**Task 2: Feature Selection Methods and**

**High Accuracy ML Model**

**Task Description:**

The task involves applying three feature selection methods on a breast cancer dataset and subsequently training and evaluating machine learning models to achieve an accuracy above 90%.

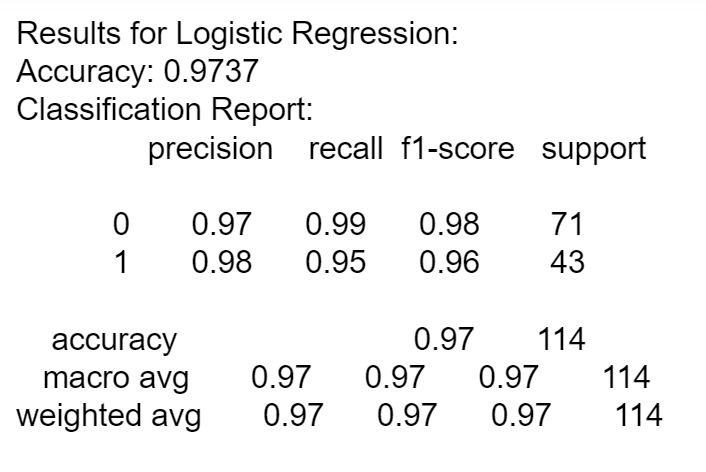
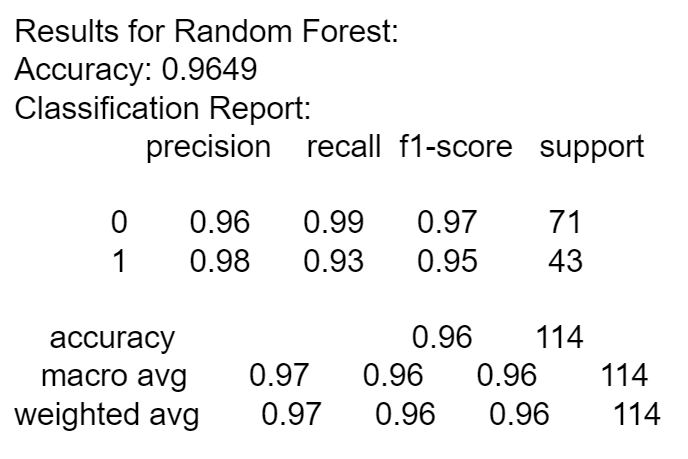
**The feature selection methods include:**

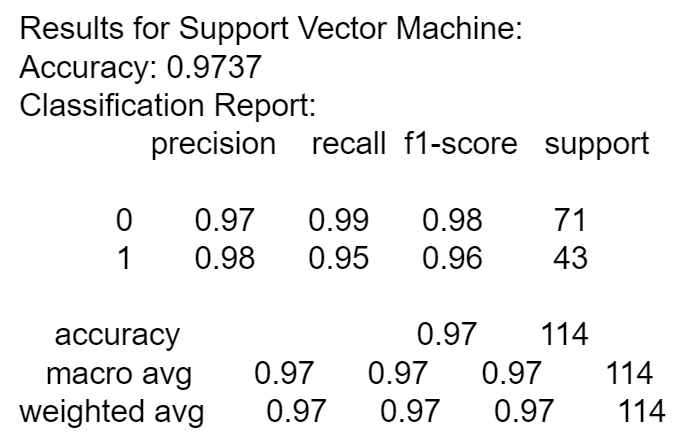
* Filter Method (SelectKBest with ANOVA)
* Wrapper Method (Recursive Feature Elimination with RandomForestClassifier)
* Embedded Method (Feature Importance from RandomForestClassifier)

**Three machine learning models are employed:**

* Logistic Regression
* Random Forest Classifier
* Support Vector Machine (SVM)

**Task Output Screenshot**



**Data Loading and Preprocessing:**

**Load the dataset:**

* Read the breast cancer dataset from Kaggle using pandas.
* Drop the 'id' column as it is not needed for analysis.

**Encode the 'diagnosis' column:** Map 'B' (Benign) to 0 and 'M' (Malignant) to 1.

**Split the data:** Divide the dataset into features (X) and target variable (Y).

**Handle Missing Values:**

* Check for missing values in the features and target variable.
* If necessary, replace missing values with the mean using SimpleImputer.

**Standardize Features:** Standardize the features using StandardScaler.

**Split into Train and Test Sets:** Split the data into training and testing sets.

**Feature Selection Methods:**

**Filter Method - SelectKBest with ANOVA:** Apply SelectKBest with ANOVA to select top k features for both training and testing sets.

**Wrapper Method - Recursive Feature Elimination (RFE) with RandomForestClassifier:** Use RFE with RandomForestClassifier to identify and select important features.

**Embedded Method - Feature Importance from RandomForestClassifier:** Train a RandomForestClassifier to assess feature importance.Select the top k features based on their importance.

**Model Training and Evaluation:**

* Choose three classifiers: Logistic Regression, Random Forest Classifier, and Support Vector Machine (SVM).

**Train and Evaluate Models:**

* For each classifier, train the model on the training set and evaluate its performance on the testing set.
* Record accuracy and classification report for each model.

**Results Display and Model Saving:**

**Display Results:** Print accuracy and classification report for each model.

**Save Trained Models:** Use pickle to save the trained models into a file named 'trained\_models.pkl'.