# JV CINELYTICS

## Intelligent Script Analysis for Smarter Filmmaking

**A PROJECT REPORT SUBMITTED IN PARTIAL FULFILMENT OF REQUIREMENT**

**FOR THE AWARD OF THE DEGREE**

**MASTER OF COMPUTER APPLICATIONS (MCA)**

**OF**

#### MAHATMA GANDHI UNIVERSITY, KOTTAYAM

**BY**

**Akash Mathew**

## Reg No : 24PMC107



### Marian College Kuttikanam Autonomous

**Peermade, Kerala – 685 531**

**2025**

# JV CINELYTICS

## Intelligent Script Analysis for Smarter Filmmaking

**A PROJECT REPORT SUBMITTED IN PARTIAL FULFILMENT OF REQUIREMENT**

**FOR THE AWARD OF THE DEGREE**

**MASTER OF COMPUTER APPLICATIONS (MCA)**

**OF**

#### MAHATMA GANDHI UNIVERSITY, KOTTAYAM

**BY**

**AKASH MATHEW**

## Reg No : 24PMC107



### Marian College Kuttikanam Autonomous

**Peermade, Kerala – 685 531**

**2025**

A Project Report on

# JV CINELYTICS

## Intelligent Script Analysis for Smarter Filmmaking

**SUBMITTED IN PARTIAL FULFILMENT OF REQUIREMENT FOR THE AWARD OF THE DEGREE**

**MASTER OF COMPUTER APPLICATIONS (MCA)**

OF

#### MAHATMA GANDHI UNIVERSITY, KOTTAYAM

**By AKASH MATHEW**

**24PMC107**

**Under the guidance of**

Dr. Sr. Italia Joseph Maria

Assistant Professor

PG Department of Computer Applications Marian College Kuttikkanam Autonomous



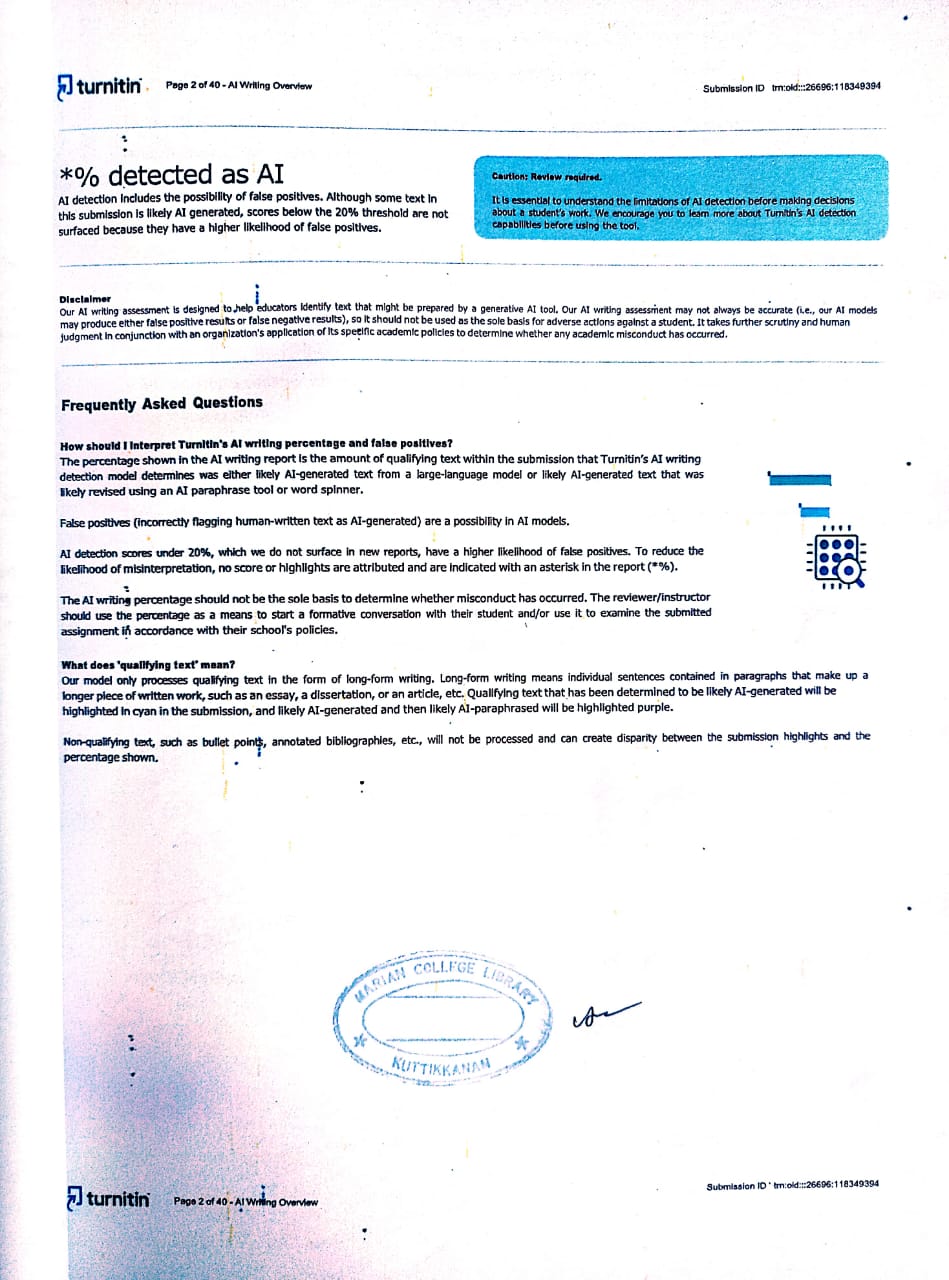
### Marian College Kuttikanam Autonomous

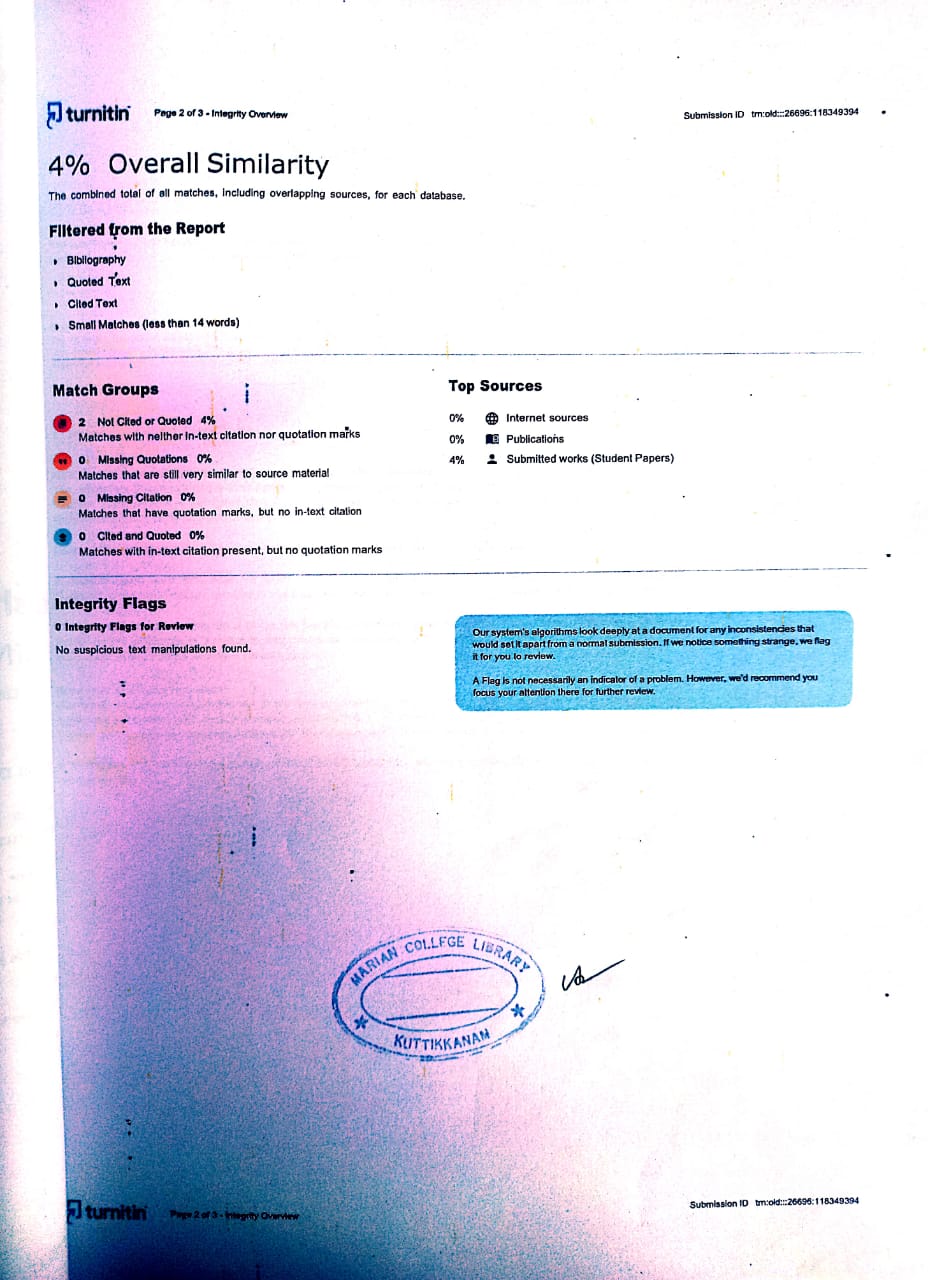
**Peermade, Kerala – 685 531**

**2025**

****

****

****

****

#### Acknowledgements

First and foremost, I thank the God Almighty for His immense grace and blessings in my life and at each stage of my project work.

I express my sincere gratitude to Prof. Dr. Ajimon George, Principal, Marian College Autonomous Kuttikanam, and Dr. Mendus Jacob, Director, PG Department of Computer Applications, for the unwavering support given throughout my project work.

My heartfelt thanks go to Mr. Win Mathew John, HOD, Department of Computer Applications, who has been a constant source of inspiration and whose advice greatly helped me to complete this project successfully.

I am deeply indebted to my project guide, Dr. Sr. Italia Joseph Maria, Assistant Professor, PG Department of Computer Applications, for her profound guidance, encouragement, and invaluable support that made this project a success.

I extend my gratitude to all the faculty members of the PG Department of Computer Applications for their timely help, support, and motivation throughout my journey.

Finally, I express my deepest appreciation to my family and friends for their moral support and encouragement, which played a vital role in the successful completion of this project work.

###### Akash Mathew

**ABSTRACT OF JV CINELYTICS**

##### Title:

JV CINELYTICS: Intelligent Script Analysis for Smarter Filmmaking

##### Problem Statement:

In the film industry, script analysis is a critical yet time-consuming task for filmmakers, scriptwriters, and producers. Manual analysis of movie scripts requires extensive effort to extract key elements such as character relationships, plot structure, emotional tone, and genre classification. Traditional methods lack automation and fail to provide data-driven insights that could inform creative and production decisions. This project addresses the need for an intelligent, automated script analysis system that can quickly process screenplay documents and provide comprehensive analytical insights including character importance ranking, location detection, plot summarization, and multi-dimensional text classification.

The choice of this domain is motivated by the growing intersection of artificial intelligence and creative industries, where machine learning can augment human creativity rather than replace it, enabling content creators to make informed decisions backed by quantitative analysis.

##### Objective of the Project:

###### Objective

The primary objective of JV Cinelytics is to develop a unified machine learning platform that:

* Trains custom multitask deep learning models for simultaneous sentiment analysis, genre classification, and emotion detection from textual content
* Analyzes movie scripts by extracting characters, locations, generating intelligent summaries, and predicting genres
* Provides a user-friendly web interface through Streamlit for seamless interaction with both training and analysis capabilities
* Delivers actionable insights through visualizations and downloadable reports to support creative decision-making in scriptwriting and production

The system combines both extractive and abstractive NLP techniques to deliver comprehensive screenplay analysis while maintaining interpretability and accuracy.

##### Dataset Description :

##### The project utilizes custom JSONL (JSON Lines) formatted datasets for training multitask models. Each training sample contains:

##### Text: Screenplay dialogue or action descriptions

##### Labels:

##### Sentiment (3 classes: negative, neutral, positive)

##### Genre (7 classes: action, drama, comedy, romance, thriller, scifi, horror)

##### Emotion (7 classes: anger, joy, sadness, fear, disgust, surprise, neutral)

##### Dataset Characteristics:

##### Source: Manually curated and collected from screenplay databases

##### Format: JSONL for efficient streaming and processing

##### Size: Flexible supports custom user-uploaded datasets

##### Vocabulary: Built dynamically from training corpus using frequency-based filtering (min\_freq=2)

##### Preprocessing: Simple tokenization with lowercase normalization

##### Script Analysis Input:

##### File Formats: .docx and .txt screenplay files

##### Structure: Standard screenplay format with character names in ALL CAPS, scene headings (INT./EXT.), dialogue, and action descriptions

##### 4. Methodology

##### 4.1 Data Preprocessing

##### Text Cleaning: Removal of script-specific formatting, extraction of dialogue and action lines

##### Tokenization: Simple white-space-based tokenization with lowercase normalization

##### Vocabulary Construction: Dynamic vocabulary building with frequency thresholding and special tokens (`<pad>`, `<unk>`)

##### Label Encoding: Multitask label encoding for sentiment, genre, and emotion classes

##### Sequence Padding: Fixed-length sequences (max\_len=256) with attention masking

##### 4.2 Feature Engineering

##### Positional Embeddings: Learnable position encodings for sequence modeling

##### Attention Masking: Binary masks to handle variable length inputs

##### Task-Specific Masking: Enables partial labeling across multiple tasks

##### Character Importance Scoring: Weighted combination of dialogue count (×2) and total mentions

##### Location Frequency Analysis: Regex-based extraction and ranking of scene locations

##### 4.3 Model Architecture

##### Multi-Task-Text-Model: A unified deep learning architecture

##### Encoder Options:

##### 1. Transformer Encoder (Default)

##### Multi-head self-attention (4 heads)

##### Position-wise feedforward networks

##### Layer normalization and residual connections

##### Embedding dimension: 128

##### Hidden dimension: 256

##### Number of layers: 2

##### 2. Bidirectional LSTM Encoder (Alternative)

##### Bidirectional sequence processing

##### Dropout regularization between layers

##### Hidden size: 256 (512 total with bidirectionality)

##### Task-Specific Heads:

##### Three independent linear classifiers for sentiment, genre, and emotion

##### Shared representation learning with taskspecific finetuning

##### Loss Function:

##### Weighted multitask crossentropy loss

##### Weights: Sentiment (1.0), Genre (1.0), Emotion (0.5)

##### Task masking for handling missing labels

##### 4.4 Training Configuration

##### Optimizer: AdamW with weight decay

##### Learning Rate: 3e4

##### Batch Size: 32

##### Epochs: 5

##### Gradient Clipping: Max norm of 1.0 to prevent exploding gradients

##### Validation: Early stopping based on validation loss

##### Device: CUDAenabled GPU when available, CPU fallback

##### 4.5 Script Analysis Pipeline

##### Character Extraction:

##### Regex pattern matching for ALL CAPS character names

##### Dialogue count and mention frequency tracking

##### Importance ranking based on weighted scoring

##### Location Detection:

##### Scene heading pattern recognition (INT./EXT. LOCATION TIME)

##### Frequencybased location ranking

##### Summarization:

##### Extractive Approach: LexRank algorithm for sentence importance scoring

##### Abstractive Approach: DistilBARTCNN126 for sequencetosequence generation

##### Hybrid Pipeline: Intelligent scene splitting, narrative element extraction, story arc analysis

##### Customizable Length: Target sentence count (default: 25 for detailed synopsis)

##### Genre Prediction:

##### Primary: Trained multitask model inference

##### Fallback: Keywordbased classification when model unavailable

##### 4.6 Evaluation Metrics

##### CrossEntropy Loss: Primary training objective

##### Validation Loss: Model selection criterion

##### TaskSpecific Accuracy: Pertask classification performance (future enhancement)

##### Qualitative Analysis: Summary coherence and completeness assessment

**TABLE OF CONTENTS**

|  |  |  |  |
| --- | --- | --- | --- |
| Chapter |  | | Page No |
| 1 | Introduction | | 1 |
|  | 1.1 | Problem Statements | 2 |
|  | 1.2 | Proposed System | 3 |
|  | 1.3 | Features of the Proposed System | 4 |
|  | 1.4 | Architecture(Block) Diagram / Workflow | 5 |
| 2 | Dataset Summary | | 6 |
|  | 2.1 | Description | 7 |
|  | 2.2 | Sample Dataset | 8 |
|  | 2.3 | Data Cleaning and Transformation Techniques | 9 |
| 3 | Model Building and Evaluation | | 10 |
|  | 3.1 | Algorithms Used | 11 |
|  | 3.2 | Model Training and Testing Results | 14 |
|  | 3.3 | Performance Metrics | 15 |
| 4 | Insights with Visualizations/ Analysis | | 18 |
| 5 | Web App Integration | | 23 |
|  | 5.1 | Home page | 24 |
|  | 5.2 | Dashboards (Admin and User) | 24 |
|  | 5.3 | Other Pages | 26 |
|  | 5.4 | Usernames and Passwords | 28 |
| 6 | GitHub Repository and Colab Link | | 29 |
| 7 | Future Enhancements | | 31 |
| 8 | Conclusion | | 34 |
| 9 | References | | 37 |
| 10 | Annexure | | 39 |
|  | A | Screen Shots | 40 |
|  | B | Ethical Clearance Form | 46 |

**FIGURE INDEX**

|  |  |  |
| --- | --- | --- |
|  | Figure | Page No |
| 1.1 | Architecture / workflow diagram | 5 |
| 3.1 | Confusion matrix diagram for genre Classification | 15 |
| 3.2 | Classification Report | 17 |
| 4.1 | Genre and User Distribution | 19 |
| 4.2 | Character Analysis | 20 |
| 4.3 | Location Analysis | 21 |
| 4.4 | Classification Breakdown | 22 |