**Python Basic (Compulsory)**

1. Count occurrence of each letter in Nishant and store it in dictionary. {n:2, i:1, ..}. Case doesn’t matter
2. Store below in df

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| game\_id | team\_name | inning | over | runs |
| 1 | India | 1 | 1.1 | 3 |
| 1 | India | 1 | 1.2 | 3 |
| 1 | Pakistan | 2 | 1.1 | 3 |
| 2 | Indian | 1 | 1.1 | 4 |
| 2 | Australia | 2 | 1.1 | 4 |

1. In the above df select all rows where game\_id = 1 and inning = 1.
2. For each game\_id,inning get total\_runs and total\_balls played .

|  |  |  |  |
| --- | --- | --- | --- |
| game\_id | inning | total\_balls\_played | total\_runs |
| 1 | 1 | 2 | 6 |
| 1 | 2 | 1 | 3 |
| 2 | 1 | 1 | 4 |
| 2 | 2 | 1 | 4 |

1. Output of df.loc[0] and df.iloc[0] for below df

|  |  |
| --- | --- |
| Index | Column |
| 49 | A |
| 48 | B |
| 47 | C |
| 0 | D |
| 1 | E |

1. How to convert a datetime object in pandas stored in string format ?

**Python Advance**

1. Now we have to store total\_runs in original\_df itself.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| game\_id | team\_name | inning | over | runs | Total\_runs |
| 1 | India | 1 | 1.1 | 3 | 6 |
| 1 | India | 1 | 1.2 | 3 | 6 |
| 1 | Pakistan | 2 | 1.1 | 3 | 3 |
| 2 | Indian | 1 | 1.1 | 4 | 4 |
| 2 | Australia | 2 | 1.1 | 4 | 4 |

**ML Basic (Compulsory)**

1. BIAS (Polynomial Regression) > or < or = BIAS (Linear Regression)

Variance (Polynomial Regression) > or < or = Variance (Linear Regression)

1. Significance of p value in linear regression.
2. Formula for cost Function of logistic regression.
3. How does decision tree decide which feature to split on?
4. Should you increase or decrease number of trees in random forest if you find test accuracy is quite lower than train accuracy.

**ML Advance**

1. Out of l1 or l2 regularization which leads to zero coefficients and why?
2. What happens if there is a multicollinearity in your dataset and you build Linear Regression, will your accuracy get impacted or something else?
3. Test f1 score is better than train f1 score, when it can happen?
4. Let’s say you build a decision tree and you find out you have high variance then what you will do?
5. A logistic regression with F features and C classes. What will be shape of weight matrix?

**Model Building Basic (Compulsory)**

1. Let’s say you are given imbalance dataset for classification please comment on how you will do following things with
   1. How many parts you will divide dataset?
   2. What all preprocessing you will do for numerical and categorical features.
   3. How will you handle Imbalance here?
   4. What Evaluation Criterion you will use?

**Model Building Advanced**

1. Let’s say you are building binary classification model and you want better precision for positive class will you increase or decrease threshold.

**DL Concepts Basic**

1. Let’s say you built a model but find training error to be quite lower than validation error what you will do?
2. Let’s say we have a very large neural network model but without any activation function except last layer (let’s say sigmoid for classification). Please comment about its performance in case when data is linearly and non-linearly separable.
3. What is Batch Normalization and how does it help?
4. What are different types of Gradient Descent and which one is better?
5. Name different activation function and their uses.
6. What is Vanishing/Exploding Gradient?

**DL Concepts Advanced**

1. How do you do dropout during inferencing?

**NLP Basic**

1. Is there any difference in methods for train and test while doing a TF-IDF transformation?
2. Explain word2vec architecture.
3. Let’s say you are given text data and you are asked to perform classification what approach you will follow?
4. How does LSTM solve vanishing gradient problem?
5. Explain self-attention.
6. What all problems are solved by transformers and how are they better
7. What is the difference BERT embeddings and word2vec embeddings?

**NLP Advanced**

1. Why does SVM (linear kernel) works better than any other kernel in case of tfidf vectors?
2. Out of skip gram and cbow which one is better for rare words.
3. Explain how to perform fine tuning of language model.

**CV Basic**

1. Max Pooling vs Avg Pooling in case of noise.
2. Let’s say input matrix size is 12\*12, filter size 3\*3 and stride of 1. What is output matrix size. Calculate without using formula.
3. Suppose you are given with very few handwritten images of 0 and 1 (let’s say 0 – 20 and 1 – 40). How will you build a model for it?
4. Explain Siamese network.

**CV Advance**

**Time Series**

1. Explain all parameters in SARIMAX modelling.
2. What metric to measure performance.

**Stats**

1. Explain Central Limit Theorem
2. Standardization vs Normalization

**Probability**

1. How can you come up with a fair coin from a biased coin?
2. P(HHT) > or < or = P(HTT)
3. Difference between probability and likelihood.

**Project Related**

**Case studies**

1. How can you predict a gender given name, provided some names with their gender?

**Aptitude Question**

1. How can you determine a fake profile in Facebook?