* **Neural Networks – *Advanced form of Machine Learning known as ‘Deep Learning’***
* **Neural Networks - *Uses interconnected nodes or neurons in a layered structure that* *resembles the human brain*.**
* **Our Goal: *Predicting Client Churn***
* **Datafile: *900 records, 6 fields used in our model.***
* **Target variable:**

*Churn*

* **Features:**

*Client Age*

*Total Purchase*

*Account Manager (0/1)*

*Years Subscribed*

*Number of Websites Used*

* Initial construction run shows strong predictive capacity: *86% .*
* Predictive accuracy decreased with additional layers added, increased neurons, and more powerful activation functions
* The initial two layers, the sigmoid activation function for each layer, and 500 & 300 neurons respectively, increased predictive accuracy 1% : *87%* .
* Neural Networks is most effective when used with *‘BIG data’.*
* Neural Networks can yield unreliable results on data that is too small/simple.
* Potential underperformance due to:

*Overfitting Training Data*

*The Vanishing Gradient Theorem (i.e. Back Propagation)*

Vanishing gradient problem is a phenomenon that occurs during the training of deep neural networks, where the gradients that are used to update the network become extremely small or "vanish" as they are backpropogated from the output layers to the earlier layers.

The vanishing gradient problem is caused by the derivative of the activation function used to create the neural network. The simplest solution to the problem is to replace the activation function of the network. Instead of sigmoid, use an activation function such as ReLU.