FlickFilter: IMDB Review Analysis

Mentor: Arun.K.S

Infosys Springboard

08-01-2025

What is the project about?

- **Overview:** Analyzing IMDB reviews to derive structured insights into user sentiment and preferences.
- **Challenge:** Unstructured textual data poses challenges in deriving actionable insights for stakeholders.
- **Goal:** Develop an efficient system to preprocess, analyze, and extract insights using advanced NLP techniques.

What did we do?

- **Methodology:**
  - Pre-process and analyze reviews using techniques like TF-IDF and word embeddings.
  - Train models to classify sentiment and visualize insights.
- **Tech Stack:**
  - Python for data pre-processing and analysis.
  - Scikit-learn for the TF-IDF matrix and Logistic Regression.
  - Streamlit for deploying a user-friendly web interface.
- **Key Benefits:**
  - High accuracy in predicting sentiment trends.
  - Modular and scalable approach for analysis across domains.

How did we work?

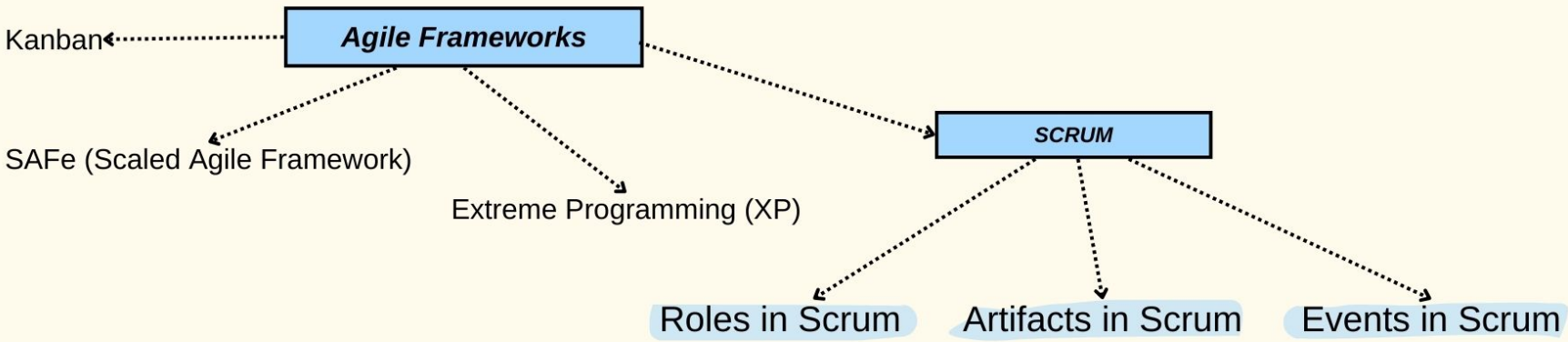

Figure 1: AGILE METHODOLOGY

What & for whom does it benefit?

- **Business Relevance:** Sentiment analysis is integral to understanding customer behavior, improving product strategies, and building recommendation systems.
- **Competitor Landscape:** Platforms like Amazon, Yelp, and Google leverage similar techniques for targeted marketing and user satisfaction analysis.
- **Opportunity:** Our tool provides a niche focus on IMDB reviews, helping filmmakers, critics, and marketers with insights to better tailor content and campaigns.

What did we achieve?

- **Sprint 1: Manassa, Rishipreeth**
  - Cleaned and pre-processed IMDB reviews dataset.
  - Generated TF-IDF matrix and word embeddings for feature representation.
- **Sprint 2: Siddharth, Vishnu, Joydip**
  - Trained Logistic Regression model for sentiment classification.
  - Delivered evaluation metrics including accuracy, F1 score, and confusion matrix.
- **Sprint 3: Aparna, Shreyash, Shrihari**
  - Created visualization reports for sentiment trends and model misclassifications.
  - Optimized model with advanced hyperparameter tuning.
- **Sprint 4: Lithikha, Aswin**
  - Built a Streamlit-based application for real-time sentiment analysis.
  - Provided comprehensive project documentation and reports.

Sprint 1

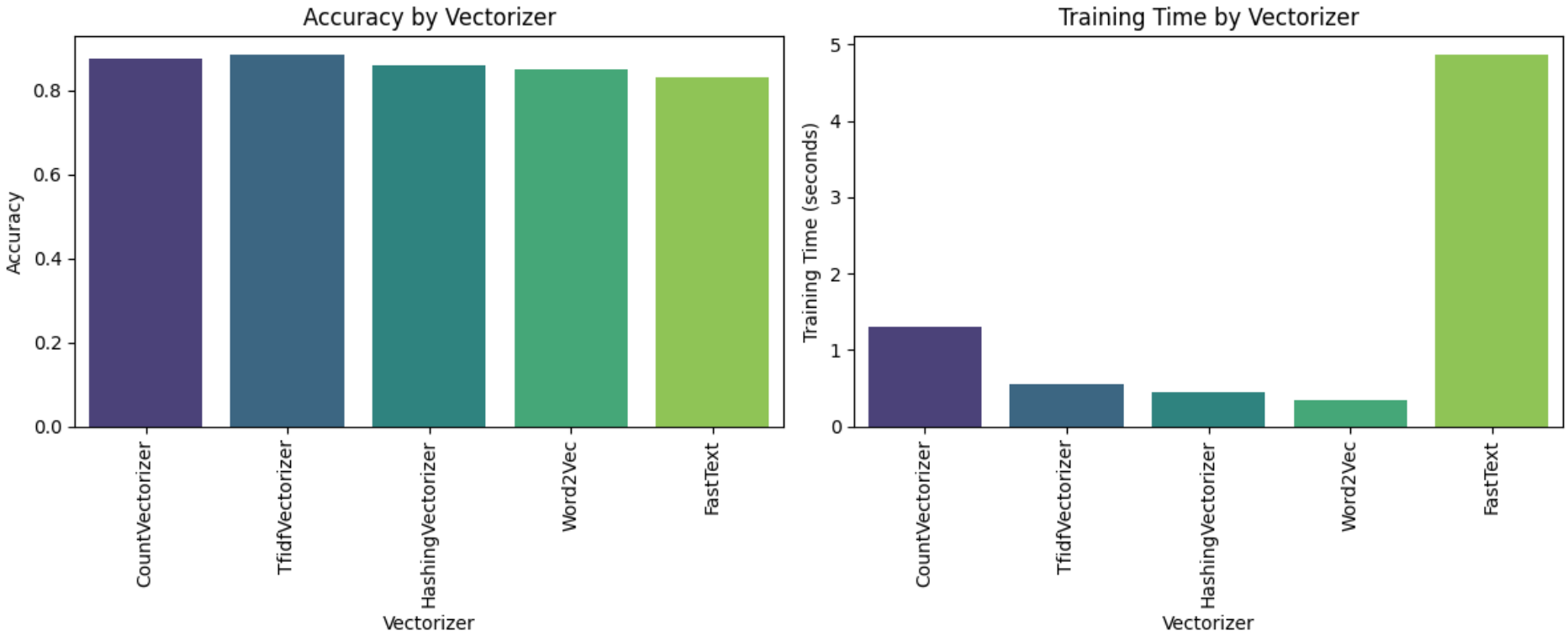

Figure 2: Vectorizer Selection

Sprint 2

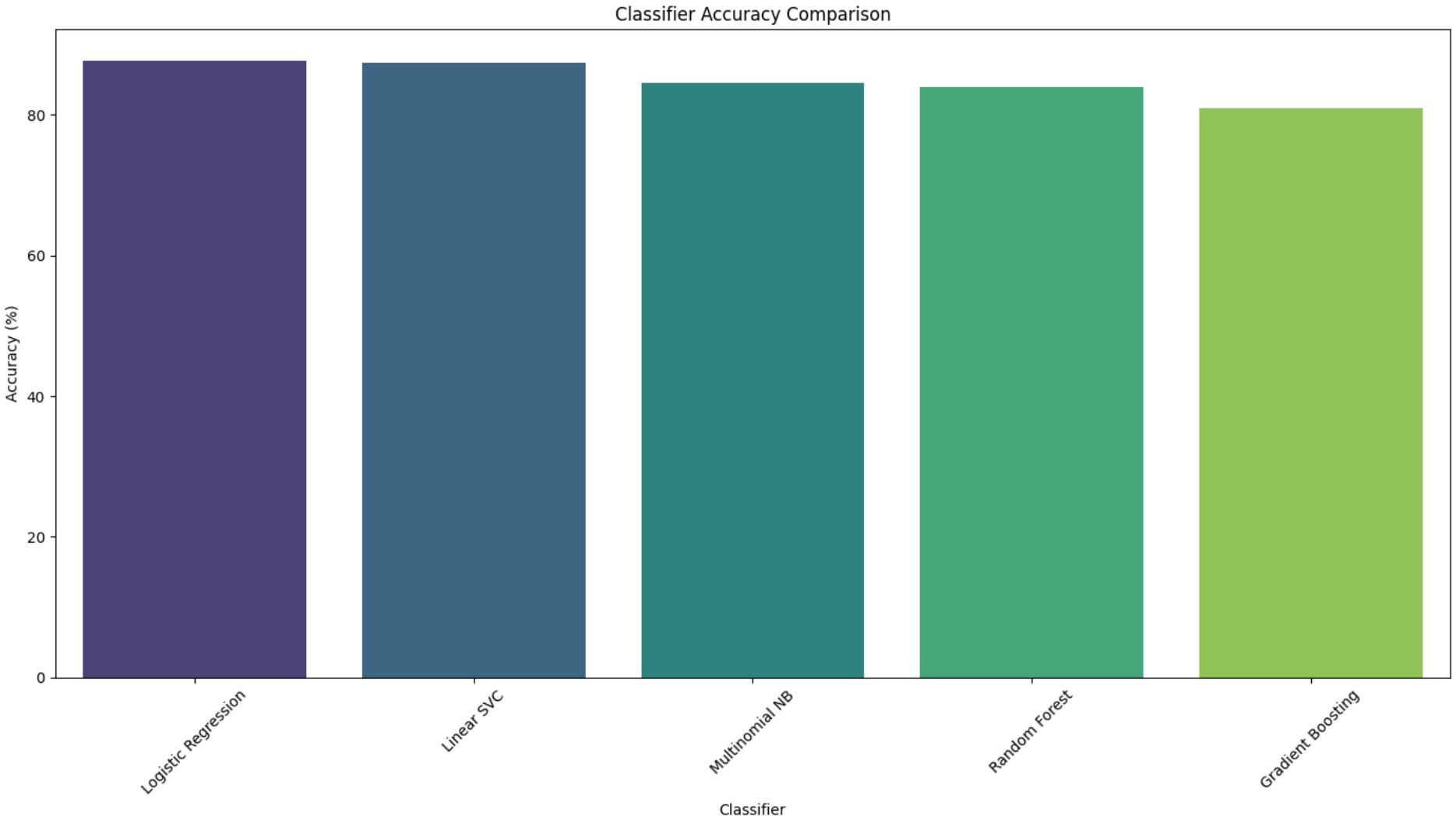

Figure 3: Model Selection

Sprint 3

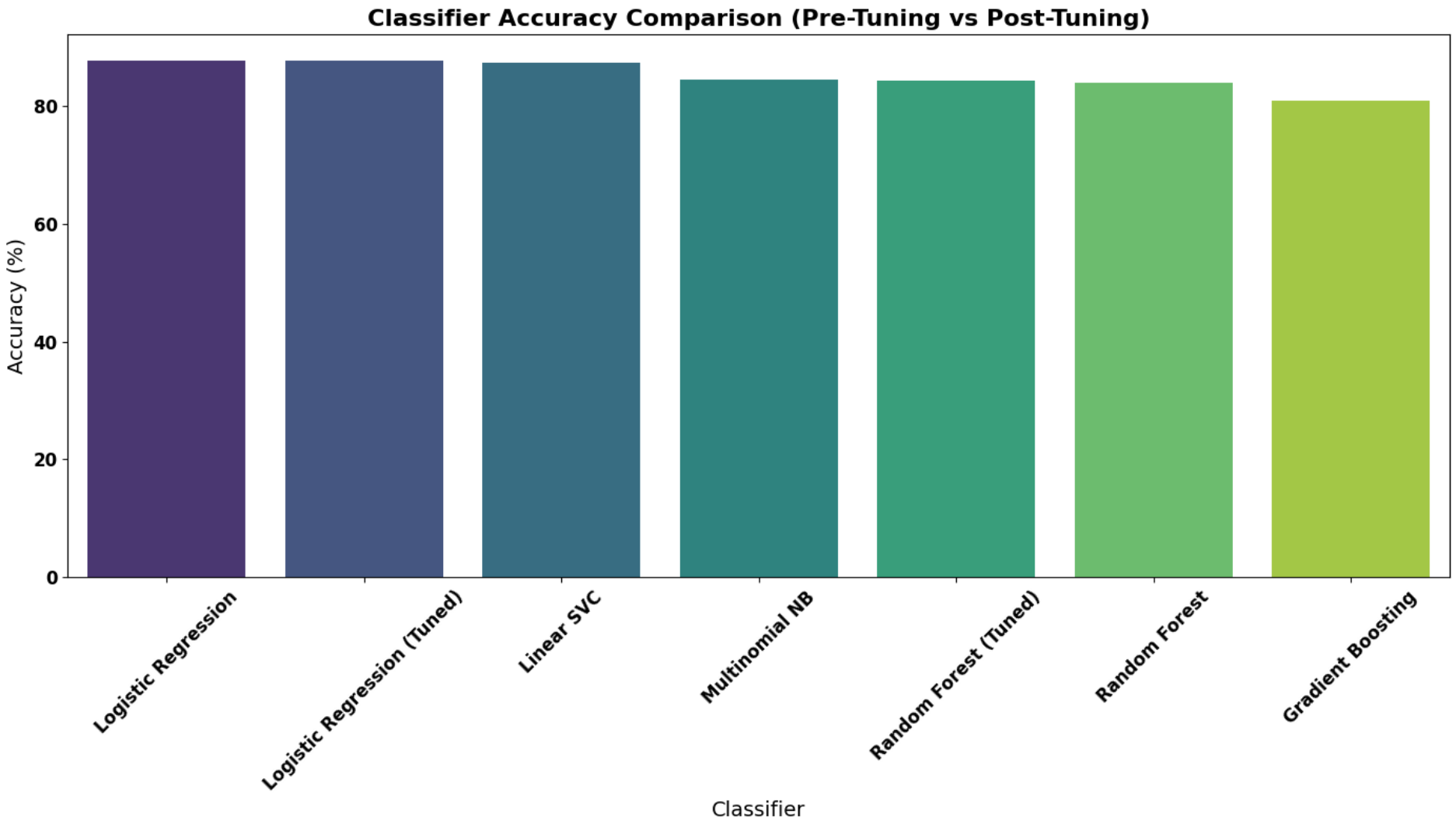

Figure 4: Models after tuning

Sprint 4

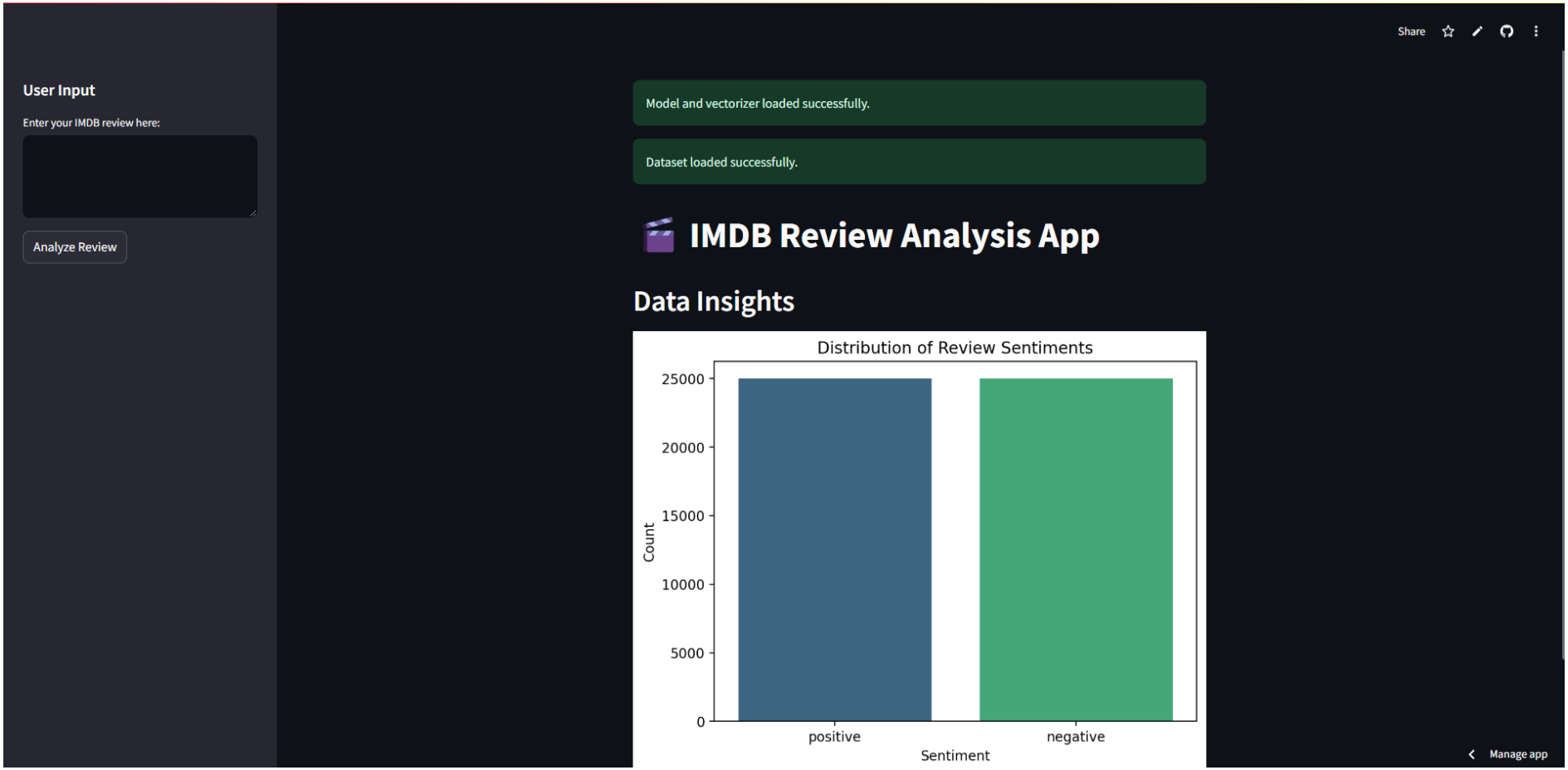

Figure 5: Click here to watch the Demo!

What next?

- **Summary:**
  - Achieved efficient sentiment analysis on IMDB reviews with actionable insights.
  - Delivered a deployable and scalable solution with a user-friendly interface.
- **Next Steps:**
  - Enhance the model to include multi-class sentiment analysis.
  - Extend to other review platforms and integrate real-time analytics.

Key Deliverables and References

- **GitHub Repository:** Project Code
- **Demo Video:** Watch Demo
- **Data Visualization:** Visual Reports
- **Sphinx Documentation:** Detailed Docs
- **Streamlit Application:** Live Web App
- **Research Paper:** Sentiment Analysis on IMDB Reviews
- Documentations

Thank You!

We are open to questions!