How hand-crafted features improve transformer networks
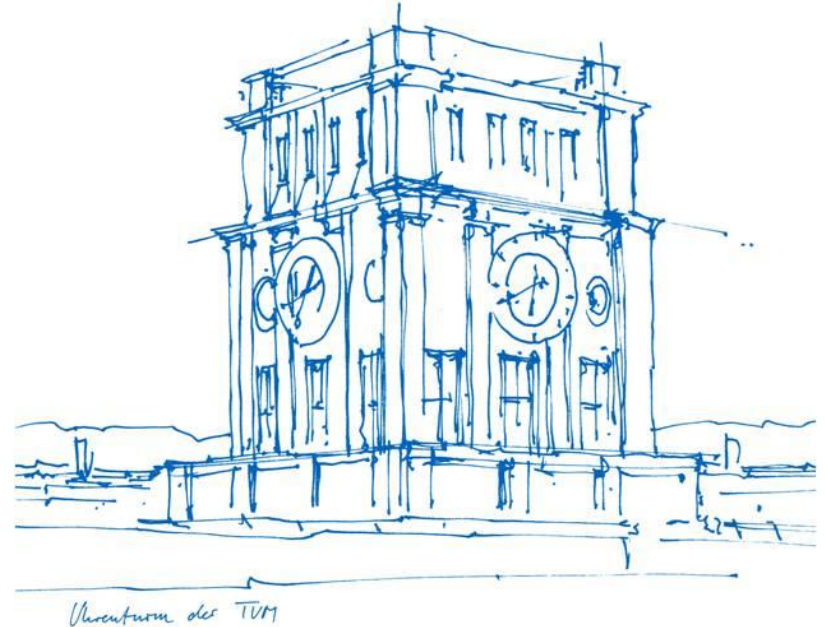
Miriam Anschütz, Georg Groh

Technische Universität München

Department of Informatics

Research Group Social Computing

Potsdam, 12th September 2022

Uhrenturm der TUM

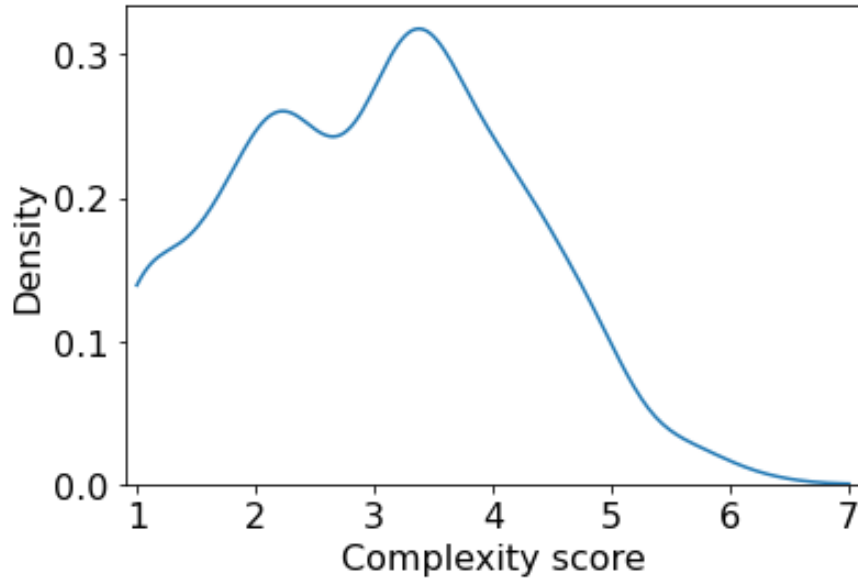# Shared task objective

Goal: Predict complexity of text [1]

Data: German sentences

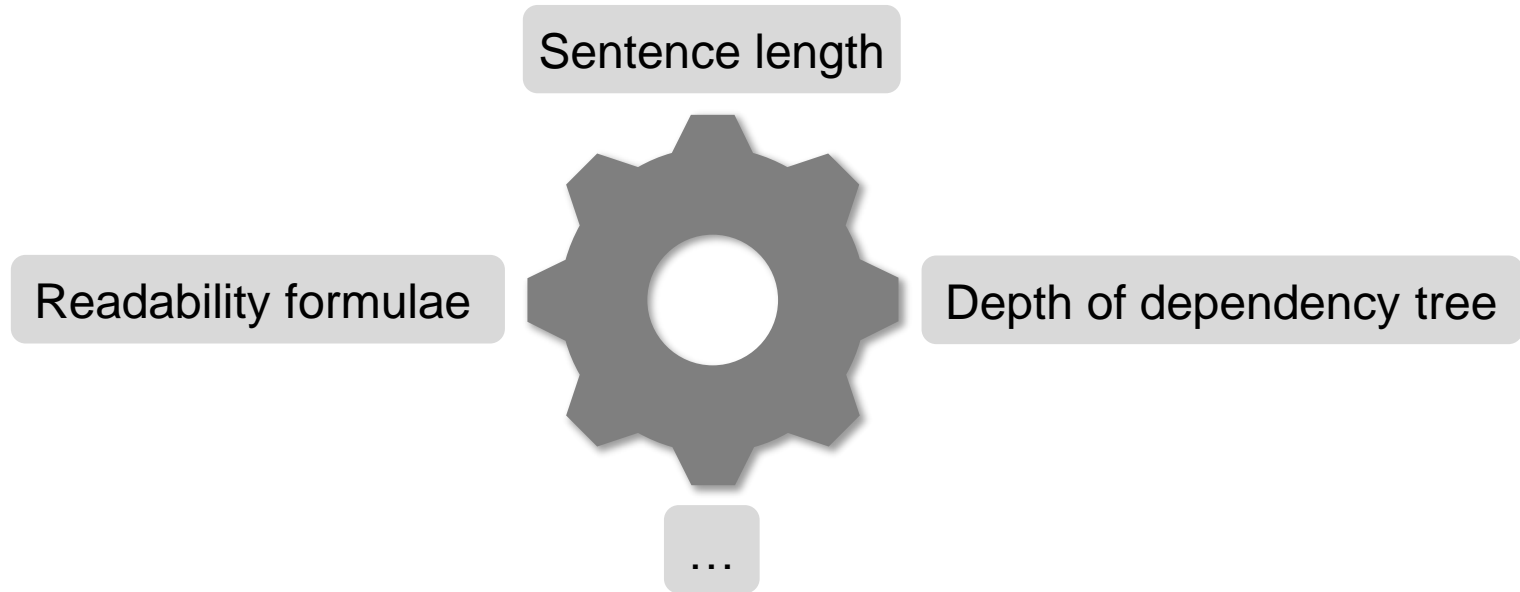Targets: Continuous complexity scores between 1 and 7

Discussion: Dataset challenges

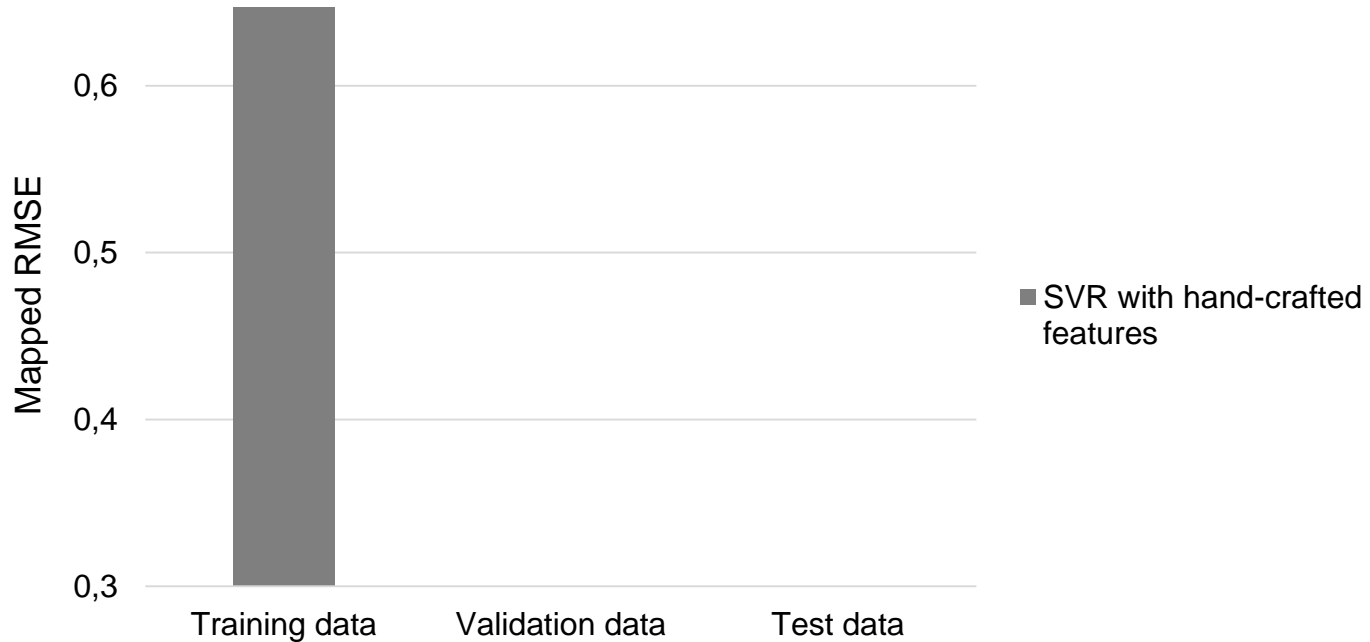The director of the Nobel Institute represents the secretary of the committee.

Complexity Score: 1.0

Selection of hand-crafted features for prediction

Sentence length

Readability formulae
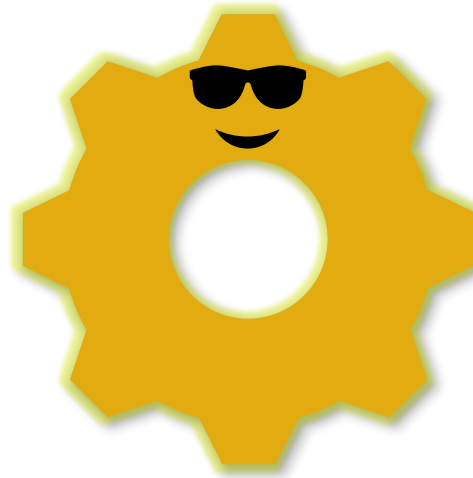
Depth of dependency tree

…

Results: Support vector regression (SVR) with hand-crafted features

State of the art: Transformer end-to-end models

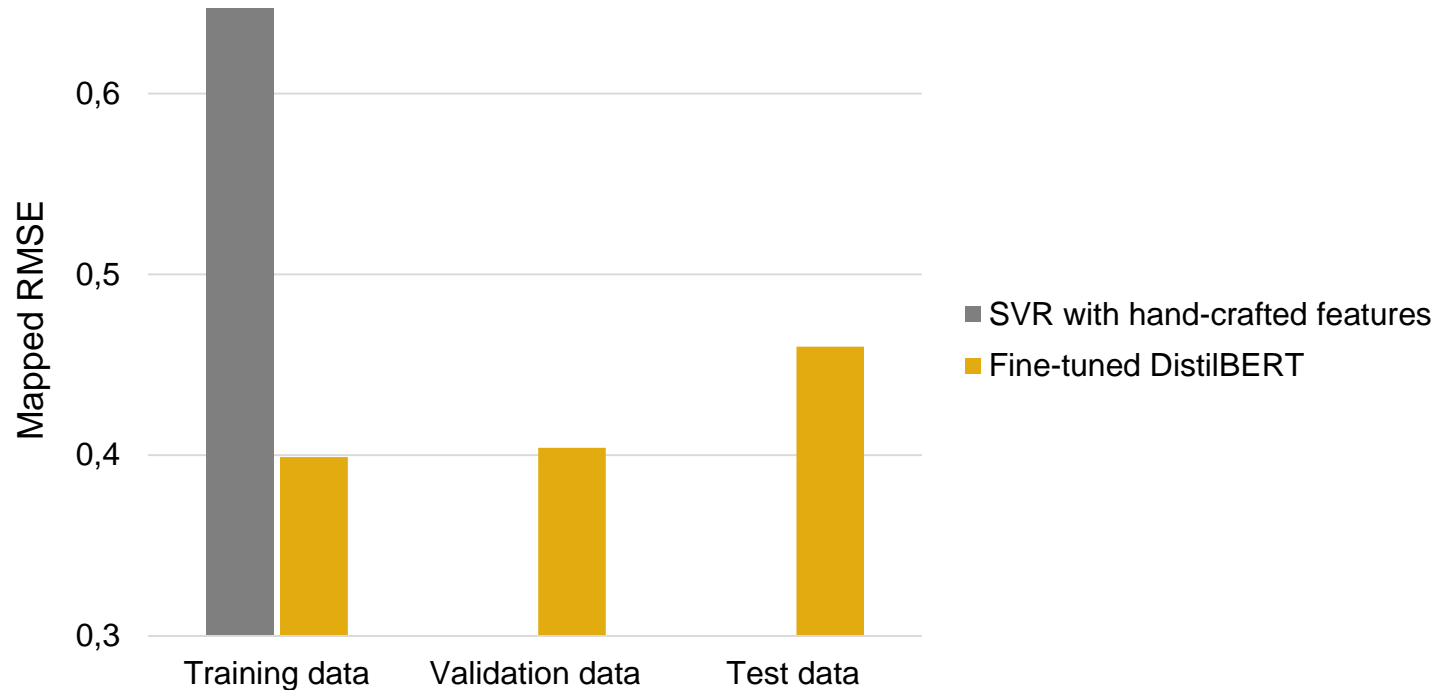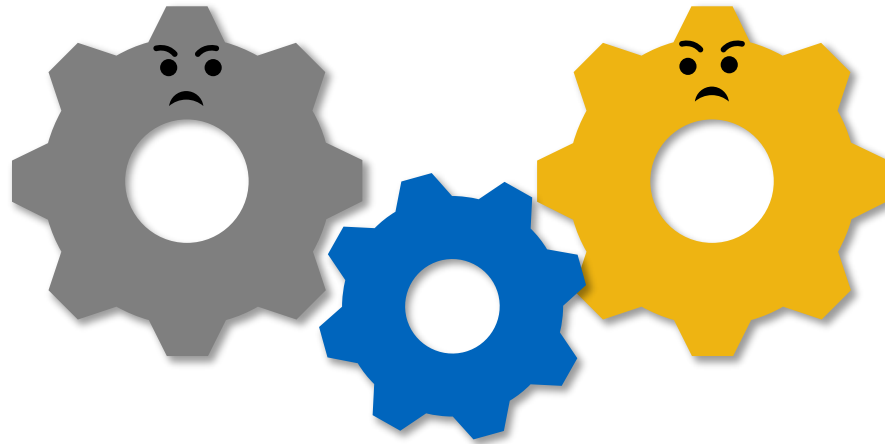# State of the art: Transformer end-to-end models
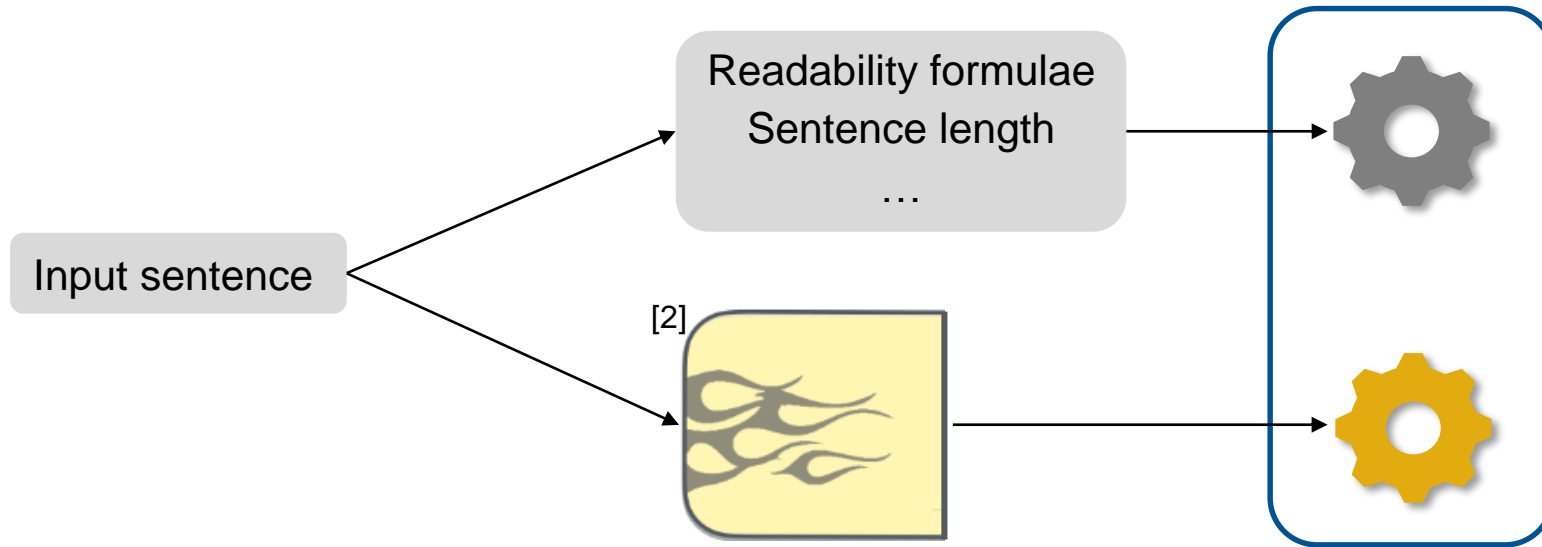

[2]

Results: Hand-crafted features vs. Transformer

Transformer model: Explaining predictions

What about combining the feature spaces?

Combined feature spaces: approach

Input sentence

Readability formulae
Sentence length
…

[2]
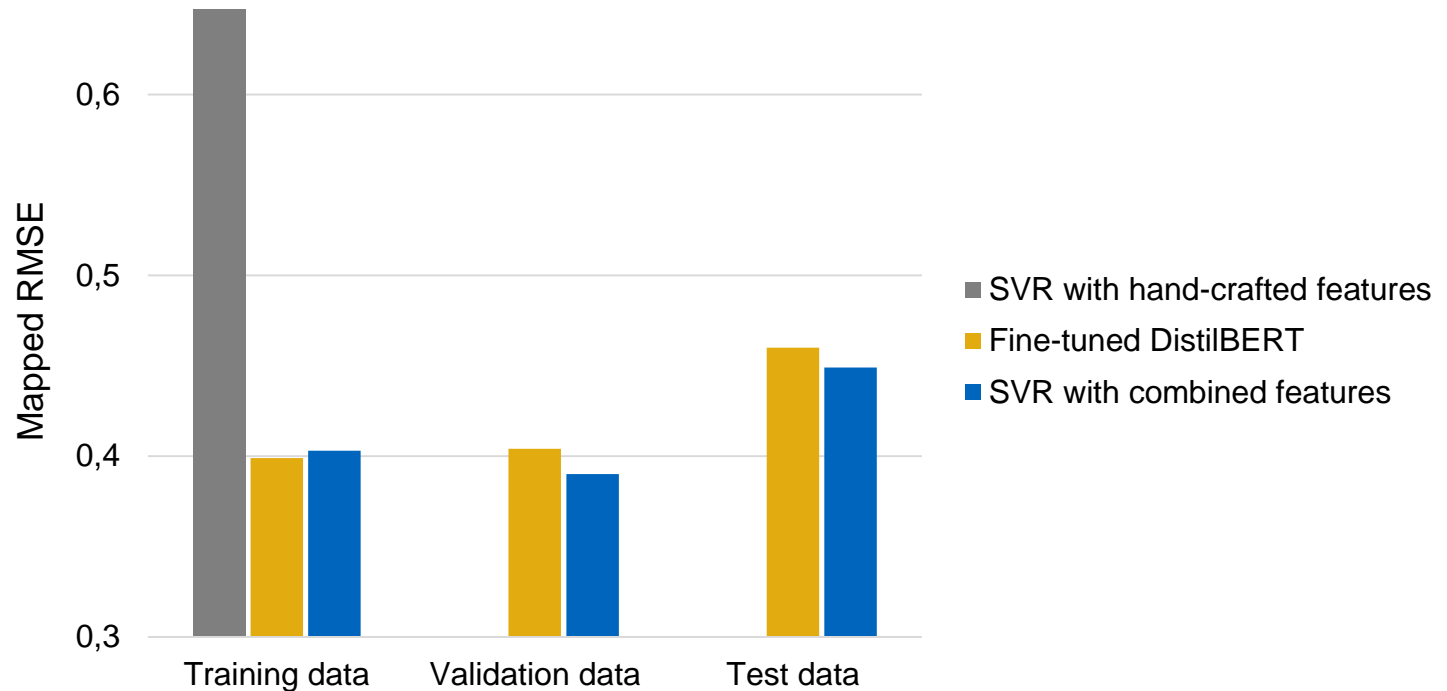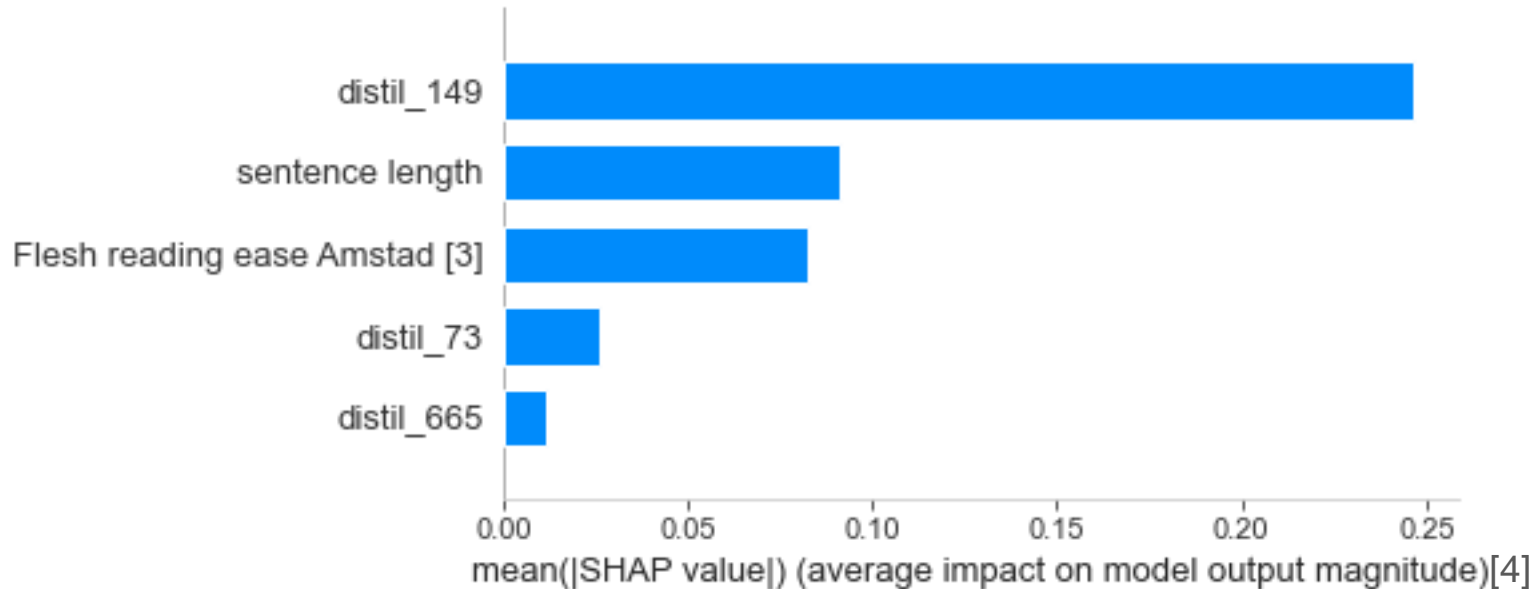
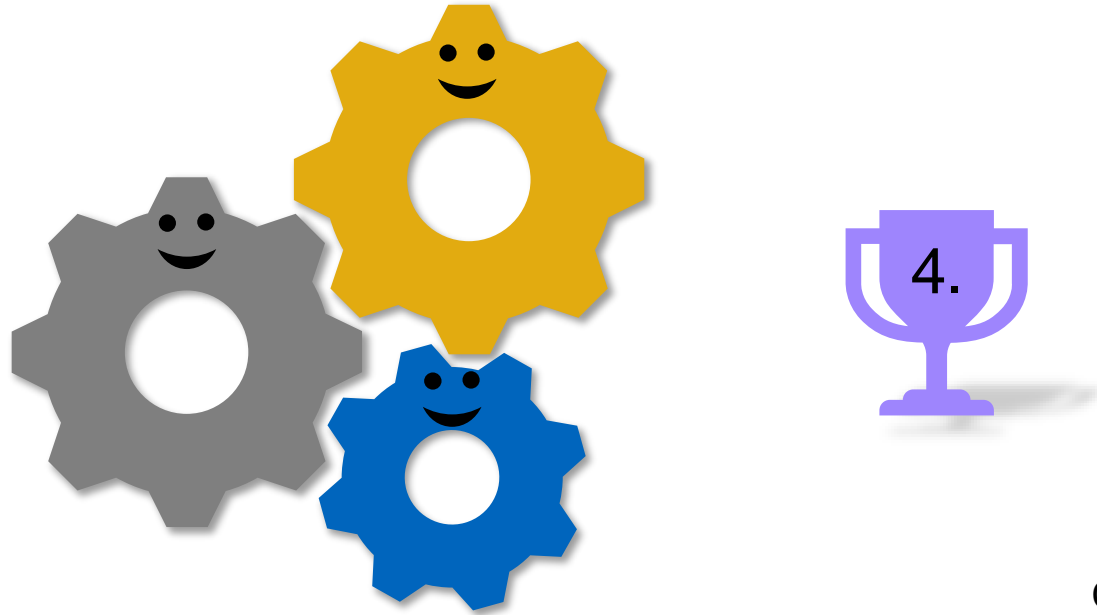# Results: Performance of combined model

Analysis: Relevant features in combined model

Conclusion: Feature ensemble most successful

4.

Check our code
on GitHub!

References

1. https://qulab.github.io/text_complexity_challlenge/assets/logo2.png
2. https://jalammar.github.io/illustrated-bert/
3. Toni Amstad. 1978. Wie verständlich sind unsere Zeitungen? Ph.D. thesis, Universität Zürich.
4. Scott M Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, Advances in Neural Information Processing Systems 30, pages 4765–4774. Curran Associates, Inc.

Appendix: Hand-crafted features

Readability formulae:

- Flesh reading ease
- Wiener Sachtextformel 1-4
- SMOG

Statistical features

- Sentence length
- Maximal dependency tree depth
- %words with ≥ 6 letters
- Average number of syllables
- %words with 1 syllable
- %words with ≥ 3 syllables

Appendix: Including hand-crafted in DistilBERT model