Natural Language Processing

Mini- Project

AI & DS

Title: Movie Genre Classification

Dhwaj Jain – S012- 60018230070

Eric Kurissery – S014- 60018230071

Hadi Gala – S015 – 60018230085

Kaustav Dedhia – S026 – 60018230009

Aim:

Develop a web app that predicts movie genres from short plot descriptions and recommends similar movies with plot summaries using NLP and content-based algorithms.

Problem:

Users struggle to identify movie genres or find similar movies based on text input; manual searching is time-consuming and subjective.
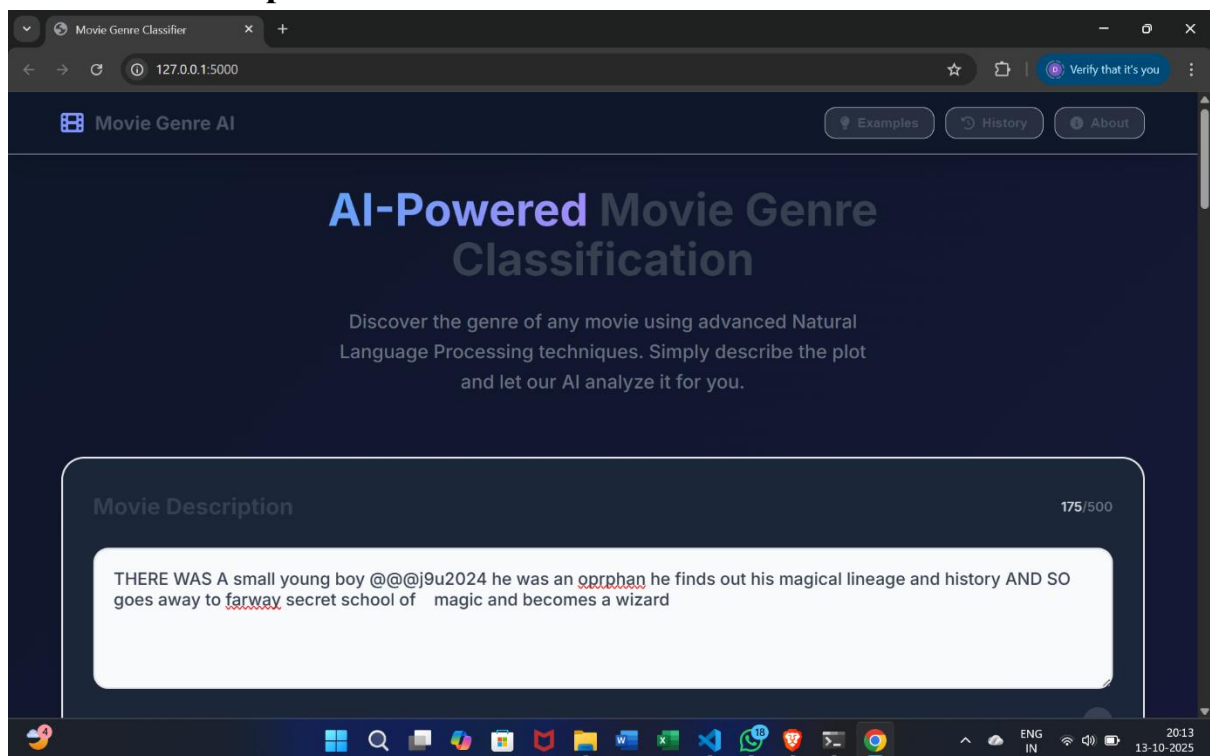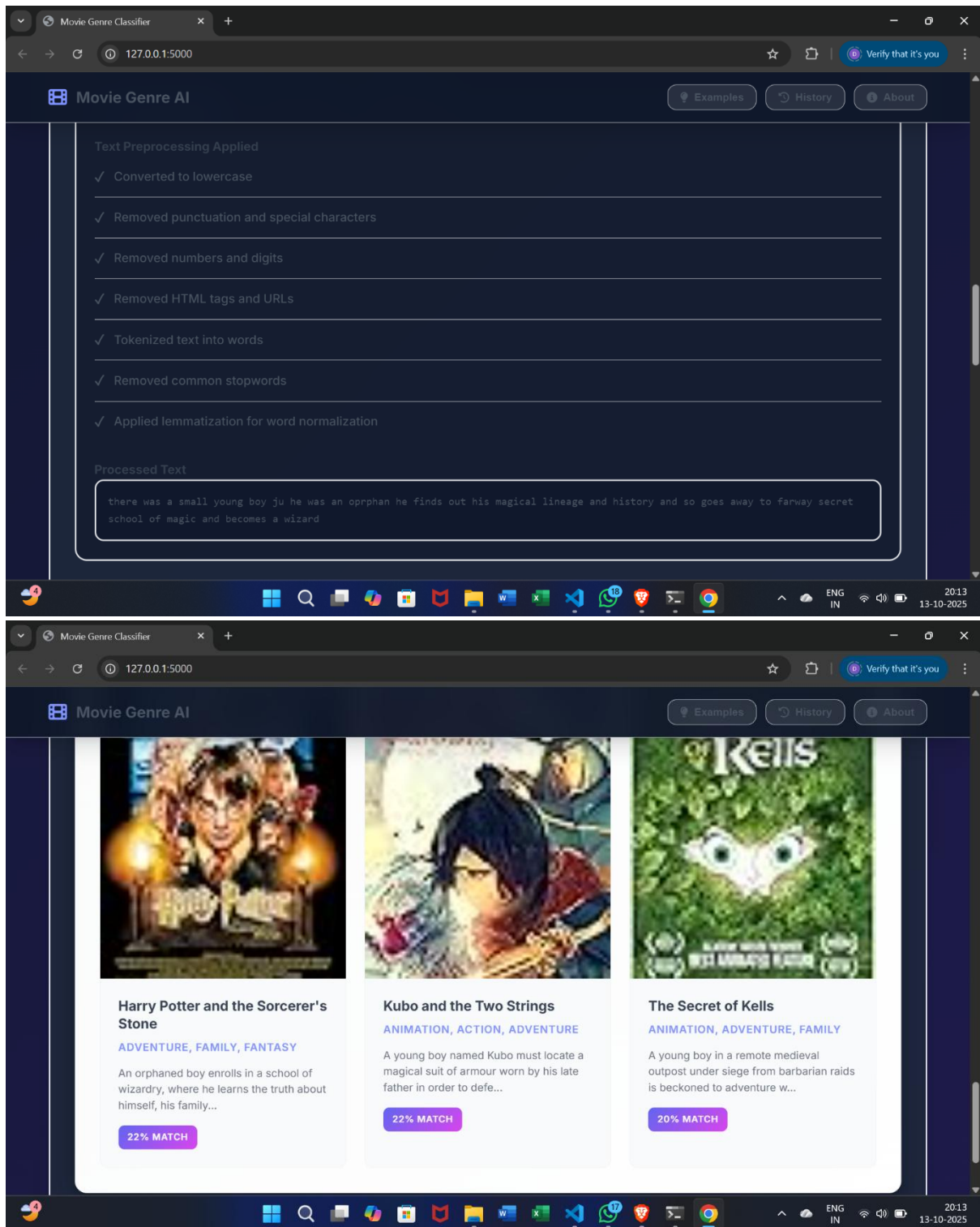
Solution:

Build an NLP model for genre classification and a content-based recommender system, accessible intuitively via a web interface.

NLP Pipeline & Implementation

- **Preprocessing:** Tokenization, stemming, lemmatization, stopword removal, and normalization techniques .

- **Feature extraction:** Discussion of N-grams, TF-IDF, POS tags, and how these representations boost genre classification performance.

- **Semantic features:** Use of WordNet (word sense disambiguation, synonyms/hypernyms) and integration into feature sets.

- **Models:** List all algorithms evaluated (Multinomial NB, Bernoulli NB, Logistic Regression, Random Forest).

- **Training/Evaluation:** Explain cross-validation, metrics such as accuracy, precision, recall, F1-score, confusion matrices, etc.

Results and Output:

Proposed Improvements/Future Work:

Planned enhancements (e.g., expand genre list, use transformers for better accuracy, or add collaborative filtering).