JMI Labs

Large Data Analysis – Project Proposal (Abstract)

This project for JMI Labs will entail predicting minimum inhibitory concentration (MIC) resistance categories based on provided patient data that includes genetic information and the existence of an extended spectrum beta-lactamase (ESBL) enzyme and/or a carbapenemase enzyme for specific isolates. The MIC values will be separated into four categories: 0 for *non-beta-lactamase*; 1 for *resistant*; 2 for *intermediate*; and 3 for *susceptible*. Categorical predictions will be calculated using a neural network. This program will evenly distribute known MIC values within the training data in order to prevent over-fitting within the model. The architecture of this model will utilize the rectified linear unit (ReLU) as the activation function for the input and hidden layers while using softmax as the activation function for the output layer. The network will also employ a dropout method within each hidden layer; the learning rate, as well as the dropout percentage, will be selected through systematic testing of learning rates between 0.0001 and 0.1 and dropout rates of 0.0 and 0.8. Two metrics will be used for the comparison of experimental models—the accuracy metric to help provide graphical visualization and the F1 score to provide an accurate statistical analysis of the precision of the output data. Final analysis of the output data will seek possible correlation between beta-lactamase values and the presence of specific genes. Further data exploration may analyze potential influence from demographic information—biological sex, location, age, etc.—as additional input nodes for enhanced prediction accuracy.