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 KUTTIKKANAM AUTONOMOUS
MAKING COMPLETE

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 KUTTIKKANAM AUTONOMOUS
MAKING COMPLETE

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 KUTTIKKANAM AUTONOMOUS
MAKING COMPLETE

Marian College Kuttikanam Autonomous

Peermade, Kerala – 685 531

2025

PG DEPARTMENT OF COMPUTER APPLICATIONS

Marian College Kuttikkanam Autonomous

MAHATMA GANDHI UNIVERSITY, KOTTAYAM

KUTTIKKANAM – 685 531, KERALA.

CERTIFICATE

This is to certify that the project work entitled

JV CINELYTICS

is a Bonafide record of work done by

AKASH MATHEW

Reg. No. 24PMC107

In partial fulfillment of the requirements for the award of Degree of

MASTER OF COMPUTER APPLICATIONS [MCA]

During the academic year 2024-2025

Dr. Sr. Italia Joseph Maria
Assistant Professor
PG Department of Computer Applications
Marian College Kuttikkanam Autonomous

Mr. Win Mathew John
Head of the Department
PG Department of Computer Applications
Marian College Kuttikkanam Autonomous

External Examiner

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

1. 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.

2. 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.

3. 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

FIGURE INDEX