Summary Last Week - 2021.4.19

Wang Mingxi

Jilin University

College of Computer Science and Technology

2021-04-18

Overview

Job summary

Git Project

- The local repository is created on GitHub and three folders are created. Local and remote repositories have also been synchronized.
- Modified the text in the bottom right corner of the file by uploading the file to GitHub and then pulling it down.

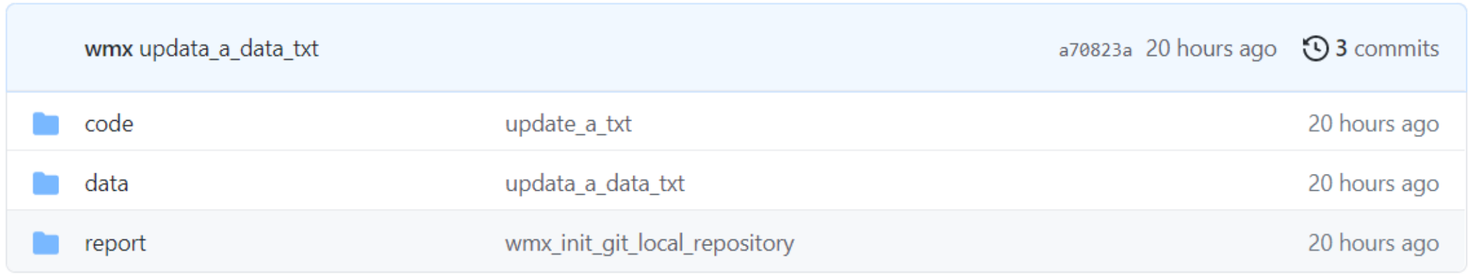

Figure 1: The repository on GitHub

Notebook Amend

I have read and studied one of XXX and have modified its contents.

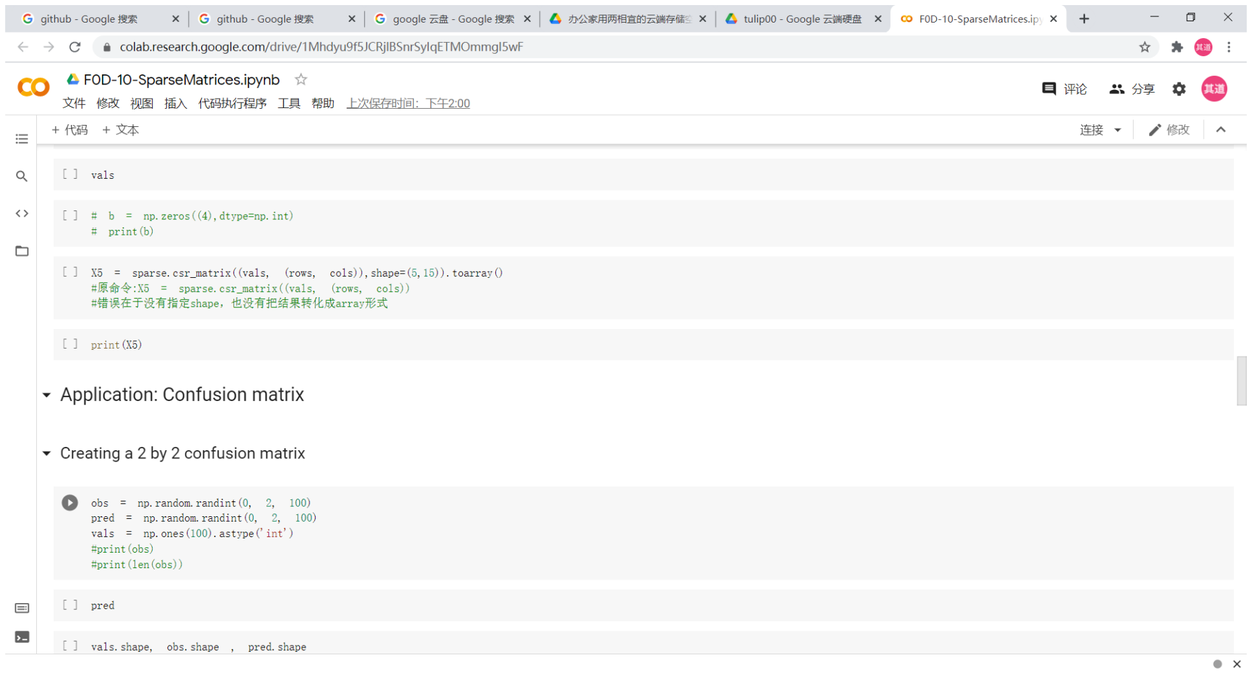

Figure 2: The repository on GitHub

Kaggle Subject

The San Francisco Crime Classification Forecast is a multi-classification problem that asks us to predict crime types.Given some characteristics, let the model predict the type of output crime based on the input.

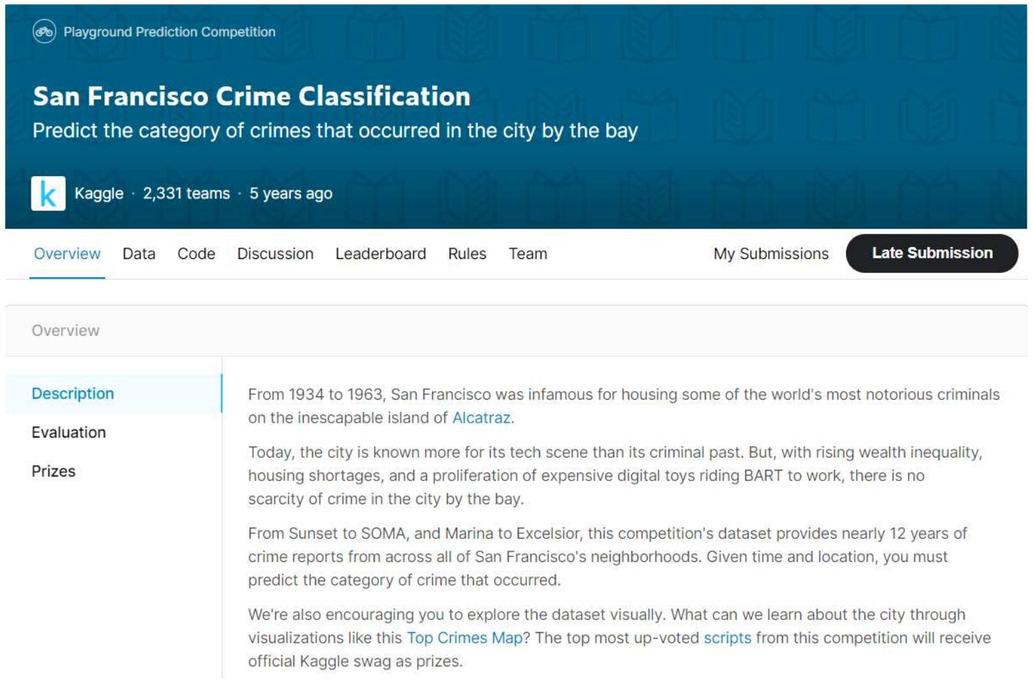

Figure 3: The repository on GitHub

Kaggle Subject

- The data analysis

The data includes the time of the crime, the place of the crime, the description of the crime (in text form), the final disposition, the type of crime, etc.Crime data correlate with crime types.The function of the model is to analyze the data set to find out the relationship and carry out the classification and prediction of crime types.

Kaggle Subject

- The data processing

The daily features can be decomposed into more detailed contents to obtain more detailed features, thus improving the accuracy of the model. For example, the date is disassembled from a comprehensive form of adults, months, days, hours, minutes, etc What we found was that there was a difference in the type of crime that was committed inside the building and at the intersection.Findings like this suggest that there are characteristics that have a direct impact on the predicted outcome, as will be explained in more detail later.

Kaggle Subject

- The data cleaning

Some data needs to be cleaned up.There are duplicate elements in the data set, and those are the ones that we want to get rid of.In addition, in some obviously against the common sense of the wrong data, we also want to delete.For example, we find that there are crimes that we know are located in the ocean by latitude and longitude analysis, and those are the few crimes that we don't care about.

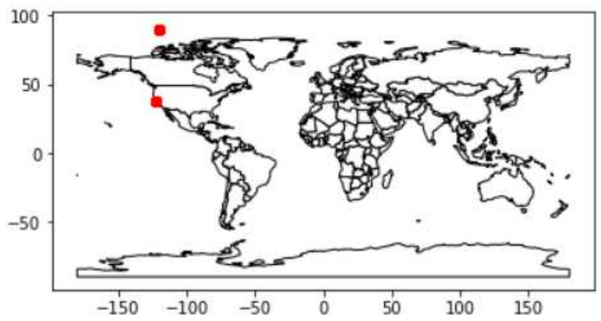

Figure 4: The repository on GitHub

Kaggle Subject

- Build a model

The decision tree was used to build the model. After 23 rounds of training, the model training was completed and the score on the test set was 2.3.However, there are still some unexplained errors and problems in the program operation. I will continue to modify them as soon as possible.

Next Week's Work

Work arrangement

I will modify the remaining notebook, and then complete the git repository. Also the Kaggle topic related process to continue to improve, and modify the code related bugs, complete the coding work

Thanks

Thankes!

Algorithm

Algorithm

Conclusion

Conclusion

Conclusion

Contact Information

Wang Mingxi

College of Computer Science and Technology

Jilin University, China

BLIU@TULIP.ACADEMY

TEAM FOR UNIVERSAL LEARNING AND INTELLIGENT PROCESSING