Mastercard AI Garage Interview Experiences

# **Slot: 1**

# **Procedure**

1. Test: Yes. (Online)
2. Interview Mode: (CV Raman Building[white building near Hoysala guesthouse])

----------------------------------- Physical ----------------------------------------------------------------------------

# Sofia Sunam

**Personal View**: Online test had 15 Aptitude, Technical MCQ 10 , Coding 2 questions (40 mins). 2 rounds of Technial Interview and 1 HR round. Asked from CV/resume. Only important conclusions from lectures of MLDS and INLP from last sem will help to clear round 1. Round 1 and 2 were of the same difficulty [moderate]. Round 2 had coding for others but I said that I know coding so as to do assignments and also have SDE2 experience. Thus, i guess, they did not ask me for DSA coding. HR round was too generic where they confirmed on location and compensation.

1. Keep CV/resume print out ready; ample number of them

2. Take water bottle umbrella and food

3. Prepare your resume well

**Status**: Attended till final round (HR) and did not get selected.

**Online Test Description**

1. **Aptitude:**

Work done, Mixture of solution, Venn diagram, long puzzle A, B, C, D, E invested X amount and save Y% , Direction and Pythagoras Theorem

1. **Tech MCQ:**

ML stuff – Bagging, Boosting, very simple

CS stuff – Gantt chart RR, stack push pop, give outputs [may be C++], pointers

1. **Coding:**  
   **1.** You are given a binary tree in the form of pairs of node locations and values. The location of each node is given in terms of a string composed of letters L (left) and R (right), indicating how to navigate from the root to the node. For example, L means the left child of the root, R means the right child of the root, LL means the left child of the left child of the root, and so on. A node is called a **Supernode** if the sum of the digits on left node equals the sum of the digits in its right node. Your task is to determine the number of Supernodes in the given tree.

2. Number of occurrences of a part of word

**Interview Description**

## **Round 1:**

Educational Background of the Interviewer (only one member): idk

Project Specific Discussions: explained the projects that I was confident about and related to role and company.

1. What makes transformer stochastic (temperature var)
2. Explain rag
3. Projects on classifier
   * Input n model working and objective
   * Why random forest over knn : Bias variance trade off
   * Is knn parametric or **non-parametric**
4. Is the graph acyclic
   * Dfs algo time complexity
5. Autoencoders
6. Cross Entropy Loss – why not MSE for classification

**Round 2:**

* Different interviewers (one member), with a long history in this domain.
  + Why did we change from rnn to transformer
  + N customers, rsD discount and k% available the discount. So total discount amt =nkd

**Round 3: HR**

* HR round was too generic where they confirmed on location and compensation.
* How do i destress myself
* Short intro and story to IISc
* Previous employer details and experience
* Bangalore experience and what challenges on first time coming out of state

---------------------------------------------------------------------------------------------------------------

# Tejas Tonde

**Personal View**: [will update in detail later] Interviews were moderate-level difficult. Round 2 interviewer focused more on the loss functions used in the projects from the resume. Keep at least 2 copies of the resume.

**Status**: Attended till Round 2. Not selected.

**Interview Description**

## **Round 1:**

Educational Background of the Interviewer: No idea

Project Specific Discussions: CNN,RNN, Transformers , Adam optimizer

* Introduce yourself
* **Adam**
* Started with one of my project related to Adam
* Asked me to write the formula for Adam. Why was Adam introduced in first place. Compare SGD(Stochastic Gradient Descent), SGD with momentum and Adam. What is bias correction term?
* Explain the entire process of how the project was done, experiments and conclusions of the project.
* **CNN and RNN**
* Explain how CNNs work. Given input image, stride,padding, filter size what is the output feature map.
* If you use a 3x3 kernel instead of a 7x7 kernel, what changes will occur? Pros and cons of each case.
* Do we encounter the problem of vanishing gradient in case of CNNs? Why is vanishing gradient more relevant in RNNs than CNNs? Explain vanishing gradient and ways to overcome it?
* Which layers in CNN will get a drastic update in the parameters? (Nothing about the layer was specified so I tried to answer with respect to the position of the layer from the output layer and size of kernel)
* Why is it preferred to use a ReLU activation instead of sigmoid or tanh? (Think in terms of the output range of these functions. Sigmoid and tanh are bounded, ReLU isn’t. Can there be any effect like vanishing gradient?)
* **Transformer**
* Have you worked on transformers? Explain different types of positional encoding in the transformer.
* Explain sinusoidal encoding in detail. Why specifically sin and cos, why not any other function? How does it ensure unique encoding for each token? In a m X n matrix of input embeddings (m = sequence length and n = embedding dimension), is it embedding applied across m or n? Explain with formula.
* **Ungrouped questions**
* Explain the difference between and auto-encoder and a variational auto encoder. What is the difference in their loss functions?
* Explain hierarchal softmax.
* **Puzzle**
* Puzzle: You have a jar of 4 litres and a jar of 9 litres. Having unlimited supply of water, how will you measure 1,2,3,4,5,6,7,8,9 litres?

**Round 2:**

1. Introduce yourself
2. Write binary cross entropy, categorical cross entropy. Write loss function for multi-class and multi-label classification.
3. Picked one project and asked to write the loss functions (dice loss and IoUloss(Intersection over Union). Which one is better and why? Are these losses differentiable?
4. How will you use VAE or an Auto-Encoder for binary classification task? Difference between VAE and Auto-encoder? Explain their loss functions.
5. How is the distance between the distributions measured?
6. If the distance is measured using KL divergence, why is it called KL divergence rather than KL distance? Write the formula for KL divergence.

---------------------------------------------------------------------------------------------------------------

# Rushikesh Pawar

**Personal View**: Emphasis on fundamental machine learning techniques and clear understanding of the projects listed on the CV.

**Status**: Cleared both technical interviews and participated in the HR interview. I was the third person to give the HR interview. Approximately 20-22 students were shortlisted after the online test, and 6-7 advanced to the HR round. The first two candidates were selected for final offers.

**Process:** Online test > Technical interview 1 > Technical Interview 2 > HR interview

Online test details:

|  |  |  |
| --- | --- | --- |
| **Subject** | **Number of Questions** | **Duration (min)** |
| Aptitude | 15 | 20 |
| Technical | 10 | 15 |
| Coding | 2 | 40 |

**Coding questions:**

1. if the sum of digits of the left child node is equal to sum of digits of right child node, then parent node is called supernode. Find sum of all the supernodes in a given tree.
2. Find out number of occurrences of given word in a given square grid (each grid[i][j] is a character). Word can be present in any direction. Left->right, right->left, top->bottom, botton->up, diagonal.

**Interview Description: Data Scientist**

**Round 1: Technical**

**Interviewers:** A person from Mastercard AI garage

**Introduction:**

* Asked me to introduce myself. (I mentioned the courses I’ve taken and my internship project.)

**Internship Project:**

* The interviewer asked for a detailed explanation of my internship project. We had an in-depth discussion covering the problem statement, challenges, my approach, and the results. (This took up most of the interview).

**Technical Questions:**

* What is a confidence interval, and how do you calculate it?
* What is SMOTE? (It was mentioned in my CV).
* Explain the Transformer architecture. Why did Transformers become popular compared to previous methods? What was the prior architecture, what were its drawbacks, and how do Transformers overcome them? Are there any limitations to Transformers?
* Asked a probability question about rolling two dice together and getting at least one "5".
* Asked for the maximum sum of a subarray (Kadane’s algorithm).

**Q&A:**

I asked 2-3 questions about his work and in general advice for a fresher joining the industry.

**Round 2: Technical**

**Interviewers:** A Data Science Manager

**Introduction**:

* Asked me to briefly introduce myself.
* Asked the difference between AI and CDS (Computational and Data Sciences) departments.
* He explained that this round would focus on the breadth of machine learning.

**Technical Questions:**

* What is SMOTE? How is a synthetic point generated, and is SMOTE always helpful?
* If given SVM, Neural Networks, and Decision Trees, in which cases would SMOTE be useful, and in which cases would it not work? Explain why. (This was an open-ended question, and I struggled a bit with it).

**Feedback:**

After I asked for feedback, the interviewer advised me to start with simple cases and gradually increase the complexity. For instance, begin with a simple hard-margin SVM (without a kernel), then progress to soft-margin, and so on. He suggested first considering cases without class imbalance, then addressing imbalanced cases. The question was open-ended and intended to test my problem-solving approach.

Although I faced some difficulty, I asked thoughtful questions, which I believe helped me qualify for the HR round.

**Round 3: HR**

* Asked about my transition from Maharashtra to Bangalore.
* Asked how I handle stress.
* Inquired if I knew about the compensation and location, and whether I was okay with them.
* Asked if I had any questions for him.

---------------------------------------------------------------------------------------------------------------