**Sentiment Analysis on Movie Reviews using SVM**

**Name:** KHANSA AHMAD QURESHI

**Date:** 16th April 2025

**Dataset Used:** IMDB Dataset.csv from Kaggle

**Objective**

This project focuses on building a sentiment analysis model that classifies movie reviews as either positive or negative. We used the IMDb movie reviews dataset (IMDB Dataset.csv) and implemented a machine learning pipeline using Support Vector Machines (SVM) to predict the sentiment of a given review.

# Steps Taken:

# Tools and Methods

We used the following tools and methods:

* Python (Colab for model training, Streamlit for UI)
* SVM classifier from Scikit-learn
* TF-IDF vectorization for text feature extraction
* Streamlit for building a user-friendly web app
* Hugging Face Spaces for deployment

# Problems Faced

**Several challenges were encountered during the development process:**

* Locating and extracting the IMDb dataset from a ZIP file in Google Colab
* Handling text preprocessing such as lowercasing, stopword removal, and tokenization
* Understanding model performance evaluation through accuracy, F1-score, and confusion matrix and ROC\_AUC

**Saving and loading trained models for use in the Streamlit app:**

* Deploying the app on Hugging Face Spaces with the correct folder structure and dependencies

# Solutions and Outcomes

To address these issues:

* We wrote code to extract the dataset from ZIP files within Colab
* Preprocessing was implemented using regular expressions
* We evaluated the model with a confusion matrix and heatmap visualization
* The model and vectorizer were saved using Pickle (.pkl) and integrated into Streamlit
* Requirements were listed and added to a `requirements.txt` for deployment

The final deployed app takes a user-inputted review and displays whether the sentiment is positive or negative instantly.

# Conclusion

This short-term project successfully demonstrates the application of machine learning to natural language processing (NLP). It helped understand the full pipeline from dataset handling to deployment and exposed us to real-world challenges and solutions in ML projects.