# PolEval 2024 Task 2: Emotion and Sentiment Recognition

Jan Kocoń, Bartłomiej Koptyra
(Wrocław University of Science and Technology)

# Task Definition

- Create a system to recognize:
  - Emotions from Plutchik's wheel
  - Sentiments in consumer reviews
- Analysis levels:
  - Whole text
  - Individual sentences
- Note: Manual labeling of test examples was forbidden

# Dataset Details

- Content: Polish consumer reviews from 4 domains:
  - Hotels
  - Medicine
  - Products
  - School
- Also includes neutral informative texts
- Annotation by 6 independent annotators
- Labels selected with ≥2 annotator agreement

# Dataset Structure

- Training set:
  - 776 reviews
  - 6,393 sentences
- Test sets:
  - 2 sets of 167 reviews each
  - 1,234 and 1,264 sentences respectively
- Random split done at review level

# Annotation Schema

Emotions (Plutchik's wheel):

- Joy
- Trust
- Anticipation
- Surprise
- Fear
- Sadness
- Disgust
- Anger

Sentiments:

- Positive
- Negative
- Neutral
- Ambivalent (both positive and negative)

# Dataset Format

- Each review:
  - Multiple sentences
  - Ends with '#' symbol
  - Individual sentence annotations
  - Whole review annotation
- Boolean values for each emotion/sentiment
- Available at: https://huggingface.co/datasets/clarin-knext/CLARIN-Emo

# Example review

"Była to pierwsza wizyta ale moze i ostatnia." - anticipation, **sadness**, **negative**

"Lakarz troche apatyczny, nie wypowiadajacy sie jasno." - **sadness**, disgust, **negative**

"Mam zrobic jakies badanie ale nie dardzo wiem jakie." - **surprise**, **sadness**, **negative**

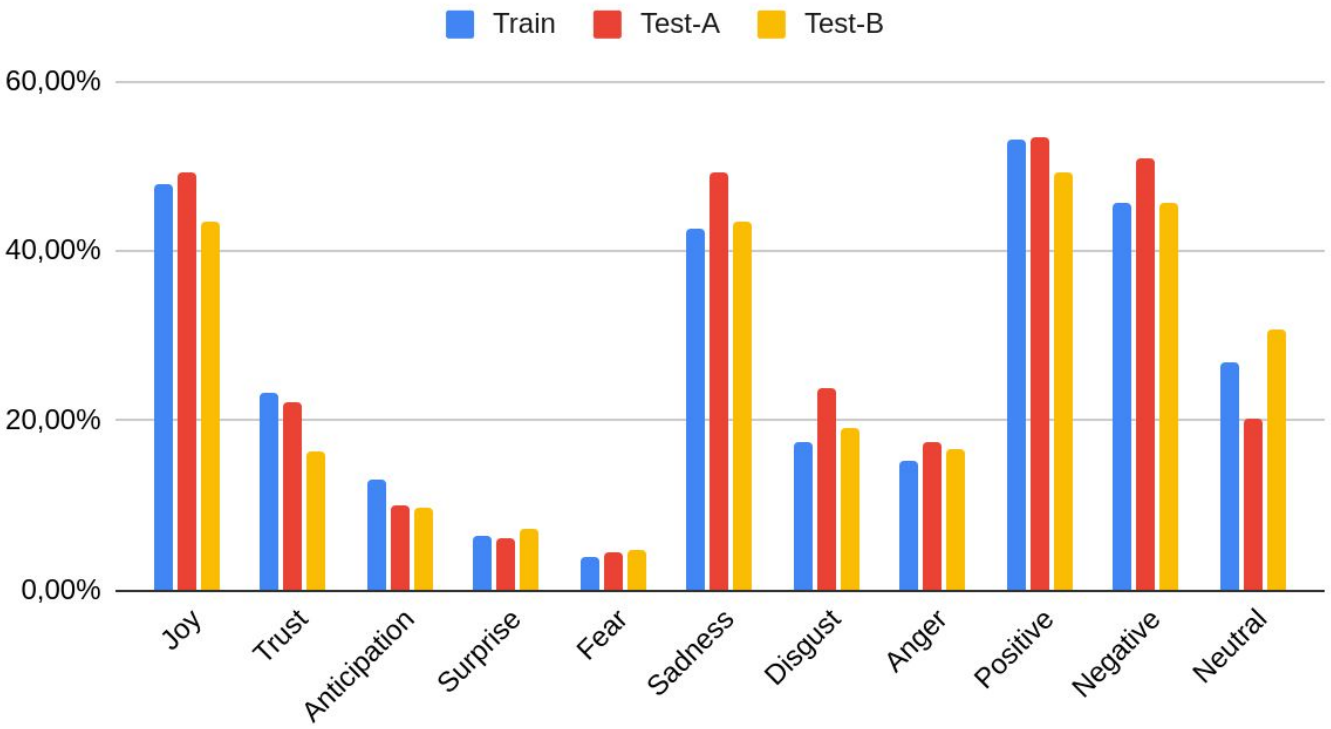
"Nie napisal skierowania/zalecenia, chyba mowil o gastrologii." - **surprise**, **sadness**, **negative**

"Powinnam byla byc bardzej wymagajaca i dopytujaca." - **surprise**, **sadness**, anger, **negative**

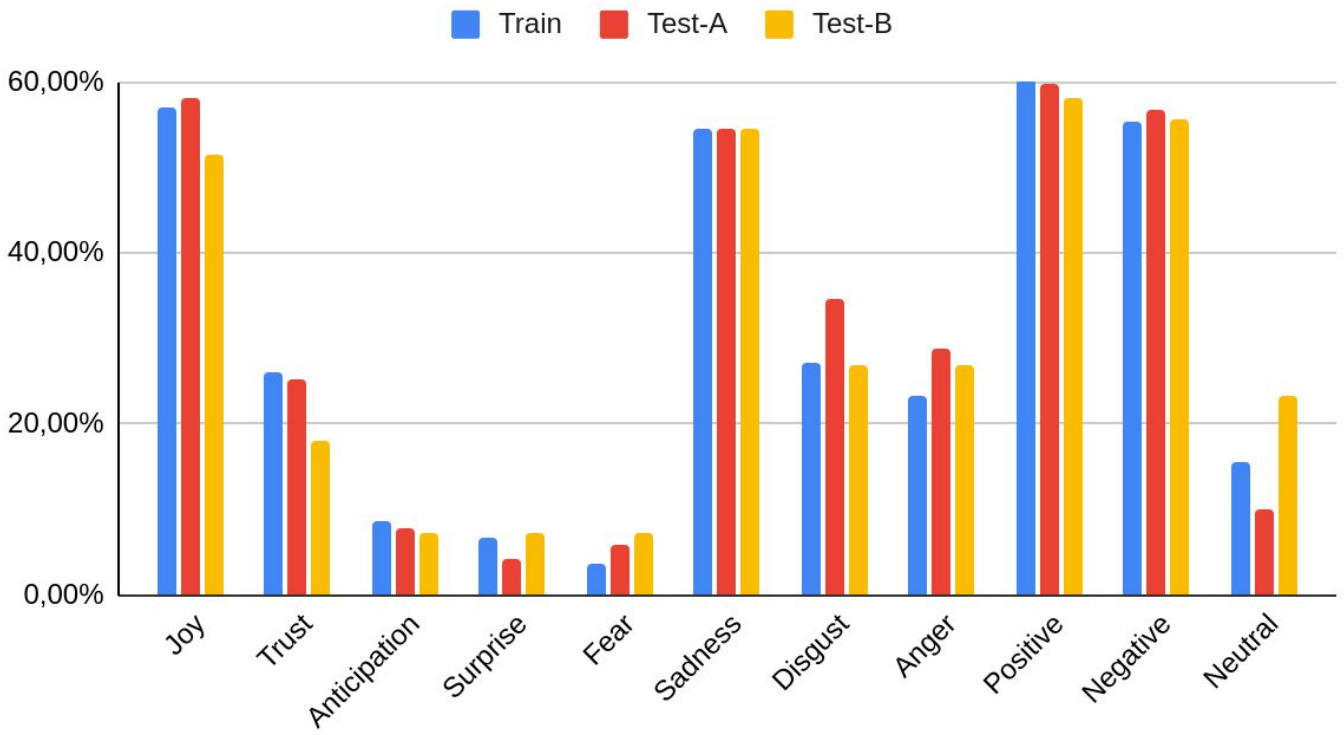
"Nie polecam tego lekarza." - **sadness**, **negative**

whole review - **surprise**, **sadness**, **negative**

# Distribution of labels - sentence level



# Distribution of labels - text level



# Evaluation Metrics

$$Final\ score = \frac{F1_{macro}\ sentences + F1_{macro}\ texts}{2}$$

# Submission Requirements

- Single tab-separated file
- 11 columns of boolean values
- Each row corresponds to input file
- Values indicate presence/absence of emotions

# Competition Results

| Rank | Submitter | Affiliation | Entries | test-A scores | | | test-B scores | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Sent. | Text | Final | Sent. | Text | Final |
| 1 | Krzysztof Wróbel | Enelpol, UJ, AGH | 10 | 81.62 | 79.40 | 80.51 | 81.51 | 78.48 | 79.99 |
| 2 | Tomasz Warzecha | - | 196 | 78.87 | 81.54 | 80.20 | 79.34 | 79.28 | 79.31 |
| 3 | Cezary Kęsik | University of Warsaw | 25 | 74.94 | 76.42 | 75.68 | 76.66 | 79.33 | 77.99 |
| 4 | Jakub Pokrywka | - | 15 | 78.65 | 75.93 | 77.29 | 79.43 | 75.77 | 77.60 |
| 5 | Paweł Lewkowicz | - | 10 | 74.29 | 77.73 | 76.01 | 77.27 | 77.20 | 77.23 |
| 6 | Katarzyna Baraniak | - | 32 | 75.94 | 77.47 | 76.70 | 76.11 | 77.76 | 76.94 |
| 7 | Cezary Kęsik | University of Warsaw | 5 | 73.62 | 79.12 | 76.37 | 75.94 | 70.43 | 73.19 |
| 8 | Jakub Kosterna | - | 4 | 50.47 | 28.71 | 39.59 | 52.19 | 28.71 | 40.45 |
| 9 | Paweł Cyrta | Metamedia Technologies | 5 | 33.04 | 32.74 | 32.89 | 31.86 | 34.28 | 33.07 |

# Key Insights

- Successful Approaches:
  - Large Language Models (LLMs)
  - Ensemble methods
  - Context-aware solutions
  - Synthetic data generation
- Challenges:
  - Label imbalance
  - Contextual understanding
  - Quality control in synthetic data

# Future Directions

- Improving rare emotion detection
- Better contextual integration
- Alternative evaluation metrics
- Enhanced quality control for synthetic data
- Applications:
  - Customer feedback analysis
  - Mental health monitoring
  - Human-computer interaction