**CIS6005 Computational Intelligence Assignment**

Student Name – P.A Malsha Madhurangi

Student ICBT Number – CL/BSCSD/25/89

****

**Table of content**

Mini project ----------------------------------------------------------------------------------------Page 9

1. Introduction to the concept of deep learning ---------------------------------------- Page 9
2. A literature review ---------------------------------------------------------------------- Page 12
3. Exploratory data analysis (EDA)------------------------------------------------------Page 14
4. System architecture ---------------------------------------------------------------------Page 28

How your application differs from other existing applications--------------------Page 28

Machine learning technique------------------------------------------------------------Page 29

1. Full model evaluation-------------------------------------------------------------------Page 30

Implementation details------------------------------------------------------------------Page 31

1. Conclusion of the final model----------------------------------------------------------Page 59

The success of using deep learning techniques in the Predict Health outcomes of horses -------------------------------------------------------------------------------------Page 59

Appendix -------------------------------------------------------------------------------- Page 62

Software Artifact ----------------------------------------------------------------------- Page 63

Reference ------------------------------------------------------------------------------- Page 69

**Table of Figures**

Figure 1---------------------------------------------------------------------------------------------Page 10

Figure 2---------------------------------------------------------------------------------------------Page 14

Figure 3---------------------------------------------------------------------------------------------Page 14

Figure 4---------------------------------------------------------------------------------------------Page 15

Figure 5---------------------------------------------------------------------------------------------Page 15

Figure 6---------------------------------------------------------------------------------------------Page 15

Figure 7---------------------------------------------------------------------------------------------Page 15

Figure 8---------------------------------------------------------------------------------------------Page 16

Figure 9---------------------------------------------------------------------------------------------Page 16

Figure 10-------------------------------------------------------------------------------------------Page 16

Figure 11-------------------------------------------------------------------------------------------Page 17

Figure 12-------------------------------------------------------------------------------------------Page 17

Figure 13-------------------------------------------------------------------------------------------Page 17

Figure 14-------------------------------------------------------------------------------------------Page 17

Figure 15-------------------------------------------------------------------------------------------Page 18

Figure 16-------------------------------------------------------------------------------------------Page 18

Figure 17-------------------------------------------------------------------------------------------Page 18

Figure 18-------------------------------------------------------------------------------------------Page 19

Figure 19-------------------------------------------------------------------------------------------Page 20

Figure 20-------------------------------------------------------------------------------------------Page 20

Figure 21-------------------------------------------------------------------------------------------Page 20

Figure 22-------------------------------------------------------------------------------------------Page 20

Figure 23-------------------------------------------------------------------------------------------Page 21

Figure 24-------------------------------------------------------------------------------------------Page 21

Figure 25-------------------------------------------------------------------------------------------Page 21

Figure 26-------------------------------------------------------------------------------------------Page 21

Figure 27-------------------------------------------------------------------------------------------Page 21

Figure 28-------------------------------------------------------------------------------------------Page 21

Figure 29-------------------------------------------------------------------------------------------Page 22

Figure 30-------------------------------------------------------------------------------------------Page 22

Figure 31-------------------------------------------------------------------------------------------Page 22

Figure 32-------------------------------------------------------------------------------------------Page 22

Figure 33-------------------------------------------------------------------------------------------Page 23

Figure 34-------------------------------------------------------------------------------------------Page 23

Figure 35-------------------------------------------------------------------------------------------Page 23

Figure 36-------------------------------------------------------------------------------------------Page 23

Figure 37-------------------------------------------------------------------------------------------Page 23

Figure 38-------------------------------------------------------------------------------------------Page 23

Figure 39-------------------------------------------------------------------------------------------Page 24

Figure 40-------------------------------------------------------------------------------------------Page 24

Figure 41-------------------------------------------------------------------------------------------Page 25

Figure 42-------------------------------------------------------------------------------------------Page 26

Figure 43-------------------------------------------------------------------------------------------Page 26

Figure 44-------------------------------------------------------------------------------------------Page 26

Figure 45-------------------------------------------------------------------------------------------Page 26

Figure 46-------------------------------------------------------------------------------------------Page 26

Figure 47-------------------------------------------------------------------------------------------Page 26

Figure 48-------------------------------------------------------------------------------------------Page 26

Figure 49-------------------------------------------------------------------------------------------Page 26

Figure 50-------------------------------------------------------------------------------------------Page 27

Figure 51-------------------------------------------------------------------------------------------Page 27

Figure 52-------------------------------------------------------------------------------------------Page 30

Figure 53-------------------------------------------------------------------------------------------Page 30

Figure 54-------------------------------------------------------------------------------------------Page 30

Figure 55-------------------------------------------------------------------------------------------Page 31

Figure 56-------------------------------------------------------------------------------------------Page 32

Figure 57-------------------------------------------------------------------------------------------Page 32

Figure 58-------------------------------------------------------------------------------------------Page 33

Figure 59-------------------------------------------------------------------------------------------Page 34

Figure 60-------------------------------------------------------------------------------------------Page 34

Figure 61-------------------------------------------------------------------------------------------Page 35

Figure 62-------------------------------------------------------------------------------------------Page 36

Figure 63-------------------------------------------------------------------------------------------Page 36

Figure 64-------------------------------------------------------------------------------------------Page 36

Figure 65-------------------------------------------------------------------------------------------Page 37

Figure 66-------------------------------------------------------------------------------------------Page 37

Figure 67-------------------------------------------------------------------------------------------Page 38

Figure 68-------------------------------------------------------------------------------------------Page 38

Figure 69-------------------------------------------------------------------------------------------Page 38

Figure 70-------------------------------------------------------------------------------------------Page 39

Figure 71-------------------------------------------------------------------------------------------Page 39

Figure 72-------------------------------------------------------------------------------------------Page 40

Figure 73-------------------------------------------------------------------------------------------Page 41

Figure 74-------------------------------------------------------------------------------------------Page 41

Figure 75-------------------------------------------------------------------------------------------Page 42

Figure 76-------------------------------------------------------------------------------------------Page 42

Figure 77-------------------------------------------------------------------------------------------Page 43

Figure 78-------------------------------------------------------------------------------------------Page 43

Figure 79-------------------------------------------------------------------------------------------Page 44

Figure 80-------------------------------------------------------------------------------------------Page 44

Figure 81-------------------------------------------------------------------------------------------Page 44

Figure 82-------------------------------------------------------------------------------------------Page 45

Figure 83-------------------------------------------------------------------------------------------Page 46

Figure 84-------------------------------------------------------------------------------------------Page 46

Figure 85-------------------------------------------------------------------------------------------Page 47

Figure 86-------------------------------------------------------------------------------------------Page 47

Figure 87-------------------------------------------------------------------------------------------Page 48

Figure 88-------------------------------------------------------------------------------------------Page 48

Figure 89-------------------------------------------------------------------------------------------Page 49

Figure 90-------------------------------------------------------------------------------------------Page 49

Figure 91-------------------------------------------------------------------------------------------Page 50

Figure 92-------------------------------------------------------------------------------------------Page 50

Figure 93-------------------------------------------------------------------------------------------Page 50

Figure 94-------------------------------------------------------------------------------------------Page 50

Figure 95-------------------------------------------------------------------------------------------Page 51

Figure 96-------------------------------------------------------------------------------------------Page 52

Figure 97-------------------------------------------------------------------------------------------Page 52

Figure 98-------------------------------------------------------------------------------------------Page 52

Figure 99-------------------------------------------------------------------------------------------Page 53

Figure 100------------------------------------------------------------------------------------------Page 53

Figure 101------------------------------------------------------------------------------------------Page 53

Figure 102------------------------------------------------------------------------------------------Page 54

Figure 103------------------------------------------------------------------------------------------Page 54

Figure 104------------------------------------------------------------------------------------------Page 54

Figure 105------------------------------------------------------------------------------------------Page 54

Figure 106-----------------------------------------------------------------------------------------Page 54

Figure 107-----------------------------------------------------------------------------------------Page 55

Figure 108-----------------------------------------------------------------------------------------Page 55

Figure 109-----------------------------------------------------------------------------------------Page 56

Figure 110-----------------------------------------------------------------------------------------Page 56

Figure 111-----------------------------------------------------------------------------------------Page 57

Figure 112-----------------------------------------------------------------------------------------Page 57

Figure 113-----------------------------------------------------------------------------------------Page 58

Figure 114-----------------------------------------------------------------------------------------Page 60

Figure 115-----------------------------------------------------------------------------------------Page 61

Figure 116-----------------------------------------------------------------------------------------Page 62

Figure 117-----------------------------------------------------------------------------------------Page 62

Figure 118-----------------------------------------------------------------------------------------Page 62

Figure 119-----------------------------------------------------------------------------------------Page 63

Figure 120-----------------------------------------------------------------------------------------Page 64

Figure 121-----------------------------------------------------------------------------------------Page 64

Figure 122-----------------------------------------------------------------------------------------Page 65

Figure 123-----------------------------------------------------------------------------------------Page 65

Figure 124-----------------------------------------------------------------------------------------Page 66

Figure 125-----------------------------------------------------------------------------------------Page 66

Figure 126-----------------------------------------------------------------------------------------Page 66

Figure 127-----------------------------------------------------------------------------------------Page 67

Figure 128-----------------------------------------------------------------------------------------Page 67

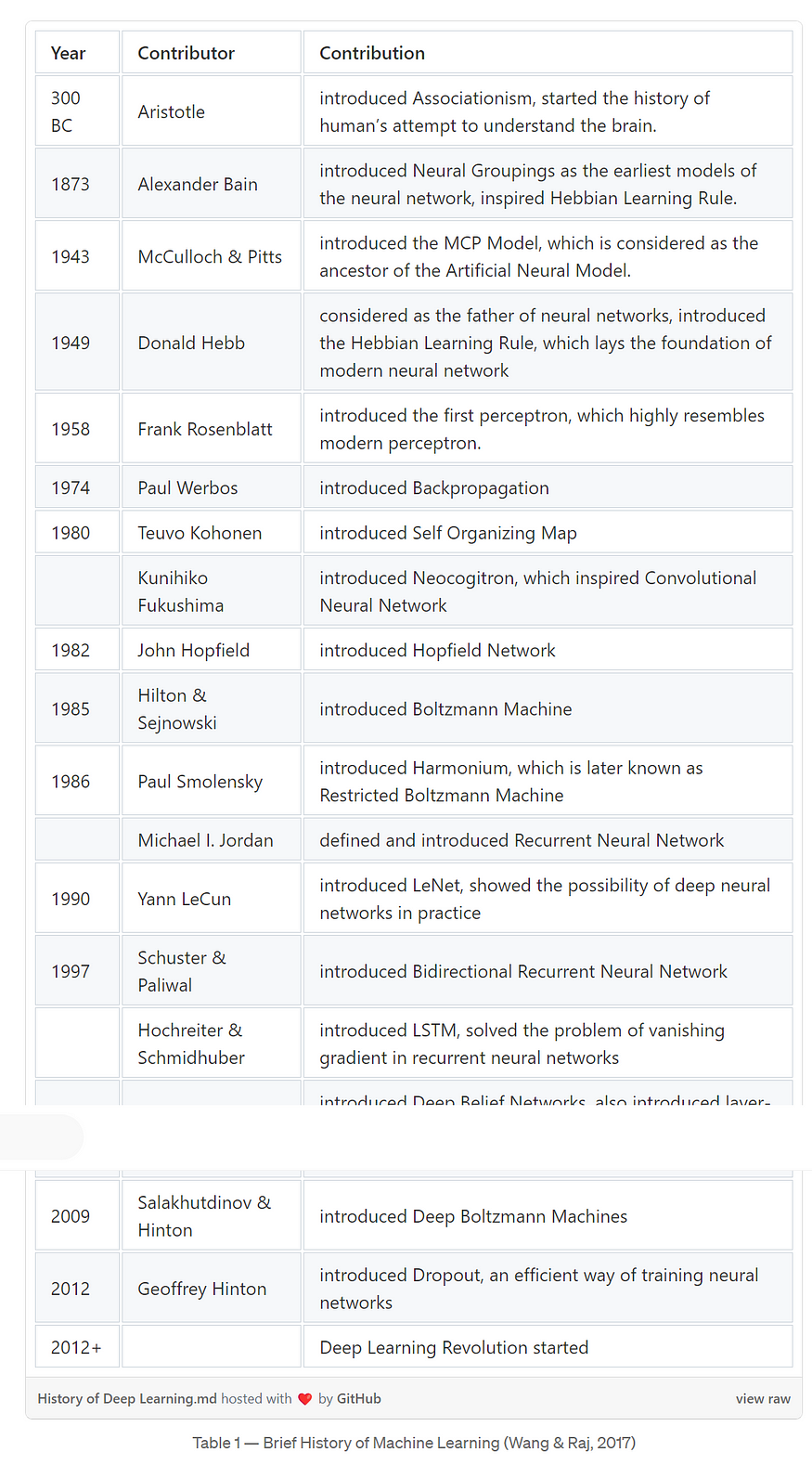
Figure 129-----------------------------------------------------------------------------------------Page 68

**Mini Project**

1. **Introduction to the concept of deep learning**

Deep learning is a method of artificial intelligence that usually refers to the way the human brain stores and stores data of a person. Just as the human brain contains millions of neurons to receive information, deep learning networks, artificial networks, and many artificial neurons work together in the computer. The deep learning models here can generate predictions, recognize complex patterns in images, sounds, and data. Deep learning methods can be used to automatically perform tasks that require human intervention, such as transcribing a few sounds, describing images. Deep learning models are defined as computer files that scientists have trained to perform an algorithm or specific task. Deep learning models are mainly used for data analysis. And a deep learning network can be divided into three parts as input layer, hidden layer and output layer. The input layer is formed using several nodes containing the data and the output layer is based on the two outputs 'yes' and 'no'. At the hidden layer, information is processed at different levels and behavior is adapted. For example, when a human is given an image of an unknown animal that needs to be classified, it compares the size, number of legs, and eye position of the unknown animal with a known animal. If the animal has cat eyes, the man guesses that it is a wild cat. In the same way, a deep learning algorithm while classifying an animal image considers each hidden layer as a different feature of the animal and tries to classify it correctly. Deep learning algorithms make machine learning methods more efficient. This is called supervised learning and it requires a large enough data set with widely varying data to increase the accuracy of the results. For example, if a training set for cat color discrimination contains more images of black cats, black cats will be correctly identified. But white cats are not recognized and thus white cat images need to be trained again in a machine learning model to recognize white cats. (Amazon Web Services, Inc., 2024)

Considering the history of deep learning, the concept of artificial neurons was proposed by Warren McCulloch and Walter Pitts in the 1940s. It is the foundation of deep learning. Here, a mathematical model based on neuron activity was created. Mcp-neuron is called In 1957, Frank Rosenblatt introduced the perceptron. In the late 1960s, deep learning was suppressed as there was little interest in binary neurons. Again, in the year 1986, with the introduction of the backpropagation algorithm by Geoffrey Hinton, David Rumelhart and Ronald Williams, deep learning made its way back to prominence. Backpropagation is known as the father of deep learning. LeCun improved the LeNet-5 architecture in the late 1990s. In 1999, Nvidia created the world's first GPU. In early 2006, deep learning models were developed on GPUs, and deep learning was used in speech recognition in the early 2010s. And in the year 2013, neural networks started to be used in computers. Thus, the demand for deep learning has increased and it has become an indispensable technology nowadays. (Sundaray, n.d.)



*Figure 1*

Deep learning architecture is also used to solve complex tasks. Discusses 6 architectural techniques used in everyday life. Namely RNN, LSTM, GRU, CNN, DBN and DSN. RNN or Recurrent Neural Network Speech Synthesis helps to maintain memory of past input, solve problems. Also, two types of RNNs can be identified as bidirectional RNNs and deep RNNs. In bidirectional RNNs, past and future information can be obtained simultaneously. And in deep RNNs, more hierarchies can be added to the DL model. LSTM or Long Short Term Memory is a type of RNN and text compression is used for tasks such as speech recognition, handwriting recognition, gesture recognition and video captioning. GRU is a type of LSTM and is used for small and sparse data sets. They show good performance. CNN or Convolutional Neural Networks are used for video recognition, video processing, video analysis and NLP. , DBN or Deep Belief Network is a multi-layer network and is used in pattern recognition and NLP. In DSN or Deep Stacking Network, a group of single deep networks contributes to improving the training problem and provides the ability to learn complex functions and classifications. Here, simple architectures such as autoencoders can be identified, which help to reduce dimensionality and naturally detect anomalies. Finally deep learning architecture is the same idea but there are different methods. Nowadays, with the development of technology, deep learning for various fields, artificial intelligence (Artificial intelligence) makes learning tasks more efficient and easier. (Lisowski, n.d.)

1. **A Literature review**

Using Artificial Intelligence to Predict Probability of Survival and Need for Surgery in Horses with Acute Colic

The purpose of this study is to explore the use of machine learning algorithms in the process of predicting the need for surgery in horses with acute colic, using the basic chemical data obtained. This study was written by Mohammad A. Frywan A. and Samih M. By Abutarbush b. Acute colic is one of the conditions most horses face. Here the effects of gastrointestinal colic run the gamut from harmless spasmodic colic to the risk of strangulation. Abdominal pain in horses can be caused by factors such as deep ulceration, ischemia, intestinal distention, and stress on the cerebral root. However, in medicine, five main criteria are used to decide whether to undergo surgery. The five criteria were findings on transrectal palpation, presence of nasogastric reflux, cardiovascular and systemic status, Pain severity (as measured by non-responsiveness to analgesics) and abdominal outcomes can be indicated. A small percentage of horses typically require surgery, 4%-10%, and rarely the final diagnosis is made without laparotomy.

With the new development of technology in recent years, with the expansion of subjects like artificial intelligence and machine learning, image analysis, processing data sets, identifying various features, helping to make accurate and quick decisions, and identifying obscure and hidden relationships for humans provide many advantages.

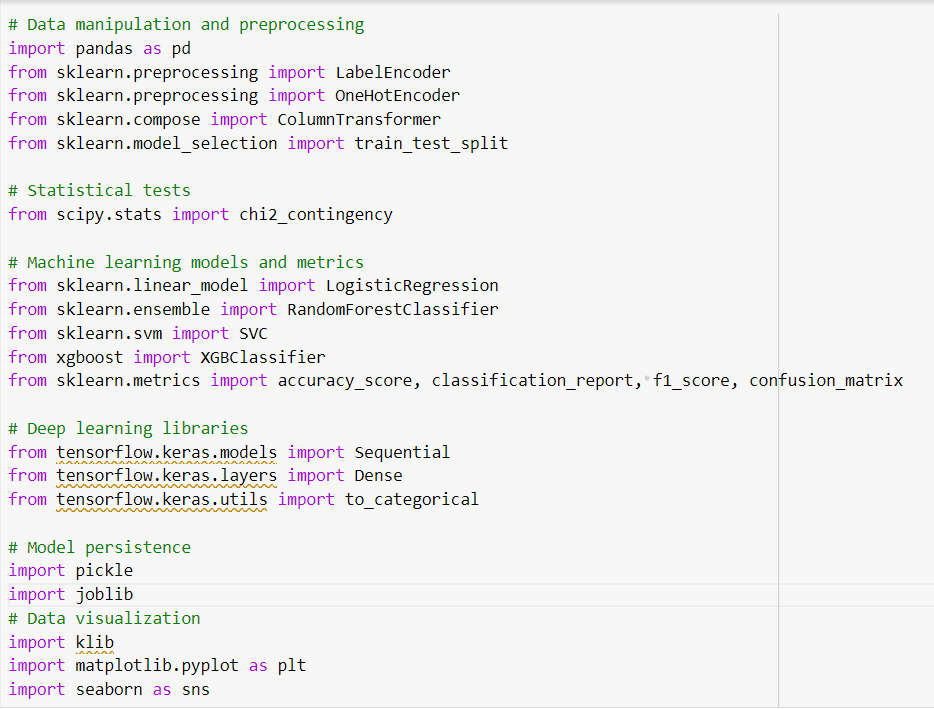
Machine learning, a subset of Ai, synthesizes mathematical models using relevant training data. The training data obtained as these samples act as the trigger for setting the model parameters. Also, Ai can provide very quick and accurate responses through image inspection. An example is veterinary robot Sofie. Ai provides advantages such as face recognition, reproduction, and quick decisions, and for example, it was able to achieve a high accuracy of 98.7% by using a deep learning method, Convolutional neural networks, to count reticulocytes in cat blood images. This kind of process saves time and money in the absence of expensive equipment. Valletta et al. also reviewed machine learning algorithms in animal behavior studies, where 3 study results are discussed. The three outcomes are identifying and counting wildlife, discovering jack social structures, and providing quail eggs from mixed clutches to the egg-laying female. Accelerometric data has been used for this study. Also, elasmobranch studies of juvenile lemon shark behavior in the Bahamas reported a high F-measure of 88%, providing good empirical evidence of shark behavior and explaining scientific observations.

The primary objective of this study was to explore the ability of the ML algorithm to predict the need for surgery and probability of survival in horses based on baseline clinical data. Survival here refers to a horse recovering and being discharged from the hospital. This provides more data on disease progression than clinical data predicts. Also, machine learning algorithms were able to predict the need for surgery and the probability of survival in acute colic disease in horses with 76% and 85% accuracy. Also, the first step in the system architecture is to implement an automated computer-based or smart-mobile tablet-based system application that can be easily used by clinicians to facilitate decision-making. By designing such an application, appropriate interfaces can be created to process data, provide appropriate outputs during prediction, and have multiple capabilities. Here, the results obtained using 4 algorithms using 285 clinical reports are summarized.

Finally, a model based on an Ai model was developed, predicting the clinical characteristics of horses using an ML algorithm. It uses Ai technology and ML algorithms to predict the probability of survival as well as the need for surgery in horses suffering from epigastric disease. Adaptation to a user-friendly interface is an important step as ML algorithms are optimized accordingly as datasets continue to expand. (Journal of Equine Veterinary Science, n.d.)

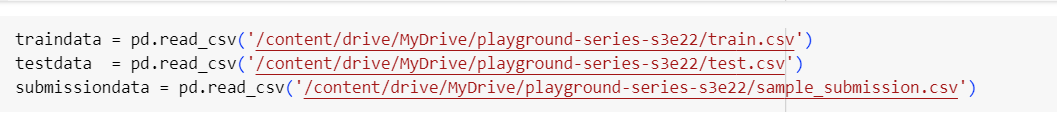
1. **Exploratory data analysis (EDA)**

Below are the imported libraries to perform all the tools and functions required to perform this prediction.



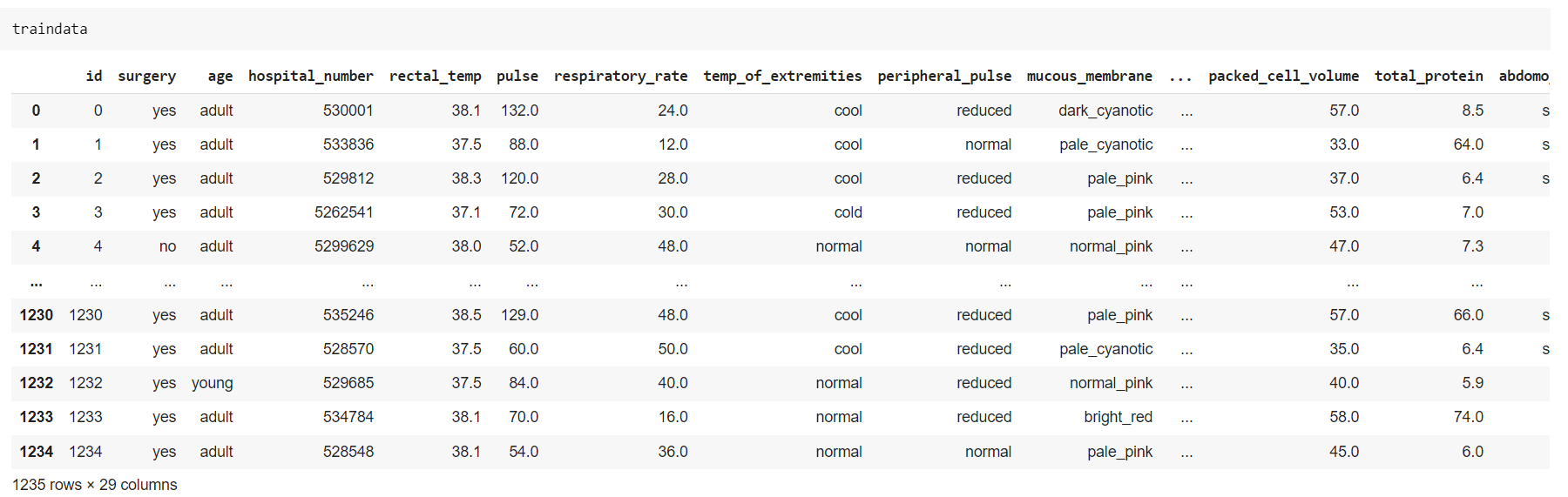
*Figure 2*

Then train data set, test data set and sample submission data set are entered into the project to start the prediction process. Here "traindata" refers to the “traindata set” used in training the model and "testdata" refers to the test data set. "Submissiondata" means the data set representing the sample submission.

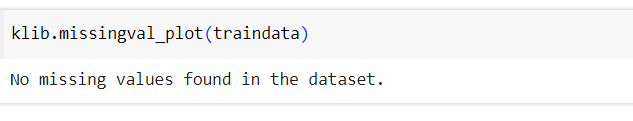


*Figure 3*

Here, an identification is made if there is a missing value in the traindata data set. It is observed that there is no missing value.



*Figure 4*



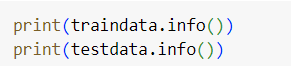
*Figure 5*

Here the unnecessary columns in the data set are removed. Here the id and hospital number columns are removed.

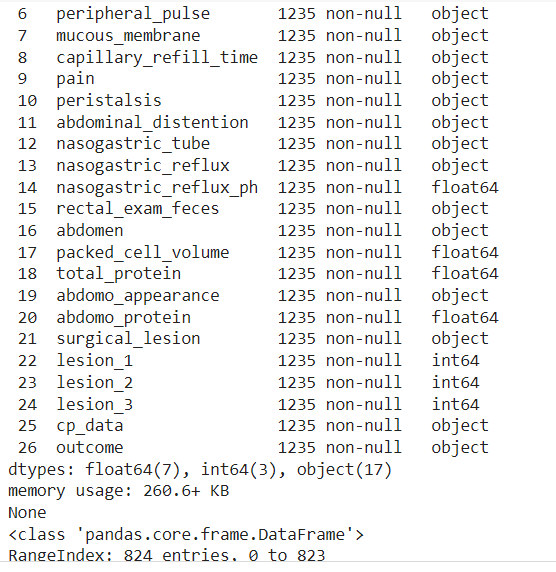
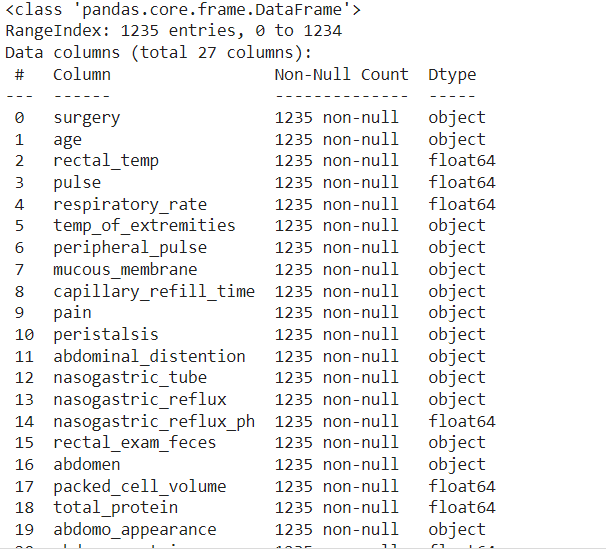


*Figure 6*

Related to viewing the structure of each variable in the dataset and its data types. This is done using the info() function and this code is used for both train and test datasets.



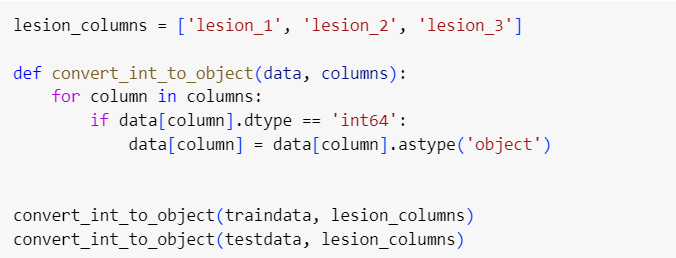
*Figure 7*



*Figure 8*

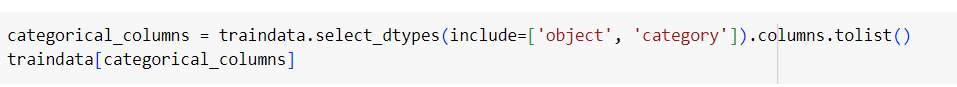
Here, lesion\_1, lesion\_2, lesion\_3 columns are considered.

These are categorical variables. Hence conversion to object or category type variables is required.



*Figure 9*

Here Categorical columns like Object, category are stored in a new data frame.

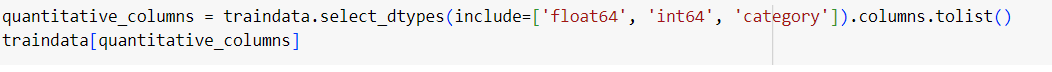


*Figure 10*

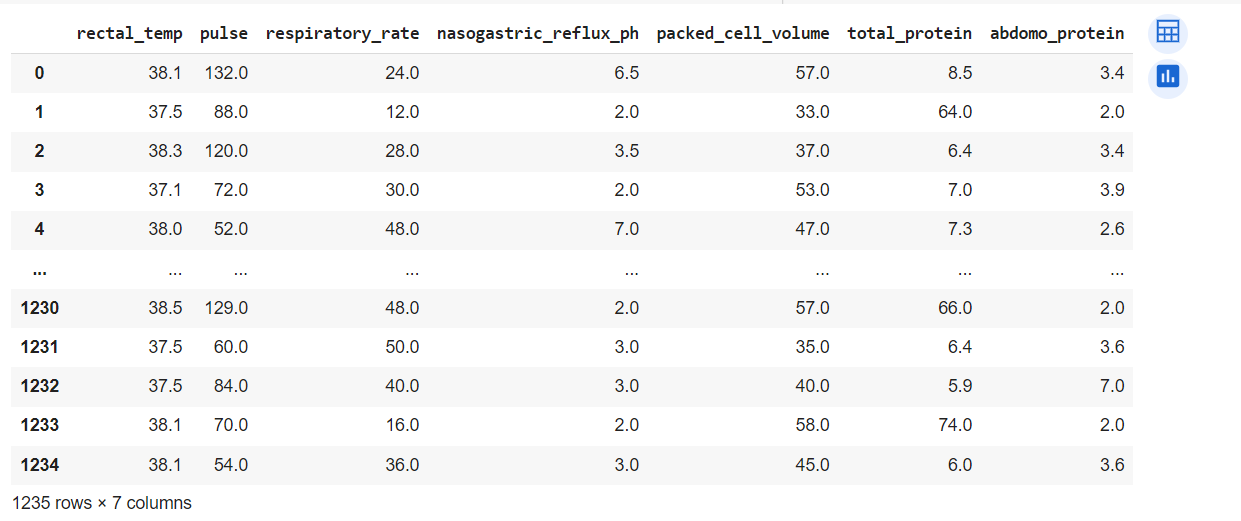


*Figure 11*

Here quantitative columns like float64, int64, category are stored in a new data frame.

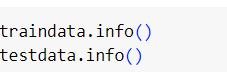


*Figure 12*

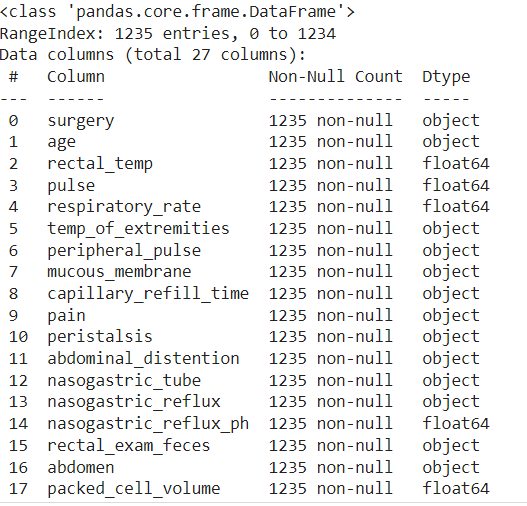
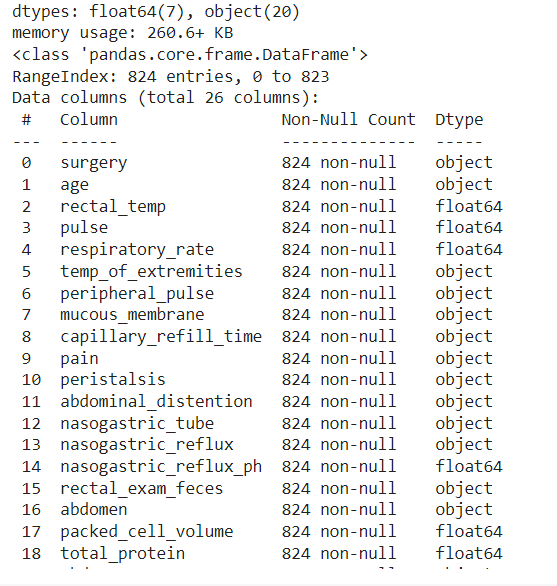


*Figure 13*

Then the train data and test data sets check the data type and structure with info().



*Figure 14*

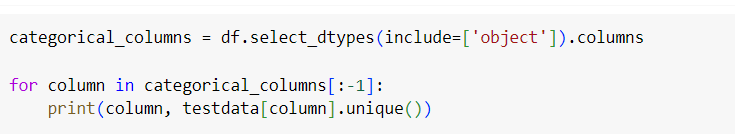
*Figure 15*

Copies data stored in traindata to df.



*Figure 16*

df is used to distribute the classification values of the data set and iterate over the training data and print the unique values in the corresponding columns of the test data.



*Figure 17*



*Figure 18*

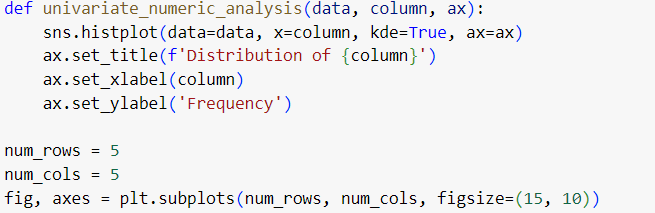
**Univariate analysis**

This field defines a function of categorical variables to perform univariate analysis.



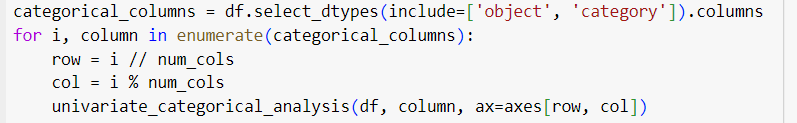
*Figure 19*

This code is used to define a function of numerical variables to perform a univariate analysis.



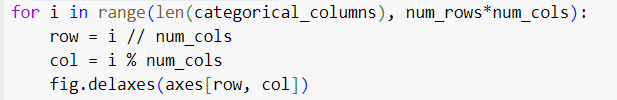
*Figure 20*

Univariate analysis is used to perform categorical variables.



*Figure 21*

This code removes unused subparts.



*Figure 22*

This code is used to set the layout.



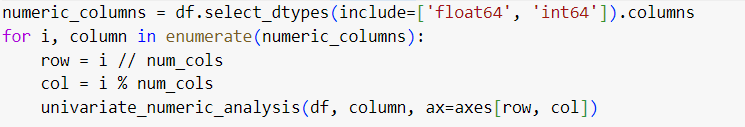
*Figure 23*

This sets up a subplot grid for numerical variables.



*Figure 24*

Here univariate analysis is done for numerical variables.



*Figure 25*

This code removes unused subparts.

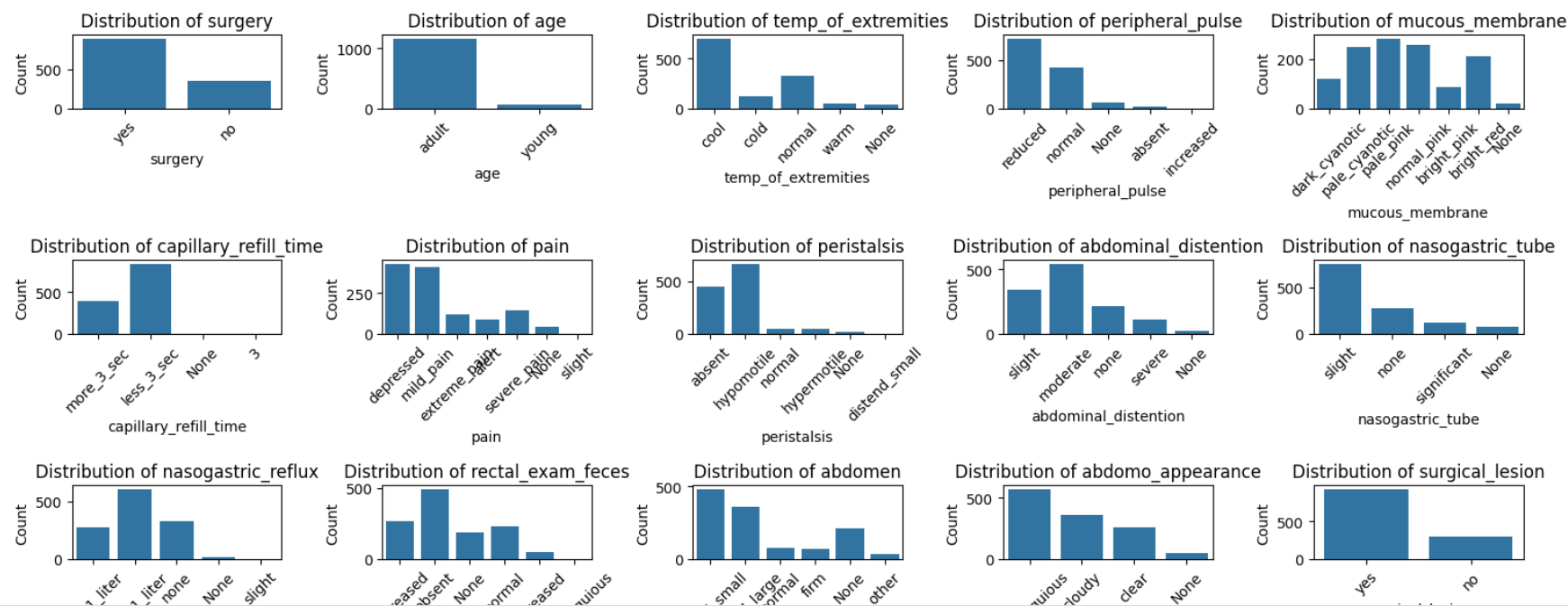


*Figure 26*

The layout is adjusted.

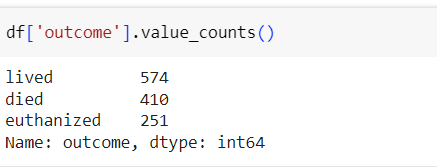


*Figure 27*



*Figure 28*

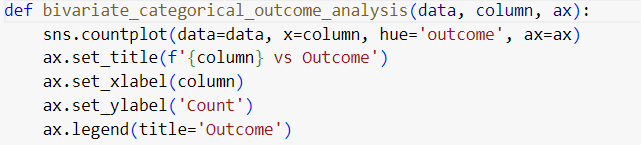
Analysis of variance class balance using outcomes of dependent variables. There is no need to balance the data set as it appears to be correct.



*Figure 29*

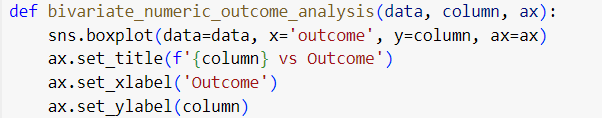
**Bivariate analysis**

For bivariate analysis a function is used to define the categorical variables with 'outcome'.



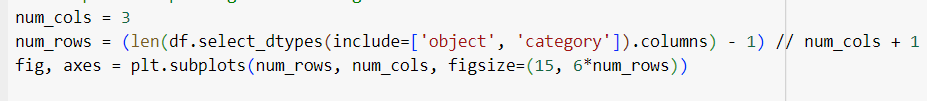
*Figure 30*

Used to define a function by numeric variables with 'outcome' for bivariate analysis.



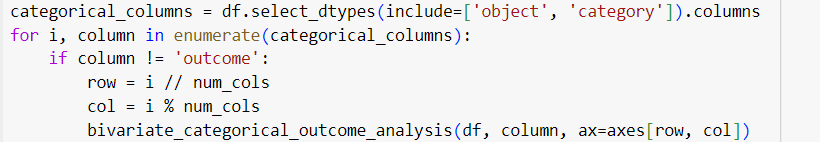
*Figure 31*

Used to set the subplot grid for categorical variables with 'results'.



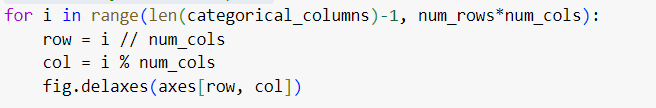
*Figure 32*

Bivariate analysis is performed for categorical variables with 'outcome'.



*Figure 33*

Removes unused partitions.



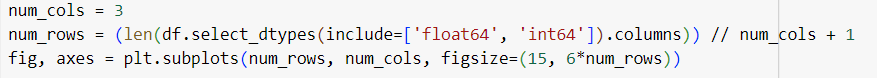
*Figure 34*

The layout is adjusted.



Figure 35

Sets the subplot grid for numerical variables with 'results'.



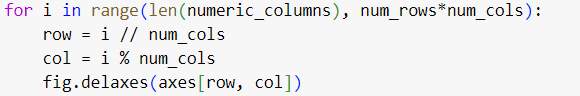
*Figure 36*

Binary analysis is performed for numeric variables with 'outcome'.



*Figure 37*

Removes unused partitions.

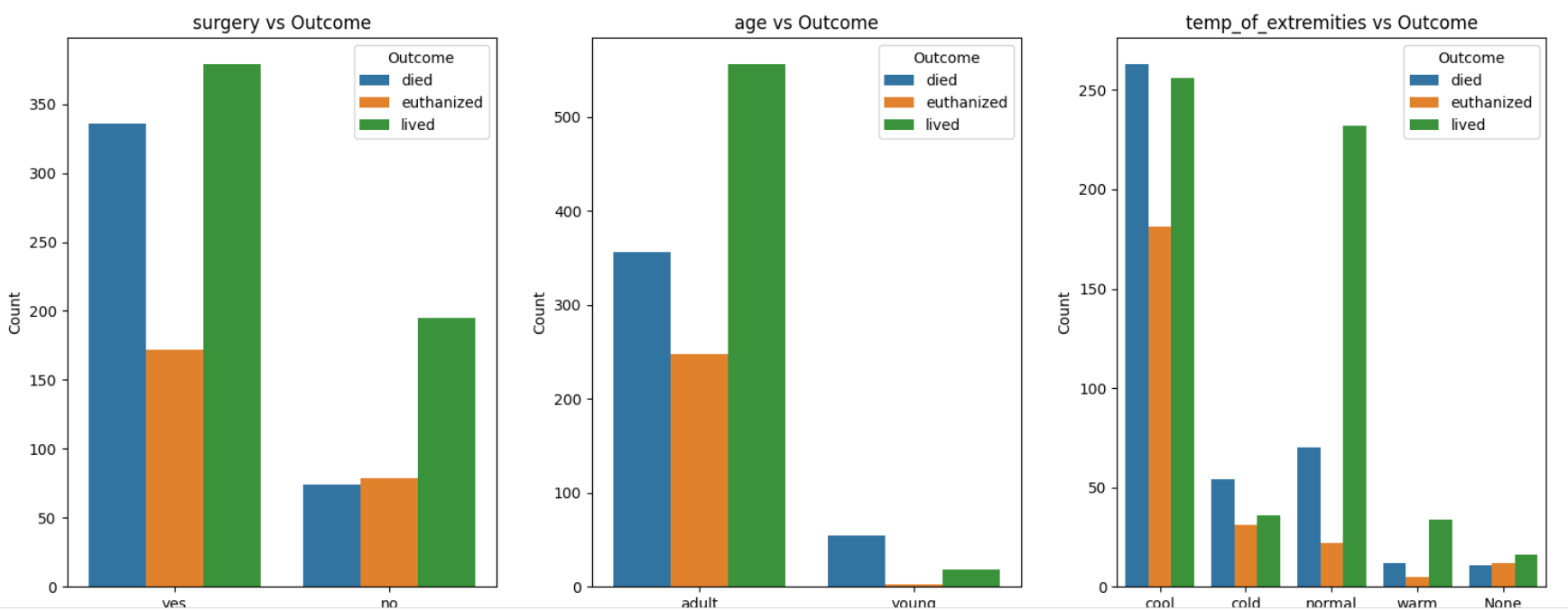


*Figure 38*

The layout is adjusted.



*Figure 39*

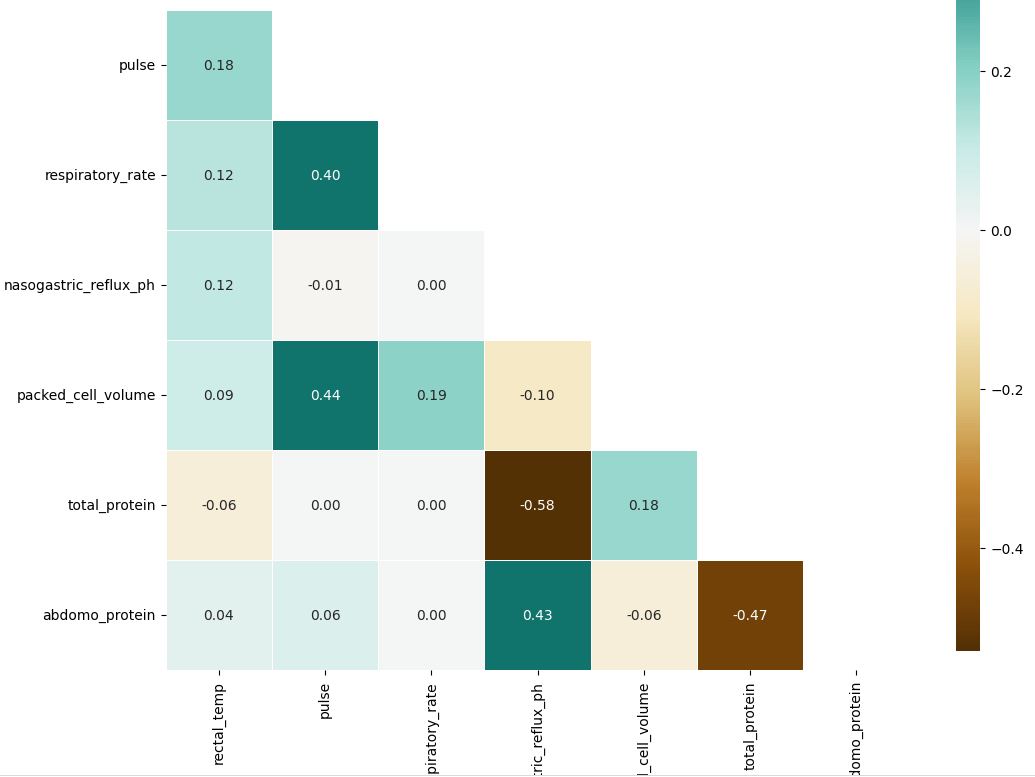
****

*Figure 40*

**Multivariate analysis**

It helps to understand the interdependencies between the variables in the data set and helps to generate a correlation plot for df.





*Figure 41*

There is a considerable negative correlation between total\_protein variable and nasogastic\_reflux\_ph variables. And also there is a positive correlation between abdomo\_protein and nasogastic\_reflux\_ph variables. Also there can be seen some negative correlation of total\_protein with abdomo\_protein. Respiratory rate and pulse variables have a positive correlation of 0.4. So when doing advance analysis or model building we need to focus of those correlation variables. if needed we must omit the correlated variables.

**Test of association for categorical variables**

The a library and a tool used for this are as follows.



*Figure 42*

Loading the data set.



*Figure 43*

Drops rows with missing values.



*Figure 44*

Only categorical columns are selected here.



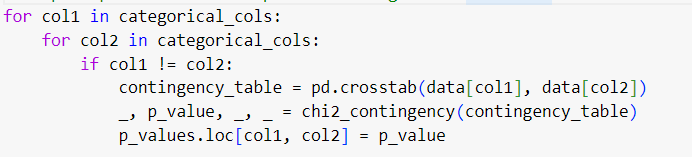
*Figure 45*

Here an empty DataFrame is created to store the p values.



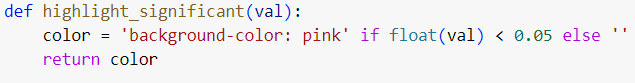
*Figure 46*

The p-value is calculated for all pairs of categorical variables.



*Figure 47*

Performs the task of highlighting significant p-values.



*Figure 48*

The p-values are formatted into the DataFrame.

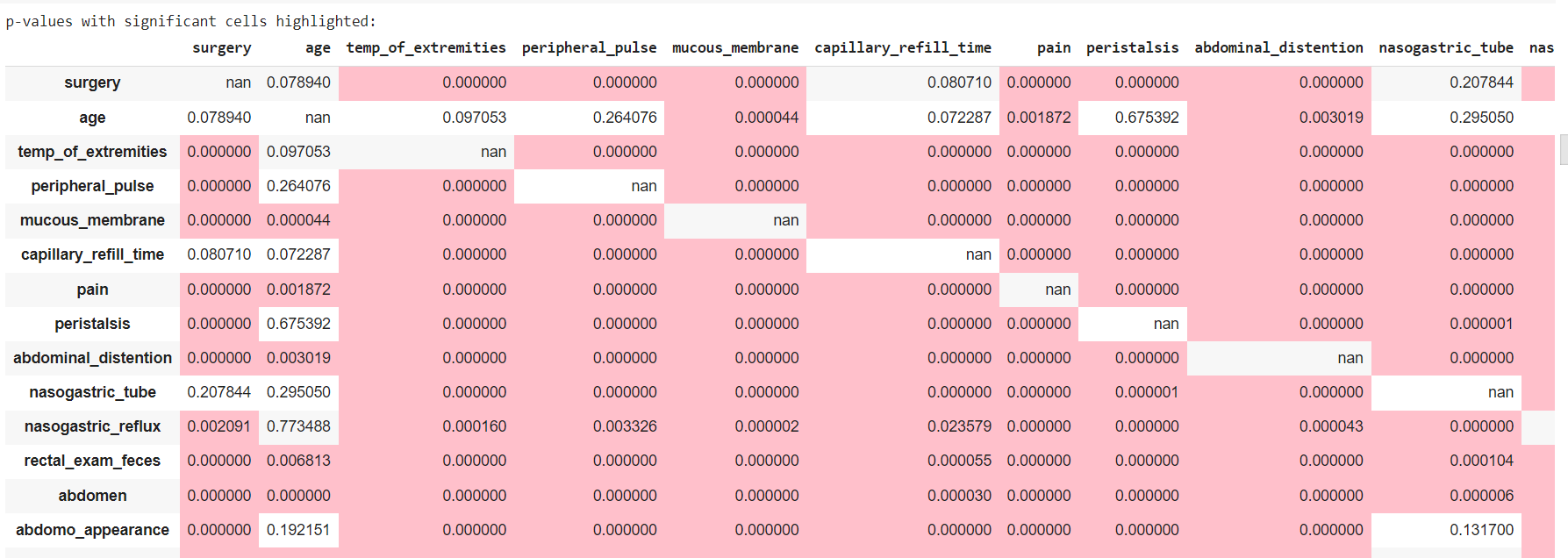


*Figure 49*

Displayed the formatted p-value.



*Figure 50*



*Figure 51*

**d. System architecture**

A complete machine learning model is used to predict the health outcomes of horses according to the developed methods. Here, based on several factors such as the horse's age, heart rate, breathing rate, pain level, abdominal division, the outcome of the horse is indicated as lived, euthanized and died. By analyzing all the factors provided in the data set, it is possible to easily determine the outcome of the horses with the results obtained from the data path trained by the model. Here data is collected based on all the factors given in the data set. The collected data is processed into training validation datasets which are separated for training and selection for the system. The model uses a user-friendly interface to make it easier for users to use the trained data. This allows users to enter horse health results and obtain horse outcome predictions using a clear interface. This process can be simply described as the architecture of a system.

**How your application differs from other existing applications**

According to the health results of the horses, decisions are made based on factors such as the behavior patterns of the horses and the appearance of the horses when considering the prediction of their outcome (lived, euthanized and died). Systems often focus on health factors such as age, pain level, respiratory rate, pulse, abdominal distension, etc. For example, a veterinarian monitors the horse's health to determine the horse's results. Looking at horse weight, observing horse behavior patterns can be another real-world application used by horse riders as well as horse owners. Although we focus on some of the same characteristics for the health status of the horses used for the advanced model in predicting the outcome of the horses from the existing applications, the methodology used in doing this is different. Because large-scale operations require a large number of people and manpower. Hence, the use of data sets obtained from real observations, training machine learning algorithms provides a sustainable solution. Using this advanced algorithm, it is possible to predict the results of horses, so it does not require direct human intervention. Therefore, this has become a very good and efficient alternative for predicting the results of horses in the real world.

**Machine learning technique**

In order to calculate the outcome of horses according to the health results of horses, several machine learning methods are analyzed using training data sets and then they are compared with each model and the best model is selected and used. This project mainly focuses on predicting an outcome such as Outcome (died, authorized, lived), so it falls under natural regression functions. Therefore, 6 methods are selected to obtain optimal results from the training data. All these ensemble methods are selected by machine learning techniques.

* Ensemble techniques

The reliability and accuracy of data using equine health outcomes is enhanced by the inclusion of cohort approaches. This involves repeatedly applying the same algorithm or combining several different algorithms among techniques to create a model that performs more accurately. The main goal of doing this is to minimize errors and reduce overfitting. Ensemble approaches are used to combine the outputs of many models and they perform well. It is important for specific tasks such as outcome prediction. Following are the various grouping techniques used in the development of the project.

That is

* SVM – Support vector machine
* Logistic model
* Random forest model
* Random forest - Reduced model
* XGBoost model
* Neutral network - sq model

**e. Full model evaluation**

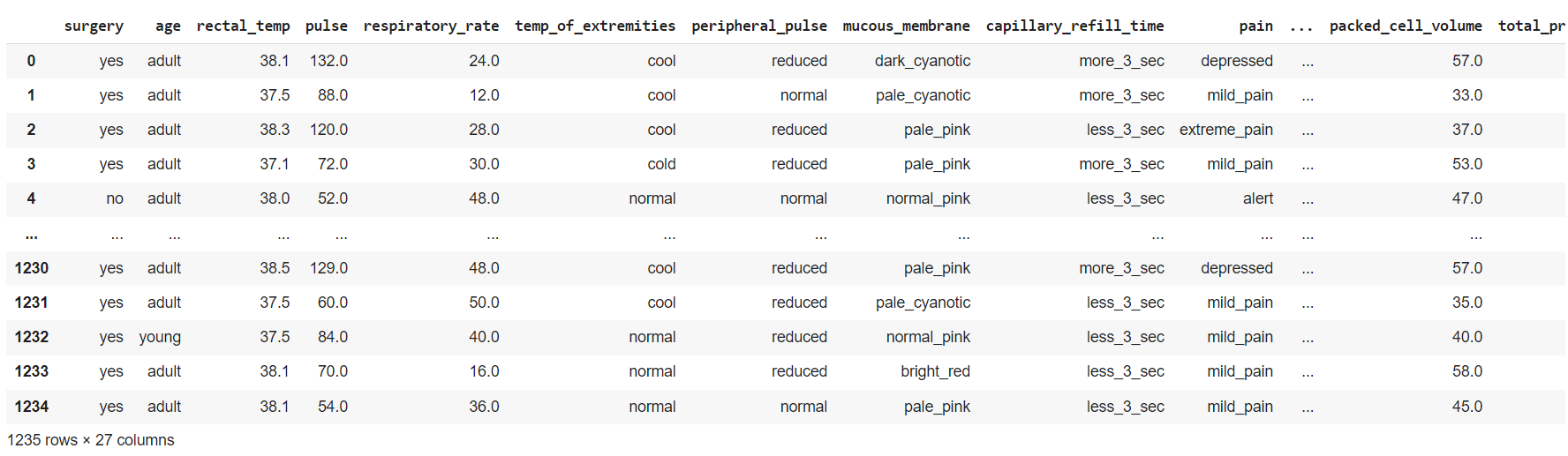
This model is concerned with predicting the correct outcome of horses based on their health status. As mentioned above EDA is implemented to analyze the data set used in training the model to get the final desired result in predicting the results of the horses. After that analysis, the best parameters to proceed with the training process are determined.

Advanced analysis - ML model building

To build Machine Learning, the data in the Train dataset has to be re-viewed.



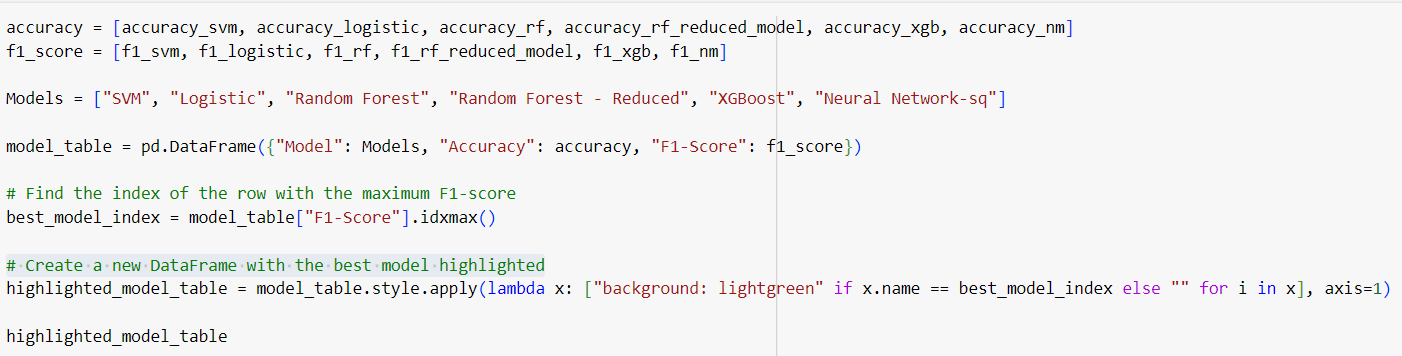
*Figure 52*



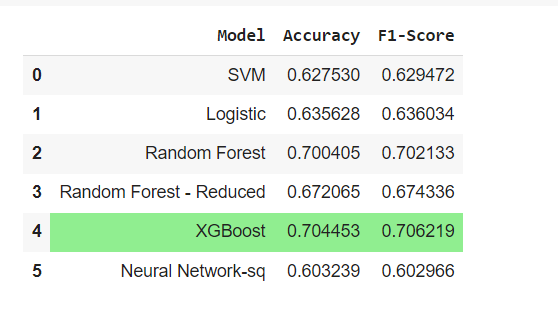
*Figure 53*

Model evaluation for best model

This code calculates the accuracy of machine learning models, and F1 score metrics. Here, 6 different regression models were proposed to predict the target variables. Namely SVM, Logistic, Random Forest, Random Forest-reduced, XGBOOST and Neural Network-sq. Here all these models are trained on datasets ready for evaluation. The XGBoost model is colored green as the best model.



*Figure 54*



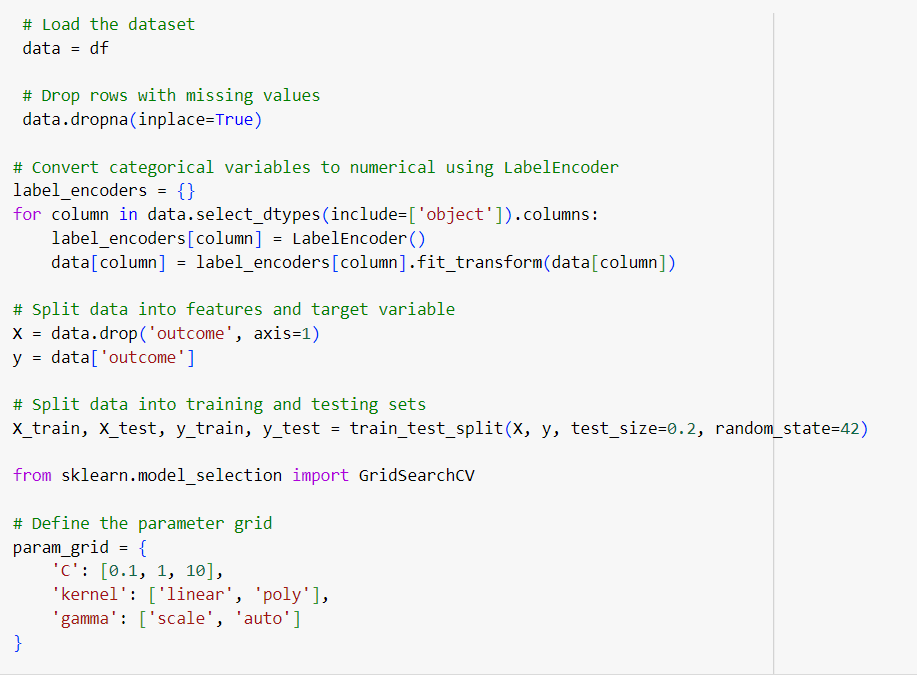
*Figure 55*

**Implementation details**

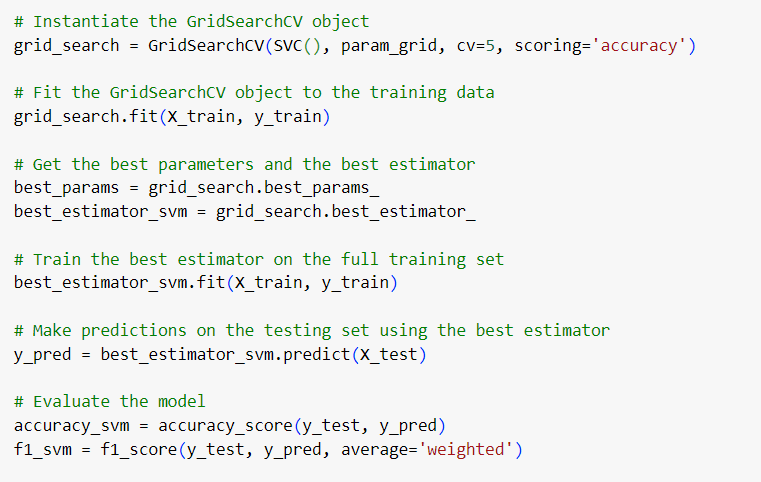
**SVM model**

* with hyperparameter optimization

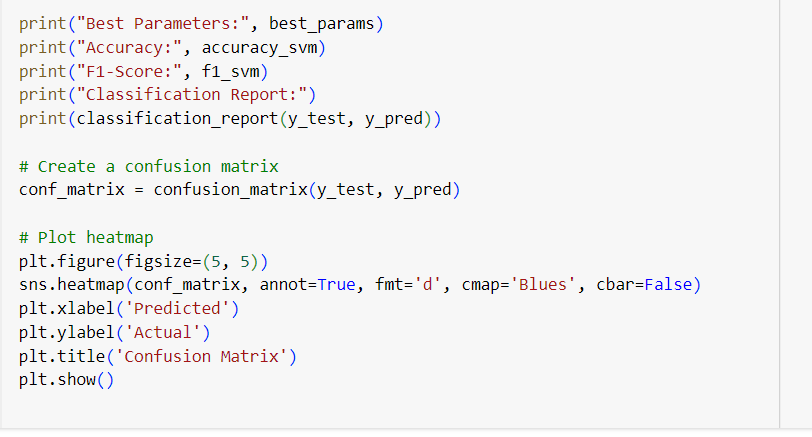
Here first, in describing the SVM model, the model has been predicted with hyperparameter optimization and without hyperparameter optimization. First let's consider model training with hyperparameter optimization. First the relevant dataset is loaded and then the missing values are removed. Then convert categorical variables to numeric variables using label Encoder. Then the data from numerical variables are mapped to feature and target variables. Then split into training and test datasets. That division then defines the parametric grid. Then instantiate it using the GridSearchCV object. It then fits the GridSearchCV object to the training data. After fitting, the best parameters and best estimates are calculated. Then the best estimator is trained on the full training set. It makes predictions on the test set using that best estimate. Then the model is evaluated. It creates a confusing matrix. And then finally plot the heat map.



*Figure 56*



*Figure 57*

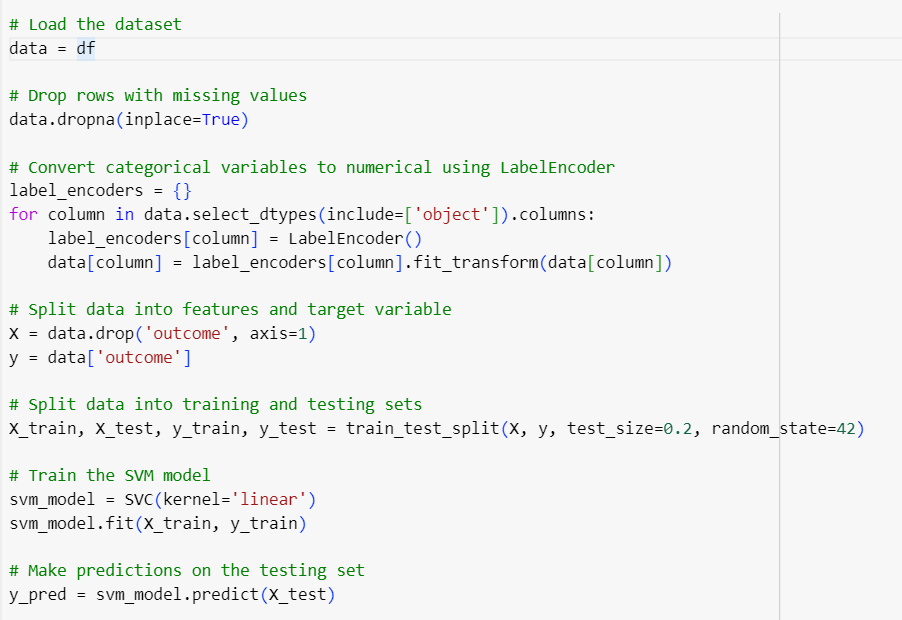


*Figure 58*

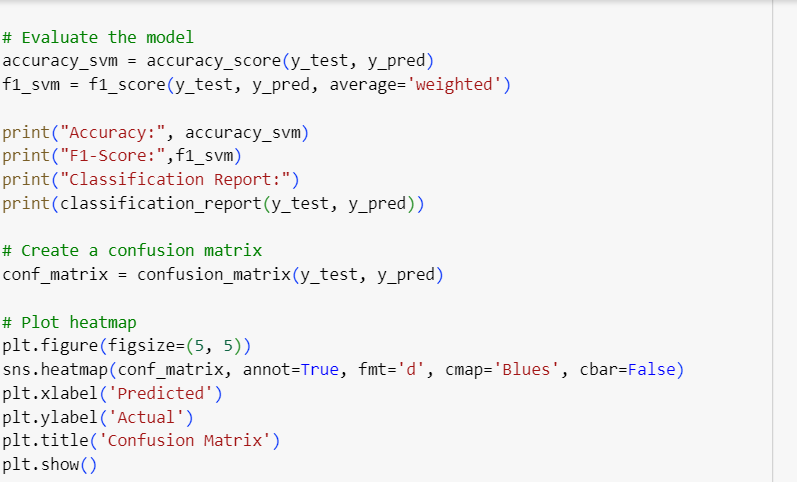
Here, when hyperparameter optimization is done, the accuracy of that model is reduced, so it has been commented in the code.

* Without hyper parameter optimization

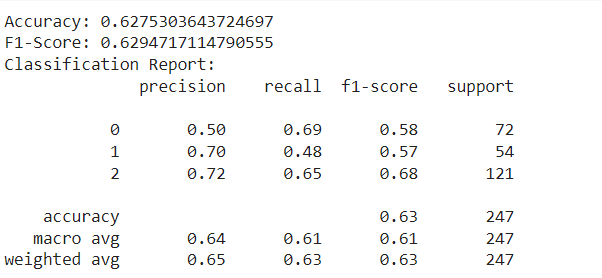
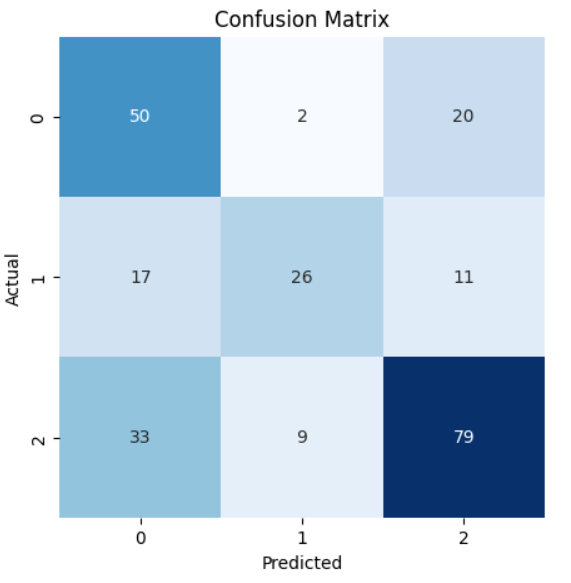
Here the SVM model performs model prediction without hyperparameter optimization. First the relevant dataset is loaded and then the missing values are removed. Then convert categorical variables to numeric variables using label Encoder. Then the data from numerical variables are mapped to feature and target variables. Then split into training and test datasets. Then the SVM model is trained. Predictions are then made on the test set. The model is then trained. After training, a confusion matrix is formed. The plots then form the heat map.



*Figure 59*



*Figure 60*

*Figure 61*

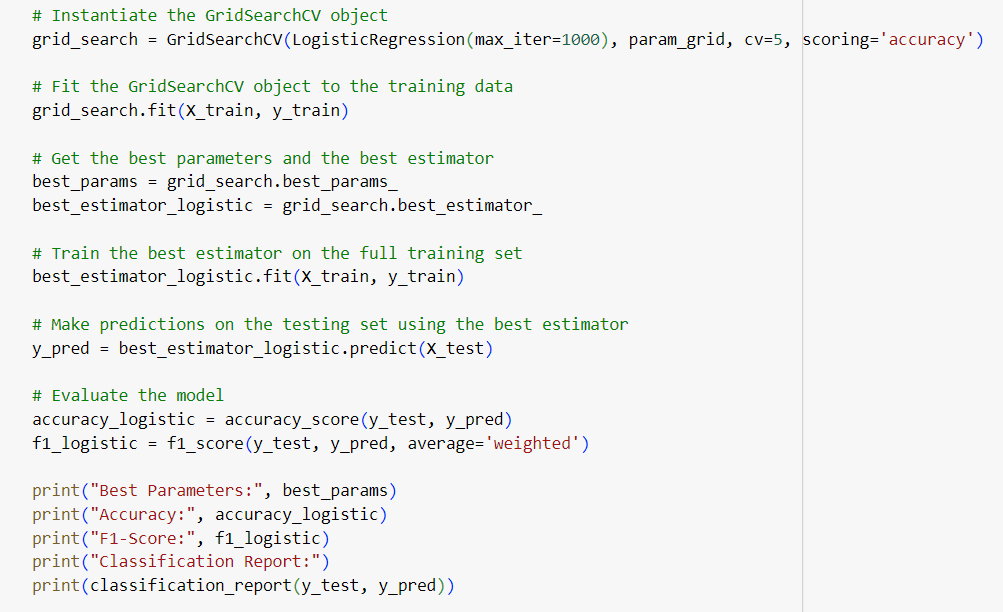
**Logistic model**

* with hyper parameter optimization

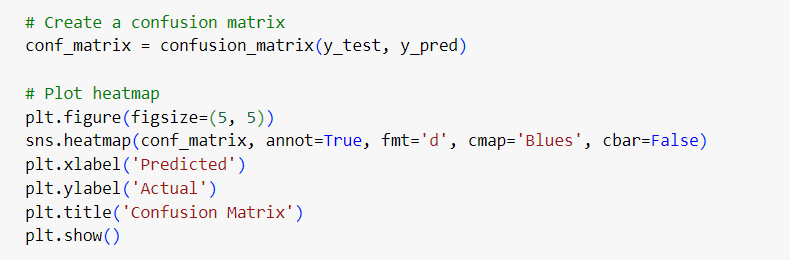
Second, let's consider how Logistic model performs one hyper parameter optimization. First the relevant dataset is loaded and then the missing values are removed. Then other features and target variables are considered. That is, X considers all the remaining columns of the data set except the outcome one, and only the outcome column is considered by y. Here x predicts y based on features. Then the classification columns are identified. Then one-hot encoding is applied to the classification columns. This converts categorical variables into numerical forms. Then split into training and test datasets. Then after that division defines the parametric grid. Then instantiate it using the GridSearchCV object. It then fits the GridSearchCV object to the training data. After fitting, the best parameters and best estimates are calculated. Then the best estimator is trained on the full training set. It makes predictions on the test set using that best estimate. Then the model is evaluated. It creates a confusing matrix. And then finally plot the heat map.



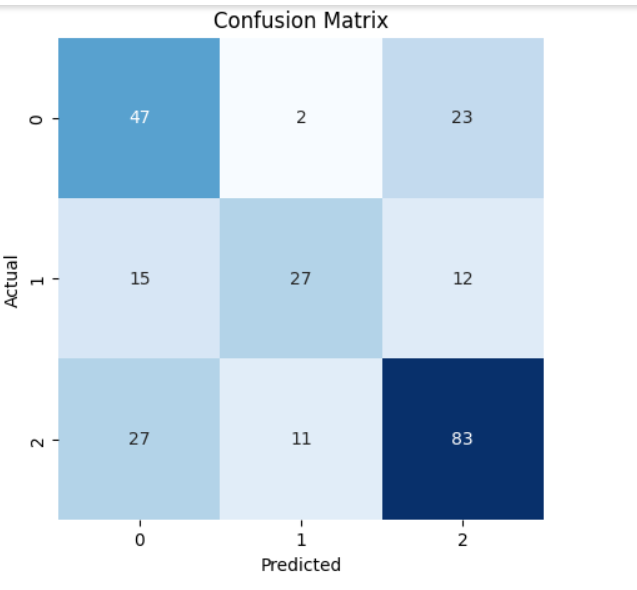
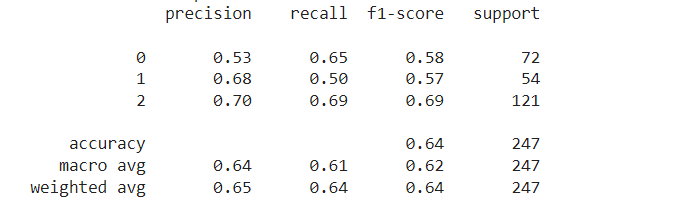
*Figure 62*



*Figure 63*



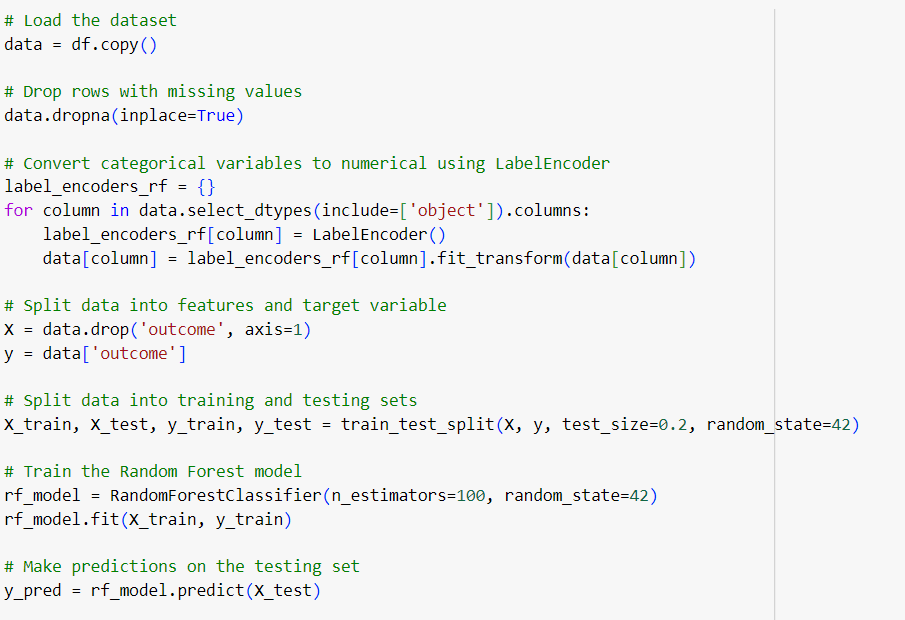
*Figure 64*



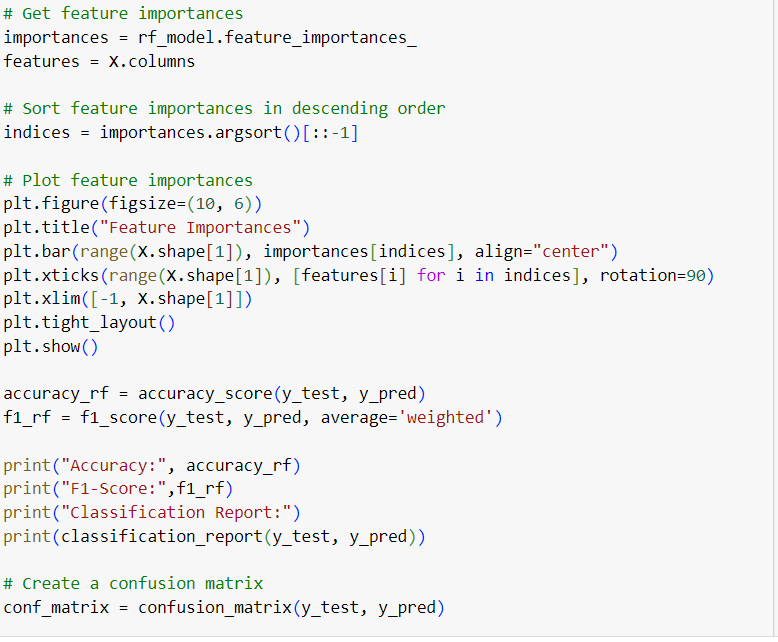
*Figure 65*

**Random Forest model**

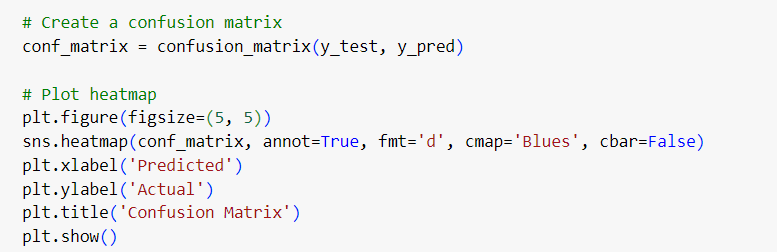
Then let's consider how to make prediction for a random forest model. First the relevant dataset is loaded and then the missing values are removed. Then convert categorical variables to numeric variables using labelEncoder. Then the data from numerical variables are mapped to feature and target variables. Then split into training and test datasets. Then the Random forest model is trained. Predictions are then made on the test set. Then consider the feature importance. This arranges the features in descending order of importance. Then the significance of the plot feature is considered with the help of Accuracy and F1 score. A confusion matrix is then formed. The plots then form the heat map.

****

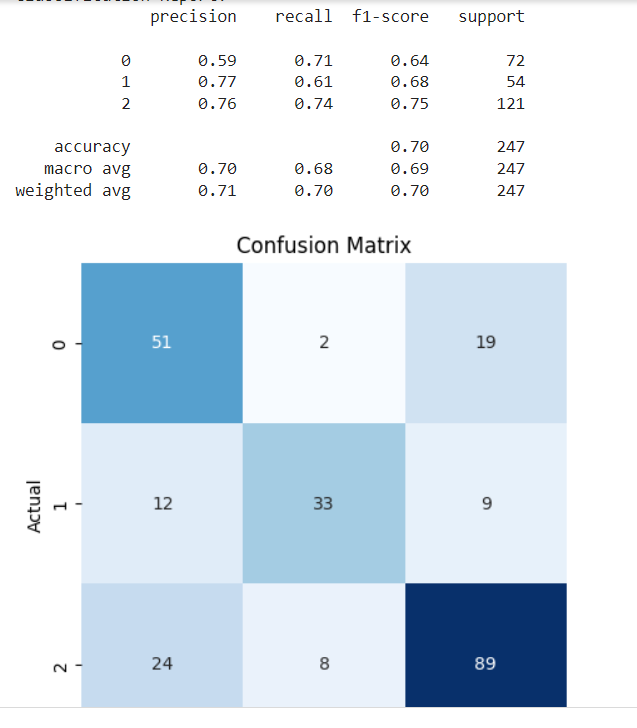
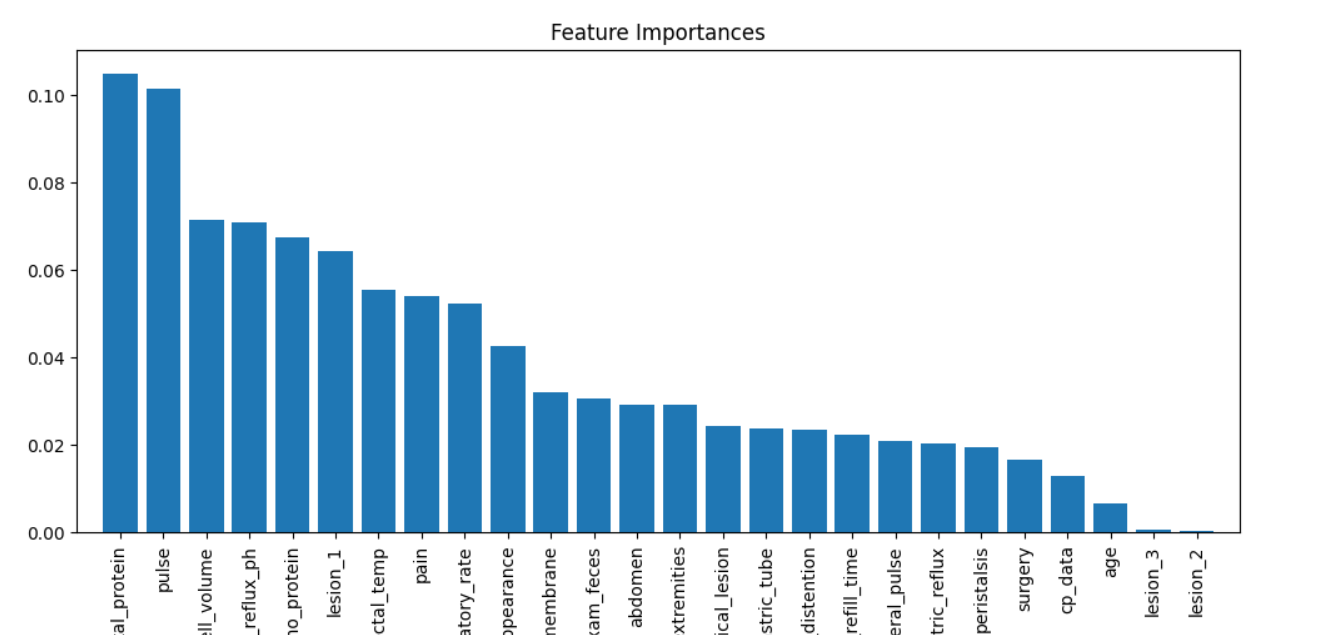
*Figure 66*

****

*Figure 67*

****

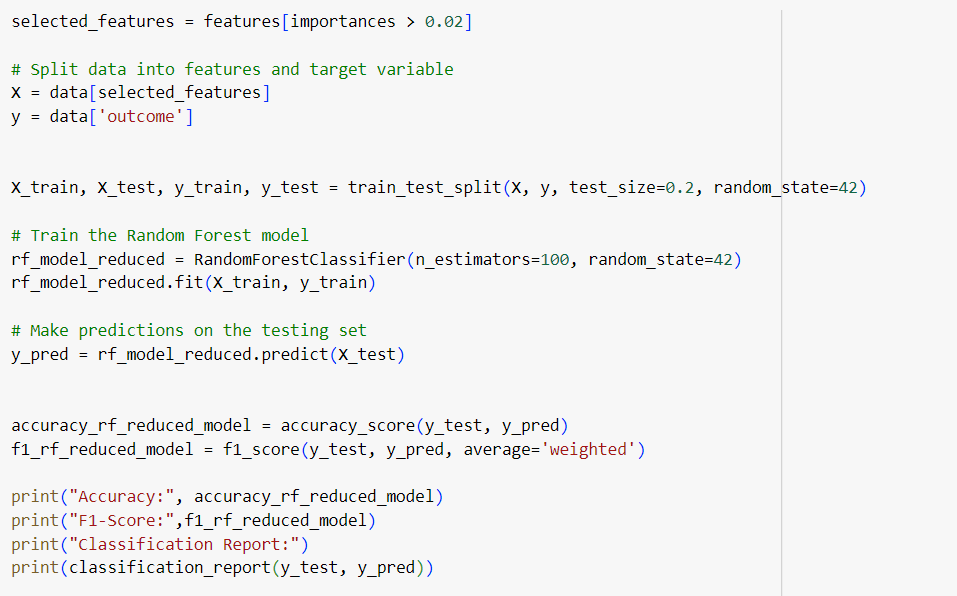
*Figure 68*

****

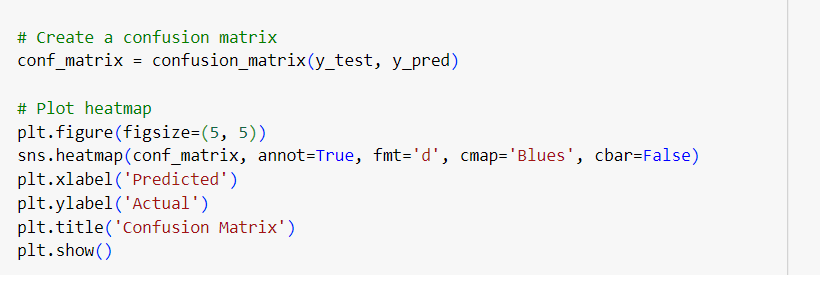
*Figure 69*

**Random forest - Reduced model**

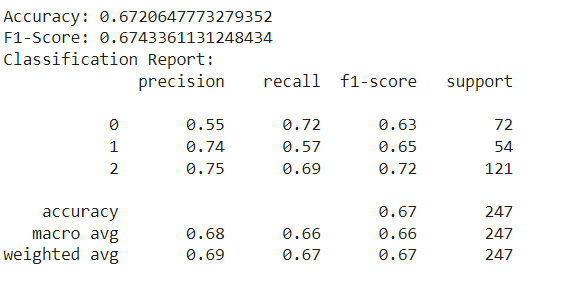
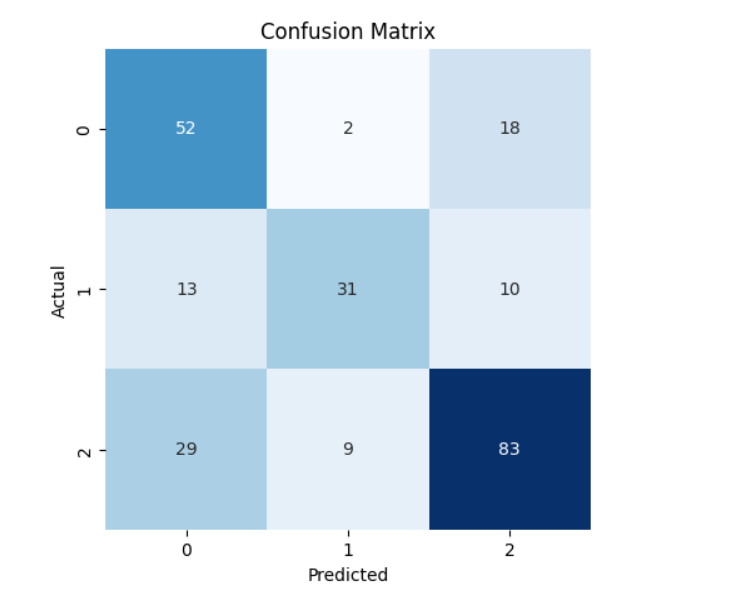
Here, let's consider the reduced model used for feature selection for random forests with the limit calculated as 0.02. First the data is partitioned into features and target variables. Then the Random forest model is trained. Predictions are then made on the test set. A confusion matrix is then formed. The plots then form the heat map.



*Figure 70*



*Figure 71*

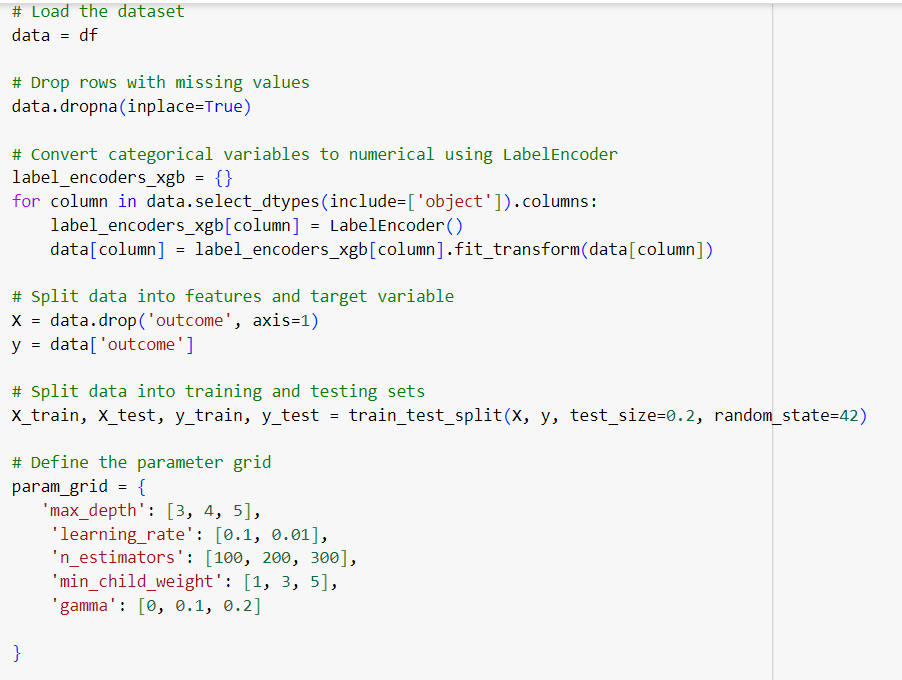
 

*Figure 72*

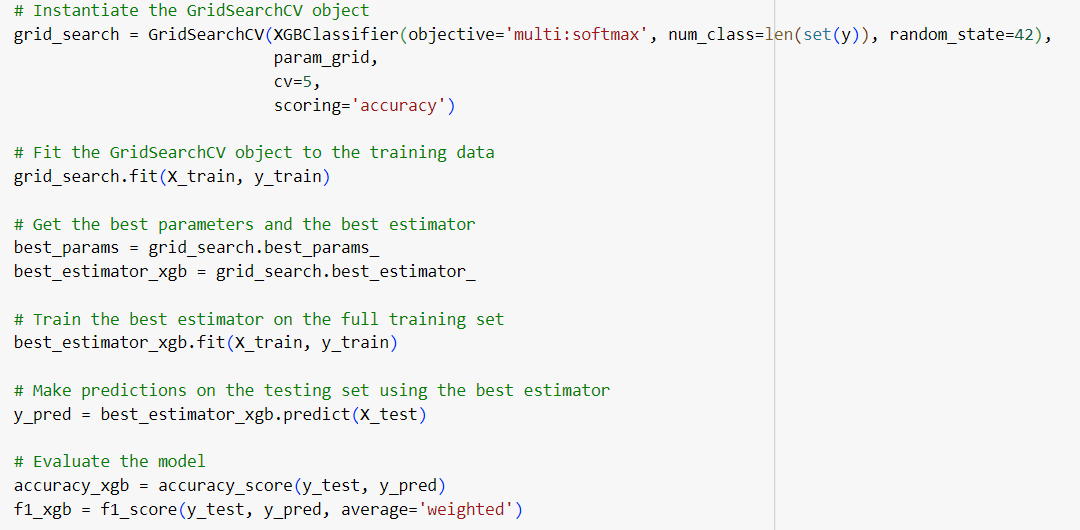
**XGBoost model**

* with hyper parameter optimization

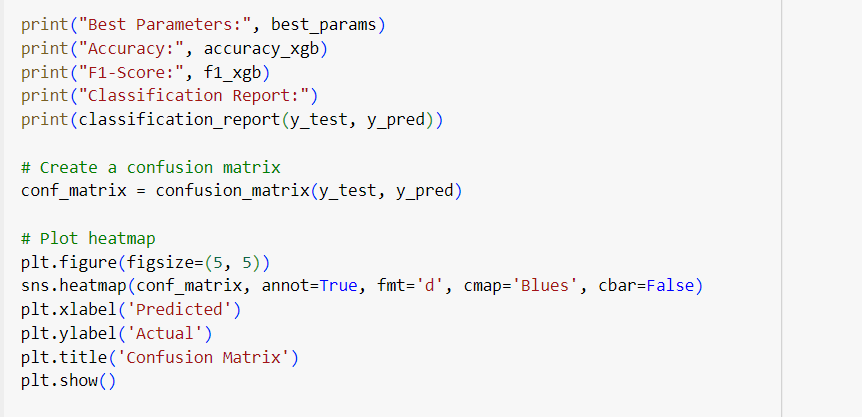
Here first, in describing the XGBoost model, the model has been predicted with hyperparameter optimization and without hyperparameter optimization. First let's consider model training with hyperparameter optimization. After first installing the XGBoost model, the related dataset is loaded and then the missing values are removed. Then convert categorical variables to numeric variables using labelEncoder. Then the data from numerical variables are mapped to feature and target variables. Then split into training and test datasets. That division then defines the parametric grid. Then instantiate it using the GridSearchCV object. It then fits the GridSearchCV object to the training data. After fitting, the best parameters and best estimates are calculated. Then the best estimator is trained on the full training set. It makes predictions on the test set using that best estimate. Then the model is evaluated. It creates a confusing matrix. And then finally plot the heat map.



*Figure 73*



*Figure 74*

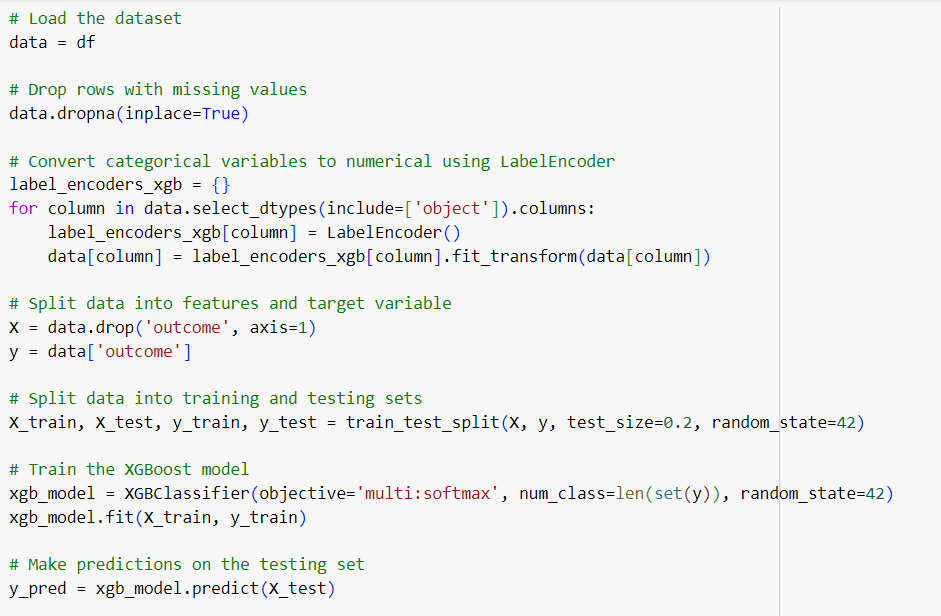


*Figure 75*

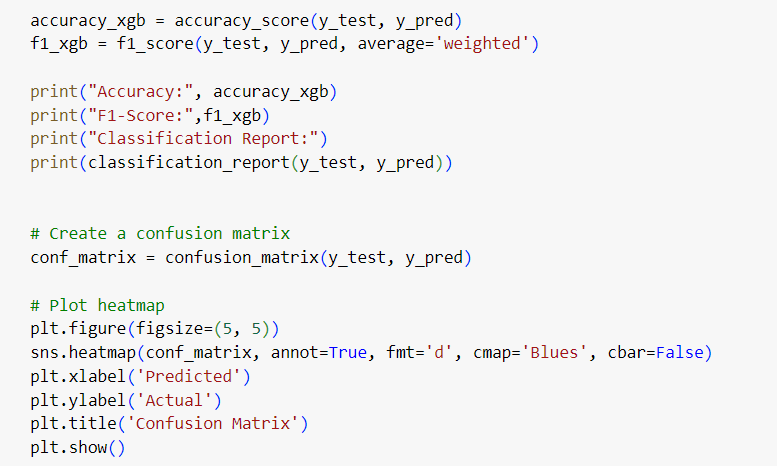
Here, when hyperparameter optimization is done, the accuracy of that model is reduced, so it has been commented in the code.

* without hyper parameter optimization

Here the XGBoost model performs model prediction without hyperparameter optimization. First the relevant dataset is loaded and then the missing values are removed. Then convert categorical variables to numeric variables using labelEncoder. Then the data from numerical variables are mapped to feature and target variables. Then split into training and test datasets. Then XGBoost performs model training. Predictions are then made on the test set. A confusion matrix is then formed. The plots then form the heat map.

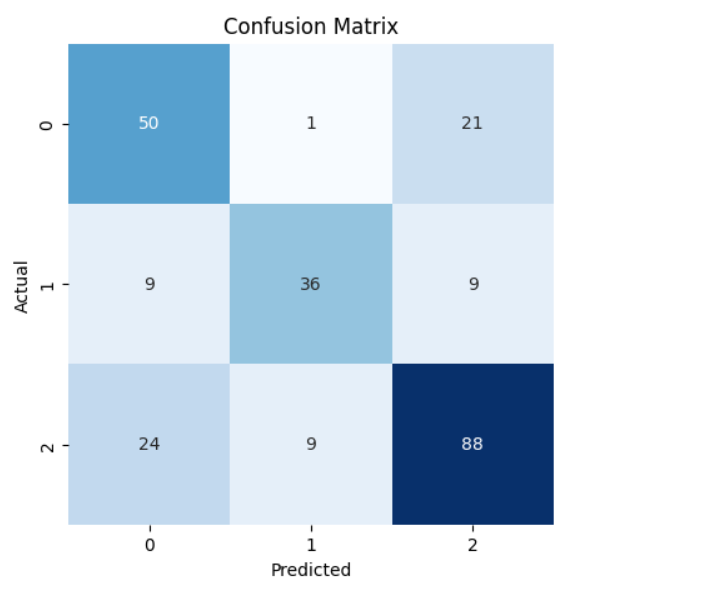
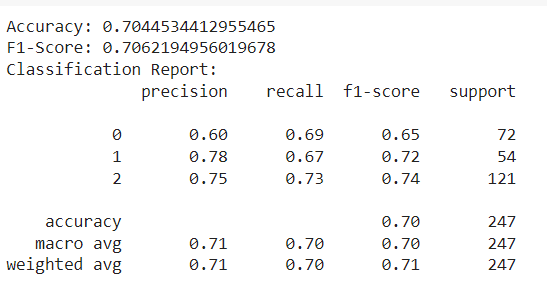


*Figure 76*



*Figure 77*

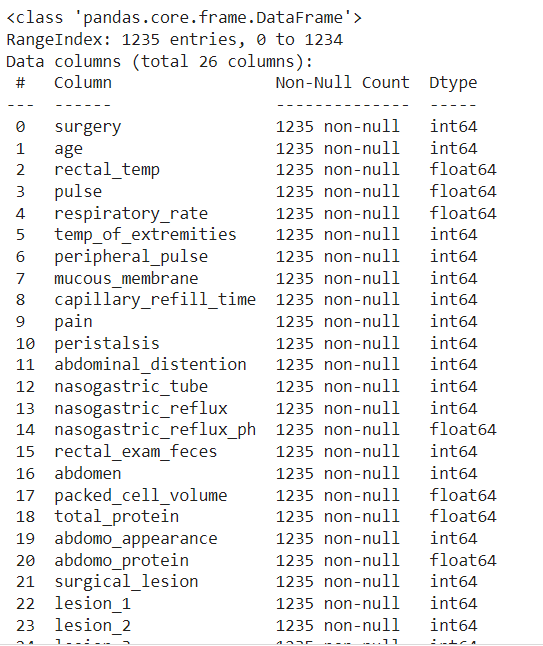
Identifying missing values, evaluating data types in different columns, is useful for understanding the structure, and provides a description of the x function. It includes non-null count and data type one.



*Figure 78*



*Figure 79*

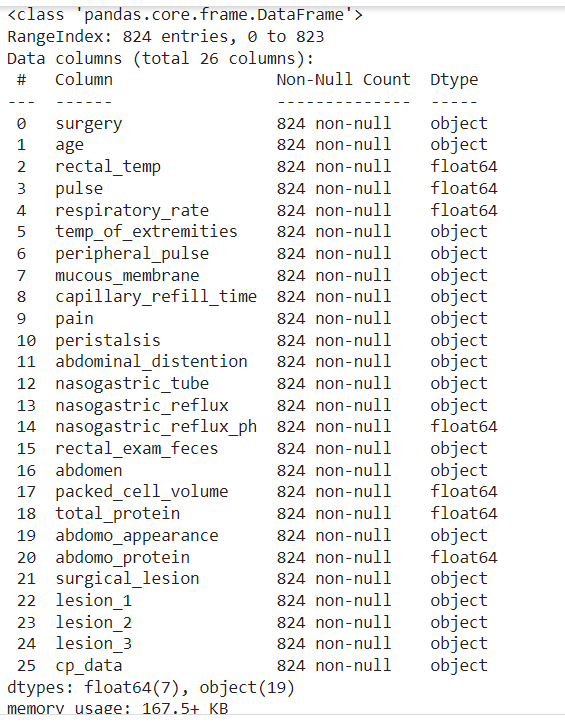


*Figure 80*

The dataset provides a description of the non-null count and data type of the data in the test dataset.



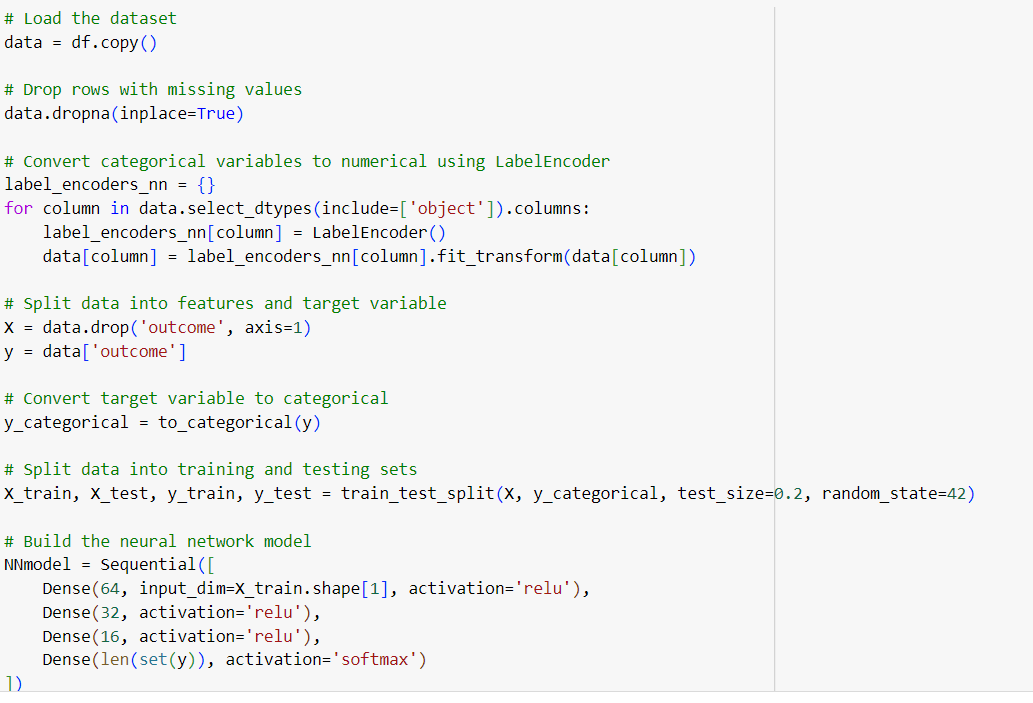
*Figure 81*



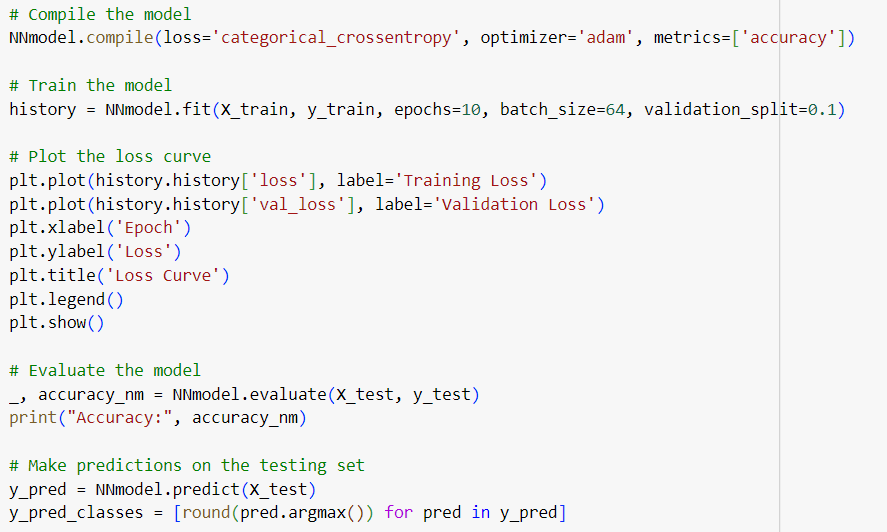
*Figure 82*

**Neural Network model**

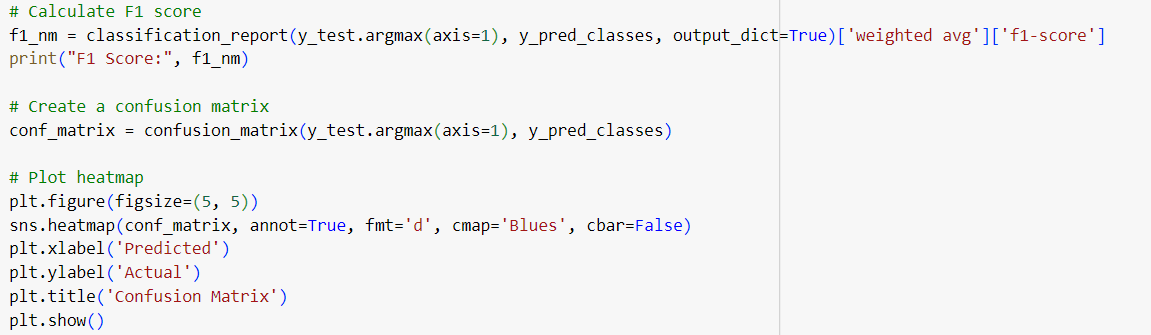
Here, when describing the neutral network model, the relevant dataset is first loaded and then the missing values are removed. Then convert categorical variables to numeric variables using labelEncoder. Here the target variables are converted to classification. Then split into training and test datasets. Then the Natural network model is trained. Then compile the model. The model is then trained. Then the model is evaluated. It then makes predictions on the test set. It calculates one F1 score. A confusion matrix is then formed. The plots then form the heat map.



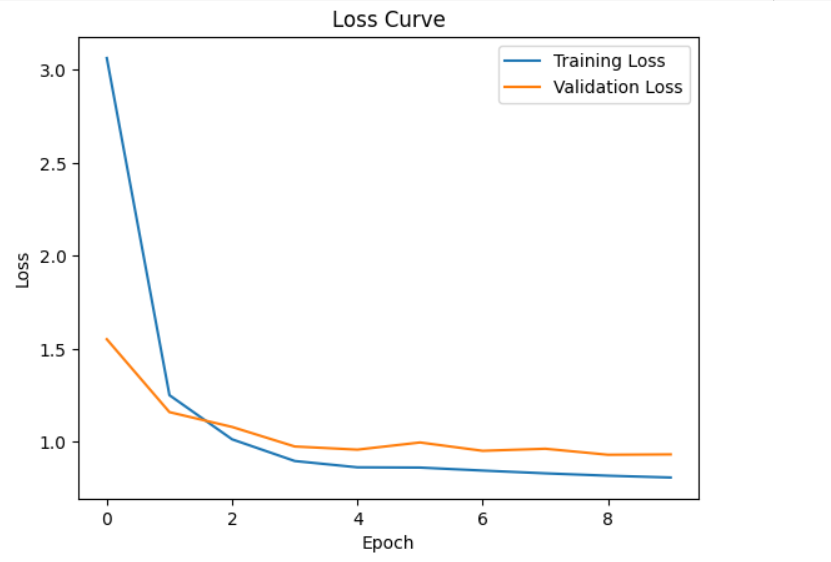
*Figure 83*

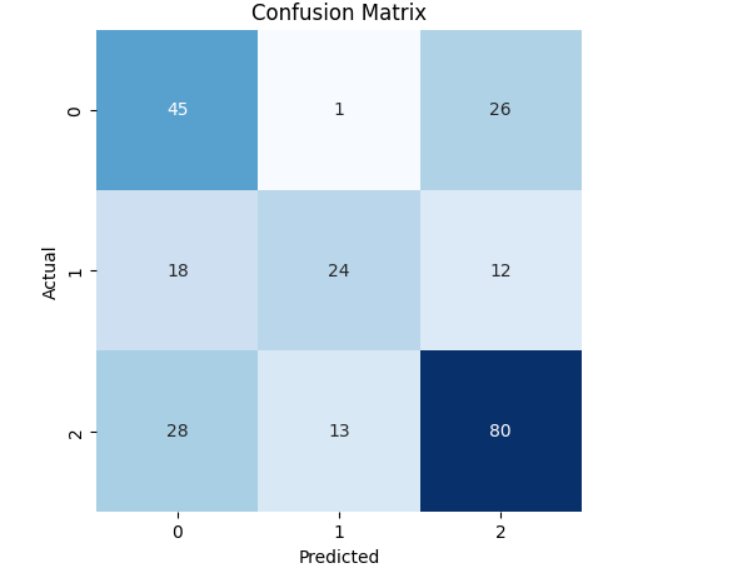


*Figure 84*



*Figure 85*





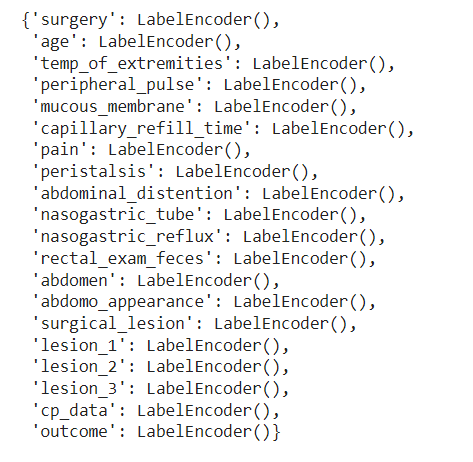
*Figure 86*

**Saving the model**

This code is used to encode numerical values of categorical features.

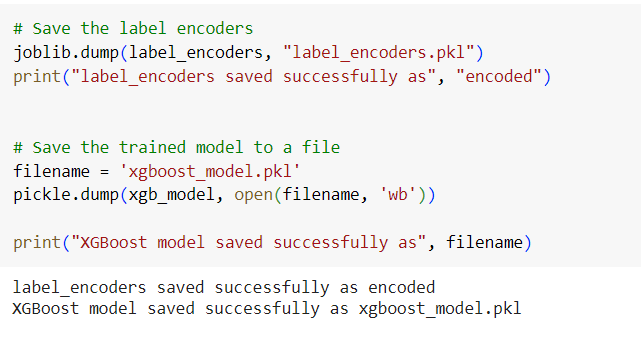


*Figure 87*



*Figure 88*

This code is used to compress the label code and compress the trained model to a file. Accordingly, label\_encoders and XGBoost model are compressed successfully.

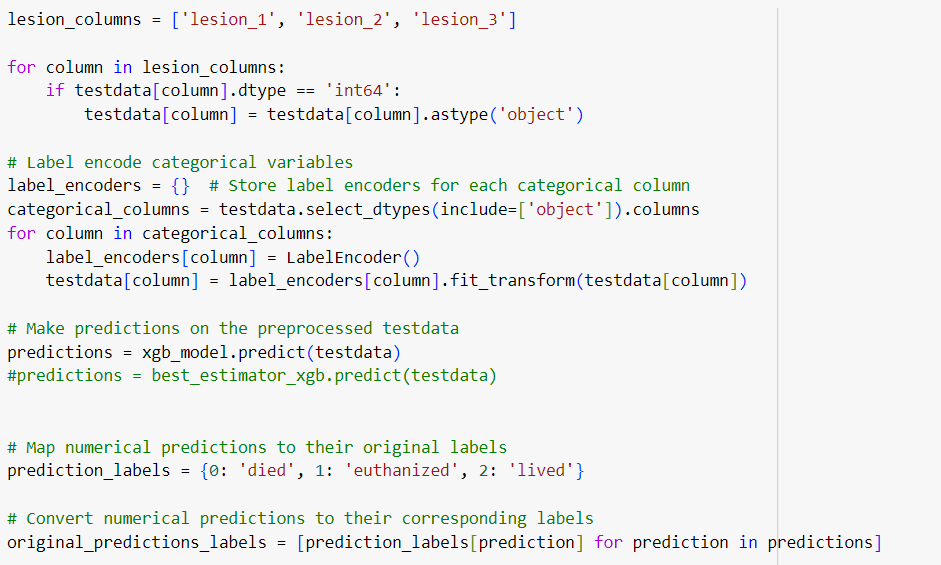


*Figure 89*

**Making predictions for new test data**

* getting predictions using XGBoost model

Here first the data type of the column is converted from int64 to object data type. and then encodes categorical variable labels. Each classification column is then stored in a label code. Predictions are then made based on the preprocessed test data. The numerical predictions are then mapped to their original labels. The numerical predictions are then converted to their corresponding labels. Then printing occurs as the original prediction.



*Figure 90*



*Figure 91*



*Figure 92*

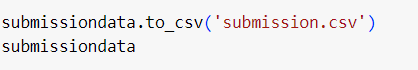
**Preparing submission dataset**

This code generates or updates the result column of the presentation data according to the results predicted by the machine learning models.

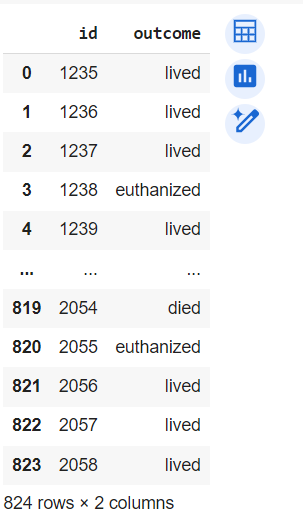


*Figure 93*

In this code, the content of the submission data is directed to a CSV file called submission.csv.



*Figure 94*



*Figure 95*

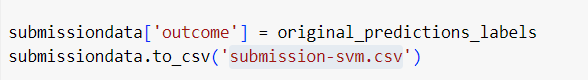
**Getting predictions using models**

Here first the data type of the column is converted from int64 to object data type. and then encodes categorical variable labels. Each classification column is then stored in a label code. Predictions are then made based on the preprocessed test data. The numerical predictions are then mapped to their original labels. The numerical predictions are then converted to their corresponding labels. Then the original forecast is printed as submission-svm.csv, submission-logistic.csv, submission-randomf.csv, submission-randomreduced.csv.

* Getting predictions using SVM model



*Figure 96*

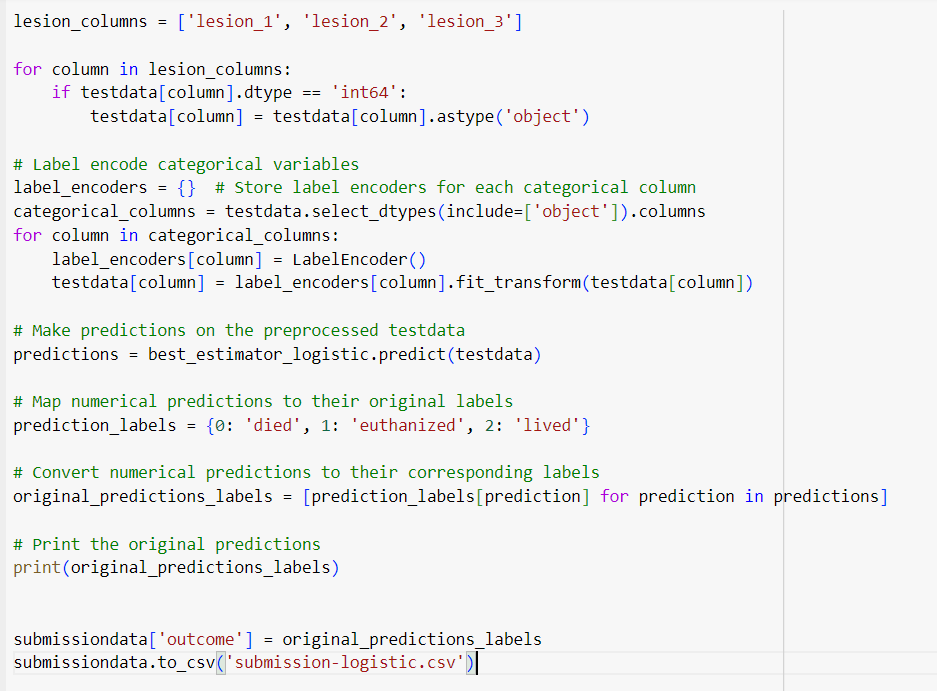


*Figure 97*

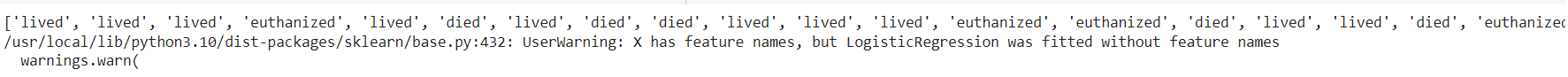


*Figure 98*

* Getting predictions using Logistic model

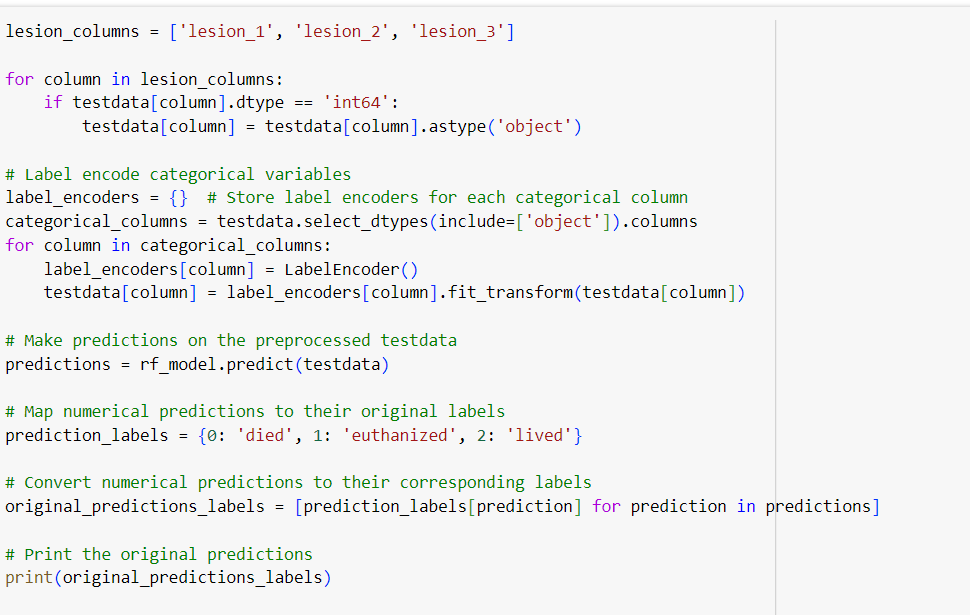


*Figure 99*

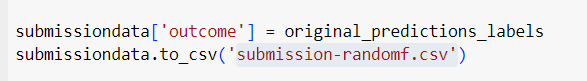


*Figure 100*

* Getting predictions using Random Forest model



*Figure 101*

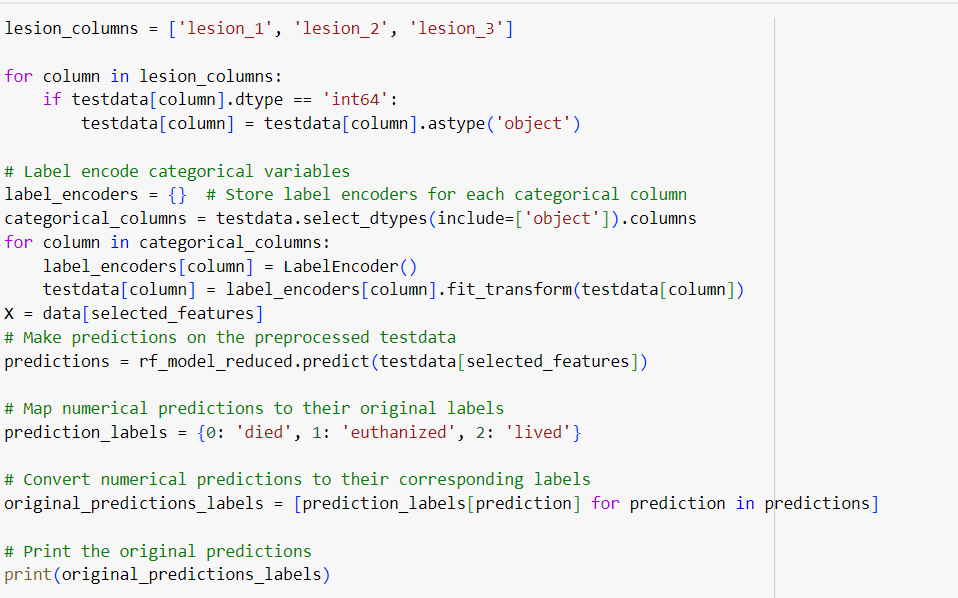


*Figure 102*

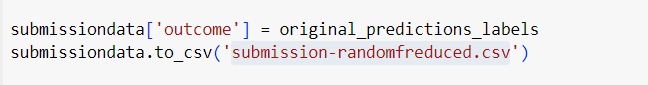


*Figure 103*

* Getting predictions using Random Forest reduced model



*Figure 104*

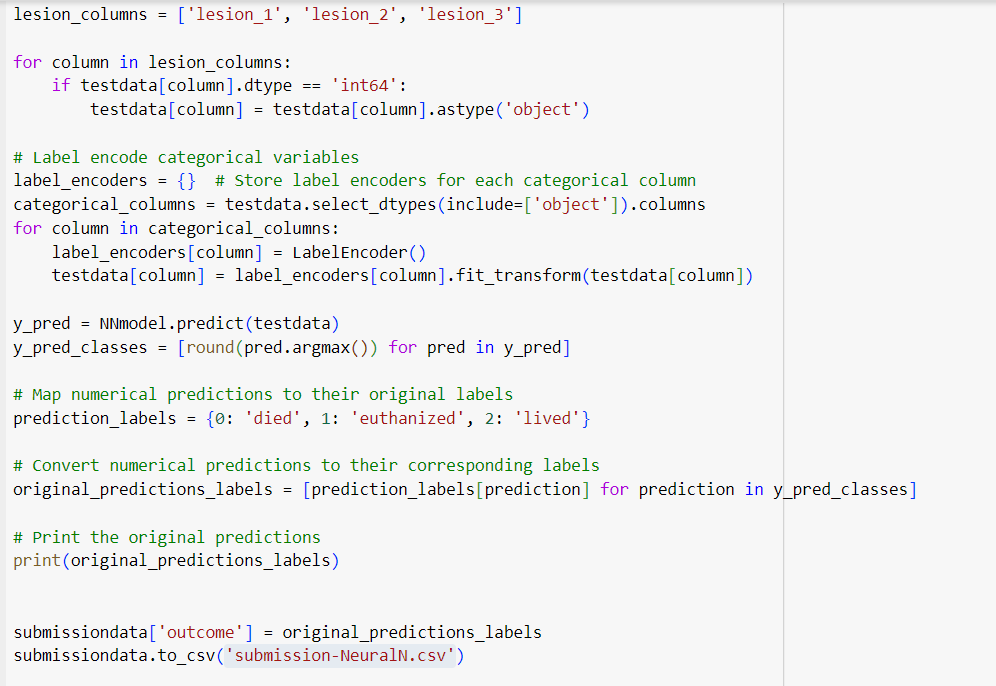


*Figure 105*

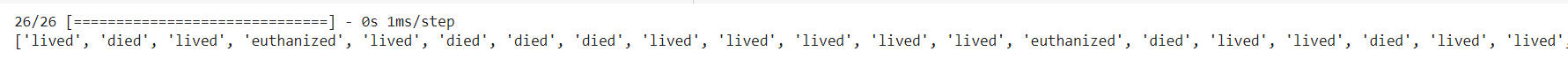


*Figure 106*

* Getting predictions using Neural Network model



*Figure 107*



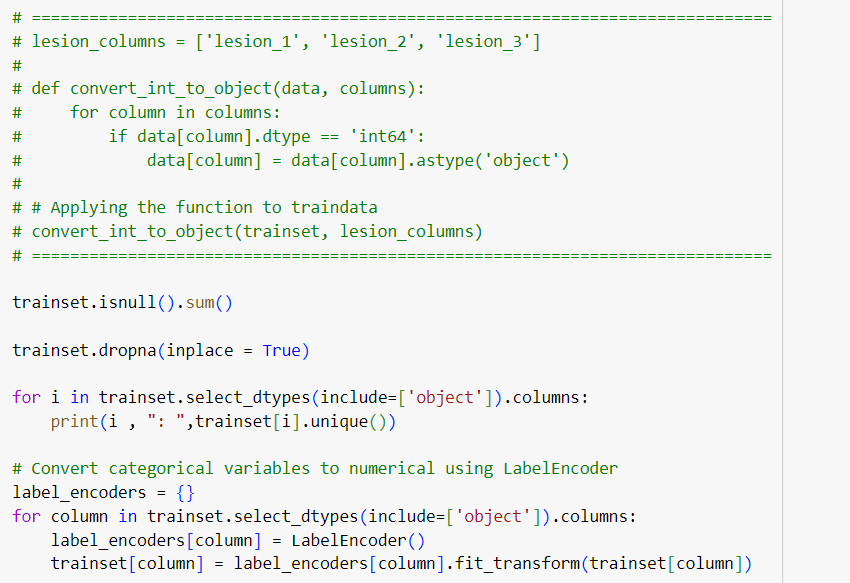
*Figure 108*

Here the numerical variables lesion\_1,lesion\_2 and lesion\_3 are associated with the XGBoost model, which makes the web application easy to build and has a significant accuracy of 76%.

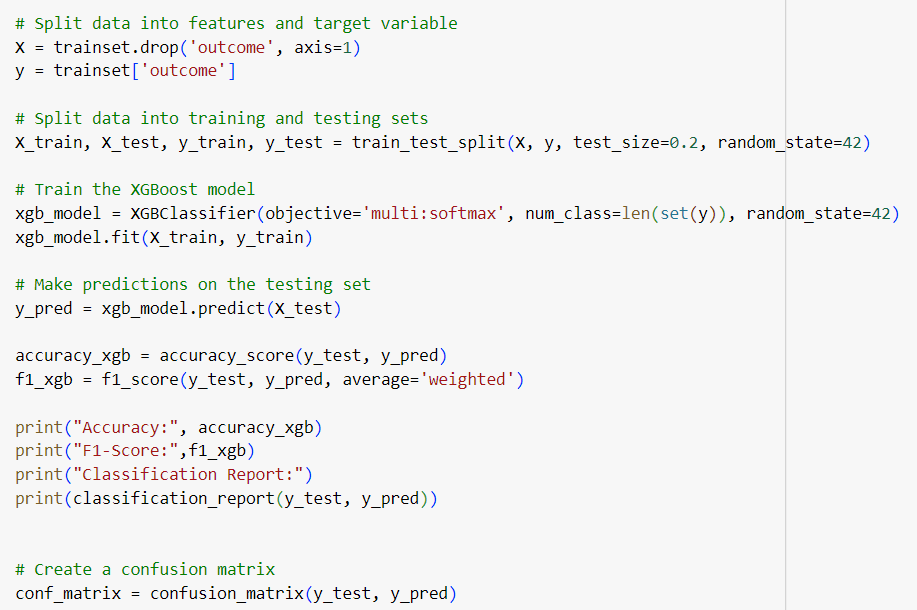
Here data loading and preprocessing are done first and missing values columns are removed. Then the numerical forms for the categorical variables are converted and the labels are coded. XGBClassifier is then trained on the training data and predicts on the test set and generates a classification report showing precision, recall and F1 score and support. Then XGBClassifier and labelEncoder, trained by pickle and joblib, compress a disc. Accordingly, create a file as label\_encoders\_new.pkl and xgboost\_model\_new.pkl.



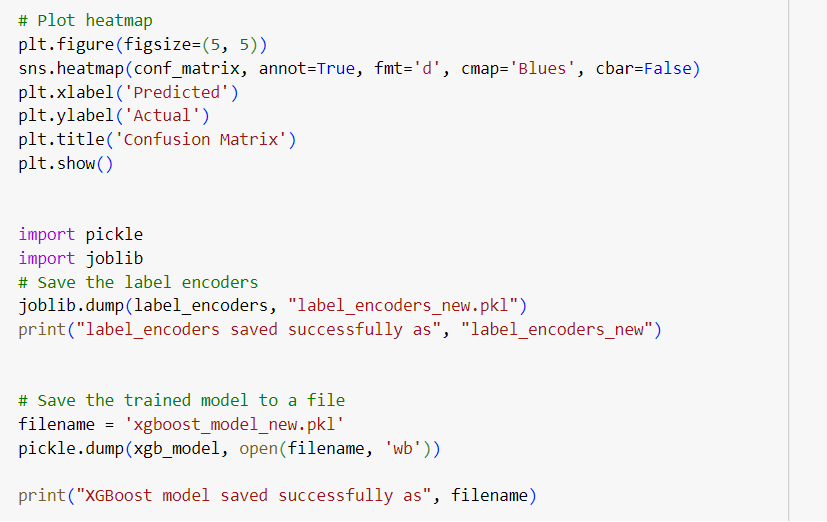
*Figure 109*



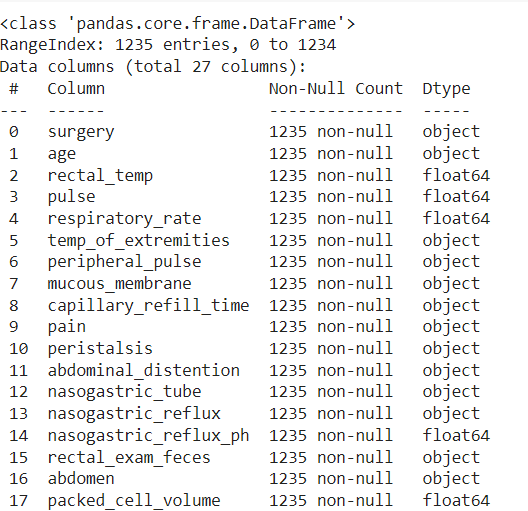
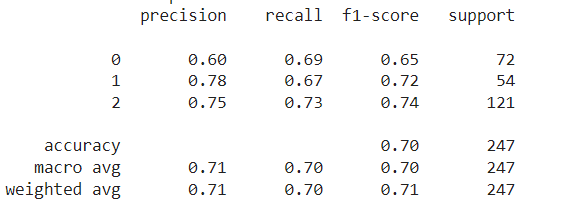
*Figure 110*

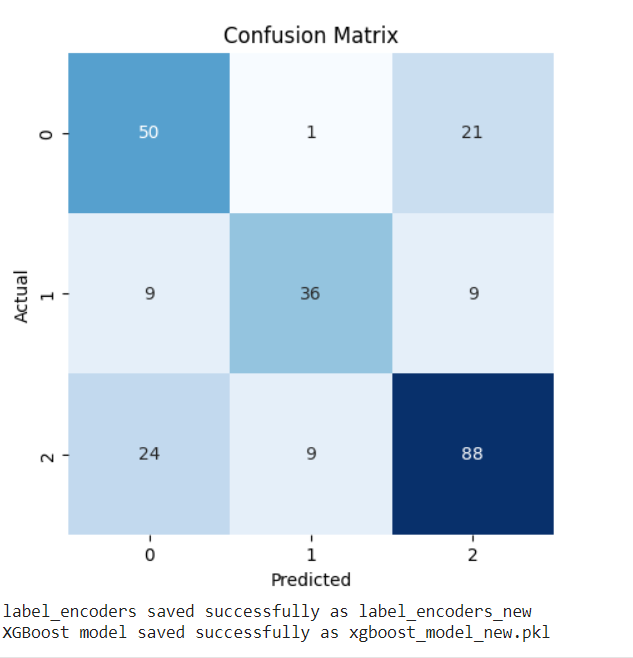


*Figure 111*



*Figure 112*



*Figure 113*

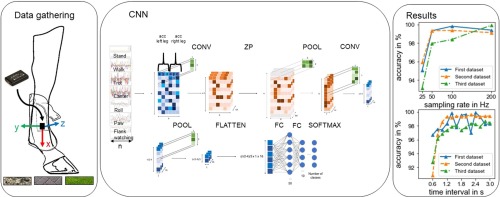
**f. Conclusion of the final model**

Throughout this prepared report, a number of different machine learning regression models have been evaluated to predict the outcome (died, authorized, lived) of horses using horse health status. From pre-processing processes as well as data exploration to model training and evaluation, it is concluded that the XGBoost model is the best model according to the comparisons based on the factors of Accuracy and F1 score related to each model. In addition to the XGBoost model, when compared with the other selected models, the main metric for predicting the outcome of the horses, the XGBoost model has an average relatively low absolute value of 0.70454. Therefore, another thing that can be concluded from this is that the essential health factors such as age, pain level, pulse, abdominal division to predict the outcome of the horses, this prediction is based on the results obtained from different models for predictions made from test and sample submission data sets. Shows correct evidence for.

**Success of using deep learning techniques in the predict health outcomes of horses**

Automatic Horse Activity Recognition by Convolutional Neural Networks Using Accelerometer Data

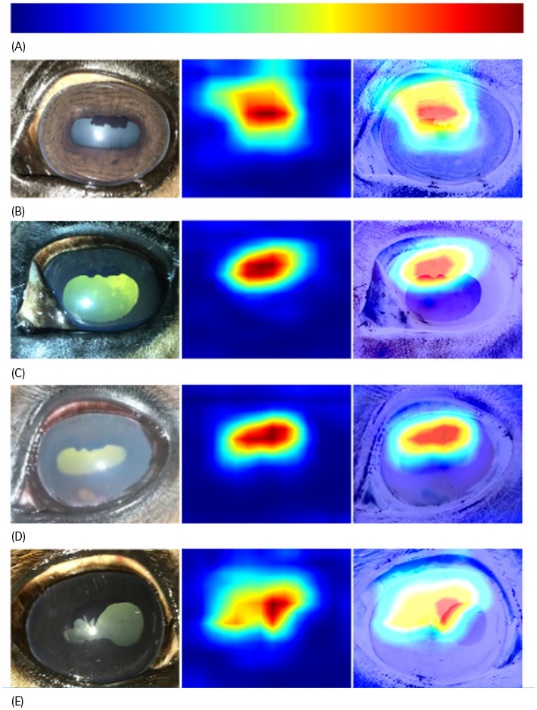
According to this report, observational behavior of horses is expected to provide important information about the health of horses. Here, the CNN model is expected to be used for tasks such as image classification that automatically extracts features using acceleration data to identify the behavior patterns of horses. Here, 7 horse activities were proposed to be classified based on deep learning approaches to identify horse activities. For the accuracy of horse activity detection, 6 test datasets were compiled based on the effect of the russified data sampling frequency, ground type and time series length factors, and 3 types of datasets are used to evaluate the model for the final conclusion. And the effect of a subject on the recognition accuracy is validated on a model with 50Hz sensor data processed at different sizes. Accordingly, at a sampling rate of 25Hz and a time interval of 2.1 seconds, 99% accuracy, at 69Hz and 2.4 seconds, four blind horse behaviors with the same accuracy, a sampling rate of 200Hz and 25Hz reduced accuracy by 4.75% and an accuracy of 0.6 seconds from 3 seconds. Decreases by 5.27%. Also the data normalization capability is valid on 50Hz data. This leads to a mean classification accuracy of 97.84% and 96.10% in ten-fold validation between a lame horse and a pony. And here it shows a 0.24% decrease in the accuracy of using the data from the baseline with one sensitivity. According to this report, it can be concluded that 3 test data sets can achieve 99% accuracy using CNN model with sampling rate of 2.1-time interval and 25 Hz. (Computers and Electronics in Agriculture, n.d.)



*Figure 114*

Artificial intelligence highlights equine uveitis as a tool for differentiating equine eye diseases

Nowadays, with the advancement of technology, the trend of experimenting in making new tools has increased. According to this report, the implementation of a mobile tool related to the eye diagnosis of uveitis, which can even lead to blindness in horses, and the visual classification system considers whether the horse exhibits uveitis or other eye diseases. In this way, the eyes affected by uveitis can be detected early and the eyes can be protected. Here, 2346 images of horses' eyes were trained by CNN or Convolutional Neural Network model and those images were increased to 9384 features. 261 unmodified images are used for the performance of the network being trained. This draws attention to less important areas of the horse's eye. It shows 99.82% accuracy on trained data and 96.66% accuracy on validated data in distinguishing equine eye disease from uveitis, other eye diseases, and healthy across 3 outcomes. (Anna May, n.d.)

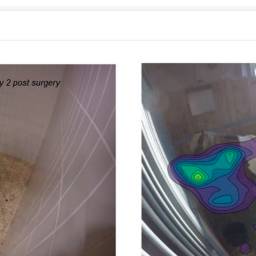


*Figure 115*

Development and validation of an automated video tracking model for stable horses

This report examines whether horses are in good or poor health by observing their behavior. In this report it has been decided that an automated video tracking model for horses should be investigated. The primary objective is to improve the ability to automatically identify horses, assess horse pain and well-being. The basic analysis done here provides the opportunity to build an algorithm to automatically observe the behavior patterns of horses. Here the video analysis process has been done by labeling 21342 video files with a duration of 11.52 minutes. Here, Convolutional neural network Loopy has been used to automatically predict the video parts, and by considering the nose, withers and tail as the main points, the results have been obtained that there is a sensitivity of more than 80% and an error rate of 7%.

(Nuray kil, n.d.)



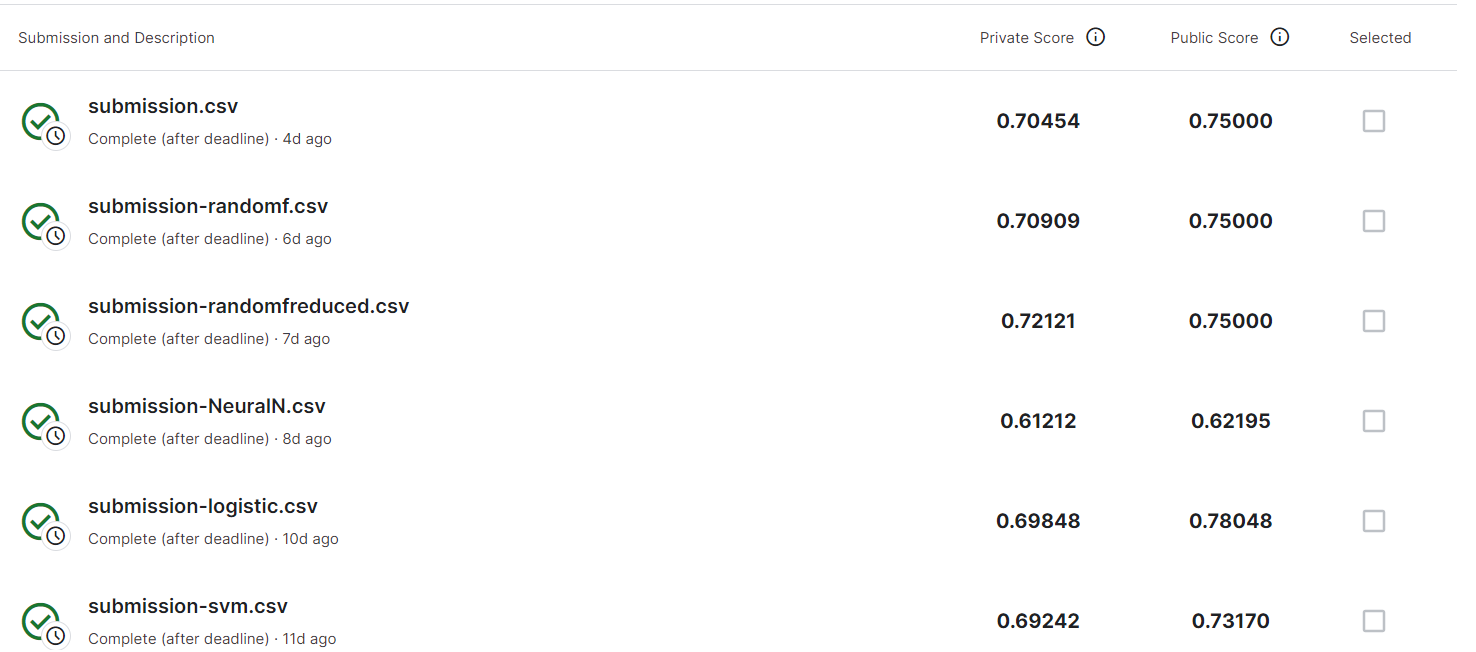
*Figure 116*

Appendix

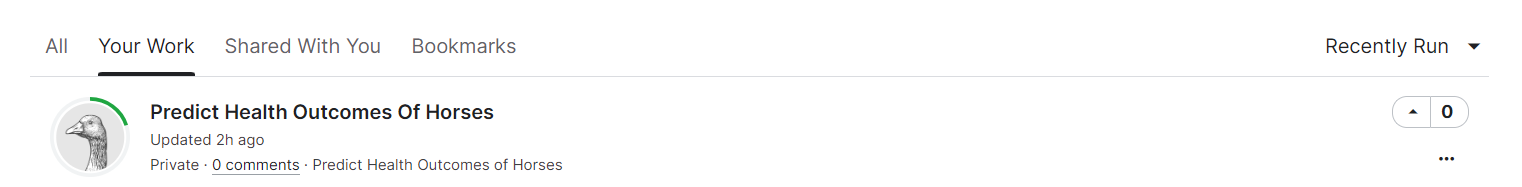
Kaggle Competition - Predict Health outcomes of Horses

Competition Link - <https://www.kaggle.com/competitions/playground-series-s3e22/overview>

Screenshots related to the competition.

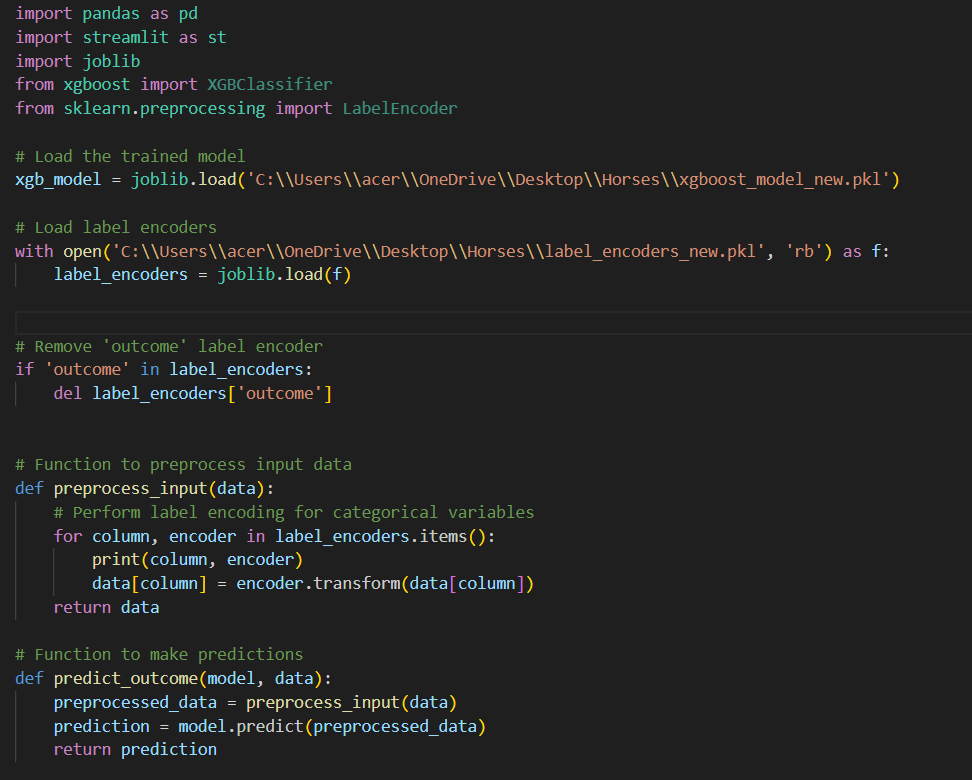


*Figure 117*



*Figure 118*

Software Artifact - Health outcomes of horses predictor



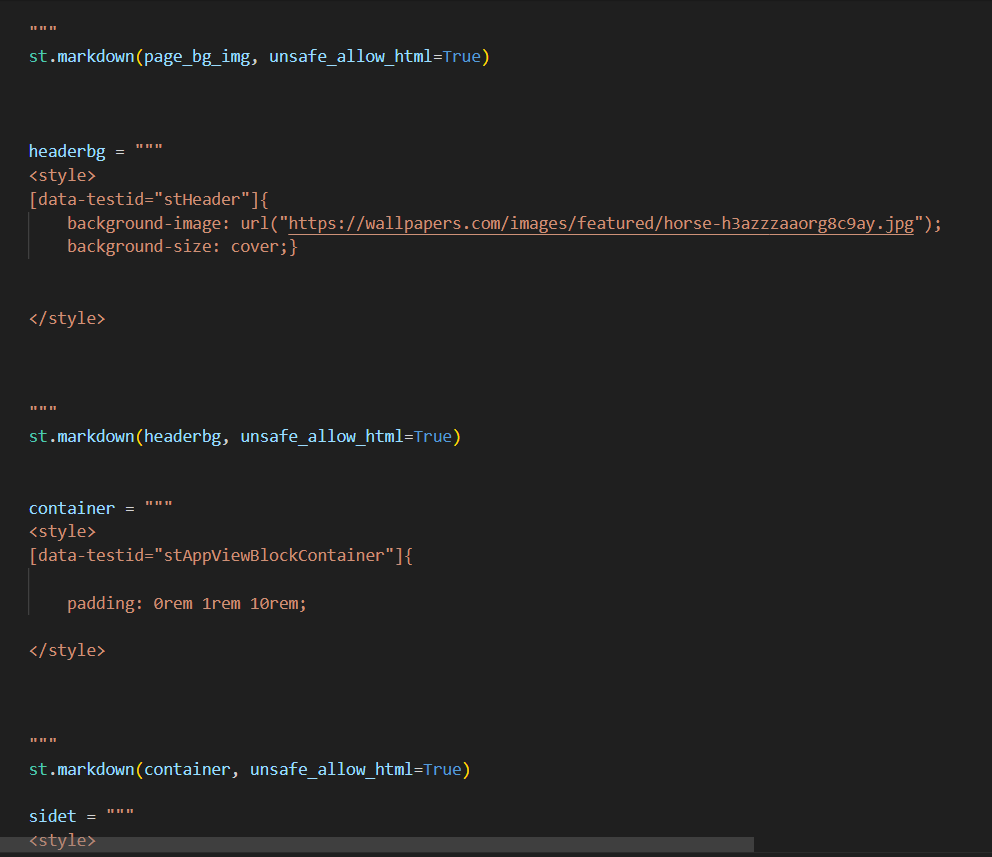
*Figure 119*



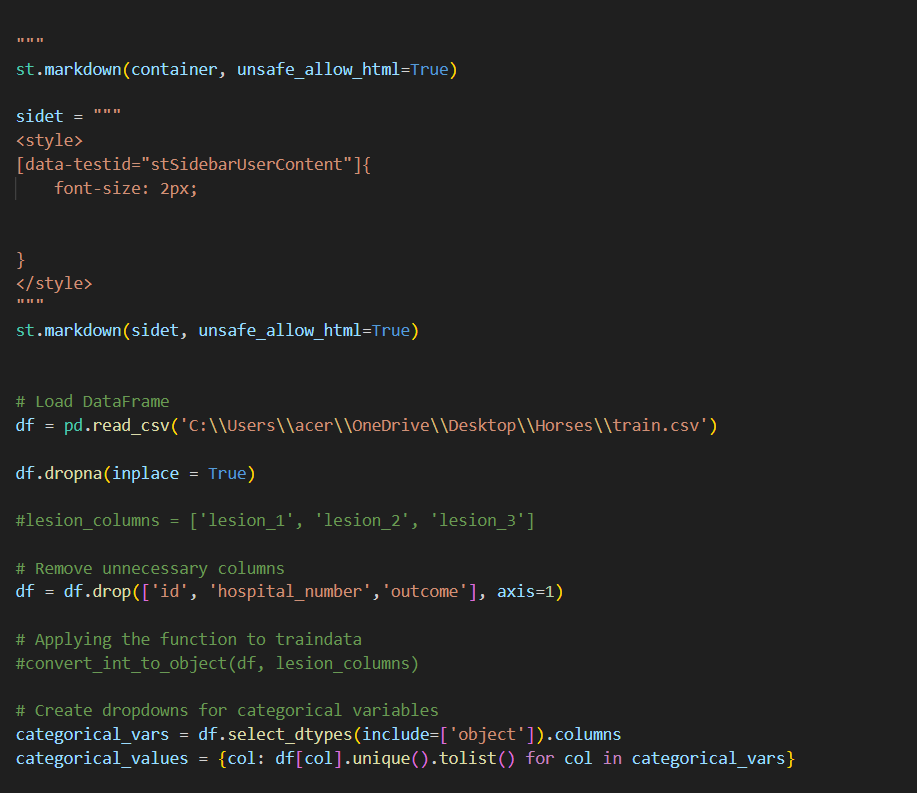
*Figure 120*



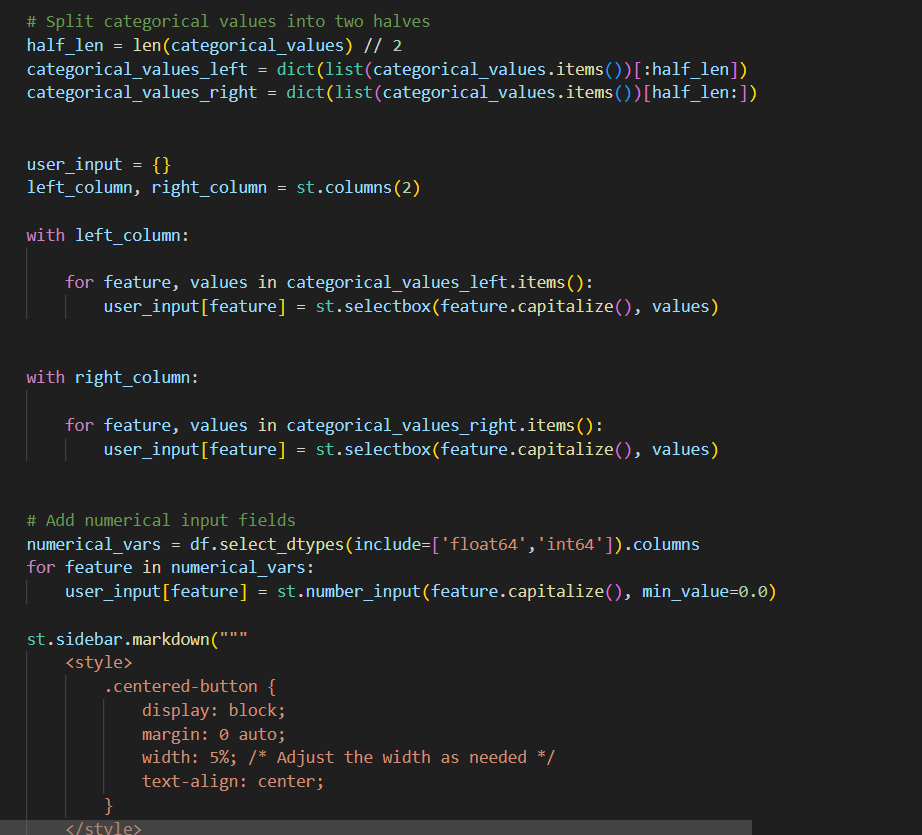
*Figure 121*



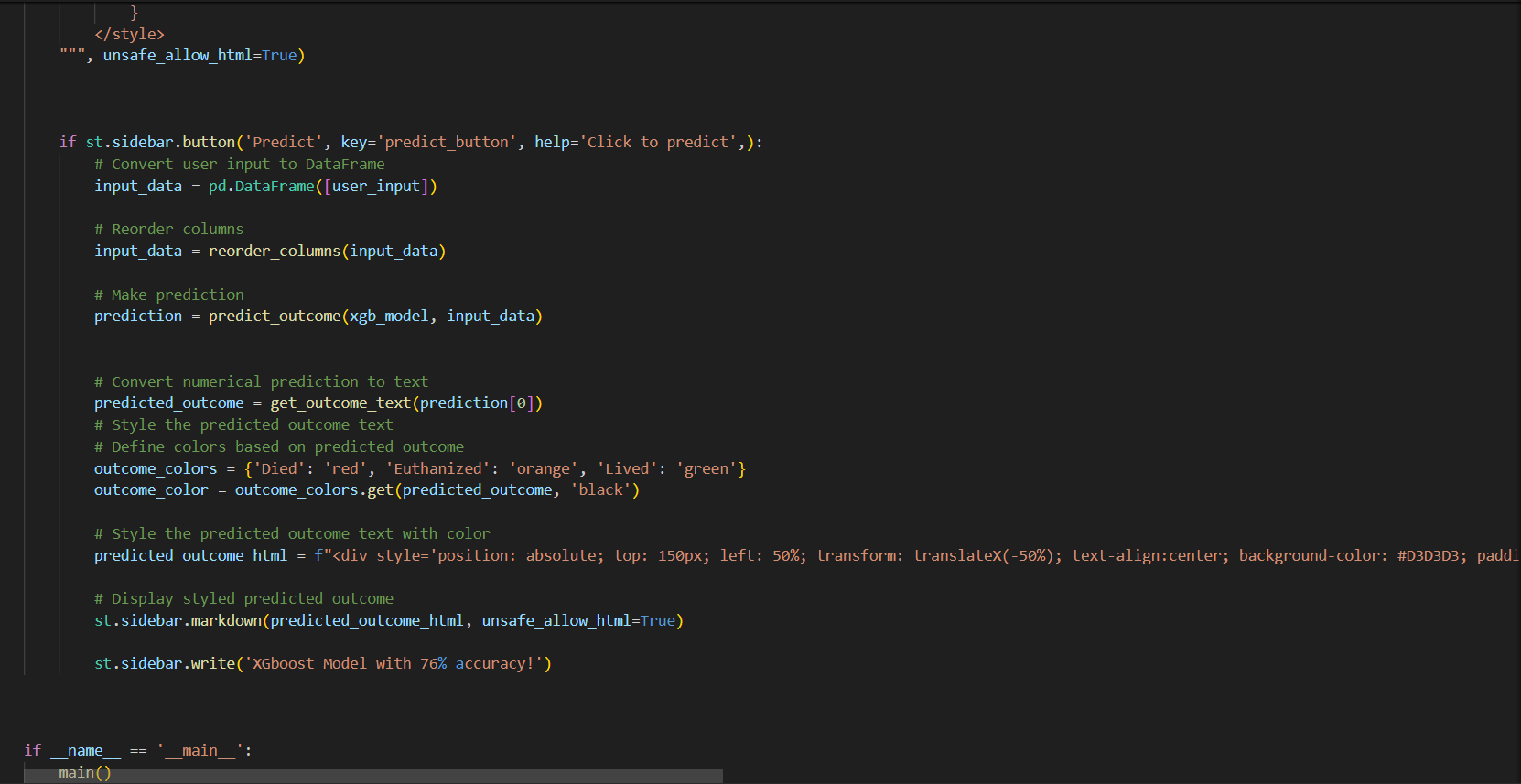
*Figure 122*



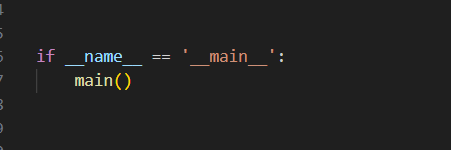
*Figure 123*



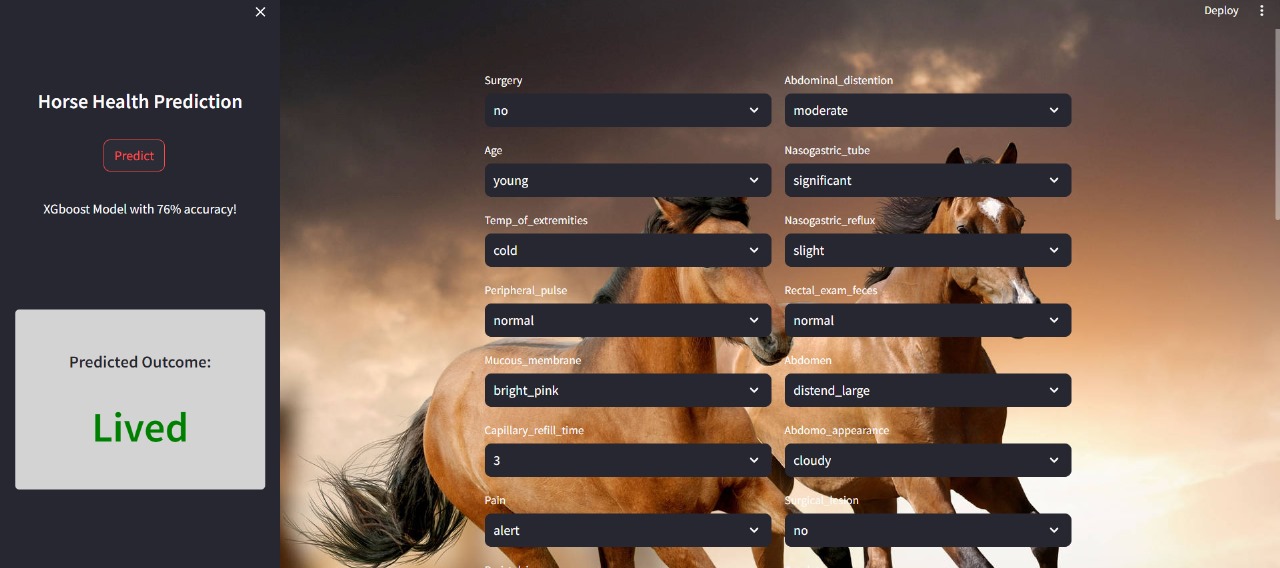
*Figure 124*



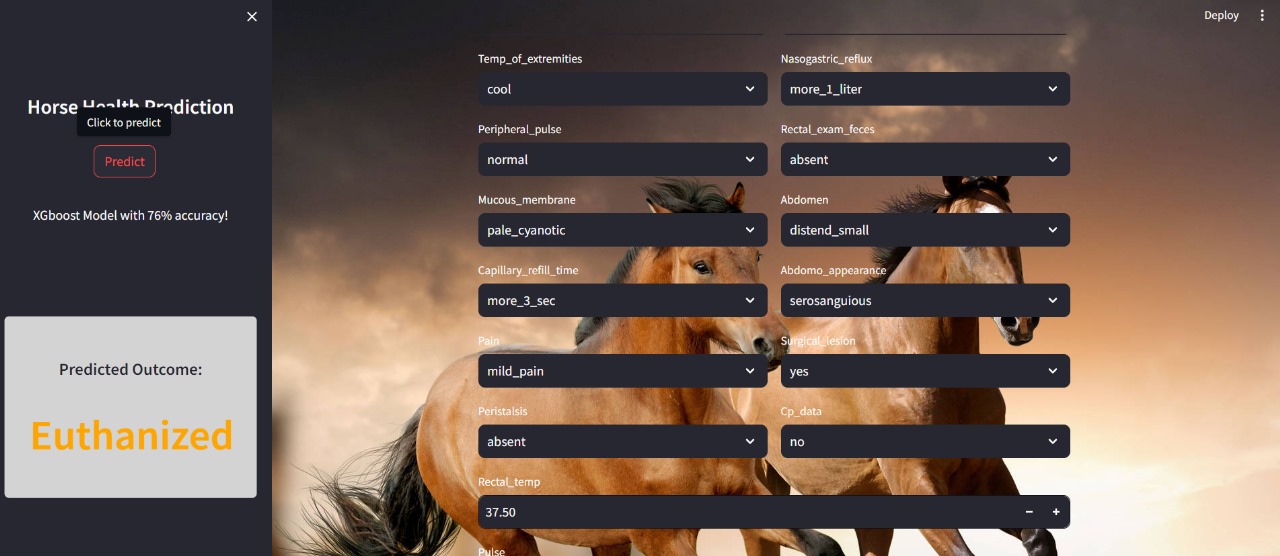
*Figure 125*



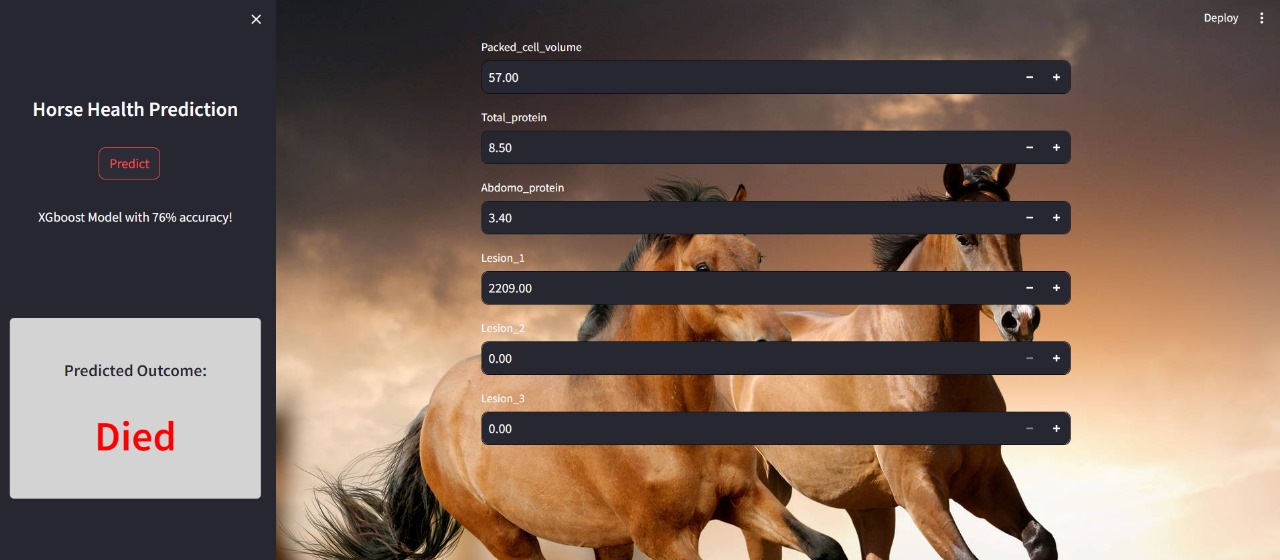
*Figure 126*



*Figure 127*



*Figure 128*



*Figure 129*

# References

Amazon Web Services, Inc., 2024. *What is Deep Learning?.* [Online]   
Available at: https://aws.amazon.com/what-is/deep-learning/#:~:text=Deep%20learning%20is%20a%20method,produce%20accurate%20insights%20and%20predictions

Anna May, S. G.-M. T. M. W. E., n.d. *Artificial intelligence as a tool to aid in the differentiation of equine ophthalmic diseases with an emphasis on equine uveitis.* [Online]   
Available at: https://beva.onlinelibrary.wiley.com/doi/full/10.1111/evj.13528  
[Accessed 28 Octomber 2021].

Computers and Electronics in Agriculture, n.d. *Automatic equine activity detection by convolutional neural networks using accelerometer data.* [Online]   
Available at: https://www.sciencedirect.com/science/article/abs/pii/S0168169919319283  
[Accessed January 2020].

Journal of Equine Veterinary Science, n.d. *Using Artificial Intelligence to Predict Survivability Likelihood and Need for Surgery in Horses Presented With Acute Abdomen (Colic).* [Online]   
Available at: https://www.sciencedirect.com/science/article/abs/pii/S0737080620300642  
[Accessed July 2020].

Lisowski, E., n.d. *Deep Learning Architecture.* [Online]   
Available at: https://addepto.com/blog/deep-learning-architecture/  
[Accessed 21 July 2020].

Nuray kil, K. E. U. A., n.d. *Development and Validation of an Automated Video Tracking Model for Stabled Horses.* [Online]   
Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7760072/  
[Accessed 30 November 2020].

Sundaray, S., n.d. *History of Deep Learning.* [Online]   
Available at: https://medium.com/@sreyan806/history-of-deep-learning-c176e2d3cddf  
[Accessed 16 July 2023].

