Reflection
Mini Project 02

KONSTANTIN - HENDRIK - NINA
AUTOMATION AND AI
HPI POTSDAM WS 2024/25

Task 1: Decision Tree

1

Measure for complexity: Flesch-Reading-Ease

Measure how hard it is for readers to understand the text

2

Hot and Label Encoding for String columns

Enable usage of string columns in decision Tree

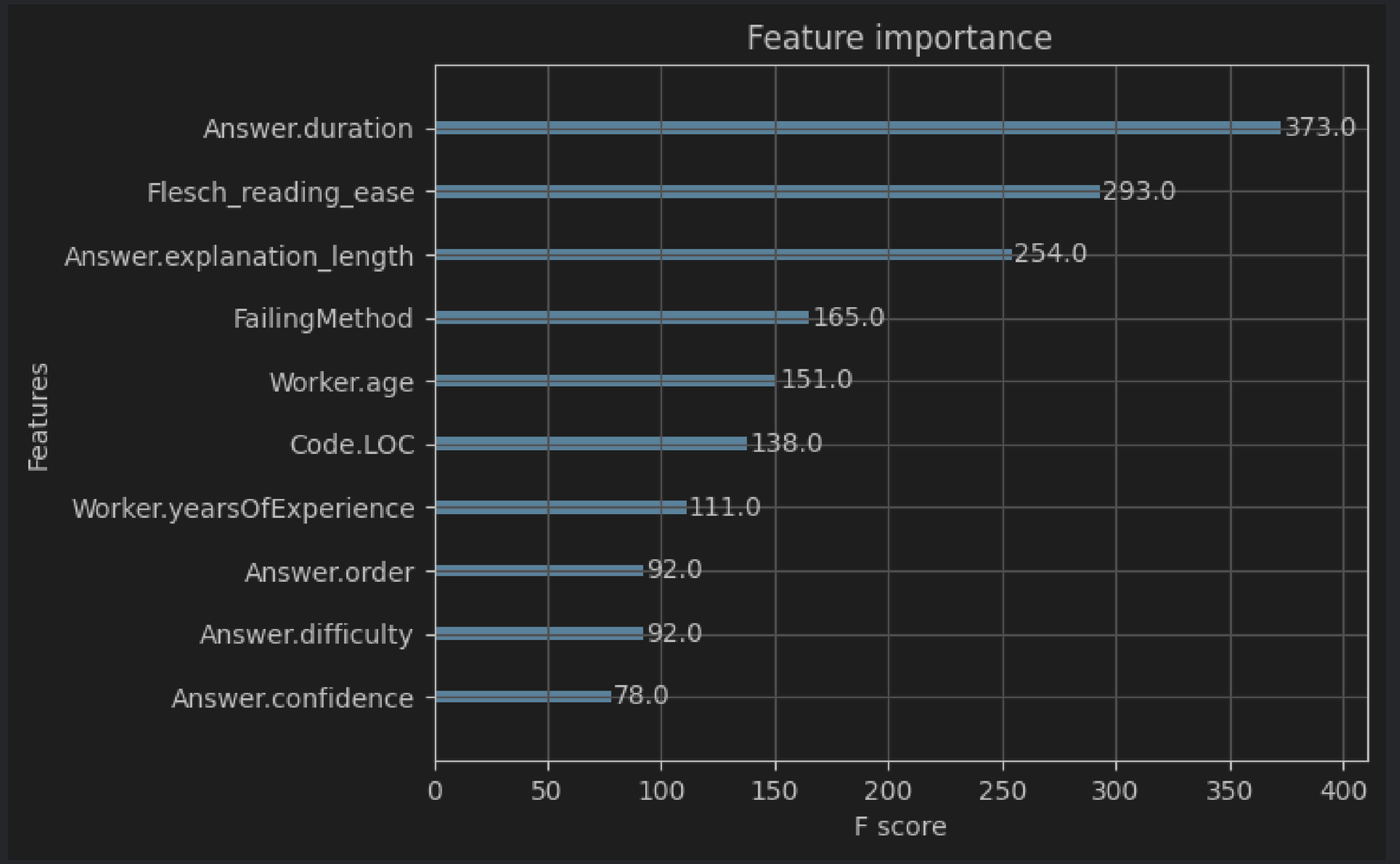
3

Results

Recall: 0.333

Precision: 0.284

Task 1: Feature Importance



Task 2

1. Using the task from our task description in a slightly adapted form

- answers were really long, including the precise code snippets
- had to prompt multiple times in order to match the format of the participants of the survey

2. Imagine to be a programmer being part of the survey, asking to consolidate the answers

- answers were rather long
- not similar to the format of the answers of the study participants

3. Precise role to be a programming buddy who consolidates bug reports

- had to prompt twice as the first answer was too long
- providing a limit in the next prompt gave a good explanation

4. Precise role to be a developer who consolidates bug reports and also a rough performance measure of their work

- worked great, using only one prompt

Task 2 Reflection

1

High BLEU, Low ROUGE

- High Precision
  - generated text overlaps with reference text
- Low Recall
  - missing content
  - does not cover key information

2

Score Alternatives

- BLEU & ROUGE evaluate based on n-gram overlap
- issues with synonyms, rephrasing
- use score that evaluates semantic similarity (BERTScore, BLEURT, ...)

Further Points:

1

⚡Data Quality

Possible mismatch between answer explanation and selection.

2

Classifier updates

Retraining is necessary if new demographics are added.

3

⚡Testing the Output

Complexity score might not be the right measure..

4

Consolidation Quality

Embeddings could be a more promising approach.

5

Debugging

⚡Trade-off between LLM offering simple interaction while being a black box regarding results.