Data Scientist Questionnaire

1. How many years of experience do you have as a Data Scientist? Take discount for the time period you are not full-time.

* 2.5 years

1. Answer the following question regarding your experience building advanced analytic solutions.

How many Proof-of-Concept you did?

* 7
* I have done 6 POC’s which are into Machine Learning, Deep Learning, Natural Language Processing (NLP) and Computer Vision (CV)

How many solutions you developed are now used in productive environment?

* 4
* The projects which re in production are Machine Learning to reduce the false positives and at the same time to make sure there are no false negatives (i.e., considering F1-Score as evaluation metric)
* Text Summarization with Deep Learning (I have used BART-LARGE-CNN pre-trained model to further finetune on downstream applications i.e. fine-tuning)
* Sentiment Analysis with Deep Learning to get the customer sentiment for the loan offers and another cross-sell offers which are proposed by bank (Used pre-trained bert-base-cased and fine-tuning on the custom dataset)
* Semantic Search Engine with Deep Learning which will used for information retrieval for bank documents which will be used by regulators from the AML team. (I have used paraphrase-mpnet-base-v2 from sentence transformers)

1. Were you responsible for being the main point of contact for the business (gathering requirements and building the architecture per those requirements)?

* Yes, I was. When I was working for the False positive reduction project using ML and for Text Summarization project with Deep Learning i.e., Bart-large-cnn model which utilizers Bert (Encoder from transformers, Denoising Autoregressive Autoencoders and the GPT (Decoder part from transformers). My roles is to explain how a particular ML or DL or any pre-trained model which you are using.
* In most scenarios, we often see the business partners/clients/Point of contact who are assigning the project expects a better solution and obviously it’s our duty to deliver it and articulate how the model works in layman terms to them. So, when I was working the person who acted as point of contact had no ML and DL expertise. So, I have explained what models I have used and how it mathematically works and why it is the best suitable model for that project.

1. Rate your programming skill (0-10, 10 is the highest) on the following:

Python: 8

SQL: 7.5

1. Rate your skill level (0-10, 10 is the highest) on following AI/Machine Learning areas you have both knowledge and professional experience (PoC or productive roll out projects):

Tabular datasets and machine learning/DL – 8.5

Unsupervised or semi supervised learning - 8

Deep learning - 8

Timeseries forecasting – 6.5

Computer vision – 6.5

Natural language processing – 8.5

1. Rate your knowledge in the following areas (0-10, 10 is the highest) and answer whether you obtained them through professional courses or self-study.

Business Statistics – 7.5

Linear Algebra and Optimization – 7

Probability and Multivariate calculus - 7