - https://github.com/laxmimerit/IMDB-Movie-Reviews-Large-Dataset-50k/blob/master/train.xlsx
2. Software - Google colab
3. References - <https://www.youtube.com/watch?v=glHWiQRgLn4>

**Individual Details :-** Nilesh Malekar,[nmalekar@gmail.com](mailto:nmalekar@gmail.com),Cell : 9325125588