International Research Journal of Education and Technology
ISSN 2581-7795

Implementation of Logistic Regression in Healthcare

Ms.Rupali.A.Zamare [1], Dr.Pankaj M.Agarkar [2], Dr.Santosh Kumar Yadav [3]

[1] *PhD Scholar, JJT University,Rajasthan*
[2] *Co-Guide, D.Y.PatilSchool of Engg,Lohagaon,Pune*
[3] *Guide, JJT University,Rajasthan*

Abstract - *Machine learning is helpful in health industry for detection and prevention of chronic diseases, thus making cure fast and easy for diagnosis .Health Industry is affected due to major death ratio due to heart diseases, which will be avoided as well as overcome by Machine learning. Machine learning having supervised and unsupervised learning techniques for classification. The support Vector machine, Random Forest, Logistic regression, Nearest Neural Network and Artificial Neural network are giving the results in the form of accuracy, recall and precision. Logistic Regression working on the logit function in mathematics use to predict the target variable and it is a supervised learning technique.*

Key Words **Support Vector Machine, Random Forest, Logistic Regression and Neural Network**

1.INTRODUCTION *(Size 11 , cambria font)*

Healthcare Industry is continuously searching for the cure of chronic diseases. According to the mortality rate, world is losing 40 % of population due to heart disease. Regression analysis of relationship between dependent and independent variables. Logistic regression is supervised learning, it analyse the relationship between categorical dependent variable and one or more independent variable. Logistic regression is similar to linear regression or linear model.[1],[8],[9]

Types of classification

1. Logistic Regression
2. Naïve Bayes
3. Stochastic Gradient Descent
4. K-Nearest Neighbours
5. Decision Tree
6. Random Forest
7. Support Vector Machine

Classification can be performed on structured or unstructured data. Classification is a technique where categorization of data is done into a given.

number of classes. The main goal of a classification problem is to identify the category or class.

- **Initialize** the classifier which is used for implementation.
- **Train the classifier**
- **Prediction of the target**.
- **Evaluate** the model of classifier[5]

1. METHODS IN CLASSIFICATION

The methods or way are also called as types of classification.

1.1. Logistic Regression

Logistic regression is a machine learning algorithm for classification. It concentrates on the influence of several independent variables on a single outcome variable.

```
from sklearn.linear_model import LogisticRegression
lr =  LogisticRegression()
lr.fit(x_train, y_train)
y_pred=lr.predict(x_test)
```

1.2.Naive Bayes algorithm

International Research Journal of Education and Technology
ISSN 2581-7795

Naive Bayes algorithm based on Bayes' theorem with the assumption of independence between every pair of features. This algorithm requires a small amount of training data to estimate the necessary parameters.

```
from sklearn.naive_bayes import GaussianNB
nb =  GaussianNB()
nb.fit(x_train, y_train)
y_pred=nb.predict(x_test)
```

3.Stochastic Gradient Descent

Stochastic gradient descent is a simple and very efficient approach to fit linear models. It is particularly useful when the number of samples is very large.

```
from sklearn.linear_model import SGDClassifier
sgd =  SGDClassifier(loss='modified_huber', shuffle=True,random_state=101)
sgd.fit(x_train, y_train)
y_pred=sgd.predict(x_test)
```

1.3.K-Nearest Neighbours

Classification is computed from a simple majority vote of the k nearest neighbors of each point. This algorithm is simple to implement, robust to noisy training data, and effective if training data is large.

```
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier(n_neighbors=15)
knn.fit(x_train,y_train)
y_pred=knn.predict(x_test)
```

1.4. Decision Tree

A decision tree produces a sequence of rules that can be used to classify the data. Decision Tree is simple to understand and visualize, requires little data preparation, and can handle both numerical and categorical data.

```
from sklearn.tree import DecisionTreeClassifier
dtree = DecisionTreeClassifier(max_depth=10, random_state=101,
                               max_features = None, min_samples_leaf = 15)
dtree.fit(x_train, y_train)
y_pred=dtree.predict(x_test)
```

1.5. Random Forest

Random forest classifier is a meta-estimator that fits a number of decision trees on various sub-samples of datasets and uses average to improve the predictive accuracy of the model and controls over-fitting.

```
from sklearn.ensemble import RandomForestClassifier
rfm = RandomForestClassifier(n_estimators=70, oob_score=True, n_jobs=-1,
                             random_state=101, max_features = None, min_samples_leaf = 30)
rfm.fit(x_train, y_train)
y_pred=rfm.predict(x_test)
```

1.6.Support Vector Machine

Support vector machine is a representation of the training data as points in space separated into categories by a clear gap that is as wide as possible.

```
from sklearn.svm import SVC
svm = SVC(kernel="linear", C=0.025,random_state=101)
svm.fit(x_train, y_train)
y_pred=svm.predict(x_test)
```

[6][7][8]

1.7.Model:Logistic Regression

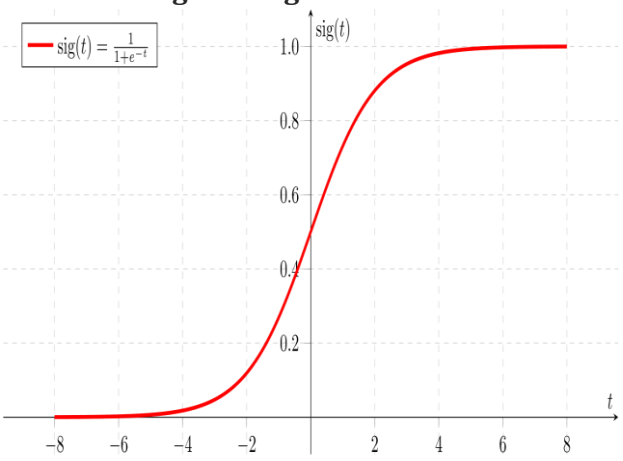

Fig.1.Sigmoid Function

The corresponding output of the sigmoid function is a number between 0 and 1. The middle value is considered as threshold to establish what belong to the class 1 and to the class 0.[8]

Thus Linear regression predict class probabilities is a modelling choice, just like it's a modelling choice to predict quantitative variables with linear regression.[10]

International Research Journal of Education and Technology
ISSN 2581-7795

2. RESULTS AND DISCUSSION

Medical data is tremendous to maintain and use in the efficient manner with different models of classification. The prediction of heart disease remains unattended if we rely on a single method or model. According to the classification models, the negative class consists of no heart disease (binary 0) and positive class of having heart disease (binary 1).Neural networks are also contributing in classification by applying concepts of brain working techniques. [1]
Heart failure is leading risk to health. The classification and existing prediction models requires features for application of selected model or emerging difficulties for estimation. The variables are available in

Electronic heath records (EHR) and data consists of 234 records of heart failure cases. The tools which are existing to cure consists of continuous medications,implanting devices ,transplantations or providing end of life care.[2]

Seatle heart failure model (SHFM) get some features which are different to estimate for example hazard ratio for certain Heart Failure medications and devices.Some features are not used in model like ethnicity ,body mass Index, certain lab measurements. The new model named as Temporal reflected logistic regression (TRLR) updates the parameters dynamically as per availability.[2]

Decision making for medical professionals is done by application of classification like Random forest, logistic regression and Bayesian classification. Dataset of 18 features and 1990 observations is used.Implementation done in RStudio platform giving accuracy as follows[3]:-

Sr. No	Algorithm	Accuracy
1	Random Forest	92.44
2	Logistic Regression	59.7%
3	Naive Bayesian	61.96%

Table 1.Algorithms and accuracy

The data patterns and features help to analyse the medical data. The heart failure prediction is done for young as well as old age human being. Observations of 4838 records is done consisting of 17 features.

Classification algorithms like logistic regression,decision trees and nueral networks are analysed on 303 subjects[5].

Sr.No	Features
1	Age
2	Gender
3	Chest Pain Type
4	Blood Pressure
5	Cholesterol
6	Fasting Blood Sugar
7	Resting ECG
8	Maximum Heart Rate
9	Thal

Table 2.Types of Features

Logistic Regression is a regression method for predicting a dichotomous dependent variable. Logistic Regression is applicable in conditions in which predictions are done depending on presence or absence of characteristics or outcome based values of set of predictor variables. In artificial neural network axons work to transmit nerve impulses from one neuron to another, whenever neurons are stimulated Multilayer Artificial Neural network is used with back propagation algorithm. Classification and Regression trees are used when quantitative and qualitative variable response is there. Thus leading to high performance modelling methods.[4]

Algorithms used	Sensitivity	Specificity	Accuracy
Neural Network	81.1%	78.7%	80.2%
Decision Tress	81.7%	76.0%	79.3%
Logistic Regression	81.2%	73.1%	77.7%

Table 3. Comparision of algorithm

Comparison of the optimal and non-optimal intervention combinations with Acute Ischemic Stroke (AIS) under matched states by applying conditional Logistic Regression(CLR) method. The method applied consist of pairing of states on Day1 or Day7 were matched as N:K while patients assigned under two groups depending on response of an optimal

combinations. Hence CLR is a valuable analytic technique in effective research.[5]

3.CONCLUSION

Logistic Regression proving beneficial for the medical data analysis in the case of heart disease.The Supervised algorithm gives better results in the form of accuracy .Implemented with Radom Forest,decision tree and having different conditions applied resulting into conditional logistic regression,thus helping the health industry to achieve success in predetection of Heart Disease.

REFERENCES

[1] Md. Jamil-Ur Rahman and et.al "Ensemble of Multiple Models For RobustIntelligent Heart Disease Prediction System", 4th International Conference on Electrical Engineering and Information & Communication Technology,IEEE2018.

[2] Mingjie Qian and et.al,"*Temporal Reflected Logistic Regression for Probabilistic Heart Failure Survival Score Prediction*", International Conference on Bioinformatics and Biomedicine 2017 IEEE.

[3] Thankgod Obasi and et.al, "Towards comparing and using Machine Learning techniques for detecting and predictingHeart Attack and Diseases" IEEE 2019.

[4] Jianxiong Cai12," Application of Conditional Logistic Regression Method in a Prospective Effectiveness Comparative Study Among Patients with Acute Ischemic Stroke"IEEE2011.

[5] Anchana Khemphila and et.al, "Comparing performances of logistic regression, decision trees , and neural networks for classifying heart disease patients"IEEE 2010.

[6] *https://analyticsindiamag.com/7-types-classification-algorithms/*.

[7] https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc.

[8] https://github.com/SSaishruthi/LogisticRegression_Vectorized_Implementation/blob/master/Logistic_Regression.ipynb

[9] https://www.sciencedirect.com/topics/medicine-and-dentistry/logistic-regression-analysis

[10] *https://www.stat.cmu.edu/~cshalizi/uADA/12/lectures/ch12.pdf*.