PROJECT D10: KAGGLE MOVIE RATINGS

INTRODUCTION

For our project we made a Movie rating predictor. This project has the potential to be useful for both movie studios and individuals, as it can help to identify movies with high ratings.

GOALS

Train a machine learning model that is capable of predicting movie ratings based on different features such as genres and keywords. Our second goal was to find interesting facts from our chosen data.

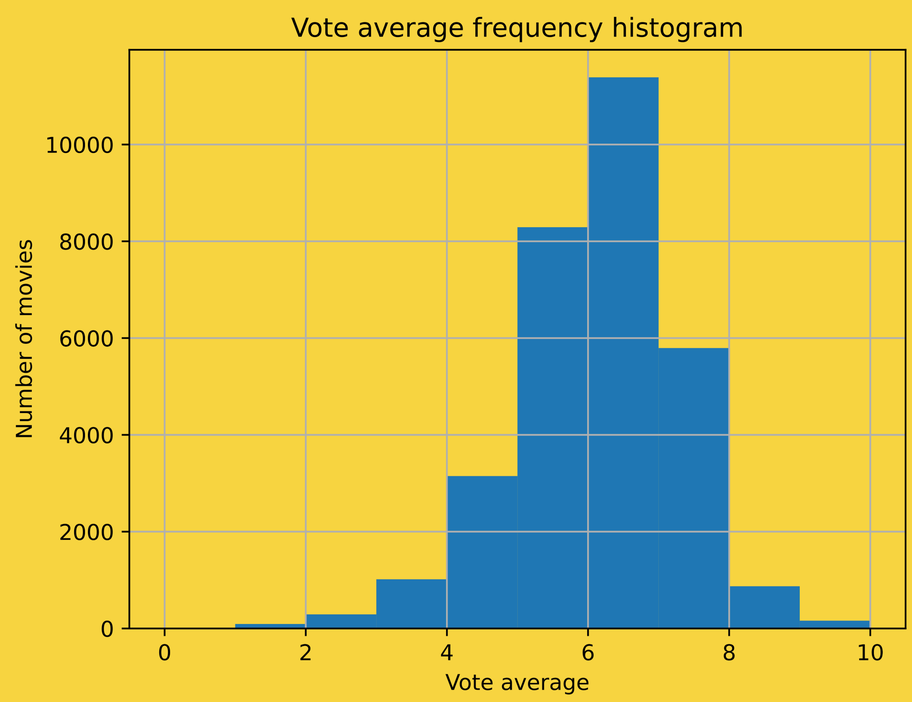
DATA

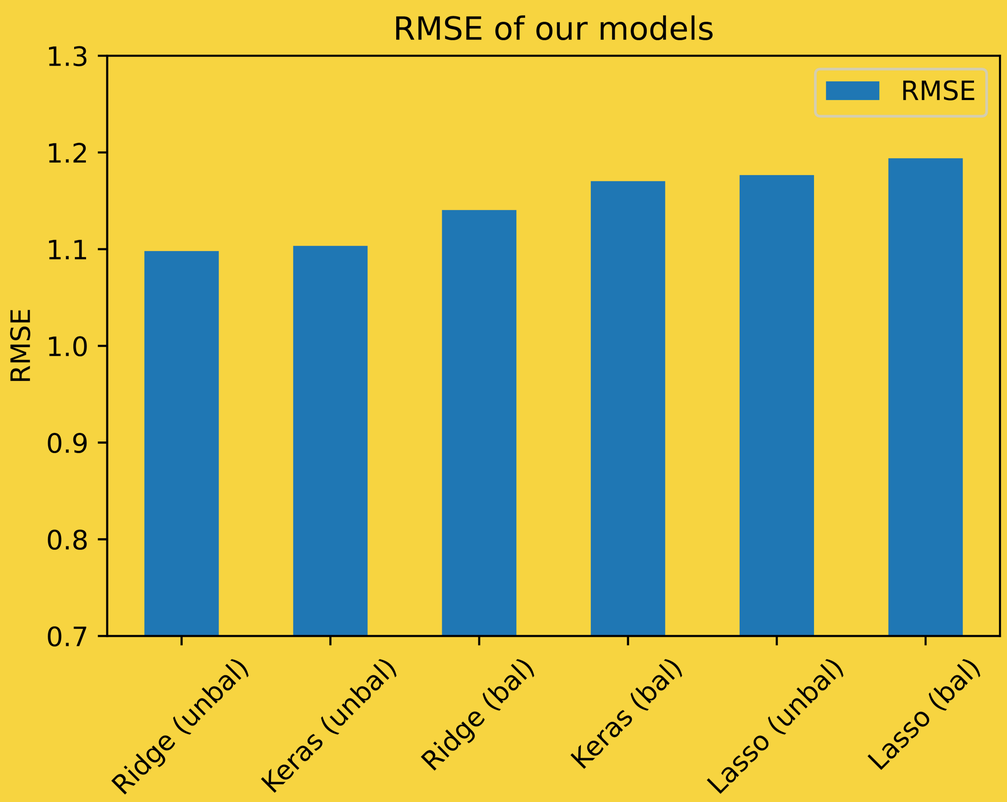
Our project is based on a Kaggle Movies Datasets, where the data was imported from IMDB. After data processing we had over 30 000 movies to train our model.

METHODOLOGY

We trained multiple different models using the following methods:

- Linear regression
- Ridge regression
- Lasso regression
- Keras neural network





PROCESS

We used standard data science techniques to preprocess and analyze the data. Then we split the project into two parts: in the first part we trained standard machine learning models, in the second part we trained a neural network.

CONCLUSION

Our most accurate model had a RMSE of 1.09. As our dataset was very imbalanced, then it was hard to get the model to predict accurately low and high ratings. Even after balancing the dataset, we did not get much improvement in performance. However, we got to experiment with neural networks, which definetly taught us a lot about deep learning.

KEVIN KLIIMASK
JENS JÄGER
TAAVI EISTRE