CSCI 4146/6409 - Process of Data Science (Summer 2023)

Assignment 2: Predictive Modelling

This task is designed to enable students to practice predictive modeling skills. In assignment 2, we'll build upon your previous data exploration work and focus on developing your predictive modelling skills. Using the same U.S. Airbnb Open Data dataset, you'll select a model, train it, evaluate its performance, interpret the results, and learn how to improve it through tuning. Additionally, you'll gain insights into the deployment process. This hands-on experience will provide you with practical knowledge of the complete data science process, from understanding the business context to deploying predictive models.

Due Friday Jun 02, by 11:59 pm - see submission details below.

Dataset: U.S. Airbnb Open Data- https://www.kaggle.com/kritikseth/us-airbnb-open-data

In this assignment, you will continue to work with the U.S. Airbnb Open Data dataset. You will go through the subsequent stages of a data science project defined in CRISP-DM.

1. [2.5] Predictive Modelling
    a. Select an appropriate predictive model for your problem. Explain your choice. [0.5]
    b. Partition the dataset into a training set and a test set. Describe the method you used for this. [0.5]
    c. Train the model on the training set. Discuss the parameters you used and why. [0.5]
    d. Evaluate the model's performance on the test set. Discuss the evaluation metrics used and the results. [0.5]
    e. Interpret the model. Explain what it reveals about the prediction subject and its domain concepts. [0.5]
2. [1.5] Model Tuning
    a. Explain how you can tune the model. What parameters can you adjust? [0.5]
    b. Tune the model. Explain the method you used and why. [0.5]
    c. Compare the performance of the tuned model with the initial model. Explain any changes. [0.5]
3. [0.5] Model Deployment
    a. What are the potential issues that might arise during deployment. Moreover, how would you address such deployment issues? [0.5]
4. [0.5] Conclusion
    a. Summarize your findings and the value of your solution to the business problem. [0.25]
    b. Identify potential areas for further analysis or improvements. [0.25]

Submission Details

Submissions should be made through Brightspace, adhering to the due date and time specified under the Assignments section. To prepare your assignment solution, make use of the provided assignment template notebook on Brightspace. The detailed requirements for writing and coding can be found within the evaluation rubric document available on the platform. Keep in mind that questions will be graded individually using letter grades, with their respective weights indicated in parentheses.

You may complete the assignment individually or with another individual. In the case of a pair submission, only one student should submit the assignment on Brightspace. Be aware that plagiarism detection tools will be employed to identify any instances of cheating or copying in both your code and the accompanying PDF.

Your submission should consist of a single Jupyter notebook as well as a generated PDF that contains the compiled results generated by the notebook. This PDF should include both the code and the results as part of the final printout. Name your files as follows:

- A2-<your_name1>-<your_name2>.ipynb and
- A2-<your_name1>-<your_name2>.pdf.

Failing to submit both files will result in a zero mark for both students.