ML, DL Questions

1. How you can define Machine Learning?
2. What do you understand Labelled training dataset?
3. What are 2 most common supervised ML tasks you have performed so far?
4. What kind of Machine learning algorithm would you used to walk robot in various unknown area?
5. What kind of ML algo you can use to segment your user into multiple groups?
6. What type of learning algo realised on similarity measure to make a prediction?
7. What is an online learning system?
8. What is out of core learning?
9. Can you name couple of ml challenges that you have faced?
10. Can you please give 1 example of hyperparameter tuning wrt some classification algorithm?
11. What is out of bag evaluation?

- What do you understand by hard & soft voting classifier?

13. Let's suppose your ML algorithm is taking 5 min time to train, How will you bring down time to 5 second for training? (Hint: Distributed Computation)
14. Let's Suppose I have trained 5 diff model with same training dataset & all of them have achieved 95% precision. Is there any chance that you can combine all these models to get better result? If yes, How? If no, Why?
15. What do you understand by Gradient decent? How will you explain Gradient decent to a kid?

- Can you please explain diff between regression & classification?
- Explain a clustering algorithm of your choice.
- How you can explain ML, DL, NLP, Computer vision & reinforcement learning with example in your own terms?

19. How you can explain semi-supervised ML in your own way with example?
20. What is difference between abstraction & generalization in your own word.
21. What are the steps that you have followed in your last project to prepare the dataset?

- In your last project what steps were involved in model selection procedure?

23. If I give you 2 columns of any dataset, what will be the steps will be involved to check the relationship between those 2 columns?

24. Can you please explain 5 diff kind of strategies at least to handle missing values in dataset?

25. What kind of diff. issues you have faced wrt your raw data? At least mention 5 issues.

- What is your strategy to handle categorical dataset? Explain with example.

27. How do you define a model in terms of machine learning or in your own word?

28. What do you understand by k fold validation & in what situation you have used k fold cross validation?

- What is meaning of bootstrap sampling? explain me in your own word.
- What do you understand by underfitting & overfitting of model with example?
- What is diff between cross validation and bootstrapping?
- What do you understand by silhouette coefficient?
- What is the advantage of using ROC Score?

34. Explain me complete approach to evaluate your regression model

- Give me example of lazy learner and eagar learner algorithms example.
- What do you understand by holdout method?
- What is diff between predictive modelling and descriptive modelling.
- How you have derived a feature for model building in your last project?
- Explain 5 different encoding techniques.
- How do you define some features are not important for ML model? What strategy will you follow
- What is difference between Euclidian distance and Manhattan distance. Explain in simple words.
- What do you understand by feature selection, transformation, engineering and EDA & What are the steps that you have performed in each of these in detail with example.
- What is difference between single values decomposition (SVD) and PCA? (hint: SVD is one of the way to do PCA)
- What kind of feature transformations you have done in your last project?

45. Have you taken any external feature in any of project from any 3rd party data? If yes, explain that scenario.

- If your model is overfitted, what you will do next?
- Explain me bias variance trade-off.
- What steps would you take to improve accuracy of your model? At-least mention 5 approach. And justify why would you choose those approach

- Explain process of feature engineering in context of text categorization.
- Explain vectorization and hamming distance.
- Can you please explain chain rule and its use?
- What is difference between correlation and covariance?
- What are the sampling techniques you have used in your project?
- Have you ever used Hypothesis testing in your last project, if yes, explain How?
- In which case you will use naïve Bayes classifier and decision tree separately?
- What is the adv & disadvantage of naïve Bayes classifier, explain
- In case of numerical data what is naïve Bayes classification equation you will use?
- Give me scenario where I will be able to use a boosting classifier and regressor?
- In case of Bayesian classifier what exactly it tries to learn. Define its learning procedure.
- Give me a situation where I will be able to use SVM instead of Logistic regression.
- What do you understand by rbf kernel in SVM?
- Give me 2 scenarios where AI can be used to increase revenue of travel industry.
- What do you understand by leaf node in decision tree?
- What is information gain & Entropy in decision tree?
- Give disadvantages of using Decision tree
- List some of the features of random forest.
- How can you avoid overfitting in decision tree?
- Explain polynomial regression in your own way.
- Explain learning mechanism of linear regression.
- What is the cost function in logistic regression?
- What is the error function in linear regression?
- What is the use of implementing OLS technique wrt dataset?
- Explain dendrogram in your own way.
- How do you measure quality of clusters in DBSCAN?
- How do you evaluate DBSCAN algorithm?
- What do you understand by market basket analysis?
- Explain centroid formation technique in K Means algorithm.
- Have you ever used SVM regression in any of your project, If yes, Why?
- Explain the concept of GINI Impurity.
- Let's suppose I have given you dataset with 100 column how you will be able to control growth of decision tree?
- If you are using Ada-boost algorithm & if it is giving you underfitted result What is the hyperparameter tuning you will do?
- Explain gradient boosting algorithm.
- Can we use PCA to reduce dimensionality of highly non-linear data.
- How do you evaluate performance of PCA.
- Have you ever used multiple dimensionality techniques in any project? if yes, give reason. If no, where can we use it?

- What do you understand by curse of dimensionality explain with help of example
- What is the difference between anomaly detection and novelty detection
- Explain gaussian mixture model.
- Give me list of 10 activation functions with explanation
- Explain neural network in terms of mathematical function
- Can you please corelate a biological neuron and artificial neuron?
- Give list of cost functions you heard of, with explanation.
- Can I solve problem of classification with tabular data in neural network?
- What do you understand by backword propagation in neural network?
- Why do we need neural network instead of straight forward mathematical equation?
- What are the different weight initialization techniques you have used?
- Can you visualize a neural network? if yes provide name of software we can use?
- How will you explain training of neural network?
- Can you please explain difference between sigmoid & tanh function.

100. Explain disadvantage of using RELU function.
101. How do you select no. of layers & no. of neurons in neural network?
102. Have you ever designed any Neural network architecture by yourself?
103. Can you please explain SWISS Function?
104. What is learning rate in laymen way and how do you control learning rate?
105. What is diff between batch, minibatch & stochastic gradient decent.
106. What do you understand by batch size while training Neural N/w with example
107. Explain 5 best optimizer you know with mathematical explanation.
108. Can you build Neural network without using any library? If yes, prove it.
109. What is use of biases in neural network?
110. How do you do hyper-parameter tuning for neural network
111. What kind of regularization you used wrt neural network.
112. What are the libraries you have used for neural network implementation?
113. What do you understand by custom layer and a custom model?
114. How do you implement differentiation using TensorFlow or Pytorch library?
115. What is meaning of epoch in simple terms?
116. What do you understand by a TensorFlow record?
117. Explain the technique for doing data augmentation in deep learning
118. List down diff CNN network you heard of.
119. List down a names of object detection algorithm you know
120. What is difference between object detection and classification?
121. List down major tasks we perform in CNN.
122. List down algorithms for segmentation
123. Which algorithm you can use to track a football in football match.

124. If I give you a satellite image data, so which algo you will use to identify image from those image data

125. Which algorithm you will use for PCB fault detection.

126. What do you understand by pretrained model?

127. Explain different types of transfer learning.

128. Explain me where your CNN network will fail with example. And where we can use RNN network.

129. Which GPU you have been using to train your object detection model?

130. How much data set you have used for this model, what was epoch, time and accuracy of the model

131. What kind of optimization you have done for training object detection model

132. How do you evaluate your object detection model?

133. List down algorithm for object tracking

134. What do you understand by FPS (frame per second)?

135. Can you please explain 2D & 3D convolution?

136. What do you understand by batch normalization?

137. Which algorithm you use for detecting handwriting detection?

138. Explain me SoftMax function.

139. What is disadvantage of using RNN?

140. List down at least 5 RNN?

141. Explain architectural diagram of LSTM, Also list Adv & dis adv

142. Explain architectural diagram of BI LSTM, Also list Adv & dis adv

143. Explain architectural diagram of stacked LSTM. Also list Adv & dis adv

144. What do you understand by TF-IDF

145. How you will be able to create a Word 2 vector of your own

146. List down at least 5 vectorization technique.

147. What is difference between RNN and Encoder-Decoder.

148. What do you understand by attention mechanism and what is use of it

149. Have you read a research paper Attention you all need? If not, then why you are claiming you know NLP

150. What do you understand by multi headed attention? explain

ML Questions Part 2

1. Tell me something about your project you have done in past?
2. What was your Dataset size for ML Project?
3. What is type of your dataset?
4. What was frequency of your dataset? (E.g. batch, streaming etc)
5. What was source system for your dataset? (E.g. sensor, satellite Kafka, cloud, etc.)
6. What was kind of derived dataset that you have mentioned in project?
7. How you have done validation dataset?
8. Have you created any pipeline to validated this dataset or you were using any tool?
9. What do you understand by data lake?
10. What do you understand by data warehousing?
11. Can you please name some validations that you have done on top of your data?
12. How you have handled streaming dataset?
13. How many different types of environments were available in your project?
14. What was your delivery mechanism for particular project?
15. Have you used any OPS pipeline for this current project?
16. How you were doing model retraining?
17. How you have implemented model retraining in your project?
18. How frequently you have been doing model retraining and what was the strategy for model retraining?
19. What was kind of evaluation you were doing in production environment
20. What was no. of request (hits) your model was receiving on daily basis?
21. How you have implemented logging in project for any failure cases?
22. How you have integrated notification (or Alarm) system for your project?
23. How you have implemented model monitoring?
24. How you have derived final KPI (Key Performance Indicator) for your client?
25. How many dashboards were there in your project?
26. On which platform you have productionised your model?
27. What kind of API you have exposed to receive data for model?
28. What was size of your final production environment (system configuration)?
29. What an all Databases you have used in project?
30. What kind of optimization you have done in your project, till what depth & explain the example
31. Can you please talk about complete team structure and team size?
32. What was duration of your complete project?
33. What was your day to day responsibility in last 2 month?

34. What kind of change request you have been receiving after you productionised project
35. What kind of testing you have done in development, UAT, pre-pod and prod?
36. Have you used some of the predefined AI-OPS pipelines if yes explain
37. Who has implemented AI-OPS in your project?
38. What was OPS stack you have been using?
39. What do you understand by CI-CD & have you implemented those in your project. If yes, what was the tech stack you used for CI-CD pipeline?
40. What was biggest challenge you faced in project and how you have resolved it?
41. Give me one scenario where you worked as team player?
42. What was your overall learning from current project?
43. How do you keep yourself updated for new technology?
44. Have you designed an architecture for this project? If yes, define a strategy wrt to your current project.

--------- Below are some questions for people with 7+ years of Exp-----------

- What kind of discussions generally wrt client?

46. What was your contribution in team building?

- How you have defined completed tech stack for AI?
- What was kind of benefit you have given to your current company in terms of cost cutting and revenue?
- What kind of new innovation you have introduced?
- How do you push your team for research or new implementation?
- How many projects you are handling?
- How many clients you have acquired?
- If new demand comes from client, how do you evaluate that requirement?
- How do you prepare costing for project?
- What kind of skillsets you look in person, to handle delivery of upcoming project?
- On an average, how much time you took for building a new team?
- What kind of stack you involved in initial project?
- How do you decide timeline for project delivery?
- How do you keep track of project progress?
- How do you handle dependencies between the team?
- How much profit you have given to your previous organization.

Deep learning & Vision Questions

- How many images you have taken to train your DL model?
- What is size of your model that you have productionised?
- Have you tried optimizing this Vision or DL model?

65. Where you have hosted your Computer Vison model?

- What was your frame per second?
- What are the data filtration strategy you have defined for CV project in production?
- Have you used any edge device in this project, if yes, why?
- What was name of camera & camera quality?
- What was final outcome you were generating from these devices?
- Have you processed the data in local system or in cloud? Give reason.

72. How many number devices you have productionised (camera, edge devices etc.)

- Let's suppose I am trying to build solution to count no. of vehicle or to detect their no. plate or track their speed. Then what is the dependency of distance, position & angel of camera on your final model? What will happen to your model? if we change position angle.
- What was your data collection strategy in CV project, have you received data from client or you have created the data? And how you have implemented it?
- What was data labelling tool that you have used for your project?
- If I have to do OCR then which API you will use or you have used in your previous project?
- Suppose if my images data is blur, what will be your strategy to enhance the image quality?

78. Have you implemented object tracking in any of your project, if yes, give me scenario?

- Suppose there are 2 mobile devices which are moving, so suppose if this two device position overlaps with each other then what will be your strategy to avoid any error while tracking those devices using camera?
- Have you implemented multicamera tracking, do you have Any idea about it?
- Explain me some of real-life use case of segmentation.
- What kind of AI application you will build for retail seller to increase their sell?
- Let's suppose I am trying to build AI solution to monitor a productivity of a kid, what kind of feature you would like to give in that product?

------- NLP-------

84. Have you productionised BERT Model, if yes, can you talk about hurdles that you have faced?

- How you have optimized your BERT based solution
- What kind of NLP tasks you were doing wrt BERT?
- Whether you have implemented BERT base & BERT large?
- What are the disadvantages of using BERT?
- Can please one of the lighter versions of transformer-based model?
- What was the accuracy that you were receiving wrt specific tasks?
- Let's suppose I have to build language-based model and there are online solution providers are there but they are costly, so what will be your strategy?
- Have you used hugging phase APIs?
- What is diff between BERT and GPT based models?
- There is no Decoder model in BERT then how do we get output if its just a encoder lever model?
- How masking is been implemented in BERT based model and what are its disadvantages
- How masking is been implemented in GPT based model and what are its disadvantages
- How backpropagation happens in BERT model. Explain.
- Can you explain Query, Key & Value in any Transformer based model?
- What are the main reasons behind success of transformer-based models?

100. Can we use BERT based model to generate embedding? if yes, How? NO, then Why?
101. What do you think about Open AI GPT3?
102. What is your thought about convolution auto encoder?
103. List down text summarization techniques, and which latest model you will prefer for text summarization?
104. What is the meaning of multiheaded attention
105. What do you understand by BLUE Score?
106. What is gradient clipping?
107. Can you please list down ways by which I will be able to split training across multiple GPUs?
108. Explain me difference between GRU and LSTM.
109. If I have to implement Any hour then what will be your approach? (Hint: using NLTK, Spacy or state of art model)
110. What do you understand by Uni-gram, bi-gram and tri-gram(N-gram)?
111. What do you understand by stemming and lemmatization?
112. For conversational AI solution will you use to predefined framework or will you create your own. In both cases what are adv and disadvantages?
113. Have you worked on google dialog flow, Azure-LUIS, IBM-Watson, or RASA- NLU?
114. What are the limitations of these respective platforms mentioned above?
115. If I have to build a ticket rerouting system for a banking client, how will you design this complete system?

116. Can you please how will you be able to design an app like In-Shorts? (hint: text summarization & etc)

117. If you have to build a solution that can generate a summary of the entire online class meeting then what will be your approach and what kind of hurdles you may face?

118. If I have to create a Gmail kind of text generation system then what will be your approach?

119. If I have to create a document parsing and validation system for legal, what will be your approach?

120. If you have to build a voice-based automation system, how you will design system architecture?

<div align="center">

----- Time Series ----

</div>

121. List down, time series algorithms that you know?

122. How can we solve TS problem in deep learning?

123. Give application of TS in weather, financial, healthcare & network analysis?

124. What is diff between uptrend and downtrend in TS?

125. What do you understand by seasonality in

126. What do you understand by Cyclic pattern in your TS data?

127. How will you find Trend in TS Data?

128. Have you implemented ARCH model in TS? If yes, give scenario?

129. What is VAR (vector autoregressive) model?

130. What do you understand by univariant and multivariant TS Analysis?

131. Give example where you have created a multivariant model?

132. What do you understand by p, d, & q in ARIMA model

133. Tell me mechanism by which I can find p, d, q in ARIMA model?

134. What is SARIMA and how it's different from ARIMA?

135. What is meaning of AR, MA and I in ARIMA model?

136. Can we solve TS problems with transformers? what is your thought on that? why do you think in that way?

137. Have you ever productionised TS Based Model using LSTM? What are adv and disadvantages

138. Can we solve TS problem using Regressive algorithm, if yes, why, if no, give a reason?