# AMERICAN INTERNATIONAL UNIVERSITY-BANGLADESH

## Faculty of Science and Technology

# Assignment Cover Sheet

| | | | |
|---|---|---|---|
| Assignment Title: | **Heart-Failure Prediction with machine learning** | | |
| Assignment No: | 01 | Date of Submission: 18 April 2022 | |
| Course Title: | Data Warehousing and Data Mining | | |
| Course Code: | **01211** | Section: A | |
| Semester: | Spring | 2021-22 | Course Teacher: **AKINUL ISLAM JONY** |

**Declaration and Statement of Authorship:**

1. I/we hold a copy of this Assignment/Case-Study, which can be produced if the original is lost/damaged.
2. This Assignment/Case-Study is my/our original work and no part of it has been copied from any other student's work or from any other source except where due acknowledgement is made.
3. No part of this Assignment/Case-Study has been written for me/us by any other person except where such collaboration has been authorized by the concerned teacher and is clearly acknowledged in the assignment.
4. I/we have not previously submitted or currently submitting this work for any other course/unit.
5. This work may be reproduced, communicated, compared and archived for the purpose of detecting plagiarism.
6. I/we give permission for a copy of my/our marked work to be retained by the Faculty for review and comparison, including review by external examiners.
7. I/we understand that Plagiarism is the presentation of the work, idea or creation of another person as though it is your own. It is a formofcheatingandisaveryseriousacademicoffencethatmayleadtoexpulsionfromtheUniversity. Plagiarized material can be drawn from, and presented in, written, graphic and visual form, including electronic data, and oral presentations. Plagiarism occurs when the origin of them arterial used is not appropriately cited.
8. I/we also understand that enabling plagiarism is the act of assisting or allowing another person to plagiarize or to copy my/our work.

* *Student(s) must complete all details except the faculty use part.*
** Please submit all assignments to your course teacher or the office of the concerned teacher.

| No | Name | ID | Program | Signature |
|---|---|---|---|---|
| 1 | PAUL, ANIK KUMAR | 19-39880-1 | BSc [CSE] | |
| 2 | HASAN, MD. JAHID | 19-39899-1 | BSc [CSE] | |
| 3 | Kaniz Fatima | 19-39712-1 | BSc [CSE] | |
| 4 | Dilruba khanam jesey | 19-39979-1 | BSc [CSE] | |
| 5 | Ridwan Mannan Rahat | 18-38990-3 | BSc [CSE] | |

***Faculty use only***

| | | |
|---|---|---|
| FACULTY COMMENTS | **Marks Obtained** | |
| | **Total Marks** | |

# Heart-Failure Prediction with machine learning

PAUL, ANIK KUMAR
19-39880-1
anikkumarpaul995@gmail.com

HASAN, MD. JAHID
19-39899-1
jahidh101894@gmail.com

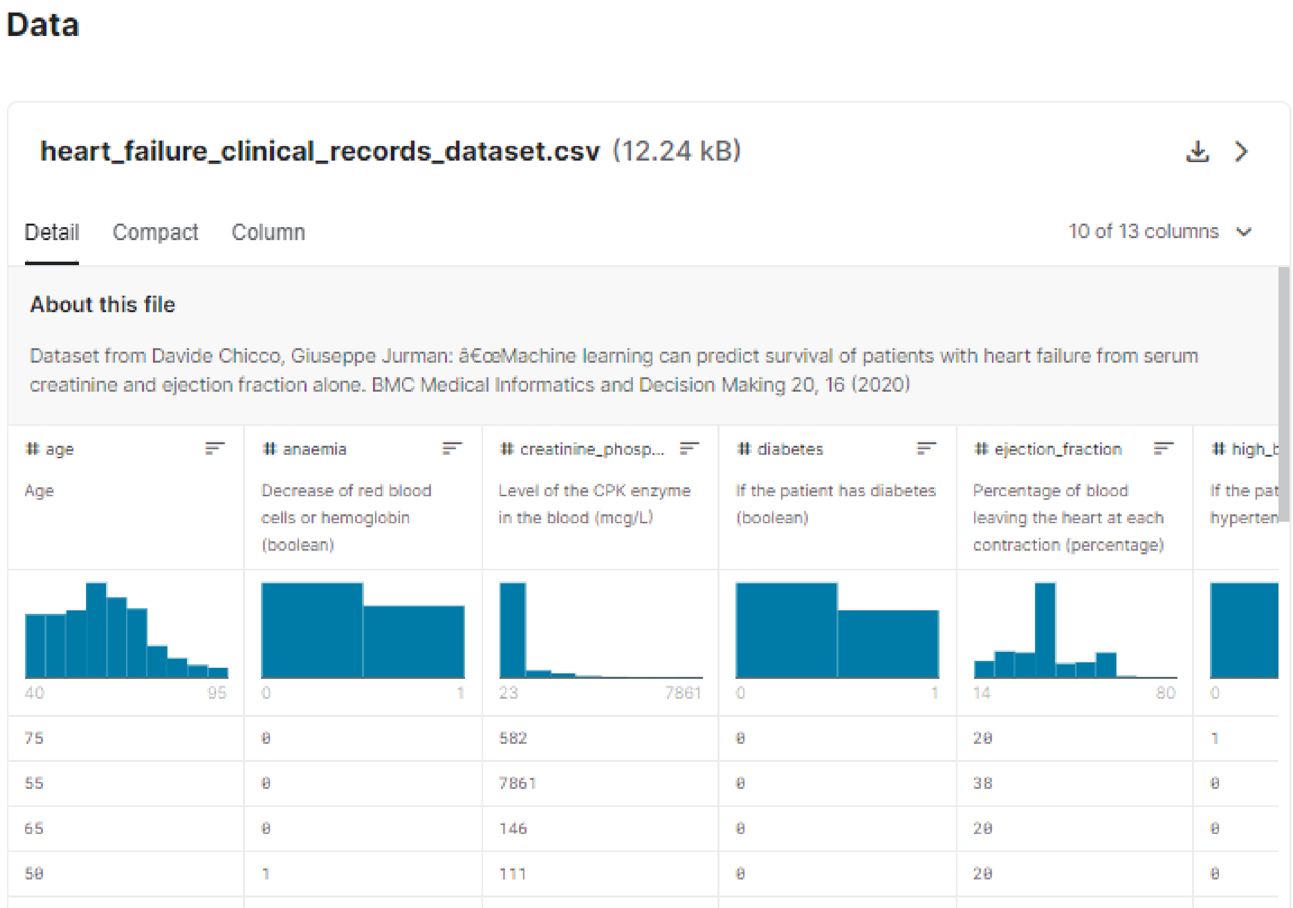
Kaniz Fatima
19-39712-1
Kanizfatima662@gmail.com

Dilruba khanam jesey
19-39979-1
Jess.jesey@gmail.com

Ridwan Mannan Rahat
18-38990-3
18-38990-3@student.aiub.edu

## Introduction

Heart diseases has increased due to smoking, lack of physical activity and other reasons which put pressure on the heart. Heart failure occurs when the heart cannot pump enough blood to meet the needs of the body. Machine learning can predict patients survival from their data by simulating them and important features can help in survival.



Data

heart_failure_clinical_records_dataset.csv (12.24 kB)

About this file

Dataset from Davide Chicco, Giuseppe Jurman: â€œMachine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. BMC Medical Informatics and Decision Making 20, 16 (2020)

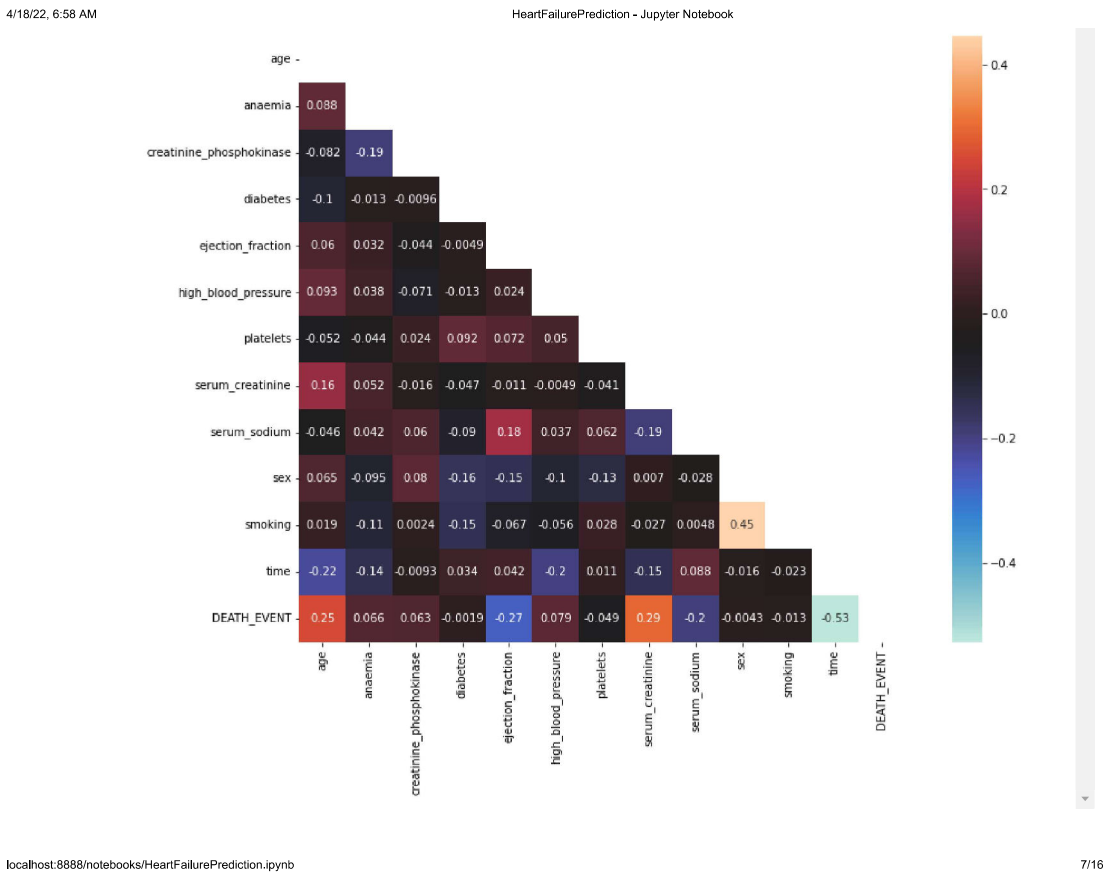
| # age | # anaemia | # creatinine_phosp... | # diabetes | # ejection_fraction | # high_b |
|---|---|---|---|---|---|
| Age | Decrease of red blood cells or hemoglobin (boolean) | Level of the CPK enzyme in the blood (mcg/L) | If the patient has diabetes (boolean) | Percentage of blood leaving the heart at each contraction (percentage) | If the pat hyperten |
| 40 95 | 0 1 | 23 7861 | 0 1 | 14 80 | 0 |
| 75 | 0 | 582 | 0 | 20 | 1 |
| 55 | 0 | 7861 | 0 | 38 | 0 |
| 65 | 0 | 146 | 0 | 20 | 0 |
| 50 | 1 | 111 | 0 | 20 | 0 |

Figure 1: Dataset
[1]

This data-set contain the medical records of 299 heart failure patients collected at the Faisalabad Institute of Cardiology, during April–December 2015. The data-set contains 13 features.

# Model Development

We analyze a data-set of 299 patients with heart failure which is collected in 2015. We apply KNN, Naive Bayes and Decision Tree to predict the patients survival and also to rank the features corresponding to the most important risk factors. We also compare these results with machine learning algorithms.



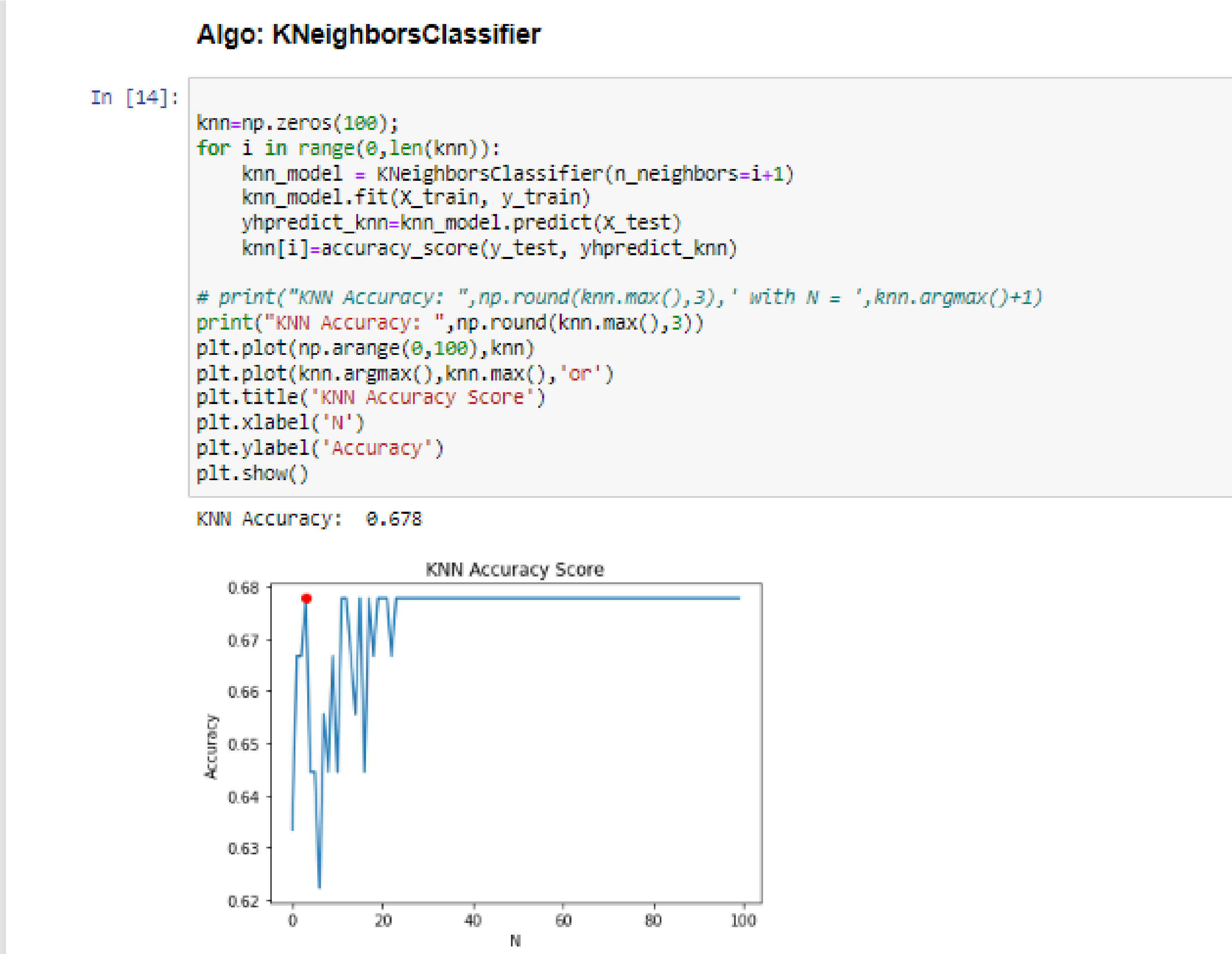
4/18/22, 6:58 AM HeartFailurePrediction - Jupyter Notebook

| | age | anaemia | creatinine_phosphokinase | diabetes | ejection_fraction | high_blood_pressure | platelets | serum_creatinine | serum_sodium | sex | smoking | time | DEATH_EVENT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| age | | | | | | | | | | | | | |
| anaemia | 0.088 | | | | | | | | | | | | |
| creatinine_phosphokinase | -0.082 | -0.19 | | | | | | | | | | | |
| diabetes | -0.1 | -0.013 | -0.0096 | | | | | | | | | | |
| ejection_fraction | 0.06 | 0.032 | -0.044 | -0.0049 | | | | | | | | | |
| high_blood_pressure | 0.093 | 0.038 | -0.071 | -0.013 | 0.024 | | | | | | | | |
| platelets | -0.052 | -0.044 | 0.024 | 0.092 | 0.072 | 0.05 | | | | | | | |
| serum_creatinine | 0.16 | 0.052 | -0.016 | -0.047 | -0.011 | -0.0049 | -0.041 | | | | | | |
| serum_sodium | -0.046 | 0.042 | 0.06 | -0.09 | 0.18 | 0.037 | 0.062 | -0.19 | | | | | |
| sex | 0.065 | -0.095 | 0.08 | -0.16 | -0.15 | -0.1 | -0.13 | 0.007 | -0.028 | | | | |
| smoking | 0.019 | -0.11 | 0.0024 | -0.15 | -0.067 | -0.056 | 0.028 | -0.027 | 0.0048 | 0.45 | | | |
| time | -0.22 | -0.14 | -0.0093 | 0.034 | 0.042 | -0.2 | 0.011 | -0.15 | 0.088 | -0.016 | -0.023 | | |
| DEATH_EVENT | 0.25 | 0.066 | 0.063 | -0.0019 | -0.27 | 0.079 | -0.049 | 0.29 | -0.2 | -0.0043 | -0.013 | -0.53 | |

localhost:8888/notebooks/HeartFailurePrediction.ipynb 7/16

Figure 2: Heatmap

## 0.1 Algorithm: KNN



### Algo: KNeighborsClassifier

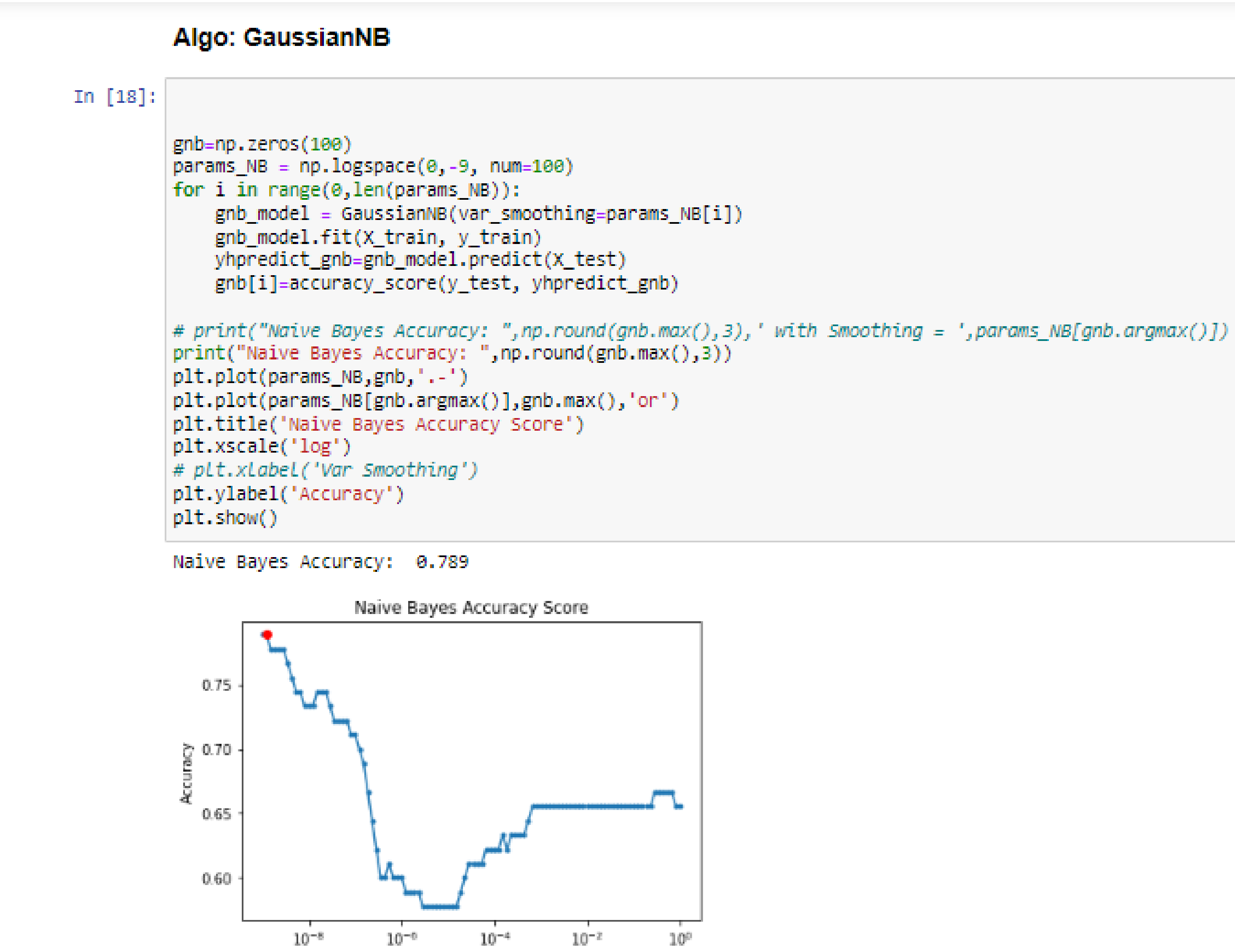
In [14]:

```
knn=np.zeros(100);
for i in range(0,len(knn)):
    knn_model = KNeighborsClassifier(n_neighbors=i+1)
    knn_model.fit(X_train, y_train)
    yhpredict_knn=knn_model.predict(X_test)
    knn[i]=accuracy_score(y_test, yhpredict_knn)

# print("KNN Accuracy: ",np.round(knn.max(),3),' with N = ',knn.argmax()+1)
print("KNN Accuracy: ",np.round(knn.max(),3))
plt.plot(np.arange(0,100),knn)
plt.plot(knn.argmax(),knn.max(),'or')
plt.title('KNN Accuracy Score')
plt.xlabel('N')
plt.ylabel('Accuracy')
plt.show()
```

```
KNN Accuracy:  0.678
```

Figure 3: KNN

## 0.2 Algorithm: GaussianNB



### Algo: GaussianNB

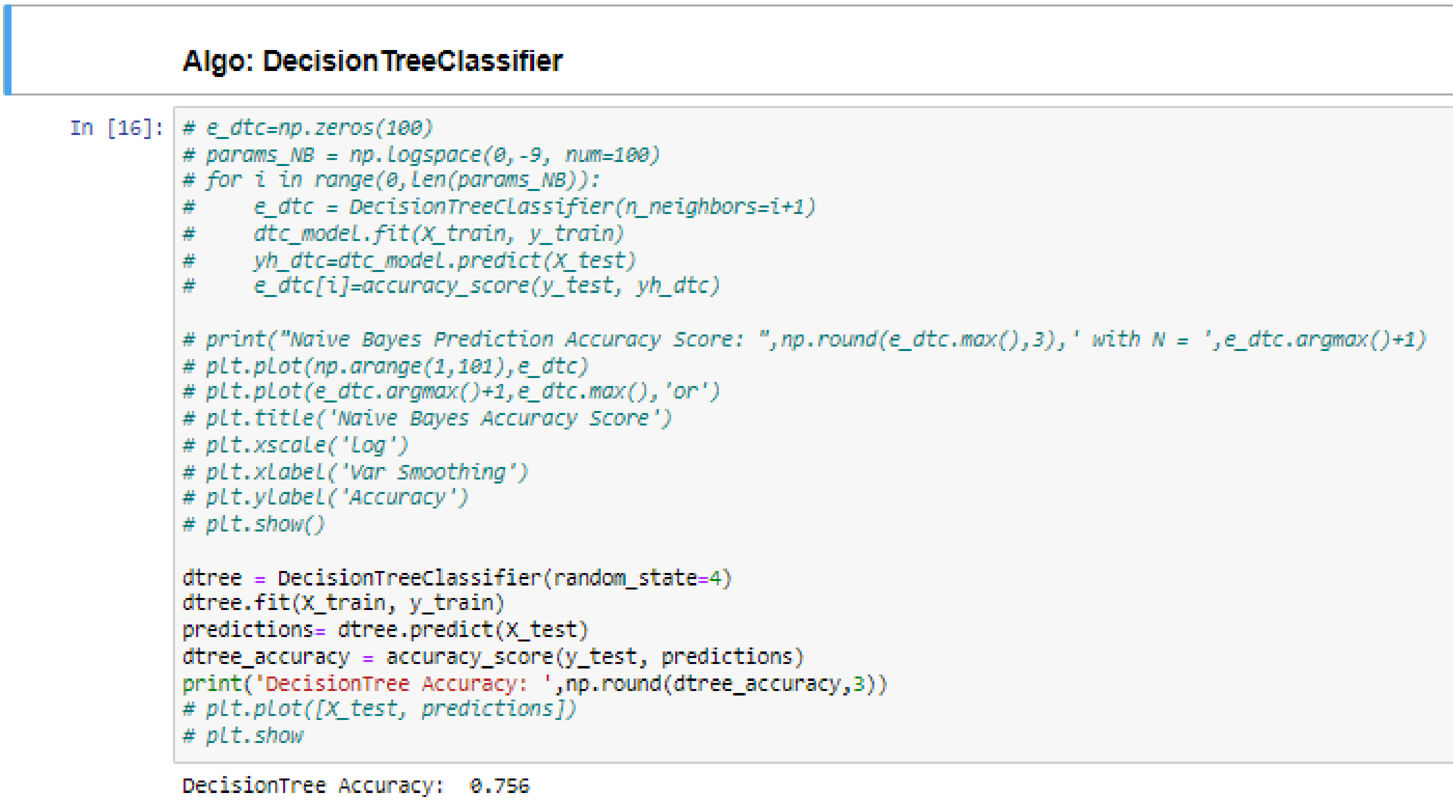
In [18]:

```
gnb=np.zeros(100)
params_NB = np.logspace(0,-9, num=100)
for i in range(0,len(params_NB)):
    gnb_model = GaussianNB(var_smoothing=params_NB[i])
    gnb_model.fit(X_train, y_train)
    yhpredict_gnb=gnb_model.predict(X_test)
    gnb[i]=accuracy_score(y_test, yhpredict_gnb)

# print("Naive Bayes Accuracy: ",np.round(gnb.max(),3),' with Smoothing = ',params_NB[gnb.argmax()])
print("Naive Bayes Accuracy: ",np.round(gnb.max(),3))
plt.plot(params_NB,gnb,'.-')
plt.plot(params_NB[gnb.argmax()],gnb.max(),'or')
plt.title('Naive Bayes Accuracy Score')
plt.xscale('log')
# plt.xlabel('Var Smoothing')
plt.ylabel('Accuracy')
plt.show()
```

```
Naive Bayes Accuracy:  0.789
```

Figure 4: GaussianNB

## 0.3 Algorithm: DecisionTree



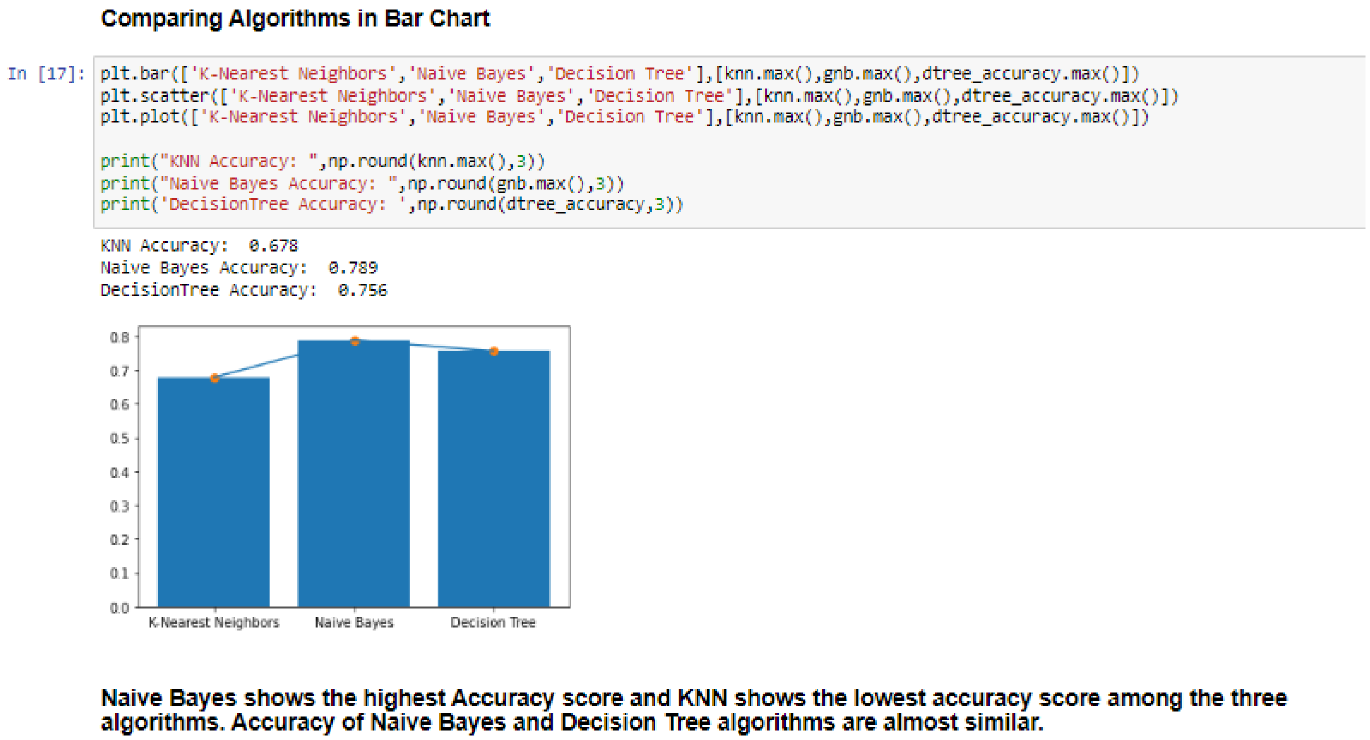
### Algo: DecisionTreeClassifier

```
In [16]: # e_dtc=np.zeros(100)
         # params_NB = np.logspace(0,-9, num=100)
         # for i in range(0,len(params_NB)):
         #     e_dtc = DecisionTreeClassifier(n_neighbors=i+1)
         #     dtc_model.fit(X_train, y_train)
         #     yh_dtc=dtc_model.predict(X_test)
         #     e_dtc[i]=accuracy_score(y_test, yh_dtc)

         # print("Naive Bayes Prediction Accuracy Score: ",np.round(e_dtc.max(),3),' with N = ',e_dtc.argmax()+1)
         # plt.plot(np.arange(1,101),e_dtc)
         # plt.plot(e_dtc.argmax()+1,e_dtc.max(),'or')
         # plt.title('Naive Bayes Accuracy Score')
         # plt.xscale('log')
         # plt.xlabel('Var Smoothing')
         # plt.ylabel('Accuracy')
         # plt.show()

         dtree = DecisionTreeClassifier(random_state=4)
         dtree.fit(X_train, y_train)
         predictions= dtree.predict(X_test)
         dtree_accuracy = accuracy_score(y_test, predictions)
         print('DecisionTree Accuracy: ',np.round(dtree_accuracy,3))
         # plt.plot([X_test, predictions])
         # plt.show
```

```
DecisionTree Accuracy:  0.756
```

Figure 5: DecisionTree

## Comparison



### Comparing Algorithms in Bar Chart

```
In [17]: plt.bar(['K-Nearest Neighbors','Naive Bayes','Decision Tree'],[knn.max(),gnb.max(),dtree_accuracy.max()])
         plt.scatter(['K-Nearest Neighbors','Naive Bayes','Decision Tree'],[knn.max(),gnb.max(),dtree_accuracy.max()])
         plt.plot(['K-Nearest Neighbors','Naive Bayes','Decision Tree'],[knn.max(),gnb.max(),dtree_accuracy.max()])

         print("KNN Accuracy: ",np.round(knn.max(),3))
         print("Naive Bayes Accuracy: ",np.round(gnb.max(),3))
         print('DecisionTree Accuracy: ',np.round(dtree_accuracy,3))
```

```
KNN Accuracy:  0.678
Naive Bayes Accuracy:  0.789
DecisionTree Accuracy:  0.756
```

**Naive Bayes shows the highest Accuracy score and KNN shows the lowest accuracy score among the three algorithms. Accuracy of Naive Bayes and Decision Tree algorithms are almost similar.**

Figure 6: Comparison

# Discussion

It might be possible to predict the survival of patients with heart failure from their serum creatinine and ejection fraction, but also that the prediction made on these two features alone can be more accurate than the predictions made on the complete dataset.
KNN Accuracy: 0.678
Naive Bayes Accuracy: 0.789
DecisionTree Accuracy: 0.756
We can see that, between three algorithms, Naive bayes shows the highest accuracy.

# Conclusion

In this project, use of machine learning in medical is shown and it has the potential to impact on clinical practice. Now it's the new supporting tool for the physicians to predict if a heart failure patient will survive or not. If a patient survive after heart failure doctors can monitor how should the live to avoid further problems with prediction of machine learning algorithm.

# References

[1] D.and Jurman Chicco. "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone". In: (). DOI: https://doi.org/10.1186/s12911-020-1023-5.