**A REVIEW OF LIVER PATIENT ANALYSIS METHOD USING MACHINE LEARNING**

**TEAM LEADER MENTOR**

ARCHANA.P DR.E.ELAKKIYA M.Sc.,M.Phil.,Ph.D.,

**TEAM MEMBERS**

ANANDHAVALLI.M

AROCKIAVINISHA.S

DESIGA SRI.S

YOGESHWARI.S

**1 INTRODUCTION**

Worldwide, liver illnesses account for almost a million fatalities. There are a number of conventional techniques for diagnosing liver disorders, but they are costly. All people who are susceptible to liver illnesses would benefit from early treatment for liver disease. Machine learning has a big impact on health care as technology in healthcare develops since it can predict illnesses at an early stage. The effectiveness of machine learning in predicting liver illness is discovered in this study. This study introduces the liver disease prediction (LDP) approach, which health professionals, stakeholders, students, and researchers can use to forecast liver illness. SVM, Naive Bayes, K-Nearest Neighbors (K-NN), Linear Discriminant Analysis (LDA), and Classification and Regression Trees (CART) are the five methods used. R and Python are used to assess accuracy and determine which classification algorithm best predicts liver illness. The autoencoder network earned 92.1% accuracy, which is over the acceptable level of accuracy and can be taken into consideration for the prediction of liver disease, whereas K-NN obtained the best accuracy from the data with 91.7%.

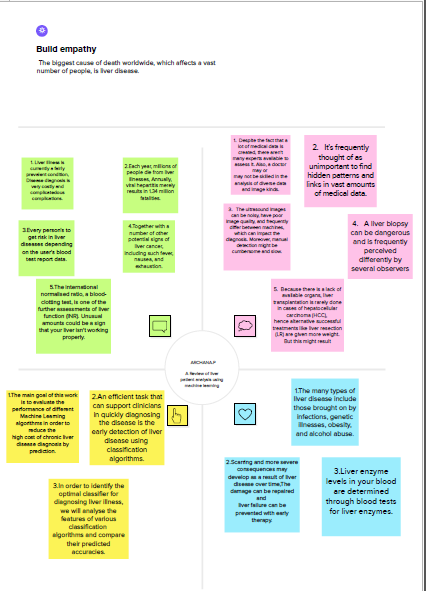
* 1. **OVERVIEW**

One of the most important organs in the human body is the liver. It is crucial to the body's ability to function. The body's toxins are eliminated, infections are fought, hormones are balanced, and bile is secreted as one of the main goals (Devikanniga et al., 2020). Many problems and liver disorders will develop if the liver's tasks are not carried out properly. Hence, if a virus damages the liver, chemicals are eaten that harm the liver, or the immune system's when dysfunction happens, the liver may suffer serious damage or malfunction, which could ultimately result in death (Nahar & Ara, 2018). One of the most dangerous and chronic diseases in the world, liver disease can have a number of negative side effects if it is not treated quickly (Dutta et al., 2022). According to a World Health Organization (WHO) report from 2018, there are around one million liver disease-related deaths worldwide, ranking liver disease as the 11th leading cause of critical mortality (World Total Deaths, n.d.). It is difficult and intimidating for medical personnel to recognise liver illness in its early stages because symptoms do not appear until the issue becomes chronic (Devikanniga et al., 2020). The standard methods of testing for liver problems, such as sonography, MRI scans, and CT scans, are also costly, dangerous, and have a number of negative effects (Joloudari et al., 2019). So, predicting liver diseases at an early stage, at low cost, and at the same time offering a better healthcare system to treat liver diseases is regarded to be a key limitation by health care professionals. Intestinal issues, dry mouth, abdominal pain, skin that becomes yellow, numbness, memory loss, and fainting issues are all symptoms of severe liver illnesses (Shaheamlung et al., 2020). These symptoms aren't noticeable in the early stages; they only become apparent after the condition becomes chronic. Yet, the liver can still function despite being partially infected (Devikanniga et al., 2020).The three phases of a liver disease's diagnosis are: liver inflammation (stage 1); liver scarring (stage 2; cirrhosis); and stage 3 (liver cancer or failure). Since these scenarios might occur with liver disease, early detection is important to ensuring New Zealanders' improved health. Early detection of liver illness increases the likelihood that it will be treated and the number of liver-related deaths will be reduced (Arbain & Balakrishnan, 2019).

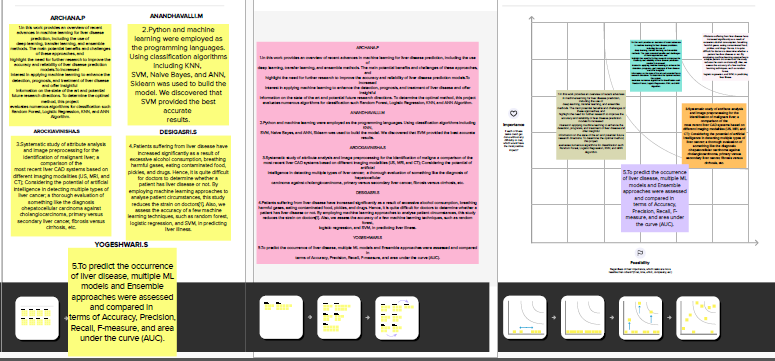
* 1. **PURPOSE**

**2. PROBLEM DEFINITION & DESIGN THINKING**

**2.1 EMPATHY MAP**

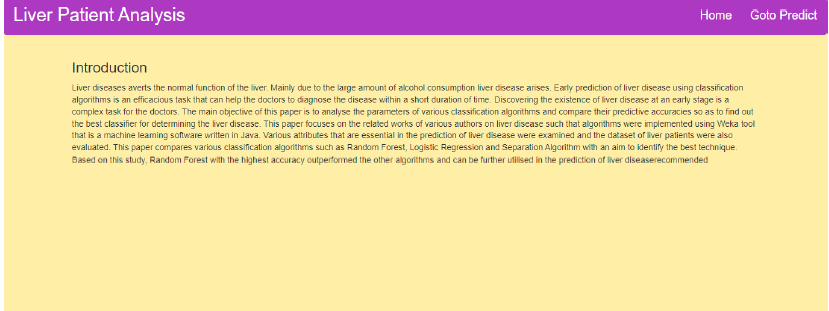
****

**2.2 IDEATION & BRAINSTORMING MAP**

****

**3. RESULT**

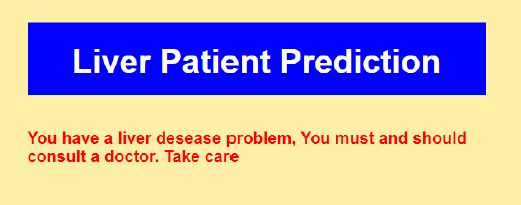
**a)**

****

**b)**

****

**c)**

****

**4. ADVANTAGES & DISADVANTAGES**

No medical knowledge is necessary: To use this programme to anticipate liver disease, you don't need any prior understanding of medical science or liver problems. You can acquire the prediction results by just entering the information that is already in the blood test report (certain details, such age and gender, are already available). High degree of accuracy: With the dataset we used to develop this application, the system predicts the outcomes with a 100% accuracy rate. Even though the accuracy might vary in some instances, it will still be sufficient to be relied upon on a wide scale. Results are projected in a matter of seconds after entering the information. In contrast to the conventional procedure, you don't have to wait for a doctor to arrive.

**5. APPLICATIONS**

The Indian Liver Patient Dataset was classified using five algorithms: Naive Bayes, Logistic Regression, Decision Tree, Random Forest, and SVM. On the test set, the models that were created for the training set were assessed. Based on prediction accuracy, it was found that SVM had a 74.09% accuracy rate.

**6. CONCLUSION**

The dataset was initially examined and prepared for feeding to the classifiers. This was accomplished by eliminating some rows with null values, altering some columns that had skewness, and converting the labels appropriately (using one-hot encoding) so that they could be used for categorization. The performance indicators that will be used to assess the models were chosen. Then, a training set and a testing set were created from the data. In order to calculate the benchmark value of accuracy, a naive predictor and a benchmark model ('Logistic Regression') were performed on the dataset. The selection of the training algorithms and the appropriate parameters for fine-tuning presented the most challenges in carrying out this project. I initially found it quite difficult to select 3 or 4 strategies from the many accessible alternatives in sklearn. By this experiment, I came to understand the value and importance of parameter tuning as a crucial component of machine learning. If we are prepared to devote more time and computer power, I believe this sector can benefit from additional advancement.

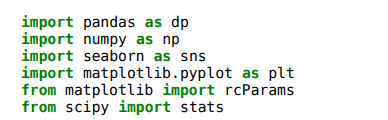
**7.FUTURE SCOPE**

The ILPD (Indian Liver Patient Dataset) from the UCI Machine Learning Repository was used as the dataset for this issue. This dataset refers to the 2017 Andhra Pradesh region.We cannot rely on this model to forecast accuracy for a large dataset because our dataset is tiny and its training dataset is identical to its test dataset. To classify values and obtain the best accuracy, we require a more accurate data set with greater values and features.

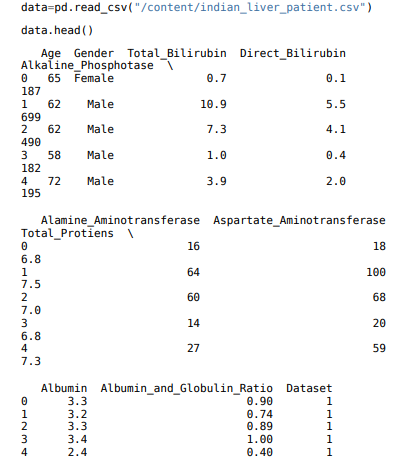
**8. APPENDIX**

**A.SOURCE CODE**

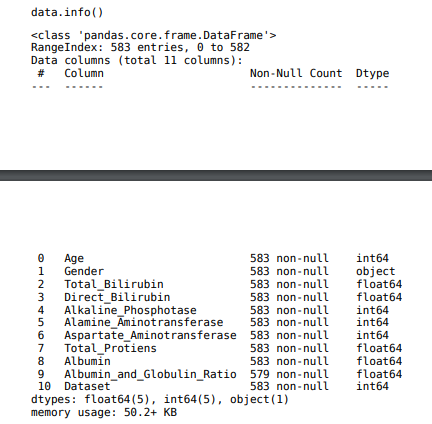
**Importing libraries**

****

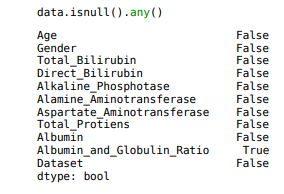
**Loading the dataset**

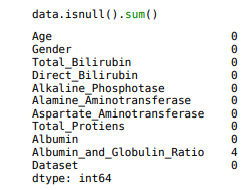
****

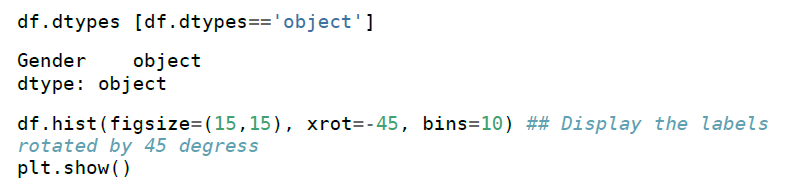
**Summary of the dataset**

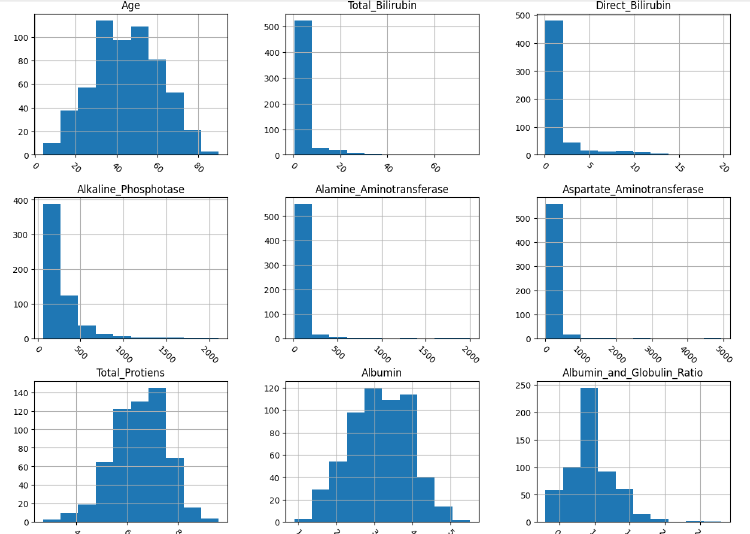
****

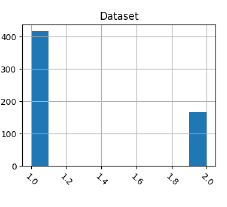
**Sum of null values**

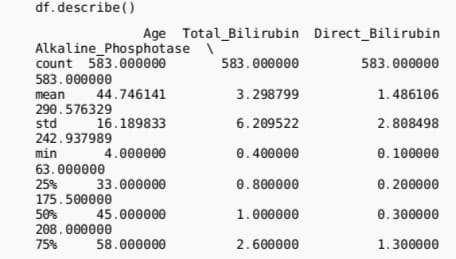
****

****

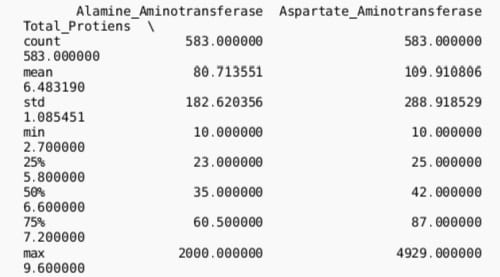
****

****

****

****

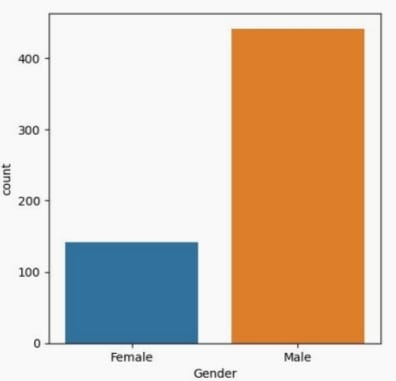
****

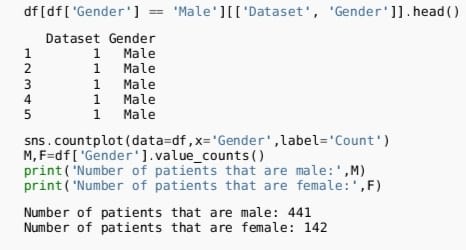
****

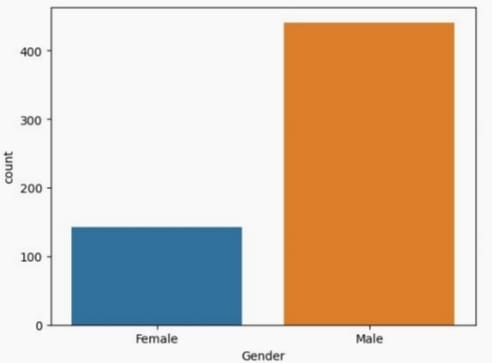
****

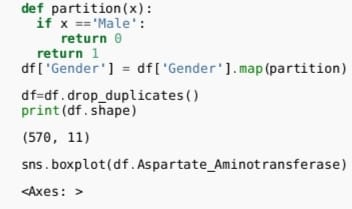
**Data Visualization**

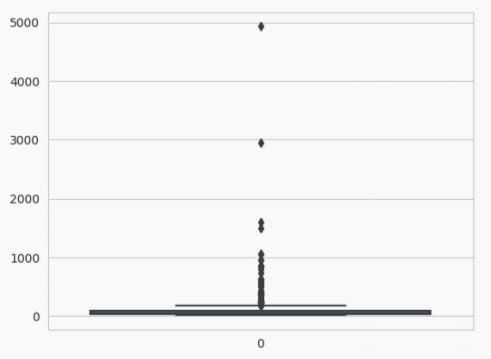
****

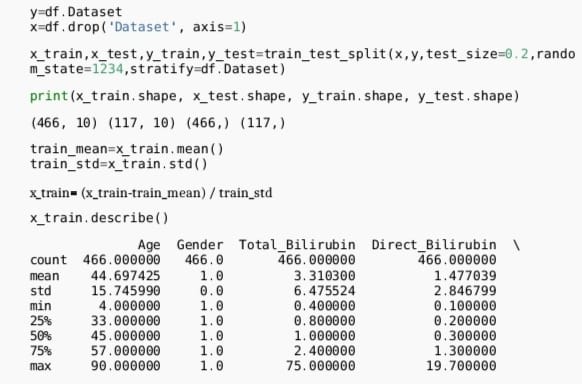
****

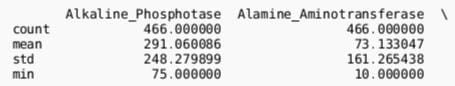
****

****

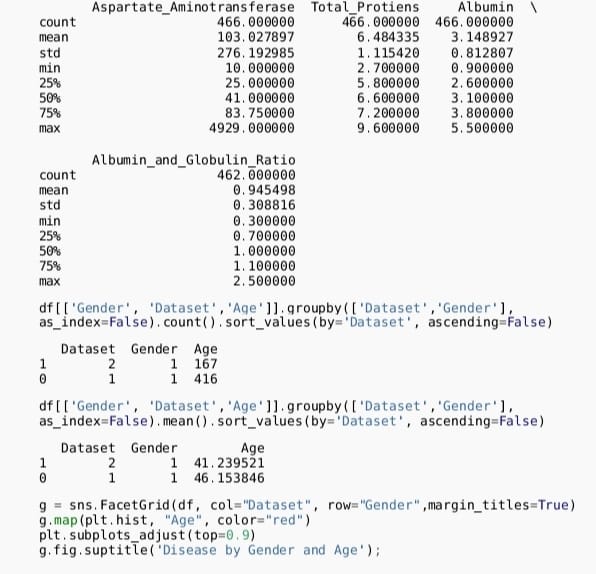
****

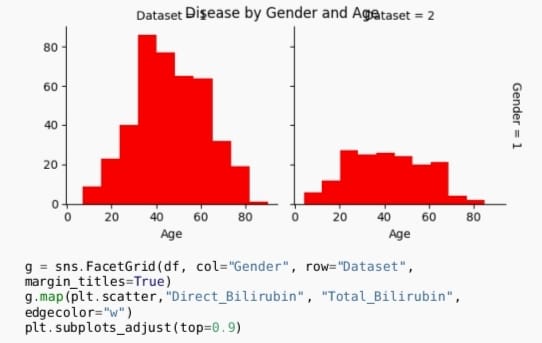
****

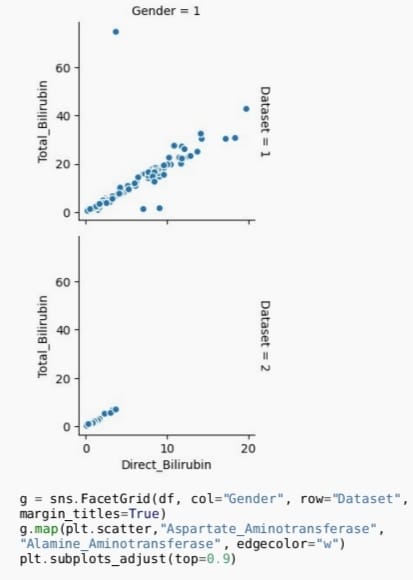
****

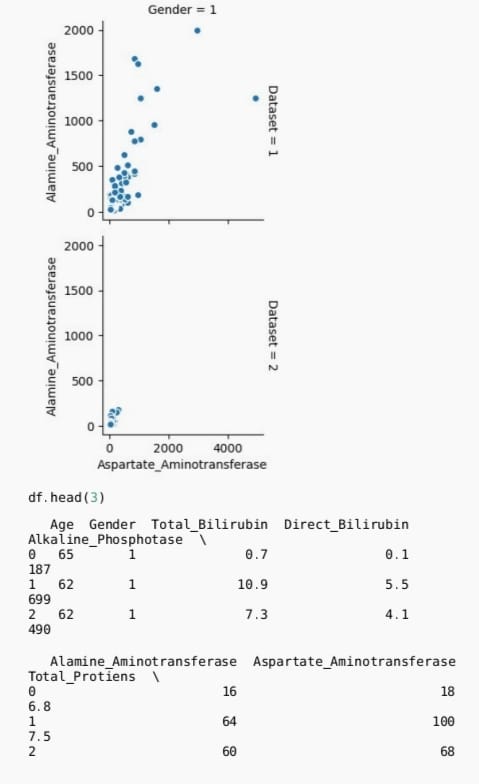
****

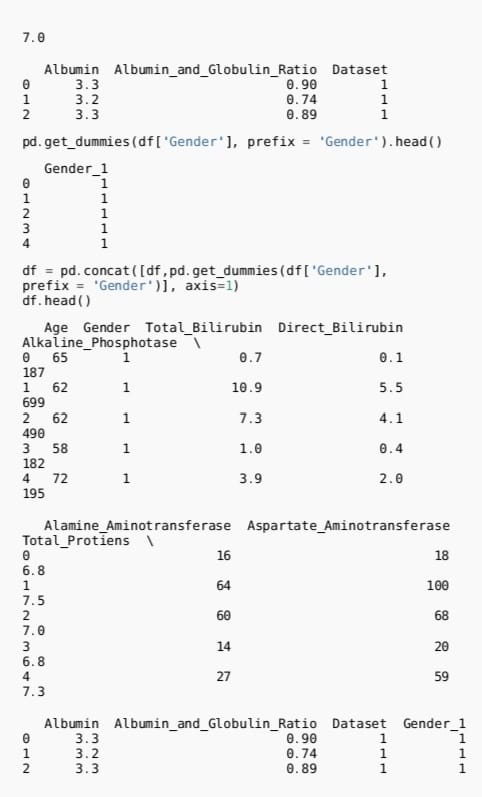
****

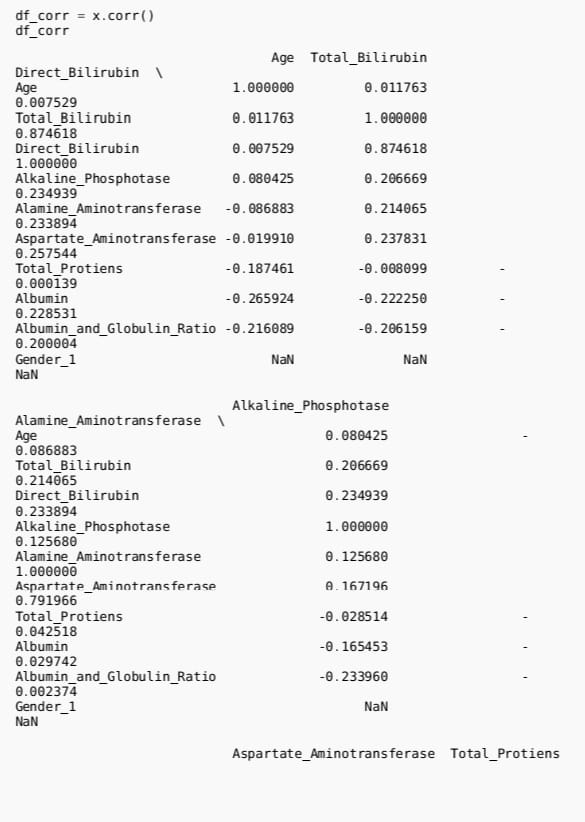
****

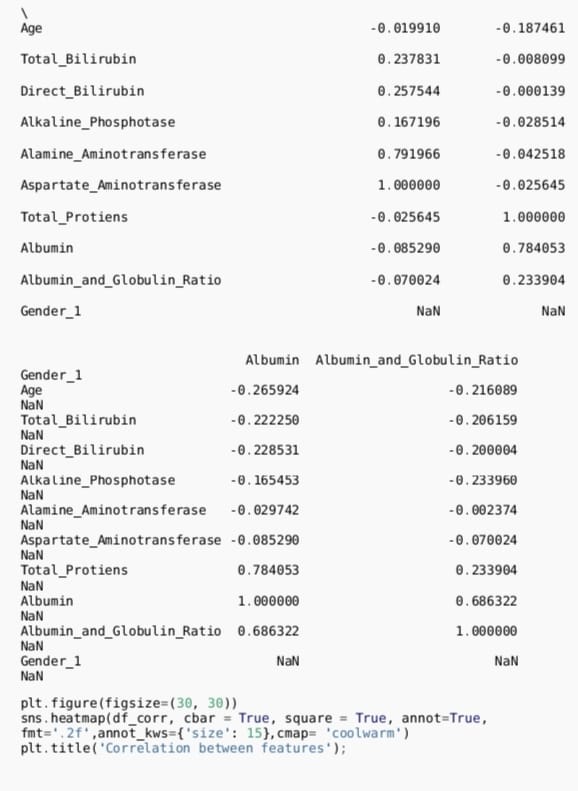
****

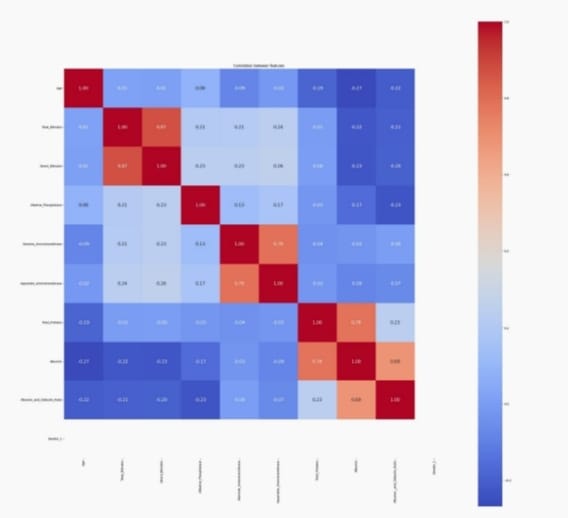
****

****

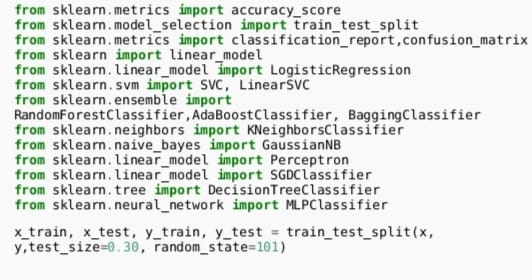
****

****

****

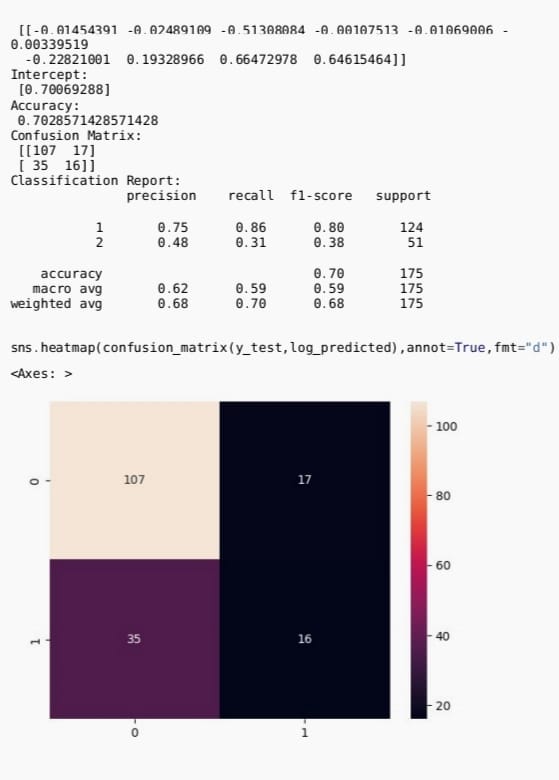
****

**Machine Learning**

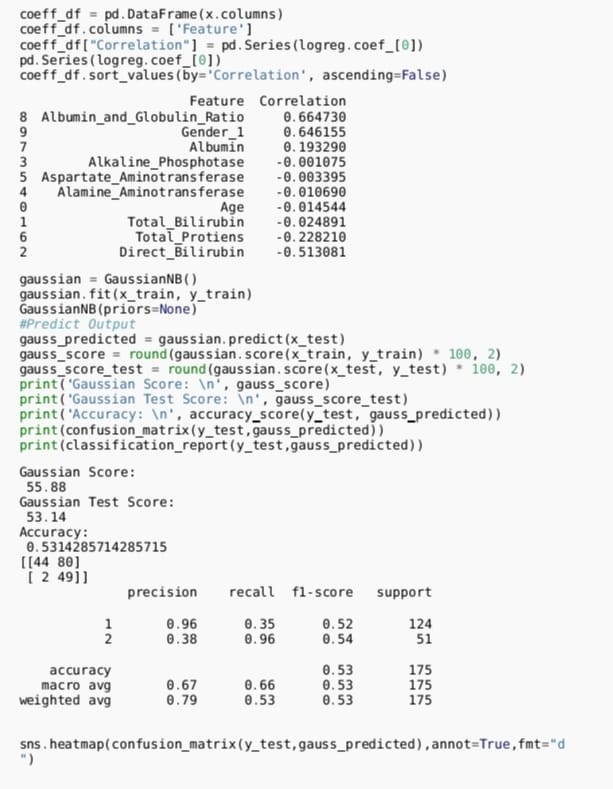
****

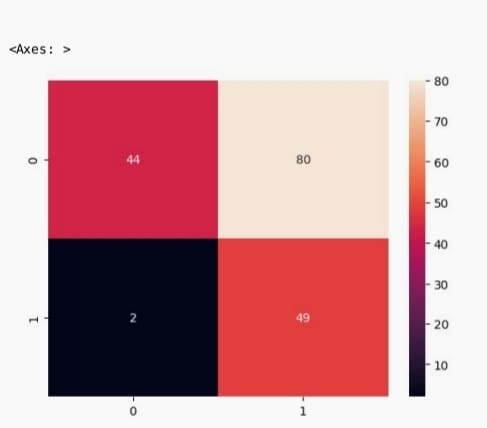
**Logistic Regression**

****

****

**Gaussian Naive Bayes**

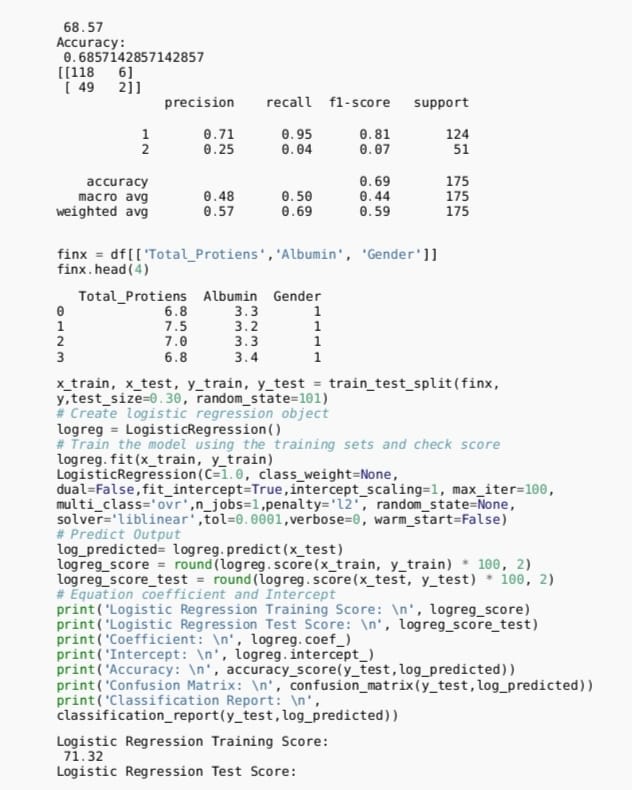
****

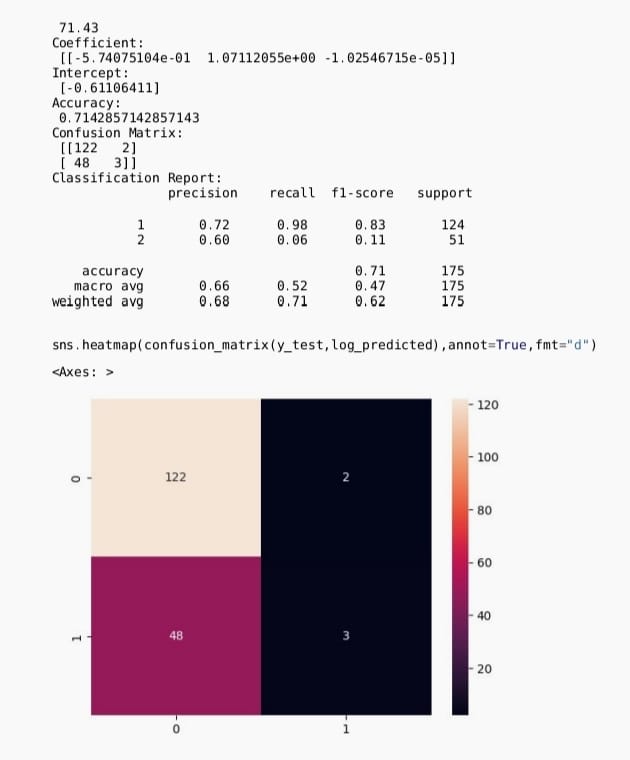
****

**Random Forest**

****

**Logistic Regression**

****

****

****