MARMARA UNIVERSITY

COMPUTER ENGINEERING DEPARTMENT

CSE4088

INTRODUCTION TO MACHINE LEARNING

FINAL PROJECT
PRESENTATION

"Mobile Price Classification"

Group Members:

Onurcan İşler 150120825
Berk Kırtay 150118043
Erkam Karaca 150118021

PROBLEM DESCRIPTION

Given various features of phones, determine their price classes:

- 0 means phone is “cheap”,
- 1 means phone is “standard”,
- 2 means phone is “expensive”,
- 3 means phone is “very expensive”.

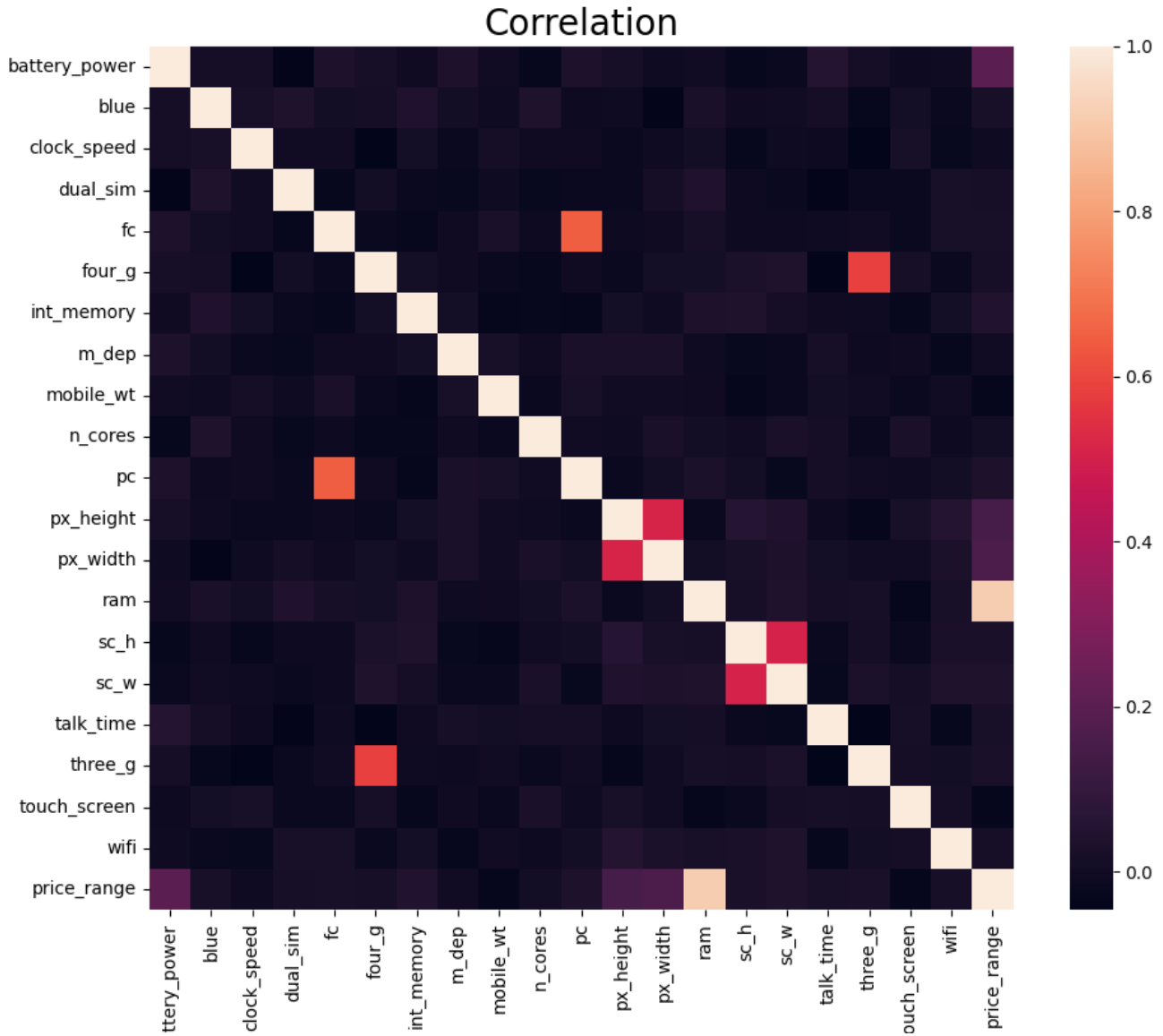
Features include Wi-Fi, 4G, Bluetooth, RAM (GB), Battery Power (mAh) etc.

EXPLORATORY DATA ANALYSIS-1

- There are 21 features. Train set has 2000 samples, and test set has 1000 samples but without labels.

- So, we will use some portion of train set as our test set.

- Correlation matrix simply tells you how pairs of variables are related. It is useful to identify the patterns in the given data.



EXPLORATORY DATA ANALYSIS-2

- Let's visualize how RAM size (GB), and Battery Power (mAh) effect the price classification of a phone.

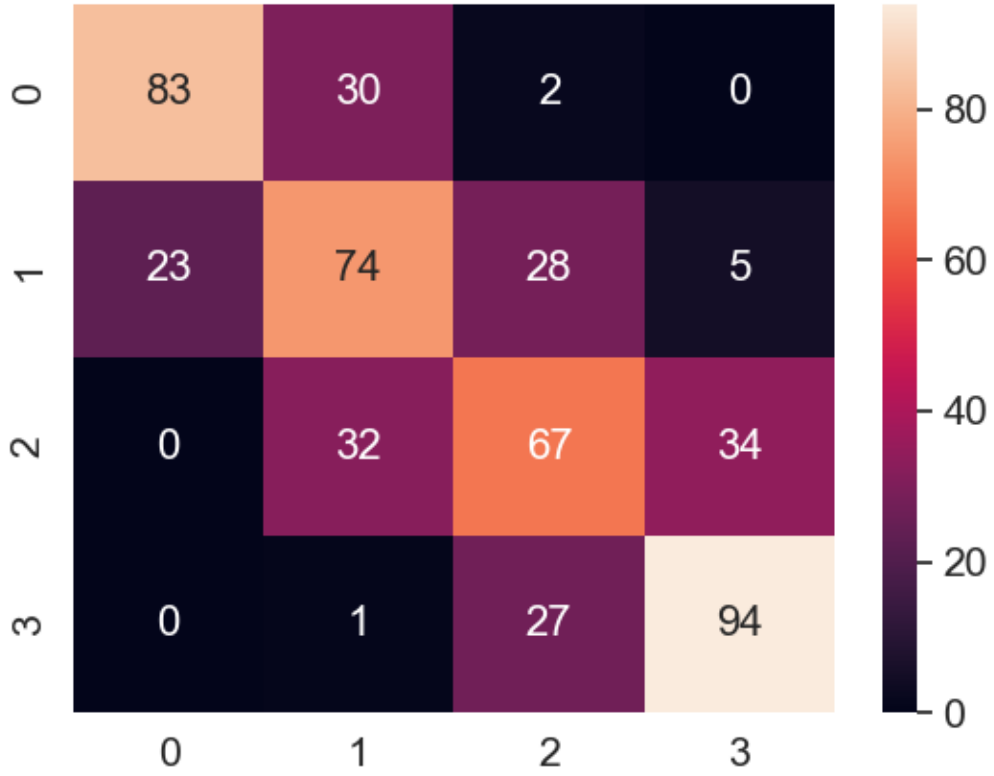
ALGORITHMS

- Since we are interested to classify the phones, we have used:
  - Logistic Regression
  - Random Forest
  - Support Vector Machines
  - K-nearest Neighbor
  - Decision Tree

Now, let’s give codes and results for each of these classification algorithms.

LOGISTIC REGRESSION

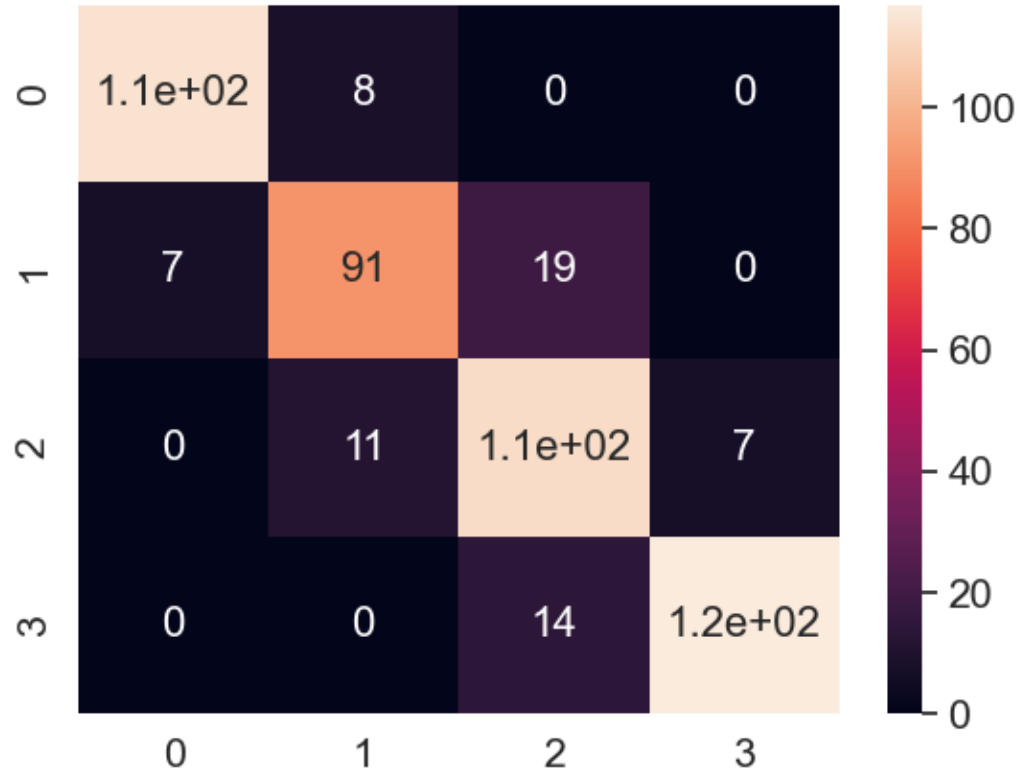
```
# 42 is the Answer to the Ultimate Question of Life...
lr = LogisticRegression(random_state = 42, max_iter=100000)
lr.fit(X_train, y_train)
y_pred_lr = lr.predict(X_test)
accuracy = metrics.accuracy_score(y_test, y_pred_lr)
```



Accuracy: 0.752

RANDOM FOREST

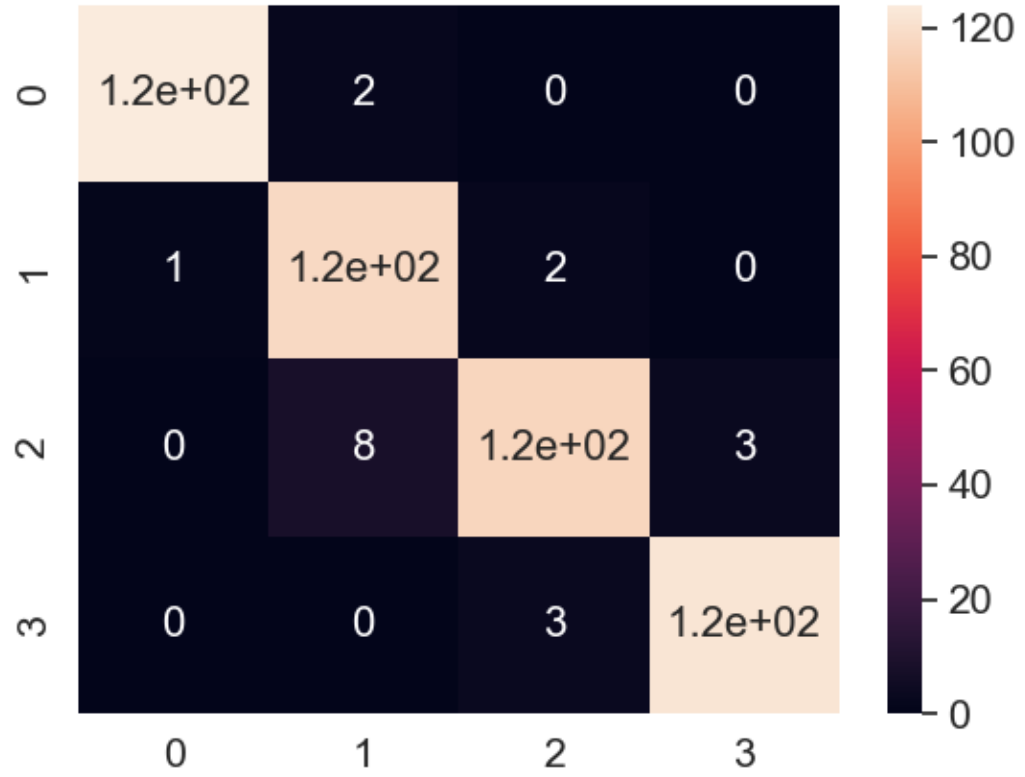
```
rc = RandomForestClassifier(random_state=42)
rc.fit(X_train,y_train)
y_pred_rc = rc.predict(X_test)
accuracy = metrics.accuracy_score(y_test, y_pred_rc)
```



Accuracy: 0.872

SUPPORT VECTOR MACHINES

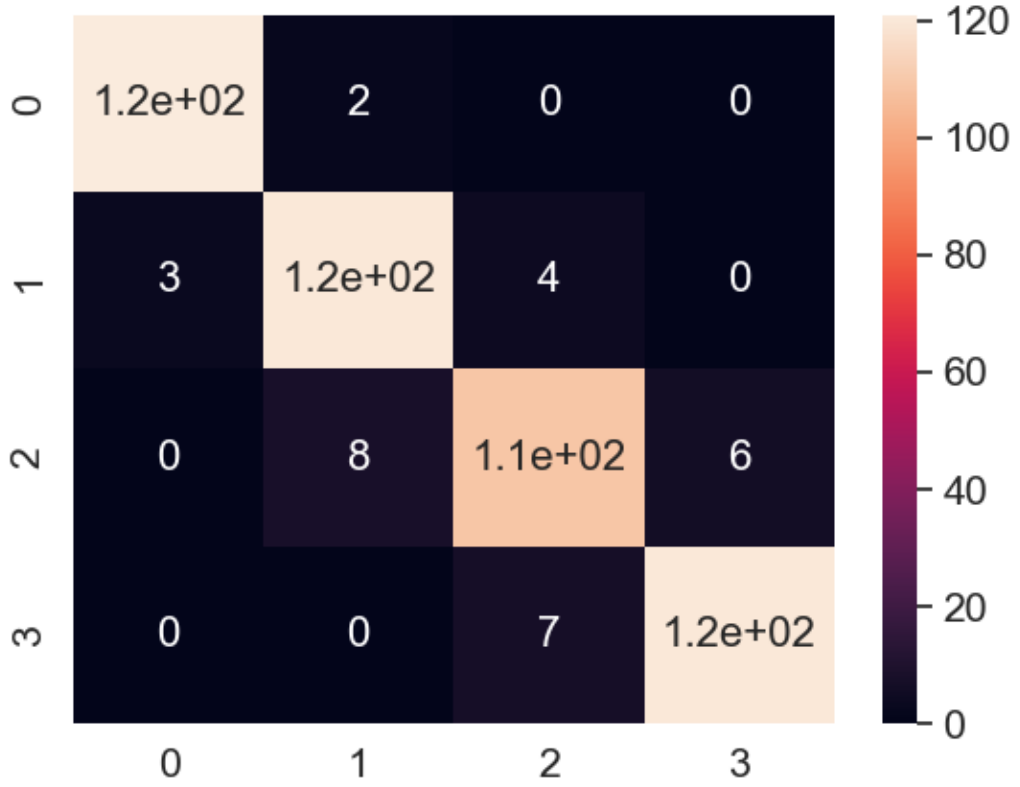
```
svc = SVC()
svc.fit(X_train, y_train)
y_pred_svc = svc.predict(X_test)
accuracy_svc = metrics.accuracy_score(y_test, y_pred_svc)
```



Accuracy: 0.964

K-NEAREST NEIGHBORS

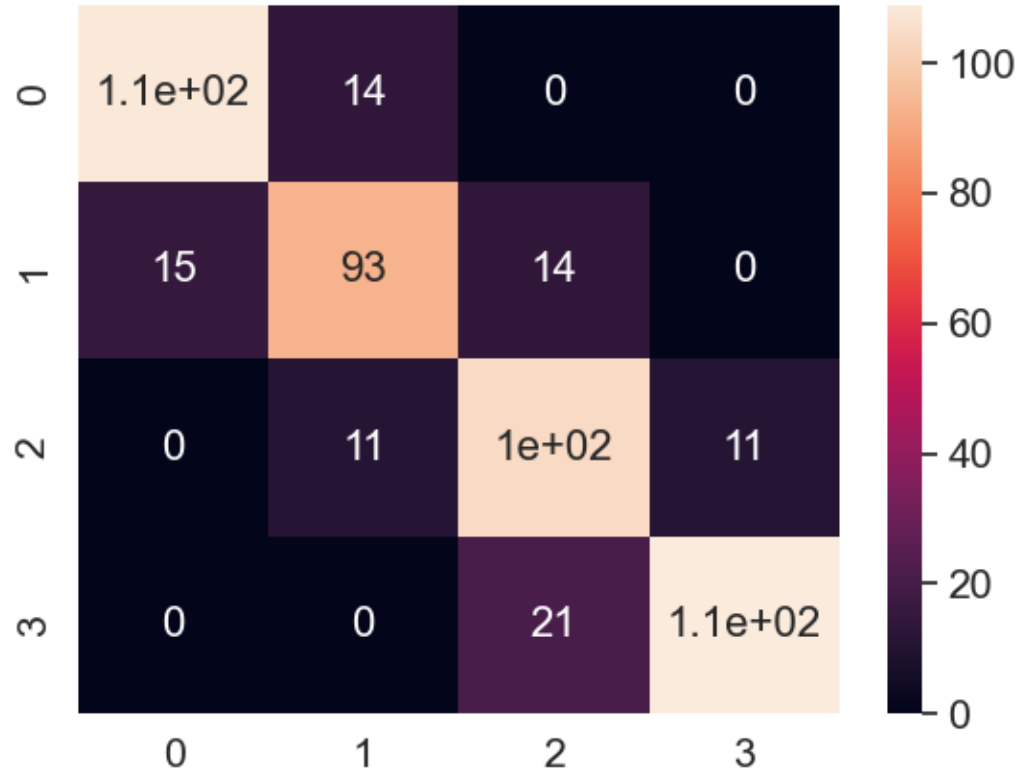
```
model = KNeighborsClassifier()
model.fit(X_train,y_train)
predicted= model.predict(X_test)
accuracy = metrics.accuracy_score(y_test,predicted)
```



Accuracy: 0.916

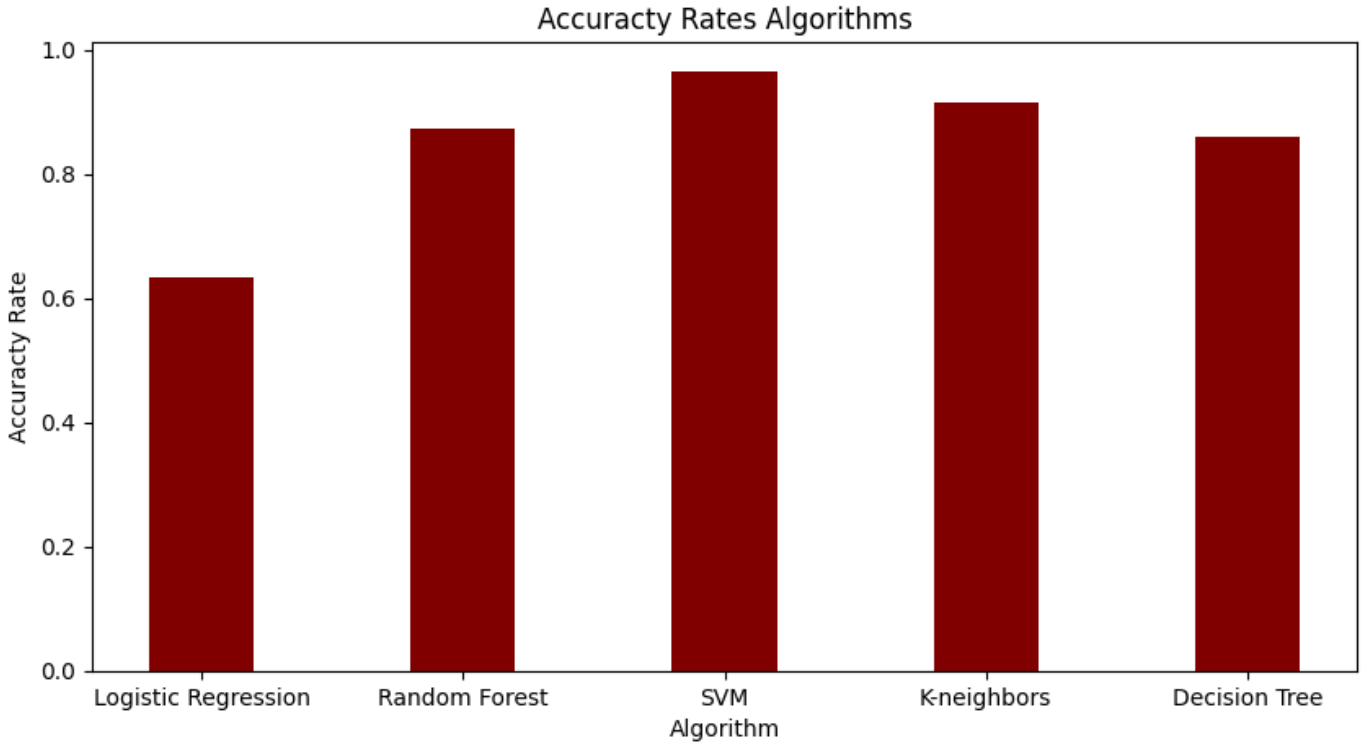
DECISION TREE

```
first_tree = DecisionTreeClassifier()
first_tree.fit(X_train, y_train)
y_pred=first_tree.predict(X_test)
accuracy = metrics.accuracy_score(y_test,y_pred)
```



Accuracy: 0.860

RESULTS



- We see SVM has slightly better accuracy rate than the other algorithms.
- In fact, SVMs are powerful on data where number of dimensions are large compared to sample size.
- In our case sample size is 2000, and dimensions is 21. So, SVM is slightly better option here.

REFERENCES

[1] Y. S. Abu-Mostafa, M. Magdon-Ismail, and H.-T. Lin, Learning from data : a short course. Pasadena, Ca: Aml Book, 2012.

[2] A. Sharma, "Mobile Price Classification, Kaggle," Kaggle, 2018. [Online]. Available: https://www.kaggle.com/datasets/iabhishekofficial/mobile-price-classification. [Accessed 2 November 2022].

[3] D. Banerjee and S. Dutta, "Predicting the housing price direction using machine learning techniques," 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI), 2017.

[4] C. Mair et al., "An investigation of machine learning based prediction systems", J. Syst.Softw., 2000.

[5] N. Hu, "Classification of Mobile Phone Price Dataset Using Machine Learning Algorithms," 2022 3rd International Conference on Pattern Recognition and Machine Learning (PRML), 2022, pp. 438-443.

[6] M. Çetın and Y. Koç, "Mobile Phone Price Class Prediction Using Different Classification Algorithms with Feature Selection and Parameter Optimization," 2021 5th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), 2021, pp. 483-487.

[7] Harshitha, B., Maria Rufina, P. \& Shilpa, B.L. Comparison of Different Classification Algorithms for Prediction of Heart Disease by Machine Learning Techniques. SN COMPUT. SCI. 4, 128 (2023).