



Final Pipeline Presentation

Group 15, polish 

Block C Year 2 2024-2025

CREATING MEANINGFUL EXPERIENCES

Domain & Use Case

Context:

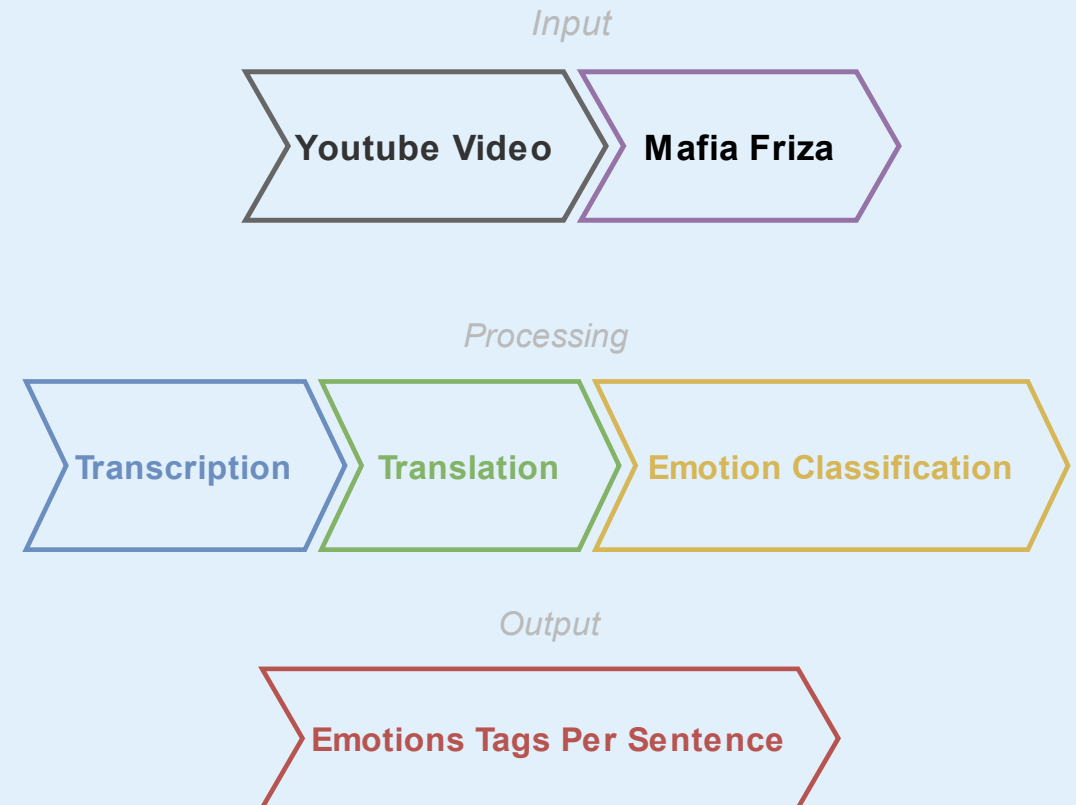
- Analyze Polish YouTube transcriptions from “Mafia Friza”—a high-stakes social deduction game with debates, accusations, and emotional outbursts.
- Data is recorded in Polish and translated to English.

Use Case:

- The Content Intelligence Agency needs reliable, multilingual emotion classification.
- Our pipeline transcribes, translates, and tags emotions per sentence, operating independently of external services.

Added Value:

- Enhances audience analysis and content planning by detecting emotional trends, offering a robust, self-contained solution for consistent, accurate emotion classification.



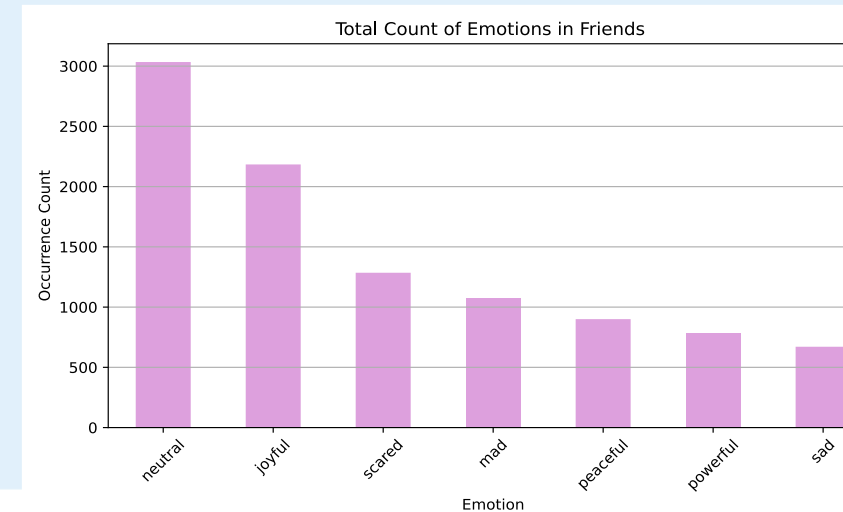
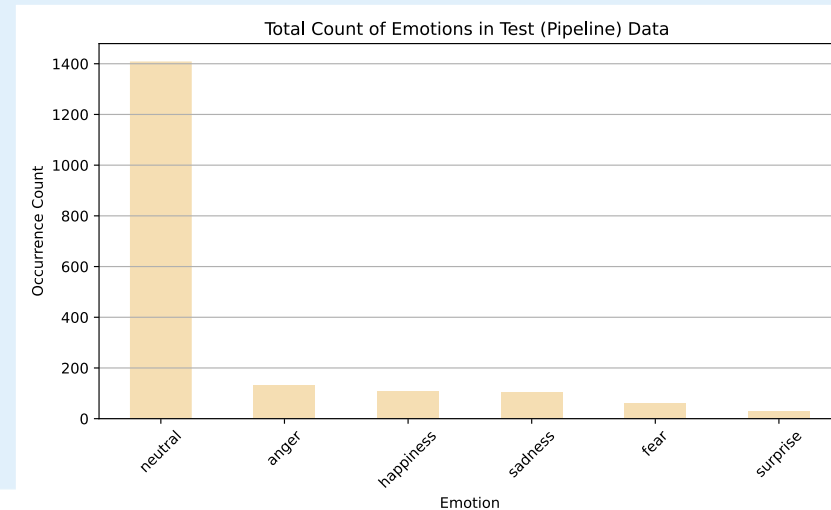
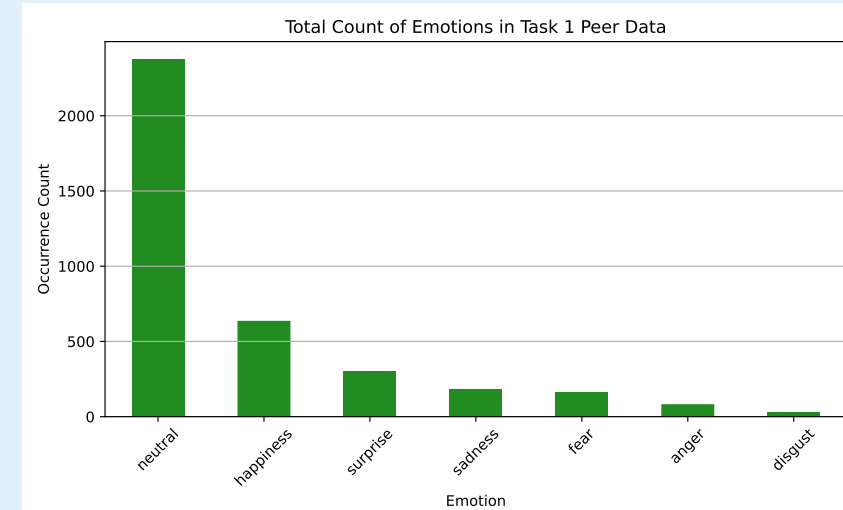
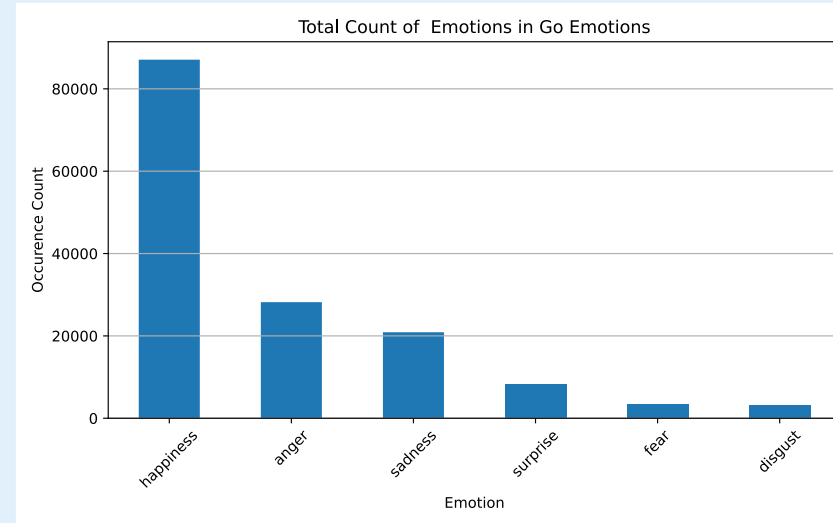
Data Characteristics

Datasets:

- **GoEmotions:** 151,344 rows
- **Task 1 (Annotated):** 3,855 rows
- **Pipeline Data (Test):** 1,845 rows
- **Friends Dataset:** 9,934 rows

Data Distribution & Class Imbalance:

- Class frequencies vary significantly.
- Neutral is overrepresented.
- Emotions like fear, disgust, and surprise are underrepresented.



Data Characteristics - Limitations

Transcription Issues:

- Noisy, overlapping dialogue (e.g., misinterpreted nicknames/slang) leads to errors.

Translation Loss:

- Machine translation from Polish to English dilutes emotional nuance.

Annotation Variability:

- Inconsistent, incomplete corrections reduce reliability.

Class Imbalance:

- Neutral dominates; fear, disgust, and surprise are underrepresented. (only 3 samples of disgust in test set)

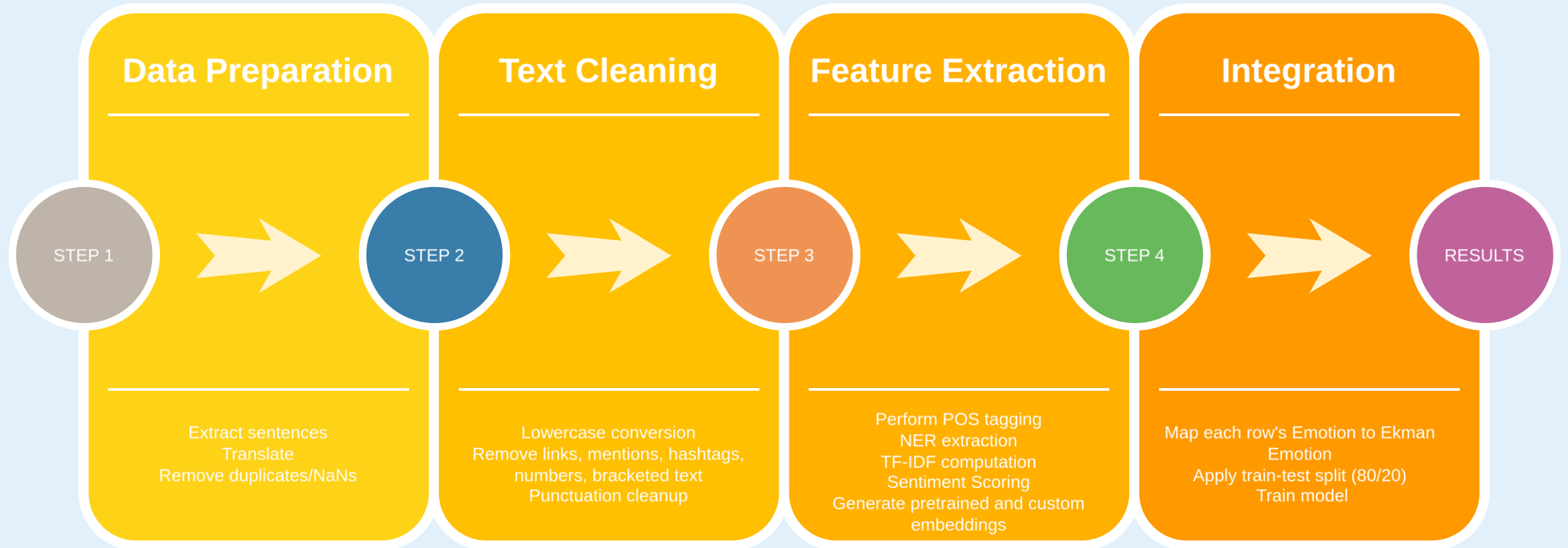
Errors in transcription:

- Residual noise complicating analysis.

Test Data Anomalies from Pipeline Output

Original Sentence	Translation	Type of error
patrz teraz, słuchaj to. Żaba robi	Look now, listen. Frog does	Incorrect nickname translation
A ja mi się wydaje że fusialka, bo ona w tym	And I think it is more fusial, because she played ...	Incorrect nickname translation
To ch**o.	To ch**o.	Did not translate to English
Co?	Co?	Did not translate to English
Tu?	Tu?	Did not translate to English
Tak	Not.	Incorrectly translated to English
A teraz przyszedł czas na rozdanie kart.	Error 500 (Server Error)!!1500.That's an error.There was an error. Please try again later.That's all we know.	Error occurred
Cisnąłeś o żabię.	You threw about the frog.	Incorrectly translated sentence and nickname to English

Data Characteristics - Processing



Evaluation Metrics

Model	Evaluation Metric
Transcription – AssemblyAI	WER rate – 1.63%
Translation - Helsinki	BLEU score – 31.2
Classification - Distillroberta	Weighted F1 Score – 0.66

- ✓ WER (Word Error rate) - shows how many words were wrong in the transcription
- ✓ BLEU Score – checks how close the translation is to 'a good one'
- ✓ Weighted F1 Score – measures how well the model handles all categories

Model choice

Classification models:

Model	Best weighted F1 score
Naïve Bayes	0.31
LSTM	0.27
RNN	0.14
Transformer (DistilRoberta)	0.66

Translation model:

Model	BLEU score
Transformer (Helsinki-NLP)	31.2

Transcription model:

Model	WER
Whisper	2,44%
AssemblyAI	1,63%

BLEU Score	Interpretation
< 10	Almost useless
10 - 19	Hard to get the gist
20 - 29	The gist is clear, but has significant grammatical errors
30 - 40	Understandable to good translations
40 - 50	High quality translations
50 - 60	Very high quality, adequate, and fluent translations
> 60	Quality often better than human

Evaluate models | Cloud Translation. (n.d.).

Google Cloud.

<https://cloud.google.com/translate/docs/advanced/automl-evaluate>

$$\text{WER} = \frac{S + D + I}{N}$$

Methodology



Video

mp4 file



Transcription

Method:
AssemblyAI

Columns:
Start Time, End Time,
Sentence



Translation

Method:
Transformer model
(Helsinki-NLP - finetuned)

Columns:
Start Time, End Time,
Sentence, Translation



Emotion classification

Method:
Transformer model
(DistilRoberta - finetuned)

Columns:
Start Time, End Time,
Sentence, Translation,
Emotion

Example output

Start Time	End Time	Sentence
00:06:30,992	00:06:32,813	Cały czas robi coś, żebym mógł wierzyć.
00:24:44,004	00:24:44,425	Dlaczego?
00:08:00,466	00:08:01,886	Wydupcę was stąd wszystkich, nie?
00:08:30,810	00:08:32,751	Ja mam taki ciężar, stary.
00:08:43,875	00:08:44,495	Nie jestem z mafii.

+

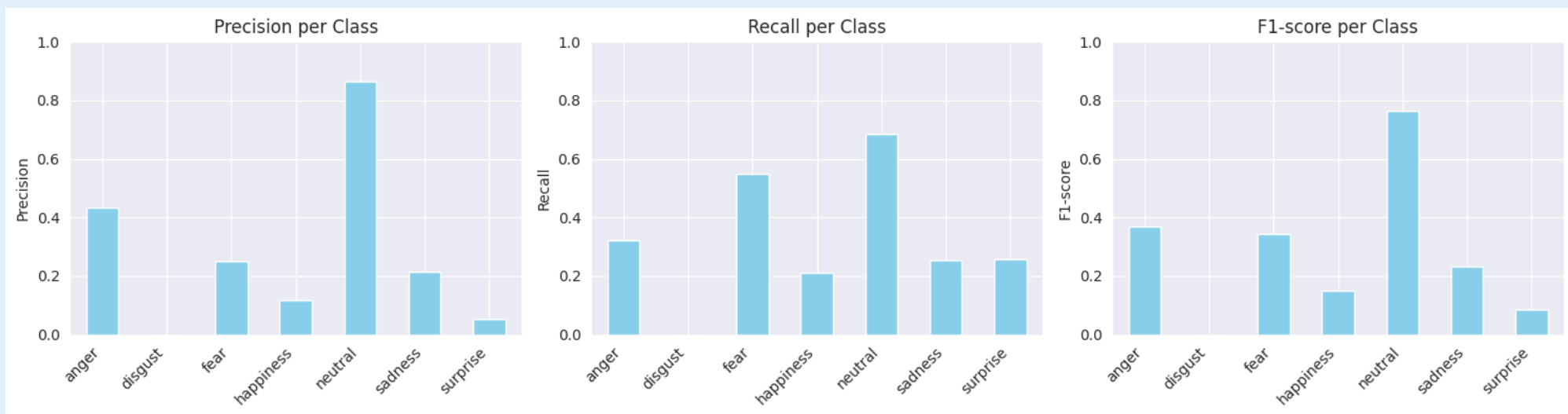
Translation
He's always doing something so I can believe.
Why?
I'm gonna fuck you all out of here, right?
I have such a burden, man.
I'm not from the mob.

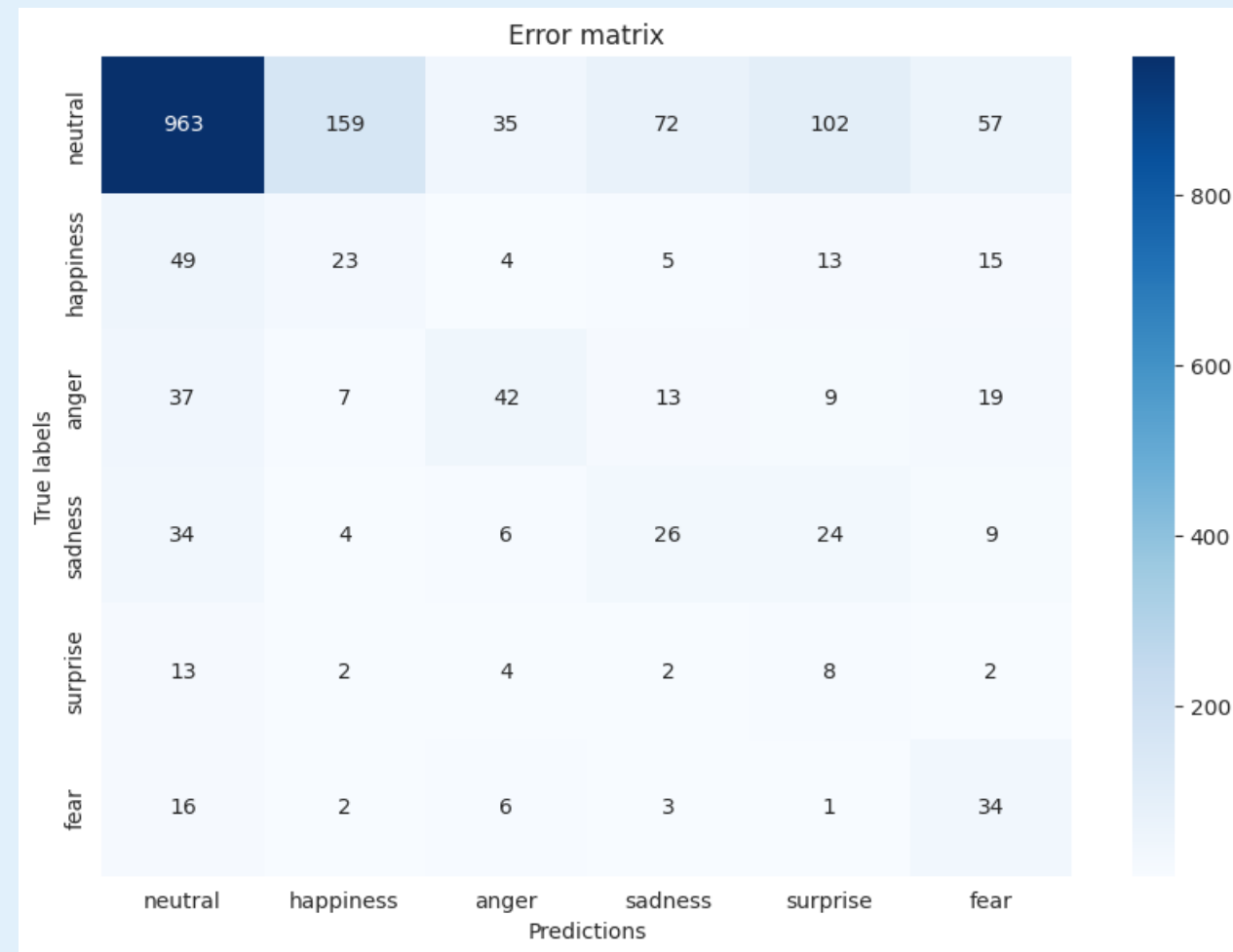
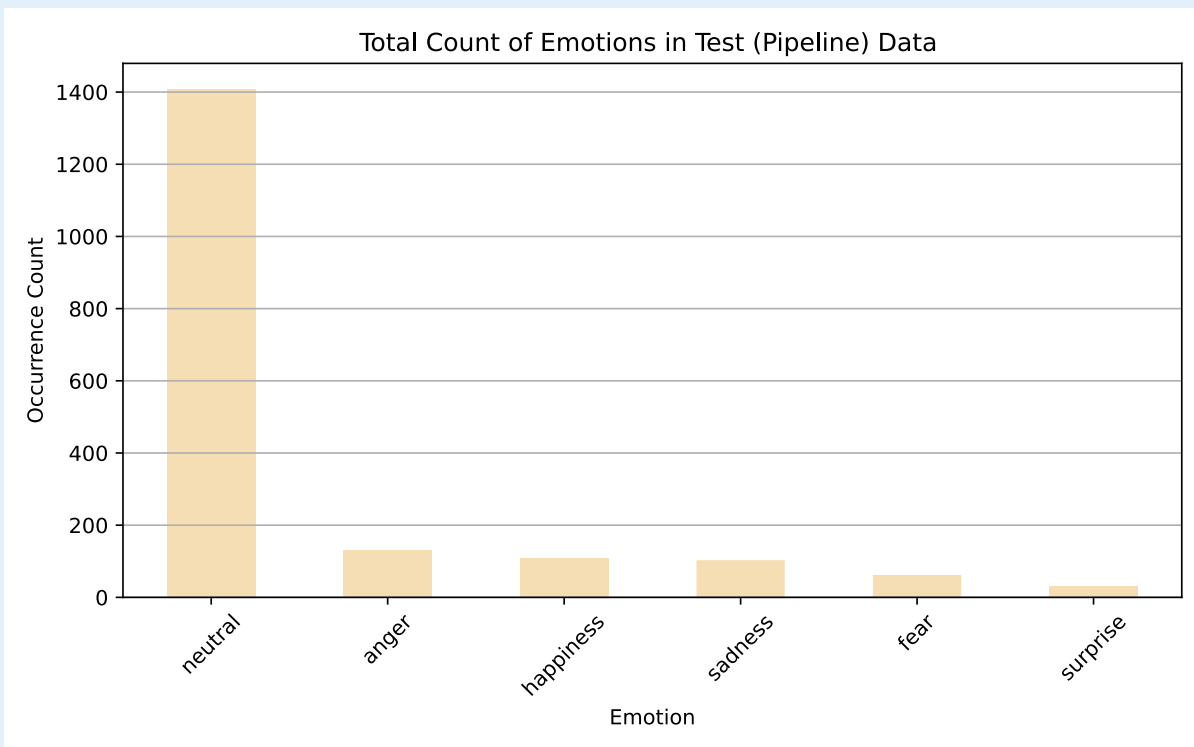
+

Emotion
happiness
surprise
anger
sadness
fear

Results & Analysis

	Worst Model	Best Model
Transcription	Whisper – 2.44% WER	AssemblyAI – 1.63% WER
Translation	Helsinki – 28.1 BLEU Score	Helsinki – 31.2 BLEU Score
Classification	RNN – 0.15 Weighted F1 Score	Transformer (DistilRoberta) - 0.66 Weighted F1 Score





Ethical & Sustainability Considerations



Fairness

Due to skewed training data classification model is less likely to output predictions for minority classes, especially emotion *disgust*.



Energy Use

Our classification model is based on DistilRoberta architecture that is 34% smaller than original Roberta base

Challenges and Limitations

Polish being low-resource language – difficulties in finding reliable datasets

Mistakes being time consuming – due to limited resources finetuning transformer models on large datasets usually takes hours

Translations are not satisfactory, which could lead to misclassification in the following step

Complexity of the language – holds the pipeline back from getting satisfactory outputs

Original Sentence	Translation	Type of error
... że próbuję to na ciebie nakręcić?	... that I'm trying to film you?	Direct translation of some words
Chciałem fakt wypełnić, ale ciężko było, po prostu, bo nie jestem mafią, więc... Aaa, ciężko było przekonać?	I wanted to fill out the facts, but it was hard, just because I'm not a mob, so...	Cutting some part of long sentences
Na niego jestem 100% pewien, że jest dobry.	I'm 100% sure he's good at it.	Adding extra, unnecessary words.

Next Steps & Recommendations

- Further research in machine translation of polish and other low-resourced languages
 - examine industry-standard strategies utilized by comparable teams to ensure data integrity
- Model improvement
 - find better transformer model
- Improve data quality
 - find new dataset
 - apply data augmentation
 - boost minority classes presence

Thank you!

Any questions?