



Mini Project - Presentation

AMC Engineering College
Department of Information Science and Engineering

MOVIE REVIEW ANALYSIS WITH NLP USING WEB SCRAPED DATA

Project Guide:

Dr. R Senkamalavalli
Professor
Department of ISE

AMC Engineering college, Bengaluru

Head of Department

Dr. R. Amutha
Associate professor & HOD
Department of ISE

AMC Engineering college, Bengaluru

Principal

Dr. K. Kumar
Principal

AMC Engineering college, Bengaluru

Presented By Group No - 09

SL.NO	NAME	USN
1	Nischitha S H	1AM22IS069
2	Poornima Narayana Hegde	1AM22IS076
3	Prachi Singh	1AM22IS077
4	Priya G	1AM22IS084

1. INTRODUCTION

- The project aims to analyze movie reviews using natural language processing (NLP) and machine learning.
- It leverages web-scraped data from social media platforms to classify sentiments.
- Sentiments are categorized as positive, negative, or neutral, offering insights into public opinion.
- An interactive dashboard allows users to explore trends by movie, genre, or time period.
- The platform is valuable for film enthusiasts, critics, and marketers.

2. PROBLEM STATEMENT

- Unstructured data from social media is abundant but challenging to analyze manually.
- Current methods are time-intensive and prone to errors.
- Existing tools often lack scalability and fail to analyze trends effectively.
- There's no comprehensive platform for exploring sentiment trends across movies or genres.
- A scalable, automated solution is essential to extract meaningful insights from vast movie review data.

3. LITERATURE SURVEY

Sl.No	Title	Authors	Year	Journal	Approach	Advantages	Limitations
1	Automatic Movie Review System Using Sentimental Analysis For Positive or Negative Reviews	Towards AI Team	2023	Towards AI	Web scraping, binary sentiment classification	Simple implementation	Ignore sentiment Neutral
2	Croatian Film Review Dataset (Cro-FiReDa): A Sentiment Annotated Dataset of Film Reviews	Gaurish Thakkar, Nives Mikelic Preradovic, Marko Tadić	2023	arXiv	New dataset, transformer-based sentiment analysis	Transformer effectiveness	Language-specific focus

3. LITERATURE SURVEY

Sl.No	Title	Authors	Year	Journal	Approach	Advantages	Limitations
3	Intro to Using Web Scraping For Sentiment Analysis	ScrapFly Team	2023	ScrapFly Blog	Scalable web scraping methods	Practical scalability	Limited NLP details
4	Sentiment Analysis on IMDB Movie Comments and Twitter Data by Machine Learning and Vector Space Techniques	İlhan Tarımer, Adil Çoban, Arif Emre Kocaman	2019	arXiv	ML algorithms compare sentiment accuracy	ML performance insights	Lacks deep learning

4 EXISTING METHOD

Manual Analysis of Reviews:

- Human effort was required to read and interpret reviews, leading to inefficiency and biases.

Limited Scope:

- Existing systems focused on structured or predefined datasets, ignoring diverse and unstructured data from social media.

Static Insights:

- Traditional methods provided static reports without capturing trends over time.

Inefficient Data Processing:

- Manual and older methods struggled to process large datasets efficiently.

Lack of Customization Options:

- Users couldn't tailor insights to specific interests, such as genre or timeframe.

5.PROPOSED METHOD

- **Data Collection:** Use web scraping tools like Selenium to gather movie reviews from social media
- **Processing Pipeline:** Clean and preprocess the text data for sentiment analysis using tokenization, stopword removal, and stemming.
- **Sentiment Analysis:** Apply machine learning models such as:
 - Naïve Bayes for basic classification.
 - Advanced transformer models like BERT for better accuracy.
- **Visualization:** Build a dashboard to present insights like sentiment trends by movie, genre, or release date.
- **Outcome:** Provide meaningful insights into public opinion, empowering decision-making for marketers and enthusiasts.

Problems in Current Methods:

1. **Scalability Issues:** Manual and traditional approaches struggle to process large volumes of unstructured social media data.
2. **Lack of Automation:** Reliance on human efforts for analysis leads to inefficiency and errors.
3. **Static Insights:** Current methods fail to provide trends or historical sentiment shifts.

How Our Proposed Solution Solves It:

1. **Scalability:** Utilizes web scraping tools like Selenium and BeautifulSoup to handle large-scale data collection efficiently.
2. **Automation:** Employs machine learning and NLP models like Naïve Bayes and BERT to automate the sentiment analysis process.
3. **Dynamic Insights:** Provides interactive dashboards to track trends over time by movies, genres, or release dates.

REFERENCES

1. İlhan Tarımer, Adil Çoban, Arif Emre Kocaman, 2019 *"Sentiment Analysis on IMDB Movie Comments and Twitter Data by Machine Learning and Vector Space Techniques."* arXiv.
2. Gaurish Thakkar, Nives Mikelić Preradović, Marko Tadić, 2023 *"Croatian Film Review Dataset (Cro-FiReDa): A Sentiment Annotated Dataset of Film Reviews."* arXiv.
3. ScrapFly Team, 2023 *"Intro to Using Web Scraping for Sentiment Analysis."* ScrapFly Blog.
4. Towards AI Team, 2023 *"Automatic Movie Review System Using Sentimental Analysis for Positive or Negative Reviews."* Towards AI.
5. Tutorials and Documentation: Web scraping tools like Selenium, BeautifulSoup, and Scrapy official documentation.
6. Transformer-based models like BERT: Research papers and implementation tutorials from Google AI.

**THANK
YOU**
