MACHINE LEARNING TASK

Name: Parth Madan

**Task 1: You are assigned to a project where you are working on a dataset provided with Career guidance which contains various details of the students and their preferred career choices. How do you select the important variables? Explain your method in detail. (You are free to assume the variables as needed).**

**SOLUTION:** Moreover, According to my there is no such rule till now made that helps to select the variable from a dataset but we can do it or we can say that it is only possible with Data Collection Process. In this process, the Quality of the data is more important than the quantity, If quality is good we can easily apply Ml Algorithms over their and easily get results out of the reach. Another method that we can use is that we take the help of the Sklearn library to inbuilt class in tree-based classifiers.Somehow in Machine learning, the step function is also there which help to separate important variable from a dataset.Some other feature selection methods are supervised and unsupervised. The scikit-learn library also provides many different filtering methods once statistics have been calculated for each input variable with the target.

For Example: In my dataset, there is one feature that is the stream and that feature is very important because on behalf of that next career of children has been fully dependent so we have to select that variable so we use some feature selection technique for example supervised to select that variable.

**Task 2: You were assigned to a project where you built a random forest model with 10000 trees. You were in cloud-nine after getting a training error of 0.00. But the validation error is 46.89. What went wrong? Does that mean that you trained your model wrong? (Explain briefly) and How would you explain Machine learning to a grade-1 kid (probably 5-6 years old)?**

**SOLUTION:** According to the question we are getting 0.0 training error at 9th cloud but on the other hand, we got 46.89 % error as validation so from that we conclude that this is all because of overfitting of the training dataset. As there is overfitting so may be machine behave poorly on the new dataset too because of a validation error of 46.89%.According to my knowledge or what I know about overfitting this will occur if training accuracy is 100 % or training error is 0%.The reasons can range from simple and insufficient training data to complex to multiple subtrees. These subtrees have the same function as the model and therefore predict the same model over and over again, making the entire algorithm very suitable for predicting data. No, it doesn't mean that we have trained our model in the wrong way the thing is that we can't take care of that there is the chance of overfitting of data too which happens by getting 0 % training error.

As For grade 1 kid definition of Machine Learning or fundamental of machine learning to understand is one of difficult task but by taking different example to tell that kid about machine learning is possible for example as machine learning is all about learning the machine so we can give example that as that as a child you learn alphabet,tables , counting by books and with the help of teacher as same a machine can learn things by training given to that machine and for that books are dataset which helps the machine to learn about fundamental.Another example we can give is Alexa or google home which helps that child to play music, help him in his studies too so we can tell them that machine learning helps Alexa to do what that child wants to do or get from that Alexa.So Alexa is an application of (Ml+AI) too. So like that, I will able to explain what is machine learning to a grade 1 kid.

**\*\*\*\*\*\*\*\*END\*\*\*\*\*\*\*\***