

LITERATURE SURVEY

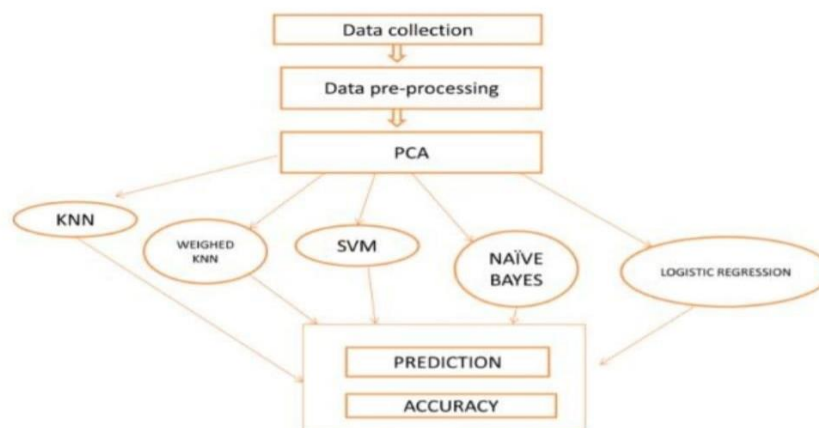
Paper Title : Prediction and Analysis of Heart Disease Using Data Mining Technique

Author : Anusha N.B, Chaitra.K ,

Publication : IRJET,may-2020,Volume:07 , Issue:05

Methodology : K-Nearest Neighbor(KNN), Navie bayes , Support Vector Machine (SVM)

In this study we have used an R studio rattle to perform Heart Disease classification of the Cleveland UCI repository. It provides an easy-to-use visual representation of the dataset, working environment and building the predictive analytics. Machine learning process starts from a pre-processing data phase followed by feature selection based on DT entropy, classification of modeling performance evaluation, and the results with improved accuracy. The feature selection and modeling keep on repeating for various combinations of attributes In K-NN, K is the number of nearest neighbors. The number of neighbors is the core deciding factor. K is generally an odd number if the number of classes is 2. When K=1, then the algorithm is known as the nearest neighbor algorithm. This is the simplest case. SVM (Support Vector Machine) is a supervised machine learning algorithm that is mainly used to classify data into different classes. Unlike most algorithms, SVM makes use of a hyperplane, which acts like a decision boundary between the various classes. SVM can be used to generate multiple separating hyperplanes such that the data is divided into segments and each segment contains only one kind of data. Weighted KNN is a modified version of KNN. A Naive Bayes' classifier may be a term addressing a simple probabilistic classification supported applying Bayes' theorem.



Paper title : Heart Attack Prediction and Visualization Using Machine learning.

Authors : Megha Banerjee, Reetodeep hazra, Suvranil saha, Megha bushan.

Publication : IJIREEICE , July 2021, Volume -09, Issue 07

Methodology : Logistic regression, Gaussian naïvebayes , Random forest algorithm

In this project, 4 machine learning algorithms are used namely Decision Tree, Random Forest, Gaussian Naive Bayes and Logistic Regression. A Decision Tree (DT) represents a tree like structure where each number considered being a branch with an outcome. DT is a fundamental component of Random Forest, which are among the most powerful ML algorithms available today. DT uses a layered splitting process, where at each layer the information data is split into two or more groups so that elements of the same group are as homogenous as possible to each other. A decision tree is a classification as well as a regression technique. It works great when it comes to taking decisions on data by creating branches from a root, which are essentially the conditions present in the data, and providing an output known as a leaf. The Random forest is basically a supervised learning algorithm which can be used for regression and classification tasks both. It is one of the most used algorithms because of its simplicity and stability. While building subsets of data for trees, the word “random” comes into the picture. A subset of data is made by randomly selecting x number of features (columns) and y number of examples (rows) from the original dataset of n features and m examples. Random forests are more stable and reliable than just a decision tree.



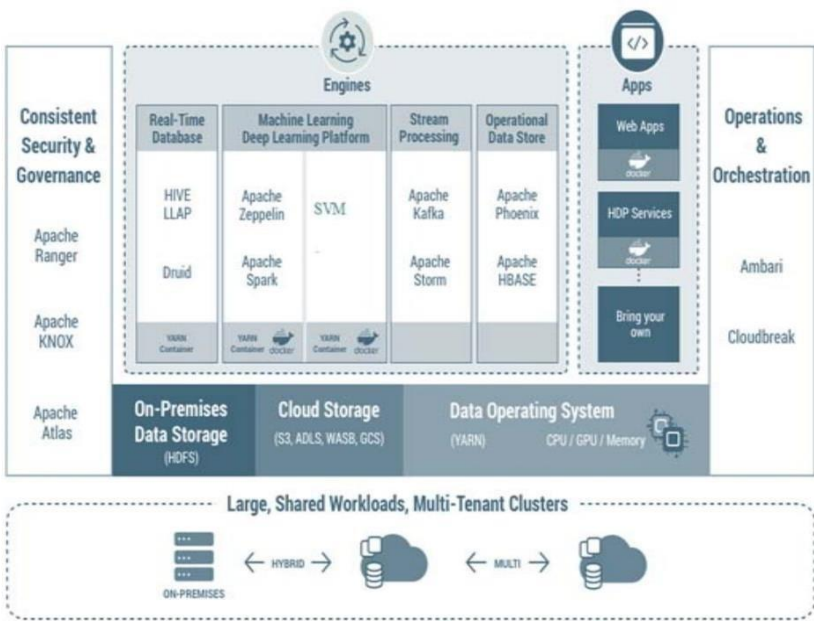
Paper title : Big Data Analytics in Heart Disease Prediction

Authors : Ahmed Imail, Samir Abdlerazek ,I.M.El-Henawy.

Publication : JATIT & LLS , June 2020

Methodology : SVM , Features reduction (FR).

The proposed methodology is using cloud structure to process medical data and to try to support doctors to decide in the diagnosis of heart diseases using SVM. This methodology is an efficient system because the proposed system applies a selection method based on the main features of the given dataset to classify the heart disease from the user profile in the cloud. The proposed framework used clusters from MapReduce. The system introduced a platform for storing heterogeneous and large data in IoT. The wearable devices send data as time series to IoT hub on Microsoft Azure . The streaming data is analyzed then can be stored on the cosmos DB.The Features Reduction (FR) is a methodology of selection that selects features with a high-class correlation (output). FR is also known as Dimensionality Reduction since it often tries to lower the number dimensions. The number of functions in FR impacts the efficiency of the classification . When a dataset is provided as data mining input with many classification features, the first target is cleaning the dataset and improving the



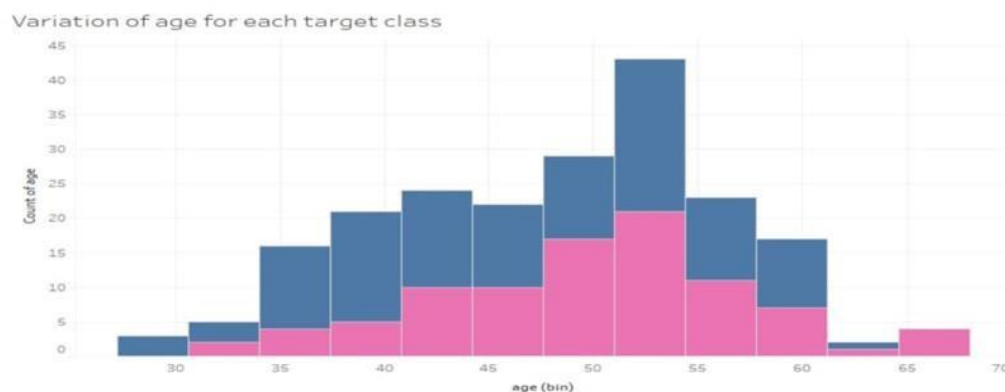
Paper title : Heart Disease Prediction Using Exploratory Data Analysis

Authors : R.Indrakumari , T. Poongodi, Soumya Ranjan jena.

Publication : ICITETM ,2020.

Methodology : SVM , KNN, ANN, Naïve Bayes, Logistic regression

Exploratory Data Analysis (EDA) is a method to analyze data using advanced techniques to expose hidden structure, enhances the insight into a given dataset, identifies the anomalies and builds parsimonious models to test the underlying assumptions. Exploratory Data Analysis (EDA) is classified into Graphical or non-graphical and Univariate or multivariate Univariate data consider one data column at a time while multivariate method considers more than two variables while analyzing. The diagnostic methods of diseases are of two types namely, Invasive and Non-invasive Invasive diagnostic method includes incise procedures in which instruments are used to cut the skin, mucus membrane and connective tissues. In contrast, non-invasive methods are used to diagnose diseases without opening the skin. Some of the machine learning algorithms based on non-invasive methods are Support Vector Machine (SVM), K- means clustering, K-Nearest Neighbour (KNN), Artificial Neural Network (ANN), Naive Bayes, Logistic Regression . Machine learning algorithms play a vital and accurate role in predicting heart disease. The advancement of technologies allows machine language to pair with big data tools to handle unstructured and exponentially growing data . In the paper, K means clustering method is proposed in big data environment and the visualization is made with the tableau dashboard.



Paper title : Big Data Analysis in Heart Attack Prediction.

Authors : Cheryl Ann Alexander and Lidong Wang.

Publication : OMICS, 2017

Methodology : Navie bayes, Decision Tree , ANN.

Different data mining techniques have been applied to predict heart disease. The accuracy of each algorithm has been verified and stated as Naïve Bayes, Decision Tree and ANN. The three-different data mining algorithms, ANN, C4.5 and Decision Trees are utilized to investigate heart related diseases using ECG signals. He analysis results clearly show the Decision Tree algorithm performs best and provides better accuracy than the C4.5 or Naïve Bayes algorithm. It requires less space when the volume of data is increased; it has a lower error rate, and minimizes the predictive error. C5.0 algorithm is the most potentially suitable algorithm for any kind of medical diagnoses. In cases of the C5.0 algorithm performs faster and provides the best accuracy with lower memory consumption Here are varied types of discourse analysis, including automated content analysis, which uses natural language processing and machine learning techniques. Running the discourse through I2A using KB De X for visualization, it was possible to classify discourse units (DUs) and their corresponding contents into distinctive idea types. First, I2A was used to detect ideas, followed by the determination of pertinent and promising ideas, then by conducting an analysis of discourse belonging to different idea types for validation purposes

