MLPC – Team Toothpaste

- Aral Cimcim, Sandro Müller, Linus Madlener, Kevin Eberl

Data Split
Best split to avoid data leakage?

Data Splitting

- 80% for training to generalize

- 10% for validation, reducing overfitting

- 10% for testing, test unbiased performance

Data Leakage

- When similar frames are in training & test set

- Potential solution: split by files first

- Another issue: Annotation too similar for certain classes and text embeddings

JOHANNES KEPLER
UNIVERSITY LINZ

Evaluation
Which evaluation criterion?

Balanced Accuracy Score

- Imbalanced data: classes not equally distributed -> no favoring majority class

- Equal weight to TPR & TNR

- $Balanced\ Accuracy = \frac{1}{2}(\frac{TP}{TP+FN} + \frac{TN}{TN+FP})$

Worst vs. best performance

- Baseline classifier predicting majority class -> 75% (drums)

Metric	Drums	Speech
Accuracy	0.747	0.534
Weighted Accuracy	0.500	0.500
Precision	0.000	0.000
Recall	0.000	0.000
F1 Score	0.000	0.000
ROC AUC	0.500	0.500
PR AUC	0.253	0.466

Table 1: Performance Metrics for labels "Drums" and "Speech"

JMU JOHANNES KEPLER UNIVERSITY LINZ