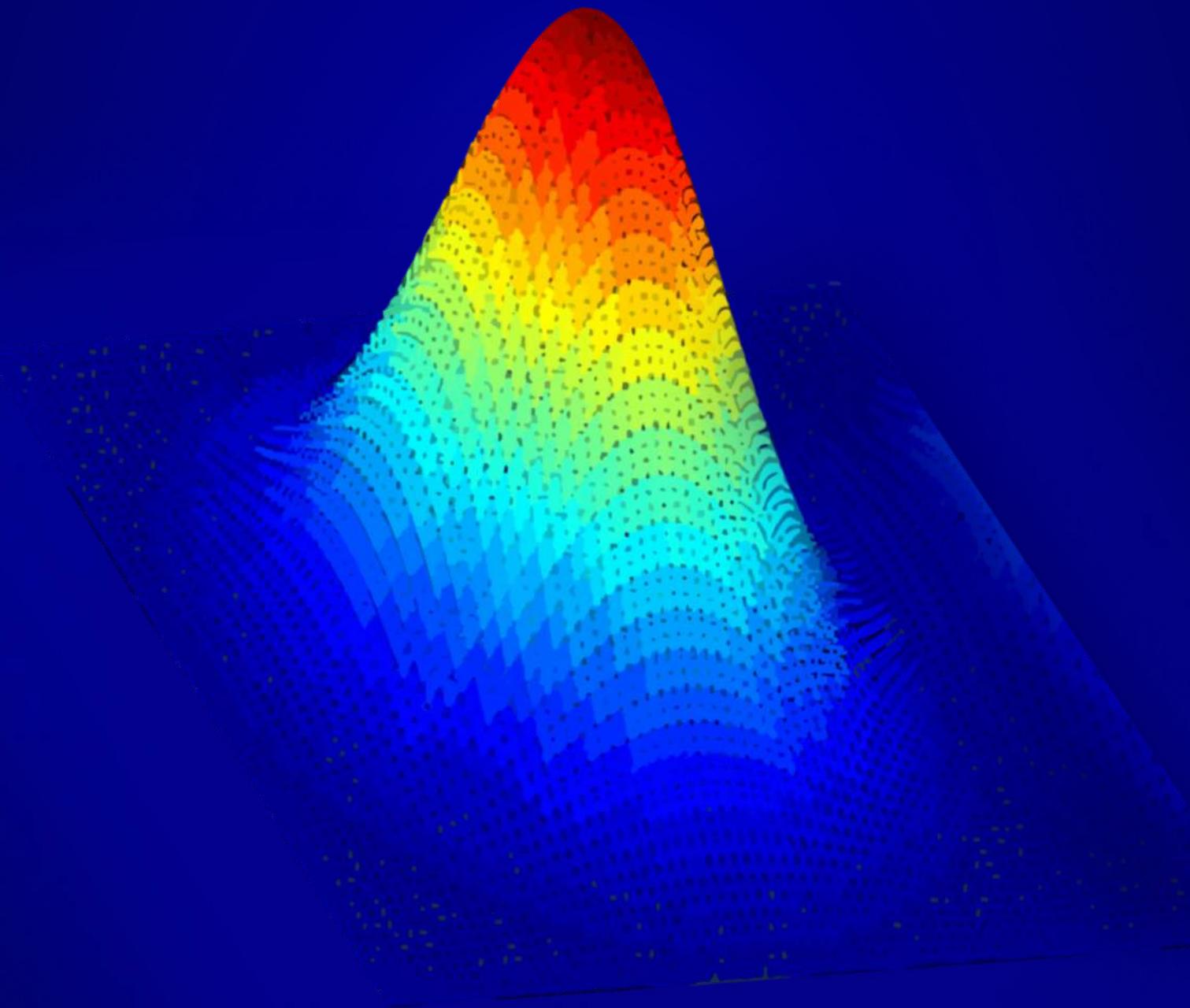




ANALYTICS INTERNSHIP GUIDE

FOR DATA SCIENCE/ANALYTICS ROLES



PREPARATION RESOURCES AND INTERVIEW EXPERIENCES



About Business Club IIT(BHU)

The Business Club, IIT(BHU) is the student-run club of Indian Institute of Technology (Banaras Hindu University), Varanasi where the confluence of finance, economics, consulting, analytics, and product management transpires. Having begun a few years with a team of a handful members, the club has since moved forward by leaps and bounds to a size of over 150+ core members and an enthused learning community of 400+ members.

As you read this, more than 20,000 GB worth of data is being generated every second by companies around the world! With the advent of a new age where Big Data and Machine Learning will revolutionize the business space, it is essential that students are equipped with skills and knowledge in these fields. Thus, we dedicate ourselves to compile a guide that can serve as a one stop preparation material for students targeting internships in data science roles.

**The Business Club
IIT (BHU) Varanasi**



Issues & Copyright

Analytics Internship Guide, The Business Club, IIT(BHU), Varanasi© 2023. All rights reserved.

Notice

No part of this publication may be reproduced or transmitted in any form or by any means – electronic or mechanical, including photocopy, recording, or any information storage and retrieval system – without permission in writing from The Business Club, IIT (BHU), Varanasi.

First edition: April 2023

**The Business Club,
IIT(BHU), Varanasi.**



INDEX

Internship Preparation	1
• SQL	2
• Probability and Statistics	3
• Linear Algebra	5
• Support Vector Machines	6
• Principal Component Analysis	7
• Convolution Neural Networks	8
• Ensemble Methods	9
• Case Studies/Guesstimates	10
Interview Experiences	12
• Microsoft	13
• American Express	24
• MasterCard	36
• CRED	54
• Goldman Sachs	61
• Publicis Sapient	70
• InfoEdge	81
• EXL	106
• Decimal Point Analytics	120
• FN MathLogic	126



INTERNSHIP PREPARATION

Analytics is a vast field with topics ranging from mathematics to case studies. Here we provide resources for some of the most important topics for internship tests and interviews.



SQL



[Intro to DBMS and SQL](#)



The thumbnail for a MySQL tutorial. It features the MySQL logo (a blue dolphin leaping) above the word "MySQL". Below the logo, it says "Tutorial 1- MySQL With Python MySQL Installation Steps". To the right is a circular profile picture of a man with glasses and a beard.

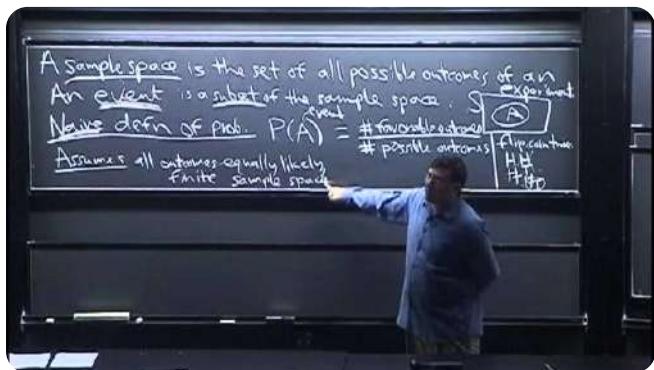
[MySQL Playlist](#)



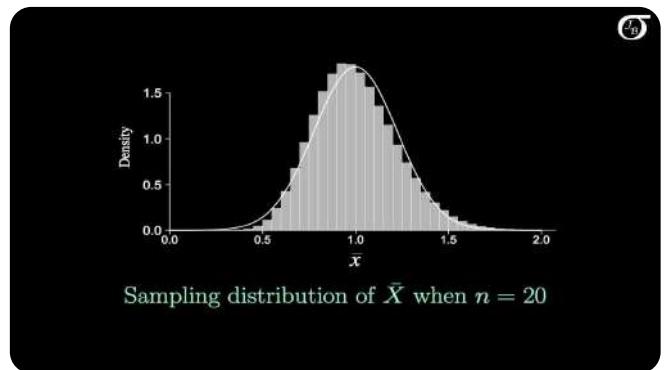
[SQL Joins](#)



Probability and Statistics



Statistics 110: Probability



Central Limit Theorem

**The Central Limit
Theorem...**



...Clearly Explained!!!

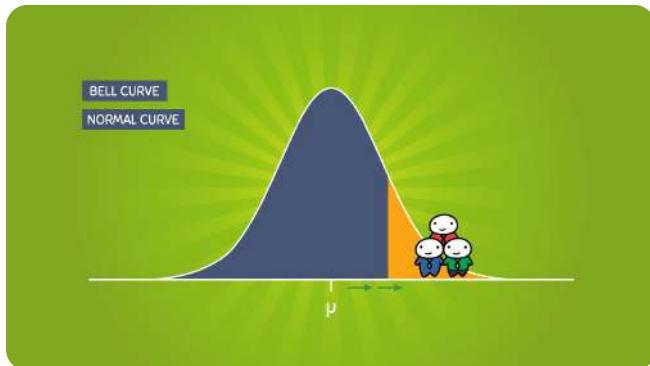
Central Limit Theorem



Law of Large Numbers



Probability and Statistics



68-95-99.7 Rule

Hypothesis Testing....



...and the Null Hypothesis!!!

Hypothesis Testing

p-values: What they are...



vs



...and how to interpret them!!!

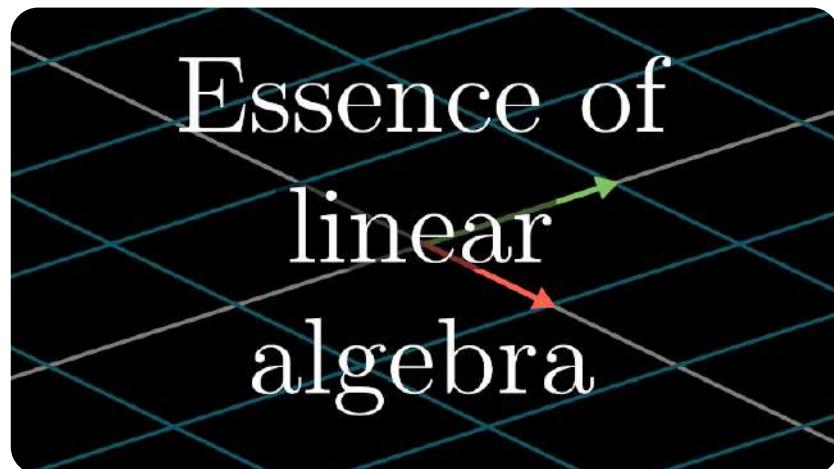
P-Values

Practice:

1. Working List of Probability Questions
2. Fifty challenging problems in probability



Linear Algebra

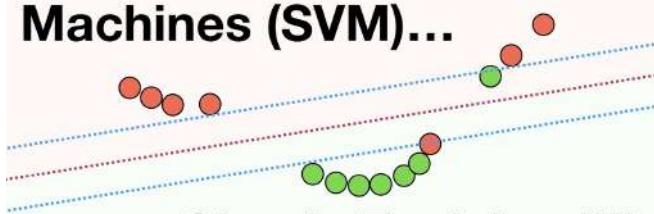


Essence of linear algebra



Support Vector Machines

**Support Vector
Machines (SVM)...**



**Support Vector
Machines Part 2:**

$$(a \times b + r)^d$$

**...The Polynomial
Kernel!!!**

Support Vector Machines

Support Vector Machines

**Support Vector
Machines Part 3:**

$$e^{-\gamma(a-b)^2}$$

...The Radial Kernel!!!

Support Vector Machines



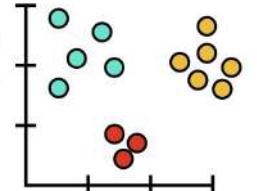
Principal Component Analysis



Dimensionality Reduction
Principal Component Analysis algorithm

Machine Learning

PCA Main Ideas...



..in only 5 min!!!

Dimensionality Reduction

PCA

```
# jupyter PCA notebook
In [1]: df = pd.read_csv('https://raw.githubusercontent.com/jbrownlee/Datasets/master/iris.csv')
In [1]: df.info()
Out[1]:

RangeIndex: 150 entries, 0 to 149
Data columns (total 4 columns):
 #   Column  Non-Null Count  Dtype  
 0   sepal.length  150 non-null   float64
 1   sepal.width   150 non-null   float64
 2   petal.length  150 non-null   float64
 3   petal.width   150 non-null   float64
dtypes: float64(4)
memory usage: 5.6 KB
In [1]: df.describe().T
Out[1]:
          count    mean    std. dev.   min   max
sepal.length      150  5.0137  3.7587  3.00  7.90
sepal.width       150  3.7587  1.7926  1.00  4.35
petal.length     150  1.7885  0.7524  0.10  4.45
petal.width      150  0.8000  0.3758  0.10  1.80
In [1]: df.corr()
Out[1]:
          sepal.length  sepal.width  petal.length  petal.width
sepal.length      1.0000    0.9515    0.9628    0.9565
sepal.width       0.9515    1.0000    0.9628    0.9565
petal.length     0.9628    0.9565    1.0000    0.9594
petal.width       0.9565    0.9594    0.9594    1.0000

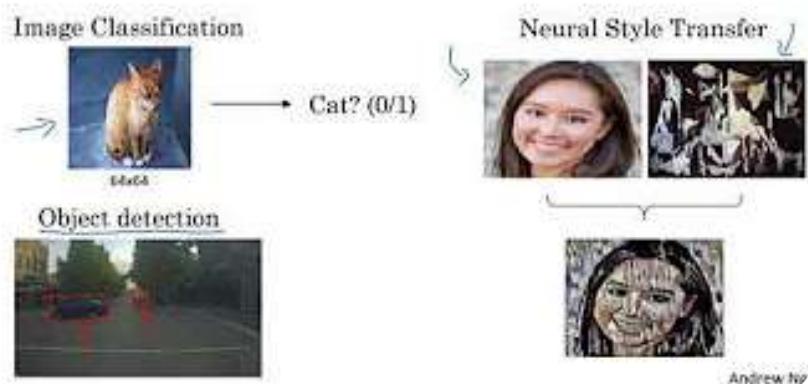
```

PCA



Convolution Neural Networks

Computer Vision Problems



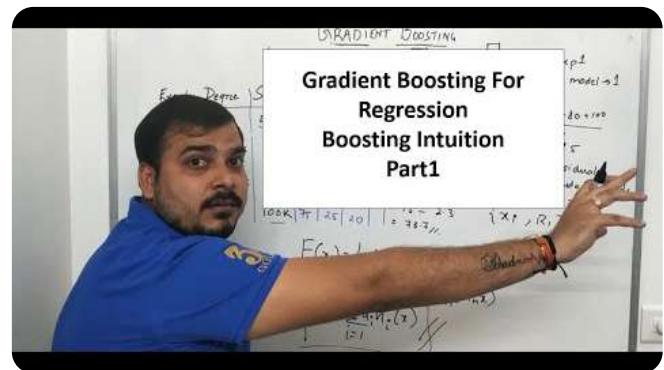
Convolutional Neural Networks



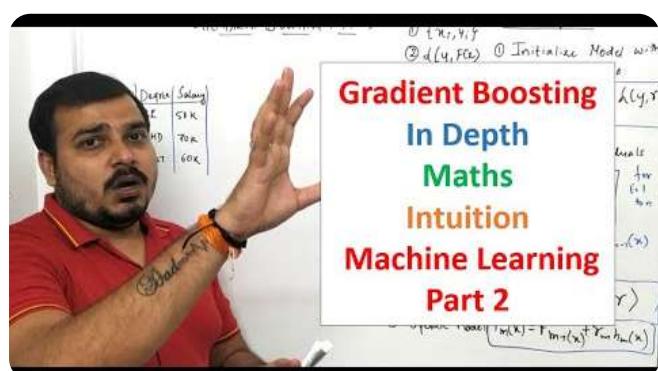
Ensemble Methods

Articles:

1. Random Forest in Python
2. XGBoost Algorithm
3. Using XGBoost
4. Ensemble methods
5. Sklearn



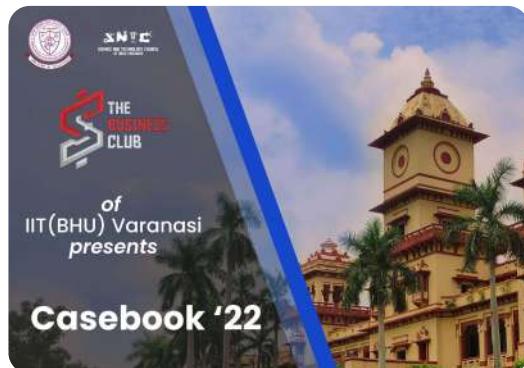
Gradient Boosting



Gradient Boosting



Case Study / Guesstimates



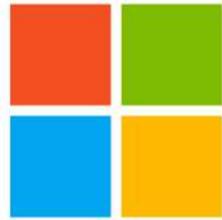
IIT (BHU) Casebook



This list of resources is by no means exhaustive, follow the DA/DS Roadmap by Business Club IIT (BHU) for a concise and comprehensive compilation of resources on Data Analytics and Science.



INTERVIEW EXPERIENCES



Microsoft

SCREENING ROUND

The screening round had 20 MCQ questions and 2 Coding questions. The coding questions were based on implementing Machine learning algorithms using Scikit-Learn, Numpy and Pandas.
Take the MCQ Round [here](#).



CODING QUESTIONS

Q1) Given a dataset, filter the dataset under certain conditions using pandas methods like mean, min, GroupBy etc. Another test case asked candidates to implement Linear Regression on the dataset and return the coefficients and intercept matrix

Q2) Given a dataset, perform K Means Clustering on it and return the dictionary with keys as clusters and values as samples in that cluster.

K mean clustering and its impleimplementation



Microsoft

INTERVIEW ROUND



STUDENT 1

ROUND 1

The interviewer started with a general introduction.

The interviewer initially asked a question, and then most of the interview revolved around follow up questions.

Q1) One Hour (60 minutes) of data related to CPU is provided to you, how will you predict the CPU usage in 61st minute?

To predict the CPU usage in the 61st minute based on the data provided, we can use time series forecasting methods. One commonly used method for time series forecasting is the Autoregressive Integrated Moving Average (ARIMA) model.

Links:

1. [What is Time Series Forcasting](#)
2. [CPU Utilization Based on ARIMA](#)

The interviewer asked if the candidate had any questions for them, and the candidate asked about their role in Microsoft if they get selected.



STUDENT 1

ROUND 2

This round was a hybrid of technical and HR rounds.

Q1) Given 2 Arrays, Find the median of both arrays.

We first combine both arrays into a single sorted array. We then find the length of the combined array n . Depending on whether n is even or odd, we either take the average of the middle two elements or return the middle element as the median of the combined array.

[Link for the article](#)

Q2) Follow up: How do you do it without using extra space?

We keep two indices i and j initially assigned to 0. Compare the i th index of 1st array and j th index of the second, increase the index of the smallest element and increase the count. Depending on whether n is even or odd, when the count reaches half the original size of the given array, we return the value of median.

[Link for the article](#)



STUDENT 1

ROUND 2

Q3) If you are given all the resources in the world, what would be 1 Machine Learning problem that you would like to work upon?

Although its subjective, below are some interesting and impactful machine learning problems that could have a significant positive impact on society if solved.

1. Disease diagnosis and prediction
2. Natural language processing (NLP)
3. Climate change modeling
4. Fraud detection
5. Autonomous vehicles
6. Personalized medicine

[Link for the article](#)

Q4) Why do you want to be at Microsoft?

Its subjective, here is a general response:
Opportunity to work on impactful projects, career growth, innovation, competitive benefits, and a culture that values diversity and inclusion.



Pro Tip

Fill out for SDE companies also, even if you are preparing for a DS/DA role.



STUDENT 2

ROUND 1

This round lasted around 50 minutes and was mostly focused on the candidate's resume. The questions were highly technical and specific. The candidate had done a previous internship in Data Science and was asked to explain their role and responsibilities thoroughly. The candidate had 3 projects around which most questions revolved:

1. IPL match prediction app
2. Solving differential equations using neural networks
3. Real vs Fake face detection using neural networks

Q1) What is Deep Learning?

[Link for the article.](#)



STUDENT 2

ROUND 1

Q2) What are CNNs and why are they used over ANNs?

Convolutional neural networks (CNNs) are a type of neural network that are specialized for processing grid-like data, such as images or audio. They use shared weight filters to extract features from input data, making them more effective for tasks such as image recognition than general artificial neural networks (ANNs).

[Link for the article](#)

Q3) What is the advantage of using LSTMs over RNNs?

Long Short-Term Memory networks (LSTMs) are a type of recurrent neural network (RNN) that are designed to address the problem of vanishing gradients, which can occur when training RNNs on long sequences of data. LSTMs use a more complex structure of cells with memory gates that can selectively store or discard information. This allows LSTMs to better handle long-term dependencies and make more accurate predictions on sequential data.

[Link for the article.](#)



STUDENT 2

ROUND 1

Q4) How do you decide whether to use LSTM or GRU?

The choice between using Long Short-Term Memory (LSTM) or Gated Recurrent Unit (GRU) networks depends on the length of sequences, amount of available data, and complexity of the task.

[Link for the article](#)

Q5) Longest increasing subsequence problem in DP

[Link for the article.](#)

[Practice here.](#)



STUDENT 2

ROUND 2

This round mostly had HR questions and a DSA question:

Q1) Reverse a Linked List

The idea is to use three pointers curr, prev, and next to keep track of nodes to update reverse links.

[Link for the article.](#)

Pro Tip

If you are guessing something in an interview, ask the interviewer first.



AMERICAN EXPRESS

SCREENING ROUND

There was no screening test conducted, shortlisting was done on the basis of Resume. Apparently, Projects and CPI were deciding factors.



AMERICAN EXPRESS

INTERVIEW ROUND

STUDENT 1

ROUND 1



The interview started with a simple introduction of the candidate. As the candidate had mentioned a case study in their resume, the interviewer asked a few questions on it. The interviewer then moved on from the resume and started asking questions on Ensemble Models.

Q1) What sampling methods did you use to sample data from the population?

There are several different sampling techniques available, and they can be subdivided into two groups: probability sampling and non-probability sampling.

Probability Sampling Methods.

Non Probability Sampling Methods.

STUDENT 1

ROUND 1



Q2) How can you be sure if a given sample gives you correct information about the population?

A well chosen sample will contain most of the information about a particular population parameter but the relation between the sample and the population must be such as to allow true inferences to be made about a population from that sample.

Population and Samples.

Q3) Do you know of any sampling errors or bias? Describe them.

STUDENT 1

ROUND 1



Two most common kind of sampling errors are:

Population Specific Sampling Errors - Such errors occur when the researcher/analyst does not understand who to survey.

Selection Error - When only survey interested participants respond to the questions, selection error may arise.

Sampling Errors

Q4) How is Bagging different from Boosting? When are each of them used?

Bagging vs Boosting

Bagging is used when individual model has a high variance and low bias while Boosting is used when it has a high bias and low variance.



STUDENT 1

ROUND 1

Q5) Describe how AdaBoost Works.

AdaBoost uses an iterative approach to learn from the mistakes of weak classifiers, and turn them into strong ones.

[Working of AdaBoost - 1](#)

[Working of AdaBoost - 2](#)

Q6) Suppose you are implementing AdaBoost using the Scikit-Learn library from Python. How do the hyperparameters of the model affect its performance?

Although such questions are rarely asked, you are expected to know few hyperparameters of AdaBoost such as n_estimators and learning_rate.

[Hyperparameters of AdaBoost](#)



STUDENT 1

ROUND 2

This round was focused on Deep Learning. Since the candidate had a project on Natural Language Processing, most of the questions revolved around it.

Q1) Given that input to a convolution 2d layers is (W,H,D) and output is (W',H',D'). Calculate the shape of the weight matrix of this layer assuming padding="none" and strides=1.

$$f_W = W + 1 - W'$$

$$f_H = H + 1 - H'$$

$$f_D = D'$$

[Understanding Input Output Shapes of Convolution layer](#)



STUDENT 1

ROUND 2

Q2) What are some drawbacks of ANNs that CNNs solve?

CNNs take images as a 2D input which preserves the spatial relationship of objects in the image while using significantly fewer parameters than ANNs.

Similarities and Differences of ANN and CNN.

Q3) How do CNNs compare with RNNs in NLP problems in your experience?

For most NLP tasks, RNNs would give better results than CNNs because RNNs process the text data sequentially while CNNs do not. Also, RNNs have significantly more parameters than CNNs and hence require more data, training time and latency.

NLP with CNNs



STUDENT 1

ROUND 2

Q4) In what case do we choose CNNs over RNNs in NLP problems?

CNNs are used to process textual data that has long sentences and many newlines.

Example: Text Summarization of Novels, News Articles etc.

Q5) Why have you used the ReLU activation function?

The rectified linear activation function overcomes the vanishing gradient problem, allowing models to learn faster and perform better. ReLU activation help the network fit to complex non linear mappings.

Usage of ReLU in neural Networks.



STUDENT 1

ROUND 2

Q6) Follow up: Does ReLU activation have no problem associated with it?

ReLU cannot learn on examples with activation zero. Moreover when large gradient flows through, a ReLU neuron will update its weight and might be ended up with a negative weight and bias which lead to production of 0 during forward propagation.

Dead Neuron.

Dying ReLU problem.

Q7) How many different weight initialization methods do you know? How are they different from each other?

Two most common types of weight initialization methods are Zero initialization and Xavier initialization.

Weight Initialization for Deep Learning Neural Networks



STUDENT 1

ROUND 2

Q8) How does Dropout Layer work? How is it handled during inference?

Dropout drops each weight of a layer with some probability equal to dropout rate. During inference, dropout layer is removed and weights are scaled by the reciprocal of dropout rate. [Dropout in Neural Networks.](#)

Q9) How do you draw a confusion matrix for multiclass classification?

Confusion matrix for n classes will be an $n \times n$ matrix.
[Confusion Matrix Guide by Analytics Vidhya](#)

STUDENT 1

ROUND 3



This was an HR Round, It was just one question.

Q1) How do you plan to contribute to American Express using your skills?

Since American Express is a bank, your answer should preferably be based on how data structures and data science skills can be utilized in banking. [Link](#)

Pro Tip

Build your Resume Properly, you should be able to answer any question based on your resume



SCREENING ROUND

The screening test had 25 MCQs and 3 SQL questions. The MCQs were divided into 5 categories:

1. Mathematics : Included simple math questions like Integration and Bayes Theorem.
2. Classical Machine Learning: Theoretical questions on classical algorithms.
3. Deep Learning: Common questions from coursera quizzes on deep learning.
4. Logical Reasoning: Basic aptitude questions
5. Python Coding MCQs

The SQL questions were pretty basic based and could be easily solved with a little practice from HackerRank.



INTERVIEW ROUND

STUDENT 1

ROUND 1



The interviewer had a stern tone, all the questions were straightforward and a short crisp answer was expected, no beating around the bush. The interviewer started by asking questions on candidate's projects:

Q1) What is the loss function used in the Logistic Regression Algorithm?

The loss function for logistic regression is **Log Loss**, which is defined as follows:

$$\text{Log Loss} = \sum_{(x,y) \in D} -y \log(y') - (1 - y) \log(1 - y')$$

STUDENT 1

ROUND 1



Q2) Explain in detail how K - Nearest Neighbours works.

The k-nearest neighbors algorithm, also known as KNN or k-NN, is a non-parametric, unsupervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point.

Detailed working of KNN.

Q3) How do optimizers work? How many Optimization algorithms do you know

Optimizers are algorithms or methods used to change the attributes of your neural network such as weights and learning rate in order to reduce the losses.

Some of the commonly used optimizers are gradient descent, stochastic gradient descent, mini-batch gradient descent, etc.

Detailed information on optimizers and its types.



STUDENT 1

ROUND 1

Q4) What is the role of an activation function in a neural network? What will happen if you do not use any?

In broad terms, activation functions are necessary to prevent linearity. Without them, the data would pass through the nodes and layers of the network only going through linear functions ($a*x+b$).

Role of activation function.

The interviewer then moved on to basic probability and statistics questions:

Q5) Some finite number of elements are given in a gaussian distribution, how will you find the true mean of the gaussian distribution?

To find the true mean of a Gaussian distribution, you can use the following steps:

STUDENT 1

ROUND 1



1. Calculate the sample mean of the given finite number of elements.
2. Calculate the sample variance of the given finite number of elements.
3. Use the formula for the confidence interval of the mean to calculate the range within which the true mean of the population is likely to fall. The formula is:
Confidence interval = sample mean \pm (t-value \times standard error) where t-value is the value from the t-distribution table for the desired confidence level and degrees of freedom ($n-1$), and the standard error is the square root of the sample variance divided by the sample size.
4. Once you have the confidence interval, you can estimate the true mean of the population by taking the midpoint of the interval.

Q7) What is a probability distribution function? What does its integral represent?

A probability distribution function (PDF), also called a probability density function, is a mathematical function that describes the relative likelihood of different outcomes in a random variable.

The integral of the PDF represents the cumulative probability up to a certain value.



STUDENT 1

ROUND 1

Q8) What is the probability distribution function for a gaussian random variable?

$$P(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(z-\mu)^2 / 2\sigma^2}$$

Q9) Tell one real life example where gaussian distribution is used.

One real-life example of where Gaussian distribution is commonly used is in the field of finance to model the distribution of stock prices. One real-life example of where Gaussian distribution is commonly used is in the field of finance to model the distribution of stock prices.



STUDENT 1

ROUND 1

Q10) Tell one real life example where the Central Limit Theorem is used.

One real-life example where the Central Limit Theorem (CLT) is used is in quality control. Suppose a factory produces a large number of items that are supposed to have a specific weight, and the goal is to ensure that the weights of the items are within a certain range. The factory may take a random sample of items and weigh them, and then use the CLT to determine whether the sample mean weight is within the acceptable range.



STUDENT 1

ROUND 2

Q1) Why are CNNs used in Computer Vision Tasks? Why not ANNs?

The Convolutional Neural Network (CNN or ConvNet) is a subtype of Neural Networks that is mainly used for applications in image and speech recognition. Its built-in convolutional layer reduces the high dimensionality of images without losing its information. That is why CNNs are especially suited for this use case.

Detailed Explaination

Since the candidate had a project on GANs, the interviewer was curious about how much the candidate knew about generative models.

Q2) Explain the architecture of GAN.

Generative Adversarial Networks (GANs), represent a shift in architecture design for deep neural networks. There are several advantages to using this architecture.

Detailed Explanation.



STUDENT 1

ROUND 2

Q3) Follow up: What is the loss function used? How is a single loss function used for two models?

The loss function used is **min-max loss**.

[Detailed explanation about the GAN loss function.](#)

Q4) When and Why is the Discriminator kept non-trainable.

The discriminator in a generative adversarial network (GAN) is typically kept non-trainable during the generator training process. The reason for this is that the discriminator has already been trained separately to recognize real data from fake data, and its parameters have been optimized for this purpose. If the discriminator were to be updated during generator training, it would quickly learn to recognize the new patterns generated by the updated generator, making the generator's job much harder.



STUDENT 1

Pro Tip

Concepts like CLT, LLN and Distributions are often overlooked but are very important.



STUDENT 2

ROUND 1

The interviewer went straight to the questions.

Q1) Explain how Boosting works.

Boosting creates an ensemble model by combining several weak decision trees sequentially. It assigns weights to the output of individual trees. Then it gives incorrect classifications from the first decision tree a higher weight and input to the next tree.

Detailed Working.

Q2) Why do we use the Mean Squared Error loss function for Regression?

The Mean Squared Error (MSE) loss function is commonly used for regression problems because it measures the average squared difference between the predicted and actual values of the target variable.



STUDENT 2

ROUND 1

Since the candidate had a project on Computer Vision + GANs, the interviewer asked a few questions on it.

Q3) What difficulties did you face while training GANs?

The difficulties while training GANs.

Q4) Why do you think are GANs so difficult to train?

GANs can be difficult to train due to various reasons like vanishing gradients, mode collapse, training instability, and hyperparameter tuning.



STUDENT 2

ROUND 1

The candidate had 2 more projects based on Robotics +Reinforcement Learning so some questions were based on them

Q5) How did you make the Environment to train your model?



STUDENT 2

ROUND 1

Q6) How do reward functions work in Reinforcement Learning?

In reinforcement learning, the agent is rewarded for taking controls that lead to successful states. The rewards can be immediate, such as receiving a point for each step taken in the right direction, or they can be delayed, such as receiving a point at the end of the episode if the goal was reached.

Detailed working.

STUDENT 2

ROUND 2



This round had a more informal tone than the previous one, the questions were mostly based on mathematics. The round only lasted 10 minutes.

Q1) What is the Curse of Dimensionality?

The Curse of Dimensionality refers to the phenomenon where the performance of many machine learning algorithms deteriorates as the number of input features or dimensions increases. The term "curse" is used because increasing the number of dimensions can lead to a rapid increase in the amount of data required to generalize accurately.

Q2) What is PCA? Explain the algorithm in its totality.

Principal Component Analysis.



STUDENT 2

ROUND 2

Q3) Explain the Bias Variance Tradeoff. How do you measure Bias and Variance, how are they related?

Explanation and relation.

Q4) What is L1 and L2 regularization? What are the mathematical expressions for them?

L1 and L2 Regularization and their expressions.

Q5) Why are Gaussian Distributions used to model errors? Why not other distributions?

Gaussian (or normal) distributions are commonly used to model errors because they have several desirable properties that make them a good choice for this purpose. Some of these properties are ease of use, symmetry, and central limit theorem.

STUDENT 2

ROUND 2



Pro Tip

Mathematics behind ML Algorithms is very important, memorize them for the interview if needed.



CRED

SCREENING ROUND

The Screening test had 2 separate rounds timed individually. There was an SQL round and an MCQ Round. SQL round had 7 SQL questions based on basic concepts like tables, joins etc. MCQ round had 30 questions based on Financial Mathematics, Probability and Statistics etc. Most candidates were able to solve all the SQL questions with a little practice from HackerRank.



INTERVIEW ROUND



STUDENT 1

ROUND 1

This was a case study round conducted in a form of Group Discussion. Candidates were divided into 2 groups and both groups were given a different case to solve.

Q1) Suppose you have a superpower that lets you know that all the people aged above 60 years will suddenly die in a single moment exactly after 1 month from now. Devise a strategy that earns you the most money in the world in this time period. It can be unethical/illegal.

- Start an institution that is an army of young diplomats, spokespeople and politicians.
- Since most of the powerful diplomats and politicians are aged above 60, all of them will die and the country will suddenly need new position holders.
- This will give our institution the power over world economies and politics.

(The interviewers were not looking for a financially sound but an innovative answer to test the creativity of candidates)



STUDENT 1

ROUND 1

Q2) Suppose you are an analyst in a company that sells coconut oil. You found out that the sales for your product have suddenly shot up in a particular retail store despite the fact that there has been no change in the product, operations or marketing of the company. Write down all possible reasons for this along with all your assumptions. No clarifying questions allowed.

Detailed Explanation of the strategy used.

STUDENT 1

ROUND 2



This was another case round. It was common for all shortlisted candidates.

Q1) Suppose it is 2011, You want to expand your online cab ride service (like Ola and Uber) to a tier-1 city with a population of 5 to 15 Lacs. You already provide services in a couple of other tier -2 and tier - 3 cities. Guesstimate the amount of people who would avail this service in 4 hours assuming these are peak hours.

- Let's assume that 20% of the population in the city has access to these facilities.
- For a city with a population of 5 to 15 Lacs, the number of potential users would be between 1 to 3 Lacs.
- Now, let's assume that during peak hours, which we can assume to be between 7 am to 10 am and 5 pm to 8 pm, people are likely to travel the most. We can further assume that on average, each user would take a cab ride for 30 minutes during peak hours.
- Therefore, the total number of cab rides during peak hours would be: Total number of users x number of rides per user during peak hours
= 1 Lakh to 3 Lakh x 1 ride per user during peak hours
= 1 Lakh to 3 Lakh rides during peak hours.
- Therefore, the service may have around 1 Lakh to 3 Lakh rides in a city with a population of 5 to 15 Lacs, assuming 20% of the population has access to smartphones and internet connectivity.



STUDENT 1

ROUND 2

Q2) Follow Up: What will be your pricing strategy if we want our revenue to be at least Rs.20,000 per hour each day?

Quite a few pricing strategies can be used for this like flat pricing, surge pricing, distance-based pricing, and time-based pricing.

(Catch is to incur loss in non-peak hours and surpass the loss and generate extra profit in peak hours)

STUDENT 1

ROUND 3



This was an HR round. The interviewer mostly focused on the candidate's engagement in student societies, festivals and how candidates were enjoying their college life. It is noteworthy that formal clothes were not allowed and cultural fit was a major factor for the interviewer.

Pro Tip

CRED looks out for cool people who have a good sense of humor and can think outside the box. They have pretty unconventional hiring methods so expect the unexpected.



Goldman Sachs

SCREENING ROUND

The test had 4 sections:

1. Coding Section (2 Questions, 30 minutes)
2. Essay Section (2 questions, 15 minutes): It consisted of 2 questions which you have to give a brief description of one of your projects where you learnt a lot and other asked to give brief description of a situation where you demonstrated leadership skills.
3. Mathematics section (12 questions, 45 minutes)
4. CS Fundamentals (7 questions, 30 minutes)

[Take the MCQ quiz here](#)



CODING QUESTIONS

Q1) You have been given coordinates of vertices of N Nachos (triangles), and M Slicers (straight lines parallel to either x-axis or y-axis). Print the number of Nachos cut by each slicer.

It is a hard difficulty problem. A similar type of question to above asked question: Count number of triangles cut by the given horizontal and vertical line segments

Q2) A milkman has to deliver milk to N locations. Milkman will start from location 1 and end up at location N. Print all possible paths a milkman can take.

It is a medium difficulty problem. A very similar question to above asked question: All Paths from source to target.



Goldman Sachs

INTERVIEW ROUND

STUDENT 1

ROUND 1



The interview had an informal environment, it started with a simple introduction including hobbies and interests. The candidate had a project on “Diamond Valuation prediction using physical features”, so the interviewer went on to ask a few questions on it:

Q1) What libraries did you use in this project?

Q2) What metrics or parameters did you use to evaluate the performance of your model?

The interviewer was then interested in the Position of Responsibility held by the candidate which was Event Coordinator.

Q3) What problems did you face while managing a team of 20 people?

STUDENT 1

ROUND 1



The interviewer then moved on to puzzles:

Q4) Monty Hall Problem

Monty Hall Puzzle

Q5) How do we measure forty-five minutes using two identical wires, each of which takes an hour to burn? We have matchsticks with us. The wires burn non-uniformly.

Puzzle



STUDENT 1

ROUND 2

The panel had 2 interviewers in it, they started with a simple introduction. The candidate had previously completed an internship project on “Modern Slavery: analyzing workplace toxicity using sentiment analysis on scraped tweets”. The interviewer asked to explain the entire project from one end to another.

Q1) What is the future scope of this project and how is it scalable and deployable?

People looking for Jobs can use this to rank companies on the basis of workplace toxicity.



STUDENT 1

ROUND 2

The interviewer then asked the candidate their hobbies apart from Machine Learning and Data Science, to which they replied Football. The candidate also had an end to end project on “Clustering Football players based on their playing style”.

Q2) Explain how you performed clustering on football players and how is this useful in real life?

Suppose you are making a FPL team, You want to choose Lionel Messi in your team and he costs 10 units which you do not have, so you chose some other player from the same cluster as Messi with less cost.

The interviewer then moved on to more informal questions.



STUDENT 1

ROUND 2

Q3) Suppose a player is ready to be sold in the market, how would you do the valuation of the player?

Demand of the position that player plays in, Previous salary of the player, Financial Condition of our club, Why current club is selling the player, Rival Clubs do not buy each other's players.

Pro Tip

- 1. Ask your friends in other institutions to provide you with questions if some company has already conducted the test there.**
- 2. Learn Process Scheduling, one question is always asked in the screening test on it.**



Goldman Sachs

SCREENING ROUND

The test had 4 sections:

1. Coding Section (2 Questions, 30 minutes)
2. Essay Section (2 questions, 15 minutes): It consisted of 2 questions which you have to give a brief description of one of your projects where you learnt a lot and other asked to give brief description of a situation where you demonstrated leadership skills.
3. Mathematics section (12 questions, 45 minutes)
4. CS Fundamentals (7 questions, 30 minutes)



publicis
sapient

SCREENING ROUND

The test had 2 sections with a common timer of 1 hour.

1. Coding section comprising of 2 questions.
2. MCQ Section which was based on Machine Learning theory, Probability and Statistics, and basic aptitude.



CODING QUESTIONS

Q1) Remove One Element to Make the Array Strictly Increasing

O(n) C++ solution.

Q2) Robot Return to Origin

O(n) Solution.



publicis
sapient

INTERVIEW ROUND



STUDENT 1

ROUND 1

Round 1 (Student 1): The interview started with a brief introduction of the candidate. The interviewer enquired about the candidate's hobbies and personal interests. The interviewer was impressed with the candidate's project on NLP (sentiment analysis) and started by asking questions on it.

.

Q1) What is a tokenizer?

A tokenizer breaks a stream of text into tokens, usually by looking for whitespace (tabs, spaces, new lines).

[Article on tokenization.](#)

Q2) Since you have used LSTMs, Tell me about the architecture of LSTMs.



STUDENT 1

ROUND 1

LSTMs deal with both Long Term Memory (LTM) and Short Term Memory (STM) and for making the calculations simple and effective it uses the concept of gates.

[Article on architecture of LSTMs.](#)

Q3) What is a word embedding?

Word Embedding or Word Vector is a numeric vector input that represents a word in a lower-dimensional space. It allows words with similar meaning to have a similar representation. They can also approximate meaning.

[Article on Word Embedding.](#)

Q4) Which metric did you use to evaluate your sentiment analysis model ?

As a classification problem, Sentiment Analysis uses the evaluation metrics of Precision, Recall, F-score, and Accuracy. Also, average measures like macro, micro, and weighted F1-scores are useful for multi-class problems.

[Article on evaluation of Sentiment Analysis.](#)



STUDENT 1

ROUND 1

The candidate had another deep learning based project on Semantic Segmentation of Roads using UNet architecture.

Q5) Explain the UNet architecture along with Upsampling and Downsampling in UNet.

U-Net is an architecture for semantic segmentation. It requires small number of training examples unlike the traditional convolutional networks that used thousands of annotated training samples.

Article on UNet architecture.

Q6) Why do we use CNN on image data?

Article on image processing using CNNs.

Image Classifier using CNNs.



STUDENT 1

ROUND 1

The candidate had completed an internship in the past winter by making a project on Time Series Forecasting. The interviewer asked a few questions about it.

Q7) What is the ARIMA model?

ARIMA is an acronym that stands for AutoRegressive Integrated Moving Average. It is a class of model that captures a suite of different standard temporal structures in time series data.

[Article on creating ARIMA model.](#)



STUDENT 1

ROUND 2

The interviewer gave a few questions on SQL based on basic concepts like joins, average, minimum etc. The candidate was able to answer a few questions but then confessed that they were not prepared for SQL coding questions. The interviewer was okay with it.



STUDENT 1

ROUND 3

This was a very short round, only a few puzzles were asked.

Q1) Rat and Poisonous Milk Bottles

GFG answer of this puzzle.

Q2) Burning candles

GFG answer of this puzzle



STUDENT 1

ROUND 4

This was a Coding Round + HR Round

Q1) Median of two sorted arrays

C++ Solution.

Q2) Group Anagrams

C++ Solution

The interviewer asked if the candidate had any questions for them. The candidate asked “Since Publicis Sapient is a consulting company and most of the other companies were at a loss because CoVID, how did Publicis Sapient consult them considering this was a one of a kind situation?”



STUDENT 1

Pro Tip

Do thorough research on the company, prepare the questions that you will ask them beforehand. Also avoid adding generic PoRs like Event Executives or Co-Coordinators rather add PoRs that show your skills like conducting workshops or preparing content for juniors



inf^{edge}

SCREENING ROUND

The screening test consisted of 40 MCQs from Probability, Statistics, Linear Algebra and Machine Learning Theory.



MCQ ROUND

Q1) If the Kurtosis of a given Data is > 3 , is it leptokurtic, mesokurtic or platykurtic?

Leptokurtic

Kurtosis

Q2) Which Command is used in SQL to select unique rows satisfying a certain condition?

To select unique rows satisfying a certain condition in SQL, you can use the **SELECT DISTINCT** statement followed by the column names that you want to select.

eg:-
SELECT DISTINCT id, name, age, major
FROM students
WHERE age > 20;



Q3) For which probability distribution are the mean and variance equal to each other?

Poisson distribution

Q4) A Rod is randomly cut at a point, what is the variance of the shorter part?

Let L be the length of the rod, and let X be the length of the shorter part after the random cut. Then, the length of the longer part is $L - X$.

The mean of X is given by: $E(X) = (0 + L) / 2 = L / 2$

The variance of X is given by: $\text{Var}(X) = E(X^2) - [E(X)]^2$

$$\begin{aligned} E(X^2) &= \text{integral from } 0 \text{ to } L \text{ of } x^2 * (1/L) dx = 1/L * \\ &\quad \text{integral from } 0 \text{ to } L \text{ of } x^2 dx = 1/L * [x^3/3] \text{ from } 0 \text{ to } L = \\ &\quad L^2 / 3 \end{aligned}$$

Therefore:

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = L^2 / 3 - (L/2)^2 = L^2 / 12$$

Thus, the variance of the shorter part is $L^2 / 12$.



Q5) A call center gets an average of 5 calls per minute. Find the probability that the call center will receive 20 calls in 5 minutes.

This scenario can be modeled using a Poisson distribution.
 $\lambda = 5$ calls per minute.

The probability of receiving exactly 20 calls in 5 minutes can be calculated using the Poisson probability mass function:

$$P(X = k) = (\lambda^k / k!) * e^{-\lambda}$$

Plugging in these values, we get:

$$P(X = 20) = (5^{20} / 20!) * e^{-5} = (9.53674 \times 10^{-7}) * 0.00673795 = 6.43077 \times 10^{-9}$$

Q6) What is the output size of a convolution layer if we use then 5x5 filters with stride 1 and pad 2 on an input with volume 32x32x3?

$$W_{out} = (W_{in} - F_w + 2P)/S + 1$$

$$H_{out} = (H_{in} - F_h + 2P)/S + 1$$

$$D_{out} = F$$

Using the formula above, we can compute the output volume as:

$$W_{out} = (32 - 5 + 2(2))/1 + 1 = 32, H_{out} = (32 - 5 + 2(2))/1 + 1 = 32, D_{out} = 5$$

Therefore, the output size of the convolution layer is 32x32x5.



Q7) On using which kernel for SVM does the dimensions of your projected data become infinite?

The kernel that can cause the dimensions of the projected data to become infinite is the "RBF" (Radial Basis Function) kernel or Gaussian kernel.

Q8) What is the time complexity of searching for an element in a set in Python?

The time complexity of searching for an element in a set in Python is $O(1)$ on average case.



inf^{edge}

INTERVIEW ROUND



STUDENT 1

ROUND 1

This round had straightforward questions and tested the candidate's understanding of Classical ML concepts.

Q1) How much do you know about Random Forests? Explain all of it.

Detailed Explanation

Q2) Explain how Support Vector Machines Work and how are they implemented?

SVM is defined such that it is defined in terms of the support vectors only, we don't have to worry about other observations since the margin is made using the points which are closest to the hyperplane (support vectors).

Detailed working and implementation.

STUDENT 1

ROUND 1



Q3) What is Gradient Descent Algorithm?

Gradient descent is an optimization algorithm which is commonly-used to train machine learning models and neural networks.

Detailed Explanation.

Q4) Explain how Hypothesis Testing works.

- Hypothesis testing is used to assess the plausibility of a hypothesis by using sample data.
- The test provides evidence concerning the plausibility of the hypothesis, given the data.
- Statistical analysts test a hypothesis by measuring and examining a random sample of the population being analyzed.

Detailed Working.

STUDENT 1

ROUND 2



The interviewer dived deep into the concepts discussed in the previous round.

Q1) What are the assumptions of Linear Regression and why are they assumed?

Detailed Explanation.

Q2) What is a Kernel in Support Vector Machines and how do kernels work?

Kernels are used in Support Vector Machines (SVMs) to solve regression and classification problems. Support Vector Machines use the Kernel Trick to transform linearly inseparable data into linearly separable data, thus finding an optimal boundary for possible outputs.

Detailed working.



STUDENT 1

ROUND 2

Q3) The interviewer asked the candidate to open Google Colab and asked to implement Filtering on a dataset under various conditions. They also asked how GroupBy works in pandas.

Filtering the dataset.

Working of GroupBy in Pandas.

Q4) A function was given, which can generate only 1s and 0s, 1 was coming out with 70% probability and 0 with 30% probability. Using this function you have to write our own function, which could generate numbers from 0 to 5 but with uniform probability.(Only given function can be used)

Answer

STUDENT 1

ROUND 3



This round was mostly focused on mathematics around classical machine learning algorithms.

Q1) What is Gini Impurity? And what is Entropy? The candidate was asked to write down the expressions for them.

Gini Index or Impurity measures the probability for a random instance being misclassified when chosen randomly. The lower the Gini Index, the better the lower the likelihood of misclassification.

Entropy is a measure of disorder or impurity in a node.

$$Gini = 1 - \sum_{i=1}^j P(i)^2 \quad E = - \sum_{i=1}^n p_i \log_2(p_i)$$



STUDENT 1

ROUND 3

Q2) What is Rank of a Matrix and how is it calculated?

The rank of matrix can be determined by reducing the given matrix in row-reduced echelon form, the number of non-zero rows of the echelon form is equal to the rank of matrix. Rank of matrix can also be calculated by finding order of the highest non-singular minor of the given matrix.

Q3) What is the expected number of coin tosses, if a coin is tossed until three heads are obtained consecutively?

The expected number of coin tosses until three consecutive heads are obtained can be found using a geometric distribution with parameter $p = 1/2$, which models the probability of obtaining a head on a single toss of a fair coin.

The answer is 8.



Round 4 was HR round.

Pro Tip

Keep your basics of ML clear, do not directly jump to deep learning if you are not comfortable with classical machine learning.



STUDENT 2

ROUND 1

The interview started with a general introduction.

The interviewer asked the candidate “On a scale of 1 to 10, how good are you with Probability and Statistics?”

The candidate answered 7.

Q1) Tell me one real application of Bayes Theorem?

The candidate answered that Bayes Theorem can be applied in Medical Diagnosis of Patients to calculate the probability of a disease given the results of medical test.

The candidate then said “Now make that rating 9 actually”



STUDENT 2

ROUND 1

Q2) Three people: You, Your Dad and Roger Federer play Tennis singles against each other. Probability that you win against your dad is more than the probability of you winning against Roger Federer. You have to play 3 matches where you can choose your opponent. Your goal is to win at least 2 consecutive matches out of 3 matches.

In which order should you choose your opponents given you cannot choose the same opponent twice in a row?

The Answer is Roger -> Dad -> Roger. The intuition here is that you must always prioritize winning the second match to ensure that you win 2 matches consecutively.

The candidate remembered this question from Statistics 110 Harvard Course and spit out the answer immediately and forgot the explanation behind the answer.



STUDENT 2

ROUND 1

Q3) What is KL divergence? The candidate was asked to write down the expression on a paper and show it.

In mathematical statistics, the Kullback–Leibler divergence (also called relative entropy and I-divergence), denoted , is a type of statistical distance: a measure of how one probability distribution P is different from a second, reference probability distribution Q.

KL Divergence

KL Divergence expression

Theoretical explanation of KL Divergence



STUDENT 2

ROUND 1

Q4) What are your views on Accuracy as a metric? Why is F1 score used instead of accuracy sometimes?

Accuracy is used when the True Positives and True negatives are more important while F1-score is used when the False Negatives and False Positives are crucial

Drawbacks of Accuracy as metric

Other metrics for classification model

Accuracy vs F1 Score

Q5) There are two regions, one region has some number of pregnant women and the other has some other number. You need to compare the fertility rate of two regions, what will be your null hypothesis in this case? How do you reject or fail to reject the hypothesis?

Using P - Value. We reject the null hypothesis if p - value is less than 5 %.



STUDENT 2

ROUND 1

**The interviewer asked why specifically 5%?
(Follow up question of question number 5)**

It is just a standard practice that we can choose other values based on the purpose of our investigation.



STUDENT 2

ROUND 2

Q1) Suppose that you are worried that you might have a rare disease. You decide to get tested, and suppose that the testing methods for this disease are correct 99 percent of the time. Suppose this disease is actually quite rare, occurring randomly in the general population in only one of every 10,000 people. If your test results come back positive, what are your chances that you actually have the disease?

To calculate the probability of actually having the disease given a positive test result, we need to use Bayes' theorem, which states:

$$P(A|B) = P(B|A) * P(A) / P(B)$$

In this case, let's define:

- Event A: You have the disease
- Event B: Your test result is positive

$P(A)$ is the prior probability of having the disease, which we know is 1 in 10,000, or 0.0001.

$P(B|A)$ is the probability of testing positive given that you have the disease, which is given as 99%, or 0.99.



STUDENT 2

ROUND 2

$P(B)$ is the overall probability of testing positive, which can be calculated using the law of total probability as:

$$P(B) = P(B|A) * P(A) + P(B|\text{not } A) * P(\text{not } A)$$

where $P(\text{not } A)$ is the complement of $P(A)$, which is $1 - P(A)$, or 0.9999.

$P(B|\text{not } A)$ is the probability of testing positive given that you do not have the disease, which is the false positive rate. Since the testing methods are correct 99% of the time, the false positive rate is 1% or 0.01.

So, we can calculate $P(B)$ as:

$$P(B) = 0.99 * 0.0001 + 0.01 * 0.9999 = 0.010098$$

Now, we can use Bayes' theorem to calculate the probability of actually having the disease given a positive test result:

$$\begin{aligned} P(A|B) &= P(B|A) * P(A) / P(B) \\ &= 0.99 * 0.0001 / 0.010098 \\ &= 0.009803 \end{aligned}$$

Q2) How do you handle unimportant columns in your dataset?

The candidate answered Principal Component Analysis using Singular Value Decomposition.

SVD is basically a matrix factorization technique, which decomposes any matrix into 3 generic and familiar matrices

Singular Value Decomposition (SVD)



STUDENT 2

ROUND 2

Q3) What are Orthonormal and Orthogonal Matrices?

The matrix is said to be an orthogonal matrix if the product of a matrix and its transpose gives an identity value.

Q4) What are eigenvalues? What is the maximum number of Eigen Values and vectors for a Linear Transformation?

In linear algebra, an eigenvector or characteristic vector of a linear transformation is a nonzero vector that changes at most by a scalar factor when that linear transformation is applied to it.



STUDENT 2

ROUND 2

Q5) What is Batch Normalization? Why is it important? How is it handled during Evaluation?

Batch normalization is a technique used in deep learning to improve the training of neural networks. It is important because it can significantly improve the training of deep neural networks by reducing the effect of internal covariate shift.

During evaluation, the batch normalization parameters are fixed, and the normalization is applied using these fixed parameters rather than the mean and standard deviation computed during training.

Batch Normalization

Q6) Suppose your model has high variance, how would you change your model and training data?

In linear algebra, an eigenvector or characteristic vector of a linear transformation is a nonzero vector that changes at most by a scalar factor when that linear transformation is applied to it.

STUDENT 2

ROUND 2



Q7) What are Type 1 and Type 2 errors?

In statistics, a Type I error is a false positive conclusion, while a Type II error is a false negative conclusion.

Know more about Type I and Type II errors.

Q8) What are kernels in Convolution Neural Networks? How does parameter sharing work?

In Convolutional neural network, the kernel is nothing but a filter that is used to extract the features from the images.

Parameter sharing refers to the fact that for generating a single activation map, we use the same kernel throughout the image



STUDENT 2

ROUND 2

Q9) What is the difference between Gradient Descent, Stochastic Gradient Descent and Mini-Batch Gradient Descent?

Difference

Q10) What are Resnets? What are skip connections in Resnets and why are they used?

ResNet
Skip connections

Q11) What are Word Embeddings? How are they trained?

It is an approach for representing words and documents.

Training Word Embeddings.

STUDENT 2

ROUND 2



Q12) Code Binary Search in Python

Code

Pro Tip

Reinforcement Learning is an impressive skill, but it would not help you in any company interviews. Very few companies use it and very few interviewers know it themselves. Focus a lot on NLP instead.



EXL

SCREENING ROUND

There was no selection test but only resume shortlisting. Apparently, preference was given to candidates who had good skills in Python, Machine Learning, SQL, Data Visualization and High CPI.



EXL

INTERVIEW ROUND

STUDENT 1

ROUND 1



Q1) How did you get interested in data science and machine learning?

I developed an interest in data science and machine learning through self-study, perhaps due to a personal interest in the potential applications of these fields in areas such as healthcare, finance, or climate science.

To know more about data science

Q2) SQL Question

Code

Q3) Explain how Support Vector Machines Work. Give a Mathematical explanation.

Know more about SVM.



STUDENT 1

ROUND 1

Q4) In an India vs Pakistan cricket match in the World Cup, Guesstimate the distance traveled by the ball.

- Assuming a standard size cricket stadium, the maximum distance from the boundary to the center of the pitch is typically around 90-100 meters
- If a batsman hits a six, the ball travels over the boundary without bouncing and the distance covered by the ball could be in the range of 80-100 meters.
- However, the actual distance covered by the ball can be affected by various factors such as the angle and velocity at which the ball is hit, the atmospheric conditions (such as wind), and the condition of the ball itself
- Additionally, the distance traveled by the ball can also vary depending on the location of the stadium, as stadiums in different parts of the world can have varying dimensions and playing conditions
- A reasonable guesstimate for the distance traveled by the ball in an India vs Pakistan cricket match could be around 80-100 meters for a six hit by a batsman

STUDENT 2

ROUND 1



The interview began with a simple introduction of the candidate. The interviewer also asked candidates about their ambitions in life and their background. It went on for 2 to 3 minutes. The candidate had a project that used core electrical methods to obtain a dataset and then use ML on it. The interviewer asked the candidate to open the project on browser and pin pointed certain sections and asked them to explain the code.

Q1) Why have you used LSTMs? Why are they better than Vanilla RNNs?

LSTMs are better than vanilla RNNs because they can effectively learn long-term dependencies in sequential data by using a gating mechanism that selectively forgets or remembers information.

Detailed comparison

STUDENT 2

ROUND 1



Q2) What made you choose Adam Optimizer over other optimizers?

Adam Optimizer is a popular choice in deep learning because it is known to perform well on a wide range of problems and is computationally efficient. It combines the benefits of both Adaptive Gradient Algorithm (AdaGrad) and Root Mean Square Propagation (RMSProp) optimizers to effectively optimize the learning rate during training.

Q3) Why did you choose Categorical Cross Entropy as Loss function?

Categorical Cross Entropy is often used as a loss function in multiclass classification tasks, as it is well-suited for measuring the difference between predicted and actual probability distributions.

Know more.

STUDENT 2

ROUND 1



Q4) Why did you choose ReLU activation?

ReLU is a popular choice for deep learning models because it is computationally efficient and can help mitigate the vanishing gradient problem. Additionally, ReLU has been shown to work well in practice for a wide range of tasks and network architectures.

[More about ReLU activation.](#)

Q5) What is clustering and Why is it used?

Clustering is a machine learning technique that involves grouping together similar objects or data points into clusters based on their characteristics or features. It is used to identify patterns or structure within the data for various purposes. Clustering is used in many areas of machine learning, including image analysis, natural language processing, and recommendation systems.



STUDENT 2

ROUND 1

Q6) What do you understand about data cleaning? What are your favorite data-cleaning methods?

Data cleaning is an important step in the data preprocessing stage of machine learning because it helps to ensure the quality and reliability of the data being used for analysis.

Some **common data cleaning methods** include Removing Duplicates, Handling missing values, Removing outliers, Standardizing and normalizing data, and Correcting data types.

Q7) The probability of hitting one projectile is 0.3, what is the probability that you will miss all three?

- If the probability of hitting one projectile is 0.3, then the probability of missing one projectile is $1-0.3=0.7$. Assuming the events of hitting or missing each projectile are independent, the probability of missing all three is the product of the probabilities of missing each projectile:
- $P(\text{miss all three}) = P(\text{miss 1st}) \times P(\text{miss 2nd}) \times P(\text{miss 3rd}) = 0.7 \times 0.7 \times 0.7 = 0.343$
- Therefore, the probability of missing all three projectiles is 0.343 or approximately 34.3%.



STUDENT 2

ROUND 1

Q8) Using four 4's, bring an answer equal to 19 and also use the four mathematical signs i.e plus, minus, divide & multiply at least once

$$4! + 4 - 4/4$$

Q9) There are 10 bags and each bag has infinite coins, but one of the bags has forged coins. So genuine coins weigh 1 gm whereas a forged coin base 1.1 gm. How would you find that bag?

- It is known that there is only one bag with forgeries. To identify that bag, Ishita can follow a simple procedure. She should take out 1 coin from the 1st bag, 2 coins from the 2nd bag, 3 coins from the 3rd bag and similarly 10 coins from the 10th bag.
- Now she should simply weigh all these picked coins together. If there were no forgeries, then the total weight should be $(1+2+3+\dots+10) = 55$ grams.
- Now, if the total weight comes out to be 55.3 then she can conclude that the 3rd bag contain forgeries. So, if the total weight is $(55.n)$, then it is clear that the nth bag contain forgeries.



STUDENT 2

ROUND 1

The interviewer then moved on to guesstimates

Suppose you have a restaurant on the
Q10) roadside of a highway, Guesstimate its
revenue.

- Let's assume that the restaurant is located on a busy highway with a daily traffic volume of 10,000 vehicles, out of which 10% stop at the restaurant
- If the restaurant has a seating capacity of 50 and is open for 10 hours a day, with an average customer spend of \$20, the estimated daily revenue would be:
- $10,000 \times 10\% = 1,000$ potential customers per day
 $50 \text{ customers} \times \$20 \text{ spend} = \$1,000 \text{ revenue per hour}$
 $\$1,000 \text{ revenue per hour} \times 10 \text{ hours} = \$10,000$
estimated daily revenue

Q11) Guesstimate the number of Avocados in Mexico.



STUDENT 2

ROUND 1

- According to the latest available data, Mexico produced around 2.4 million metric tons of avocados
- Assuming an average weight of 150 grams per avocado, and considering that 85-90% of avocados are exported, it can be guesstimated that there were around 12.8 billion avocados produced in Mexico

Q12) Guesstimate the number of iPhone users in India.

Suppose, we are considering all models of the iPhone.

- There are 1.39 billion Indians. Out of which, 40% of the population includes children and senior citizens. This means that children and senior citizens will be excluded which leaves us with 834 million people that can own an iPhone.
- Now, out of all these people, we will consider the upper class and upper-middle class who can own an iPhone. For this, we will exclude the lower middle class, which is around 14%.
- This brings the probable number to 717 million. As per the statistics, the market share of the iPhone is 3.2%. This means there can be 22 million possible iPhone users in India.



STUDENT 2

ROUND 1

Q13) What do you think is the market share of Apple in phones?

As of 2021, Apple's market share in smartphones worldwide was around 14%, according to reports.

STUDENT 2

ROUND 2



This was an HR round.

Q1) What do you think the EXL's possible challenges were? How do you think COVID affected EXL?

Like many other companies, EXL faced several challenges during the COVID-19 pandemic, including operational challenges, economic challenges, cyber-security risks, and employee well being.

Q2) What do you think the EXL's possible challenges were? How do you think COVID affected EXL?

STUDENT 2

ROUND 2



- EXL is a global business process outsourcing (BPO) company that offers a range of services to clients in industries such as finance, healthcare, insurance, and utilities.
- Based on the fact that you are hiring a data analyst, it is likely that EXL is looking to leverage data and analytics to drive business insights and improve performance.
- By harnessing the power of data, EXL can gain a competitive advantage and help its clients achieve their business objectives more effectively.

Pro Tip

Always read up in advance about whichever company you are applying for. What do they do and what fields are they involved in? What are the potential problems they can face? Who are their competitors? Why do you want to work at EXL? How does this role suit you? These always help.



DECIMAL POINT

Innovative Data & Research Solutions

SCREENING ROUND



CODING QUESTIONS

The screening round had some MCQs and 2 Coding Questions. The MCQs were based on Logistic Regression, Bias Variance tradeoff and Probability and Statistics.

Q1) There are n number of islands numbered from 1 to n. A person wants to jump from island 1 to island n but there is a condition that the person can jump from one island A to island B if $\text{modulo}(A,B) = 0$. Find the number of ways in which the person can reach the last island.

[Link](#)



DECIMAL POINT

Innovative Data & Research Solutions

INTERVIEW ROUND



STUDENT 1

ROUND 1

The interviewer mostly focused on the candidate's projects and wanted to know the nitty gritty details of the algorithms used. The interviewer asked why the candidate used the specific algorithms and did they try any alternative algorithms. The projects were:

1. Prediction of discount prices of items listed on Amazon and Flipkart.
2. Exploratory Data Analysis based on Taxi Service like its pick up point, duration and leave point.

STUDENT 1

ROUND 1



Q1) What are the drawbacks of Random Forest and Decision Tree Models ?

An article containing the pros and cons of Random Forest

Key differences between Random Forest and Decision Tree Models

Blog on key disadvantages of Random Forest

Q2) What is Bagging? Explain the mathematics behind it.

Bagging, also known as bootstrap aggregation, is the ensemble learning method that is commonly used to reduce variance within a noisy dataset. In bagging, a random sample of data in a training set is selected with replacement—meaning that the individual data points can be chosen more than once.

Medium Article explaining math behind bagging

STUDENT 1

ROUND 1



Q1) Name all the metrics you know for Classification and Regression.

Medium article explaining 20 metrics for classification and regression.

Performance metrics for classification and regression.

Pro Tip

You should always know the drawbacks of the model you used in your projects.



Math Logic

INTERVIEW ROUND



STUDENT 1

ROUND 1

It started with a short introduction of the candidate. The interviewer asked questions on the projects.

Q1) The candidate had used AUC-ROC metric in their project, interviewer asked what do you understand about AUC-ROC and why is it used?

When we need to check or visualize the performance of the multi-class classification problem, we use the AUC (Area Under The Curve) ROC (Receiver Operating Characteristics) curve. ROC is a probability curve and AUC represents the degree or measure of separability. It tells how much the model is capable of distinguishing between classes.

[Article on AUC-ROC curve.](#)

[StatQuest video on AUC-ROC curve.](#)



STUDENT 1

ROUND 1

Q2) Explain the Bias Variance Tradeoff

If the algorithm is too simple (hypothesis with linear eq.) then it may be on high bias and low variance condition and thus is error-prone. If algorithms fit too complex (hypothesis with high degree eq.) then it may be on high variance and low bias. In the latter condition, the new entries will not perform well. Well, there is something between both of these conditions, known as Trade-off or Bias Variance Trade-off.

[Article on Bias Variance Tradeoff.](#)

[Video on Bias Variance Tradeoff.](#)

Q3) Why did you use CNN? How are CNNs different from RNNs and ANNs?

CNNs are specially used for image recognition and tasks that involve pixel data.

[Article on difference between CNNs, ANNs and RNNs.](#)



STUDENT 1

ROUND 1

Q4) There are 6 ants on each vertex of hexagon they started moving in random direction towards each other. What is the probability that no two ants meet ?

Ans . 1/32

Explanation of similar problem

Q5) Code a function to find the GCD of Two numbers in least time complexity.

```
1 # Iterative function in python to return
2 # gcd of a and b
3 def gcd(a, b):
4
5     # Everything divides 0
6     while(a > 0 and b > 0):
7         if (a > b):
8             a = a % b
9         else:
10            b = b % a
11
12     if (a == 0):
13         return b
14     return a
```

GFG link



STUDENT 1

ROUND 2

Q1) There is an unmarked die .We need to mark the faces of the die such that when two such dice are thrown , we need to get numbers from 1 - 12 with equal probability)

Mark one dice with {0,0,0,1,1,1} and other dice with {1,3,5,7,9,11}. The intuition is that each number should have probability 1/12 so each number should have 3 possible combinations.

Q1) Why did you use the TensorFlow framework for Deep Learning in your project? Why not PyTorch or any other?

There are a lot of advantages that TensorFlow enjoys over PyTorch when it comes to production. Better performance due to the static computation graphs, packages/tools which help in fast development over platforms like browser, mobile, and cloud.

[Article on Tensorflow vs Pytorch](#)



STUDENT 1

ROUND 2

Q3) Describe the entire pipeline implementation of your project.

Pro Tip

Do not underestimate the importance of DSA in the screening rounds of the intern test.



We wish you all the best for the upcoming internship season and hope you land your dream job!
Please fill out this [feedback form](#) so that we can improve our future publications.

