Midterm Project Review

Professor: Maryam Ravan

Presenter: Bahaduri Prachiti Jagdish

ABOUT SLEEP QUALITY

Sleep quality can be measured by 3 criterias:

❏ Sleep latency
❏ Sleep efficiency
❏ Percentage of deep sleep

To calculate these parameters we need to distinguish between 3 sleep categories:

❏ Wakefulness
❏ light sleep + REM (S1, S2, REM)
❏ deep sleep (S3, S4)

ABOUT THE DATASET

- DATASET - SC Sleep-EDF Database
- There are a total of 800 observations in each sleep category with 15 features
- There are three sleep categories according to the given dataset
  - SC1 - contains features for light sleep + REM category
  - SC2 - contains features for deep sleep category
  - SC3 - contains features for awake category

PROBLEM STATEMENT

❏ Compare the accuracies across 6 classifiers
❏ Perform 10 fold cross validation
❏ Demonstrate a curve showing the accuracy of all methods versus the number of features
❏ Provide the test performance and calculate the Confusion Matrix, Accuracy, Sensitivity, Specificity, F1-score, False Positive Rate, and False Negative Rate for each class using the best number of features for that class.

Naïve Base

Decision Tree

Random Forest

Linear discriminant analysis

Quadratic discriminant analysis

Multilayer Perceptron

AN OVERVIEW OF THE DATASET

```
SC = pd.concat([sc1, sc2, sc3], axis = 0, join='inner', ignore_index=True)
SC.head()
```

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	-1.13770	-0.78326	-0.72821	-1.30810	0.71241	-1.15460	-1.21720	1.74920	0.60935	-1.37700	-1.30780	-0.697440	-1.06690	1.86660	-1.03540	1
1	-0.93703	-0.16375	0.32545	-1.16010	0.13397	-0.93057	-1.11390	1.22920	0.52222	-0.87295	-1.15980	-0.258450	-1.04140	2.01890	-0.95348	1
2	-0.87882	-1.03170	-0.74039	-1.00300	0.57247	-0.78881	-0.94716	1.76390	0.96157	-0.58891	-1.00280	-0.074423	-0.66255	1.31660	-0.83350	1
3	-0.81328	-0.47618	-0.01379	-0.92353	0.80455	-0.81035	-0.90316	1.71100	0.72966	-0.69230	-0.92341	-0.618610	-0.78337	0.94421	-0.80556	1
4	-0.49330	-0.63534	0.29655	-0.31032	-0.30882	-0.22440	-0.43386	-0.59472	0.84042	0.29545	-0.30986	-0.347410	-0.35576	0.48155	-0.51734	1

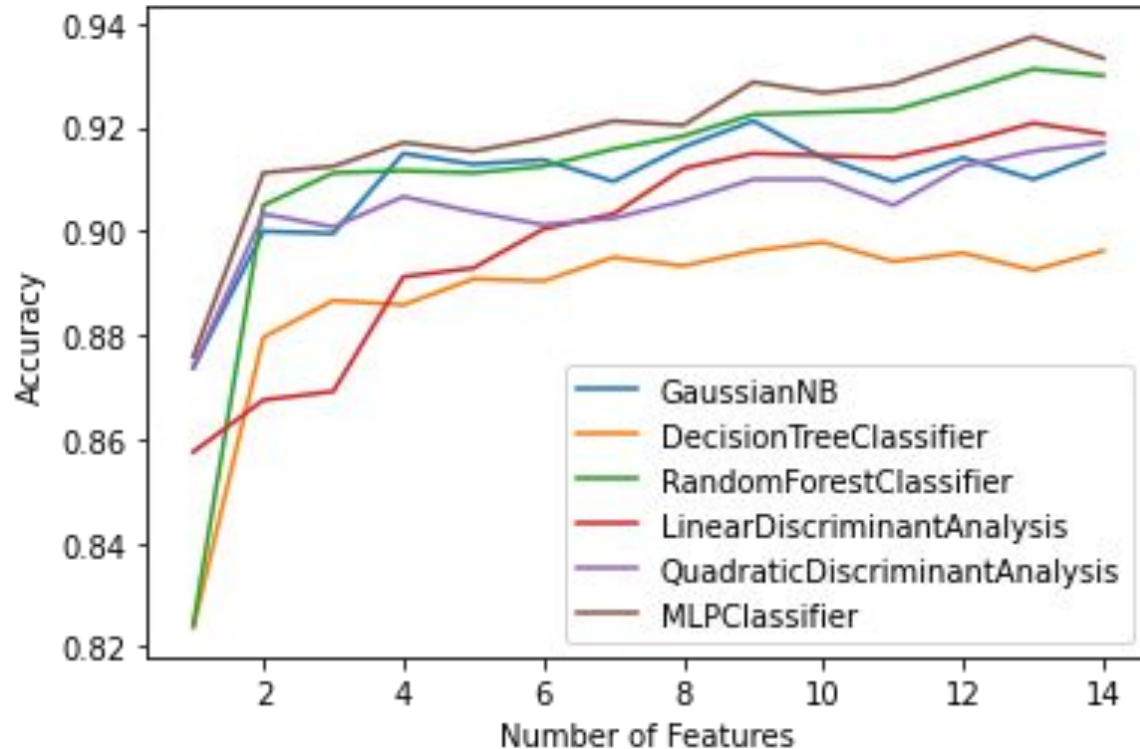# ACCURACY VS THE NUMBER OF FEATURES CURVE

TABLE INDICATING THE BEST NUMBER OF FEATURES FOR EACH CLASSIFIERS AND THEIR CORRESPONDING ACCURACIES

CLASSIFIERS	BEST NO. OF FEATURES	ACCURACIES
Naïve Base	9	0.92
Decision Tree	10	0.90
Random Forest	13	0.93
Linear discriminant analysis	13	0.92
Quadratic discriminant analysis	14	0.92
Multilayer Perceptron	13	0.94

CONFUSION MATRICES AND THE TEST PERFORMANCE METRICS USING THE BEST NO. OF FEATURES FOR EACH CLASSIFIERS

CLASSIFIER: NAIVE BAYES WITH 9 BEST FEATURES

- ❏ **Accuracy: 0.9213**
- ❏ **Sensitivity: 0.9252**
- ❏ **Specificity: 0.9039**
- ❏ **F1-score: 0.9112**
- ❏ **False Positive Rate: 0.0961**
- ❏ **False Negative Rate: 0.0748**

Confusion Matrix for 9 Features with GaussianNB()
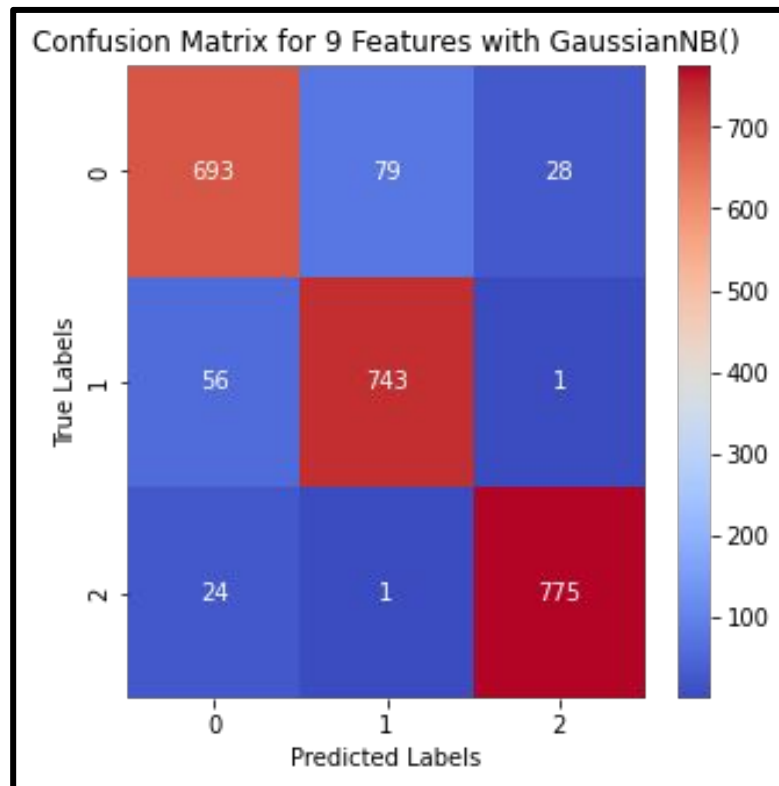
CLASSIFIER: DECISION TREE WITH 10 BEST FEATURES

- ❑ **Accuracy: 0.8975**
- ❑ **Sensitivity: 0.8737**
- ❑ **Specificity: 0.8784**
- ❑ **F1-score: 0.8743 False**
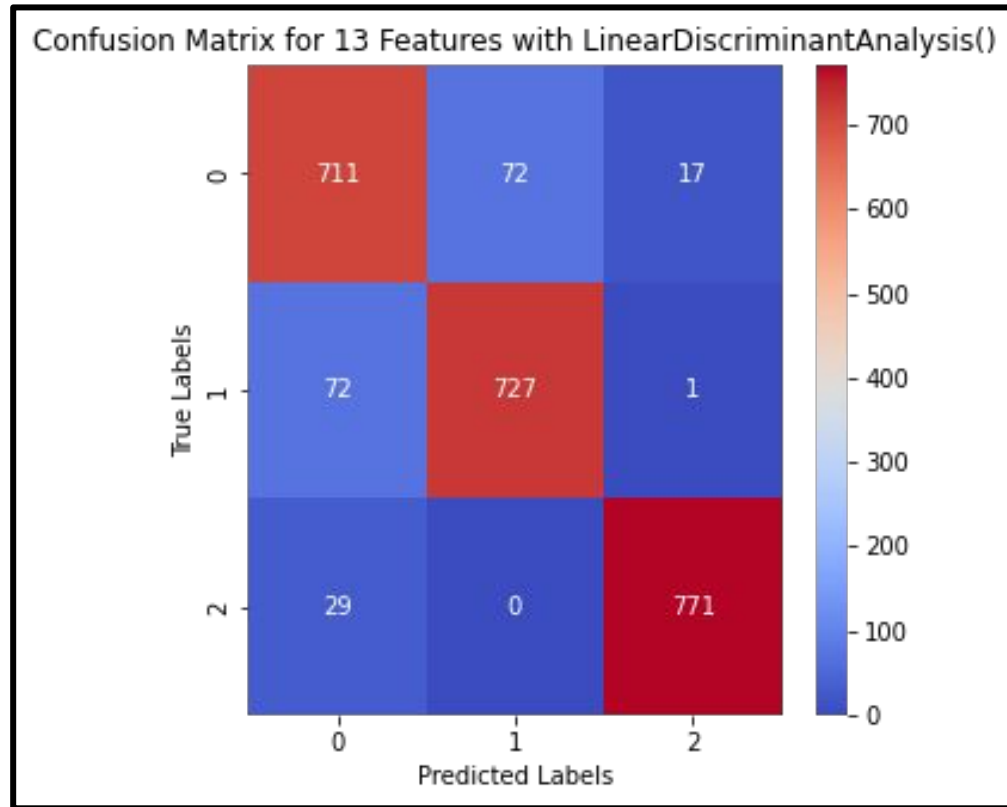- ❑ **Positive Rate: 0.1216**
- ❑ **False Negative Rate: 0.1263**

Confusion Matrix for 10 Features with DecisionTreeClassifier()
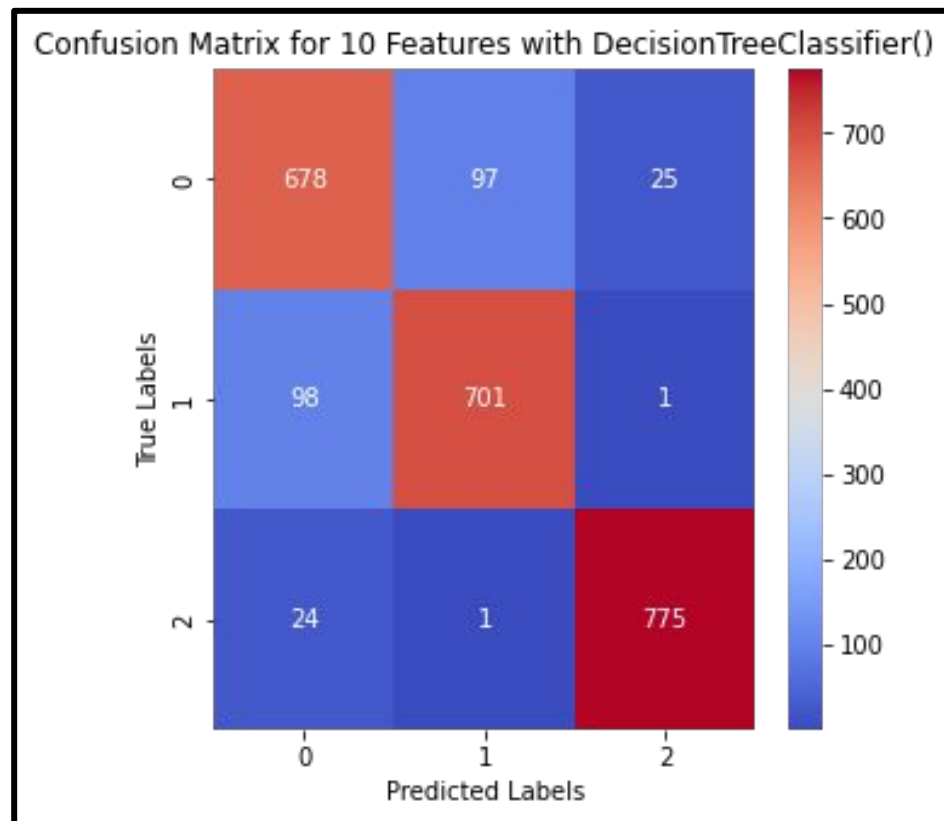
CLASSIFIER: RANDOM FOREST WITH 13 BEST FEATURES

- ❏ **Accuracy: 0.9317**
- ❏ **Sensitivity: 0.9271**
- ❏ **Specificity: 0.9160**
- ❏ **F1-score: 0.9199**
- ❏ **False Positive Rate: 0.0840**
- ❏ **False Negative Rate: 0.0729**

Confusion Matrix for 13 Features with RandomForestClassifier()
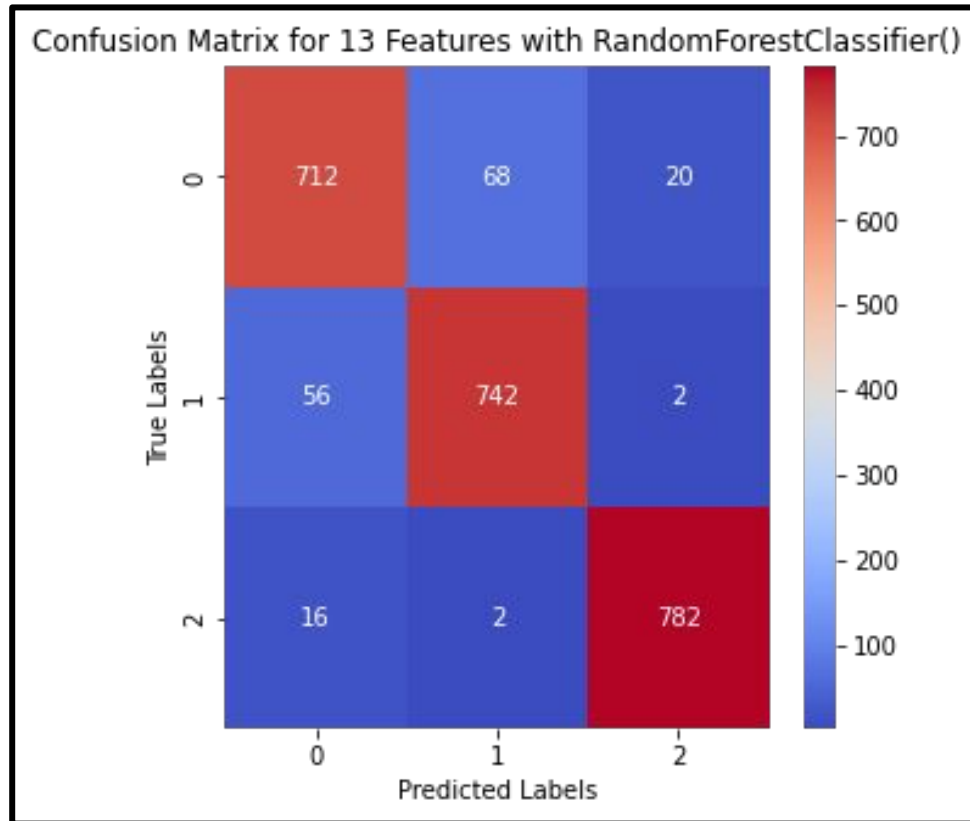
CLASSIFIER: LINEAR DISCRIMINANT ANALYSIS WITH 13 BEST FEATURES

- ❏ **Accuracy: 0.9204**
- ❏ **Sensitivity: 0.9080**
- ❏ **Specificity: 0.9099**
- ❏ **F1-score: 0.9080**
- ❏ **False Positive Rate: 0.0901**
- ❏ **False Negative Rate: 0.0920**

Confusion Matrix for 13 Features with LinearDiscriminantAnalysis()

CLASSIFIER: QUADRATIC DISCRIMINANT ANALYSIS WITH 14 BEST FEATURES

- ❏ **Accuracy: 0.9154**
- ❏ **Sensitivity: 0.9431**
- ❏ **Specificity: 0.8810**
- ❏ **F1-score: 0.9033**
- ❏ **False Positive Rate: 0.1190**
- ❏ **False Negative Rate: 0.0569**

Confusion Matrix for 14 Features with QuadraticDiscriminantAnalysis()
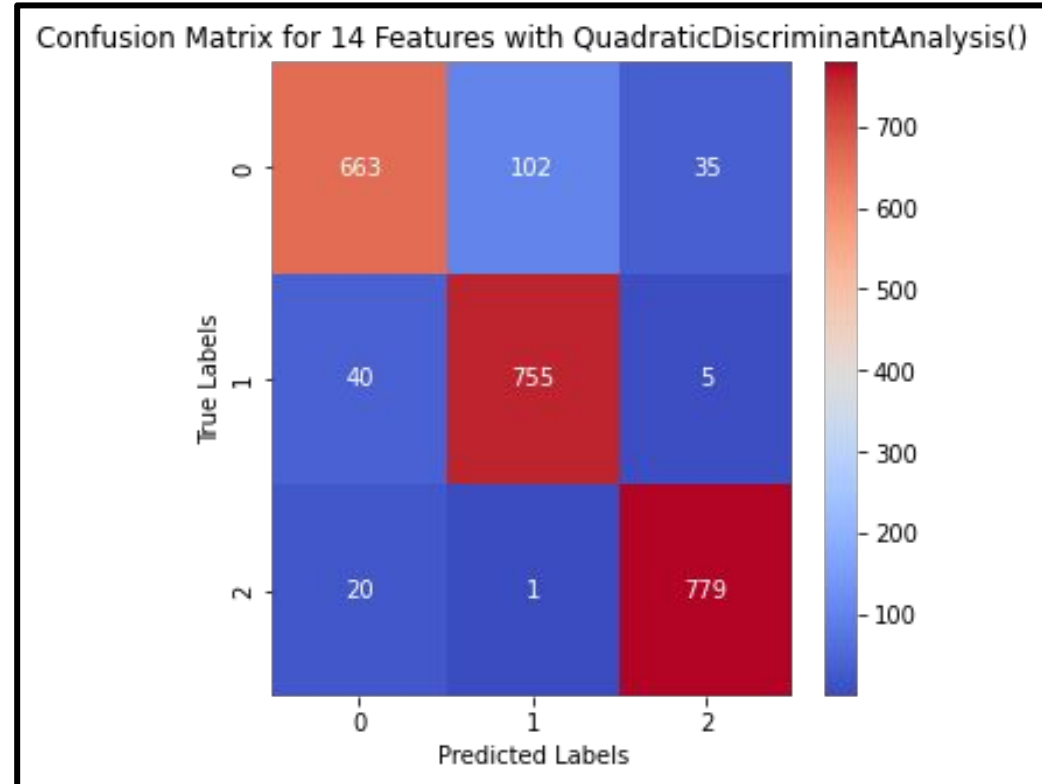
CLASSIFIER: MLP CLASSIFIER WITH 13 BEST FEATURES

- ❑ **Accuracy: 0.9350**
- ❑ **Sensitivity: 0.9392**
- ❑ **Specificity: 0.9146**
- ❑ **F1-score: 0.9246**
- ❑ **False Positive Rate: 0.0854**
- ❑ **False Negative Rate: 0.0608**

Confusion Matrix for 13 Features with MLPClassifier()
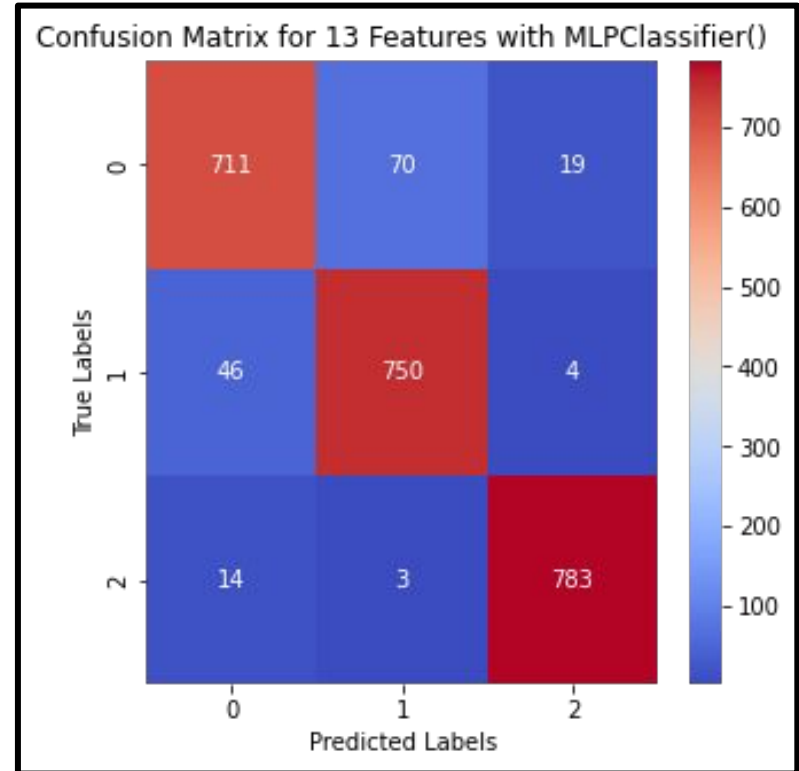
Which classifier has the best performance with less number of features?
– MLP with 13 best features

THANK YOU!