CS 440 Artificial Intelligence Project Proposal
Project 26: Bioactivity Of Small Molecules Prediction

Team Members:
1. Laasya Vajjala - 11848603
2. Lalith Mohan Midde -11848601
3. Sri Varsha Adavath - 11860642
4. Surya Pramod Vadapalli - 11861342

I. Project Statement:

Project Goal:

The project aims to create an intuitive tool employing a Bayesian network to swiftly predict the bioactivity of small molecules against target proteins, thereby expediting the drug discovery process. We have access to a dataset containing information on known 100 small molecules, their molecular descriptors, target protein activity, and bioactivity. The dataset will be preprocessed and split into training, validation, and testing sets.

The application of this project lies primarily in the field of drug discovery and development. The developed tool can assist researchers in identifying potential candidates for further experimental validation and drug design. Additionally, the tool could also find utility in virtual screening of compound libraries, prioritizing compounds for experimental testing, and optimizing lead compounds for improved efficacy and safety profiles.

For evaluating the performance of our Bayesian network model, we will employ several key metrics commonly used in binary classification tasks. These metrics include accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC). In our comparative analysis, we will compare the performance of our Bayesian network model with other machine learning algorithms commonly used for bioactivity prediction tasks, such as support vector machines (SVM), random forests, and deep learning approaches.

As an alternative method, if the Bayesian network model encounters limitations, we will consider employing a deep learning approach, specifically a convolutional neural network (CNN) or a recurrent neural network (RNN). Additionally, ensemble learning techniques such as stacking or boosting could be explored to combine the strengths of multiple models for improved accuracy and robustness.

II. Team Roles:

1. Lalith Mohan Midde - Incharge of data preparation and processing
2. Surya Pramod Vadapalli - Incharge of model development
3. Sri Varsha Adavath - Incharge of model Integration
4. Laasya Vajjala - Incharge of final testing, evaluation and documentation

III. Timeline:

Week 1: Data Preprocessing and Feature Engineering

Team Meeting 1: Discussion regarding data preprocessing tasks, including cleaning, handling missing values, and standardizing features.

Week 2: Model Implementation and Tuning

Team Meeting 2: Determine the architecture and parameters of the Bayesian network model. Discussing the tools for implementation and parameters for hyperparameter tuning, and model validation.

Week 3: Development of User Interface and Integration

Team Meeting 3: Plan the design and functionalities of the user interface for the tool. Discussion regarding incorporating input fields for molecular descriptors and displaying prediction results.

Week 4: Final Model Evaluation and Documentation

Team Meeting 4: Reviewing the performance of the Bayesian network model using evaluation metrics such as accuracy, precision, recall, and AUC-ROC. Discuss any refinements or adjustments needed. including cross-validation experiments and comparative analyses with other machine learning methods. Finalize the documentation, including detailed descriptions of the model architecture, implementation steps, and experimental results.