**Drug Usage Prediction Program**

**Goal of the Project**

The Drug Usage Prediction Program was made with the goal of monitoring the usage of drugs throughout the United States and predicting the future rise of the usage of the substance. The project should be able to use pre-existing data about the usage of drugs and other substances in the United States and visualize the current and future usage of drugs throughout the country.

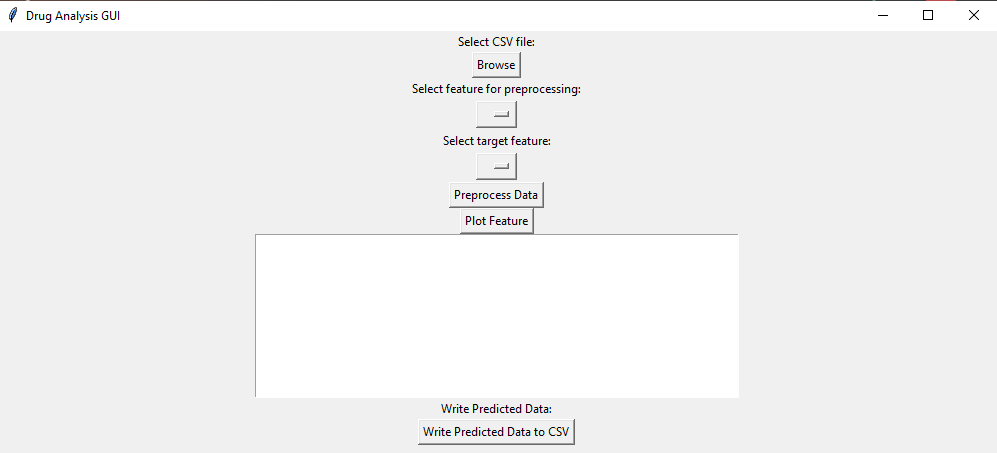
**Significance of the Project**

The Drug Usage Prediction Program is meant to aid addiction specialists, especially those who specialize in drugs in specific. The usage of the Drug Abuse Project will allow them to determine in the future what areas need help the most by allowing them to look at statistics such as how many users there may be in a specific state, or by which rate or usage demographic is most likely to use drugs or substances. Allowing them to look into the future of what may happen in the future may allow them to curb the growth of users in any means necessary.

The Drug Abuse Project does this by use of artificial intelligence, which is done by a model provided by Scikit-Learn. The model in question is the linear regression model, which specializes in cases of regression and classification.

**Installation and Instruction on How to Use**

The project will not be able to run unless you have installed specific python modules using pip. Alternative methods are available at the links provided.  
Pandas: <https://pandas.pydata.org/docs/getting_started/index.html>Scikit-Learn: <https://scikit-learn.org/stable/install.html#install-official-release>Matplotlib: <https://matplotlib.org/stable/users/getting_started/>Portalocker: <https://pypi.org/project/portalocker/>

Once the modules are installed, the project can be run.  
To allow the program to do its work, select a CSV file after selecting the ”Browse” feature.

Code Structure

Our code structure is built to read the csv file first, then preprocess the data to clean up.

Then, we process that data and split it into training and testing sets. Our model is a linear regression model that we use to feed the data into and predict trend of drug usage rates in the upcoming future. And then we apply operating systems concepts to the project and demonstrate file management, and file locking as well as process termination implemented as confirmation box before the program stops completely.

Functionalities and Test Results

The project calculates the mean squared error while predicting the trend in the future usage of drug. It also writes the predicted data into a csv file.

A screen shot of a graph

Description automatically generatedA screenshot of a computer

Description automatically generated

A close up of a computer screen

Description automatically generated

A screenshot of a computer

Description automatically generated

Discussion and Conclusion (Includes things like limitations and such)

So, there are some limitations that this project runs into. When it came to applying to multiprocessing, we wanted to see if we could apply that into processing data, but it couldn’t work due to the nature of the csv file and the data that is contained in it. We left that out. As for locking, it is used excessively like how preprocessing, model training, and evaluation are wrapped in locks, which could lead to unnecessary performance.

An error message pops up when plotting data. It pops up because there is a column of data that's in string form. We’ve tried dropping the data column to resolve this issue, but it causes another error saying the column doesn’t exist and will not proceed with plotting the data. We would have like to resolve this issue but due to time constraint, we could not entirely resolve this problem.

The preprocessing buttons will teleport down to the bottom of the message box when a csv file is uploaded by the user. The top button is the X column and the bottom button is the Y column respectively. I believe it is an issue with the drop down menu function in our structure.

Contributions:

Liam: worked on data model and training with preprocessing as well contributing to presentation and report

Binh: Worked on file handling, locking, feature extraction, GUI, data visualization, presentation and report.

Note: Binh submitted almost all the contributions because Liam’s machine could not push commits to the repository due to configuration of his machine. He gave me the code from his branch and Binh pushed it