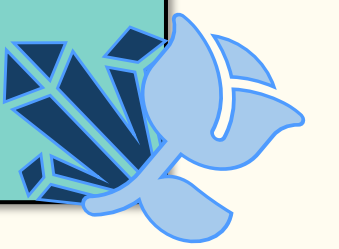


Among the ravens in the shadow court from the academy of dark ice and crystal roses



Does this title make your book popular?

Introduction

For our project, we investigated whether the presence of certain words in a **books title** influences their **popularity**. To address our research question, we used data from two **online reviewing platforms**.

Essential Decisions

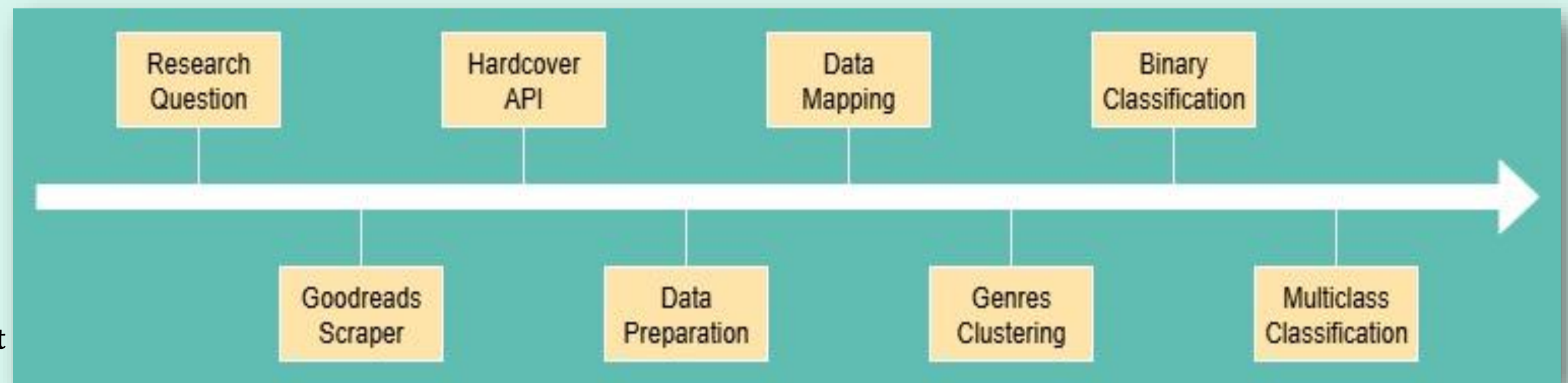
- **Goodreads-lists** as main data source
- **Hardcover** as additional source of data
- **English** books
- **Huggingface** for the models

goodreads

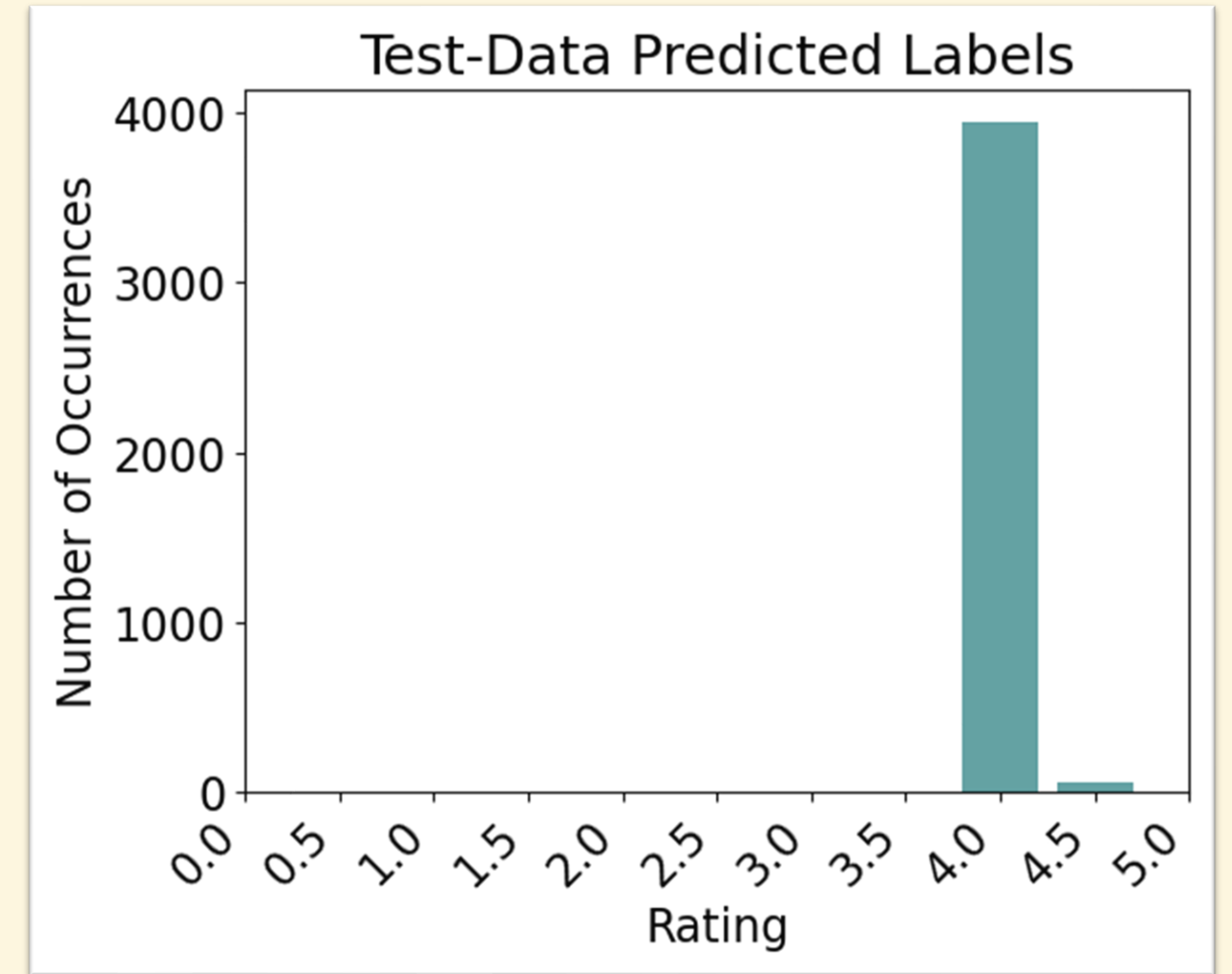
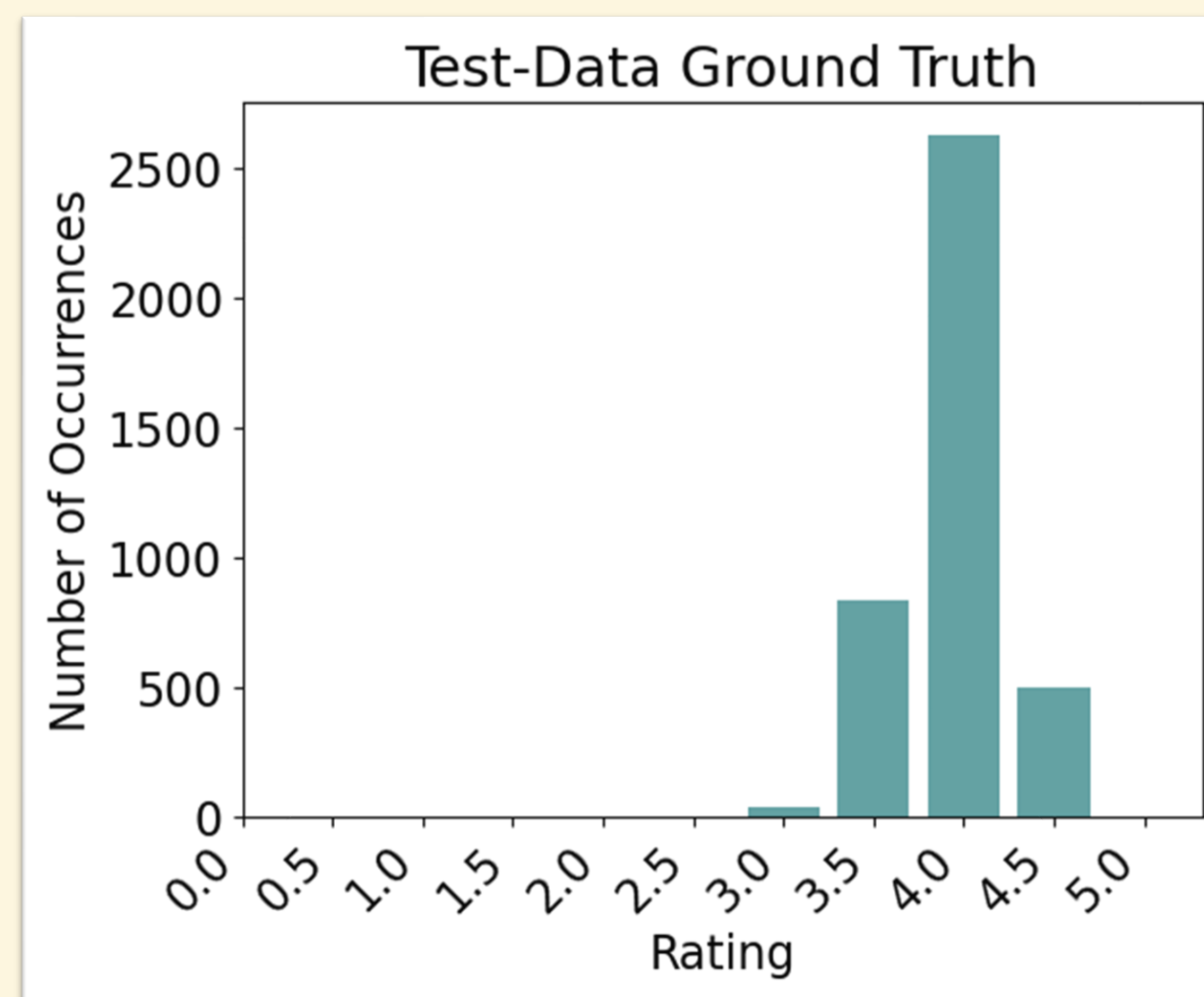
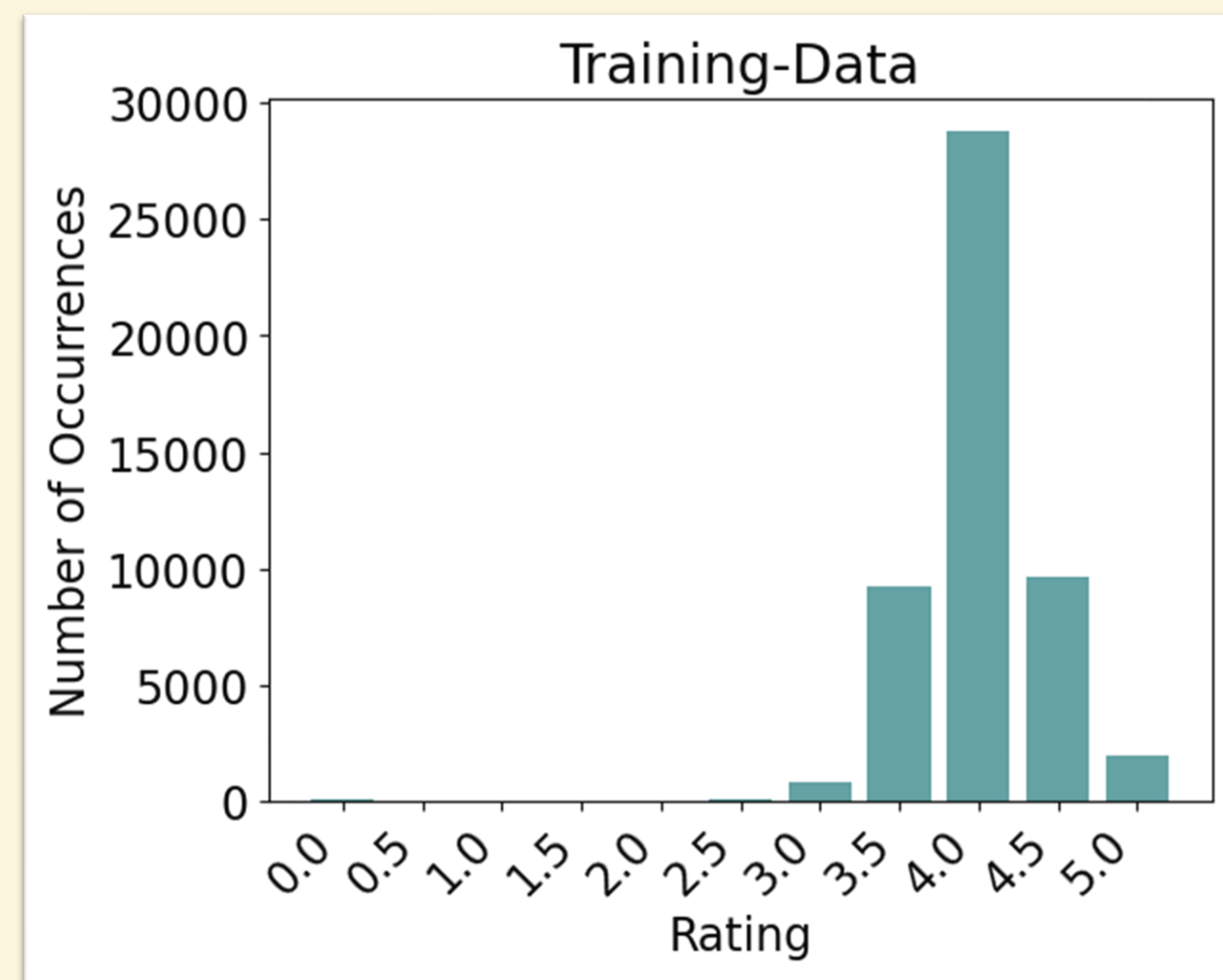


Timeline

- Updating a **Goodreads scraper** and scraping data
- Getting more data with the **Hardcover API**
- Sorting out data with **missing values**
- **Mapping** the Goodreads and Hardcover data with the **ISBNs** of the books
- **Clustering Genres** by hand
- **Binary** model to predict if book is **popular or not**
- **Multiclass** model to **predict rating**



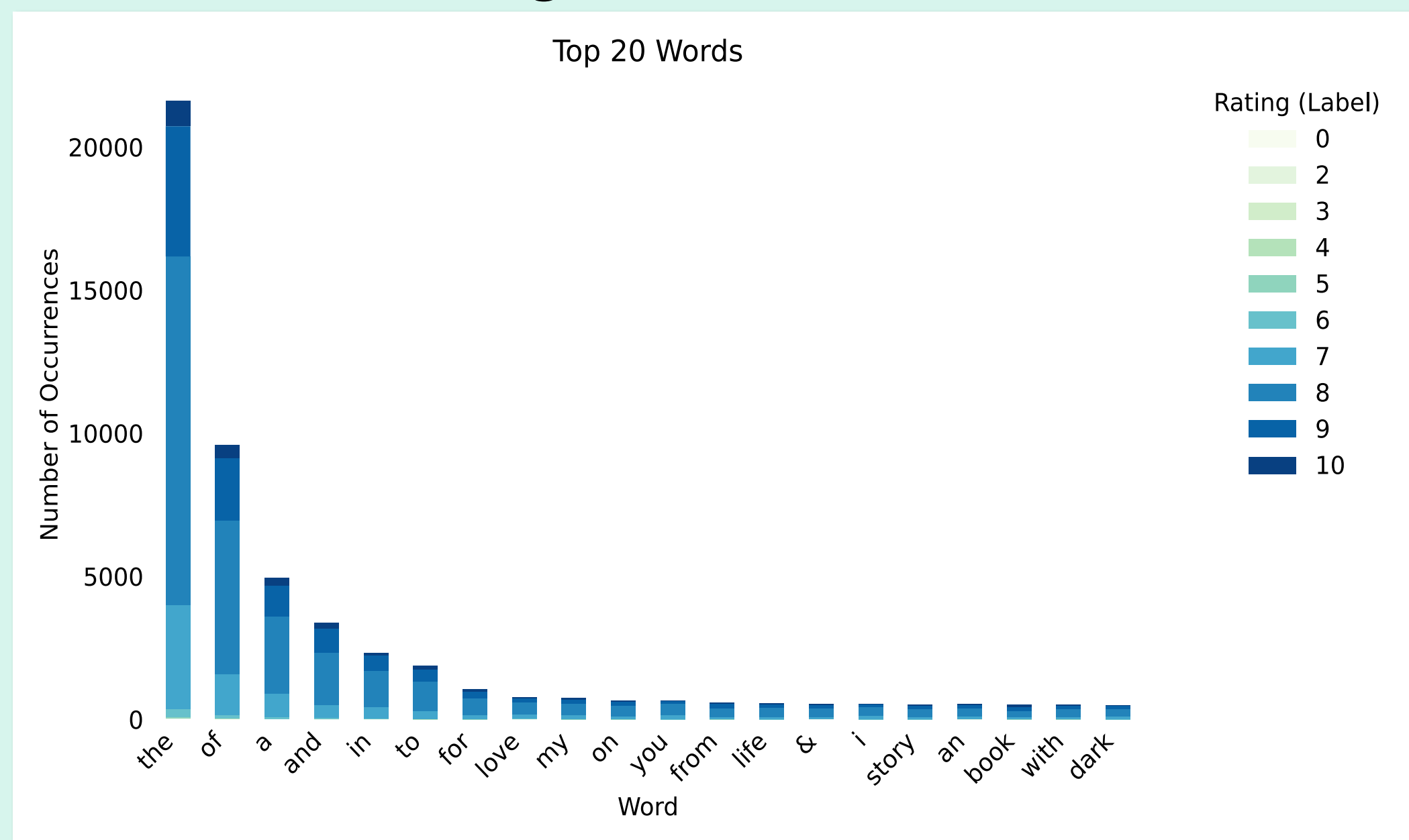
Multiclass Classification



- Classification of the Ratings in **0.5-steps**
- **Distillbert-Base-Uncased**
- **Dummy-Model** which uses the most frequent rating of 4.0

- **Dummy-Accuracy: 65%**
- **Modell-Accuracy: 66%**
- **Weighted-F1: 54%**

Which are the deciding factors for the models?

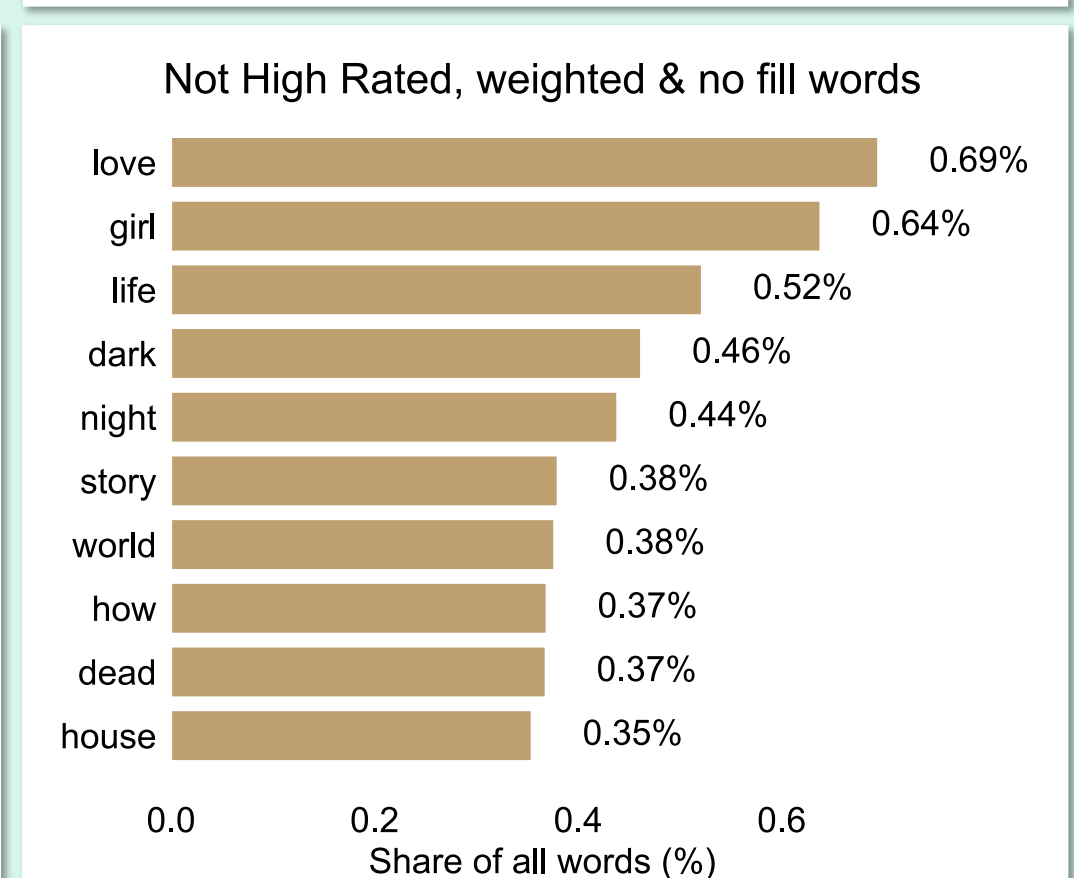
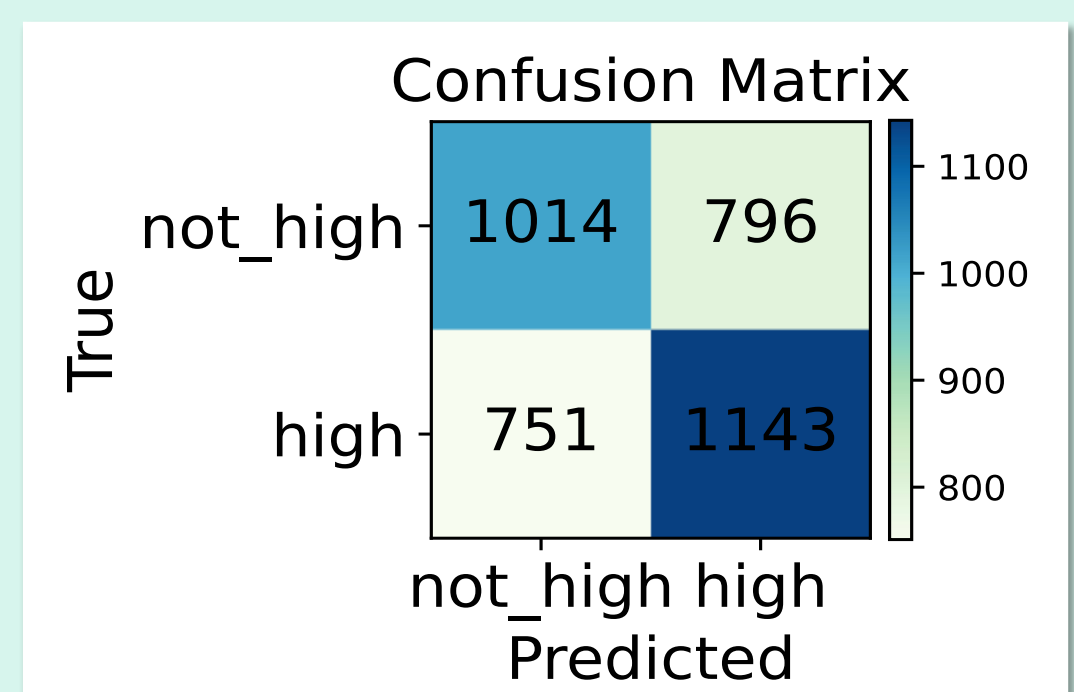
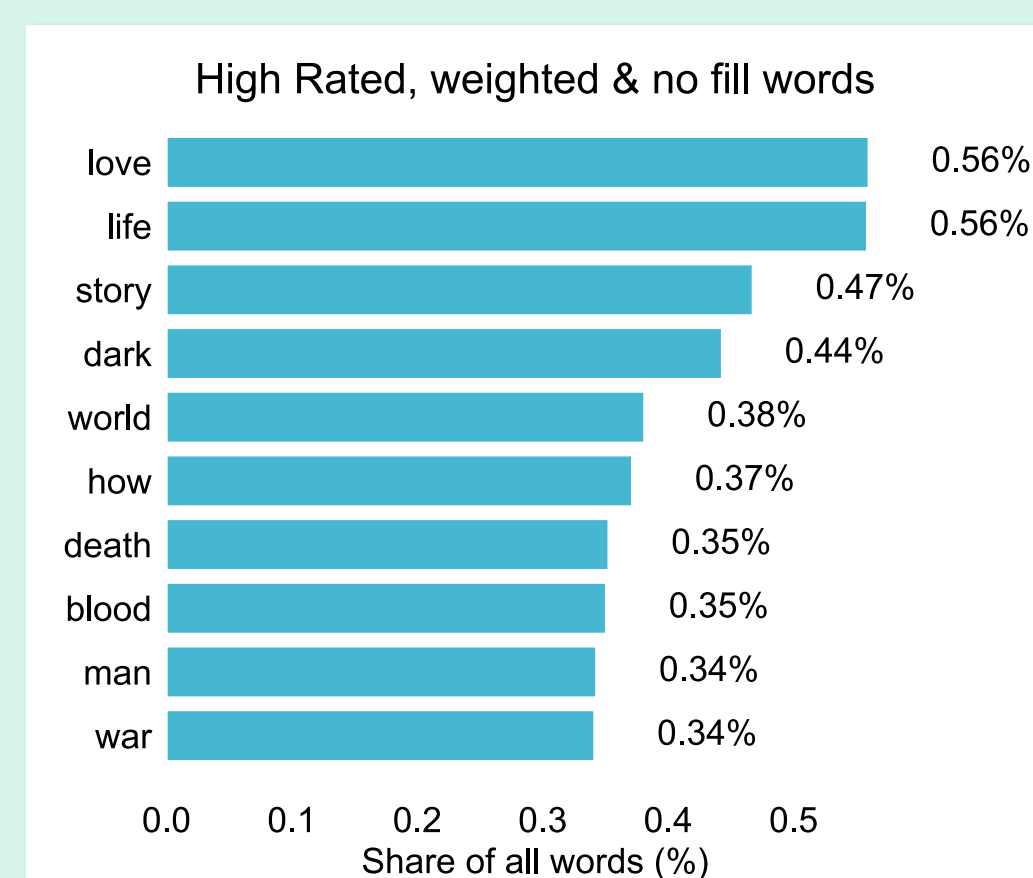


Filler words have a big influence on the models

Most of the words are labeled with a **rating of 4.0** (label of 8)

Binary Classification

- Threshold **100 ratings**
- Labeling classes (**Median 3.97**)
- **Bert-Base-Uncased** model
- **F1-Score: 60%**
- **Modell Accuracy: 58%**
- **Dummy-Accuracy: 51%**



Conclusion, Outlook & Learning Outcome

The model itself is **not better than guessing**, meaning the **titles are insufficient for popularity-prediction**. Other features (like author, description, genre, rating) could **improve the prediction** significantly. The **results** are most likely **influenced** by the fact that we choose **books from "Best of ..." Lists**, already hinting that these books probably **won't have a bad rating**.