Predicting side effects of unknown drugs using the Simplified Molecular Input Line Entry System (SMILES)

By Adhitya Balaji and Mohit Manchella

Background

What does our application do?

- Displays the structural representation and the SMILE of a compound given the compound ID number assigned to it
- Predicts the presence of side effects of a particular drug
- Reports data regarding the accuracy of the model's prediction

Practical Applications

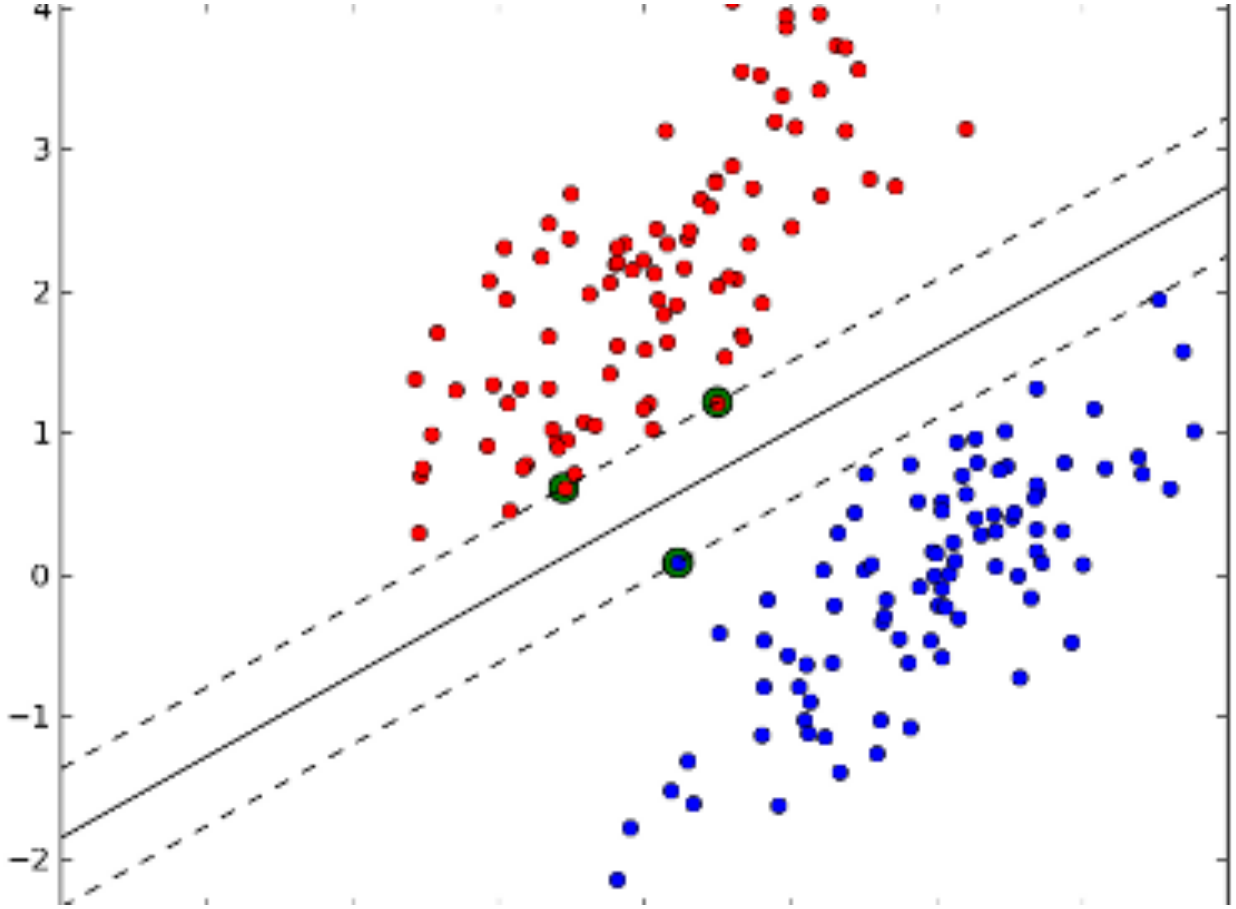
Why use this application?

- Often times side effect testing for drugs is an expensive and lengthy procedure
- Saves time by cutting down on side effect clinical testing expenses
- Warns the producers of a possible side effect to look for
- Confirms discovered side effects of FDA approved drugs

Methods

- Used a Support Vector Machine (SVM) classification model to predict side effects
- Separates data points into two groups based on a "best fit line" or ***decision boundary***
- The value of the unknown data point can be predicted based on it's location relative to the decision boundary

SVM Example



Code

- All code was written in Python
- Imported Python libraries
  - RDKit, Scikit, Tkinter
  - Purpose of importing libraries
- SVM
  - Divide the dataset into training and testing models
  - Train the model
  - Predict and find four metrics
  - Import

Live Demonstration

Further Improvements

- Incorporated more machine learning techniques
  - Linear, K-Nearest Neighbors, Decision Trees, Ridge, Lasso, etc.
- Expand the dataset
  - Use data, such as the compound structural representation, besides SMILES for side effect predictions
- Minor GUI Adjustments
  - Make the program more user-friendly and organized

Acknowledgements

We would like to thank the CTSI, Project STEM, and the IUPUI School of Science Department of Computer and Information Science for the generosity and opportunity this summer. Without their support, this research done by all of us would not have been possible.