CS3 - DDoS Anomaly Detection Case Study Rubric

DS 4002 - Spring 2025 - Christine Tsai
Submission Format: Presentation and GitHub Repository

Purpose: Completing this case study will allow you to practice various time series analysis techniques to show you potential applications to solve real-world problems. Your goal is to compare two modeling techniques, specifically ARIMA and an autoencoder, and determine which model is the most effective at detecting anomalies. By the end of this assignment, you will have had the opportunity to perform exploratory data analysis and model-building on multiple datasets obtained from real server metrics.

Task: You will use the supporting materials in the GitHub repository to follow a process of analyzing the data and building two models using Python on the three provided datasets. Once you complete this project, you will be able to report to your company which approach you think is best for detecting anomalies that are indicative of a DDoS attack, and use visualizations related to your models to support your argument.

Tips for Success: Document your thought process when making certain decisions. Comment your code especially in areas that might be harder to interpret upon an initial glance. Make sure your visualizations effectively communicate what you want them to depict.

Deliverables: You will generate a report that discusses your process that led you to your decision on which model you think is best for the purposes of anomaly detection. Your report willbe organized and contain ample amounts of visualizations especially for the exploratory data analysis and results portion of your analysis.

Spec Category	Spec Details
Repository and Submission	You will submit a complete and organized GitHub repository that contains the following components: ● All of the scripts that were used ○ Include comments throughout, and for each script, make sure to include a docstring or heading that gives an overview of the purpose of the script ● All of the data ○ Include the different versions of the data that you use

	throughout the analysis process if you end up cleaning the data ● Exploratory data analysis visualizations ○ Include plots to illustrate characteristics of the data such as the distirbution, etc. or whateber you think is most fitting ● Analysis visualizations ○ Create plots that depict the model in action. For example, your plots can distinguish data points that the models deemed as anomalies from non-anomalous data ● Intuitive document names and good file organization ○ A tree of the structure of the repository that shows where documents can be located should be embedded in the README file for the repository ● References to articles or websites that you used throughout your assignment should be included ● Give a high-level overview of the main steps you followed to reproduce the results in the case study in your README
Data Preparation and Cleaning	Determine whether the data needs to be cleaned and then produce visualizations for each of the data sets
Exploratory Data Analysis	Visualizations outputted with each of the scripts for plotting ● Create informative plots by executing the scripts corresponding to each dataset ● Discuss any initial trends found in the data
Time Series Preparation	Determine whether the data needs to be

	transformed after performing a stationarity analysis
Modeling/Anomaly Detection/Evaluation	Successful ARIMA model fitting and anomaly detection and successful autoencoder model fitting and anomaly detection. ● Ensure optimal model parameters were selected ● Ensure threshold for what value is considered an anomaly is calculated accordingly ● Discuss the confidence in each of the models