e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 5, May 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

International Journal Of Multidisciplinary Research In Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 5, May 2024 |

| DOI:10.15680/IJMRSET.2024.0705143 |

Python -Powered Predictive Analysis in Cardio-Insights

G S Naveen Kumar, Jagannadha Rao D B, Chaumwal Priyanka Radheshyam, M D.Salauddin , L. Sai Swapnik

Associate Professor & Head, Department of Data Science and IT, Malla Reddy University, Hyderabad, Telangana, India[1]

Associate Professor, Department of Data Science, Malla Reddy University, Hyderabad, Telangana, India[2]

Assistant Professor, Department of Data Science, Malla Reddy University, Hyderabad, Telangana, India[3]

B. Tech, Department of Data Science, Malla Reddy University, Hyderabad, Telangana, India[4 5]

ABSTARCT: Cardio-Insights is a comprehensive analytical project leveraging Python to unravel the complex landscape of cardiovascular health. By scrutinizing a diverse dataset encompassing demographic and lifestyle factors such as age, gender, blood pressure, cholesterol levels, and more, the initiative aims to identify key determinants influencing the prevalence of cardiovascular diseases. Using Python's powerful libraries, including Pandas, NumPy, Matplotlib, Seaborn and plotly the project seamlessly processes and models the dataset. Employing advanced analytical techniques, including Exploratory Data Analysis Cardio-Insights seeks to construct predictive models that forecast cardiovascular risks based on an individual's unique set of health-related attributes. The anticipated outcomes hold promise for healthcare professionals, policymakers, and individuals alike, providing valuable insights into the intricate dynamics shaping cardiovascular health. The Python-centric analytical approach not only ensures efficiency and accuracy in handling large datasets but also fosters reproducibility and transparency, encouraging collaborative research and further exploration in the field. Cardio-Insights stands as a testament to the transformative potential of interdisciplinary collaboration, utilizing technology to unlock critical insights that contribute to the proactive management of heart health and the advancement of public health outcomes.

KEYWORDS: Cardio-vascular disease, Machine Learning, Data Analytics.

I. INTRODUCTION

Human body is made up of different organs, but the organ that keeps a person alive is the heart. It is important that the heart must pump blood throughout the body. In recent times, the functioning of the heart is constrained by various factors, either it may be genetics or because of the habits a person has which affect the heart. Cardiovascular health must be maintained to live a long and healthy life. Unfortunately, people around the world are facing cardiovascular diseases. Any research that can help in diagnosing the disease before much damage is done would save people's money and, most importantly, their lives. Data analytics and machine learning in this field are potentially useful and fruitful tools which can help to derive insights, trends, and patterns of cardiovascular diseases. Data analytics is the scientific process of transforming data into meaningful insights for making better decisions. Machine learning algorithms can predict the probability of a person having heart disease.

II. LITERATURE SURVEY

Research has been done in this filed and people have found methods to predict the heart disease using machine learning algorithms. In one of the papers, a survey is done which includes using of one or more algorithms of data mining techniques to predict the rate of heart disease [1]. A survey has been presented in the form of a paper which analyses performance of various models based on machine learning algorithms and techniques [2]. In one of the papers XG boost (machine leaning model) was used for heart disease prediction [3]. In another paper, python with matplotlib and seaborn libraries were used to derive insights for decision making on cardio diseases [4].

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 5, May 2024 |

| DOI:10.15680/IJMRSET.2024.0705143 |

III. METHODOLGY & IMPLEMENTATION

The above figure [Fig-1] is our methodology. First we obtained the dataset [5] and performed exploratory data analysis using python's pandas so that we can get meaningful insights and relation between variables. We then performed descriptive analytics using python's matplotlib and seaborn libraries. These libraries are majorly used for visualization purposes[6]. Through the visuals we understood the overall population variations in categories of Smoking, Alcohol Consumption, Blood pressure levels, Age, Weight, Gender, Cholesterol level etc.

We used machine learning algorithm, XGBoost to perform predictive analytics on the dataset and created a correlation matrix to understand how one variable is correlated to another.

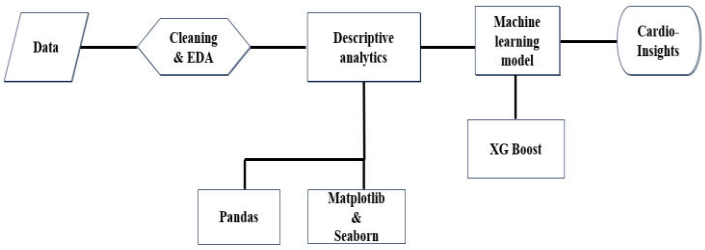

Fig-1: Methodology

The below table [Table-1] is our dataset consisting of 70,000 records and 12 variables. Here is the glimpse of our raw dataset before any formatting

We obtained plots on different variables, the countplot [Fig-2] here is used for indicating all the male population who don't consume alcohol with the ages between 30-65 years.

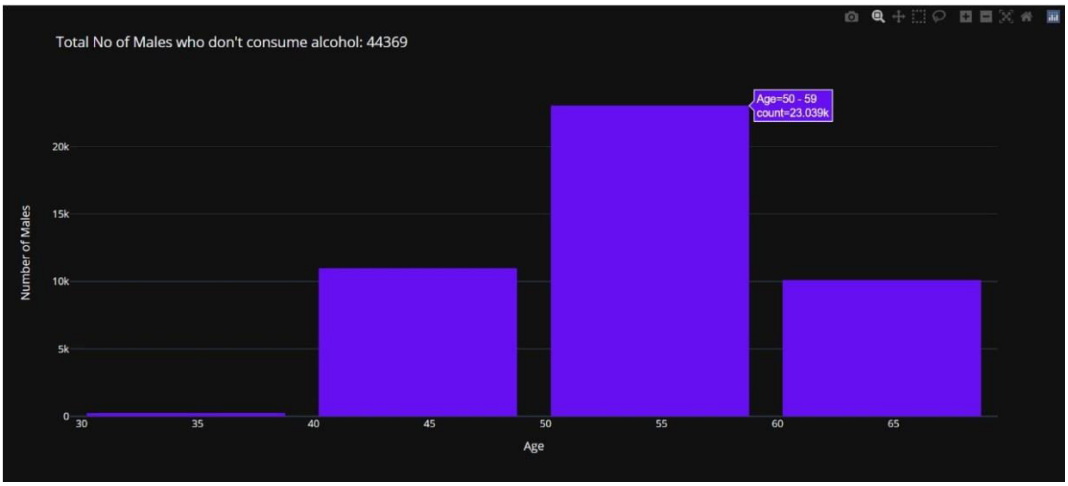

Fig-2: No. of males who don't consume alcohol

age	gender	height	weight	bp_hi	bp_lo	cholesterol	gluc	smoke	alco	active	cardio
18393	2	168	62	110	80	1	1	0	0	1	0

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 5, May 2024 |

| DOI:10.15680/IJMRSET.2024.0705143 |

20228	1	156	85	140	90	3	1	0	0	1	1
18857	1	165	64	130	70	3	1	0	0	0	1
17623	2	169	82	150	100	1	1	0	0	1	1
17474	1	156	56	100	60	1	1	0	0	0	0
21914	1	151	67	120	80	2	2	0	0	0	0
22113	1	157	93	130	80	3	1	0	0	1	0
22584	2	178	95	130	90	3	3	0	0	1	1
17668	1	158	71	110	70	1	1	0	0	1	0
19834	1	164	68	110	60	1	1	0	0	0	0
22530	1	169	80	120	80	1	1	0	0	1	0
18815	2	173	60	120	80	1	1	0	0	1	0
14791	2	165	60	120	80	1	1	0	0	0	0
19809	1	158	78	110	70	1	1	0	0	1	0
14532	2	181	95	130	90	1	1	1	1	1	0
16782	2	172	112	120	80	1	1	0	0	0	1
21296	1	170	75	130	70	1	1	0	0	0	0
16747	1	158	52	110	70	1	3	0	0	1	0
17482	1	154	68	100	70	1	1	0	0	0	0
21755	2	162	56	120	70	1	1	1	0	1	0
19778	2	163	83	120	80	1	1	0	0	1	0
21413	1	157	69	130	80	1	1	0	0	1	0

Table-1: Dataset

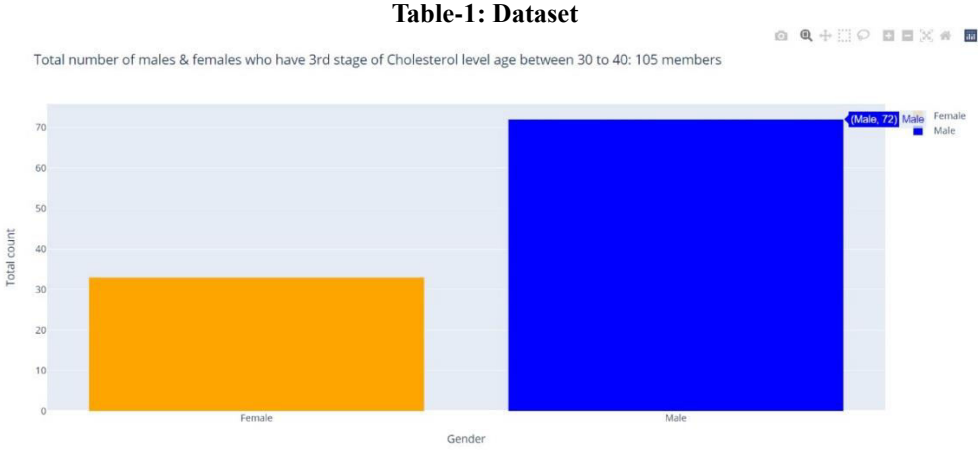

Fig-3: 3rd stage of cholesterol level

The above countplot **[Fig-3]** on cholesterol level indicates the total number of people who has 3rd stage of cholesterol level of both male and female population

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 5, May 2024 |

| DOI:10.15680/IJMRSET.2024.0705143 |

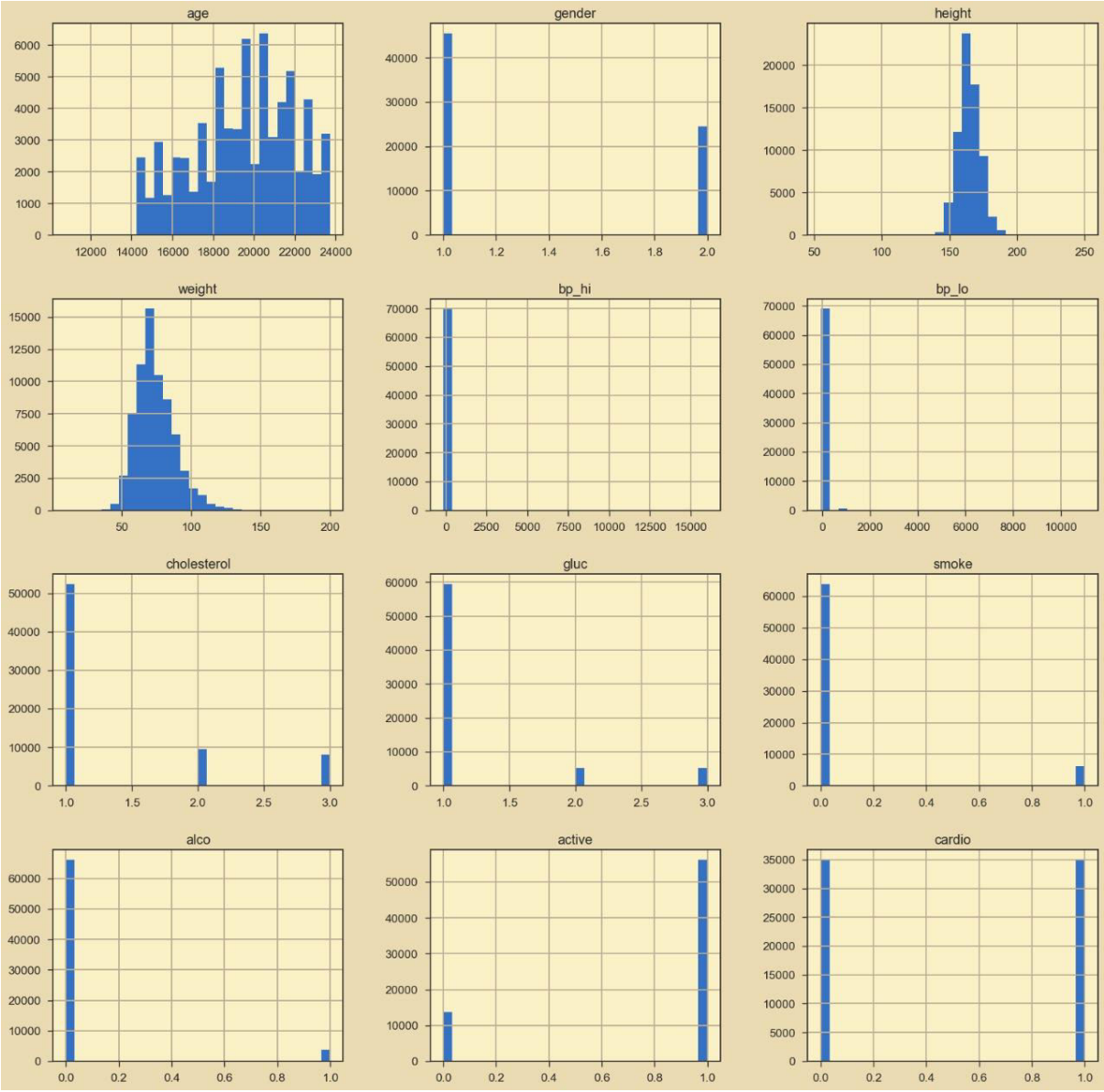

Fig-4: Histograms of overall variables

The above histogram [Fig-4] shows the observations frequencies of all the attributes. After checking that the data is balanced, the correlation between the data is found out and is plotted as a heat map using the Seaborn library. The below figure [Fig-5] is the correlation matrix.

The colours represent the strength and direction of the correlation[7]. The scale on the right side of the image shows that red tones indicate a positive correlation and yellow tones indicate a negative correlation[8]. The intensity of the colour corresponds to the strength of the correlation.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 5, May 2024 |

| DOI:10.15680/IJMRSET.2024.0705143 |

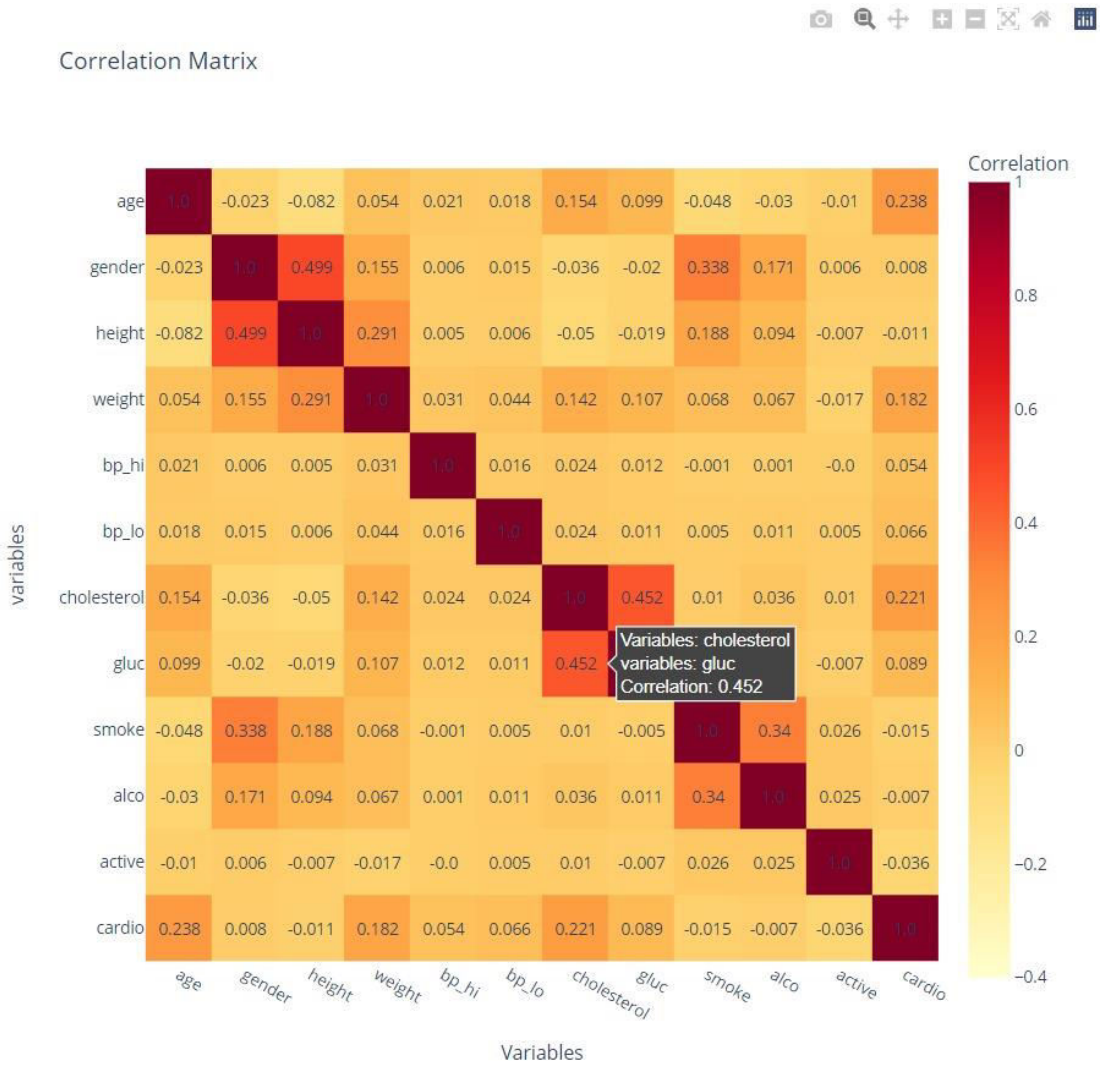

Fig-5: Correlation matrix of all attributes

IV. RESULTS

The results are observed from the XGBoost model that gives the output in the form of a confusion matrix. **Confusion matrix** which is a visualization tool typically used in machine learning to evaluate the performance of a classification algorithm. The matrix compares the actual target values with those predicted by the machine learning model, allowing you to see how well the model is performing in terms of correctly and incorrectly predicting outcomes.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 5, May 2024 |

| DOI:10.15680/IJMRSET.2024.0705143 |

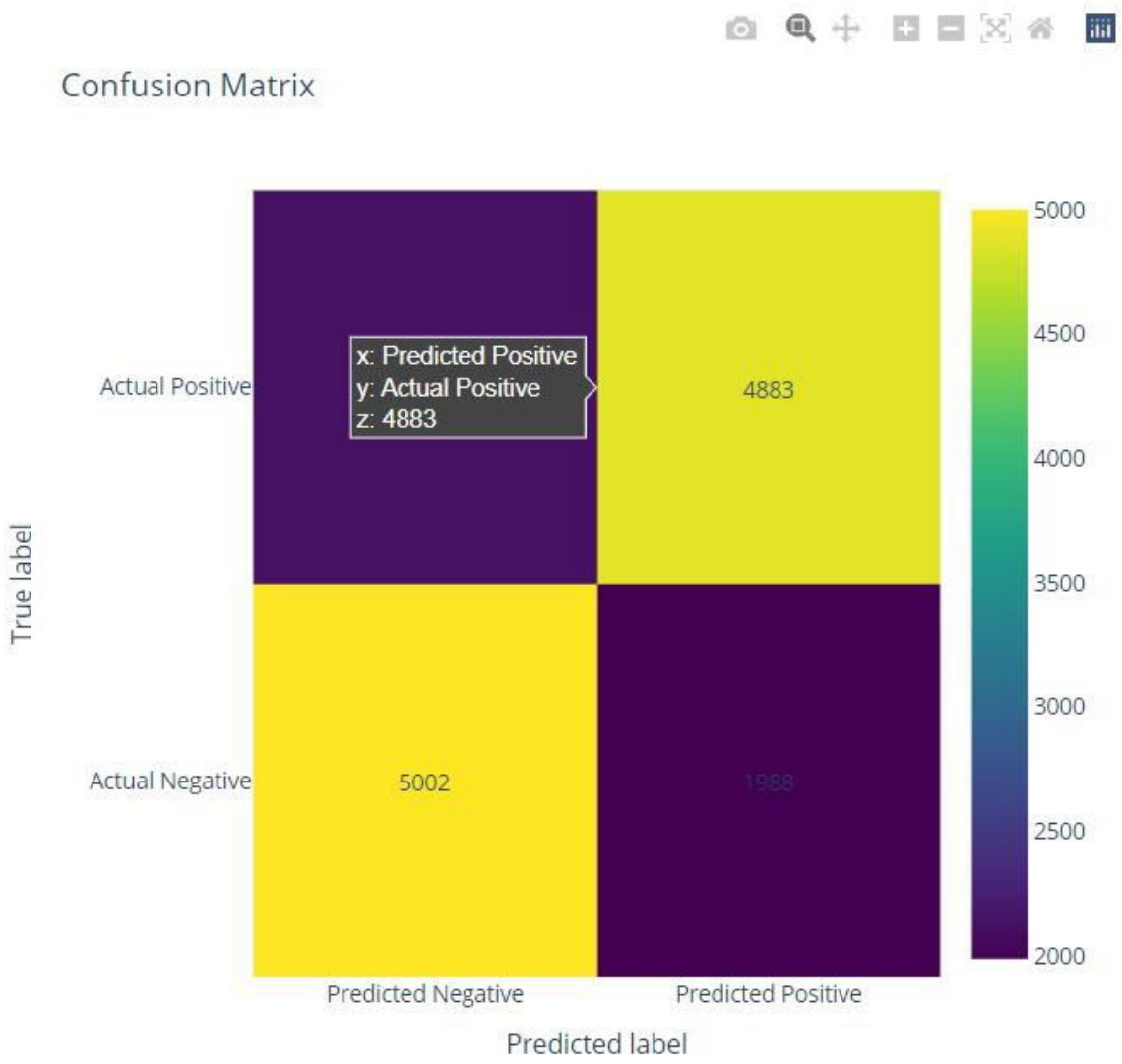

Fig-6: Confusion Matrix

The above figure [Fig-6] is the confusion matrix. Where,

- The x-axis represents the predicted labels (or classifications) that the model has output.
- The y-axis represents the actual true labels as they are in the dataset.
- 4,883 individuals were correctly predicted to have cardiovascular disease.
- 5,002 individuals were incorrectly predicted to have cardiovascular disease (Type I error).
- 2,127 individuals were incorrectly predicted not to have cardiovascular disease (Type II error).

The accuracy for the model is calculated as

Accuracy = (TP + TN) / (TP + TN + FP + FN) **-1**

Where TP is True Positive Values

TN is True Negative Values

FP is False Positive Values

FN is False Negative Values

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 5, May 2024 |

| DOI:10.15680/IJMRSET.2024.0705143 |

```
In [18]: from sklearn.metrics import classification_report
         print(classification_report(y_test, y_predict))
         # precision is the ratio of TP/(TP+FP)
         # recall is the ratio of TP/(TP+FN)
         # F-beta score can be interpreted as a weighted harmonic mean of the precision and recall
         # where an F-beta score reaches its best value at 1 and worst score at 0.
```

	precision	recall	f1-score	support
0	0.68	0.80	0.74	7027
1	0.76	0.63	0.69	6973
accuracy			0.71	14000
macro avg	0.72	0.71	0.71	14000
weighted avg	0.72	0.71	0.71	14000

Fig-7: Performance metrics

The [Fig-7] shows the performance metrics of XGBoost model and has the accuracy of approximately 71%.

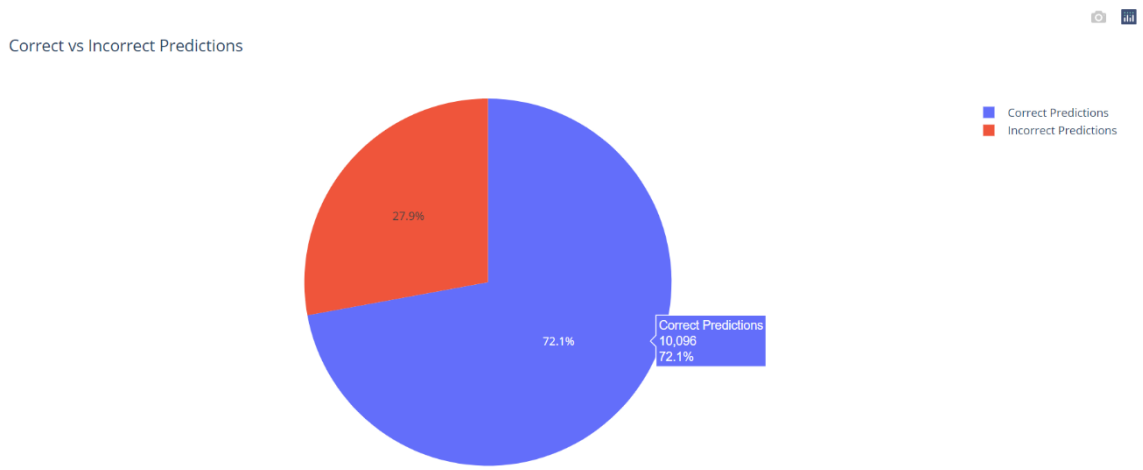

Fig-8: Pie Chart For Final Predictions

The [Fig-8] Pie Chart For Final Predictions of the XGBoost model 72.1% were Predicted as True and 27.9% were predicted as False.

V. CONCLUSION

In Conclusion, the machine learning models are getting more accurate in predicting cardio vascular diseases which can be said to be the most prominent problems in the society. As the work in the field of machine learning is being done, there soon may be new and efficient ways to improve the prediction of heart disease. The machine learning algorithm in our project has used all available attributes well and performed better. The conclusion can be drawn that data analytics and machine learning can be able to predict the chances of a person going to get heart disease by correlating their habits.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 5, May 2024 |

| DOI:10.15680/IJMRSET.2024.0705143 |

REFERENCES

[1] Kaur, B., & Singh, W. (2014). Review on heart disease prediction system using data mining techniques. International journal on recent and innovation trends in computing and communication, 2(10), 3003-3008
[2] Ramalingam, V. V., Dandapath, A., & Raja, M. K. (2018). Heart disease prediction using machine learning techniques: a survey. International Journal of Engineering & Technology, 7(2.8), 684- 687.
[3] Jason Brownlee. XGBoost with python.
[4] Fabian Nelli. Python Data Analytics.
[5] Coursera
[6] Wijdicks EF. The history of neurocritical care. Handb Clin Neurol. 2017;140:3–14.
[7] Rincon F, Mayer SA. Neurocritical care: a distinct discipline? Curr Opin Crit Care. 2007;13(2):115–21.
[8] Samuels O, Webb A, Culler S, Martin K, Barrow D. Impact of a dedicated neurocritical care team in treating patients with aneurysmal subarachnoid hemorrhage. Neurocrit Care. 2011;14(3):334–40.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com