Vivekanand Education Society's
Institute of Technology
(Autonomous Institute Affiliated to University of Mumbai, Approved by AICTE & Recognised by Govt. of Maharashtra)
NAAC accredited with 'A' grade

Semester: VI

Title of the Project: AI in Social Media Trend Prediction

Domain: Artificial Intelligence

D15C Sandesh Yadav (61)

Content

- Introduction
- Problem statement
- Algorithm description
- Implementation
- Results
- Application
- Conclusion
- References

Introduction

- **Overview:**
  - Rapid growth of social media generates massive data daily.
  - AI techniques can analyze this data to predict emerging trends.

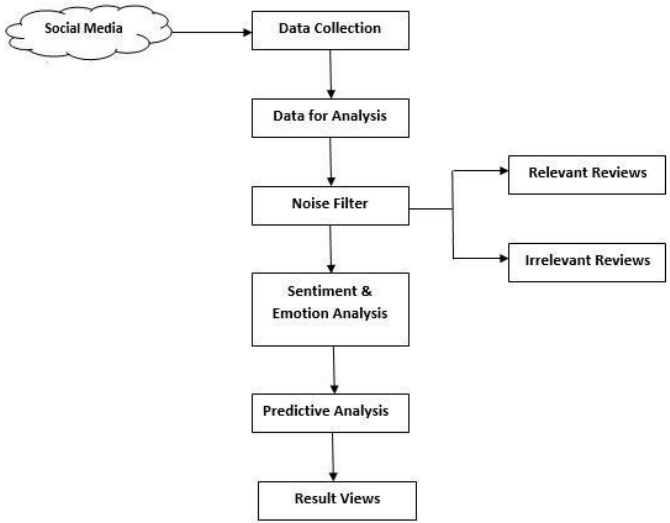
- **Importance:**
  - Helps brands, marketers, and influencers tailor strategies.
  - Enables proactive content creation and audience engagement.

Problem statement

- **Key Challenges:**
  - Extracting relevant signals from noisy, unstructured data.
  - Handling real-time data flows and rapidly changing trends.
  - Accurately forecasting trends to support decision-making.
- **Objective:**
  - Develop an AI-driven model to analyze and predict social media trends effectively.

Algorithm Description

- **Data Preprocessing**
  - **Cleaning Data:**
    - Remove duplicate entries to ensure unique records.
    - Skip problematic rows when loading the dataset.
  - **Handling Missing Values:**
    - Fill missing tags with an empty string.
    - Convert essential numerical columns (view_count, likes, dislikes, comment_count) to numeric format.
    - Drop rows with missing values in these key numeric columns.
- **Filtering Data:**
  - Remove rows where view_count is zero, as they don't provide meaningful insights.



Algorithm Description

- **Feature Extraction**
  - **Statistical Feature Engineering:**
    - Define median threshold values for view_count, likes, dislikes, and comment_count.
    - Use these thresholds to classify whether a video is Trending (1) or Not Trending (0).
- **Regression Model Features:**
  - Use view_count, likes, dislikes, and comment_count as independent variables for Logistic Regression.
  - Compute and display the regression equation to understand feature contributions.
- **Text-Based Features (Potential Enhancements):**
  - Process tags column for keyword analysis.
  - Apply Natural Language Processing (NLP) to analyze video titles/descriptions for sentiment and engagement potential.

Implementation

- **Load & Clean Data**: Read dataset, remove duplicates, and handle missing values.
- **Convert Data Types**: Ensure numeric values for key features; remove zero view_count.
- **Set Trending Thresholds**: Use median view_count to label videos as **Trending (1) or Not Trending (0)**.
- **Prepare Features & Labels**: Select view_count, likes, dislikes, and comment_count as predictors.
- **Train Model**: Split data (80/20) and train a **Logistics Regression** model.
- **Evaluate Model**: Predict on test data, compute accuracy, precision, recall, and display regression equation.
- **User Prediction Interface**: Take input, predict trending status, and display results with thresholds.

Result

```
--- Video Trend Prediction Interface ---
Enter view count: 2222222
Enter number of likes: 222
Enter number of dislikes: 2222
Enter number of comments: 2
C:\Users\Sandesh\AppData\Local\Programs\Python\Python313\Lib\site-packages\sklearn\utils\validation.py:2739: U
serWarning: X does not have valid feature names, but LogisticRegression was fitted with feature names
  warnings.warn(

--- Prediction Details ---
View Count: 2222222.0 (Threshold: 947661.0)
Likes: 222.0 (Threshold: 37003.5)
Dislikes: 2222.0 (Threshold: 0.0)

Prediction: The video is predicted to be 'Trending'.
PS C:\Users\Sandesh\Downloads\IN_youtube_trending_data.csv>
```

Application

- **Digital Marketing & Advertising :** AI identifies trending topics, enabling brands to craft viral campaigns and optimize ad targeting based on engagement patterns.
- **Influencer & Content Strategy :** Helps influencers create relevant content, maximize visibility, and determine the best posting times for higher engagement.
- **Media & News Industry :** Detects emerging news trends early, allowing media houses to prioritize and cover high-impact stories efficiently.
- **E-Commerce & Brand Management :** Predicts consumer demand based on social media trends and enhances brand reputation management through sentiment analysis.
- **Crisis & Reputation Management :** Monitors negative trends in real-time, helping companies respond proactively to potential PR crises before they escalate.

Conclusion

The AI-driven model effectively analyzes and predicts social media trends using key engagement metrics. By leveraging data preprocessing, feature extraction, and logistic regression, it accurately classifies trending content. The system helps brands, marketers, and influencers make data-driven decisions, optimize content strategies, and improve audience engagement. Future enhancements, such as NLP for text analysis and real-time processing, can further refine trend forecasting capabilities.

References

Dataset :

- Kaggle. (2023). *IN_youtube_trending_data* [Data set]. Retrieved from https://www.kaggle.com/datasets/rsrishav/youtube-trending-video-dataset/

Research Papers:

- Wu, B., Liu, P., Cheng, W.-H., Liu, B., Zeng, Z., Wang, J., Huang, Q., & Luo, J. (2024). *SMP Challenge: An Overview and Analysis of Social Media Prediction Challenge*. arXiv preprint arXiv:2405.10497. Retrieved from https://arxiv.org/abs/2405.10497

Thank You!!!