**Assignment On Breast Cancer**

**Data Analysis Report**

**By**

***Vijaya Venkata Santosh Kumar Gonela***

**The consumption of**

**The Chief Oncologist**

***Seattle Grace Hospital***

**Executive Summary**

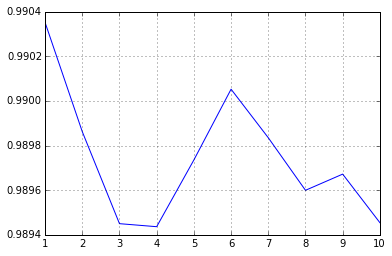
The presence of malignant tumours in human breast tissues is identified by the Scan Ware Systems developed by the West coast Technologies Inc. which gives indepth analysis and recommendations. Seattle Grace Hospitals analysis and recommendation considers the new technology reliable in predicting the malignant tumours. The data used in this analysis is provided by the *Seattle Grace Hospitals* and recommendations are provided after exhaustive data analysis using widely used and industry standard data analysis techniques and methodologies. The accuracy of the models and settles on the random Forest regression method uses various regression techniques.

After performing data analysis on the provided data we found that the new technology *Scan Ware Systems* is highly reliable in predicting the presence of malignant tumors, as the data study shows high confidence level with an accurate fit and the model came out with high accuracy. Thus, the analysis recommends using the *Scan Ware Systems* developed by the *West Coast Technologies.*

**Introduction**

The early development of random forests was influenced by the work of Amit and Geman who introduced the idea of searching over a random subset of the available decisions when splitting a node, in the context of growing a single tree. The idea of random subspace selection from Ho[4] was also influential in the design of random forests. In this method a forest of trees is grown, and variation among the trees is introduced by projecting the training data into a randomly chosen subspace before fitting each tree. Finally, the idea of randomized node optimization, where the decision at each node is selected by a randomized procedure, rather than a deterministic optimization was first introduced by Dietterich.

The introduction of random forests proper was first made in a paper by Leo Breiman. This paper describes a method of building a forest of uncorrelated trees using a CART like procedure, combined with randomized node optimization and bagging. In addition, this paper combines several ingredients, some previously known and some novel, which form the basis of the modern practice of random forests.I have used categorical values that gives 99 percent accuracy to predict the malignancy of the breast cancer. However, there is 0.1 percent of not predicting the breast cancer.

****

**Questions and Answers**

1. Were you successful in attempting to create a machine learning model to predict malignant tumors.

*Yes – the model clearly establishes that it’s an excellent machine model and can successfully predict malignant tumors.*

1. The hospitals lawyers are VERY careful and are worried about the accuracy of your model.  You'll need to inform them of the risks of using your system, and possibly convince them of its safety.

*This model, like any model, carries some risk. But within the given parameters of the data science, this is certainly one of the top accurate models we can have using the Random Forest model.*

*The model is very good. It is good to the point that the C-Score stands at 0.99, which is almost perfect.*

*There could be a chance of around 1 out of 100 cases where the model cannot find the presence of malignant tumor.*