December 30, 2017

**Ishrak Islam Zarif**

Roll: 1307025

**Assignment on Machine Learning**

Course ID: CSE 4211

**INTRODUCTION:**

Machine learning is a field of [computer science](https://en.wikipedia.org/wiki/Computer_science) that gives [computers](https://en.wikipedia.org/wiki/Computer) the ability to learn without being explicitly programmed. Machine learning is closely related to (and often overlaps with) [computational statistics](https://en.wikipedia.org/wiki/Computational_statistics), which also focuses on prediction-making through the use of computers. It has strong ties to [mathematical optimization](https://en.wikipedia.org/wiki/Mathematical_optimization), which delivers methods, theory and application domains to the field. Within the field of [data analytics](https://en.wikipedia.org/wiki/Data_analytics), machine learning is a method used to devise complex models and algorithms that lend themselves to prediction; in commercial use, this is known as [predictive analytics](https://en.wikipedia.org/wiki/Predictive_analytics). These analytical models allow researchers, [data scientists](https://en.wikipedia.org/wiki/Data_science), engineers, and analysts to "produce reliable, repeatable decisions and results" and uncover "hidden insights" through learning from historical relationships and trends in the data.

**DATA SOURCES AND DESCRIPTION:**

In this assignment, we are provided with a dataset. This dataset is composed of a range of biomedical voice measurements from 31 people, 23 with Parkinson's disease (PD). Each column in the table is a particular voice measure, and each row corresponds one of 195 voice recording from these individuals ("name" column). The main aim of the data is to discriminate healthy people from those with PD, according to "status" column which is set to 0 for healthy and 1 for PD.

Our tasks are to:

1. Find the best set of features from the dataset that gives the best accuracy in prediction.

2. Create a useful machine learning model to predict Parkinson Disease (PD) given a test set of features. The models you must try. The accuracy must be validated by 5-fold Cross Validation.

a) Random Forest

b) SVM

c) Neural Network

d) Nearest Neighbor

Based on our self-interest, we can try any other methods. Our efforts will be appreciated.

3. Present our result with following measures:

a. Confusion Matrix

b. Accuracy Curve for different feature set. X axis: Feature Set, Y axis: Accuracy

**MACHINE LEARNING METHODS:**

In machine learning, tasks are generally classified into broad categories. These categories are based on how learning is received or how feedback on the learning is given to the system developed.

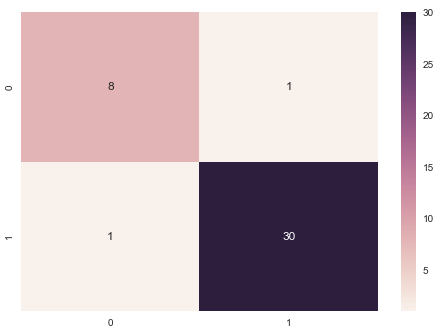
Two of the most widely adopted machine learning methods are **supervised learning** which trains algorithms based on example input and output data that is labeled by humans, and **unsupervised learning** which provides the algorithm with no labeled data in order to allow it to find structure within its input data.

Some machine learning algorithm are:

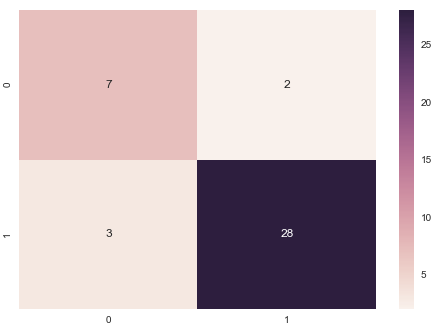
* Naïve Bayes Classifier Algorithm
* K Means Clustering Algorithm
* Support Vector Machine Algorithm
* Apriori Algorithm
* Linear Regression
* Logistic Regression
* Artificial Neural Networks
* Random Forests
* Decision Trees
* Nearest Neighbours

**RESULTS:**

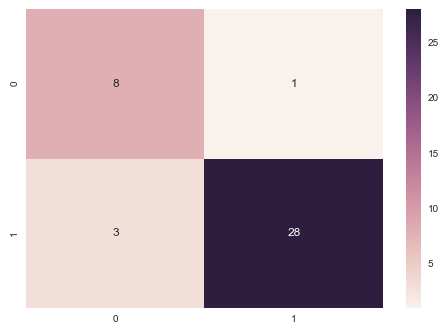
Using SVM:



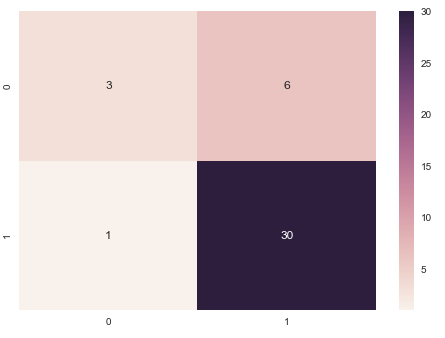
Using Decision Tree:



Using KNN:



Using ANN:



**CONCLUSION:**

Machine learning is a type of artificial intelligence ([AI](http://searchcio.techtarget.com/definition/AI)) that allows software applications to become more accurate in predicting outcomes without being explicitly programmed. The basic premise of machine learning is to build [algorithms](http://whatis.techtarget.com/definition/algorithm) that can receive input data and use [statistical analysis](http://whatis.techtarget.com/definition/statistical-analysis) to predict an output value within an acceptable range. The processes involved in machine learning are similar to that of [data mining](http://searchsqlserver.techtarget.com/definition/data-mining) and [predictive modeling](http://searchdatamanagement.techtarget.com/definition/predictive-modeling). Both require searching through data to look for patterns and adjusting program actions accordingly.

**REFERENCES:**

* [**https://archive.ics.uci.edu/ml/datasets/parkinsons**](https://archive.ics.uci.edu/ml/datasets/parkinsons)
* [**https://www.wikipedia.org/**](https://www.wikipedia.org/)
* [**https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/**](https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/)
* **https://www.datacamp.com/community/open-courses**