Consulting & Analytics club
IIT Guwahati

2024-25

# Gradient Ascent

## INTERVIEW EXPERIENCES

A Detailed guide for Internships and Placements focusing on Machine Learning & Data Science

## Varun Tej

Data Scientist | Intern | 2024 | Zepto

The **interview** began with a coding challenge where I was given a data frame with three columns: emp_id, dept, and salary. The task was to identify all employees who had the second-highest salary within their department, and if two employees shared the same second-highest salary, to return the one whose first name appeared in ascending order. This could be implemented in Python or SQL. Unfortunately, as I hadn't revised SQL thoroughly, I was unable to solve the problem effectively.
Following this, the interviewer asked if I was more comfortable with deep learning or machine learning, to which I responded that I had primarily worked with deep learning. The questions that followed covered a range of deep learning and machine learning topics, including:

1. Why parameters are necessary for ML/DL models.
2. Activation functions, their purpose, explanations of the various activation functions I knew, and their derivatives.
3. How ReLU introduces non-linearity.
4. The loss functions I've used, along with their explanations and derivatives.
5. Evaluation metrics for regression problems, their advantages, and disadvantages.
6. An explanation of backpropagation and the equations involved.
7. How backpropagation is implemented in RNNs.
8. The vanishing gradient problem in RNNs, how to mitigate it, and examples where RNNs might avoid this issue.

The interview was challenging, and I struggled to answer 2-3 of these technical questions. Given my difficulties with both the coding task and some of the deep learning questions, I did not progress to the HR round.

## Soham Karak

Data Scientist | Intern | 2023 | Adobe

The **OA** was mostly on basic theoritic ML with MCQ question in testing, decision trees, PCA, SVM, linear algebra, probability as far as i can remember. There was no DSA based coding questions in the round.

Some questions that i remember:
1. What factors affect while choosing between PCA and t-SNE?
2. Some optimisation that you can think of in t-SNE.
3. Role of Bayesian Algorithm in Naive Bayes.

The **interview** consisted of two rounds. The first was primarily technical, where the interviewer asked me questions on linear algebra(eigen value decomposition)and probability(a preety basic quant probability qs).He also posed two questions on machine learning fundamentals. Following that, we discussed the projects listed on my resume. The second round focused mainly on HR topics.

## Eshu Patel

Data Scientist | Placement | 2023 | Tiger Analytics

**Round 1.** The interviewer asked binary search and asked to prove how it's complexity is log(n). Other than this he discussed projects in my CV(not in detail just a brief overview). Then he asked me about my BTP (I am from Chemical Engineering, don't know why he asked this but we had a good discussion on this).
**Round 2.** The interviewer told me I have a dataset with house pricings but some data is missing how can I get that. What data is useful for this like house location, size, etc. I gave an answer based on basic mathematics using mean and median then he asked which is better mean or median and some similar things. Basically +2 maths only.
Some questions that he asked:
1. Describe how central tendency measures such as mean and median can be used as features or to handle data imbalance.
2. What imputation techniques do you know and how do you choose between them based on the datasets or algorithm you are going for.
3. In a large dataset with many columns, how would you choose the relevant columns. The dataset can have columns which have string values.
**HR Round.** HR asked why I want to join for this just go through their JD and tell something good about them and the work they are doing. She also asked that the job location is Chennai are u ok with this. I said yes as for Chennai only they were hiring.
They asked some other questions like rotate a matrix by 90 degrees. This was not asked to me but to some other persons.
Just show some curiosity about company and show as you know something about them. This will show your interest in the company.
That's all from my side

## Sai Teja Reddy

Data Scientist | Placement | 2023 | Tiger Analytics

The overall process consisted of an online assessment and 3 rounds of interviews.

**Online Assessment**
**2 sections**
**Section 1:** aptitude, mathematics, and few technical questions.
**Section 2:** 2 coding questions, both easy, only Python was allowed. Although the questions were easy and didn't require any complex algorithms, each was timed with a relative short window, making it challenging to both read and write the code within the given time.

**Interview Round 1**
The first interview began with a brief introduction where I was asked to talk about myself. Afterward, the interviewer reviewed my CV and asked me to explain a few terms listed under my projects, such as the variance inflation factor, log-likelihood score, and basics of decision trees.
Then he passed me a white sheet, and asked to write complete code for a coding question. The question was simple: a string of lowercase English alphabets was given, and I needed to return the frequency of all characters that appeared in alphabetical order. When I asked which programming language to use, the interviewer told me that I could choose any. I opted for C++ since I was most comfortable with it. I used unorderd_map for the question. After I explained my code, the interviewer asked about unorderd_map and how it works.

After that, he requested I solve the problem using a different approach, so I implemented it using vectors. Then, he asked about the complexity of the solution's space and time. Further, he asked what modifications I would make if the string included special characters or uppercase letters, and what changes to make if the question was to print according to the order of occurrence instead of alphabetical order.

**Interview Round 2**

The interviewer greeted me warmly when I entered the room, and the session began with a brief introduction. He then inquired why I was not shortlisted and was on the extended shortlist. I wasn't sure how to respond, but shared my thoughts on why I might have missed the first round of shortlisting. He gave me a blank sheet and started with aptitude questions and puzzles.

The questions were primarily based on probability and statistics. The early questions were fairly straightforward, but as we progressed, the difficulty gradually increased. After every answer, we briefly discussed the approach I had taken. He asked me around 13 questions before I could not answer, after which he stopped and asked if there were any questions for him on my part. I asked him how to solve the last question, to which he first refused, remarking "tum bahar jaa kar sab ko iska solution bata doge, fir maja kharaab ho jayega". After a lot of requests, he finally agreed and explained the solution.

**Round 3 (HR round)**

The HR round was quite friendly and conversational. I was asked to Introduce myself, then the first question she asked was, "Why Tiger?" and it was the only tricky question on her part. Then she inquired if I was willing to reallocate to Chennai. Towards the end, she gave me the opportunity to ask any questions I had. I asked her a couple of questions about the company and role, after which the interview was over

## Swateya Gupta

Data Scientist | Placement | 2023 | Turing

The main selection was done based on 5 rounds. After that we had an **HR telephonic interview** which was sort of formality because they did the main selection who cleared all 5 rounds, it consisted of two coding questions one on each round, then a round on analytical and general reasoning questions. 4th round consisted of two model responses and were asked which model is better and why. In the last round, we had to analyze data of a CSV file and 11 MCQs were asked for that CSV file, we had to use python data analysis libraries like pandas, NumPy, Matplotlib etc.

1. What methods would you use to identify outliers in a CSV dataset?
2. How and when do you use transformations like log, square root, or Box-Cox.
3. Difference between one hot encoding and label encoding.

# Maharsh Raval

Data Science | Intern | 2023 | Mastercard

**Round 1** (45 mins)

- In the coding round, we had 2 hours to solve two basic questions from strings and arrays, with a difficulty level similar to 900 to 1000 rated Codeforces problems.
- The first technical round primarily concentrated on probability and logic puzzles. Standard puzzles like the burning rope puzzle, a pair of socks problem, and a probability question involving dice were asked.
- I was questioned on fundamental machine learning concepts and basic SQL, such as various types of joins and I was also asked to solve some problems including joins.

**Round 2** (30 mins)

- The second round mainly revolved around my resume. However, the interviewer was particularly interested in my project on cab fare estimation, and most of the interview was focused on discussing that project in detail.
- After that, he asked me about Mastercard and what I could bring to their team if I got hired.

**Round 3** (HR Round)

The last round was more of a formality. It was mainly to inform us about our selection and check if we knew about the stipend and location.

# Souraja Kundu

ML Intern | Intern | 2023 | Samsung Noida

**Round 1**

- He went through my CV while I was giving my intro. One suggestion is don't make your CV too much stuffed, leave enough spaces.
- Then he asked me to explain in details my last year's research experience in Japan and also one of my self projects. Having at least one good research project completed in depth helps to give you an edge over others. You may try to work under a professor if you don't have any unique project idea. Then he started with Machine and Deep learning general knowledge.
- Then he started with Machine and Deep learning general knowledge. Some questions that I remember are: What is salt and pepper noise and how to remove that from images, multi layer perceptron weight update equations using backpropagation, different edge detecting filters and CNN stuff, dimensionality reduction etc. Then he moved on to DSA part. Leetcode medium level and gfg would be enough to practice for this part.

**Round 2( HR Round )**

She asked me to explain 2 projects from my CV.Then asked my background and asked to tell the story of my journey focusing on academics and research. Finally she asked me to tell the contribution of Samsung other than androids and home applications. You should be ready to answer this question for HR round and should tell some points which generally people can't think of. This shows your dedication for the company.

## Ankit Varshney

Data Science | Intern | 2023 | Mastercard

**Round 1** (60 mins)

- I used an Introduction to Statistical Learning by springer to learn stats, Pattern Recognition and Machine Learning by CHristopher Bishop for ML, and ANdrew Ng's deep learning specialisation for DL. I also used MIT's lin alg courses.
- The aptitude test for InfoEdge consisted of mcqs related to stats/math, DL and ML. The technical interview had theoretical questions related to lin alg, analytics, stats, probability, DL and ML. It also had a coding round where we had to make an ML model. The CV round was next and consisted of questions on your skills, followed by the HR round about your interests.

## Daksh Kaushik

Data Science | Intern | 2022 | Publicis Sapient

**Round 1** (60 mins)

- A DSA question based on shortest path DP. Although DSA questions don't carry much weight in the data science interview still you should try to crack these questions
- Moving on they asked me about my project, I told them one of my projects which was license plate detection where me and my peers used the YOLOV5 algorithm. So they asked me about the same in deep. I want to say that whatever project you do you must have a deep understanding of the underlying algorithms used in that project.
- There were questions about situational ML which I had to answer on the spot. These types of questions are easy if we know the model-building pipeline.
- The interviewer asked me to estimate the number of customers coming to the mall in a year. He gave me some parameters and asked me which algorithm I would use for the same, how will you check outliers, and check the efficiency of the model.
- Then we moved on to some standard HR questions.

# Lokesh Nahar

Data Science | Intern | 2022 | Publicis Sapient

**Round 1** (60 mins)

- I was asked about OOP's, DBMS particularly SQL, and the Basic difference between SDE and data science.
- They also asked some technical questions on ML like Bias Variance, supervised unsupervised learning, Use of NLP, and what all was covered in the NLP course I took( like word embeddings, Word2Vec)
- Then we moved towards project discussion. They asked me about my favorite project(about face recognition)and then questioned me about the same.
- Then they asked some basic SQL questions along with some jee types probability questions. I also told them about my JEE Adv rank and some research work I was doing. I also told them about my case study under the CNA club
- Then we discussed in what ways I can go deeper into the fields. For e.g., We can use Recommender systems and tools like NoSQL, Azure, etc.

Be aware as well as humble in the HR round. , they asked me about 5 company values. It was mainly based on a resume. Sidenote puzzles can also be useful

# Shamith K

Data Scientist | Intern | 2024 | Zepto

I recently participated in an **interview round** that followed a machine learning (ML) assignment based on time series analysis using the company's data. Out of the candidates, 10 were shortlisted for this round, which lasted for 45 minutes.
In the initial 5 minutes, I introduced myself, emphasizing my experience with finance-related projects from my first and second years and mentioning that I was still in the early stages of my ML journey. This led the interviewer to ask more about the probability and statistics aspects, along with some questions related to the assignment provided.
I then explained one of my finance projects, which involved deploying an ML model. Up until this point, the interview was going smoothly. However, as the discussion delved deeper into the technical intricacies of ML, my performance had some hits and misses.

Unfortunately, I was not able to clear this round. Out of the 10 candidates, only two progressed further, with one eventually being selected.

# Amish Agrawal

ML Intern | Intern | 2022 | Decimal Point

**Round 1:** Aptitude Test and Coding Test

- He started with a basic introduction and then moved on to my Resume.
- He asked about one of the projects and asked me to explain it to him. Ensure you know everything about your project and can respond with any cross-questions based on implementation details. Whatever I was explaining, he cross-questioned me on some things related to the project. This went on for about 20 mins.
- After that, he asked me about ML basics. Questions include the difference between bias and variance, what we mean by high bias/variance, how it can be tackled, and so on. He expected a fundamental understanding of the concepts, cross-questioning me whenever I used a big word and asking me to break it down in simpler terms.

(.Try to use examples to explain your point, I used the whiteboard to draw and explain high bias/variance; similarly, you can take the help of such tools to explain yourself.)

- Finally, he asked one question that he was asking everyone in the end. It was based on PCA. On the whiteboard, he made 2D axes and plotted some data points. Then he drew 3 lines naming them PC1, PC2, and PC3. He asked which ones were the principal components of the data. The one capturing the max variance(you can look at the diag and tell) was the first, and we know that PCs are orthogonal to each other, and for n-dimensional space, there is max. N PCs, so the 2nd PC was perpendicular to the 1st one, and there is no 3rd PC.

In the end, he asked me if I had any questions, so I asked about the kind of projects that are allotted to interns.

**Round 2**( HR Round )

I got a call from HR after all the other interviews were completed. It was basically an HR round in which they asked typical questions like what are your strengths, gave you a situation, and asked what you will do. It was just formality, and it meant that you were selected.

# Akshay Chintala

ML Intern | Intern | 2022 | Providence

**Round 1:** Aptitude Test and Coding Test

- They asked me what I'm good at to which I said probability. They asked some basic questions in probability and statistics.
- They asked linear algebra although not that much.

**Round 2:**

- They asked me about the project I made.
- Mainly stuff like what is the purpose of the project and the problems I faced during it.
- I made a project on Braille to speech converter which was a CNN model and I got questions related to that project only.
- They asked me first what the project does and why I made it. They asked what problems I faced when I was working on it and what was my contribution in that project.

# Umang Jain

ML Intern | Intern | 2022 | Nimbleedge

**Round 1**

- In the first half, he asked me various ML-related questions, like Regularization, Bias Variance, Kmeans, and KNN, etc.
- In the second half, he asked me to explain him any one of my projects. So I had a project which was a combination of both NLP and CV domains, so I explained that project to him , and he asked me various questions that he had about the project.
- This round wasn't that difficult (Also they were mainly searching for people with NLP backgrounds)

**Round 2** (90 mins)

The second round was with the founder. In that round, they were mainly checking how strong the basic knowledge is, and they asked me several situation-based questions.

- Suppose you have two decision trees D1 and D2, and a dataset T1 and T2 which were a subset of a common large dataset. D1 and D2 are trained on T1 and T2 respectively. Now you have to make a third decision tree D3, such that it is robust to both datasets T1 and T2, but T1 and T2 are not accessible to us, and only have access to D1 and D2.

This question was asked to check the thinking in case of an unseen problem. There was no right or wrong answer in this case.

- Another question he asked me was, the difference between SGD and GD, which one is better, and why it is better, on an intuitive understanding

- Next was, to consider SVM. Suppose we train an SVM multiple times, with different random initializations at the start, still, why do we get the same weights once the training is complete every time? The answer was, that the loss function of SVM is Convex, having a global minimum, and SVM always reaches the global minimum perfectly, giving the same weights at the end.
- Then he asked me to write pseudo code of a two-layered neural network for binary class classification, from scratch, both forward and backward propagation.

# Pragyan Banarjee

MDSR Intern | Intern | 2022 | Adobe

**Round 1** (45 mins)

- I already had a research internship before that, so for the first 30 mins , the discussion mainly revolved around that only , we discussed what I did in that internship and what more could have been done. They asked me what was the outcome of the project? Did anyone continued the work after my internship was over, or if there was any chance of a publication.
- Then they asked simple questions on CNN (I told them that I was primarily interested in CV)
- They also asked me a question in which there were 2 matrices each representing part of an image ... What could have been the best ways to concatenate them to make a single image
- There were lots of variations to this question... And we sort of had a discussion on this topic... Discussed some ways to do them and their pros and cons
- simple stats-related questions like suppose there is a flow of data (data that is coming real-time), how to calculate its mean and variance on the go (real-time)

I'd suggest focusing on the basics and how you'd apply them to solve simple real-world problems because apart from projects they usually ask basic questions only.

# Debarshi Chanda

Data Science | Intern | 2022 | Amazon

**Round 1** (45 mins)

- Write a function to print the Spiral Order Traversal of a Binary tree. (Already knew the solution, said the solution using 2 stacks. Interviewer asked if the same can be implemented using a single queue.Said about doubly ended queue but hadn't used it before so told the interviewer honestly that I hadn't used deques before. Went on to code the 2 stack solution)
- An array is given, Find the length of the subarray having maximum sum. (Told the Brute Force approach O(n). Modified Kadane's algorithm to maintain length of the array as well)

**Round 2**

Discussion on Project (Amazon ML Challenge):

- Explain Problem Statement of Amazon ML Challenge
- Explain the Approach, Follow-up: Mathematical formulation of Arcface, Told him didn't know and gave the intuitive explanation.

Asked me about the performance of various models we tried. Follow-up: XLNet performed better than BERT. Why? ML Design (Interviewer said that there was no right or wrong answer):

- Suppose you want to use all the columns (Title, Description, Bullet Points). What can you do?
  - 1st Design: Use 3 BERT models for the 3 columns respectively, concatenate embeddings and use a fully connected layer.
  - 2nd Design: Use 1 BERT and use [SEP] token as a separator between different columns.
- Pros and Cons of Both the design.
- Which model will you choose for your leaderboard? 1st
- Is there any case when the 2nd model will outperform the 1st model? When the columns have a similar distribution.
- Which model should we use if we only have 3000 rows instead of 30 lakh?

ML Fundamentals:

- Metrics: Suggest a metric other than accuracy. F1-score. Follow-up: Define F1-score. Define Precision or recall. How would you calculate F1-score for 10000 classes?
- Suggest something other than F1-score. AUC score. Follow-up: Define AUC. Statistical significance of AUC.
- Regularization in NN: Dropout, How to deal with different behavior of dropout in train and test time. Scaling.
- Batch-Normalization: What is Batch-Normalization, Why is it useful? Same scale and minimize covariate shift.
- Difference between SGD, Batch Gradient Descent and Mini-batch Gradient Descent.
- Linear Models: Difference between L1 and L2 regularization. Why L1 induces sparsity and L2 doesn't? Why use L2 at all?
- Can linear models overfit to the training data, it is just a line. (Outliers)
- Unsupervised: Tell any Unsupervised learning algorithm to decompose 10000 dimensions in 10. PCA. Explain PCA in 2 lines. How do you get the eigenvectors? You have the 10 eigenvectors, now how do you convert the 10000 dimensions to these 10 dimensions.

# Shreya Sajal

Data Science | Intern | 2022 | Amazon

**Round 1** (45 mins)
2 DSA questions, interviewer shared the link to the live coding platform. Fortunately I was able to solve both.

- Intro
- Longest Common Substring (brute force, then explained DP approach, Time-space complexity, code and dry run)
- Density of Binary Tree in one traversal (started thinking using DFS but BFS approached clicked at the right moment, wrote the code)

**Round 2** (70 mins)
The interviewer started with his intro, he was a Senior Applied Scientist at Amazon and he mainly worked in the NLP domain.

- Moved onto my intro, ML exposure overview, Brief overview of 2 CV projects + Amazon ML Challenge approach and my contribution.
- Went on with discussion on my Amazon ML Challenge approach further as we had used 2 approaches (End-to-End Classification and Transformer model head with KNN, and and ensemble for the final predictions):
- What was KNN over embeddings exactly doing (Inference time Euclidean distance calculation and voting)
- How did you train 3M points? (Stratified K-Fold)
- You have a sentence and you have to do POS tagging for every word in that (3 POS were there noun, verb and pronoun. I started with basic RNN approach and explained, again the main focus was on output activations, dimensions and loss function at every time step)
- If you didn't have any memory limitations and could train all the points, then I claim End-to-End Classification should work better than KNN + embeddings. Justify my claim. (He assumed that the extracted embeddings were not finetuned and we were directly using the pretrained weights for embedding extraction, so I just justified it based on his assumption only as it was easier that way)
- If the max_length was some fixed value for your classification model that couldn't be changed and text length > max_length what would you do? (Trim- info loss, so best I could come up with text length // max_length rows for one row with labels same for training, max vote for segments during inference, he seemed satisfied)
- What else could you have done if there was no such time limit for the hackathon? (Stack embeddings + KNN and few more points from our Future prospects sections mainly)
- What if you were not allowed to use BERT or any transformer based models? (I explained him a solution using Bi-LSTM many-to-one, the main focus was on the basic architecture not much depth into the u, f, o gates, and also he focused on the output dimensions and loss function for training)
- What if it was multilabel and not multiclass? (He mainly wanted to know the output layer dimensions and activations to be used)
- You have a sentence and you have to do POS tagging for every word in that (3 POS were there noun, verb and pronoun. I started with basic RNN approach and explained, again the main focus was on output activations, dimensions and loss function at every time step)
- Design a search engine for Amazon, you have product descriptions and a text query given. (I explained a sentence embedding + semantic similarity + ranking approach, he was satisfied)
- Follow-Ups: What if you had a binary column as well for every product? (I first said that we can multiply each row's binary with it's cosine similarity we had calculated, so the disliked one's get filtered, he said it was okay as a baseline but there will be some info loss in this, I then said we can change the weight from 0 to 1 and from 1 to 2, then none of the rows will be completely lost and the not similar-liked would not get an edge over the similar-disliked ones mostly, he sounded satisfied)
- Explain BERT architecture (I explained mainly how it is trained and the tasks it is trained on- NSP and MLM and the input it takes, input word embeddings and some basic transformer encoder stack attention overview, he was satisfied)
- Follow-Ups: What is masked language model? What if the masking was not done? (I couldn't answer) What is language modelling?

# Akanksh Khandelwal

Data Scientist | Intern | 2024 | KLA

The **eligibility criteria** for this interview was a CPI of 7.5 or higher, and only students from CSE, MnC, DSAI, EEE, ECE, and EP were allowed to participate. The **online assessment** consisted of multiple-choice questions focused on machine learning (ML) and data science, along with one easy coding problem. Six candidates were shortlisted for the interview round.

During the **interview**, I was first asked to introduce myself, followed by questions on principal component analysis (PCA), matrices, linear regression, and the time complexity of matrix inversion. The questions were largely centered on data science and algorithms. As I was not specifically targeting ML services, I repeatedly requested the interviewer to ask questions related to High-Performance Computing (HPC) or Competitive Programming (CP). Toward the end, he asked a simple dynamic programming (DP) question, which I was able to solve in two minutes.

Although the interviewer was pleased with my performance, I believe they were seeking someone with deeper expertise in AI and ML. Ultimately, no candidates were selected from the interview process.

# Arsh Kandroo

Data Science | Intern | 2022 | Microsoft

**Round 1** (55 mins)

- Picked up a project from my resume, and asked me to explain the approach. Inquired about the model that I used and how it works, which was XGBoost.
- Basic questions on Gradient Descent (slope of derivative etc)
- Asked me to share the screen and write the formula for Gradient Descent optimization with momentum. Further, he asked questions about the intuition of the same and the significance of every term in the formula. Moving ahead, he asked me to recursively open up the formula and asked some questions based on that.
- Basic questions about Mean, Variance and their formulas.
- Eigenvalues and Eigenvectors, positive definite matrix and Hessian matrix.
- Lastly, he asked me to share my screen again and code the function for cosine similarity between two vectors.
- Interview ended with me asking him about his projects and stuff.

This guy's camera was off and didn't seem jovial. So, I tried to talk to him in the end about his projects and work to ease things off a bit.

**Round 2** (50 mins)

- Discussed 2 projects from my resume thoroughly, from motivation to the models used. Make sure you know the functioning and intuition behind each and every model or the library that you have used in the project. Since one of my projects was based on NLP, which resulted in him asking me questions about BERT and LSTMs. Also, there were some questions based on shortcomings of the project; make sure you have already tapped them and you should know how to overcome them.

- Gave me a situation in which I've to classify queries on a search engine as technical or not. Told him my approach and the models that I could use for better results.
- Asked me to share my screen and gave me a small DataFrame. He wanted me to implement an easy method in pandas on the DataFrame. Though my answer wasn't what he was expecting but he was satisfied.
- Again, the interview ended with me asking him about his work in the MS and what team he has been working under.

This guy was really cool. Smiled throughout the interview. He had done his MTech in Data Science from IIT Madras. Also, we talked about his journey in MS.

**Round 3**

- Started with him asking me about my journey in ML and explaining all my projects in chronological order
- He picked up a project from my resume and asked me to tell its shortcomings and how'd I handle them.
- Gave me an ML case study.
- "Whenever we search diseases related to symptoms on a search engine, it may show us extreme diseases on top, hiding the real disease. We want to develop an app when given symptoms as input shows us all the probable diseases along with their probabilities.
- I explained my approach from data collection to deployment. He asked me the challenges I'll face and how will I handle them.
- Towards the end, we talked about life at MS and the projects that he had worked on. Also, I asked him for feedback of my interviews, which turned out to be pretty good.
- This guy was really cool as well with 17 years of experience at MS.
- General Tips
- Whenever explaining your approach, make sure you start from the beginning like data collection and move ahead step by step towards data cleaning, preprocessing, model development and metrics.
- Make sure to tap the shortcomings of your project beforehand and the answers to tackle them.
- Communicate clearly and try to discuss thoughts with your interviewer. He/She will definitely guide you towards the correct answer.

- I explained my approach from data collection to deployment. He asked me the challenges I'll face and how will I handle them.
- Towards the end, we talked about life at MS and the projects that he had worked on. Also, I asked him for feedback of my interviews, which turned out to be pretty good.

This guy was really cool as well with 17 years of experience at MS.

**General Tips**

- Whenever explaining your approach, make sure you start from the beginning like data collection and move ahead step by step towards data cleaning, preprocessing, model development and metrics.
- Make sure to tap the shortcomings of your project beforehand and the answers to tackle them.
- Communicate clearly and try to discuss thoughts with your interviewer. He/She will definitely guide you towards the correct answer.

# Rohan Shaji

Data Science | Intern | 2022 | Microsoft

**Round 1** (40 mins)

- First asked me about one of my projects. Described what it was, when it was done(as part of Summer Analytics final hackathon). Described what was special about the data(duplicates, class imbalance). Told him I used SMOTE to solve the imbalance. Described how it works in brief. He asked a counter question, which was actually one of the drawbacks of the SMOTE method.
- Asked if I know Python. Said yes, but not OOPS. Asked me to code a basic function on Notepad (find cosine similarity).
- Asked me about momentum in gradient descent. I did not remember the formula(and I told him that). Opened Paint and drew the diagram of what momentum does. He realized I knew how it works, so he told me the formula and asked me to explain why momentum works(i.e. Why those mathematical formulas speed up gradient descent). After a few hints, I was able to solve it.
- Asked a question on statistics, was not able to answer it satisfactorily so he moved on. Asked me to write the formula for expectation, variance of a Continuous Random Variable.
- Asked me if I have any questions for him, I asked him about the work he does.

**Round 2**

- Chillest interview ever. Interviewer was a friendly, young guy. Asked me to describe one of my projects and what I had done(data cleaning). Explained the scraping and cleaning part in detail.
- What is supervised, unsupervised, semi-supervised. Give a real life example of supervised(eg: learning what each color is called)
- Explain bias and variance. Explained it, and drew the graph of bias-variance tradeoff.
- What are activation functions? Why use them? List the common ones.
- Gave me a diagram showing some data in 2 dimensions, asked if I would use logistic regression or decision trees for classification. The key was that the data points were not linearly separable.
- Suppose you want to detect diabetes, cancer, heart disease etc. using neural networks.
- A person could have multiple diseases. Do you build one model for each disease or one model to detect all the diseases? Why?
- Most interesting question: Given the weight, length, region and color of a mushroom, predict if it is poisonous. For such questions, make sure to ask questions(do we already have the data? If yes, do we have a lot of data? Do we have time and resource constraints? etc.). Told him my approach, how I'd preprocess(nulls, duplicates, outliers), what model I'd use(and which ones you would not), and why. Make your thought process known to the interviewer. Communicate.
- Then he asked if I had questions. I asked him what work he does, how interns are allotted teams, and what an intern’s day looks like.

**Round 3**

- This guy was the boss of my second interviewer. Again, very chill. Asked me to give an intro. Told him how I started doing ML(because I like math). He then asked me why I still stuck to ML after one year. While answering this, I made sure to talk about Microsoft for brownie points :D
- This interview was discussion-based, with him giving me a problem to solve. I had to describe the entire pipeline, from obtaining data, to model selection and tuning, to deployment. He asked questions in between. Asked me how I’d assess my product. Had a brief discussion on that.

- Discussed additional features I could create. He seemed satisfied, and said we were done(25 mins into the interview). Then he asked me if I had questions for him. Asked him about the work he does, his opinion on things like AutoML, Github CoPilot etc, and finally asked for feedback. In the end, it was clear to me that I was through.

This guy was really cool as well with 17 years of experience at MS.

**General Tips:**

- Do not fool around if you do not know something. Tell them you do not know it, they'll help you.
- Communicate clearly. Explain all your ideas.
- Be yourself, do not act fake. The interviewers want you to do well, they are not your enemies.

# Rishon Dsouza

Data Science | Intern | 2022 | Microsoft

**Round 1**

- (CODING SECTION) Very standard parentheses matching question based on the stack DS was asked.
- (ML SECTION) In this section you control the interview. He asks you whether you are familiar with a topic and then questions you based on the terms you use while explaining the topic. They asked me to explain Linear & Logistic regression. Since this was a very basic question I was expected to go into the advanced mathematical ideas of it. Some topics that were asked: Polynomial LR, Convex optimization problems, Bias Variance, Metrics used in classification, Closed form analytic solution

**Round 2**

- This interview while technical didn't go into direct theory questions in ML. These were application based questions designed to test innovative thinking and good practices as a practical data scientist. Question was open-ended. You were expected to ask for further clarifications. Ask as many questions as you want and modify your approach based on how the interviewer responds. Every step of the problem you solve the interviewer throws in another piece of the problem.

**Question:** You are a wildlife photographer and you have captured 20 million photos with 5 million photos from each continent. A lot of the photos are random trees, grass etc. There could be multiple photos of the same species. Devise a system (taking the scale of the problem in mind) to identify exotic

species (species not seen before in that continent). You also have a wildlife expert at your disposal for consultation. However, being a human he can only look at a limited number of photos and charges about 500rs per photograph given for consultation.

**Answer:** This was a complicated question and required several clarifications. It has to be solved respecting the scale (compute capabilities given size of dataset) and integrity (accuracy and non-redundancy) of the problem.

I used a pre-trained CNN (removing the last few layers) to extract features from the images. Then used the unrolled feature vectors on a clustering algorithm to form object clumps of the same type. This way similar species and unwanted images(like grass) are grouped. Taking a representative from each group we can manually weed out the bad clusters. Now we can use the expert. Interviewer then added a constraint that you can only ask the expert 10 questions. Here we further reduce the clusters to only the “exotic” species of that continent by using Anomaly Detection (using Gaussian similarity).

**Round 3**

- This was the final interview taken by the head of the ML division. This was a mix of an HR and Technical interview. We went deep into my projects (about 30 min). Since my projects had a visual component I offered to run them in front of him. He also asked how I would approach building a chat bot for IITG freshers. I had to mention what sort of data I would need, where I would collect the data from and what features I would add. He didn't expect theoretical knowledge, just wanted to know my thinking process. Overall this interview relied on soft-skills and thinking out loud.

# Ishant Khurana

Data Science | Intern | 2022 | Fractal Analysis

**Round 1** (Aptitude Test and Coding Test)

- In the first step of this round, we have to solve some basic Aptitude questions
- The test contains 70 Aptitude questions, separated by 4 sections ( in 75 min ) Data Analysis, Reasoning Ability, Quantitative Ability, and Verbal Ability.
- (According to me, this is the first and most important part of the interview process as you have to divide your time wisely and should invest your time in the right questions, else you can waste a lot of it.)
- In the next step, we have to solve three programming questions
- There are 3 coding problems, level of these problems is medium level, there’s one more thing that you have to solve these questions in some specific language, (you can’t use C++ or C) ( in my case the language we are supposed to do code was python)
- (This is an elimination round, so please be sure you do all the three questions else your chances are less to get selected for the further interview process.)

**Round 2** (Technical Interview)

- This interview round lasted for around 30 minutes. Most of the questions were related to my projects. They are expecting that you should have detailed information about your project, from it’s codebase to it’s working. Then the interviewer asked me some simple probability and stats problems, kinda medium level. (In this round they just check your confidence and your knowledge in the things you showed in your resume. Be prepared with a nice Introduction for yourself, to set a good tone for the interview at the start.)

**Round 3** (HR Round)

- This is the last round of the interview progress. In this round too first there’s a bit of question on my projects. After that it’s all normal HR round questions.

# Aryan Meshram

GenAI Data Scientist | Intern | 2022 | Fractal Analysis

**Round 1** (Aptitude Test and Coding Test)

- It was an online test consisting of two rounds. First round was an aptitude test which had questions of logical reasoning, mathematical thinking, English as well as statistics.
- Second round was a coding round which had moderate to tough questions. I was able to solve 2 out of 3 questions.
- My performance in the coding round was enough to make me believe that I would not get selected (once again) for sure.

**Round 2** (Technical Interview)

- Technical round was the one I was most scared of. It was a 15 minute round (much to my relief that I would not be asked to code a question). I nervously joined the call at 4 pm.
- The interviewer was a man who didn't turn on his video. He started off with my introduction. I introduced myself and told him the past experience I had in DL. Much to my amusement he asked me about the second project I had in my CV. I explained to him whatever I could. He seemed satisfied and based on my answers he asked me some basics of neural networks, loss function, gradient descent etc.
- By the time I finished my answer, time had already ended for my interview and seeing his hurriedness, I chose not to ask him any questions at the end (which I regretted a lot at that time).

**Round 3** (HR Round)

- This was a very interesting round. The interviewer was a lady who asked me to introduce myself. After she heard my prepared answer, she started a casual conversation about me, how my friends thought of me, my hobbies, what makes me different, what expectations I had from the company, where do I see myself in the next five years etc.

# Udit Jethva

Data Scientist | Intern | 2024 | Adobe

**Round 1: Online Assessment**

Consisted of problems from basic statistics and probability fundamentals. MA225 should be enough for the kind of questions asked. There was no question based on ML fundamentals as far as I remember. 2-3 Logical Reasoning problems were also asked. There were 2 DSA questions at the end, easy ones. People who scored well were shortlisted irrespective of CGPA, Branch. (My guess)

**Round 2: Technical Interview**

Everyone had a 30 min long online interview with the time slots fixed. My interviewer was a PhD researcher working at the Bengaluru office. We began with an introduction. [Your introduction is crucial as it can literally give direction to how your interview goes. Of course companies will ask what they have to ask, but this may sway some questions to the topics in which you are comfortable. Mention your strong topics, courses, relevant hobbies (like Competitive Programming, Kaggle, CTFs).] He then asked me to explain a project which I was most confident about. I picked my 4th project which was a Wordle solving bot [Whenever an interviewer asks you to pick up a project, pick the one on which you can talk the most. This will consume interview time and project discussion is something in which you have the reins of the interview and can be used to make a solid impression. In my case, the project also involved some mathematics which made it much more relevant.]

# Varun Yerram

MDSR Intern | Intern | 2022 | Adobe

**Round 1** (ML Round)

- Explanation of one of the projects, with cross questions on it.
- From a stream of input numbers, find the following at any point of time:
- Mean
- Variance
- A Sample number from the stream with probability 1/n (n = number of numbers read)
- Given an unbiased coin, design an experiment to select one of the three numbers uniformly:
- Follow up - what type of distribution will it form.
- Follow up - Calculate the probability mathematically and prove that it converges to 1/3.
- Given a sequence of heads, calculate and find the number of times you need to flip to get HH.
- Given a matrix where rows and columns are sorted independently, design an algorithm to search a number in an optimal way.(O(2n) i.e. O(n) possible)
- Explain cross connections in UNET and what is the difference between them and residual connections in ResNet.
- Are the parameters shared in a Unet encoder and Unet Decoder, what is the effect of sharing them?
- Explain Backward pass of Unet, more specifically how do gradients propagate through decoder to encoder.
- What does loss.backward() do in pytorch? how will its functioning change in the above question. (Depending on your choice of framework question may vary)
- At last I was asked to explain an NLP project I had done, as he had an NLP background.

# Pranjal Verma

Data Science | Intern | 2022 | Envestnet Yodlee

**Round 1**

- Sort numbers in an array
- Select all those cells in the data frame under given conditions and change the value of only those
- Reverse the string using a one-liner
- How to deal with Missing data?
- Techniques one can use to find outliers, and how do we treat them?
- Some technical questions about seasonality, trends in a time series data, and how to find them?
- What is True positive, false negative, sensitivity, recall?
- Define one case when a recall is preferred over precision and explain regularization.
- Explain the working of Random forest
- Explain the working of linear regression and how the loss is calculated?
- Name some of the charts and explain the boxplot
- Rest all were resume-based.

.

I tried explaining to him the project which took a while because we didn't have the luxury of a pen and paper. The project involved greedily optimising the next move in the game which required something similar to probability distributions (You can check it out on my github profile: u-d-ash). This math interested him and we also had lengthy discussions on the methods which I was using and alternatives which could've been used to improve results. He looked very happy with the work I had put in and expressed a desire to work with me.
Note: My project was pretty elaborate which took most of my interview time but my peers were asked probability, ML fundamentals, math puzzles as well.
And yes, I got in.

# Lokesh Nahar

AI Consultant | Placement | 2023 | SiriusAI

I walked inside the room for an **offline Interview Round**. The panel consisted of the CEO, the AI Director, and the HR representative. We started with introductions, where I briefly shared my background in Data Science, Deep Learning, and Natural Language Processing. I mentioned my passion for exploring and implementing AI solutions, which seemed to resonate well with the panel.

As part of my introduction, I highlighted my internship experience, particularly a project focused on Responsible AI. They were intrigued by the topic and asked me to explain it in detail. I walked them through the problem statement, the methodology, and my contributions to the project. This led to some follow-up questions about the frameworks I used and why I chose specific evaluation metrics. Initially, I was a bit nervous and fumbled, but they encouraged me to stay calm, and I took a moment to compose myself before continuing. By the end, they seemed satisfied with my answers.

The next question was about my favorite project, for which I chose my course project using Large Language Models. I explained the architecture and key concepts, and as they probed deeper into the technical details, they asked me to explain the Neural Net Transformer, the Multi-head Attention, and BERT architecture. I was able to answer their questions, and they appeared pleased with my responses.

Once the technical part of the interview was done, HR asked about my leadership experience, specifically my role as the Convener of SPICMACAY IIT Guwahati, where I led the team organizing the Virasat’24 festival. I shared my experience managing the event, working with artists, and handling budgets.

# Vatan Narang

Data Scientist | Placement | 2023 | HiLabs

**Test :**
MCQ questions on Data Science. Questions were from varied topics of ML and algorithms. Less weightage on DL and more on ML and data science.
3 levels of questions of marks 2,4 and 6 respectively.
2 DSA questions. Easy - 1 from Binary search, 1 from general string manipulation.

**Interview :**
3 rounds. Level increases at each round. Even the interviewer level has the same order. 1st round was taken by some person with 1-2 yrs experience, 2nd with 4-5 yrs and my third round was taken by head of data science in HiLabs

**Round 1:**
Started on CV. I took time to explain my projects in great detail because I saw that he was interested in it. Specially if you have an intern in ML field, huge plus.
I tried to impress him with my CV and projects explaining them in great detail.
Some general basic level ML questions on standard and simple ML algorithms and techniques.
1 DSA question - Rotate 2D Matrix by 90 degrees.

**Round 2:**
Started on with general ML questions increasing the level question by question.
Also included some case study and situation type questions. I think he wanted to know upto what extent I can think as a Data Scientist.
An example question - I work in a company that sells credit card to college going students. As a data scientist, what type of data would best suit me to have such that I can train model on it or what data is the most insightful from business perspective.
I was also asked about how would I handle large datasets, specifically how to optimise the dataset so that it doesn't become an issue in production. They also asked about how would i handle underrepresented classes in the dataset while splitting it into test and train.
General DSA questions - 1-2 theory questions on some data structure. 1 coding question on sliding window.

**Round 3:**
This round was taken by Head of Data science in HiLabs, so my experience would seem to completely revolve on ML questions only. But for some other people I heard there were some questions on DS as well as DSA , and even some general GFG puzzles.
Although in my case it was completely ML but practicing DSA and Puzzles for this round is suggested.
My experience was I would say complete ML. He started to pick things from my CV and going deeper and deeper to what extent of knowledge do I have in those things. To understand my knowledge of any topic he gave me questions to test theory directly, then asked me to give some examples and how do they match the theory, then some practical implementation scenarios too and how can they affect my results. Basically everything. It's in his hands upto how much deep and which topic he wants to talk on.
General suggestion - This round was very tough. But I think I was able to clear it because I kind of tried to keep the interview in my lines and speaking words which I knew and could explain. Second, I tried to think and give answers of his questions, even if not sure of correctness I tried to give logically correct and some reasonable explanations which I felt were satisfactory after thinking on the ML basics.

# R Kailash

Data Scientist | Placement | 2023 | Hilabs

- Be well-prepared and confident about your CV projects and technical skills.
- The **Online Assessment** will include multiple-choice questions on Machine Learning (ML) and statistics, along with two basic Data Structures and Algorithms (DSA) questions: one on binary search and one additional easy DSA problem.
- The **interview process** consists of three rounds: the Coding Round with two easy to medium-level DSA questions, the Technical Round involving discussions on your Computer Vision (CV) projects and ML algorithms, and the HR + Technical Round combining HR questions with technical queries.
- Effective communication is crucial, so practice explaining your thoughts clearly.
- Engage in mock interviews with seniors or batchmates to simulate real interview conditions.
- Have a solid understanding of ML algorithms, focusing on their basic workings rather than deep technical details.
- If faced with a tricky question, avoid going blank and show your intent to solve the problem; asking for hints is acceptable.
- Demonstrate a strong desire to learn and grow during the interview.
- Take the HR interview seriously and prepare accordingly.
- Practice puzzles and quantitative problems on platforms such as GeeksforGeeks, Brainstellar, and PrepLeaf to enhance your problem-solving skills.