Enterprise Analytics Office

Open Ended-Modeling Assessment

Las Vegas Restaurant Inspections

INSTRUCTIONS

The data you will be looking at represents restaurant inspections that took place in the Las Vegas metropolitan area. The original source of the data is located at the City of Las Vegas Open Data Portal. Inspections are performed using a Food Establishment Inspection Report.

For this exercise, you will be asked to explore two subsets of this data that have been manipulated for this exercise: TRAIN_SET_2021.csv and TEST_SET_2021.csv. **Explore the possibility of building a minimally viable product (MVP) model** to predict the outcome of a restaurant's next inspection.

TASKS TO BE COMPLETED

(1) **Conduct exploratory data analysis** of the Training Set. Provide an overview of the data set and underlying patterns you may identify. Without a thorough data dictionary, you may have to make some assumptions about the data. Document any transformations you perform.

(2) **Attempt to build an MVP model that predicts the outcome of a restaurant's next inspection, using NEXT_INSPECTION_GRADE_C_OR_BELOW as the response** – General restaurant information and data from the restaurant's most recent inspection has been provided. Use your knowledge of the data and your own statistical expertise to determine if an MVP model can be built using the available data. Document your thought process, the model techniques you considered, and your recommendation about whether a workable model can be built.

(3) **Apply your model to the test set** – Predict NEXT_INSPECTION_GRADE_C_OR_BELOW for the data in the TEST_SET_2021.csv file.

(4) **Provide recommendations for how you would enhance the data set to improve the predictive power of the model** – Assume "the sky's the limit" and write a one-page recommendation on how the data set could be improved. Provide at least 8-10 additional attributes and prepare to defend your choices. (It may be helpful to perform additional research online as time allows).

YOUR DELIVERABLES

(1) **Report** – Develop a report including your exploratory data analysis, modeling process and recommendations. There are no length requirements for this report - the amount of relevant insights you derive will determine its length. Your report should include your original code and be generated in HTML format via R Markdown, Jupyter notebooks, or similar.

(2) **Predictions** – For your selected model, save your predictions to a CSV file, with **only** comma delimiters. The file should include only three columns: RESTAURANT_SERIAL_NUMBER, CLASSIFIER_PROBABILITY, and CLASSIFIER_PREDICTION. The serial number should be a character data type and the predictions should have real values. The file name must take the form: predictions_*Lastname_Firstname_Intern.csv.* For example, Bob Smith would submit his predictions in the file predictions_Smith_Bob.csv.