**DPR-3 Chronic Disease Detection**

**Q1- Tell me about your current Project?**

**Ans-** The project objective is to check the chronic disease on given data by classification methodology. We can figure out the condition of the disease in advance and take respective prescription to it.

**Q2- What was the size of data?**

**Ans-** The number of batches usedin that each batch consists of 201 data.

**Q3- What was the data type?**

**Ans-** Target column value would be categorical.

**Q4- What was the team size and distribution?**

**Ans-** The team consisted of:

- 1 Product Manager,

- 1 Solution Architect,

- 1 Lead,

- 2 ML-Ops engineers,

- 2 QA engineers,

- 2 UI developers, and

- 3 Data Scientists.

**Q5- What Hadoop Distribution were you using?**

**Ans-** Not Applicable.

**Q6- What is the version of distribution?**

**Ans-** Not Applicable.

**Q7- What was the size of cluster?**

**Ans-** We took 8 clusters but we can extend as well as per requirement increases.

**Q8- How many nodes were there in all the Dev,UAT ,Prod environments?**

**Ans-** Mainly there were three nodes but we divided them as per sub nodes required in an application**.**

**Q9- How were you creating and maintaining the logs?**

**Ans-** We maintain logs for each and every activity in an application. We checked the logs regularly.

**Q10- What techniques you were using for pre-processing for various data science use cases and visualization?**

**Ans-** We used various techniques to do pre-processing of data.

**1)Data Validation:** Good data/Bad data**.**

**2)Statistical Graphical Analysis:** Over given data we would apply visualization to check what is right or wrong.

**3)Data Transformation:** Apply all the transformation techniques (like missing values,null value,nan value,outlier).

**4)Standardization & Normalization:** Apply normalization and standardization over data if required.

**Q11- How were you maintaining the failure cases?**

**Ans-** We maintain separate log for this which will cover all failure cases.

**Q12- What kind of automation have you done for data processing?**

**Ans-** When data introduce to system then it will check validations(Name validation ,Number of columns, Name of columns, Datatype of columns ,Null values in column) if any of these not matched then data reached to respective data folders (Good\_data\_folder/Bad\_data\_folder).We prepare separate pipeline for all the task that we are performing.

**Q13- Have you used in scheduler?**

**Ans-** Yes, a scheduler was used for schedule different job simultaneously.

**Q14- How are you monitoring your job?**

**Ans-** We regularly checks our logs to see everything is working properly or it requires to retraining of data.

**Q15- What were your roles and responsibilities in the project?**

**Ans-** My responsibilities mainly deal with data gathering and do the exploratory data analysis .After putting it to pipeline check the result . And maintain proper logs working and close one batch.

**Q16- What was your day to day task?**

**Ans-** My daily working hours and respective work is recorded in JIRA as per sprints decided in scrum meeting. We got all the knowledge and requirements during standup meetings about our on-going project.

**Q17- In which area you have contributed the most?**

**Ans-** I contributed mainly in exploratory data analysis of data and get ready data for performing model operation. To finalize model with highest accuracy and do the deployment of the project with respect to respective cloud.

**Q18- In which technology you are most comfortable?**

**Ans-** I am comfortable with Machine Learning ,Deep Learning, Natural Language Processing. I can do work well with Machine Learning ,Deep Learning, Natural Language Processing depending upon project requirements.

**Q19- How you rate yourself in big data technology?**

**Ans-** I have worked with big data technologies but my main focus to work as a data scientist.

**Q20- In how many projects you have already worked?**

**Ans-** I have worked in many project small as well as big. Regression problem ,Classification problem, NLP Project ,Object Classification, Object Identification etc.

**Q21- How were you doing deployment ?**

**Ans-** The mechanism of deployment depends on the client’s requirements. Some clients wants that at the respective cloud and others wants in API calls respectively.

**Q22- What kind of challenges have you faced during the project?**

**Ans-** The biggest challenge to find right amount of data that will solve our current problem statement. Next step to find out the right algorithms for our solution. That will help to achieve accuracy as well as optimize results.

**Q23- What will be your expectations?**

**Ans-** I am looking to explore on new technologies and gain excellence with time. I try to put all my knowledge and efforts to my work to achieve our productive goal.

**Q24- What is your future objective?**

**Ans-**

As data science field is keep changing day by day ,so we have learn new technologies and keep in an adaptive mode .We have to explore data science field more and more and provide better solution for customer /user satisfaction.

**Q25- Why are you leaving your current organization?**

**Ans-** Unemployed.

**Q26- How did you do Data validation?**

**Ans-** Datavalidations(Name validation ,Number of columns, Name of columns, Datatype of columns ,Null values in column) if any of these not matched then data reached to respective data folders (Good\_data\_folder/Bad\_data\_folder).We prepare separate pipeline for all the task that we are performing.

**Q27-How did you do data enrichment?**

**Ans-** By collecting data from different vendors as well and try to test on different training sets.

**Q28- How would you rate yourself in machine Learning?**

**Ans-** Out of 10 is 8.5 .

**Q29- How would you rate yourself in distributed computation?**

**Ans-** Not Applicable.

**Q30- What are the areas of machine learning algorithms that you already have explored?**

**Ans-** I have explored various machine learning algorithms like Linear Regression, Logistic Regression, L1 and L2 Regression, Polynomial Regression, Multi Linear Regression,Decision Trees, Random Forests, Extra Trees Classifier, PCA, TSnE, UMAP, XG Boost,CAT Boost, ADA Boost, Gradient Boosting, Light Boost, K-Means,K-Means ++,LDA, QDA, KNN, SVM, SVR,Naïve Bayes, Agglomerative clustering, DBScan, Hierarchical clustering, TFIDF, Word to Vec, Bag of words, Doc to Vec, Kernel Density Estimation are some of them.

**Q31- In which part of machine learning have you already worked on?**

**Ans-** I have worked on both supervised and unsupervised machine learning approaches and building different models using the as per the user requirement.

**Q32- How did you optimize your solution?**

**Ans-** Model optimization depends on a lot of factors.

-Train with better data(increase the quality), or do data pre-processing steps more efficiently.

- Keep the resolution of the images identical.

- Increase the quantity of data used for training.

- Increase the number of epochs for which the model was trained.

- Tweak the batch input size, the number of hidden layers, the learning rate, rate of decay, etc. to produce the best results.

- If you are not using transfer learning, then you can alter the number of hidden layers, activation function.

- Change the function used in the output layer based on the requirement. The sigmoid functions work well with binary classification problems, whereas for multi-class problems, we use a sigmoid model.

- Try and use multithreaded approaches, if possible.

- Reduce Learning Rate in plateau reasons optimizes the model even further.

**Q33- How much time your take to get trained?**

**Ans-** It took around 10 hours to train the model using Nvidia Pascal Titan GPU.

**Q34- At what frequency are you retraining and updating your model?**

**Ans-** The model gets retrained every 10 days.

**Q35- In which mode have you deployed your model?**

**Ans-** I have deployed the model both in cloud environments as well in the on-premise ones based on the client and project requirements.

**Q36- What is your area of specialization in machine learning?**

**Ans-** I have worked on various algorithms. So I can work in almost all algorithms as per customer requirements.