

For Acute Pancreatitis using Supervised Machine Learning Algorithms

*G. Karteek, #K. Yasudha

*PG Student, #Assistant Professor, Department of Computer Science, GIS, GITAM (Deemed to be University), Visakhapatnam, India, *gurugubeelikarteek@gmail.com, #yasudha.p@gmail.com

Abstract - The data present in healthcare industry is very huge and delicate which requires to be managed watchfully. There are multiple fatal diseases which grow rapidly all over the world pancreatitis is one among them. Medical professionals want a reliable prediction system to diagnose Pancreatitis. Getting useful information out of the data which has been examined using diverse perspective and various machine learning methods and grouping the required information is a bit difficult task. When various data mining methods are applied on a huge and accessible data which will definitely provide us with the required information to the users. Pancreatitis contributes to Infection, Kidney failure, Breathing problem, Diabetes, Malnutrition, Pancreatic cancer. So, mining the Pancreatitis data in efficient way is a crucial concern. An outcome feature has to be predicted using a dataset where the outcome may contain only two constants that is either 1 or 0. 0 refers to the sufferer having Acute Pancreatitis and 1 refers to the sufferer may have chronic pancreatitis. Thus, an outcome feature with exemplary accuracy has to be predicted using the test dataset and classification algorithms. In order to realize this data is very necessary and then diverse classification techniques can be experimented. Then a finest model can be preferred which gives the maximum accuracy among all others.

Keywords — Acute Pancreatitis, Supervised Machine Learning Algorithms, KNN Algorithm, Decision Tree Algorithm, Logistic Regression, Random Forest.

I. INTRODUCTION

Pancreatitis is pathologic condition in which pancreas becomes red, swollen, and painful. Pancreas sits behind stomach, near small intestine. Enzymes which help in digestion of food are released by the pancreas; also adjust how a human body administers glucose. Pancreatitis is considered as a chronic problem which occurs to a human body more often. Treatment will be built upon the severity of the disease if pancreatitis is intense or persistent. Indications of pancreatitis may comprise: pain that swaddle round the upper abdomen, hyper acidity, puking, abdominal irritation, unforeseen weight reduction, Inflation with a bulged abdomen and raise in temperature of the body. Pancreatitis is generally acute or chronic. Acute pancreatitis may be a main explanation for hospital admissions for gastrointestinal issues.

The arrival of intense pancreatitis is usually unexpected. The tenderness generally disappears within various days behind the medicaments begin, but some cases could require a hospital stay. Acute pancreatitis is much more common in adults than in children. The main reason behind intense pancreatitis in adults is Gallstones. This condition may also develop into chronic pancreatitis, especially if anyone does smoke or drink alcohol regularly. Persistent pancreatitis is tenderness of the pancreas which comes back consistently or occurs over a long period of time. Sufferers with persistent pancreatitis may also have everlasting injury to the pancreas and other complications. There is a development of a tissue named scar tissue which comes from continuous inflammation. Stable alcoholic consumption results in almost 70 percent of persistent pancreatitis in adults. Ancestral diseases, like CF, also can cause chronic pancreatitis in some people. Acute and chronic pancreatitis shares many of the same causes. Bile stones are the foremost basic reason behind intense pancreatitis. Bile stones are tiny, rigged masses that are formed by the gallbladder.

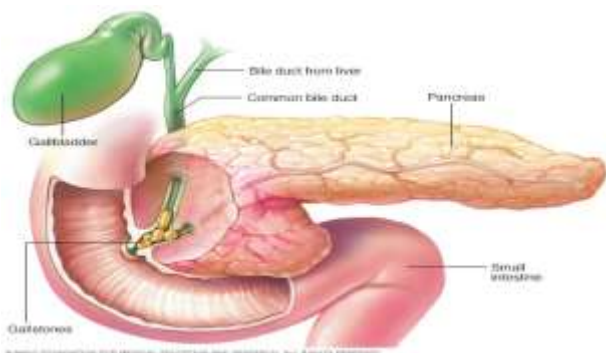


Fig.1. Acute Pancreatitis

II. ALGORITHMS DESCRIPTION

A. KNN Algorithm

KNN is one among the supervised machine learning algorithm which is put to use on classification and

regression prognostication obstacle. KNN algorithm makes use of feature indistinguishable to anticipate the constants of fresh data points which will be additionally allocate a value constructed on how firmly it is equivalent to the points in the training dataset. KNN algorithm contains all accessible cases and groups and groups fresh cases based on a homogeneous measure. It is broadly used in real life scenario as it is distribution free which represents that it doesn't make any underlined presumptions regarding distribution of data.

B. Logistic Regression Algorithm

Linear regression and logistic regression are alike, but the gigantic variation lies in what they are utilized for. Linear regression algorithms are utilized to anticipate or forecast values. But logistic regression is put to use on classification task. Logistical regression is a statistical framework that in its basic form which makes use of a logistic function to change a binary reliable variable although many more uncomplicated extensions live.

C. Random Forest Algorithm

A Random Forest is a classification algorithm contains numerous decision trees. Random forest algorithm builds the forest with numerous trees. The maximum the number of trees in the forest gives higher precision outcomes. Random forest algorithm establishes decision trees on data test pieces and then get the forecasting from each of them and ultimate typic the finest solution by means of elections. It is an ensemble procedure which is superior to a single decision tree since it lessens the over fitting by equating the outcome.

D. Decision Tree Algorithm

Decision tree constructs classification or regression framework in the form of a tree formation. It fragments the dataset into tiny subsets; parallelly a corresponding decision tree is additionally developed. A decision tree is a blueprint in the form of a tree structure, where each interior node represents a test on a feature, each branch signifies a final output of a test, and each leaf node grasps a class label. Constructing a decision tree which is compatible with the dataset is simple. The ultimate challenge present is constructing a better decision tree, which regularly means a tiny decision tree.

E. Support Vector Machine Algorithm

Support Vector Machines is a supervised machine learning algorithm which can be used for classification and regression problems. It uses a method called kernel trick to reform the data and then based on these changes it discovers a desirable boundary in the middle of the feasible outcomes. The plan of SVM is plain: i.e., the algorithm builds a line or hyper plane which distinguishes the data into classes. Kernel methods have been instituted for series data chart, wordings, pictures as well as vectors. Image categorization,

identification of handwritings is the difficulties that can be resolve using SVM.

F. XGBoost Algorithm

XGBoost is a decision-tree-based ensemble Machine Learning algorithm that makes use of a gradient boosting substructure. It uses more precise approximations to discover the best tree representation. The two causes to use XGBoost are its execution pace and model efficiency. XGBoost is ensembles of a decision tree algorithm where fresh tree selects bugs of those trees that are previously slice of the model. Trees are attached until no supplementary developments shell be built to the model. Some important characters of XGBoost comprise brilliant correction of trees and a corresponding reduction of leaf nodes.

III. ABOUT DATASET AND ATTRIBUTES

Pancreatitis disease dataset is taken in this paper which solves the classification trouble using machine learning algorithms. Pancreatitis sickness dataset consists of various information including diverse sufferers' medical information which contains some attributes such as Id, sex, smoking, alcohol, bp, serum calcium, lipid profile, hb, age and Type of disease. The data set actually containing pancreatitis sufferers which was modernized a month ago and was extracted from GEMS hospital for a far better précised outcome to be produced. Trained data samples and test data samples are the two segregations of the data set. Cleaning the data, Transforming the data and splitting the data are some of the major methods used for the Preparation of the data to initiate a data model. The cleaning and transforming the data are already performed in the dataset. Dividing the data into two segments is the main focus in the preparation of dataset. The dependent features i.e. A and output feature B are the two segments after the division of data samples. In the data sample, A and B are divided into training data sample i.e. A_train and B_train. And test data sample i.e. A_test and B_test. The dataset is being divided into 75% as the training data and 25% as the testing data.

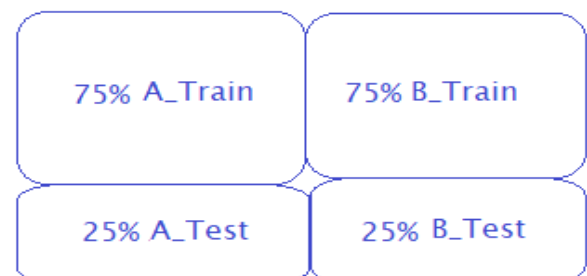


Fig.2. Representation of Training Data and Testing Data.

IV. RESULTS

The graph for K-nearest neighbor algorithm for the value of train data and test data accuracy is:

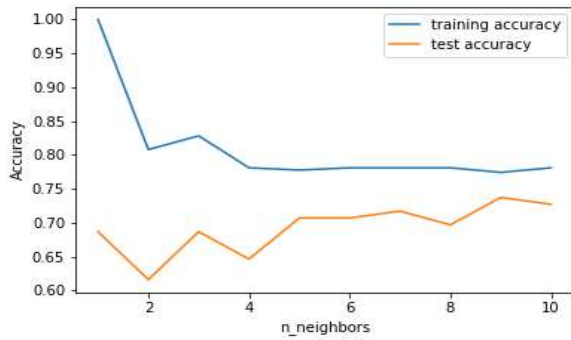


Fig.3. Representation of KNN Algorithm.

Other algorithms results are contrasted with each other algorithms for a better précised result. Accuracies of each algorithm is produced for both training data sample and testing data sample with regard to the pancreatitis sufferer’s data sample.

Table.1. Comparison Values

Algorithm	Trained data accuracy	Test data accuracy
KNN	0.77	0.74
Logistic Regression	0.751	0.747
Decision tree	1.000	0.960
SVM	1.000	0.990
XGBoost	0.989	0.989
Random forest	0.997	0.747

V. CONCLUSION

By studying various classification algorithms on the given dataset as input in Table-1. 0.989 is the accurate outcome of the training data sample and testing data sample which is generated by the XGBoost classification algorithm. This accurate constant of the training and testing data samples hear by state that the data sample is divided perfectly and also has produced a considerable classification model with the expected outcomes.

REFERENCES

[1] Hepatobiliary and Pancreatic Diseases International Volume 16, Issue 2, 15 April 2017, Pages 155-159.

[2] Krishna, S.G., Kamboj, A.K., Hart, P.A. et al. The changing epidemiology of acute pancreatitis hospitalizations: a decade of trends and the impact of chronic pancreatitis. *Pancreas*. 2017; 46: 482–488.

[3] Banks, P.A., Bollen, T.L., Dervenis, C. et al. Classification of acute pancreatitis—2012: revision of the Atlanta classification and definitions by international consensus. *Gut*. 2013; 62: 102–111.

[4] Zhao, G., Zhang, J.G., Wu, H.S. et al. Effects of different resuscitation fluid on severe acute

pancreatitis. *World J Gastroenterol*. 2013; 19: 2044–2052.

[5] Mullady, D.K., Yadav, D., Amann, S.T.,..., and NAPS2 Consortium. Type of pain, pain-associated complications, quality of life, disability and resource utilisation in chronic pancreatitis: a prospective cohort study. *Gut*. 2011; 60: 77–84.

[6] Lowenfels, A. B., Maisonneuve, P., Cavallini, G., Ammann, R. W., Lankisch, P. G., Andersen, J. R., et al. (1993). Pancreatitis and the risk of pancreatic cancer. *N. Engl. J. Med*. 328, 1433–1437.

[7] Yu, A., Woo, S. M., Joo, J., Yang, H.-R., Lee, W. J., Park, S.-J., et al. (2016). Development and validation of a prediction model to estimate individual risk of pancreatic cancer. *PLoS ONE*. 11:e0146473. doi: 10.1371/journal.pone.0146473

[8] Classification Algorithms in Machine learning. Available:<https://analyticsindiamag.com/7-types-classification-algorithms/>