**User interview**

**Participant: 1**

**Interview length: 15 minutes**

**Response Summary:**

**What is the occupation of the interviewee?**

Interviewee is a full-time working employee as a data scientist and a researcher. He is also a graduate student doing specialization in Imaging and Machine Learning. He also has various research papers published in reputable conferences and Journals.

**What kind of data have you worked on?**

I have worked on Imaging Datasets including Medical, X-RAY, CT scan and various Tabular Datasets including Heart Disease datasets and Cancer Detection datasets.

**What kind of analysis have you done in your career uptil now?**

I have worked on Semantic Image Segmentation to segment lungs imagery from Chest X-RAY, Neural Network Image Classification, parameter value selection (hyperparameter selection) and various Clustering methods.

**What tools and softwares have you used?**

For data visualization and business intelligence, I have used Tableau and Power BI. For hyperparameter optimization and data pre-processing, I have used various Python machine learning libraries including scikit-learn, TensorFlow, NumPy, Pandas, Keras, SciPy and PyTorch.

**What basic steps you take when you have a machine learning problem to solve?**

The first step is to analyze the problem in a sense to figure out the input and the output. Once it is done, I see which model is best suited for the problem and make a list of them from model of least complexity to model of highest complexity. Then I check which type of data should I use, on basis of my inputs and outputs I search if I have data available online or should I need to collect it myself. Then I preprocess data to be fed to model and train the least complex model to check the performance and move up to the ladder and stop where I get the desired result without exhausting my resources.

**What is the most painful process you think in training a model?**

According to my experience, the most painful process would be to prepare and pre-process the data. Sometimes I collect the data and it is impossible to pre-process it according to my needs. Or if I pre-process it and extract features, then these features are not correct for getting desired output and the accuracy turns out to be bad.

**Are you aware of hyper-parameter tuning?**

Sure, I am. It is one of the important things in ML to get optimal solution by controlling the learning process

**What is the most painful thing about hyper-parameter tuning?**

Whenever I build a model, I want the solution to be optimal, but there are a lot of parameters that needs to be tuned to get that optimal solution. So, I have to try all the possible values to get a better fit for my model and dataset. Even if I tune the model, whenever I change the data, accuracy decreases and I have to tune the parameter values again.

**What do you think is the most troublesome step or process for you as a machine learning specialist?**

Personally, I find the data pre-processing most troublesome in the process of training a model.

**What tools do you use for pre-processing and hyper-parameter tuning?**

I have not used any built-in software or tool to do that yet. For hyper-parameter tuning I have used sklearn libraries, hit a trial and sometimes my own experience in the field helps me for simple models.

**Why do you think pre-processing is more painful than hyper-parameter tunning?**

There are many models which apply brute force to get optimal hyper-parameter values for optimal accuracy. Moreover, I know that there is some optimal value but in case of pre-processing I’m never sure, the data can be wrong or the pre-processing technique I’m using can be wrong. In my career so far, I have come across problem of hyper-parameter tuning but I’ve solved it easily, but pre-processing is always time consuming.