Final presentation

Group 22, Polish: Monika Stangenberg(231648), Michał Bątkowski(232079), Kamil Łęga(232753)

Block C Year 2 2024-2025

Classifying emotional tone from spoken Polish language using NLP.

Domain: Audio & text emotion recognition with AI.

The team created a pipeline which:

1 Extracts and transcribes speech from Polish videos

2 Translates text from Polish language to English

3 Classifies sentences into one of seven emotion categories based on the model by Ekman and Friesen (1971): happiness, sadness, anger, surprise, fear, disgust + neutral.

Context:

- Project done for **Content Intelligence Agency**
- Supports research and helps develop Polish NLP tools.
- Focus on real, unscripted speech for more authentic results

Data Characteristics & Limitations

Test data:

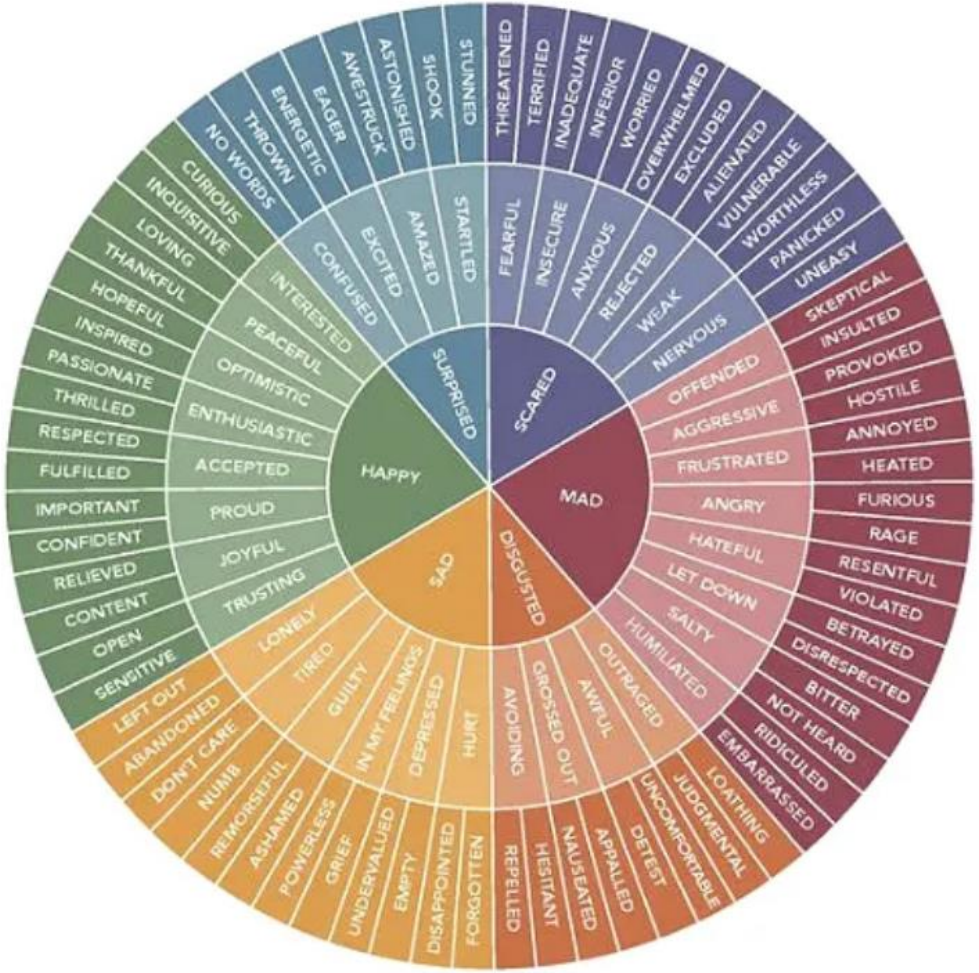
- Transcription of two street interviews from YouTube. (Polish language)
- They are street interviews that cover different topics (vacation, weather, and politics).
- Unscripted and informal language, with a dominance of neutral sentences.
- Transcription and emotion predictions were generated using Content Intelligence Agency Pipeline.

Translation:

Since language is a very delicate factor, achieving good results requires highly accurate translations. However, the company's pipeline does not perform well on Polish translations.

example: (Żywiec is a city)
original: Z daleka, z Żywca aż.
translation: From a distance, from Żywiec to.
real meaning: From far away, from Żywiec even.

Emotion Wheel



Data Characteristics & Limitations

Test data v1 – emotions were labelled according to the emotion wheel
Because of this approach, data labels were questionable.

example, (this sentence is clearly neutral, but it is classified to happiness because of curiosity)
What to do when you can't sunbathe and bathe? We will ask tourists about it.
company pipeline label: **curiosity**
emotion wheel label: **happiness**

Test data v2 – emotions were manually corrected, which improved performance of the model.

Additional limitations (Test & Train):
- Class imbalance (in test data: neutral – 85567, fear – 174)
- Small dataset (it is hard to find data similar to our street interviews data)

Data Characteristics & Limitations

Train data:

- The final model is trained on data shared by other teams working on the same problem.
- Their data was collected by extracting and transcribing speech from videos in various languages with unscripted language.
- Their transcription and emotion predictions were generated using Content Intelligence Agency Pipeline.
- The "Daily Dialogue" dataset was combined with this data.

Translations were based on different original languages, which affected the overall text quality. Most teams used TV shows as their data source, which is very different from our street interviews.

Previous experiences – the team tried different data to train the model (Friends emotion-labelled dialogues, MELD dataset, Daily Dialogue, Affect data, fairy tales). Because these data was very different to our train data, models were performing poorly.

Why We Chose Transformer model

- **Context-aware**: Understands emotion beyond keywords
- **Cross-lingual**: Robust to Polish to English translation
- **Pretrained**: Efficient fine-tuning on limited resources
- **Handles imbalance**: Supports class-weighted loss

Performance and Accuracy

Adjusted Test Set (Corrected Labels):

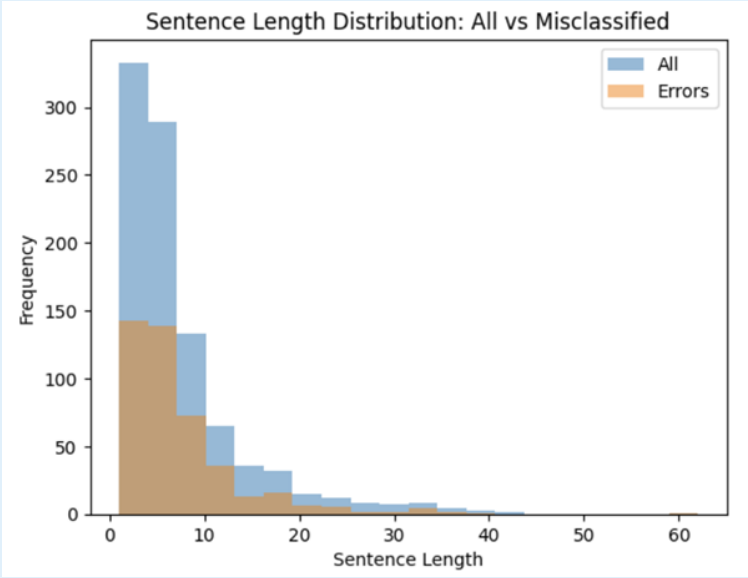
- **Accuracy**: 53.3%
- **F1 Score**: 0.53 (weighted)

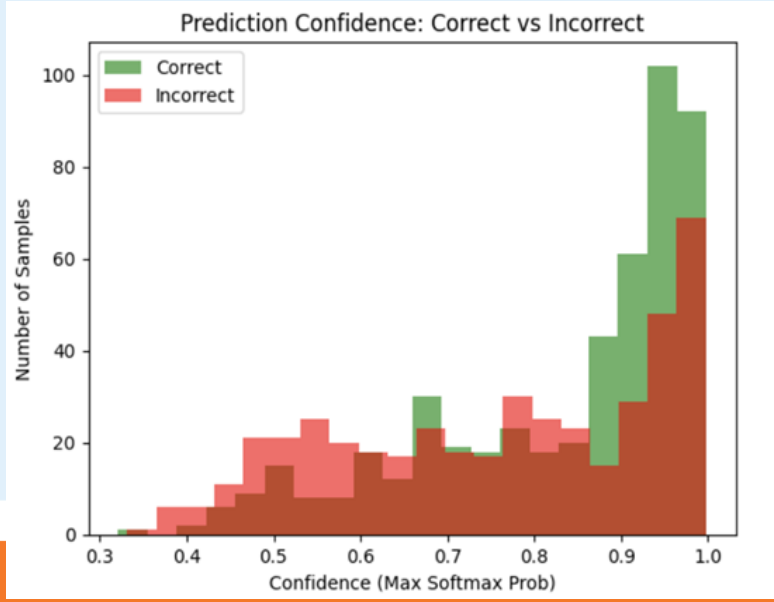
Original Test Set (Noisy Labels):

- **Accuracy**: 43.6%
- **F1 Score**: 0.42

Key Takeaways:

- Label correction improved metrics
- Confusion in **surprise/sadness**, **disgust/anger**
- Overconfidence in errors (≥ 0.9 confidence)
- Short inputs = high error rate





Example generated text

Input Sentence	Predicted Emotion
He will come later during this day	Neutral
I do not want this to happen ever again	Fear
Stop what you're doing right now	Anger
I never want to see you again	Disgust
I've never seen anything like this	Surprise
Yesterday was such an interesting day, it gave me so many positive insights	Happiness
I will do it later because now I feel majorly down	Sadness

Model performance analysis (XAI)

- Model often focuses on key emotional words
- Wrong predictions: misinterpretation of emotional tone/focusing on punctuation
- Model is quite robust, because it can recover when key words are removed.

example:

When I look at him that this bandit, not a minister.

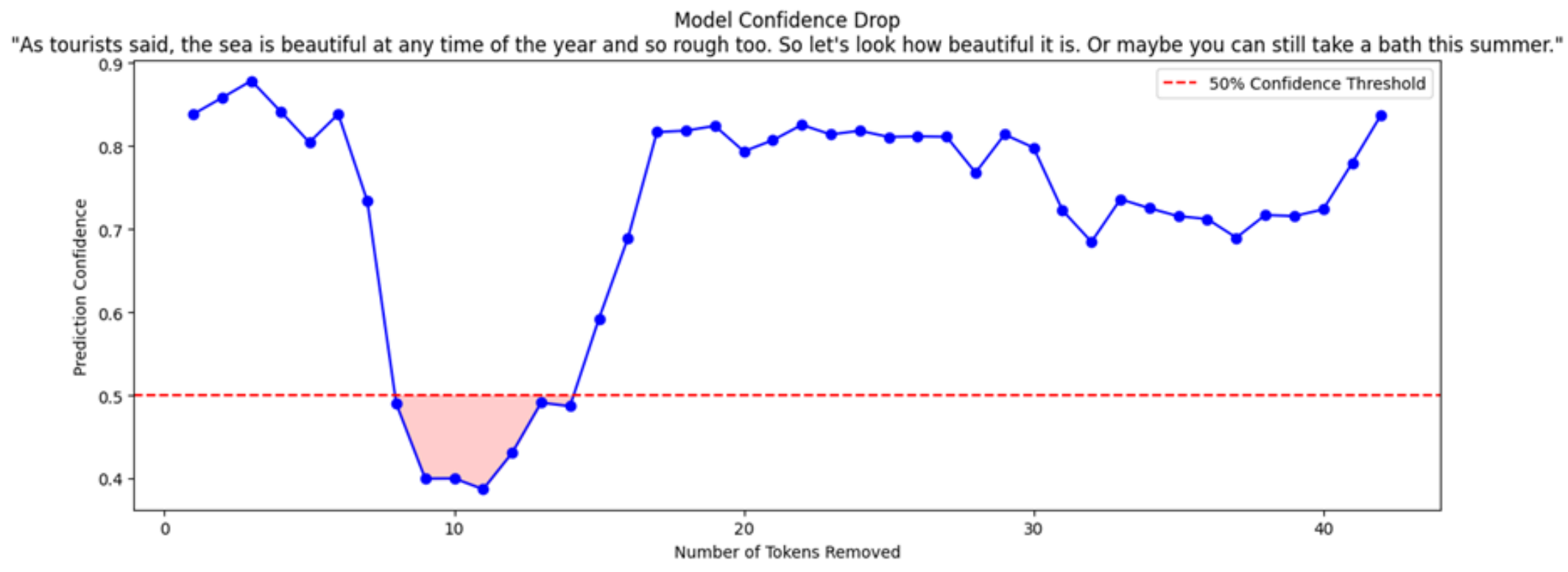
confidence: 0.42

model pays a lot of attention to "bandit"

Input Peturbation

The model loses confidence after removing 8 tokens (around the word "beautiful"), but quickly recovers.

This means it relies on key words but can still make accurate predictions using the rest of the sentence. It is robust because even when important words are missing, it can adapt.



Translation process

- After transcription of polish audio.
- Use of pretrained model of "Helsinki-NLP/opus-mt-pl-en" as it performed the best.
  - Three different iterations of model, tried hyperparameter tuning and additional model training.
  - No additional actions taken to modify it as it only decresed the performance of the model.

Translation process

- Machine translation struggles with context, idioms, and local terms.
  Example:
    - *"W jaki sposób trzeba sobie zasłużyć, żeby zostać ustczaninem?"*
    - *"How do you earn to be a mouthman?"*
    - *“Ustczanin”* (a person from Ustka) was mistranslated as *“mouthman”* — showing the model's failure to recognize demonyms and cultural context.
- Word-by-Word substitution without semantic coherence (Lack of context)
    - Example:
        - *"Z Żywca aż."*
        - *"From Life to Life."*
        - *"Żywiec"* is a **place name** (of a town[or beer]), but the model misinterpreted it as the noun *"life"*.
- These issues point to:
    - Insufficient handling of named entities.
    - Lack of cultural/language-specific awareness.
    - Challenges in disambiguating meaning in short, decontextualized phrases.
    - Struggles with informal, spoken language structures.

Final project pipeline

- **Audio Input** – Polish `.mp3` files with spoken content
- **Transcription** – Sentences and timestamps via **AssemblyAI**
- **Translation** – Polish → English using **Helsinki-NLP models**
- NLP Feature Extraction – Using spaCy and custom scripts
- **Emotion Detection** – Transformer model classifies emotion (e.g., joy, fear)
- **Final Output** – CSV with aligned times, translations, and emotion labels

	Start Time	End Time	Sentence	Translation	Emotion
0	00:00:23,227	00:00:26,288	Studio Promenada dzisiaj z promenady.	Promenade Studio today with the promenade.	happiness
1	00:00:26,448	00:00:34,609	Przy takiej pogodzie na plaży niespecjalnie mo...	In such weather on the beach you can not espec...	neutral
2	00:00:34,668	00:00:40,189	Mimo, że koniec wakacji, to turystów wciąż jes...	Even though the holidays are over, there are s...	neutral
3	00:00:40,270	00:00:43,831	Co robić, gdy nie można się opalać i kąpać?	What if you can't sunbathe and bathe?	neutral

Final project pipeline

- Common problems that can occur during usage:
  - **Noisy Audio**: Poor audio quality can reduce transcription reliability.
  - **Emotion Drift Post-Translation**: Emotional cues in Polish may not map directly to English.
  - **Emotion Model Generalization**: Model may struggle with informal or regional expressions.
  - **Pipeline Latency**: Real-time use is limited due to cumulative processing time.
  - **Limited Training Data**: Emotion classifier may underperform on underrepresented emotions.

Ethical Considerations

- **Human-verified labels**: Improved fairness & reduced label bias
- **Overconfidence detected:** thresholds alone not enough; human review or calibration advised
- **Risk-aware design**: Not intended for clinical/mental health use
- **Transparent AI**: XAI tools used to expose model reasoning
- **Sustainable AI**: Low energy training & local inference

Sustainable AI Approach

- **Pretrained model reused** (roberta-base)
- **Early stopping** to reduce training waste
- **Moderate energy use**: ~0.13 kWh per training run
- **Efficient tuning** via 5-fold Optuna & stratified sampling
- **Conscious design** despite cloud-based training

Next Steps

- Find/create better dataset, which is suitable for our problem.
- Find better guide to classify emotions.
- Improve translations/train model on Polish data.
- Keep data anonymized and remove it after the end of the project.
- Deploy functional application.

Sources

- Alishahi, A. (2010). *Computational modeling of human language acquisition*. Synthesis Lectures on Human Language Technologies, 3(1), 1–107. https://doi.org/10.2200/S00233ED1V01Y200912HLT005
- Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: Analyzing text with the natural language toolkit*. O'Reilly Media.
- Ekman, P., & Friesen, W. V. (1971). *Constants across cultures in the face and emotion*. Journal of Personality and Social Psychology, 17(2), 124–129. - emotion wheel
- Jurafsky, D., & Martin, J. H. (2025). *Speech and language processing* (3rd ed., draft). Retrieved January 12, 2025, from https://web.stanford.edu/~jurafsky/slp3/
- OpenAI. (2024). *ChatGPT (GPT-4)* [Large language model]. https://chat.openai.com/

We used **ChatGPT (OpenAI, 2024)** to help rephrase and structure content based on our original project work from GitHub. All insights and analysis are our own.

For your attention!