Exploratory Data Analysis

HARVEY KWONG and JACOB DEROSA, University at Buffalo, USA

1 Introduction and Problem Statement

The rise of antibiotic resistant infections poses a significant challenge to global public health, and gonorrhoea is one of the most concerning pathogens due to its rapidly increasing resistance to available treatments and its difficulty to be detected. This project focuses on predicting antibiotic resistance in Neisseria gonorrhoeae, the bacteria responsible for gonorrhoea, using subsets of its DNA sequences as predictive features. The primary objective of this project is to explore the relationship between bacterial DNA segments and resistance patterns, and to identify key genetic markers that could possibly predict resistance. By performing data cleaning and exploratory data analysis on a dataset containing DNA sequences and resistance outcomes, we aim to address the following question: Which DNA segments are most associated with antibiotic resistance?

This project's contribution is very important as it helps to clarify biological data, making it easier for future studies to focus on effective solutions. By gaining a deeper understanding of the data and identifying key variables, this project serves as a step towards more advanced research on antibiotic resistance.

2 Data Sources

We utilized the "Predicting antibiotic resistance in gonorrhoea" dataset from Kaggle: (https://www.kaggle.com/datasets/nwheeler443/gono-unitigs/data).

3 Data Cleaning/Processing

4 Exploratory Data Analysis

5 Title Information

Authors' Contact Information: Harvey Kwong, harveykw@buffalo.edu; Jacob DeRosa, jderosa3@buffalo.edu, University at Buffalo, Buffalo, New York, USA.