

# Movie Finder

**Group: 08**

**Patrick Eckel:** Design Document, Frontend, Data Transfer

**Marcus Gugacs:** Design Document, Frontend, Recommender, CLI, Evaluation, Questionnaire, Report, Presentation

**Martin Tobias Klug:** Design Document, Subtitle Fetching, Summarization Pipeline, Report

**Lukas Leitner:** Design Document, Data Preprocessing, Report, Presentation

**<https://www.github.com/Ilmpaq/air-2024>**

**01.2025**

# Introduction

## Motivation:

- Lot's of video content online across different streaming services
- Central place to get personalized recommendations to value user time

# Data

## Movie Dataset:

- 180.000 Rows left after preprocessing
- Columns reduced from 20 to 11 + rich features

## Subtitle Data:

- Subtitles provided by API (Key required)
- Raw Subtitles downloaded & processed on demand

# Methods

## Sequence Transformer:

- Model: all-mpnet-base-v2
- Used to similarity between user query and movie features

## Emotion Classifier:

- Model: emotion-english-distilroberta-base
- Classify emotions in english text to measure alignment with user mood

## Movie Introduction Summarization:

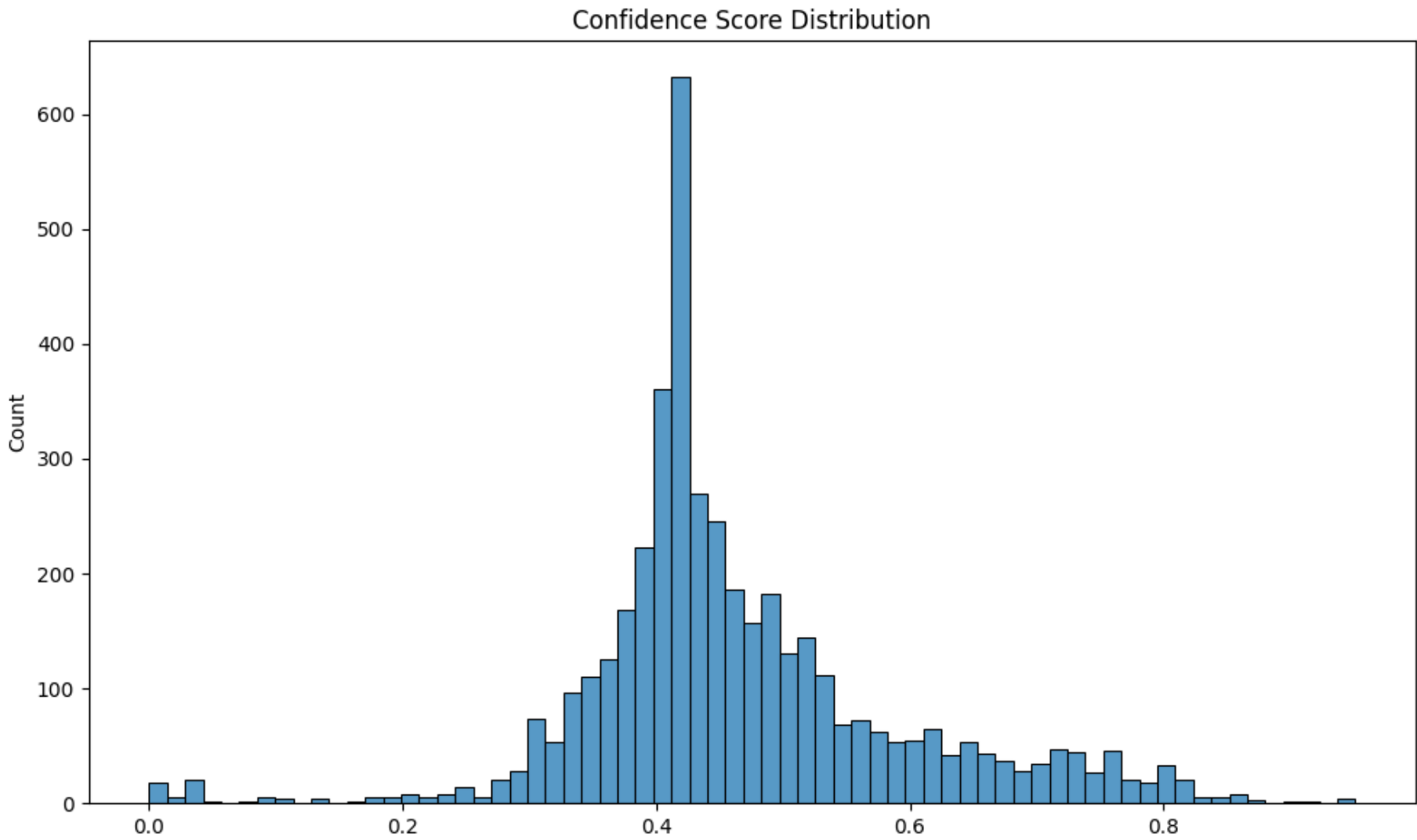
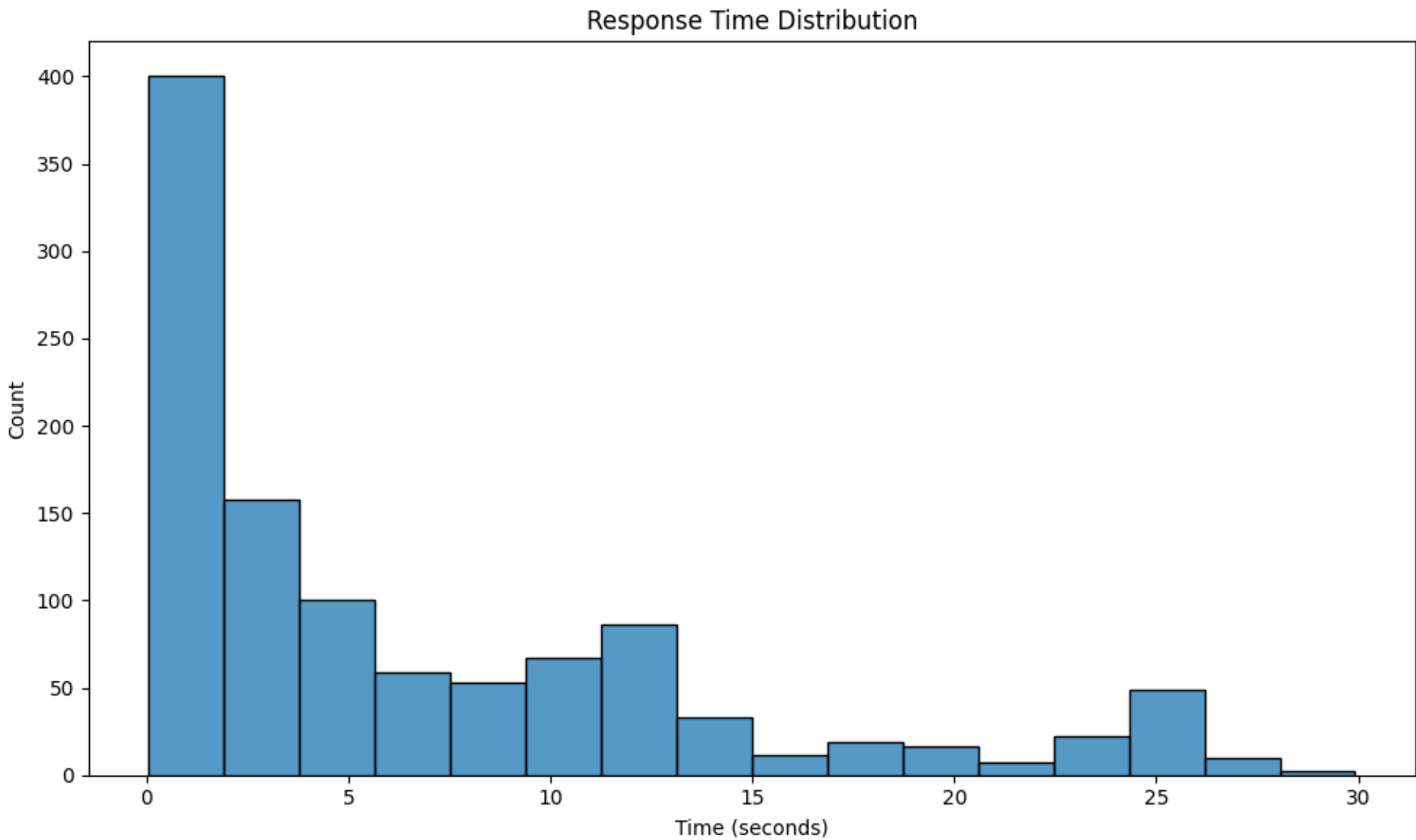
- Model: facebook/bart-large-cnn
- Creates introduction from first 1024 tokens of subtitles

## Keyword Extraction:

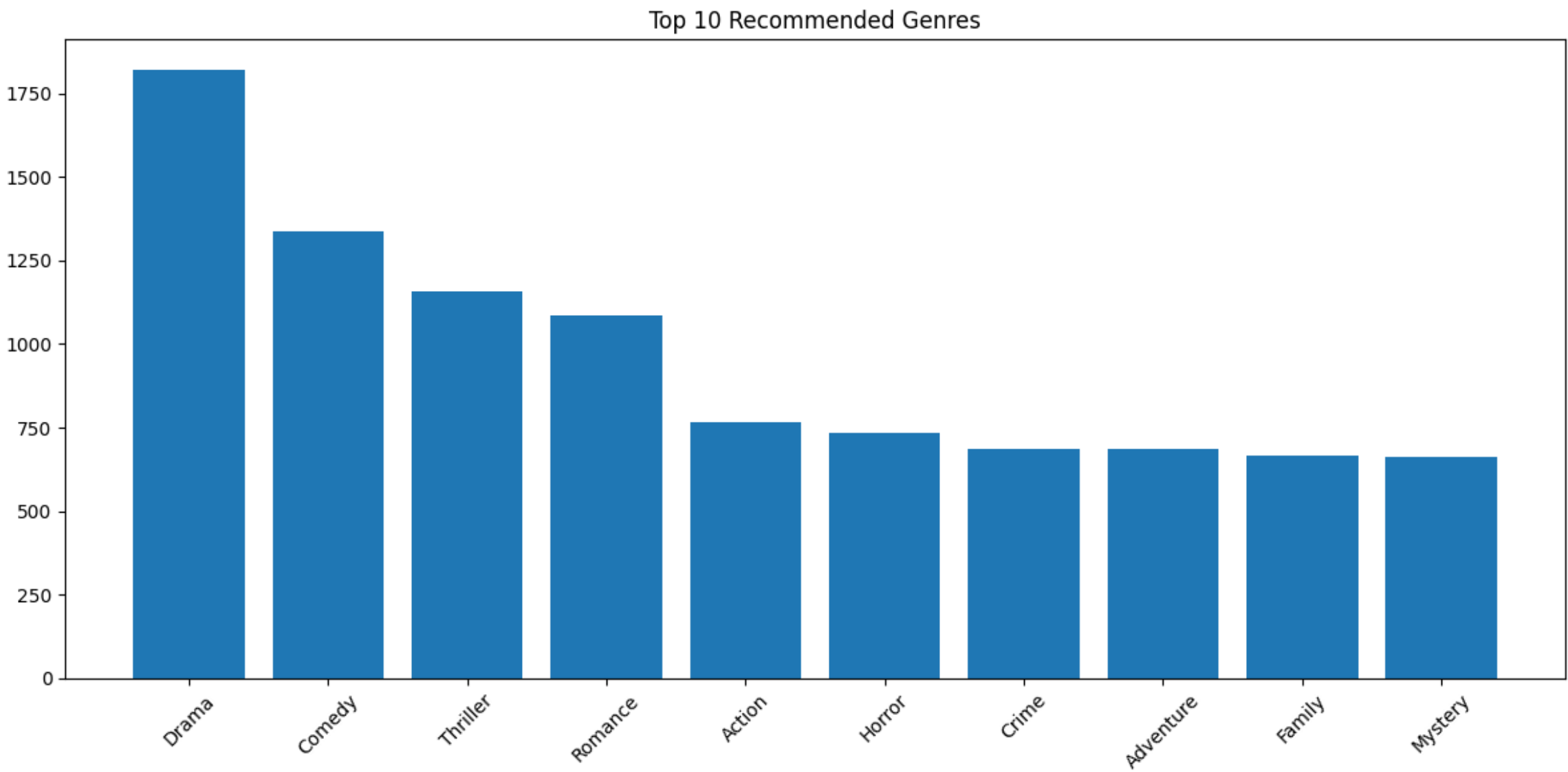
- Model: KeyBERT
- Extracts key themes from cleaned movie subtitles

# Results

	Average
Average Precision (in %)	99.08
Average Recall (in %)	10.10
Average F1-Score (in %)	14.38



	Average	Standard Deviation	Min	Max	Median
Response Time (in seconds)	6.69	7.32	0.06	29.92	3.60
Genre Diversity	3.8	0.51	1	4	4
Confidence (in %)	46.55	12.94	0.00	95.13	43.36
Rating	6.84	0.81	3.00	9.75	9.60



# Conclusion

## **Findings:**

- Usable and efficient recommendations
- Tweaking and fine-tuning for further improvements
- Minor tweaks in testing lead to significant changes

## **Limitations:**

- Limited of movies used per recommendation
- Limited tokens used per introductory summary