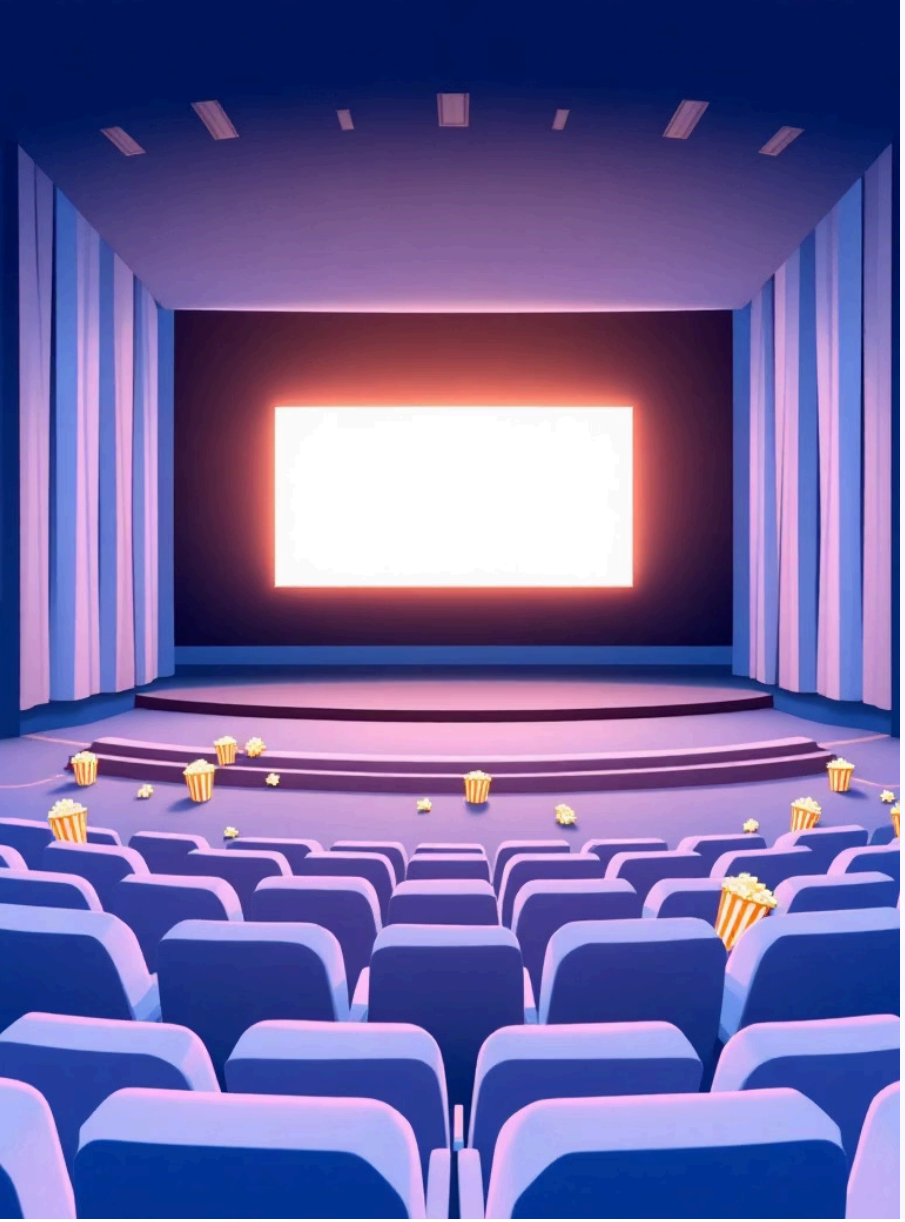
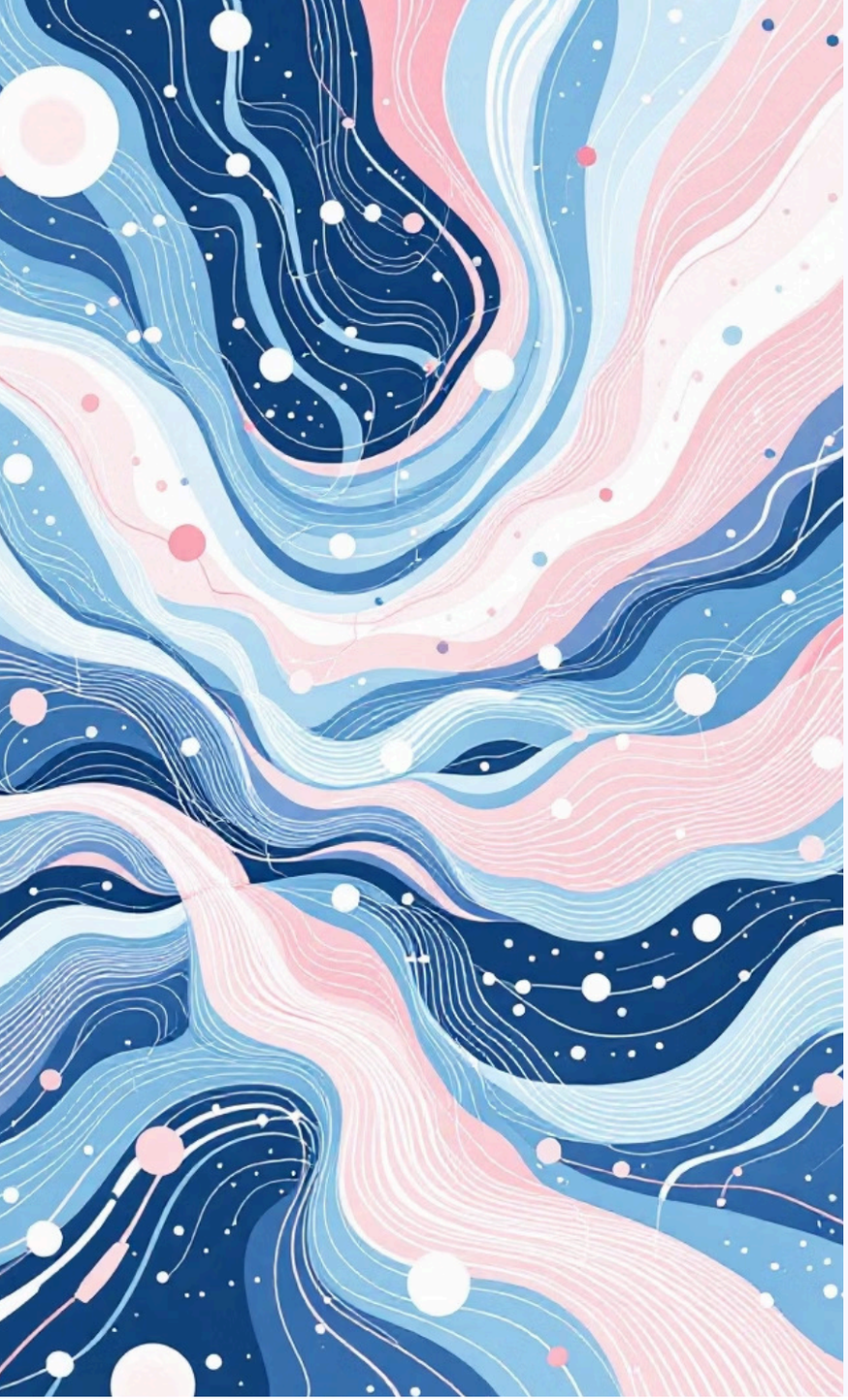


Movie Review Sentiment Analysis AI Project

Leveraging AI to decode audience emotions and unlock insights from millions of movie reviews



Problem Statement: Understanding Movie Reviews at Scale

Volume Overload

Millions of movie reviews flood online platforms daily, expressing diverse opinions that are impossible to analyze manually

Language Complexity

Reviews contain informal language, slang, sarcasm, and mixed sentiments that challenge traditional interpretation methods

Business Urgency

Filmmakers and entertainment businesses need automated tools to gauge audience sentiment quickly and make data-driven decisions

Proposed Solution: AI-Powered Sentiment Classification

01

Machine Learning-Based Sentiment Classifier

An ML model (Naive Bayes / Logistic Regression / SVM) trained on IMDB reviews to classify reviews as *positive* or *negative*.

02

Automated Text Pre-processing Pipeline

Implement a pipeline for cleaning reviews (removing noise, stopwords, punctuation, normalization) to improve model accuracy

03

Real-time Analysis

Automate sentiment detection to provide instant insights on movie reception and audience reactions

Goals and Objectives

1

Handle Complexity

Process noisy, unstructured text data through effective preprocessing and feature engineering

2

Actionable Insights

Deliver meaningful sentiment insights to support strategic decision-making in the entertainment industry

3

Scalability

Ensure the system can analyze large datasets containing 50,000+ movie reviews efficiently

Technology Stack

Python

Core programming language powering data processing, model development, and deployment workflows

ML Libraries

Training and testing sentiment classifier

NLP Tools

NLTK library for stopword removal, stemming, tokenization, and comprehensive text normalization

Dataset

IMDB Movie Review Dataset - Large labeled dataset of movie reviews used for training and validating the model

Key Features & Highlights

Data Cleaning

Comprehensive preprocessing including HTML tag removal, special character filtering, and case normalization

Real-Time Review Analysis

User inputs any review → system instantly predicts Positive / Negative sentiment.

Scalable & Modular Architecture

Each component–data preprocessing, feature extraction, training, and prediction–is separated, making the system easy to update or upgrade.

50.6%

Model Accuracy

Real-World Ready

The trained model can predict sentiment on new, unseen movie reviews with high confidence, making it production-ready for business applications

Thank You!

ARJUN CHAUHAN

S24CSEU1756

NIRVAN SINGH

S24CSEU1763