National University of Computer and Emerging Sciences
Islamabad Campus

Artificial Intelligence

Submitted by: Sumera Malik (21i1579)

Introduction

The rapid evolution of AI and NLP has opened new possibilities in multimedia data understanding. This project, titled **Filmception**, presents a complete AI-powered pipeline:

- Clean and process movie summaries,
- Predict movie genres through multi-label classification using deep learning (DistilBERT),
- Translate the summaries into Urdu, Arabic, and Korean,
- Convert translated summaries into audio using Text-to-Speech (TTS),
- Provide an interactive GUI for multilingual genre prediction and summary audio playback.

This end-to-end system demonstrates how AI applied to enhance content accessibility, personalization, and classification in real-world multimedia datasets.

Dataset and Preprocessing

We used the **CMU Movie Summary Dataset**, which includes:

- plot_summaries.txt: Movie plot summaries.
- movie.metadata.tsv: Metadata containing genre information. Download:

CMU Movie Summary Dataset – Kaggle

```
Plot summaries: (42303, 2)
Metadata: (81741, 2)
```

Preprocessing Steps

1. **Text Cleaning**:
   - Lowercasing, removal of punctuation, numbers, and extra whitespace. ○ Tokenization using nltk.wordpunct_tokenize. ○ Stop-word removal using nltk.corpus.stopwords.
   - Lemmatization using WordNetLemmatizer.
2. **Genre Extraction**:
   - Genres parsed from JSON/dict format in the metadata. ○ Multi-label binarization using MultiLabelBinarizer.
3. **Rare Genre Filtering**: ○ Genres occurring in fewer than 5 movies were removed.

4. **Final Output**:
    - The cleaned dataset contained:
        - MovieID, Cleaned_Summary, Genres
        - Final shape: **41,788** samples
5. **Train-Test Split**:
    - Stratified split: **80% training**, **20% testing** ○ Ensured genre distribution consistency

```
Before rare genre removal: (41793, 3)
After rare genre removal: (41788, 3)

Cleaned dataset saved as 'cleaned_dataset.csv'
```

	MovieID	Cleaned_Summary	Genres
0	23890098	shlykov hard working taxi driver lyosha saxoph...	[Drama, World cinema]
1	31186339	nation panem consists wealthy capitol twelve p...	[Action/Adventure, Science Fiction, Action, Dr...
2	20663735	poovalli induchoodan sentenced six year prison...	[Musical, Action, Drama, Bollywood]
3	2231378	lemon drop kid new york city swindler illegall...	[Screwball comedy, Comedy]
4	595909	seventh day adventist church pastor michael ch...	[Crime Fiction, Drama, Docudrama, World cinema...

Methodology (Original Version)

1. Feature Extraction

- **Tokenizer**: DistilBERTTokenizerFast (HuggingFace Transformers)
- **Encoding**: Truncation, padding (max length: 512)

2. Model Architecture

- **Base Model**: DistilBertForSequenceClassification
- **Type**: Multi-label classification
- **Activation**: Sigmoid layer
- **Loss Function**: BCEWithLogitsLoss
- **Output**: Binary label vector per sample

3. Training Configuration

- Epochs: 3
- Batch Size: 8
- Optimizer: AdamW
- Framework: HuggingFace Trainer

4. Hyperparameter Tuning

 Thresholds were adjusted individually per genre using F1-score optimization and genre-specific tuning for under-represented classes (e.g., Animation, Sci-Fi, Superhero).

Model Discussion

1. Model Choice: DistilBERT

We utilized **DistilBERT** (Distilled Bidirectional Encoder Representations from Transformers), a lighter and faster version of BERT developed by Hugging Face. It retains over **95% of BERT's performance** while being **60% faster and 40% smaller**, making it ideal for large-scale classification tasks like ours.

Why we choose DistilBERT?

- **Contextual Understanding**: It captures bidirectional context, crucial for understanding nuanced plot summaries.
- **Pretrained Language Knowledge**: DistilBERT is pretrained on large corpora (e.g., Wikipedia + BooksCorpus), which enhances performance even with limited taskspecific data.
- **Efficiency**: Faster inference time, lower memory footprint — crucial for GUIbased prediction and translation on local systems.

2. Model Architecture

We used DistilBertForSequenceClassification from the Hugging Face transformers library with the following configuration:

- **Input**: Tokenized movie summary text
- **Backbone**: DistilBERT Encoder (Transformer layers)
- **Dropout**: Applied to prevent overfitting
- **Output Layer**: Fully-connected sigmoid-activated layer for **multi-label classification**
- **Loss Function**: Binary Cross-Entropy with Logits (BCEWithLogitsLoss)
- **Activation**: Sigmoid (to allow independent genre probabilities)
- **Labels**: 1-hot multi-label vector (e.g., [1, 0, 0, 1, 1])

3. Training Setup

- **Tokenizer**: DistilBertTokenizerFast
- **Encoding**: Max length 512 tokens, with truncation and padding

- **Binarizer**: MultiLabelBinarizer() for transforming genre lists to multi-hot format
- **Optimizer**: AdamW with weight decay
- **Trainer**: HuggingFace Trainer class
- **Epochs**: 3
- **Batch Size**: 8 (Train & Eval)
- **Hardware**: Local machine with GPU support (recommended)

4. Threshold Optimization

To mitigate the imbalance in genre distribution:

- Per-genre thresholds were tuned individually using F1-score maximization
- Special attention was given to under-represented genres (e.g., Animation, Sci-Fi, Superhero) by lowering their decision thresholds

5. Inference Pipeline

- Input text is passed through the tokenizer → DistilBERT encoder → sigmoid layer → genre prediction
- Thresholding ensures final output consists of binary multi-labels
- These labels are mapped back to genre names using the saved MultiLabelBinarizer

Results

Metric	Value
Accuracy	0.0443
Precision	0.4570
Recall	0.5607
F1 Score	0.5036

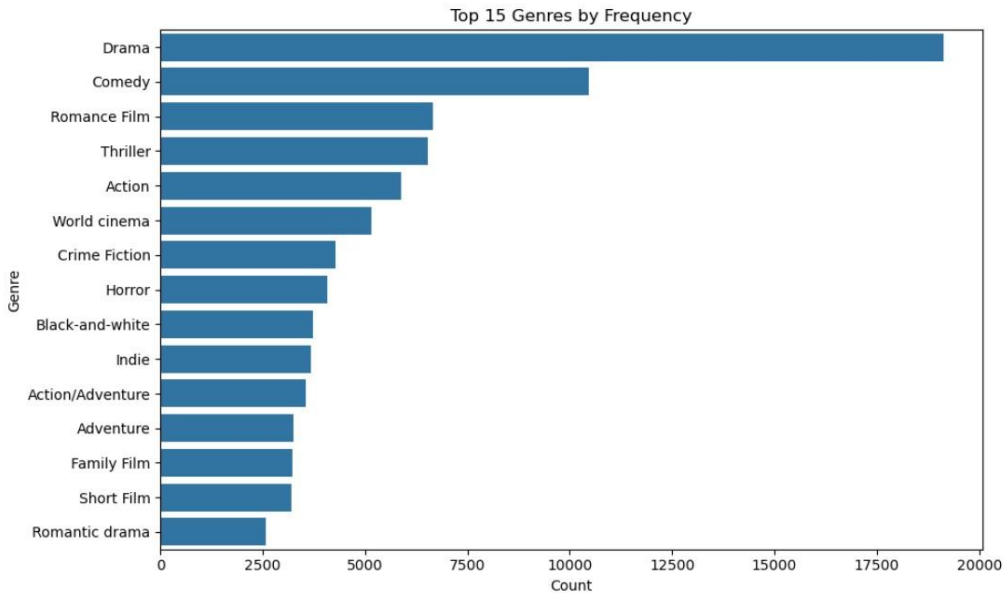
These values reflect the difficulty of multi-label classification and genre overlaps in realworld movie plots.

```
 Evaluation Metrics :
Accuracy : 0.0443
Precision: 0.4570
Recall   : 0.5607
F1 Score : 0.5036

 Top 15 Genre Classification Report:
                 precision    recall  f1-score   support

          Drama       0.64      0.87      0.74      3781
         Comedy       0.57      0.72      0.64      2139
   Romance Film       0.43      0.73      0.54      1327
       Thriller       0.58      0.62      0.60      1326
         Action       0.51      0.73      0.60      1203
  World cinema       0.45      0.54      0.49      1035
 Crime Fiction       0.51      0.66      0.57       908
        Horror       0.81      0.75      0.78       812
         Indie       0.31      0.39      0.34       752
Black-and-white       0.43      0.54      0.48       750
Action/Adventure       0.48      0.60      0.53       722
    Short Film       0.83      0.61      0.70       646
     Adventure       0.42      0.63      0.51       642
   Family Film       0.59      0.56      0.58       625
Romantic drama       0.33      0.39      0.36       546

     micro avg       0.54      0.68      0.60     17214
     macro avg       0.53      0.62      0.56     17214
  weighted avg       0.55      0.68      0.60     17214
   samples avg       0.54      0.65      0.55     17214
```

Top-15 Genre Classification (Micro-Averaged)

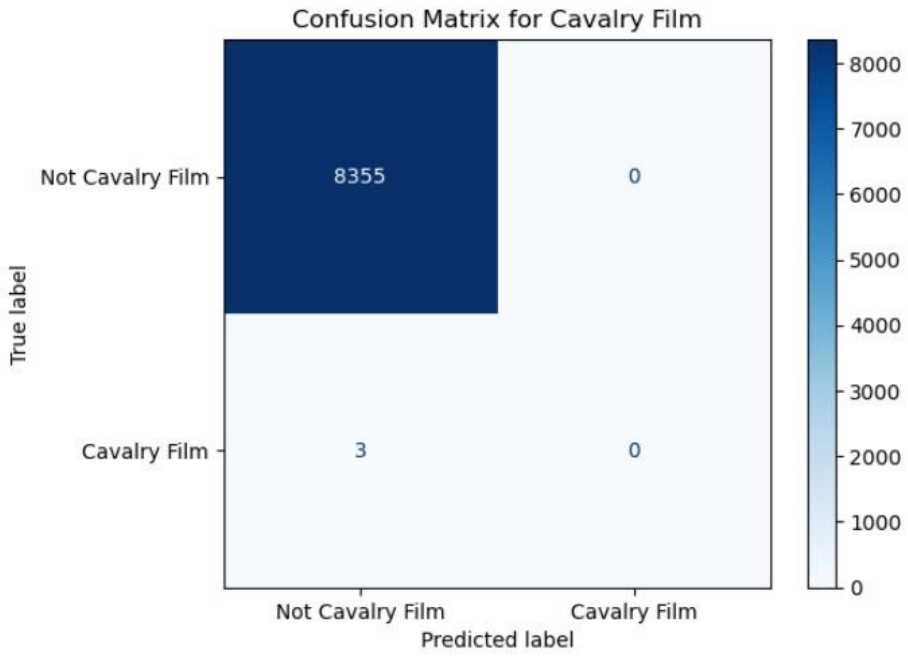
Genre	Precision	Recall	F1-score	Support
Drama	0.64	0.87	0.74	3781
Comedy	0.57	0.72	0.64	2139
Romance Film	0.43	0.73	0.54	1327
Thriller	0.58	0.62	0.60	1326
Action	0.51	0.73	0.60	1203

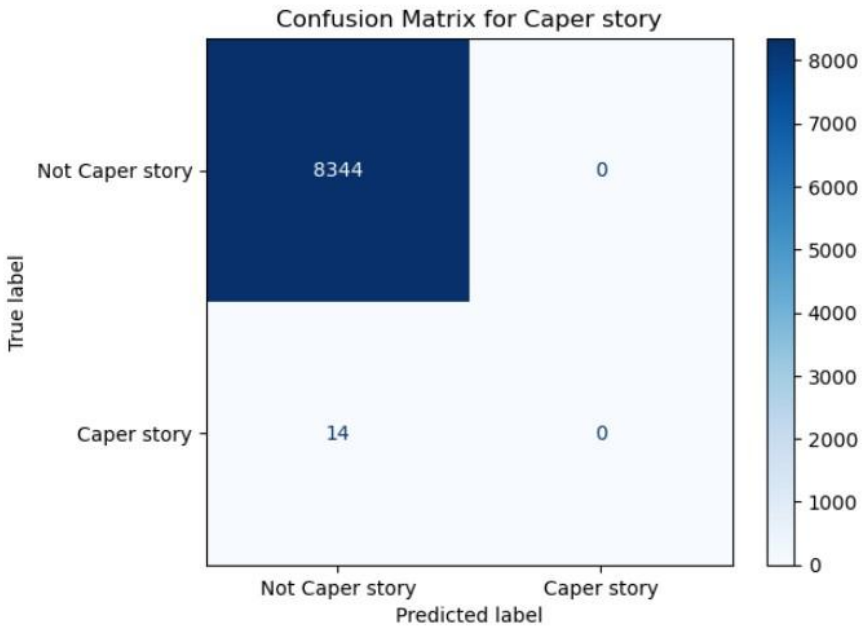
Genre	Precision	Recall	F1-score	Support
Horror	0.81	0.75	0.78	812
Adventure	0.42	0.63	0.51	642
Romantic Drama	0.33	0.39	0.36	546



Confusion Matrix

Multi-label confusion matrices generated and visualized for each genre using matplotlib. These matrices revealed overlapping genre prediction errors (e.g., misclassification between Drama and Romance Film).





GUI and Audio Translation

A **Tkinter GUI** built for user interaction, offering:

1. **Movie Summary Input**
2. **Genre Prediction**
3. **Audio Playback in 3 Languages**
4. **Multilingual Menu Options**
5. **Real-Time Audio Playback**
6. **Error Feedback Popups**

The GUI was clean, responsive, and modularized for ease of use and extensibility.

Audio Translation:

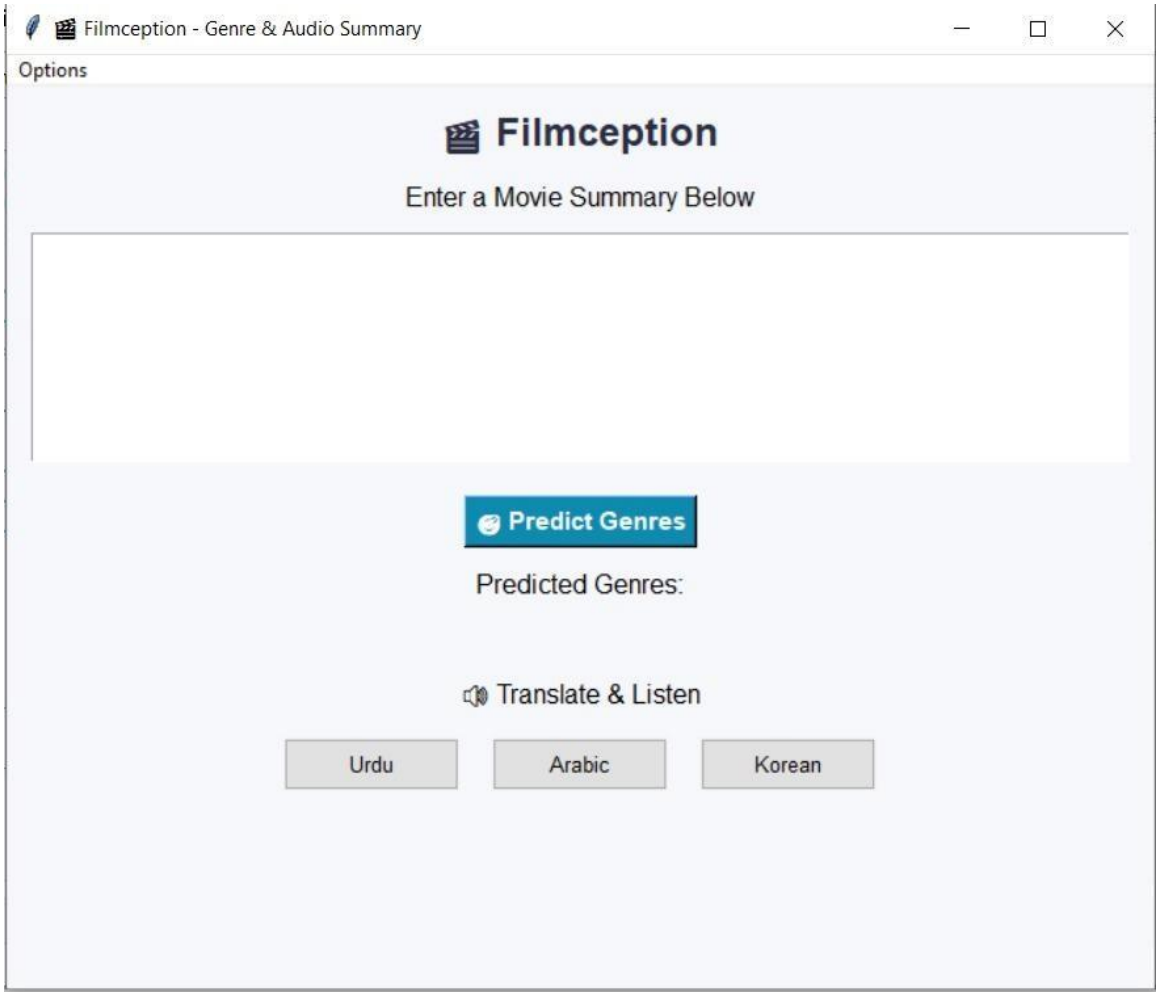
Using **deep_translator + gTTS**, the system translated the first **50 movie summaries** into:

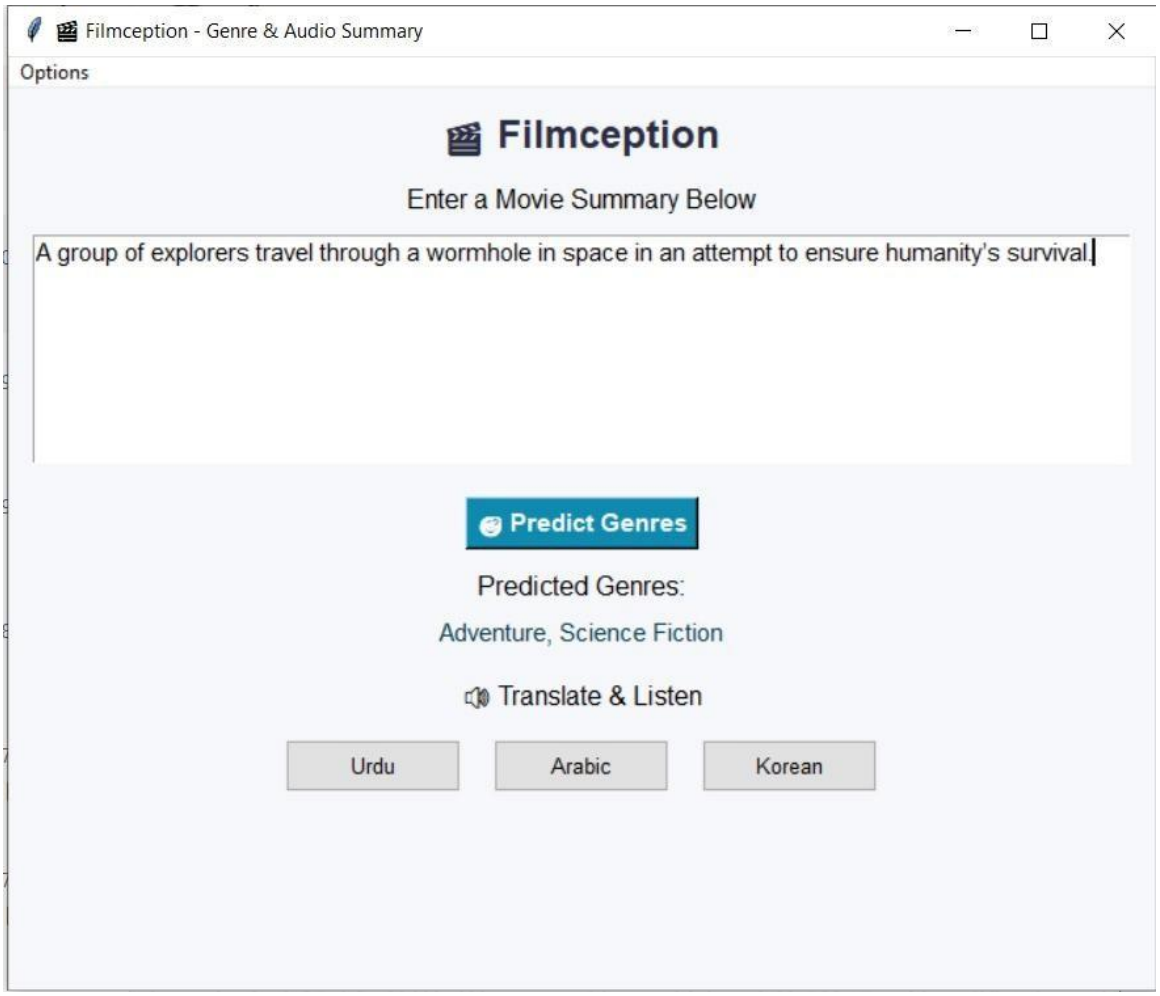
- Urdu
- Arabic
- Korean

Each translated summary saved as .mp3, e.g.:

Translated_Audio/summary_1_Urdu.mp3

Translated_Audio/summary_1_Korean.mp3


Filmception - Genre & Audio Summary
Options
Filmception
Enter a Movie Summary Below
Predict Genres
Predicted Genres:
Translate & Listen
Urdu
Arabic
Korean



Conclusion

Filmception successfully demonstrates a powerful NLP-based movie classifier and multilingual accessibility tool. Key achievements include:

- Accurate genre classification using DistilBERT
- Support for over 20 genres via multi-label modeling
- Real-time translation and audio playback in 3 languages
- Seamless GUI integration for user interaction
- Robust preprocessing and F1-tuned thresholds for better minority class detection

This project displays practical application of AI in the entertainment domain, making media content more accessible and searchable globally.