**Machine Learning – Assignment 2(KNN Prediction)**

**Ganesh G (25MML0043)**

**https://drive.google.com/drive/u/3/folders/1mJeHBra-Ux5RHaG8-ymCq7bd67hj4NzU**

**KNN - K-Nearest Neighbors.**

**AIM:**To develop and simulate an online learning model for wine quality prediction that can incrementally update its parameters as new samples arrive, while effectively handling missing values and adapting to concept drift in wine quality standards over time.

**INTRODUCTION:**

In traditional machine learning systems, models are trained once on a fixed dataset and then deployed for prediction. However, in many real-world scenarios, data arrives continuously, and the underlying patterns may change over time — a phenomenon known as concept drift. Wine quality prediction is one such domain where continuous updates are valuable. New wine samples, changes in production techniques, and evolving consumer preferences can all alter the relationships between chemical features and perceived quality.

**REAL WORLD APPLICATIONS:**

**Dynamic Quality Control in Food & Beverage**

* Similar techniques can be applied to beer brewing, coffee roasting, or dairy production, where chemical properties shift due to seasonal changes in raw materials.

**Agricultural and Vineyard Management**

* Continuous learning models can process weather, soil, and grape growth data to predict future wine quality and guide harvesting decisions.

**Supply Chain Optimization**

* Distributors can forecast quality degradation over time, helping decide optimal storage conditions and delivery schedules.

**Fraud Detection in Wine Industry**

* Models can be trained to detect anomalies in chemical profiles, helping prevent counterfeit wine production.

**ALGORITHM:**

1. Import the necessary libraries: pandas, numpy, matplotlib, seaborn, and modules from sklearn.

2. Load the dataset using loading function.

3. Check for missing values and Handle missing values using imputation techniques.

4. Apply different scalers and observe how different scaling methods impact model performance.

5.Split data into training and testing sets(e.g., 80% train, 20% test).

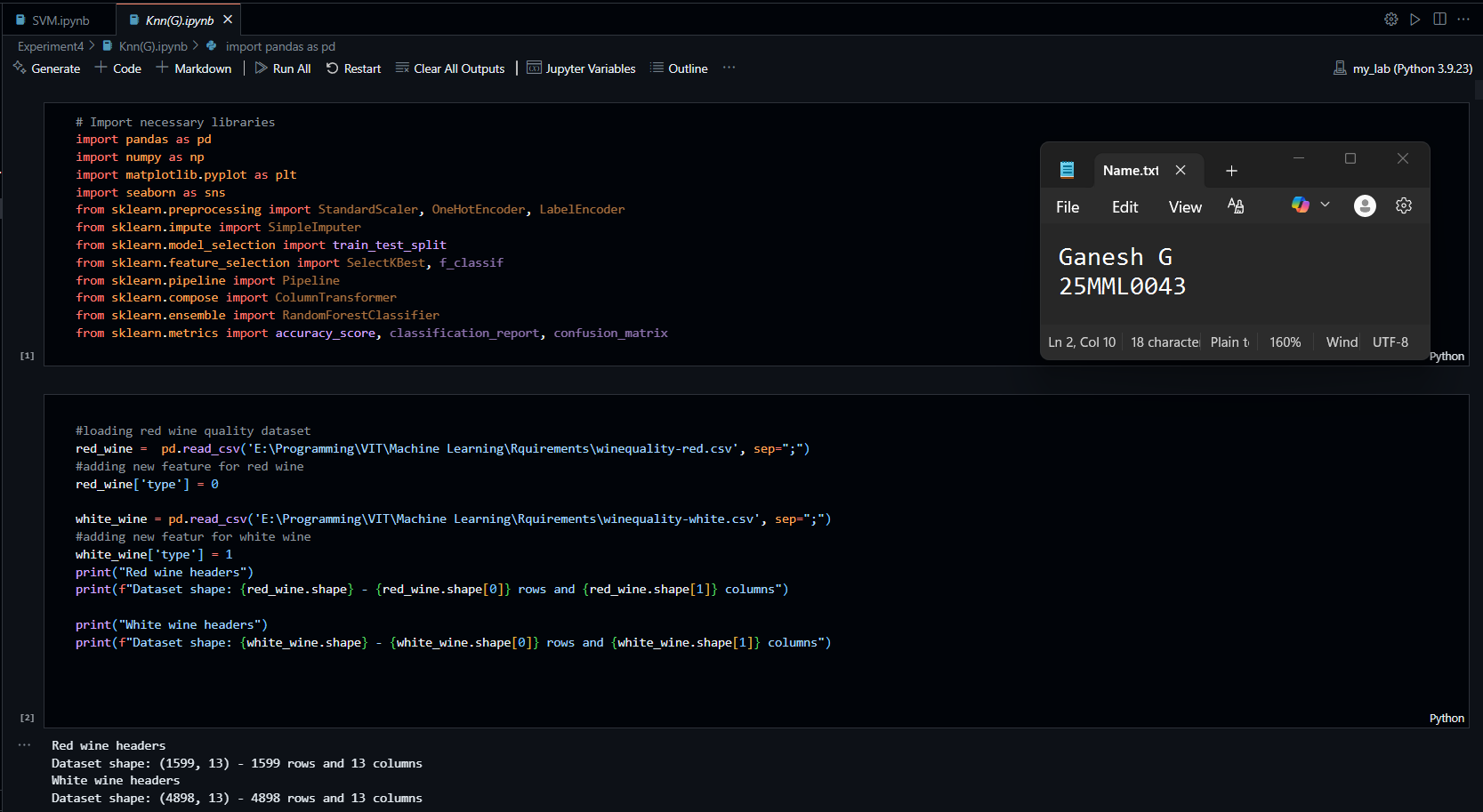
6. Initialize KNN classifier and train the model on the training set.

7.Predict the quality of wine on the test set and Evaluate model using metrics such as Accuracy, Confusion Matrix.

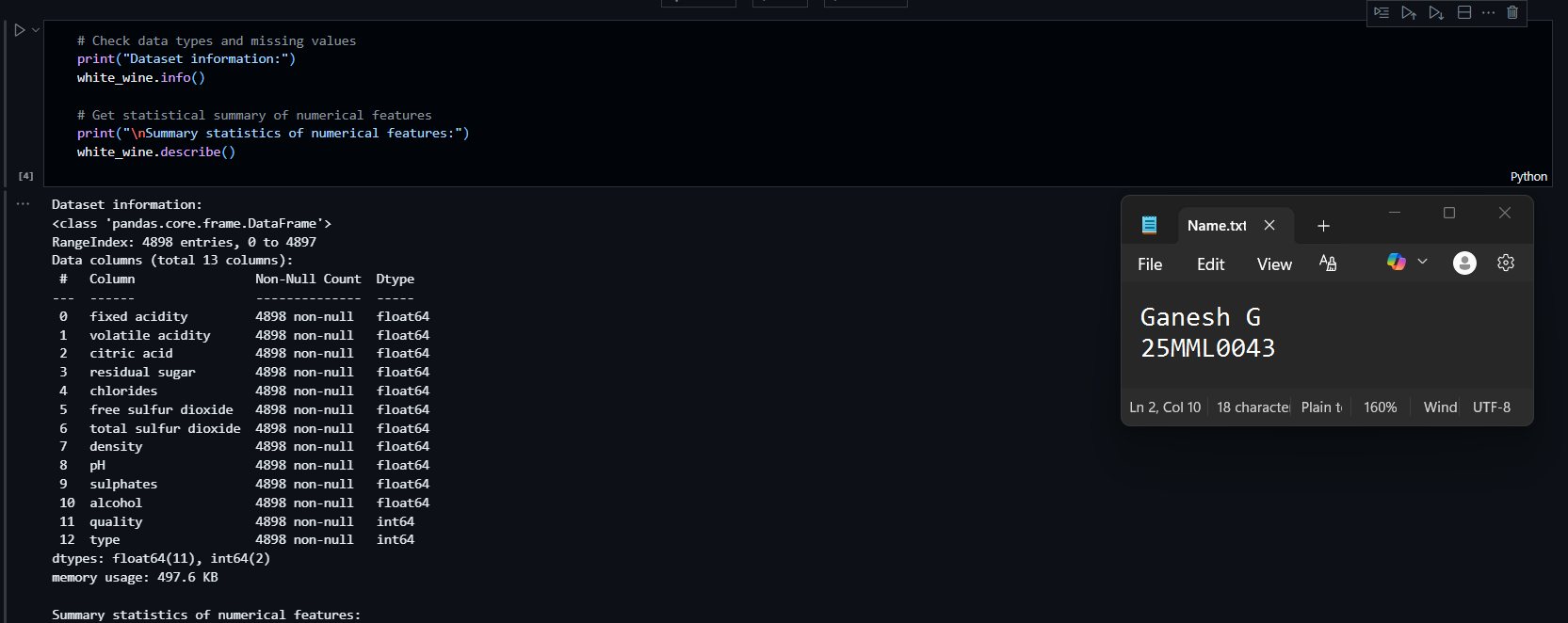
8. Compare performance across different scaling methods and identify which scaler works best with KNN for this dataset.

**Implementation and results**

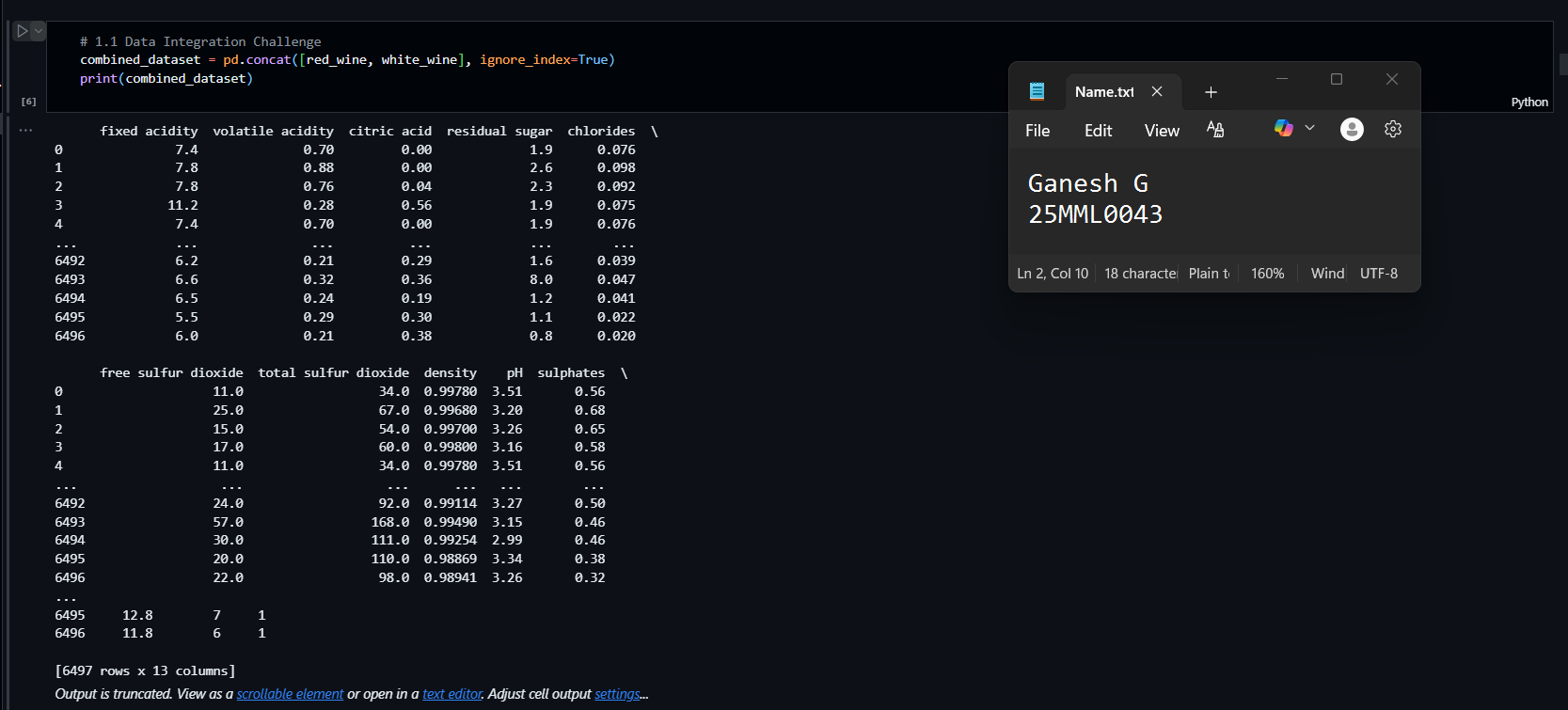
* 1. Load dataset

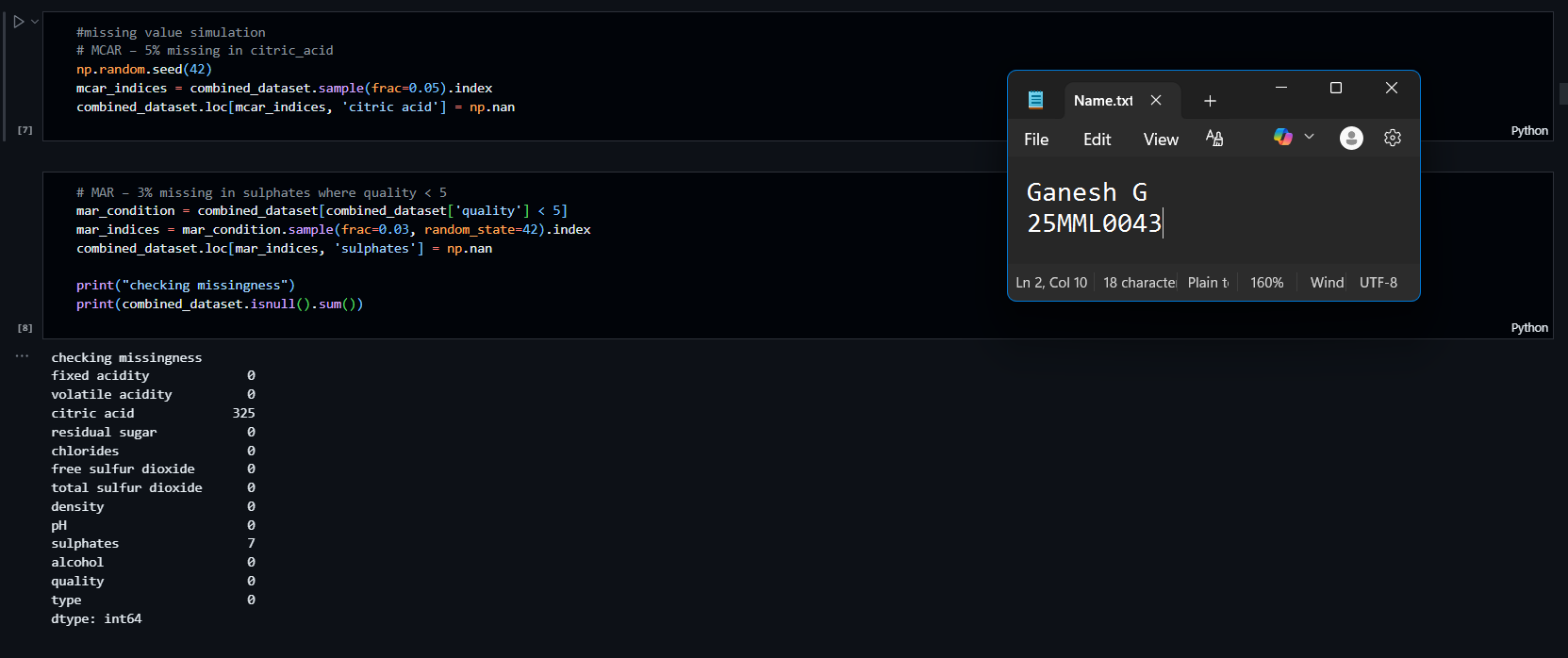
****

* 1. Checking missing values



* 1. Data Preprocessing

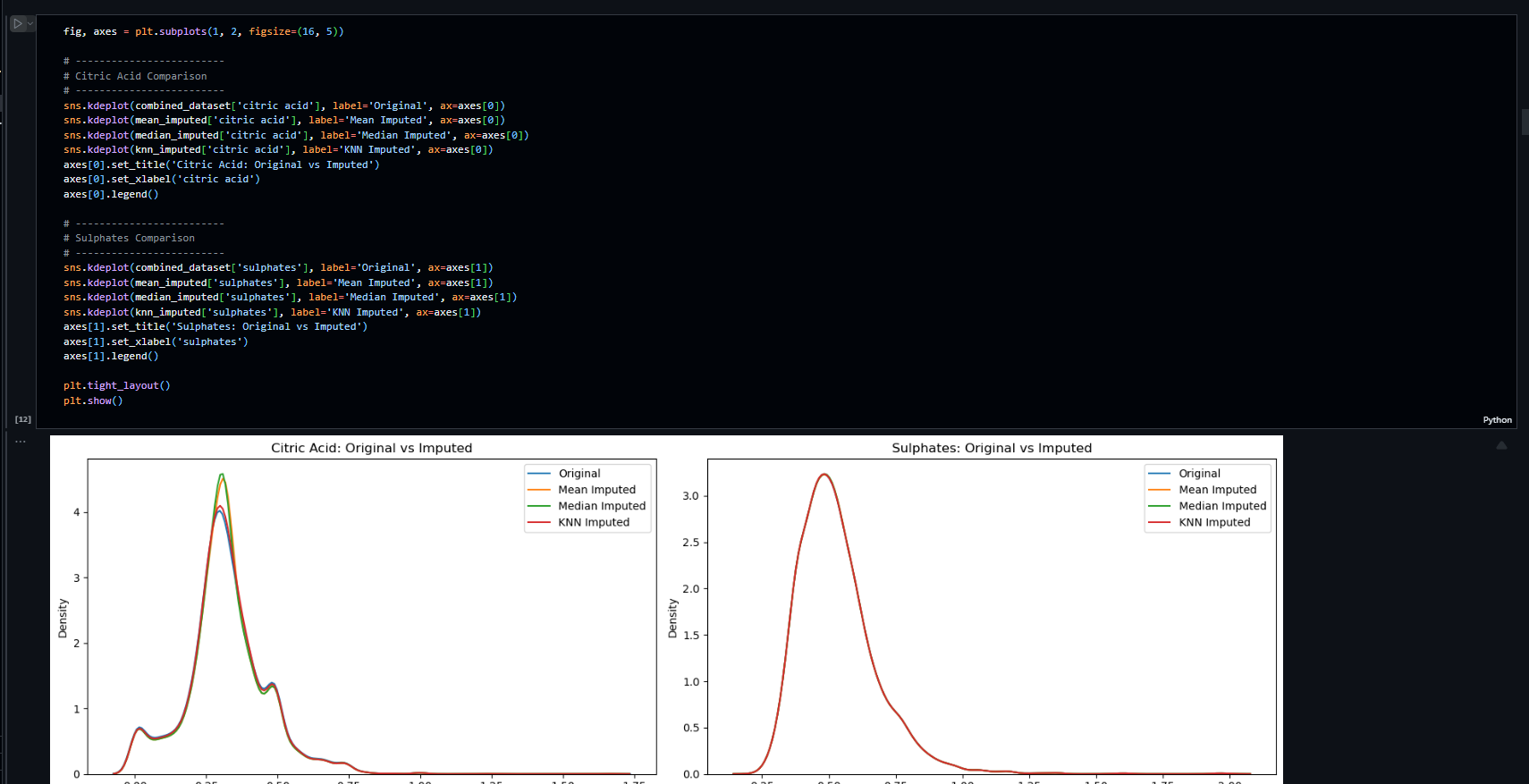




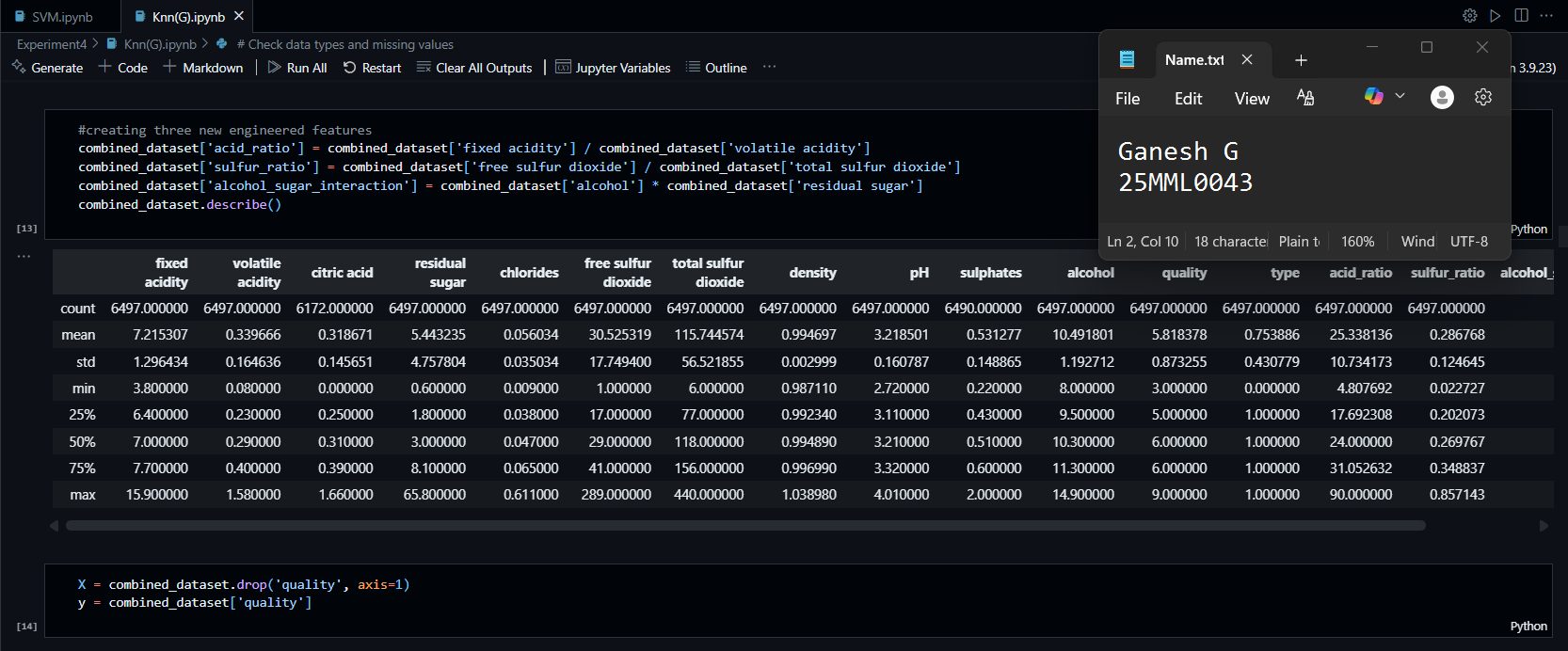
* 1. Imputation strategies



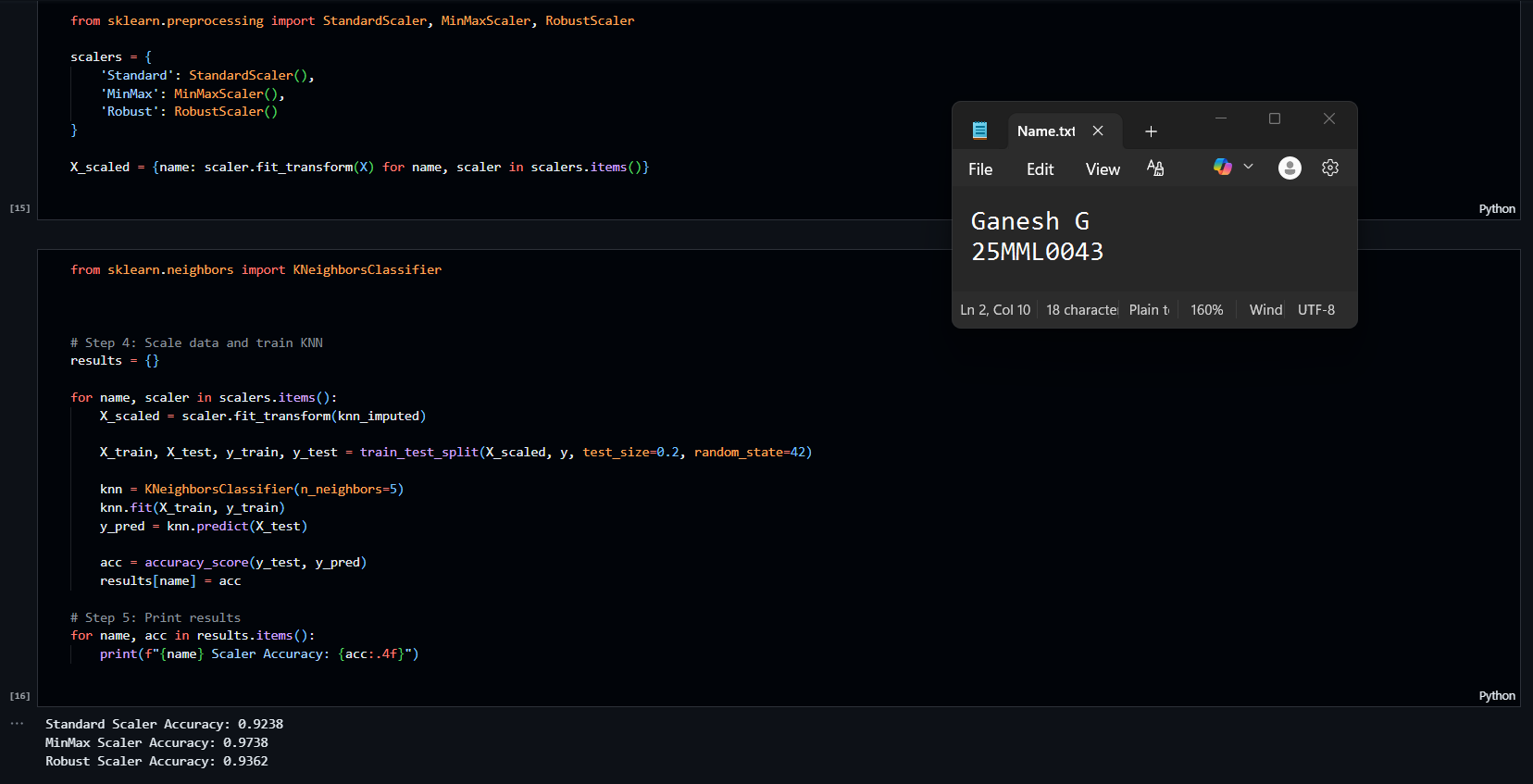
* 1. Comparison



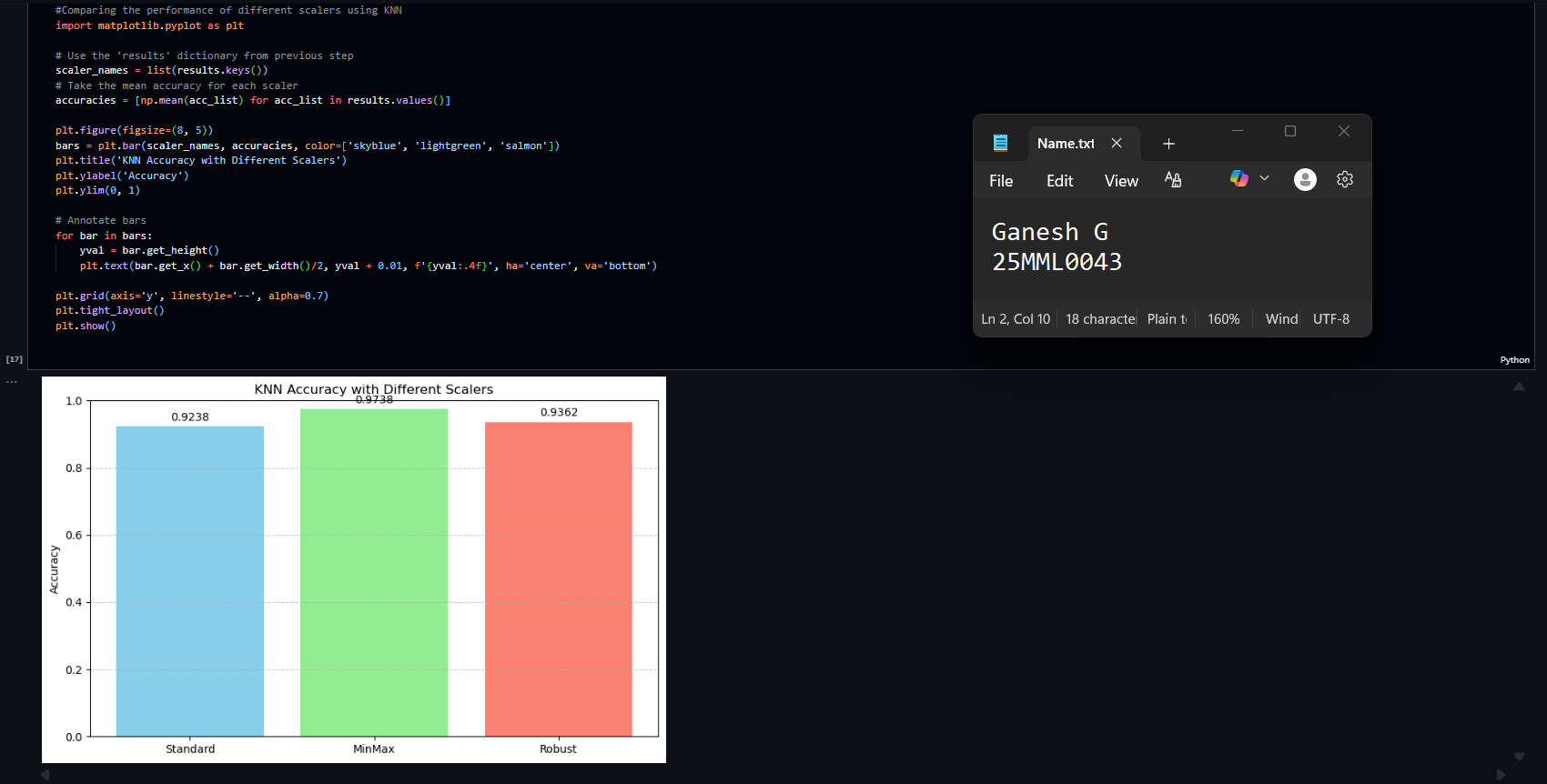
* 1. New Engineered Features



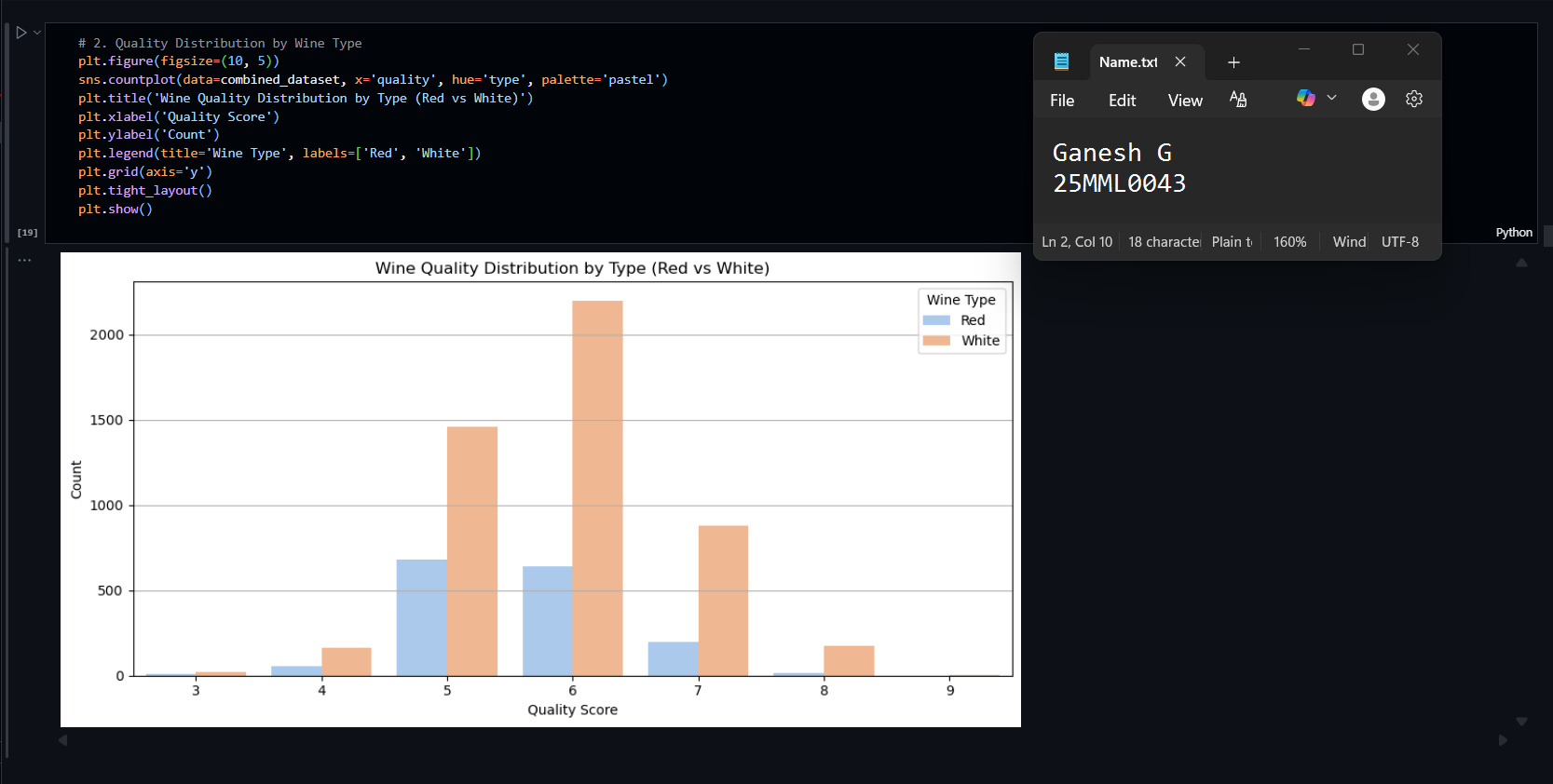
* 1. Scale data and train KNN



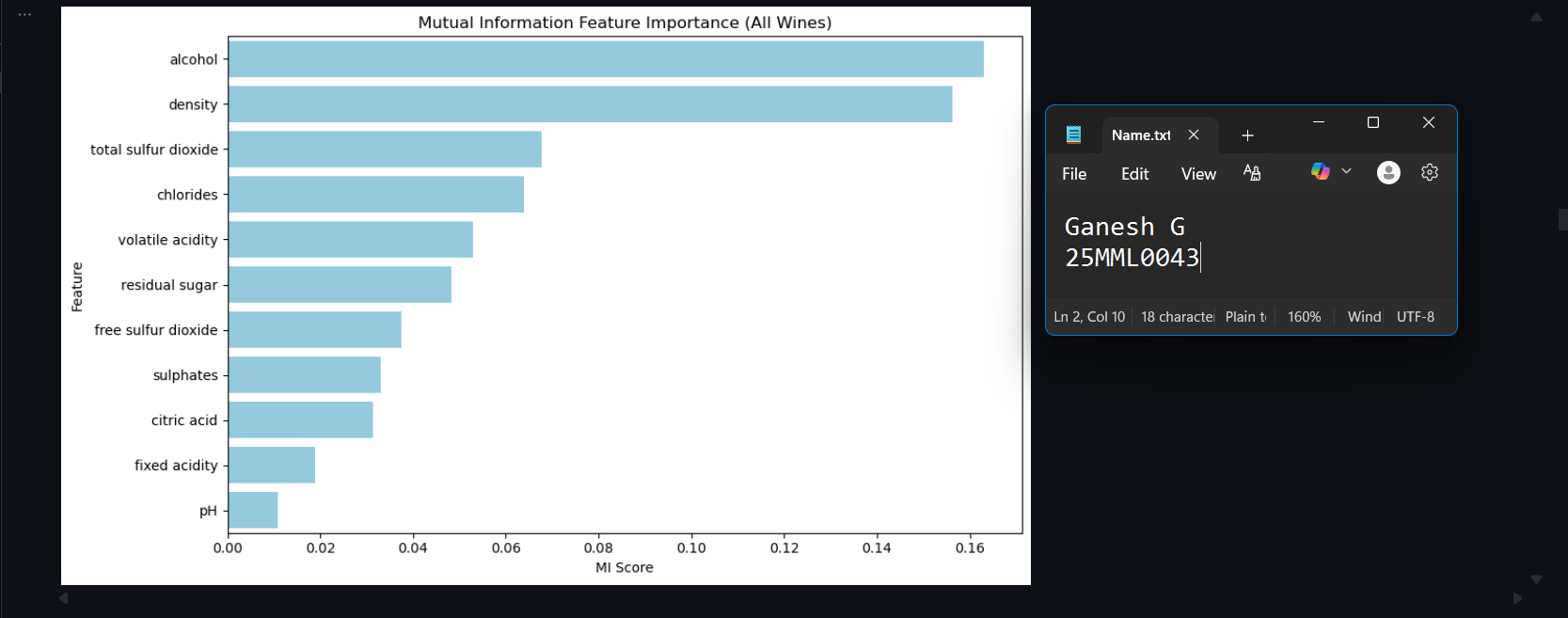
* 1. Comparing different Scalars

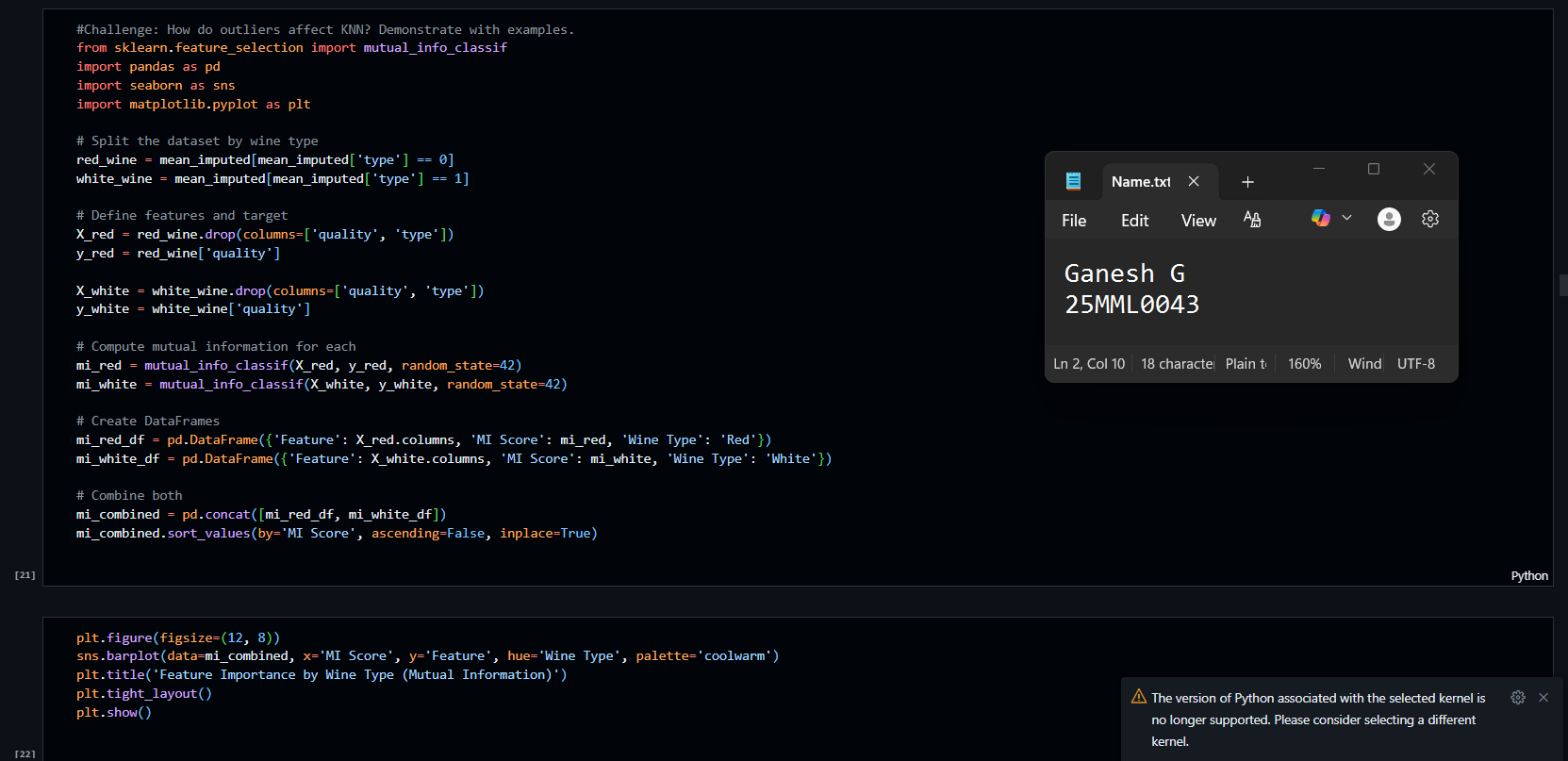


* 1. Analyze quality Distribution

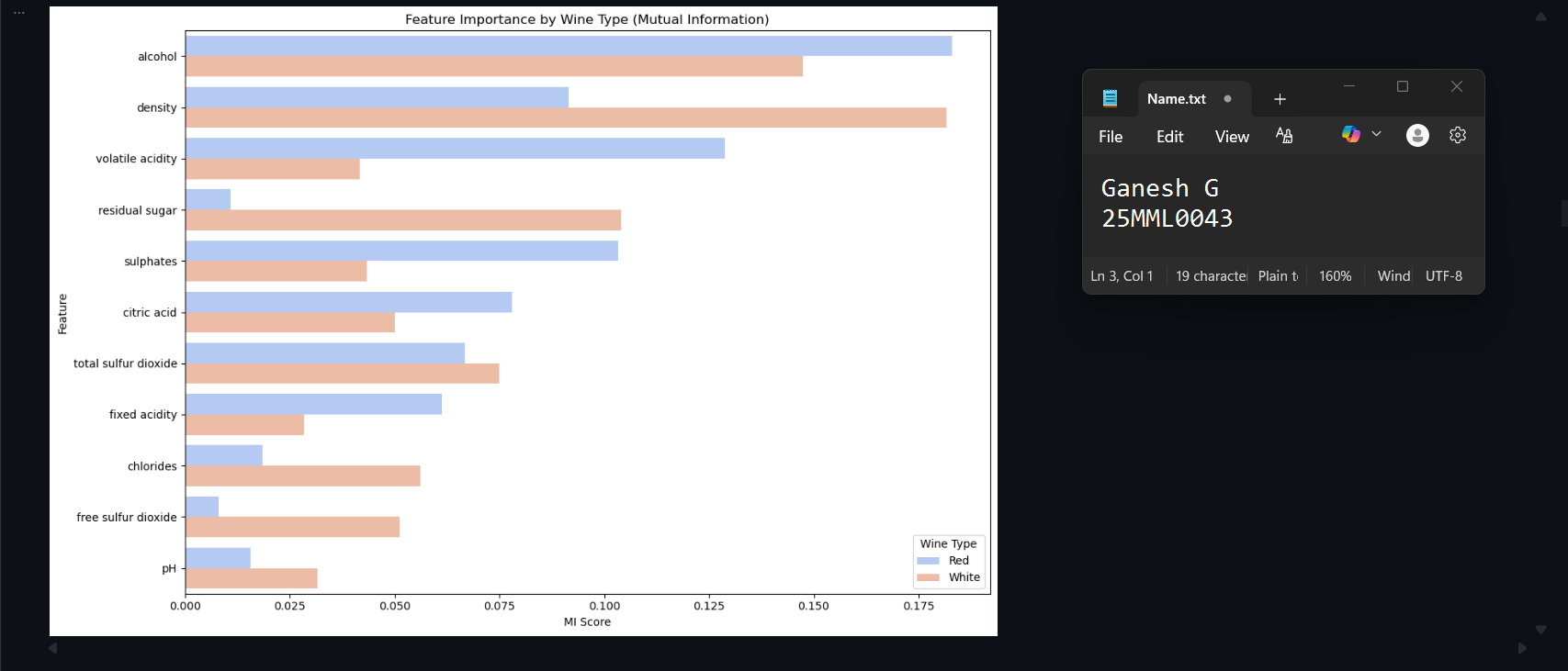




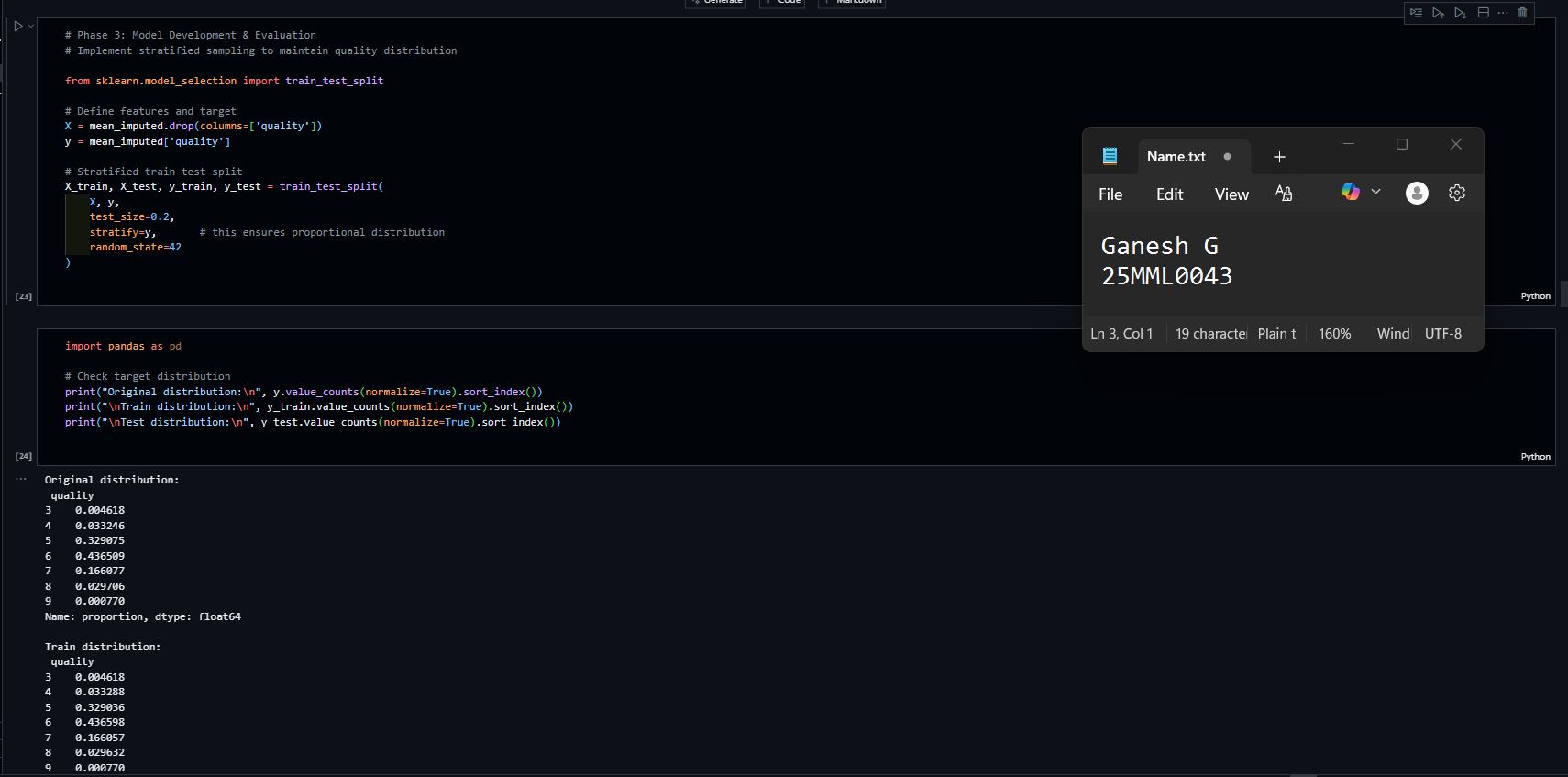


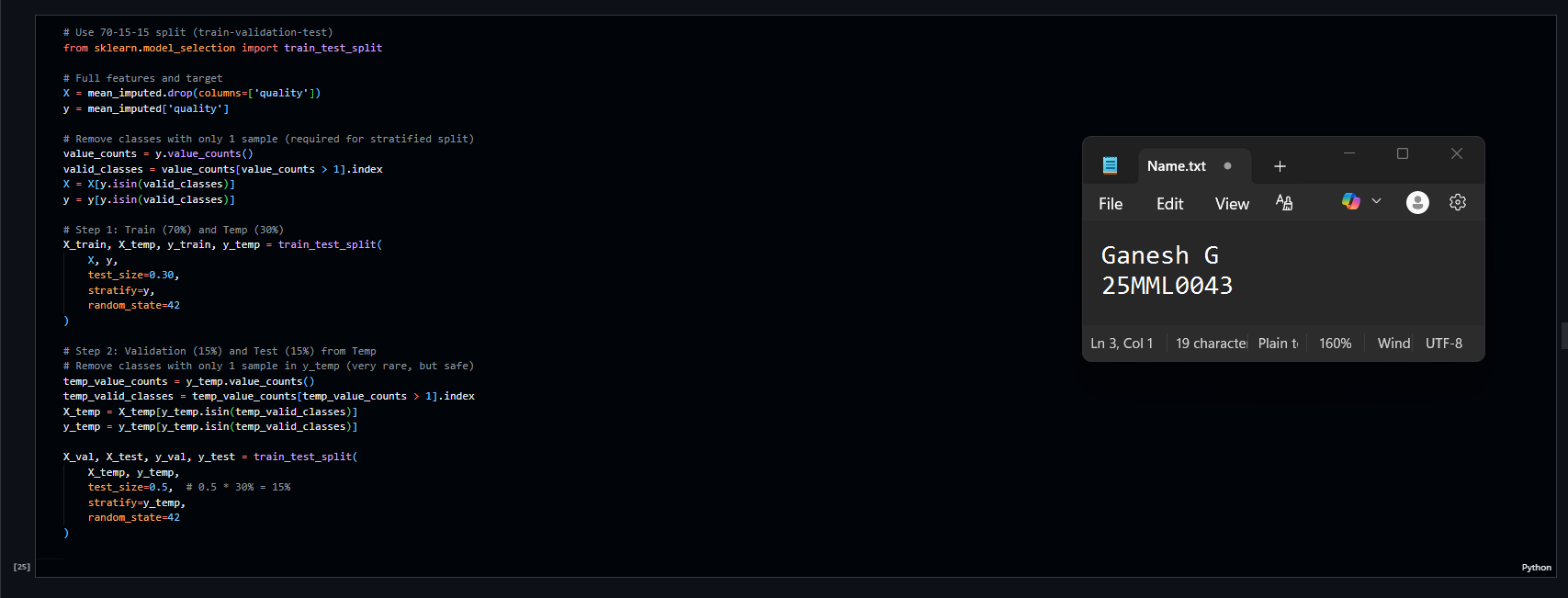


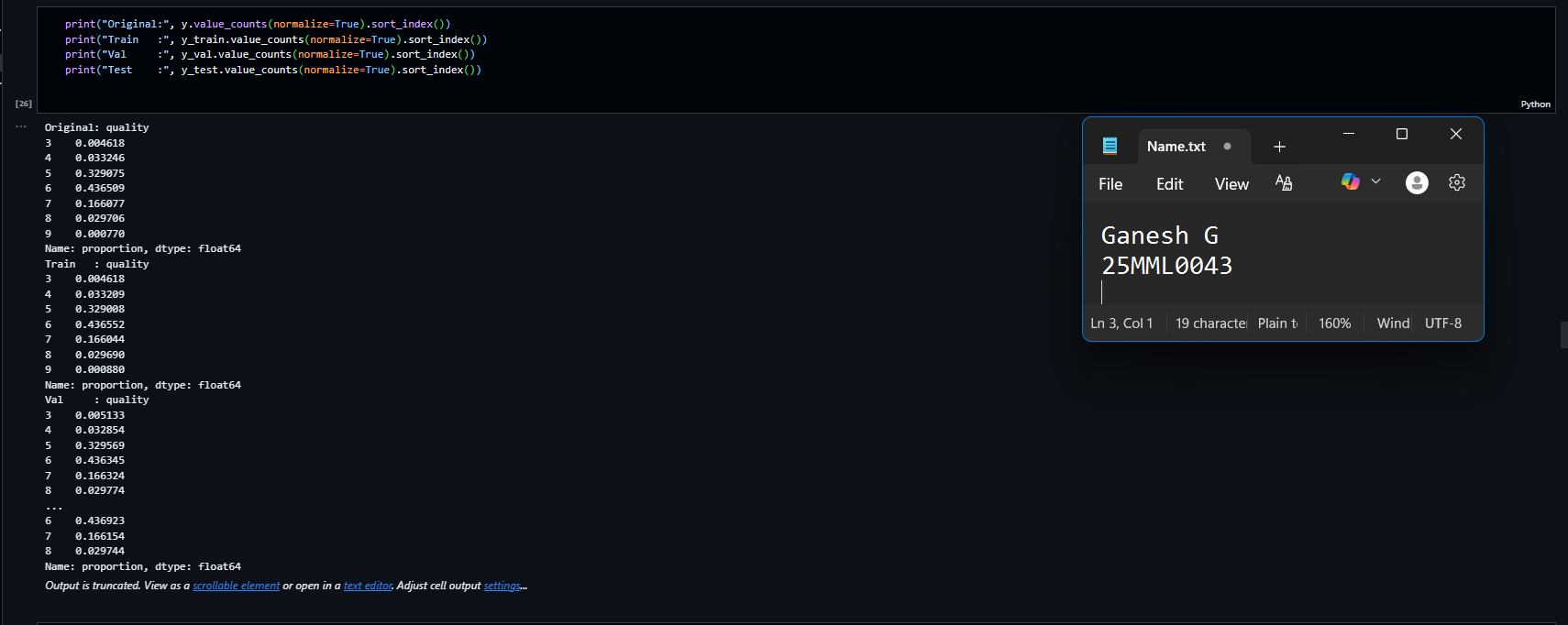




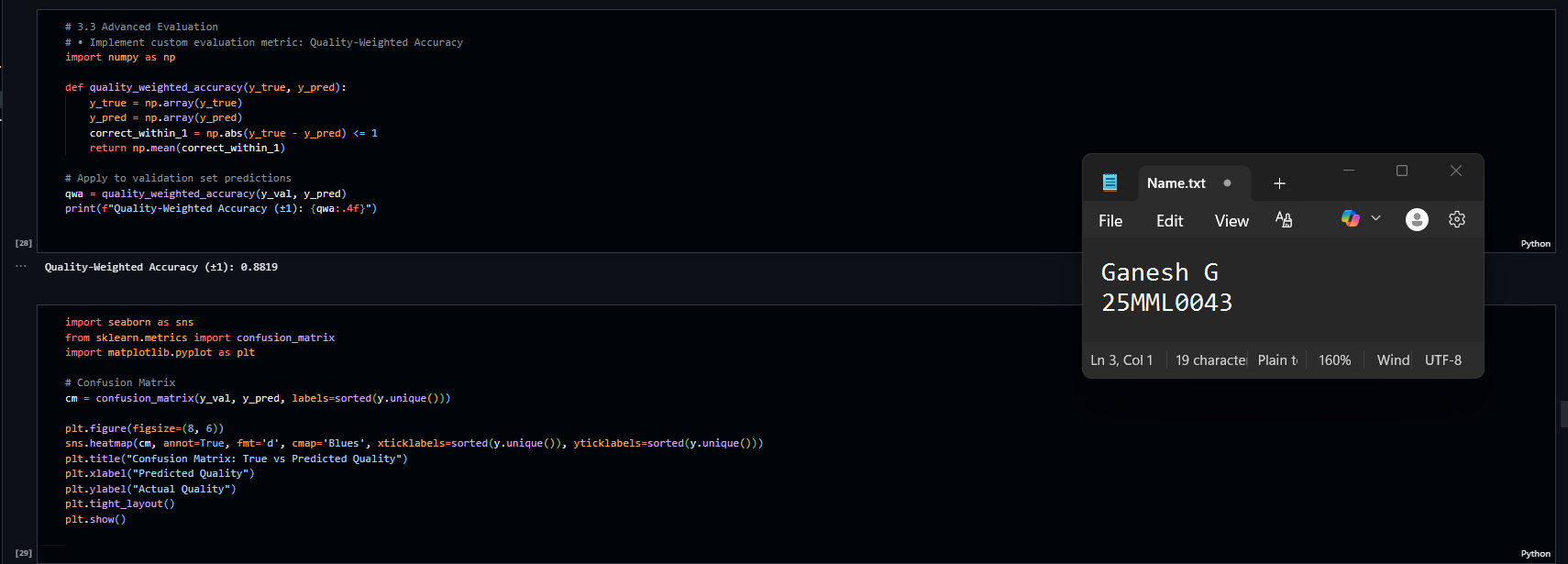
* 1. Model Development & Evaluation







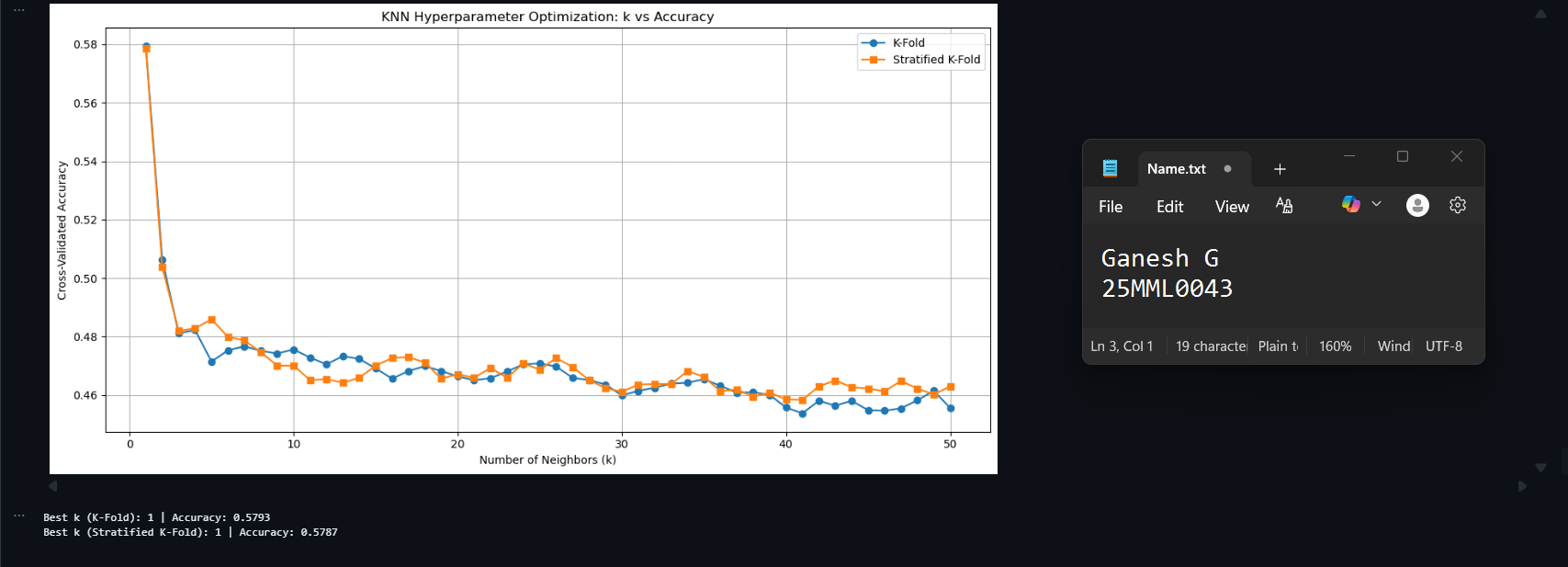


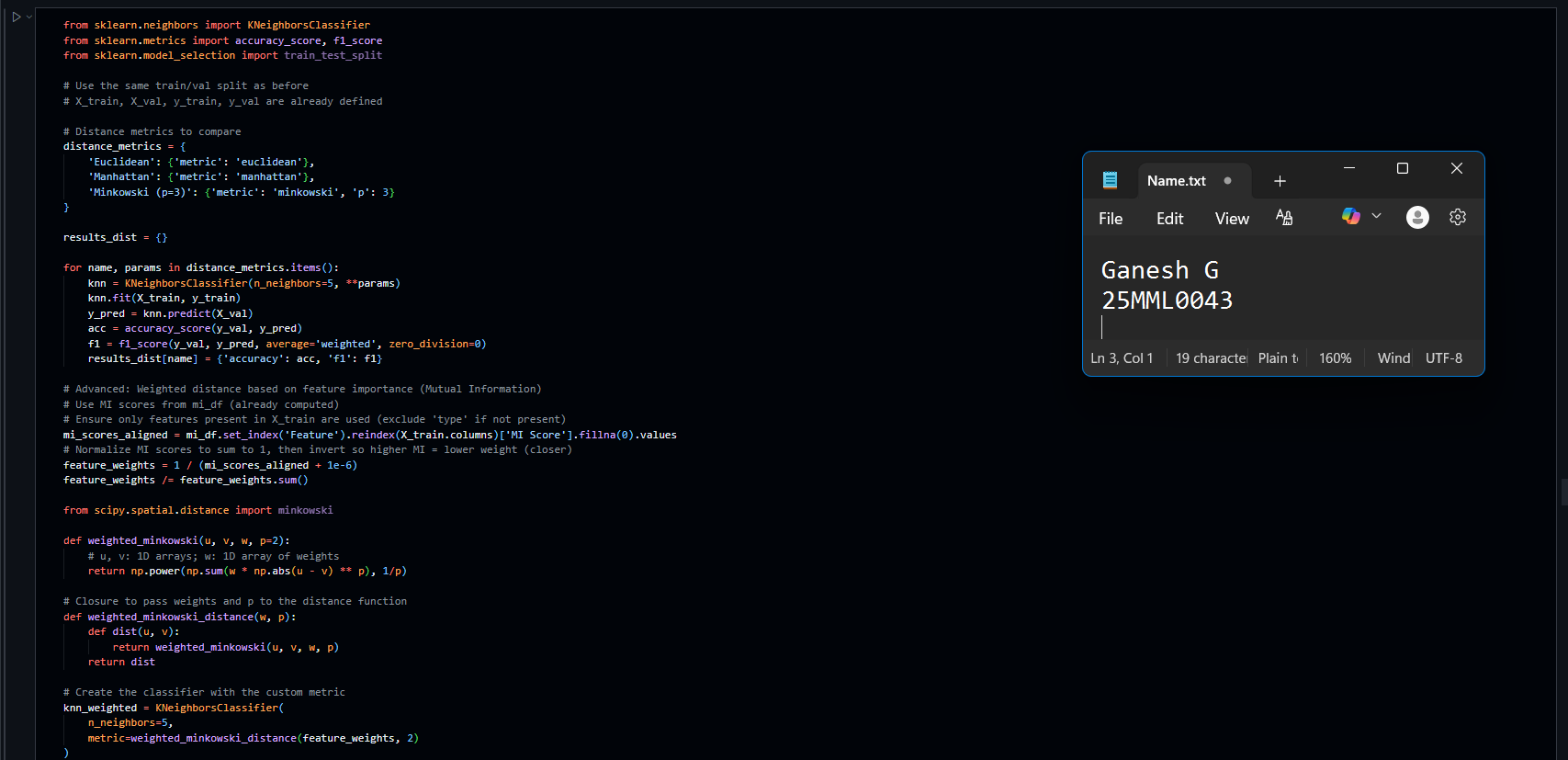




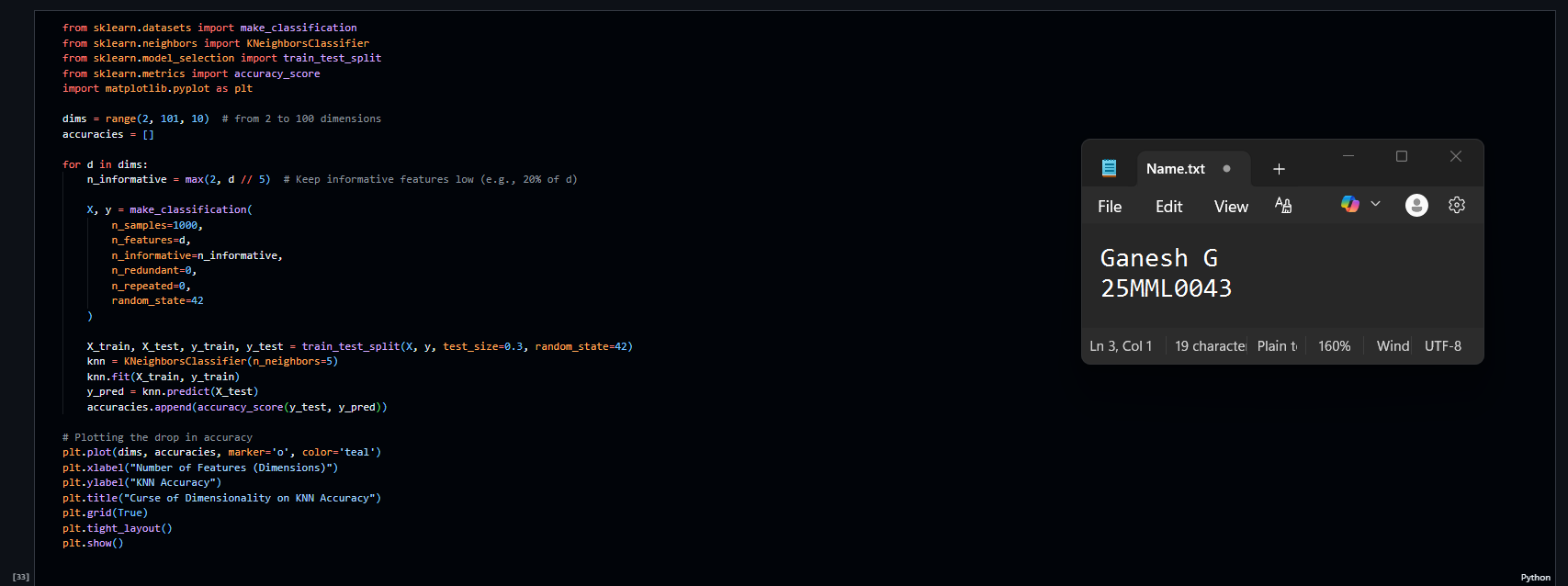
* 1. Hyper Parameter Optimization

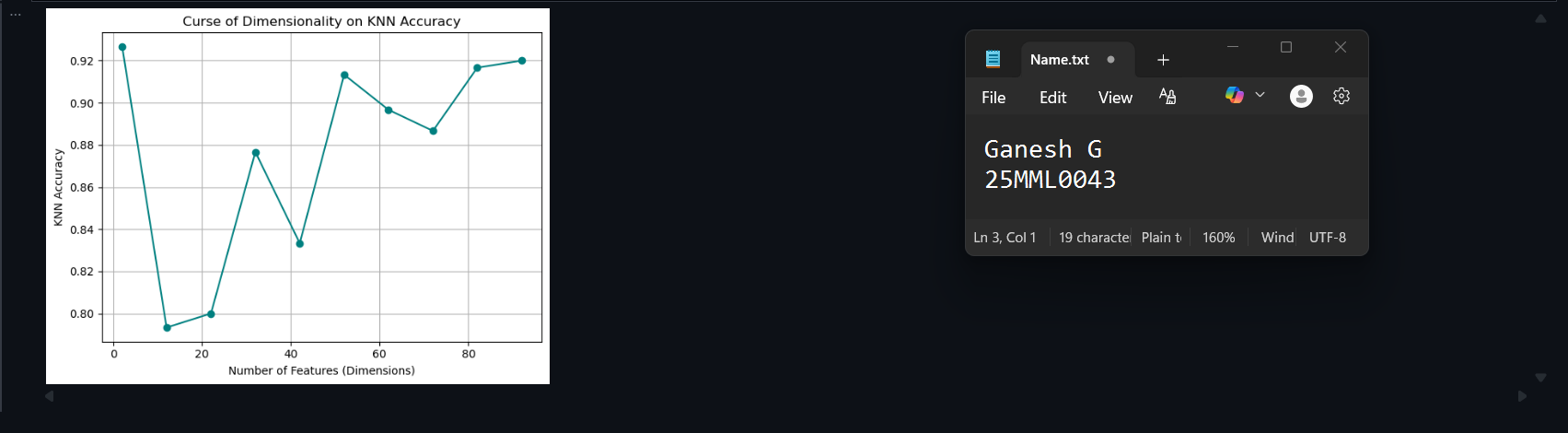


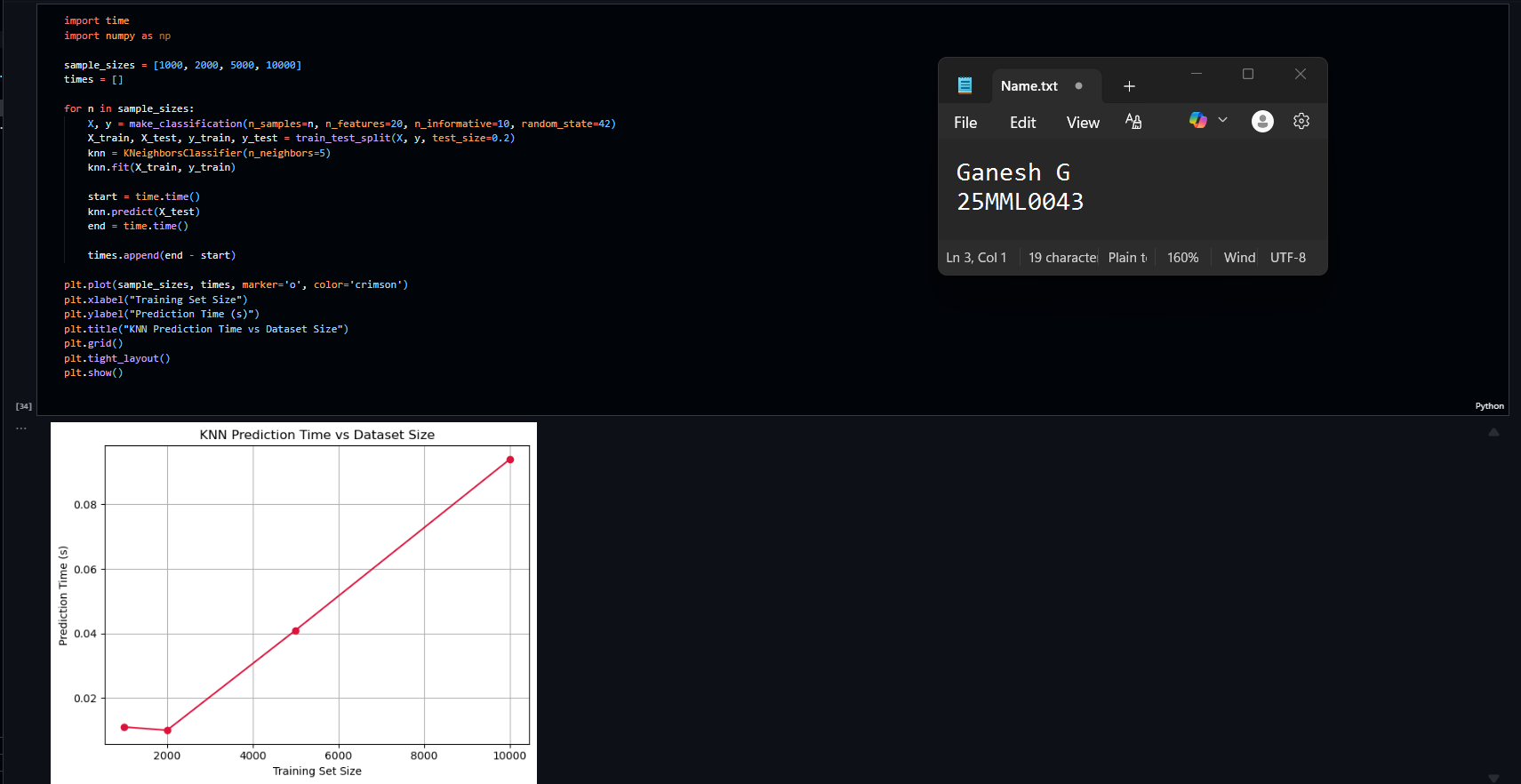




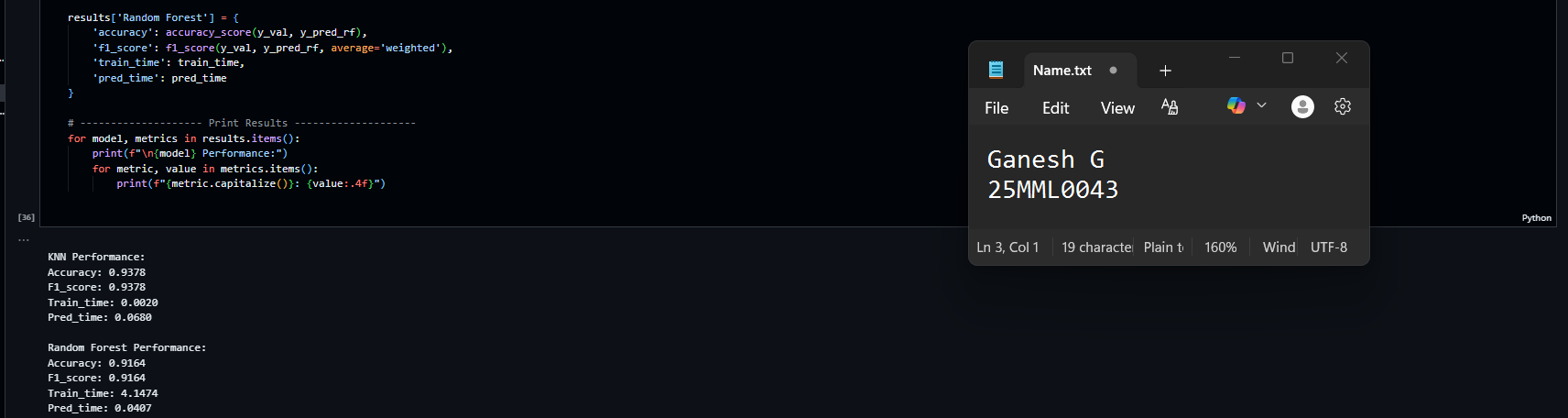


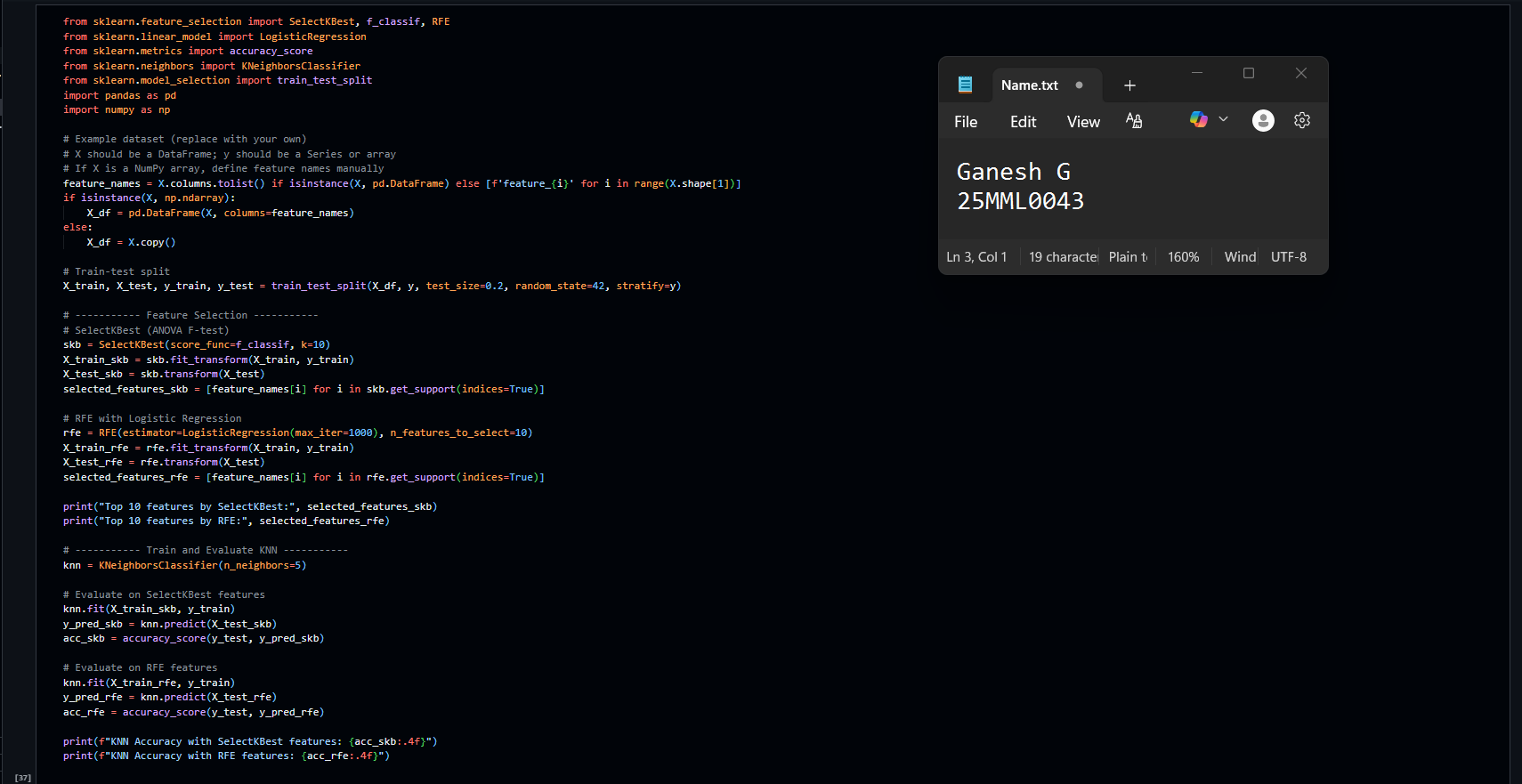


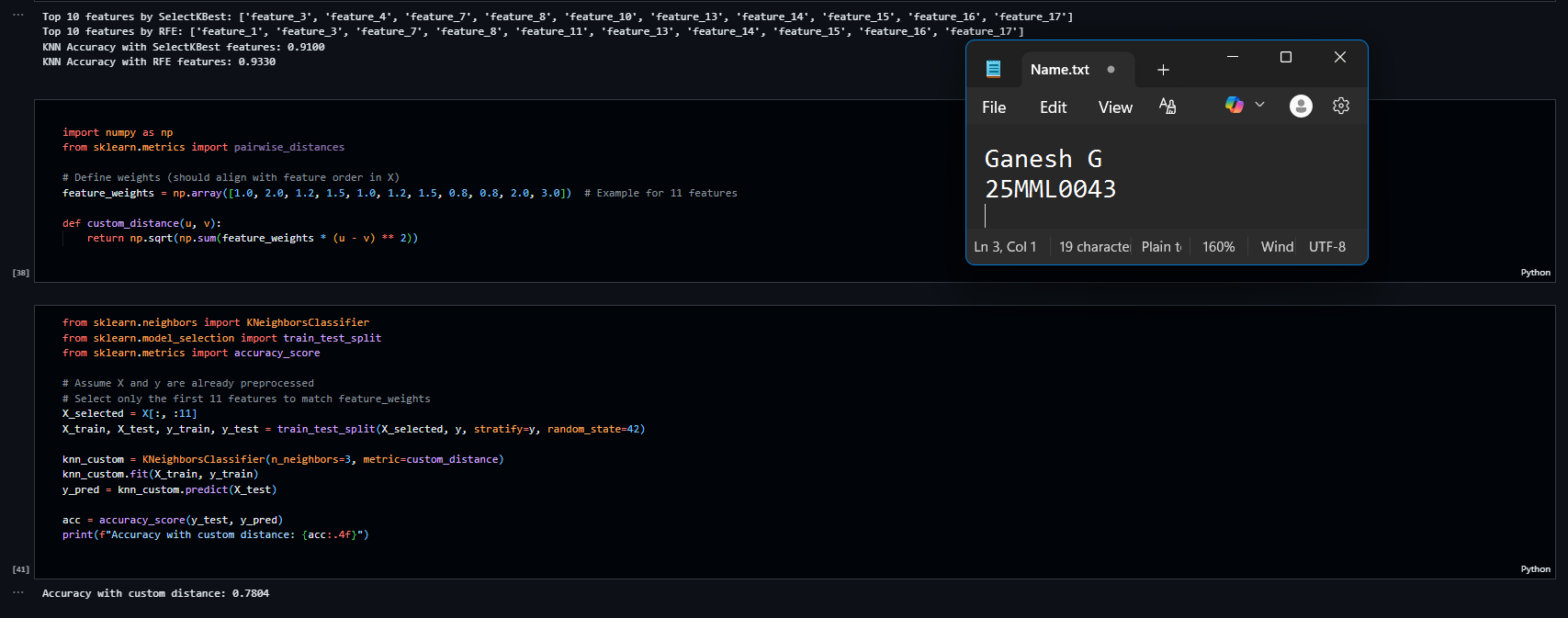


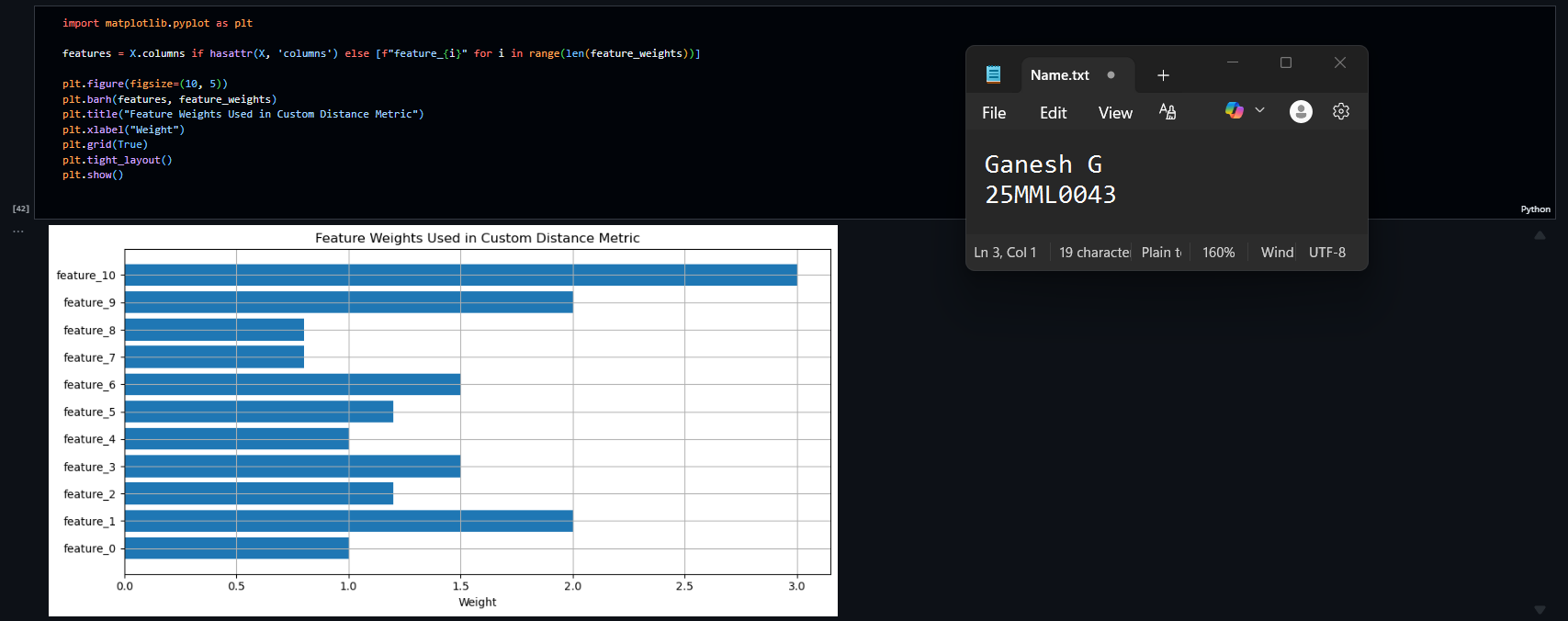


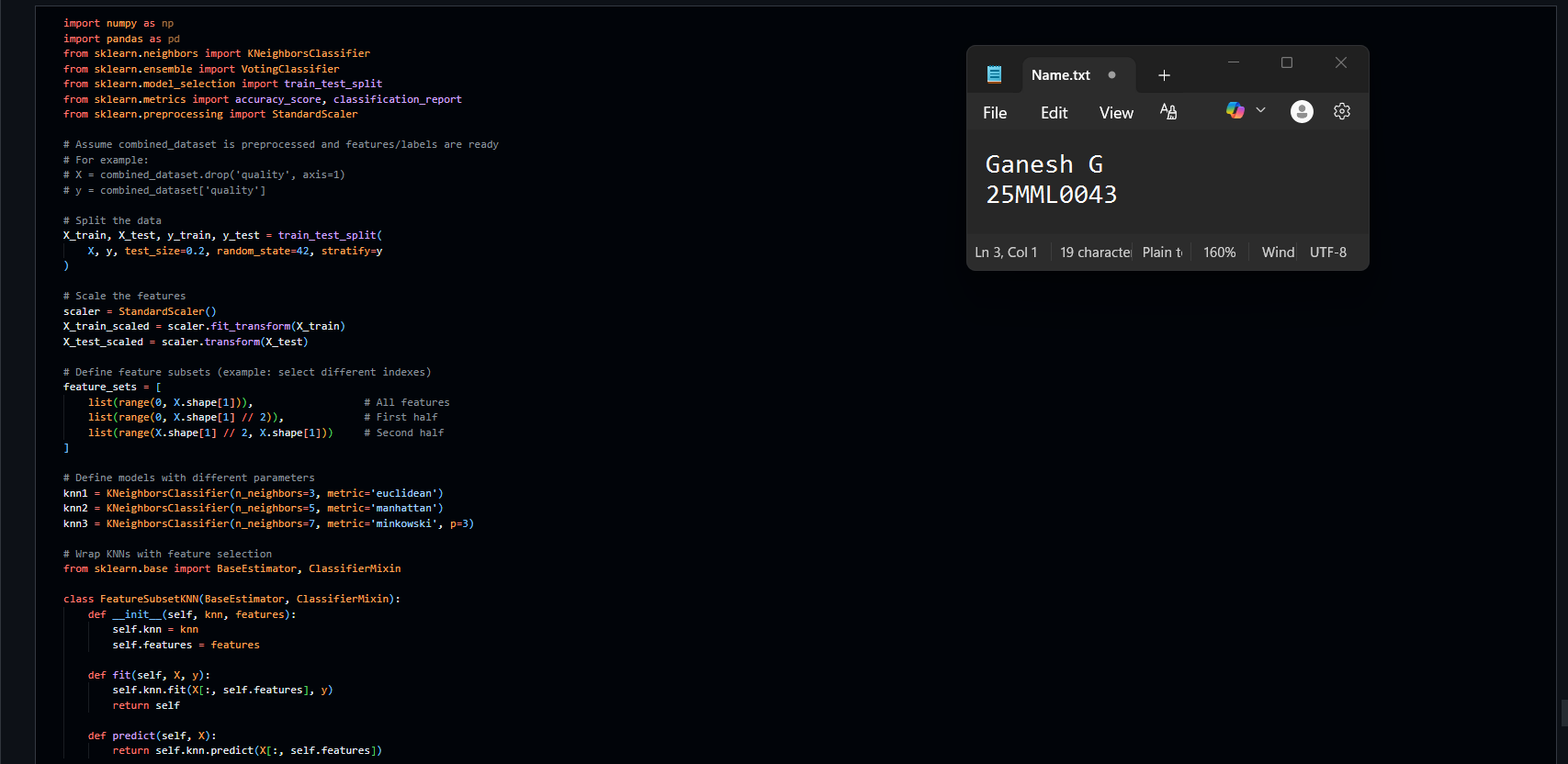


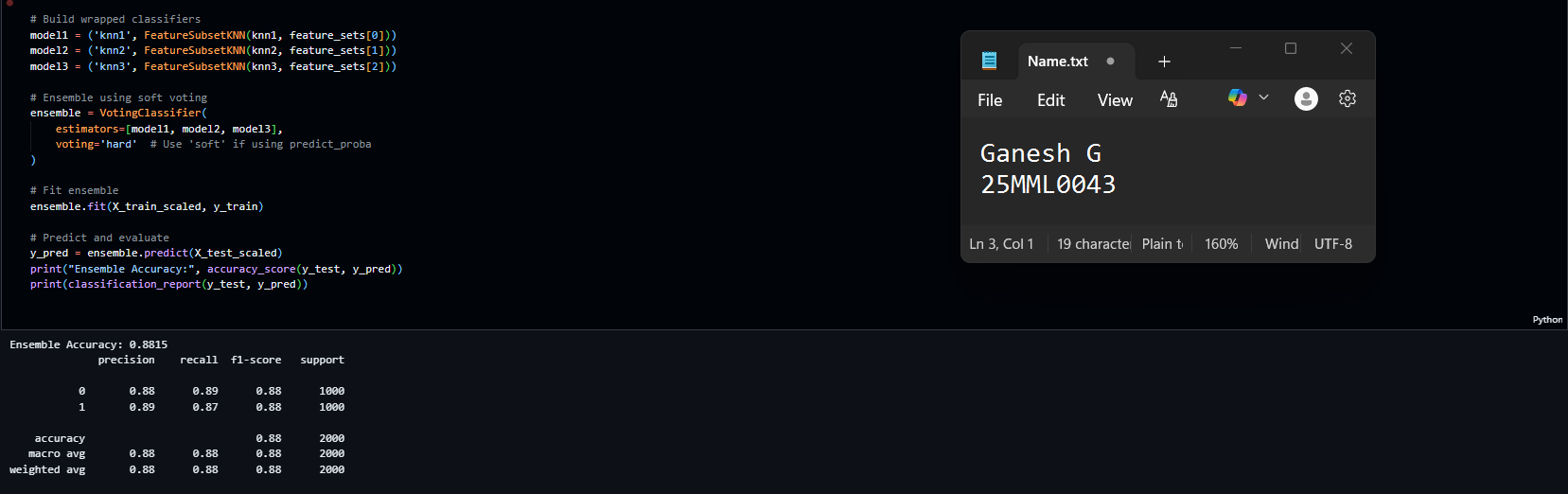


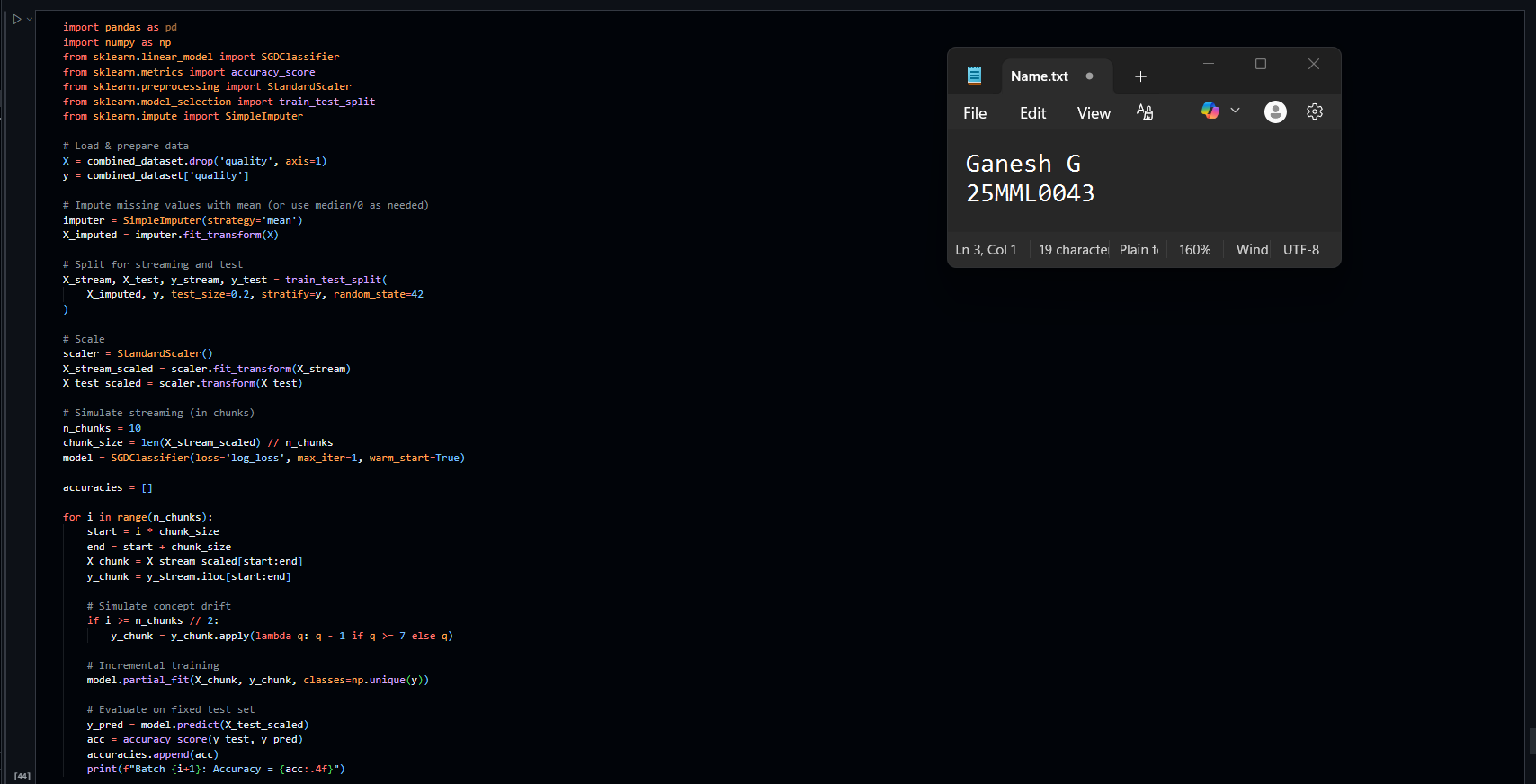


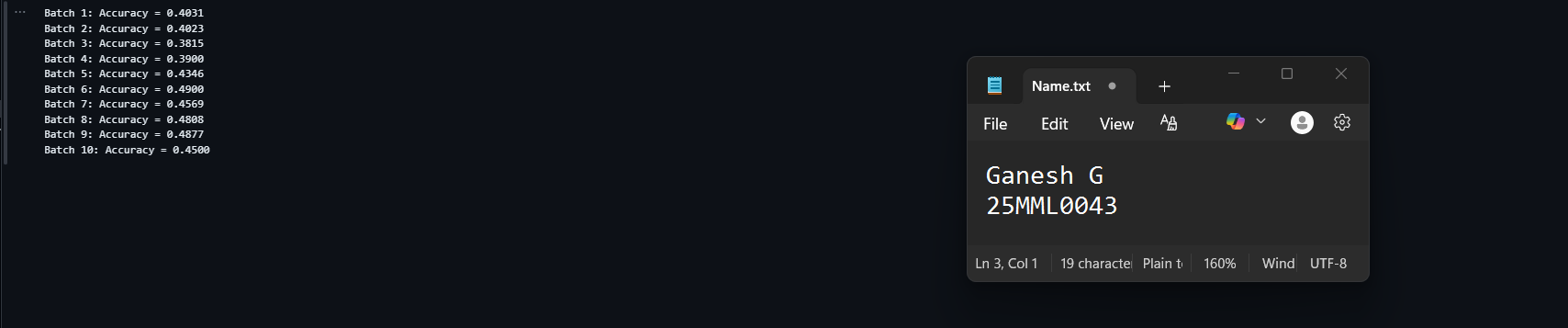












**SVM - Support Vector Machine.**

**AIM:**

To implement Support Vector Machine (SVM) classifiers with different kernels for various synthetic datasets (linear, moon-shaped, and circular), perform feature scaling, tune hyperparameters using GridSearchCV, and evaluate the models using accuracy, precision, recall, F1-score, confusion matrices, and decision boundary visualization.

**INTRODUCTION:**

Support Vector Machine is a supervised machine learning algorithm used for classification and regression tasks. It works by finding the optimal separating hyperplane that maximizes the margin between classes in the feature space. If the data is not linearly separable, SVM uses the kernel trick to project data into a higher-dimensional space where separation is possible.

**REAL WORLD APPLICATIONS:**

SVM is widely used in various domains, including:

1. **Medical Diagnosis** – In Medical diagnosis, it is used for detecting diseases such as cancer, diabetes, heart disease.
2. **Image Recognition** –It is widely used for Face detection and object classification.
3. **Text Categorization** – Spam filtering, sentiment analysis.
4. **Bioinformatics** – Protein classification, gene expression analysis.

**ALGORITHM:**

1.Import the necessary libraries.

2.Load the Dataset - Import the Breast Cancer Wisconsin dataset from sklearn.datasets.

3. Explore the Data - Check number of samples, features, and class distribution.

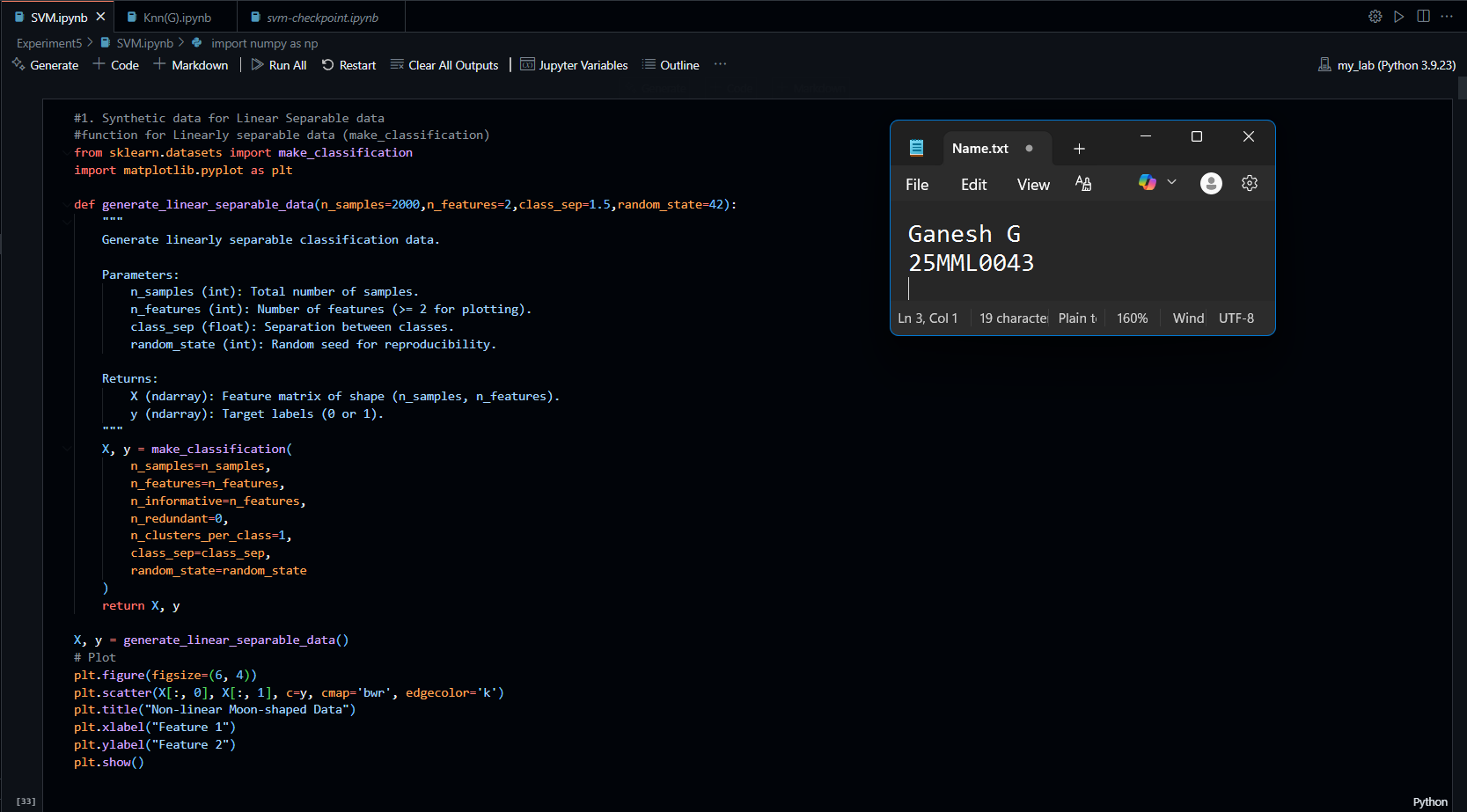
4.Split the Data and divide into training set (80%) and testing set (20%).

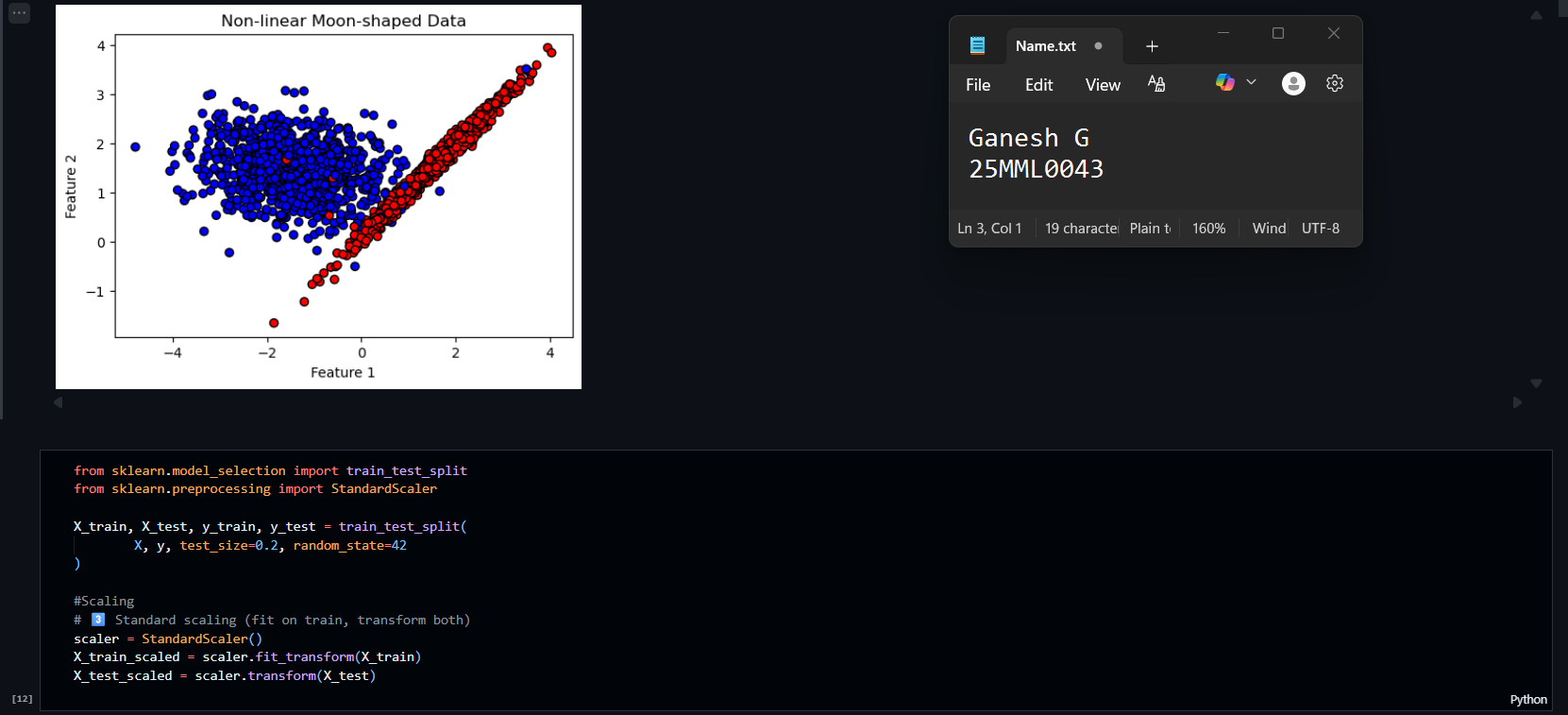
5. Feature Scaling - Use Standard Scaler to normalize the features for better SVM performance.

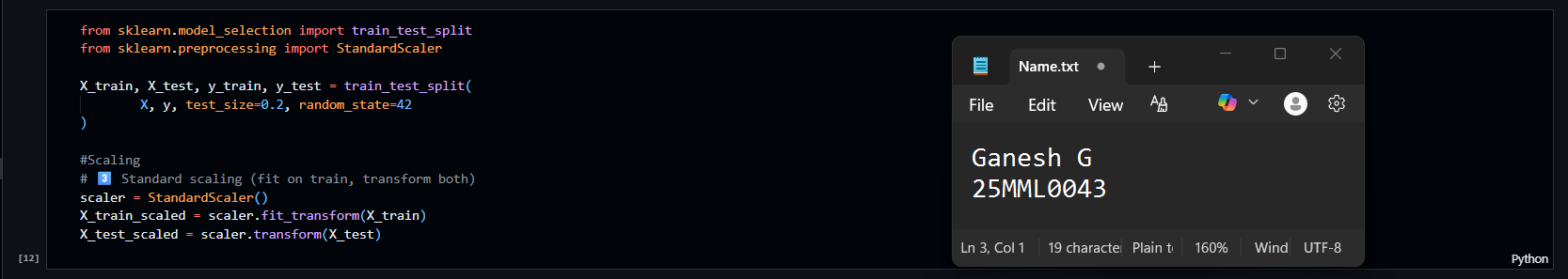
1. Choose Kernel - Use linear, poly, rbf, and sigmoid kernels for experimentation.
2. Train the SVM Model - For each kernel, fit the SVM on training data.
3. Make Predictions - Use the trained model to predict on test data.
4. Evaluate the Model - Calculate confusion matrix, accuracy, precision, recall, and F1 score.
5. Visualize Decision Boundaries - Reduce features to 2D using PCA and plot boundaries for each kernel.
6. Hyperparameter Tuning - Use GridSearchCV to find best C and gamma values for the RBF kernel.

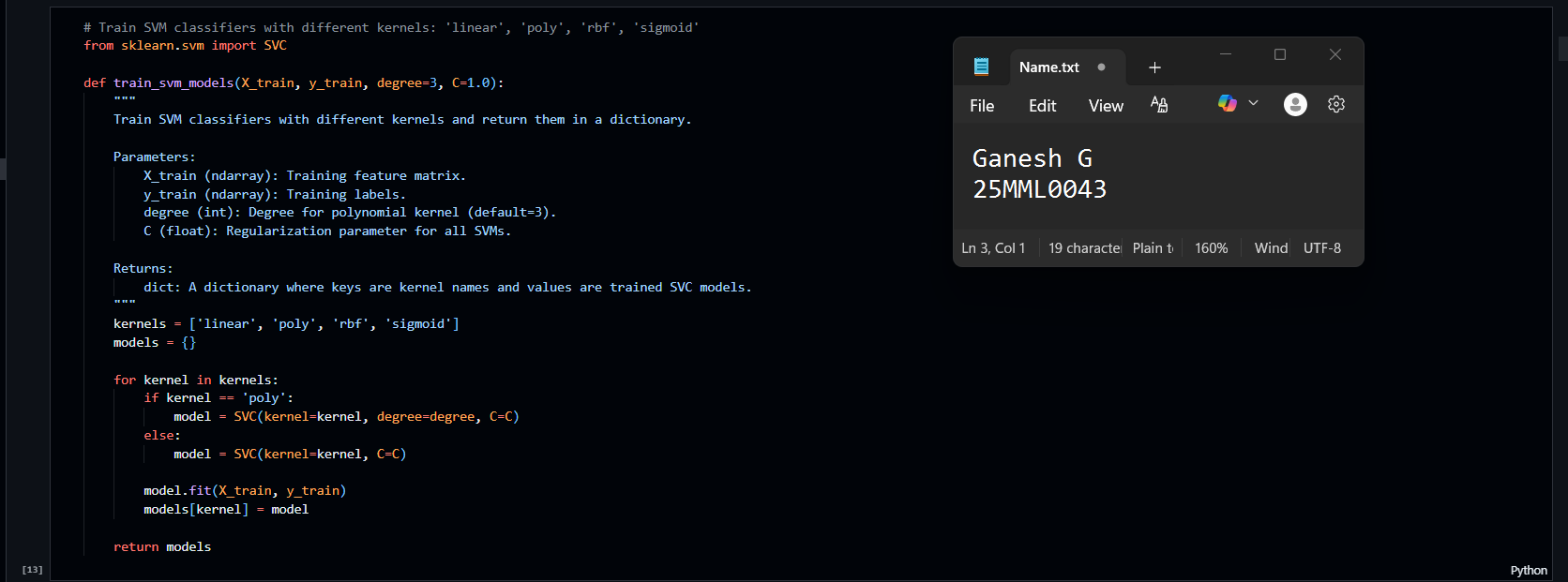
**IMPLEMENTATION AND RESULTS:**

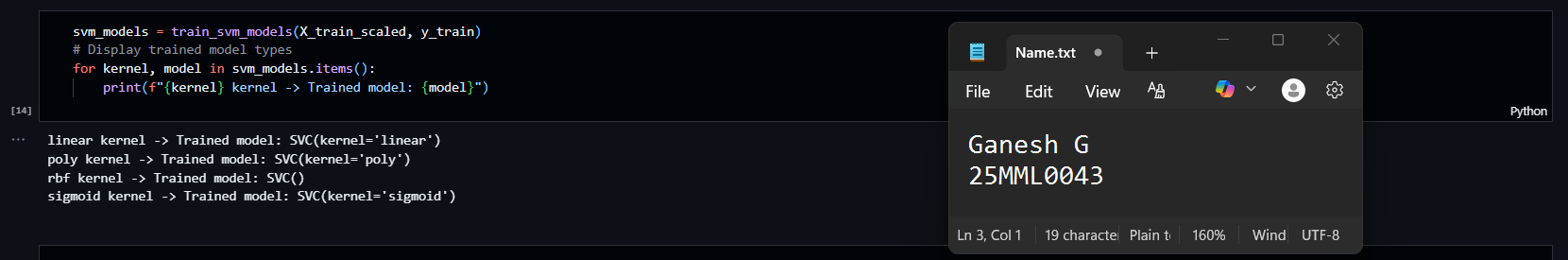
* 1. Linear Separable data

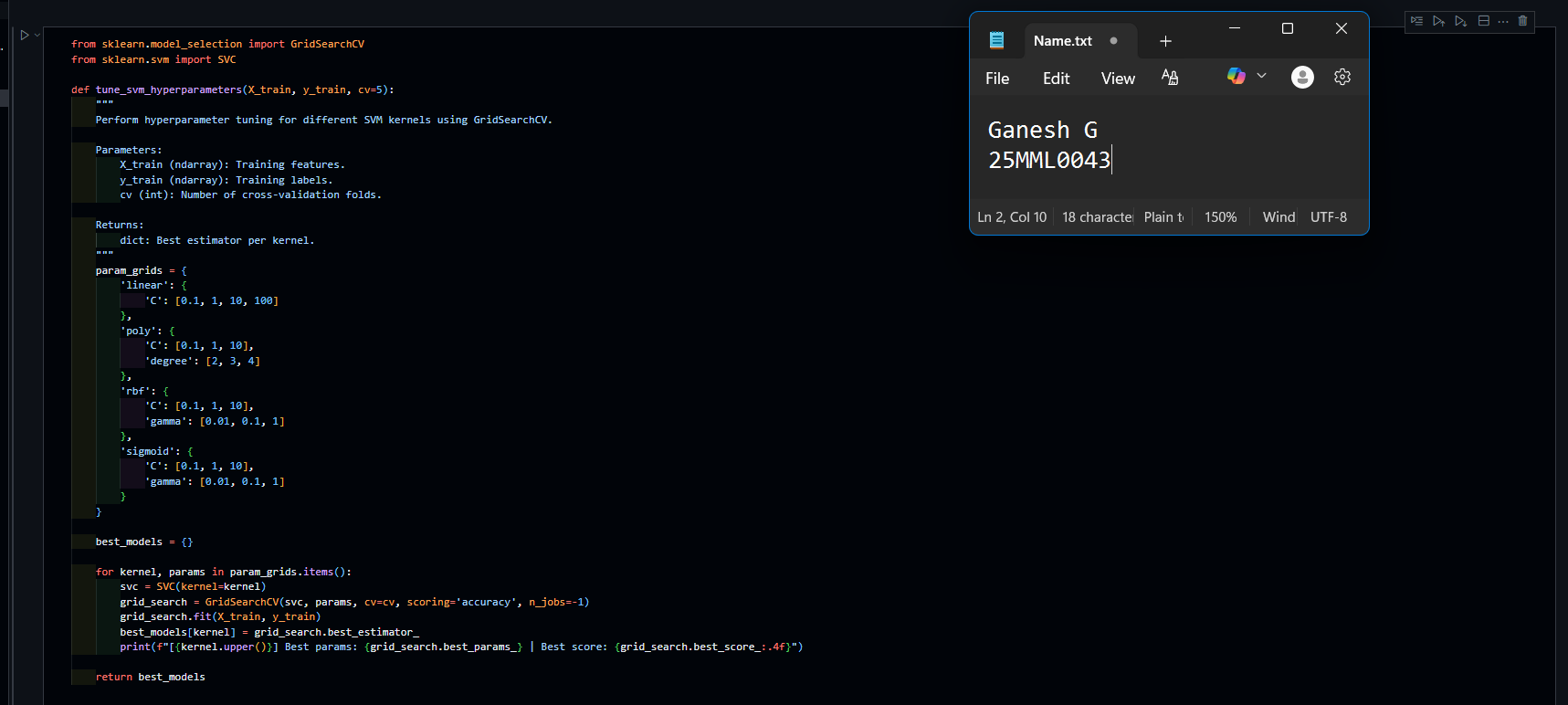


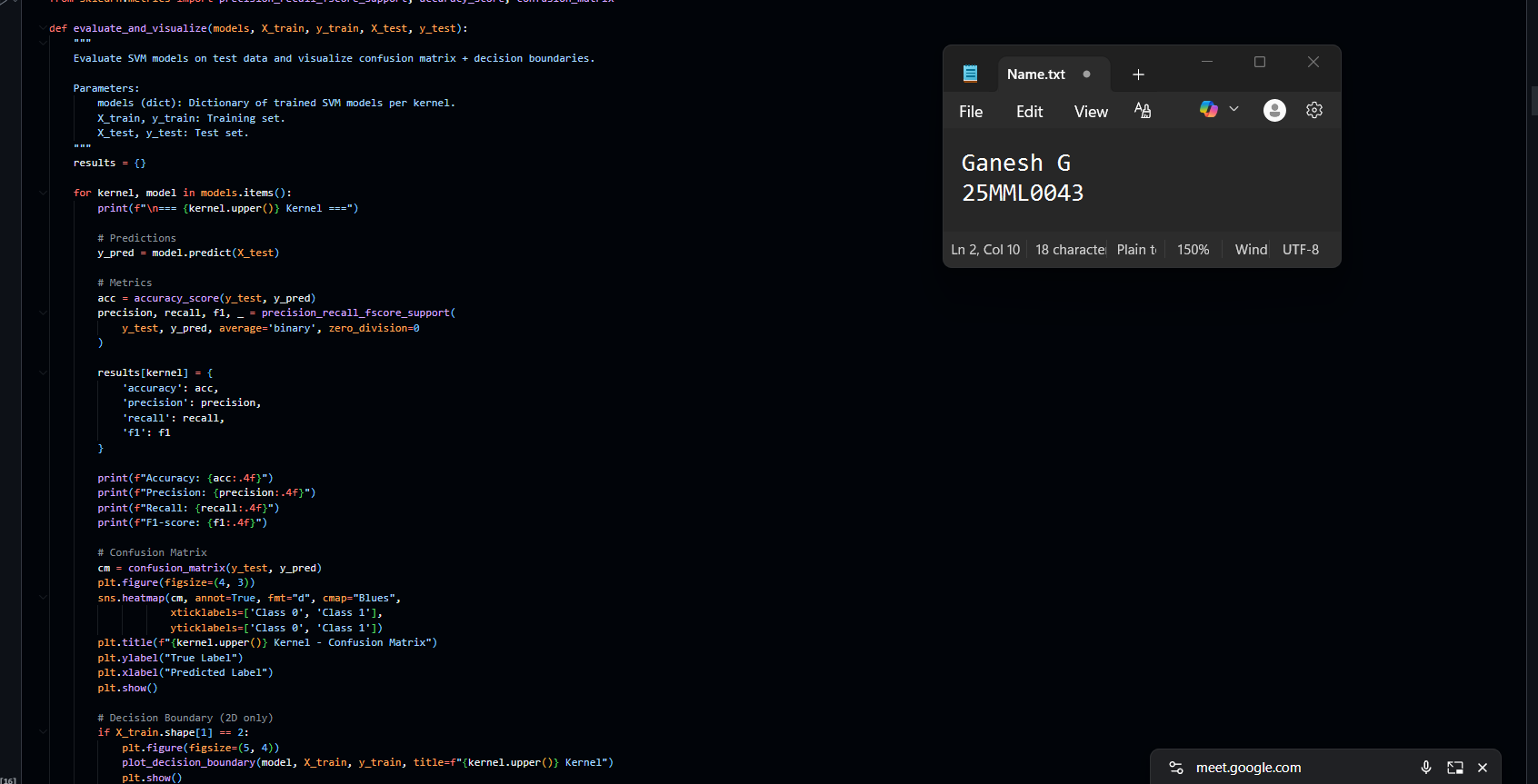




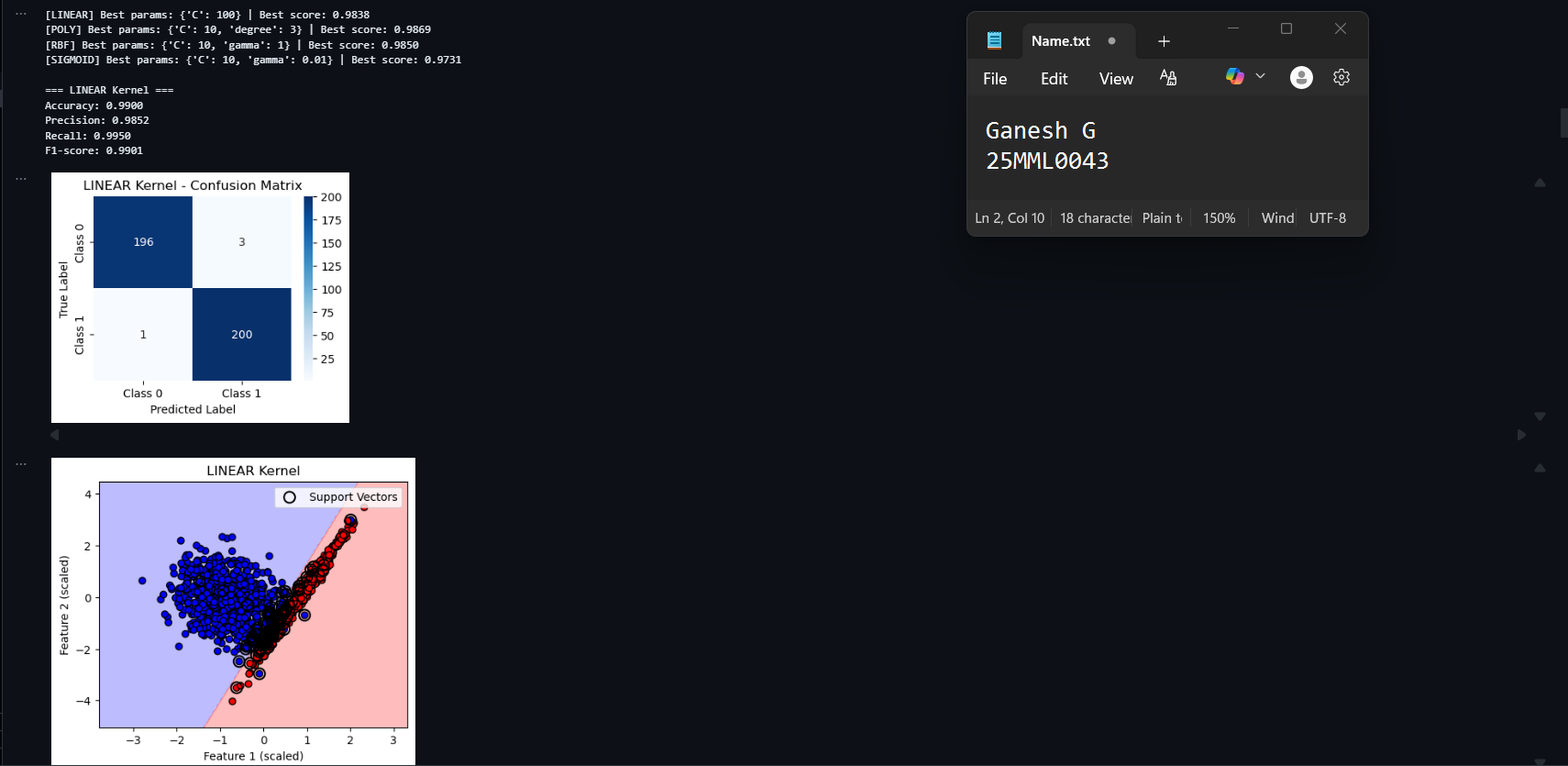


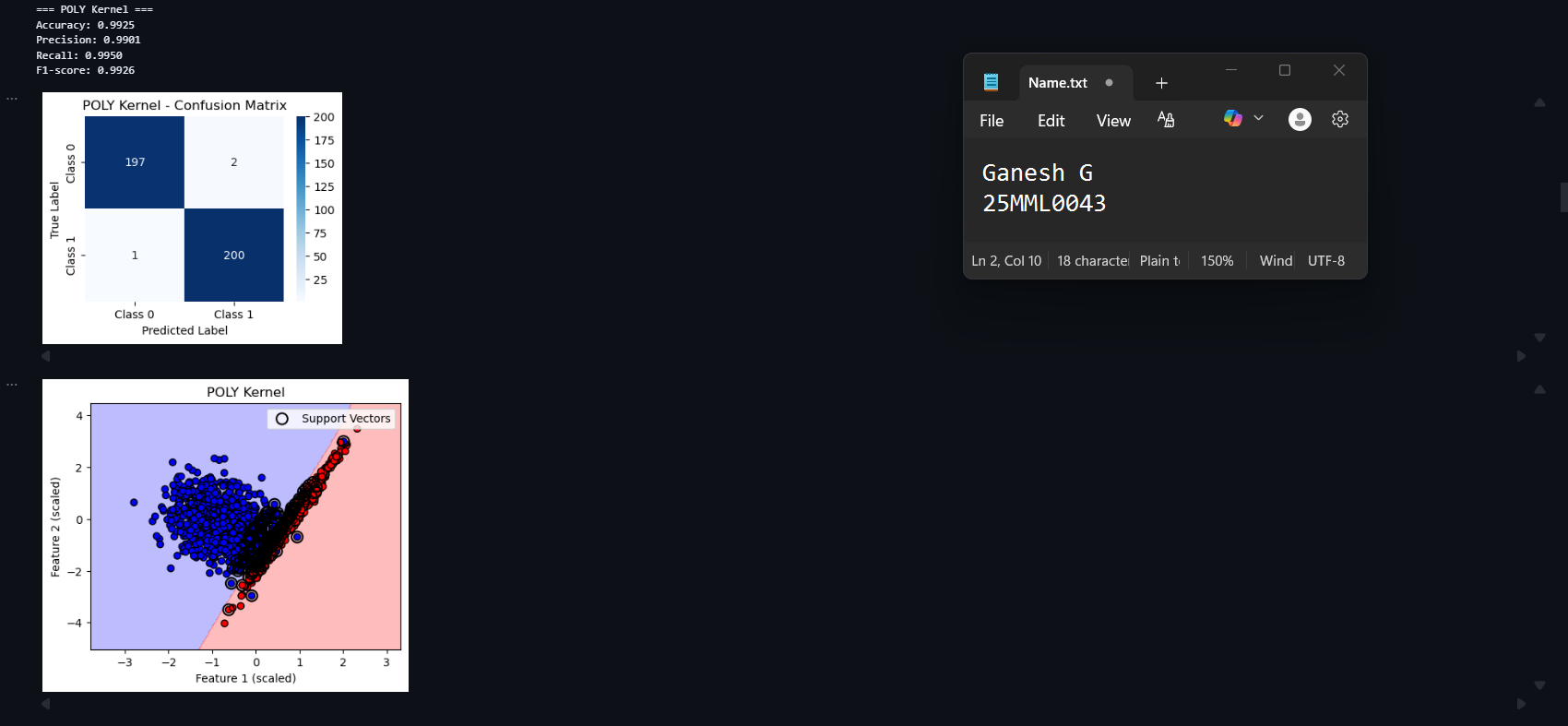


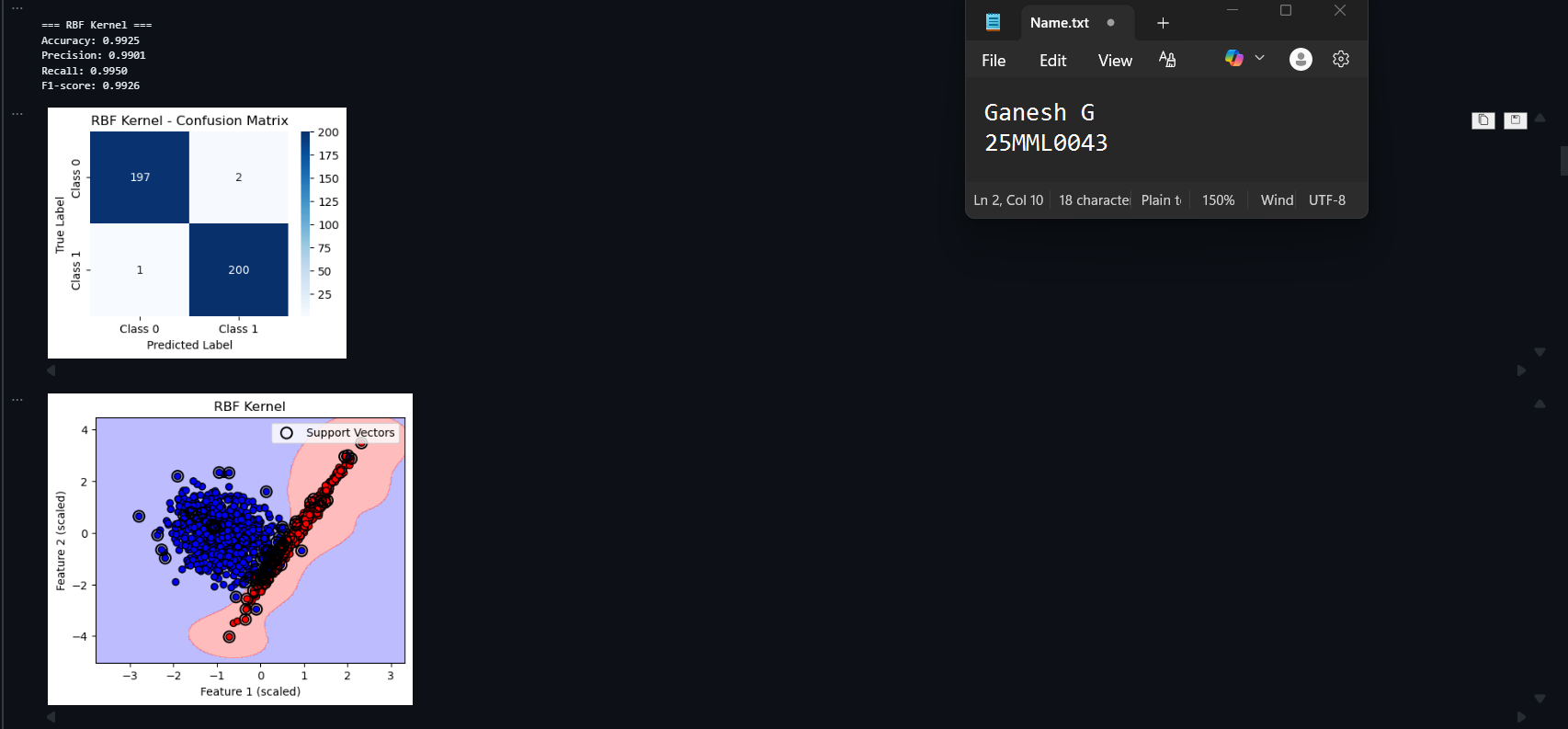


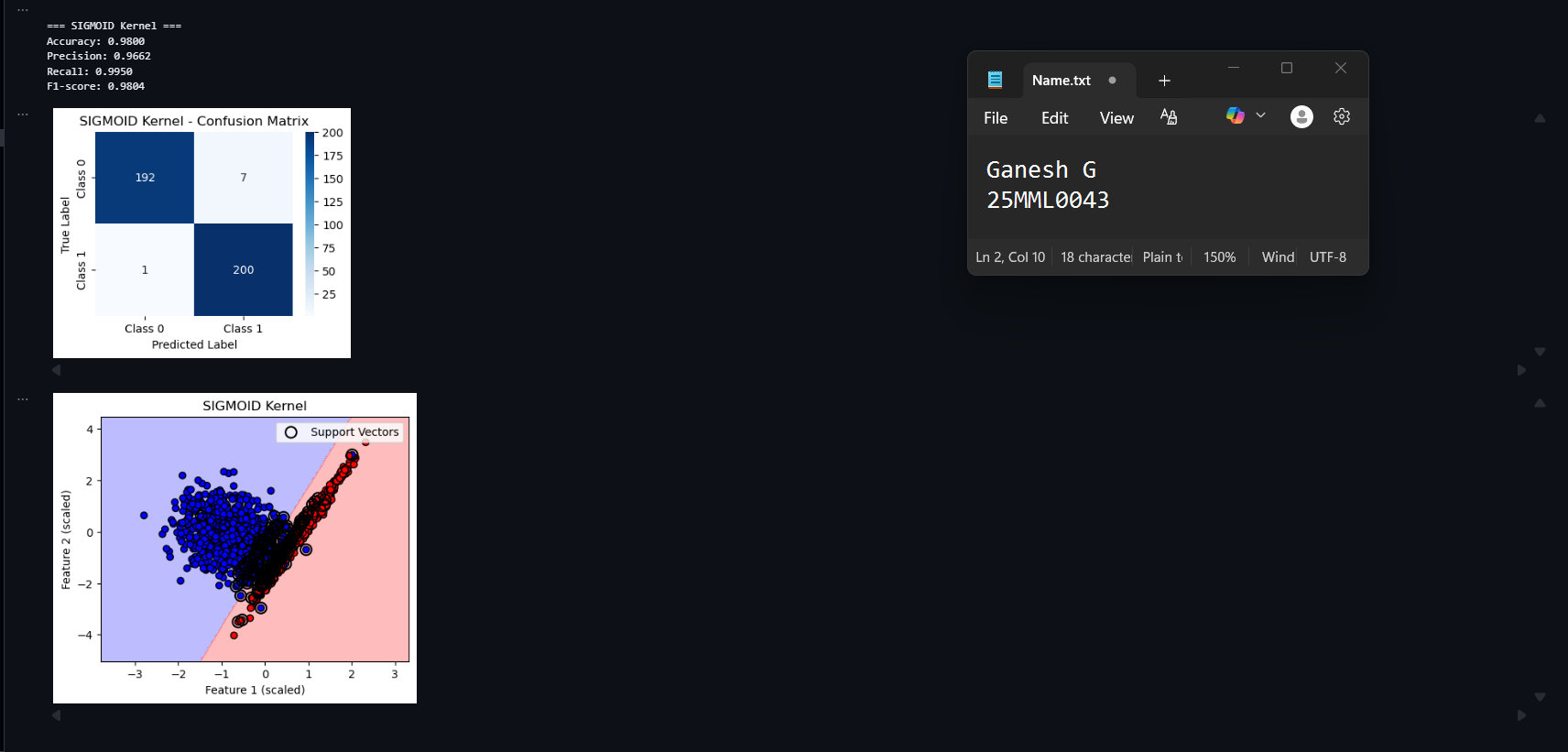




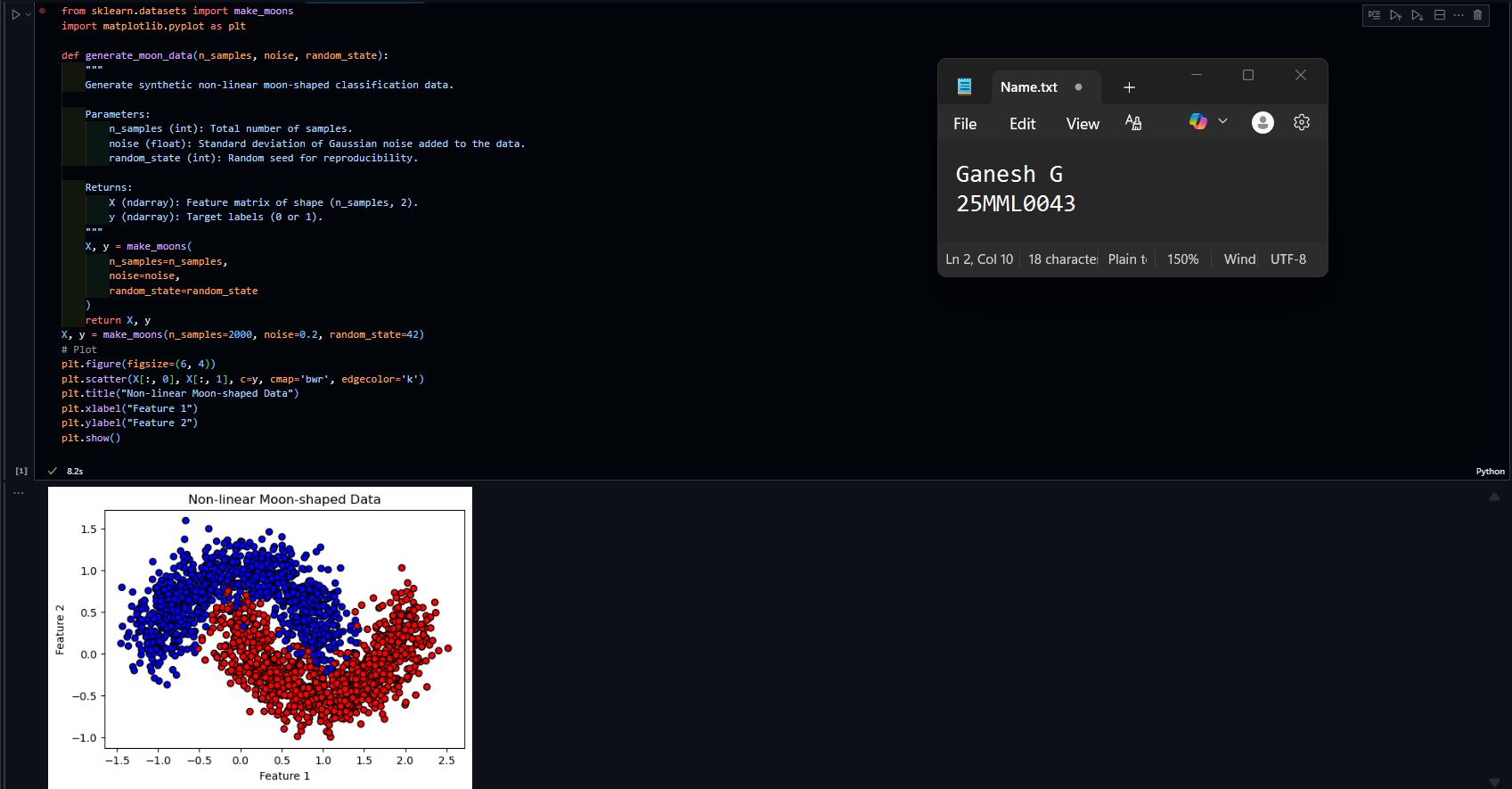


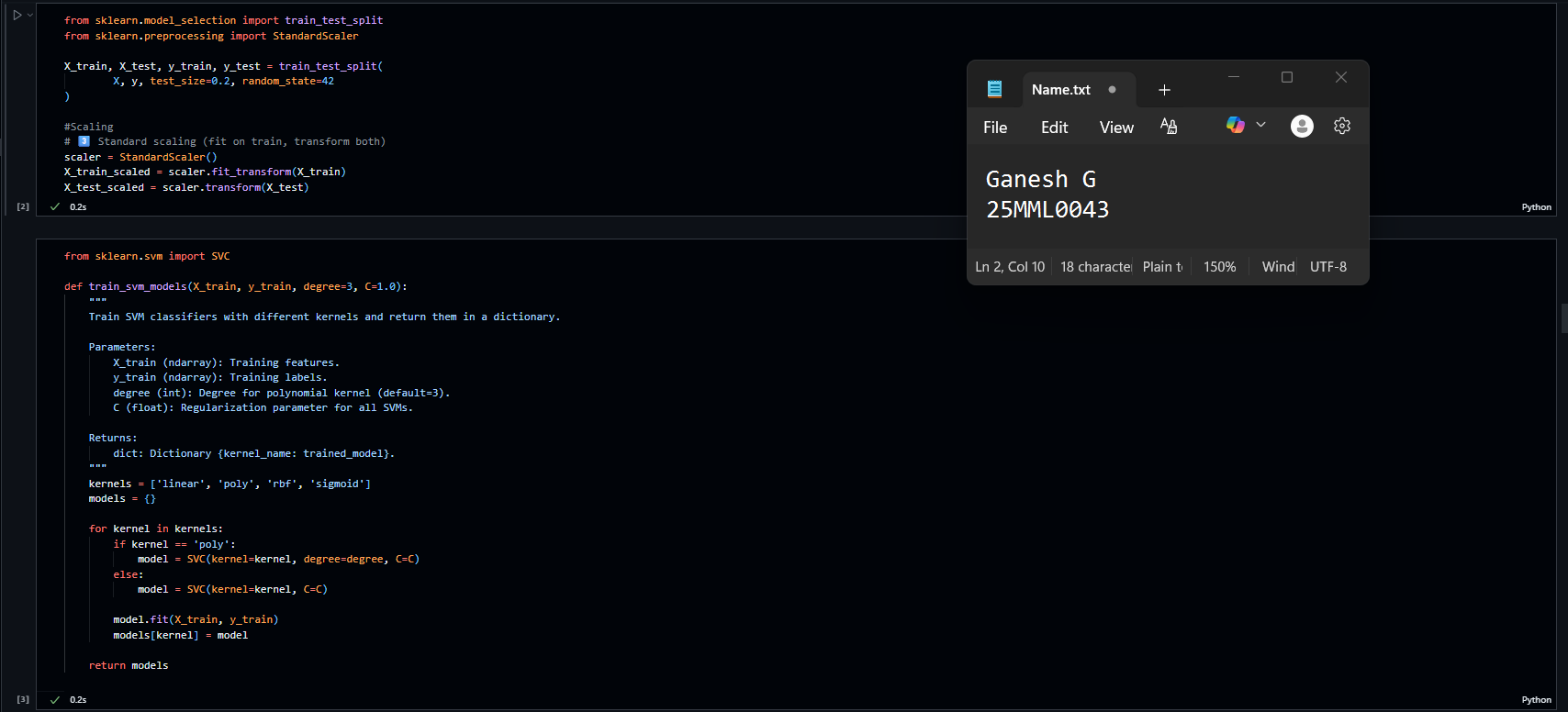


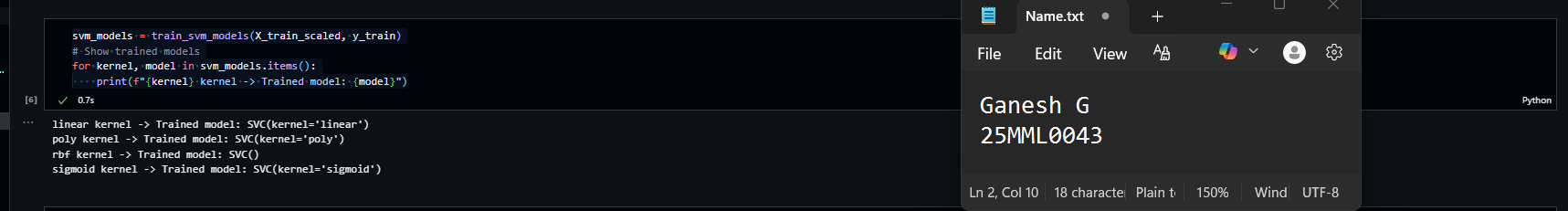


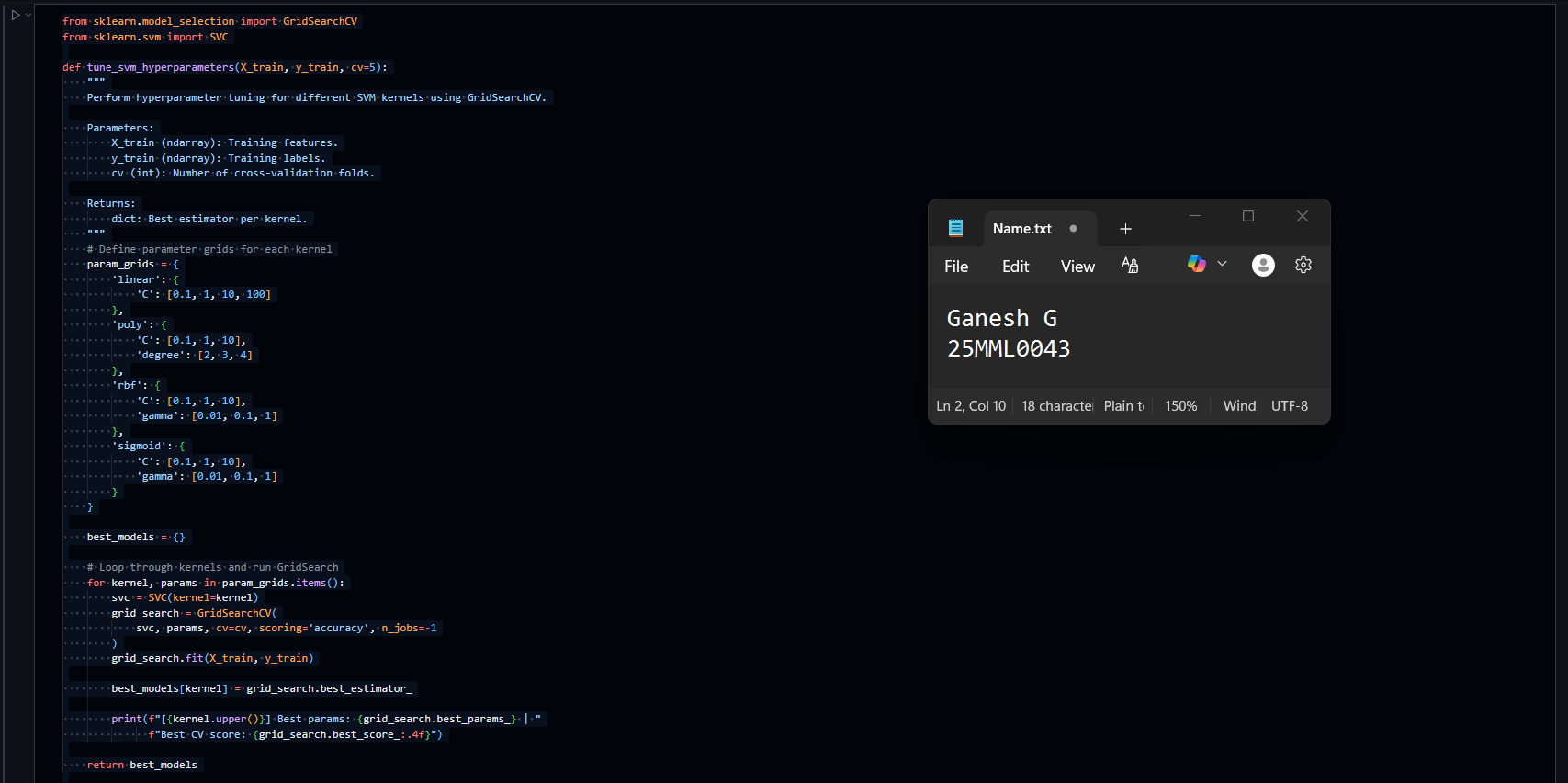


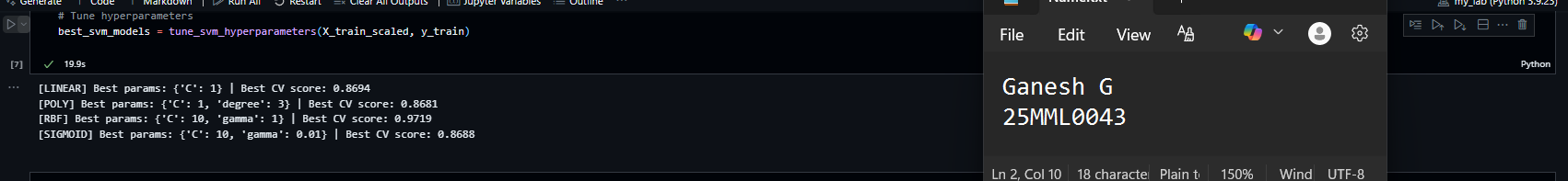
* 1. Nonlinear moon data

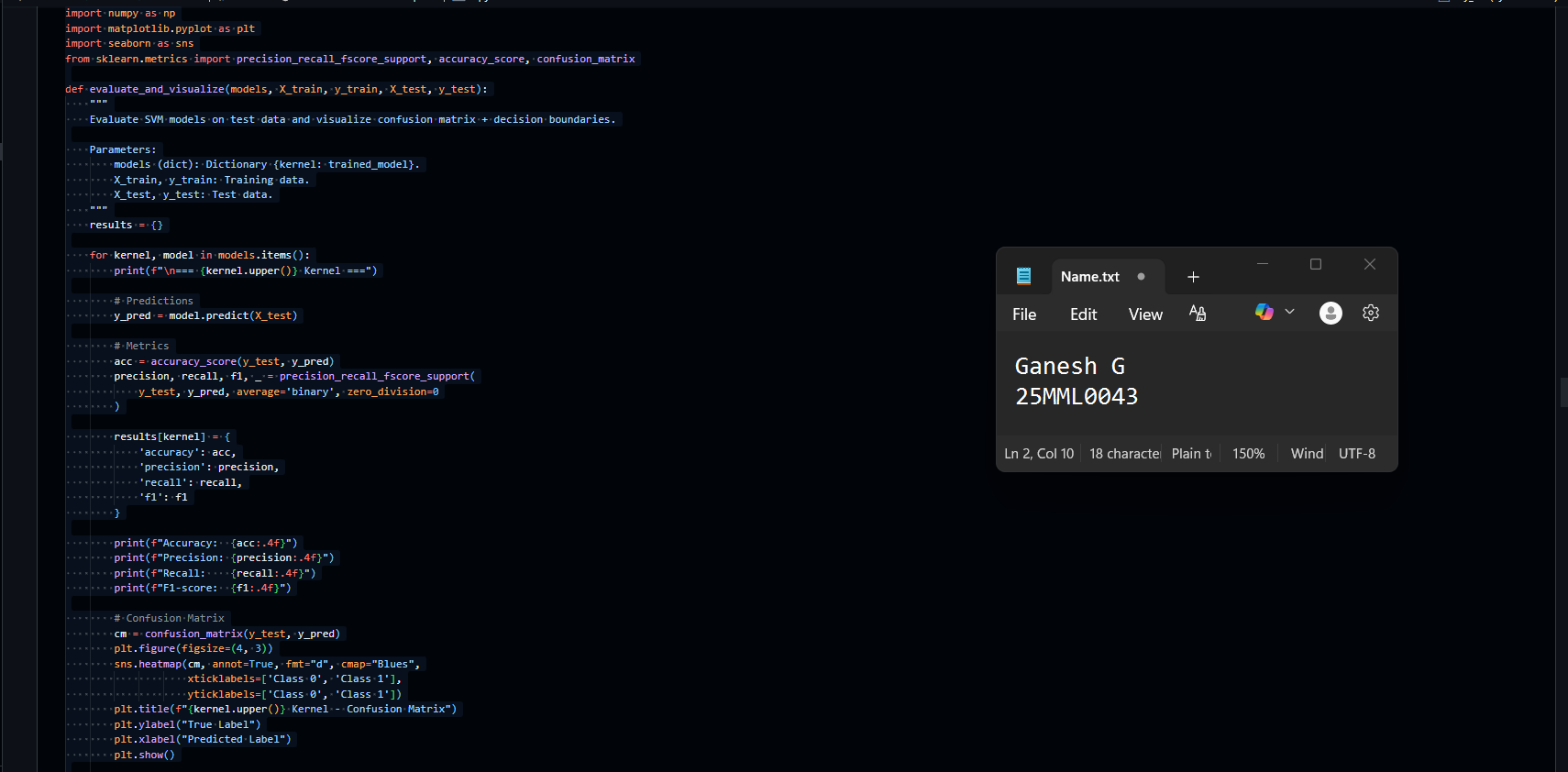


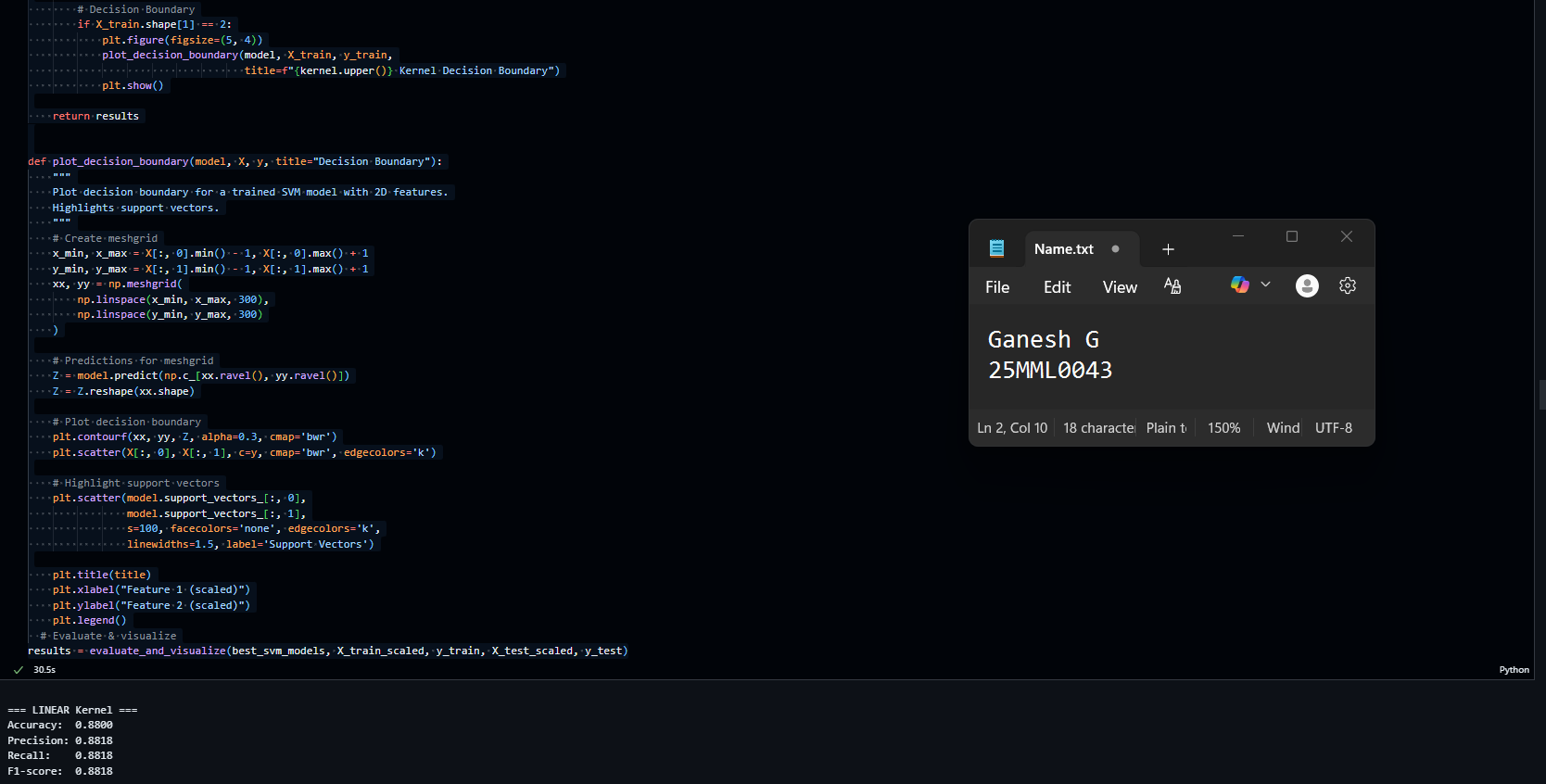


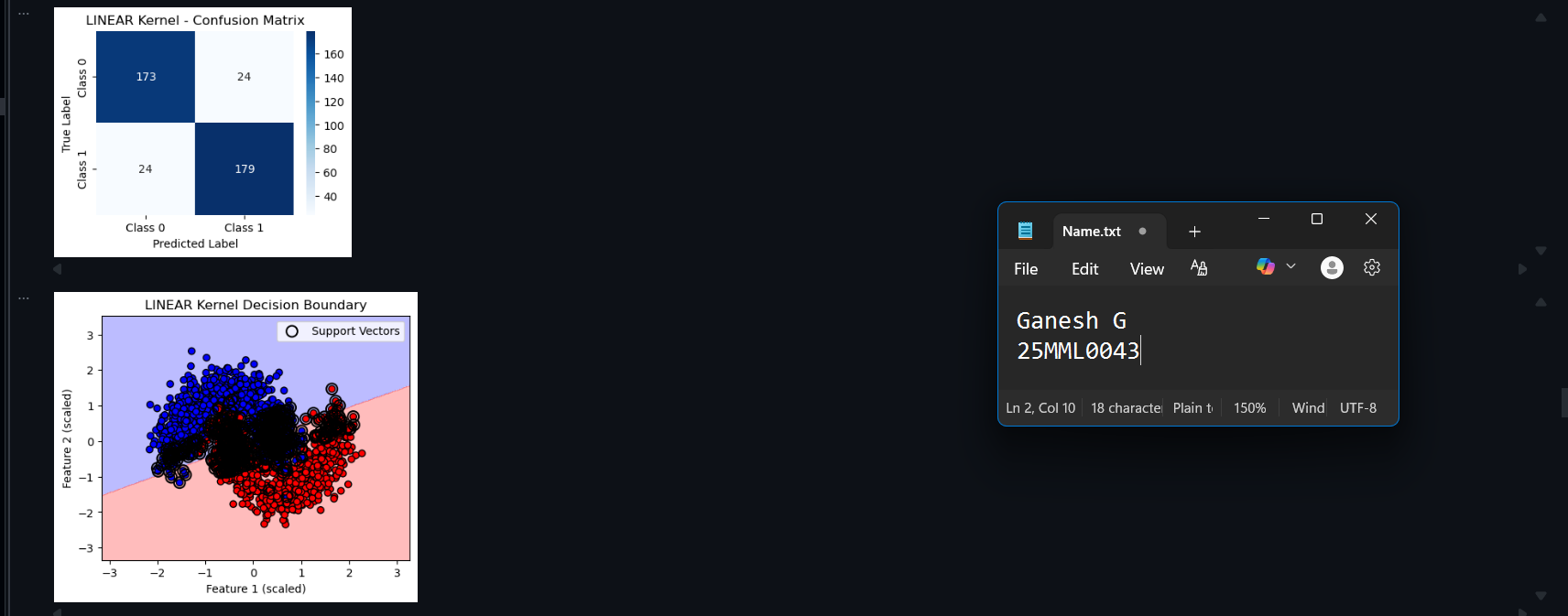




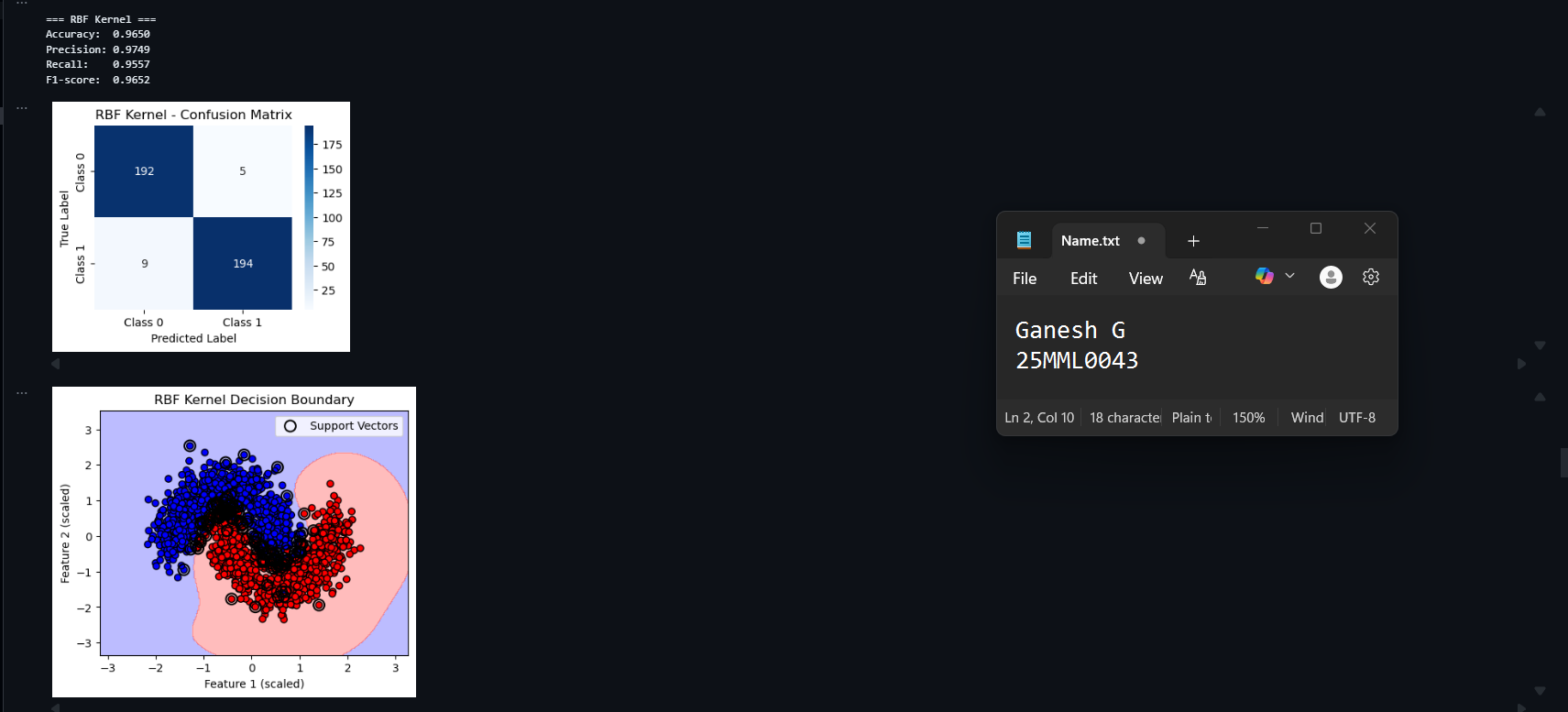


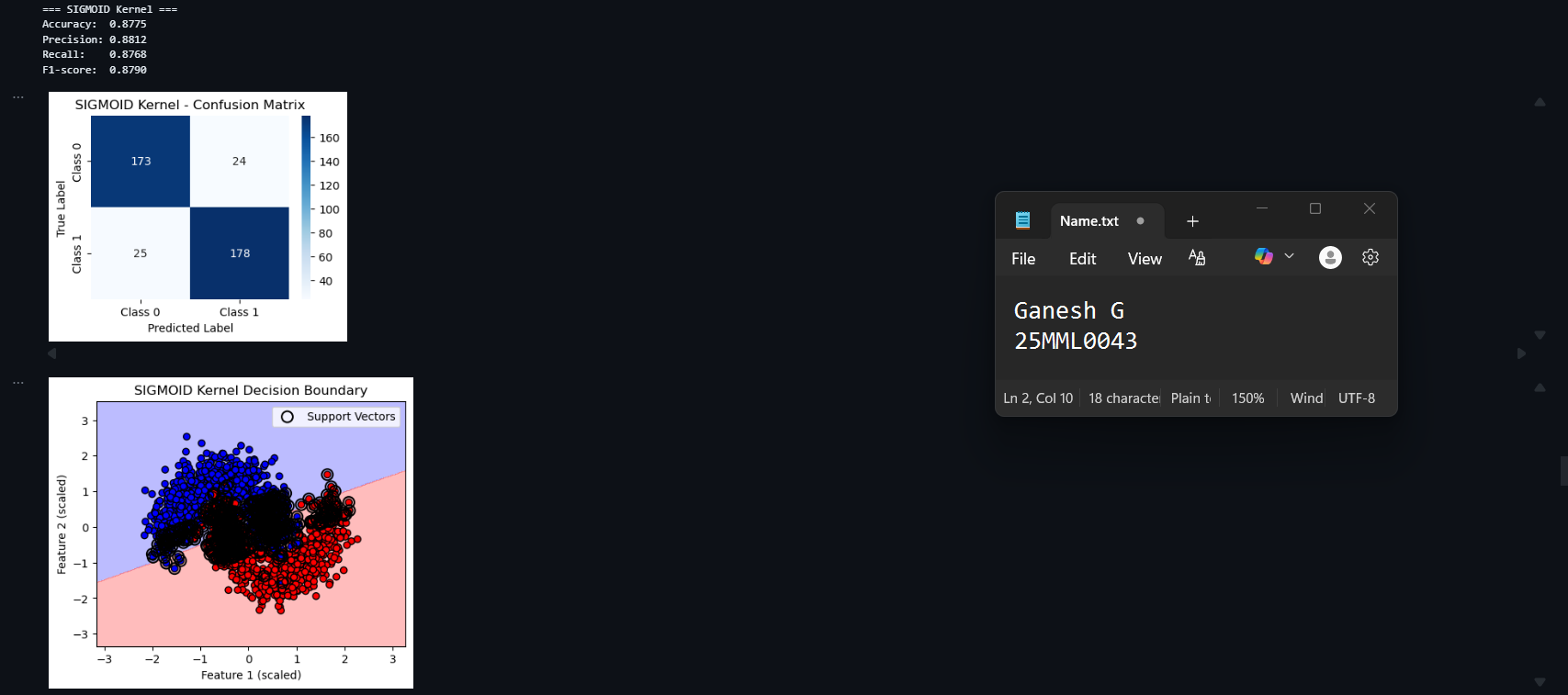












* 1. Nonlinear Circular Data



