PILOT STUDY PROPOSAL

485 words

In a layman language Machine learning means ‘a machine learning from data’. Machine learning algorithm gathers knowledge from the data and then it performs the task on unseen data same like a man learning to drive a car, man practice driving on limited number of roads, learn meaningful insights and after his learning is complete, he can drive on any road in the world.

The manager has historical data of customers which includes past policies, information about the insured, the purpose and destination of the travel and whether a claim is filed or not. Data provided will have target variable that is claim, so clearly it is a case of supervised learning. Insurance company wants to predict whether a customer will file a claim on their travel or not, so clearly it is binary classification problem. It is difficult to tell in advance which ML model will perform well on data. There are two approaches to select best ML model. First one is to test every ML model on the data. By doing this at the end we will be sure which ML model to pick but this method is not feasible as real life problems have lot of data and it will take a lot of time and computational power. Second approach is to pick ML model which is best for the problem. This can be done if ML consultant has deep understanding how algorithm works can interpret which algorithm is best for the problem. In this classification problem, I will use decision tree as it is simple model, uses less computational power and easily interpretable. If I will not get expected accuracy then I will construct SVM model as it can learn complicated relationships from the data. Travel insurance generally covers luggage loss, emergency medical insurance, lost\ stolen luggage, flight delays, flight cancelation. Informative features that I would like to be provided is age, destination, crime rates in destination country, duration of travel, BMI, smoker, applied claims in previous travels. BMI, smoker and age, all these features can tell about traveller’s health. If traveller is healthy he/she will less likely to claim for reason medical emergency. Duration of travel will also matter, longer the duration chances of filing a claim will also increase. Destination of travel and crime rates in destination country will tell about chances of mishappening that traveller could face.

Evaluation of model is also important part of model building. Simplest way to evaluate a classification model is to calculate accuracy. Accuracy metric is not always true representation of how good model does because for example if the data provided has 95% True values and 5% False values then in this case no one can rely only on accuracy metrics and one has to calculate precision and recall to correctly evaluate the model. So on analysing the data I will decide exactly which evaluation metrics I will use.