**Name: Nikhil Jadhav**

**Student ID: 801075504**

**Cloud Computing for Data Analysis**

**VIDEO CASE 09 : Decision Trees**

Watch following videos:

**Video 1:** <https://youtu.be/RG4FYHfAQJQ>

**Video Case Questions:**

1. What are the purposes of decision trees?
2. What are some disadvantages that you see in decision trees when built for larger datasets?
3. Mention some other applications of decision trees.

**Answers:**

1. Follwing are the purpose of the decision tree:

* Decision trees are algorithms organizations use to assist them in making the best decision for their given objective.
* This approach is recommended for problems that have several solutions or contains several nested decisions. By using decision trees, organizations have a visual illustration of their decision's affects allowing them to picture what the costs and benefits will be for each possible outcome.
* Benefits associated with decision trees include that they "are simple to understand and interpret, have value even with little hard data, possible scenarios can be added [to them], worst, best and expected values can be determined for different scenarios, [they] use a white box model, and they can be combined with other decision techniques".
* In machine learning decision tree are a supervised learning method which can be used for classification and regression. They can also be used for identifying feature importance.

1. Decision tree suffers from one of cardinal sins in analytics and machine learning i.e. overfitting. Apart from overfitting decision tree suffer from following disadvantages:
   1. **Tree structure prone to sampling** – While Decision Trees are generally robust to outliers, due to their tendency to overfit, they are prone to sampling errors. If sampled training data is somewhat different than evaluation or scoring data, then Decision Trees tend not to produce great results.
   2. **Tree splitting is locally greedy** – At each level, tree looks for binary split such that impurity of tree is reduced by maximum amount. This is a greedy algorithm and achieves local optima. It may be possible, for example, to achieve less than maximum drop in impurity at current level, so as to achieve lowest possible impurity of final tree, but tree splitting algorithm cannot see far beyond the current level. This means that Decision Tree built is typically locally optimal and not globally optimal or best.
   3. **Optimal decision tree is NP-complete problem** – Because of number of feature variables, potential number of split points, and large depth of tree, total number of trees from same input dataset is unimaginably humongous. Thus, not only tree splitting is not global, computation of globally optimal tree is also practically impossible.
2. Following are some of the application of the decision trees:
   1. Predicting and reducing customer churn across many industries.
   2. Fraud detection in the insurance sector.
   3. Credit risk scoring in the banking and financial services.
   4. Whether to increase capacity vs outsourcing to fulfil demand
   5. Improving First Call Resolution (FCR)
   6. Reducing Average Handling Time (AHT)
   7. Improving Customer Satisfaction Rate (CSAT)