

University of Aveiro

Classification problem with data compression

**Identify if a text was rewritten by
ChatGPT**

Goal

The goal of our work is to be able to identify if a piece of text was rewritten by chatGPT or not, using a compression algorithm.

The compression algorithm that we will use is the Markov model, identifying patterns in our language. It's also an algorithm that can be trained on other data.

The decision will be made based on the level of compression of each model (AI model and Human model)

Data pre-processing

- We had nearly 300 MB of data after removing irregular characters and downsampling the data (initially with 1.3 GB)
- 150 MB of human written text and 150 MB of GPT generated text
- We divided the data in two datasets: a training one with 90% of the data and a testing one with 10% of the data
- We also created variations of the datasets with no capital letters, no numbers and/or no spaces.

Implementation

HashTable

HashMap with counts
Calculation of probabilities
Save/Load to a file

MainClass

Two MarkovModels
Predict the origin of the file
Test models and fill confusion matrix

MarkovModel

Amount of bits to compress a text
Model training

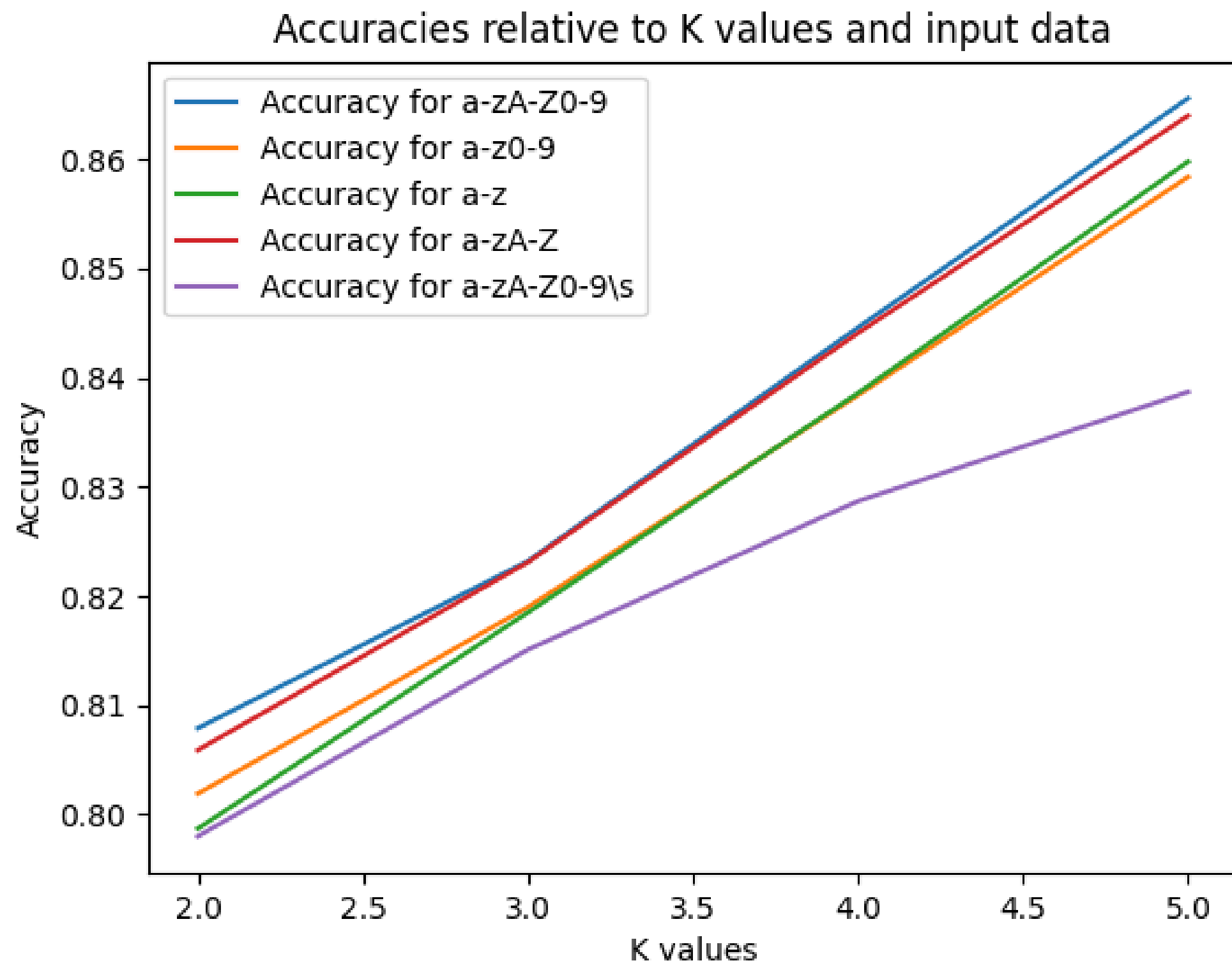
Metrics

Confusion Matrix
Metrics calculation

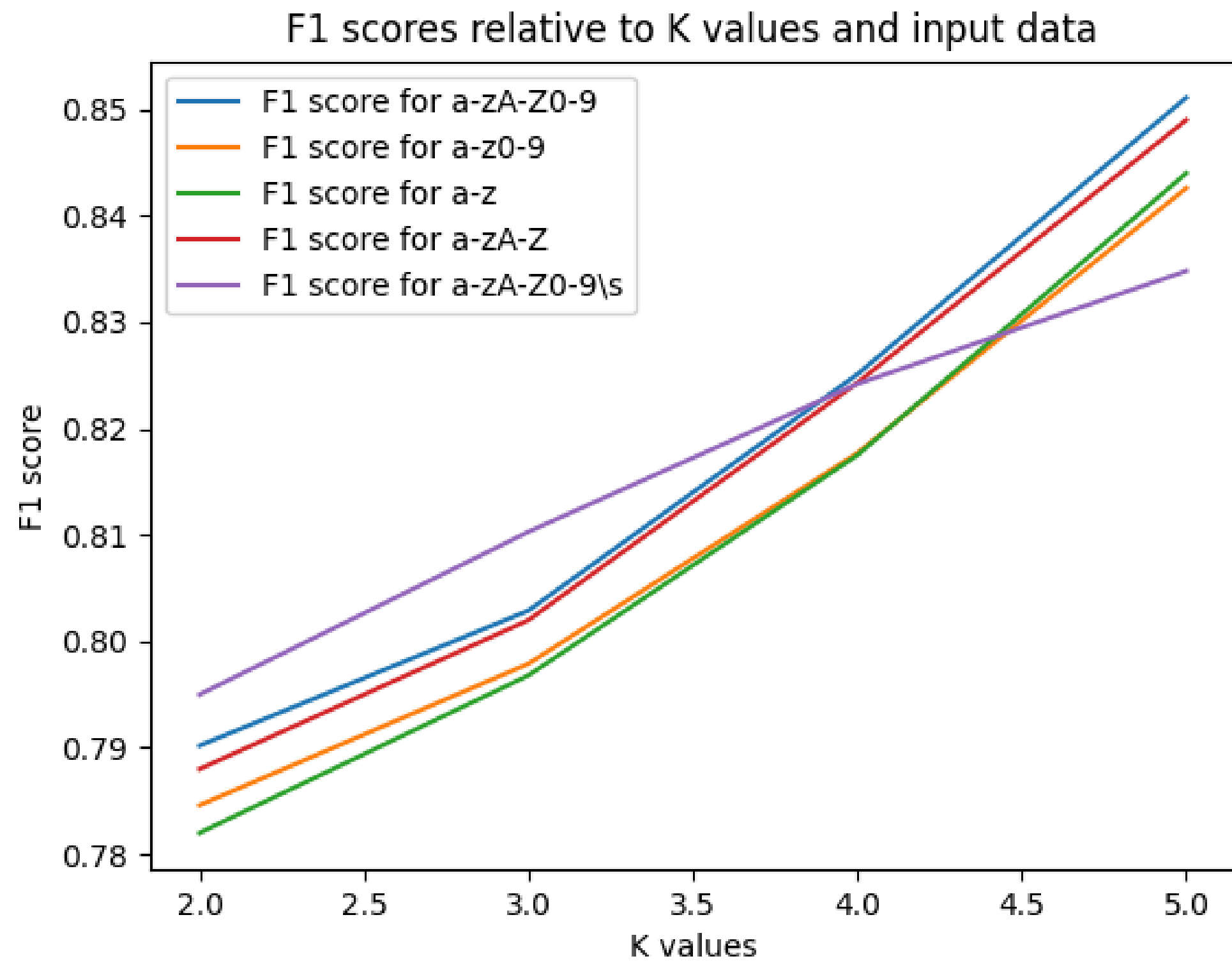
Experiments

- We varied the context length between 2 and 5.
- We also varied the input data by lowering the case of capital letters, removing numbers or removing spaces.

Results



Results



Results

We created our best model with K=5 and a-zA-Z0-9

Recall: 0.7566

Precision: 0.9725

Accuracy: 0.8656

F1 score: 0.8511

Results

Confusion Matrix for K=5 and a-zA-Z0-9

Actual \ Predicted	Positive	Negative
Positive	1980	637
Negative	56	2482

Conclusion

- Overall we believe the model is fairly reliable to use, but could be improved in the classification of GPT written texts since it still misses a considerable amount of times compared to the human texts classification
- Besides that, if we trained the models for higher values of K , maybe the model would have been even better