



Filmception

**Static aAn AI-powered Multilingual movie summary
translator and genre classifier**

Artificial Intelligence

Submitted by: Aleena Fatima, Muhammad Haseeb

Roll number: 22i-2353, 22i-1698

Date: 6 May 2025



National University of Computer and Emerging Sciences

Islamabad Campus

Introduction

The Filmception project aims to develop an AI-powered multilingual movie summary translator and genre classifier. This report outlines the steps taken in the project, focusing on data preprocessing and cleaning, text translation and audio conversion, movie genre prediction model development, and the creation of an interactive user interface.

1. Data Preprocessing and Cleaning

Summary Extraction and Cleaning

Preprocessing Steps:

- **Removing Special Characters:** All non-alphanumeric characters were eliminated to ensure clean text.
- **Stopword Removal:** Commonly used words that do not contribute to the meaning (e.g., "and", "the") were removed.
- **Lowercasing:** All text was converted to lowercase to standardize the data.
- **Tokenization:** The summaries were broken down into individual words or phrases for easier analysis.
- **Stemming/Lemmatization:** Words were reduced to their base or root form to unify different variations of a word.
- **Removing Non-Relevant Words:** Numbers and punctuation that do not add value to genre classification were removed.

Metadata Extraction

Genre Information: Extracted from the movie.metadata.tsv file, the genre information was identified as multi-labels, allowing each movie to belong to multiple genres (e.g., "Action", "Adventure", "Comedy").

Final Output: A cleaned dataset was created containing the Movie ID, cleaned summary, and genres. This dataset will be used for training the genre prediction model.



National University of Computer and Emerging Sciences

Islamabad Campus

```
cleaned_plot_summaries.csv
1  movie_id,summary,genres
2  23890098,"shlykov, a hardworking taxi driver and lyosha, a saxophonist, develop a bizarre lovehate relationship, and despite their prejudices, real
3  31186339,"the nation of panem consists of a wealthy capitol and twelve poorer districts. as punishment for a past rebellion, each district must pro
4  28663735,"poovalli induchoodan is sentenced for six years prison life for murdering his classmate. induchoodan, the only son of justice maranchery
5  2231378,"the lemon drop kid , a new york city swindler, is illegally toutting horses at a florida racetrack. after several successful hustles, the k
6  595909,"seventhday adventist church pastor michael chamberlain, his wife lindy, their two sons, and their nineweekold daughter azaria are on a camp
7  5272176,"the president is on his way to give a speech. while he is traveling there a man shows up with a camera. a reporter tries to ask a member o
8  1952976,"plot the film opens in 1974, as a young girl, dahlia, stands outside after school in the rain, waiting for her mother. flash forward to 20
9  24225279,"the story begins with hannah, a young jewish teen, as she is completing her senior year of high school. her small neighborhood in brookly
10 2462689,"infuriated at being told to write one final column after being laid off from her newspaper job, ann mitchell prints a letter from a fictic
11 20532852,"a line of people drool at the window of the shop of market butcher buzz buzzard. a short series of gags ensues about how buzz dishonestly
12 15401493,"lola attempts to gain her fathers trust fund by hiring a hispanic husband, bo , offering him 100,000 per year if he marries her. bo accep
13 18188932,"milan and goran are two criminals who smuggle illegal immigrants. one night after they complete a smuggle, they discover that one of the
14 2940516,"bumbling pirate crewman kills his captain after learning where he has hidden buried treasure. however as he begins to lose his memory, he
15 1335380,"the film is based on the events that happened on the ship exodus in 1947 as well as events dealing with the founding of the state of israe
16 1488747,"plot following the sudden death of kids father pop the local church has donated a scholarship fund to kid so that he can go to college. pl
17 24448645,"despite lucys reservations, she agrees to her husbands request to attend a wild party of lust and sexual exploration on a remote island.
18 15072401,"alan colby, heir to a vast fortune, reappears after a seven year absence, only to be murdered before he can claim his inheritance. the lo
19 4018288,"debbies favorite band is dream street, and her favorite member is chris troussdale. when chris gets a fever while travelling on the dream s
20 4596602,"ashes to ashes is set in london, england. the film is a surreal journey into the dreams and misadventures of gabriel darbeaux a young scre
21 15224586,"the film follows the experience of vietnamese refugees in the united states immediately following the fall of saigon at the end of the vi
22 15585766,"three friends are droving cattle in australia in 1939 the restless bluey donkin, easygoing milo trent and english peter linton, who is in
23 1760737,"set in a lonely city on a rainy night, the film takes place in a bicycle shop that is closed for the night. in the corner of the shop slee
24 29062594,"a group of teenagers at catholic boarding school in the provinces spend their time trying to adapt to the demands of the priest in charge
25 20326153,"the film follows the lives of several west point cadet classmates who find themselves on opposite sides of the war. the film also follows
26 9252321,"this story of a zulu family is a composite story of events enacted by africans whose experiences resembled the storys events. forced out o
27 8388648,"the film opens with madhav and sreenu playing a football cup final in kolkata. madhav and sreenu are best friends, football players, and n
28 13455076,"the stooges play three sets of identical triplets, born one year apart. all nine brothers lose track of each other after world war ii, un
29 33269085,"ranchhod rai patel is an egotistical and arrogant selfmade gujarati man who arrived from india and settled in the new york area years ago
30 2254451,"preetam , a struggling cartoonist, meets anita at a tennis match, where she is watching her favorite tennis star. anita, a wealthy and wes
31 447194,"samson, a hebrew placed under nazirite vows from birth by his mother, is engaged to a philistine woman named semadar. at their engagement,
```

Train-Test Split: A careful train-test split was performed to prevent data leakage and ensure robust model evaluation. The dataset was divided into training and testing sets, maintaining a balanced representation of genres.

```
... Input: The Quick Brown Fox Jumps Over The Lazy Dog! 123
Output: quick brown fox jump lazy dog

Input:
Output:

Input: None
Output:

Input: the the is are
Output: the

Dataset integrity checks passed.
Total: 41796, Train: 33614 (80.42%), Test: 8182 (19.58%)
Sample Train Row:
  movie_id          31186339
cleaned_summary    nation panem consists wealthy capitol twelve p...
genres_list        (Action, Action/Adventure, Drama, Science Fict...
Name: 0, dtype: object
Sample Test Row:
  movie_id          23890098
cleaned_summary    shlykov hardworking taxi driver lyosha saxopho...
genres_list        (Drama, World cinema)
Name: 0, dtype: object
```



National University of Computer and Emerging Sciences Islamabad Campus

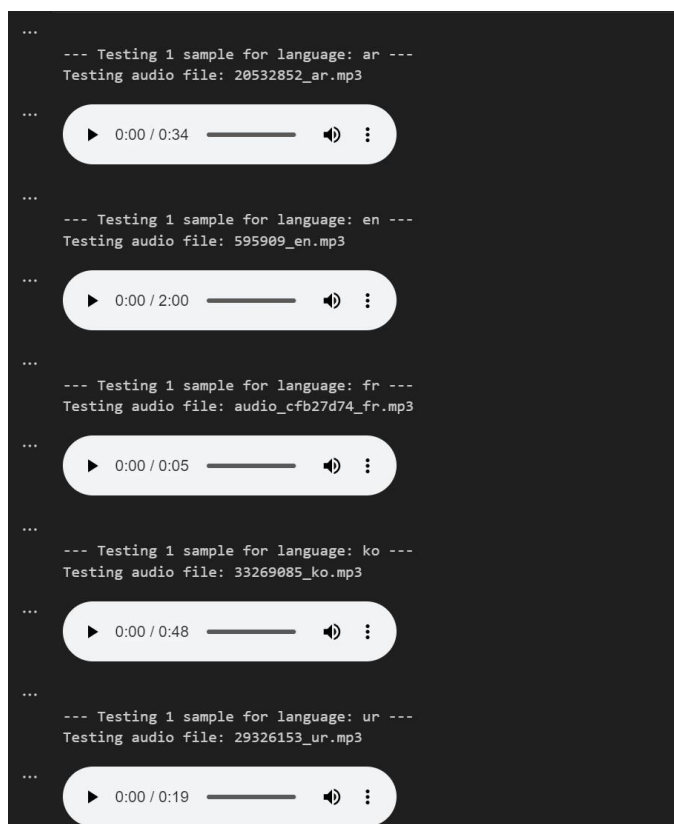
2. Text Translation and Audio Conversion

Text Translation

- Translation Process:
 - The cleaned movie summaries were translated into Arabic, Urdu, and Korean using translation tools such as Google Translate and MarianMT (Hugging Face).
 - A minimum of 50 movie summaries were translated to test the capability to work with multilingual data.

Audio Conversion

- Text-to-Speech (TTS):
 - The translated summaries were converted into audio using TTS engines like gTTS and pyttsx3.
 - The audio files were saved for playback, allowing users to select their preferred language for listening to the movie summaries.



3. Movie Genre Prediction Model

Model Development

National University of Computer and Emerging Sciences

Islamabad Campus

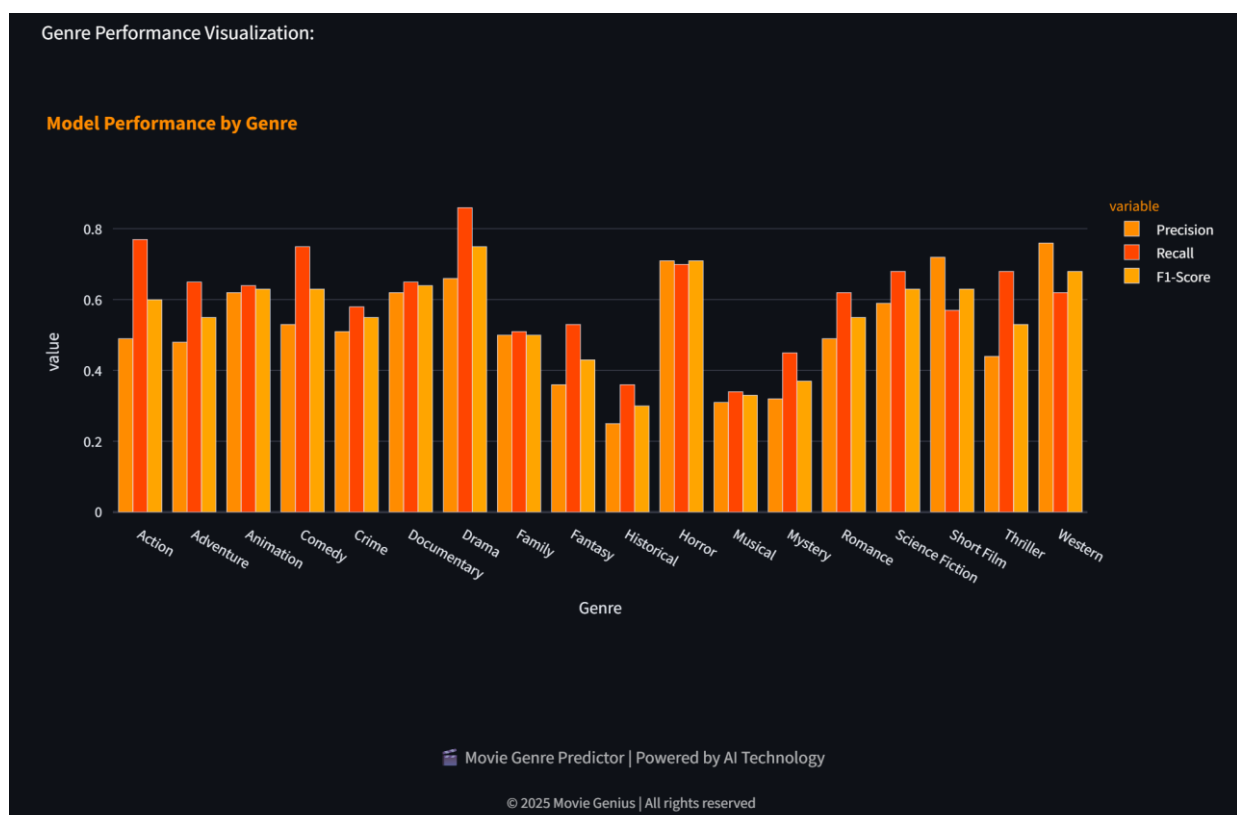
- Machine Learning Model:
 - A multi-label classification model was built to predict movie genres based on their summaries.
 - Various model architectures were considered, including Logistic Regression, Random Forests, and more complex models like LSTMs and Transformer-based models (e.g., BERT).

Feature Extraction

- Techniques Used:
 - Features were extracted from the movie summaries using methods such as TF-IDF and word embeddings (Word2Vec, GloVe).
 - Pre-trained transformers were also utilized for text representation.

Evaluation

- Performance Metrics:
The model's performance was evaluated using:



- Accuracy: Overall accuracy of the model.
- Precision, Recall, F1-Score: Metrics to assess the model's ability to identify specific genres.



National University of Computer and Emerging Sciences Islamabad Campus

- Confusion Matrix: Visualization of classification results for both training and test sets.

```
===== Evaluation =====
```

	precision	recall	f1-score	support
Action	0.49	0.77	0.60	1388
Adventure	0.48	0.65	0.55	1268
Animation	0.62	0.64	0.63	504
Comedy	0.53	0.75	0.63	2495
Crime	0.51	0.58	0.55	1060
Documentary	0.62	0.65	0.64	254
Drama	0.66	0.86	0.75	4438
Family	0.50	0.51	0.50	750
Fantasy	0.36	0.53	0.43	433
Historical	0.25	0.36	0.30	168
Horror	0.71	0.70	0.71	815
Musical	0.31	0.34	0.33	488
Mystery	0.32	0.45	0.37	428
Romance	0.49	0.62	0.55	1506
Science Fiction	0.59	0.68	0.63	467
Short Film	0.72	0.57	0.63	635
Thriller	0.44	0.68	0.53	1372
Western	0.76	0.62	0.68	213
micro avg	0.54	0.69	0.61	18682
macro avg	0.52	0.61	0.56	18682
weighted avg	0.55	0.69	0.61	18682
samples avg	0.56	0.70	0.58	18682

```
Hamming Loss : 0.1103  
Subset Accuracy: 0.1399  
F1 Score (Micro): 0.6077  
F1 Score (Macro): 0.5551  
Precision (Micro): 0.5400  
Precision (Macro): 0.5195  
Recall (Micro): 0.6948  
Recall (Macro): 0.6097  
PS C:\Users\cb26h\Desktop\umer help\model>
```

4. Interactive User Interface (Menu-Based System)

The project features an interactive, menu-driven system designed to enhance user experience. The workflow includes:

Input: Users input a movie summary.

Options:

- Option 1: Convert Summary to Audio ~ Users select the audio language, and the system generates and plays the audio.
- Option 2: Predict Genre ~ Users classify the genre of the movie based on the summary, and the system outputs the predicted genre(s).



The system allows users to explore both features sequentially, enhancing engagement and usability.

The screenshot shows the 'Movie Genre Predictor' web application. At the top, there's a header with the title and a description: 'Discover the genres of any movie based on its plot summary. Get predictions, listen to audio summaries, and more with our AI-powered tool.' Below this is a large input field for the plot summary. Underneath, there's a 'Model Highlights' section with three cards: 'High Accuracy' (Achieves 60.77% F1 Score (Micro) across 18 genres), 'Multi-Genre Support' (Predicts multiple genres simultaneously with confidence scores), and 'Robust NLP' (Uses TF-IDF and neural networks for precise text analysis). The main section is 'Predict Movie Genres', which includes a 'PREDICT GENRES' button. Below the button, there's a green box indicating 'Predicted Genres:' followed by a list of genres: Drama, Horror, Mystery, Psychological thriller, Thriller, and World cinema. At the bottom, there's a 'Confidence Scores' table.

	GENRE	CONFIDENCE
0	Drama	87.55%
1	Horror	71.92%
2	Mystery	56.54%
3	Psychological thriller	52.84%
4	Thriller	83.04%
5	World cinema	56.52%

- There is translation service with deepl API that gives accurate translation



National University of Computer and Emerging Sciences Islamabad Campus

>

RUNNING... Stop Deploy

Convert Summary to Audio

Convert your plot summary to audio in your preferred language.

Choose Language

KR Korean (Clear Hangul pronunciation)

Speech Speed

Volume

0.50 1.00 2.00 0.10 0.50 1.00

Original Text:

Opening Scene: A gothic cathedral in 19th-century London, bathed in flickering candlelight. LUCIA (mid-20s, sharp-witted, a thief with a tragic past) steals an ornate silver locket from a nobleman's corpse during a funeral. As she touches it, ghostly whispers flood her mind, revealing a vision of a hidden royal crypt beneath the city. LUCIA (muttering): "Bloody hell. Not this again." Cut to flashback: Lucia as a child in a Roma caravan, her grandmother warning her of the "Veil"—a thinning barrier between the living and dead. Lucia's family was slaughtered by shadowy figures seeking the locket. Present day: Lucia is hunted by the Order of the Hollow Crown, a secret society of aristocrats who use the dead as spies. She teams up with ELIAS (40s, a disgraced ex-detective turned occultist) to decipher the locket's secrets. Their investigation reveals a plot: the Order plans to resurrect a long-dead king to restore absolute monarchy, using a ritual at the next lunar eclipse.

Midpoint twist: Lucia's locket is one of three "Keystones" needed for the ritual. The others are held by a rival thief (a charming rogue with his own agenda) and the Order's leader, LORD VESTHAM (50s, a charismatic villain who believes tyranny is humanity's natural state).

Climax: Lucia and Elias infiltrate the crypt during the eclipse, fighting reanimated corpses and Vestham's zealots. Lucia must choose: destroy the locket to stop the ritual (and sever her last link to her family's legacy) or use its power to control the dead—risking her sanity.

LUCIA (gripping the locket): "I won't let the past own me."

The ritual is disrupted, but the final shot shows Vestham's corpse twitching—hinting the dead king's spirit escaped.

GENERATE AUDIO

Translated Text (Korean):

오프닝 장면: 19세기 런던의 고딕 양식 대성당, 깜빡이는 촛불이 환하게 비춥니다. 루시아(20대 중반, 예리한 재치, 비극적인 과거를 가진 도둑)는 장례식 도중 귀족의 시신에서 화려한 은제 목걸이를 훔칩니다. 그것을 만지자 유령의 속삭임이 그녀의 머릿속을 가득 채우며 도시 아래에 숨겨진 왕실의 지하실에 대한 환상을 드러냅니다. 루시아 (중얼거림): "빌어먹을. 또 이러면 안 돼." 플래시백으로 전환됩니다: 로마의 캐러밴에 타고 있던 어린 시절의 루시아, 할머니가 이승과 저승 사이의 얇은 장벽인 '베일'에 대해 경고하는 모습. 루시아의 가족은 로켓을 노리는 어둠의 세력에 의해 학살당했습니다. 현재: 루시아는 죽은 자를 스파이로 이용하는 귀족들의 비밀 단체인 할로우 크라운 교단에게 쫓기고 있습니다. 그녀는 불명예스러운 전직 형사에서 오컬티스트로 변신한 40대 엘리아스와 팀을 이루어 로켓의 비밀을 해독합니다. 조사 결과, 이 교단은 다음 월식 때 의식을 통해 죽은 왕을 부활시켜 절대 군주제를 회복하려는 음모를 꾸미고 있다는 사실을 알게 됩니다.

Conclusion

The Filmception project successfully integrates data preprocessing, multilingual translation, audio conversion, and genre classification into a cohesive system. The interactive user interface further enhances user engagement, making it a comprehensive tool for movie summary translation and genre prediction. Future work may involve refining the model and expanding the dataset for improved accuracy and user experience.