**Parkinson's Disease Prediction by SVM**

**Introduction**

Parkinson's disease is a neurodegenerative disorder that affects movement. Early detection and diagnosis of Parkinson's disease are crucial for timely intervention and treatment. In this report, Support Vector Machine (SVM) is utilized for the prediction of Parkinson's disease based on relevant features.

**Importing Libraries**

The necessary libraries are imported to facilitate data manipulation, visualization, and model implementation. These include pandas for data handling, numpy for numerical computations, matplotlib and seaborn for data visualization, and scikit-learn modules for model training and evaluation.

**Exploratory Data Analysis**

**Moving Status Column**

The 'status' column, representing the disease status, is moved to the last position in the dataset to facilitate analysis.

**Descriptive Statistics**

Descriptive statistics are generated to gain insights into the distribution and summary statistics of the dataset's numerical features.

**Univariate Analysis**

Univariate analysis involves examining the distribution of individual features in the dataset to identify any patterns or outliers.

**Bivariate Analysis**

Bivariate analysis explores the relationships between pairs of features to uncover potential correlations or dependencies.

**Correlation Comparison with Heatmap**

A correlation heatmap is generated to visualize the pairwise correlations between features. This analysis helps identify the most relevant features for predicting Parkinson's disease.

**Applying Models**

**Data Preparation**

The dataset is divided into features (X) and the target variable (Y), where X comprises all features except 'status', and Y represents the 'status' column indicating disease status.

**Splitting Data into Training and Testing Sets**

The data is split into training and testing sets with a ratio of 70% training data and 30% testing data. This partitioning allows for model training on a subset of the data and evaluation on unseen data to assess generalization performance.

**SVM (Support Vector Machine) for Classification**

**Model Training and Prediction**

The Support Vector Machine (SVM) classifier is utilized for predicting Parkinson's disease based on the features extracted from the dataset. The following steps are performed:

SVM Initialization: An instance of the SVC (Support Vector Classifier) is created.

Model Training: The SVM classifier is trained on the training data (X\_train and y\_train) using the fit() method.

Prediction: The trained model is used to predict the disease status for the testing data (X\_test) using the predict() method.

Model Evaluation: The accuracy of the SVM classifier is calculated by comparing the predicted labels (y\_pred) with the actual labels (y\_test) using the accuracy\_score() function.

**Confusion Matrix**

A confusion matrix is generated to visualize the performance of the SVM classifier in predicting the disease status. The confusion matrix provides insights into the number of true positive, true negative, false positive, and false negative predictions made by the classifier.

**Classification Report**

The classification report provides a comprehensive summary of the classification performance, including precision, recall, F1-score, and support for each class (positive and negative). It offers valuable insights into the SVM classifier's ability to correctly classify instances belonging to different classes.

**SVM Kernel Functions for Classification**

Kernel functions are essential components of Support Vector Machine (SVM) algorithms, particularly for non-linear classification tasks. They play a crucial role in transforming input features into a higher-dimensional space where the data becomes more separable. This transformation allows SVMs to find optimal decision boundaries that effectively separate different classes in the dataset.

Kernel functions enable SVMs to handle complex patterns and relationships in the data that may not be linearly separable in the original feature space. By mapping the data into a higher-dimensional space, kernel functions facilitate the creation of non-linear decision boundaries that can accurately classify instances belonging to different classes.

In classification tasks where the relationship between features and target is non-linear, kernel functions are indispensable for capturing and modeling complex data patterns. Different types of kernel functions, such as linear, polynomial, Gaussian RBF, sigmoid, and string kernels, offer flexibility in modeling various types of non-linear relationships, allowing SVMs to achieve high accuracy and generalization performance.

Overall, kernel functions are necessary under SVM Kernel Functions for Classification to enable the effective handling of non-linear data patterns and improve the classification performance of SVM algorithms.

**Linear Kernel**

The Linear Kernel is a basic kernel function that computes the dot product of the input features. It is suitable for linearly separable data and works well when the relationship between features and target is linear. The linear kernel is chosen for its simplicity and interpretability, making it a good starting point for SVM classification.

**Gaussian RBF Kernel**

The Gaussian Radial Basis Function (RBF) kernel is a non-linear kernel that computes the similarity between data points in a high-dimensional space. It is capable of capturing complex relationships between features and target and is suitable for non-linearly separable data. The Gaussian RBF kernel is chosen to handle more complex patterns and relationships in the dataset.

**Polynomial Kernel**

The Polynomial Kernel is another non-linear kernel that computes the similarity between data points using polynomial functions. It is effective for capturing non-linear relationships between features and target. The polynomial kernel is selected to explore the potential benefits of higher-order polynomial transformations in improving classification accuracy.

**Sigmoid Kernel**

The Sigmoid Kernel is a non-linear kernel that computes the similarity between data points using sigmoid functions. It is suitable for binary classification tasks and can capture non-linear decision boundaries. The sigmoid kernel is included to investigate its effectiveness in handling non-linear relationships in the dataset.

**String Kernel**

The String Kernel is a specialized kernel function designed for handling string or text data. It computes the similarity between sequences of characters or words and is commonly used in natural language processing tasks. The string kernel is included to explore its potential applicability in capturing patterns and relationships in the dataset.

**Conclusion**

After evaluating the performance of various SVM kernel functions for Parkinson's disease prediction, it is observed that the Linear Kernel outperforms the others in terms of accuracy. The model accuracy achieved with the Gaussian RBF kernel is also noteworthy, but the Linear Kernel demonstrates superior predictive performance.

Therefore, for this specific task of Parkinson's disease prediction, the Linear Kernel function is selected as the most suitable choice. Its simplicity, interpretability, and high accuracy make it a reliable option for accurately classifying Parkinson's disease based on the provided dataset.