3. Study of the Classifier with respect to Statistical Parameters.

AIM:-

To study and Compare the performance of various classifiers using Statistical parameters Such as accuracy, Precision, recall, fi-score and confusion matrix.

OBJECTIVE:-

* To understand how different machine Learning classifiers performs on a given dataset
* To Evaluate the Classifiers using Key Statistical metrics.
* To compare and interpret the performance of each classifier for model Selection

PSEUDOCODE:-

```
1. Import required Libraries (pandas, Sklearn, etc....)
2. Load the data set.
3. Preprocess the data (handling missing values, Encoding, Scaling).
4. Split the dataset into training and testing sets.
5. Define a List of classifiers - KNN, SVM, Decision tree
6. For each classifier:
     a. Train the model on the training set.
     b. predict using the test sets.
     c. Calculate Statistical Parameters:
          - Accuracy
          - Precision
          - Recall
          - F1-Score
          - ~~Confusion Matrix~~.
7. Compare results in a tabular format.
8. Analyze and determine the best-performing classifier based on Content.
```

1. Accuracy

The proportion of correctly predicted Samples out of the total Samples

$$\text{Acearary} = \frac{\text{True Possitives} + \text{True Negative's}}{\text{Total Samples.}}$$

2. Precision

Of all Samples predicted as positive, how many were actually positive

$$\text{Precision} = \frac{\text{Tree Positives.}}{\text{True positives} + \text{False positive.}}$$

3. Recall:

Shows how well the model detects actual positives.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True positives} + \text{False Negative}}$$

4. F1-Score

$$F_1\,\text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} * \text{Recall}}$$

Evaluation Matrix

Data set used:- breast. Cancer dataset using Sklearn

Algorithm used:-

→ KNN K-nearest neighbours.

→ SVM

→ Decision tree.

KNN matrics:-

Accuracy :- 95.00%

| | Precision | recall | F1-score | Support |
|---|---|---|---|---|
| 0 | 0.93 | 0.93 | 0.93 | 43 |
| 1 | 0.96 | 0.96 | 0.96 | 71 |
| accuracy | | | 0.95 | 114 |
| macro avg | 0.94 | 0.94 | 0.94 | 114 |
| weighted avg | 0.95 | 0.95 | 0.95 | 114 |

SVM classification Report: accuracy :- 96%.

| | Precision | recall | F1-score | Support |
|---|---|---|---|---|
| 0 | 0.93 | 0.95 | 0.94 | 43 |
| 1 | 0.97 | 0.96 | 0.96 | 71 |
| accuracy | | | 0.96 | 114 |
| Macro avg | 0.95 | 0.96 | 0.95 | 114 |
| weighted avg | 0.96 | 0.96 | 0.96 | 114 |

Decision Tree Classification Report: accuracy :- 0.95

| | Precision | recall | F1 score | Support |
|---|---|---|---|---|
| 0 | 0.93 | 0.93 | 0.93 | 43 |
| 1 | 0.96 | 0.96 | 0.96 | 71 |
| accuracy | | | 0.95 | 114 |
| macro avg | 0.94 | 0.94 | 0.94 | 114 |
| weighted avg | 0.95 | 0.95 | 0.95 | 114 |

- ~~TP [True positive].~~
~~The number of instances correctly predicted as belonging to Positive class.~~

Inference:-

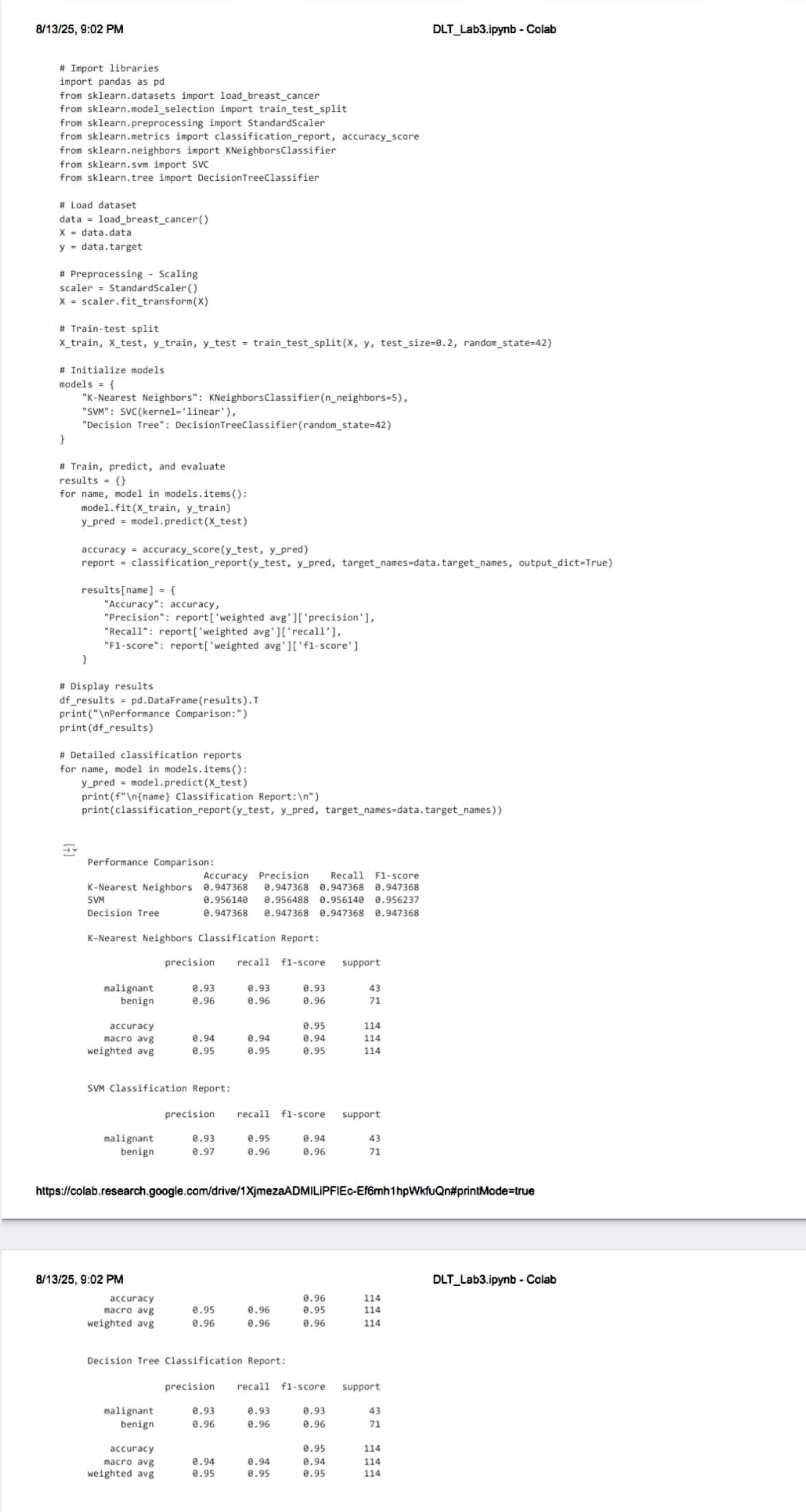
(i) Here SVM performance Best model with accuracy 96% along with balanced precision, recall and f1 Score.
It Shows Similar performance for both training and testing set is called generalization.

(ii) KNN -> Is minimal overfitting.
Training Slightly higher than testing but Still close.

(iii) Decision tree :- Training ≃ 100%. Testing ≈ 94%. -> clear overfitting (Common in decision trees without pruning).

Result:-

The Breast cancer dataset has be used to train KNN, SVM and decision tree and the Statistical parameters has been Compared and inference.



```
# Import libraries
import pandas as pd
from sklearn.datasets import load_breast_cancer
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import classification_report, accuracy_score
from sklearn.neighbors import KNeighborsClassifier
from sklearn.svm import SVC
from sklearn.tree import DecisionTreeClassifier

# Load dataset
data = load_breast_cancer()
X = data.data
y = data.target

# Preprocessing - Scaling
scaler = StandardScaler()
X = scaler.fit_transform(X)

# Train-test split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Initialize models
models = {
    "K-Nearest Neighbors": KNeighborsClassifier(n_neighbors=5),
    "SVM": SVC(kernel='linear'),
    "Decision Tree": DecisionTreeClassifier(random_state=42)
}

# Train, predict, and evaluate
results = {}
for name, model in models.items():
    model.fit(X_train, y_train)
    y_pred = model.predict(X_test)

    accuracy = accuracy_score(y_test, y_pred)
    report = classification_report(y_test, y_pred, target_names=data.target_names, output_dict=True)

    results[name] = {
        "Accuracy": accuracy,
        "Precision": report['weighted avg']['precision'],
        "Recall": report['weighted avg']['recall'],
        "F1-score": report['weighted avg']['f1-score']
    }

# Display results
df_results = pd.DataFrame(results).T
print("\nPerformance Comparison:")
print(df_results)

# Detailed classification reports
for name, model in models.items():
    y_pred = model.predict(X_test)
    print(f"\n{name} Classification Report:\n")
    print(classification_report(y_test, y_pred, target_names=data.target_names))
```

```
Performance Comparison:
                     Accuracy  Precision    Recall  F1-score
K-Nearest Neighbors  0.947368   0.947368  0.947368  0.947368
SVM                  0.956140   0.956488  0.956140  0.956237
Decision Tree        0.947368   0.947368  0.947368  0.947368

K-Nearest Neighbors Classification Report:

              precision    recall  f1-score   support

   malignant       0.93      0.93      0.93        43
      benign       0.96      0.96      0.96        71

    accuracy                           0.95       114
   macro avg       0.94      0.94      0.94       114
weighted avg       0.95      0.95      0.95       114


SVM Classification Report:

              precision    recall  f1-score   support

   malignant       0.93      0.95      0.94        43
      benign       0.97      0.96      0.96        71
```

```
    accuracy                           0.96       114
   macro avg       0.95      0.96      0.95       114
weighted avg       0.96      0.96      0.96       114


Decision Tree Classification Report:

              precision    recall  f1-score   support

   malignant       0.93      0.93      0.93        43
      benign       0.96      0.96      0.96        71

    accuracy                           0.95       114
   macro avg       0.94      0.94      0.94       114
weighted avg       0.95      0.95      0.95       114
```