SE 3006 Software Architectures
Project Progress Report
TEAM 8

Report Summary

Between February 26th and March 19th, 2025, several separate meetings were held as part of the project development process. The first meeting took place physically in the classroom, while the others were conducted either online or in a school environment. The main purpose of these meetings was to plan the project workflow, assign responsibilities, and establish the technical foundation.In this report, the progress made so far and the tests conducted have been explained along with graphs, and an interim report has been created.

Initial Planning

During the first meeting, the team decided on what tasks to undertake:

- **Halil İbrahim Akça**
- **Elif Yıldız**
- **Ramazan Ediz**
- **Talha Çetinkaya**

It was decided that team communication would be conducted via WhatsApp. Additionally, it was noted that more research was needed to finalize the project structure.

Technical Discussion

The second meeting focused on technical discussions. The potential of the Whisper API for transcribing audio files to text was evaluated in detail. Pre-trained models available on Hugging Face (such as T5, BART, GPT, and LLaMA) were discussed in terms of their integration into the project. Steps for deploying the application on a web platform were also planned. The meeting concluded with the decision to prepare datasets and further investigate APIs and similar tools.

Model Selection and Development

In the third meeting, previous technical research was expanded upon and model selection was finalized. Analyses using the Whisper API for audio-to-text transcription were reviewed, and the integration process of Hugging Face models was discussed in more detail. During this meeting:

- User interface development was initiated.
- Backend progress was shared.
- Preparations were made for future Computer Vision tasks.

Challenges and Experimentation

Several challenges were encountered while using Whisper. Initially, audio was extracted from a meeting video using `VideoFileClip`, and transcription was attempted using VOSK. Due to unsatisfactory results, a more advanced VOSK model was tried, but performance remained inadequate. Whisper was then retried and yielded more promising results.

```python
from moviepy import VideoFileClip


video_path = "videoplayback.mp4"
audio_path = "audio.wav"

video = VideoFileClip(video_path)
video.audio.write_audiofile(audio_path)

print("Ses başarıyla kaydedildi:", audio_path)
```

Figure 1: Extracting sound from video recording

```python
import whisper
import os

os.environ["PATH"] += os.pathsep + r"C:\ffmpeg"

model = whisper.load_model("base")

result = model.transcribe("audio.wav")

with open("toplanti_metni.txt", "w", encoding="utf-8") as f:
    f.write(result['text'])

print("Metin 'toplanti_metni.txt' dosyasına kaydedildi.")
```

Figure 2: Converting sound to text and recording it

The transcript generated by Whisper was summarized using the `T5-small` model from Hugging Face. Sentiment analysis was conducted using the `CardiffNLP` sentiment model,

and predictions were evaluated. Additionally, audio-based emotion detection was attempted using a pre-trained model specialized for voice recordings.

```python
import re
from transformers import pipeline
import matplotlib.pyplot as plt
from collections import Counter

sentiment_analyzer = pipeline("text-classification", model="cardiffnlp/twitter-roberta-base-emotion")

with open("toplanti_metni.txt", "r", encoding="utf-8") as f:
    metin = f.read()

cumleler = [c.strip() for c in re.split(r'(?<=[.!?])\s+', metin) if c.strip()]

results = [sentiment_analyzer(cumle)[0] for cumle in cumleler]

with open("duygu_analizi_sonuclari.txt", "w", encoding="utf-8") as f_out:
    for cumle, res in zip(cumleler, results):
        output = f"Cümle: {cumle}\nDuygu: {res['label']}, Güven: {res['score']:.2f}\n\n"
        f_out.write(output)
```

Figure 3: Extracting emotion from text

```python
if __name__ == "__main__":

    analyzer = AudioEmotionAnalyzer()

    result = analyzer.predict_emotion(r"C:\Users\akcah\Desktop\arhtecture\audio.wav", visualize=True)
    if result:
        print(f"Predicted emotion: {result['emotion']} with {result['confidence']*100:.2f}% confidence")

    print("\nAudio Emotion Analyzer Ready!")
    print("Use analyze_directory() to process multiple files or predict_emotion() for a single file.")
```

Figure 4: Extracting emotion from sound

```python
from transformers import pipeline


summarizer = pipeline("summarization", model="t5-small")

with open("toplanti_metni.txt", "r", encoding="utf-8") as f:
    toplanti_metni = f.read()

# 1024 tokenlık parçalara bölüp özetliyoruz
chunks = [toplanti_metni[i:i+1000] for i in range(0, len(toplanti_metni), 1000)]
summary = " ".join([summarizer(chunk, max_length=200, min_length=50, do_sample=False)[0]['summary_text'] for chunk in chunks])

print("Toplantı Özeti:\n", summary)
```

Figure 5: General summary with T5 model

Emotion Analysis Visualization: Audio vs Text

The two images show different methods of emotion analysis:

Image 1: Emotion Analysis with Audio Waveform

At the top, there is the audio waveform, showing the changes in the sound over time (about 1:35 minutes).

At the bottom, there is a graph showing the probability of different emotions detected in the audio.

The dominant emotion here is ”fear” (13.7).

Other emotions are similarly distributed: sad, happy, angry, calm, neutral, surprised, and disgust.

The emotions are fairly evenly distributed, with none being strongly dominant.

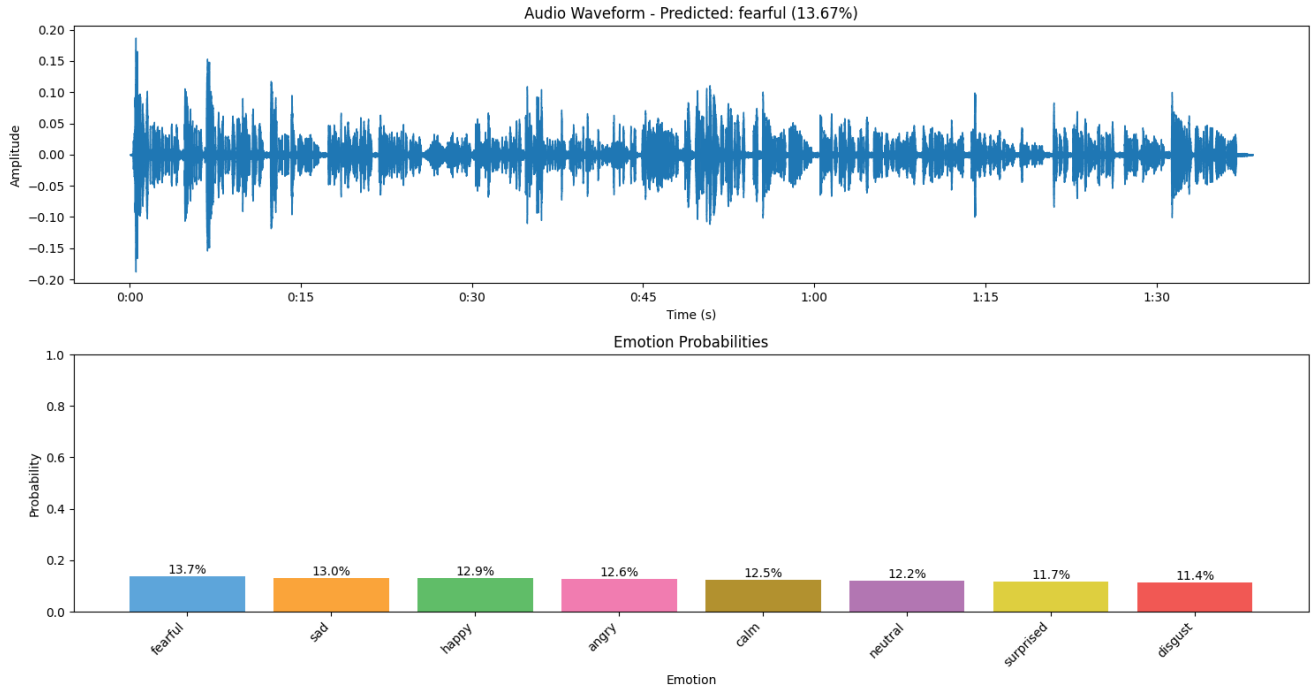

Figure 6: Image 1: Extracting emotion from text

Image 2: Sentiment Distribution

This chart shows the probability distribution of four emotions: optimism, anger, sadness, and joy.

”Sadness” has the highest probability (42).

”Anger” follows closely (39).

”Optimism” is lower (14), and ”Joy” has the least probability (3).

Negative emotions (sadness and anger) are clearly dominant here.İ

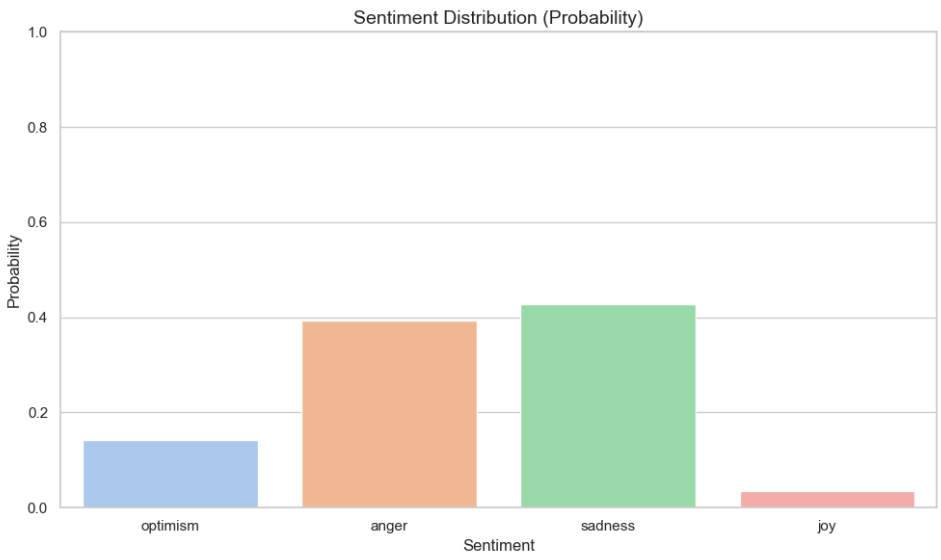

Figure 7: Image 2: Extracting emotion from text

Key Differences:

- Image 1 analyzes audio data, while Image 2 analyzes text. - Image 1 shows 8 different emotions, with a more balanced distribution. Image 2 focuses on negative emotions. - Image 1 uses more granular emotion categories, while Image 2 focuses on broader sentiment categories.

Conclusion and Outlook

The results obtained so far are considered promising but open to improvement. Overall, the project is progressing in a planned and research-oriented manner:

- Tool and model selection has largely been finalized.
- It has not been added because we do not have a prototype for the user interface work yet.
- Valuable experiments have been conducted in audio processing and analysis.

Future meetings will focus on model efficiency, integration challenges, and performance optimization.