

MOCK QUESTIONS

①

PYTHON; PL → (Interpreted) → HLL

Q1. Data Types

=)

1. Character Data Type → String

2. Numeric Data Types

{ 1. Int → whole number
2. Float → Decimal
3. Complex → Real + Imaginary

3. Boolean Data Types

1. True (1)
2. False (0)

Q2. Data Structure

=)

Mainly 4 Data Structures

1. Lists → mutable

2. Tuples → Not mutable

3. Dictionary → Partially mutable

4. Sets → Only Unique Elements

Q3. Self

⇒ is the basic parameter used while defining instance of a class.

It is also useful to access variables that belong to the class

Q4. **while vs For loop**

⇒ while → condition based
For → Iteration based

Q5. **Decorators** OOPS

⇒ is a function that takes existing function and modifies its behaviour without changing its structure.

example Assigning function to a variable

```
def func(x);  
    return x*x
```

Multiplier = func * 10

multiplier(10)

Q6. Memory Management;

⇒

Function calls /
References

=>

Stack
memory

Objects

=>

Heap
memory
(private heap)

Q7.

Types of Functions

⇒

1. Self Defined
2. Lambda
3. Recursive
4. Inbuilt

Q8.

Lambda Functions

⇒

is an anonymous function which means it can be used without even naming the function

Lambda x: x * x

Q9. **Keywords**

⇒ Reserved words that cannot be used as variable names. They have a functionality associated.

Q10. **Identifiers**:

⇒ Naming functions & variables.

②

STATISTICS branch of applied mathematics which deals with data analysis for better decision making

Q1. **Types**

⇒ 1. Descriptive → Summary
2. Inferential → Conclusions

Q2. **Types of Tests**

→ **T-Table**

⇒ 1. **T-Test** → Statistical Test
• When SS is less than 30
• When population variance unknown

2. **Z - Test** → Statistical Test

- when sample size > 30
- when population variance is known
- we use **Z-table**

Q3. ↓ **P-value**; is the probability value used in credibility test.

which is used to prove credibility of null hypothesis against calculated values

Q4. **STD** → Standard Deviation

$$s_{(\text{Sample})} = \sqrt{\frac{(x - \bar{x})^2}{n-1}} ; \sigma = \sqrt{\frac{(x - \bar{x})^2}{N}}$$

Q5. **CLT**

Increase in Sample \Rightarrow Normal Size distribution

Q6. **Distributions** → a. **Discrete** b. **Continuous**

Q2. Confusion MATRIX

		(Actual)		
		TP	TN	← Type II
(Predictions)	TP	FP FN		
	FP	TN		

↑
Type I

$$1. \text{ Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

↳ correct predictions.

$$2. \text{ Precision} = \frac{TP}{TP + FP}$$

↳ correct positive predictions

Also klas positive predictiveness

- Opposite Negative predictiveness
 $= \frac{TN}{TN + FN}$

3. Recall / Sensitivity

$$= \frac{TP}{TP + FN}$$

Ratio of Total positives to actual positives

4. F1 Score;

↳ Measure of model's accuracy / performance

$$F_1 = 2 * \frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}$$

↳ Harmonic mean of beta.

③

OOPS

1. Four Pillars

Get

Set

1. Inheritance → Single multi
2. Encapsulation → Self. — Acc.
3. Polymorphism → Many forms
4. Abstraction

Polymorphism

Many Forms

1. Operator Overloading; Example; we can use '+' operator

2. Method OVERRIDING ;
- 1. Inheritance ✓
 - 2. Between two classes
 - 3. Same / Name / parameter

3. Method Overloading ;
- 1. Inheritance (not necessary)
 - 2. within class
 - 3. Same name / different parameters

2. Constructor

For initializing instance variables

3. Object Oriented ;

python is a famous programming language which supports OOP

- => OOP is the programming which makes use of concepts ; classes and objects
- => Basically, it helps us interact with real world entities (objects) and build functions around them.

(4) MACHINE LEARNING:

is a set of algorithms to help our machine learn different patterns

1. Supervised

- Requires Labels
- Classification / Prediction

Regression = Linear, DTR, RFR

Classification = Logistic, SVM, DT, RF
KNN, Image classifiers

2. Unsupervised

- NO Labels
- Finds hidden patterns

①

Clustering = K-means, Hierarchical

- a. Divisive
- b. Agglomerative

②

Association = • Market Basket Analysis
= • Recommendation Systems

③

Anomaly Detection = • Fraud Detection
• Fault Detection

④ Dimensionality Reduction

3. Reinforcement

Learning through Trial & error.

- No data only environment is given
- Reward system

Robotics

Self Driving Cars

Q1. SUPERVISED;

↳ R^2 Score Goodness of Fit

Variance of y explained by x

$$\left(1 - \frac{RSS}{TSS}\right) \text{ or}$$

$$\frac{\text{Explained Variance}}{\text{Total Variance}} = \frac{\text{Unexplained Variance}}{\text{Total Variance}}$$

Q2. R^2 -Squared vs R^2 Adjusted

↳ The main difference is that adjusted R^2 takes into consideration the number of independent variables used to predict y .

=. This way R^2 helps us identifying **optimal** number of **x -variables** to be used

=. If new variable is added
that means

R^2 will increase which means
 R^2 adjusted will also increase

But

If on addition R^2 doesn't increase ; It means adjusted R^2 score will decrease

=. R^2 score will always be greater than adjusted R^2 score.

Q3. **Variance**:

Difference when new dataset is fed

Bias; Systematic error that occurs due to incorrect assumptions

Q4. **Entropy** measure of impurity

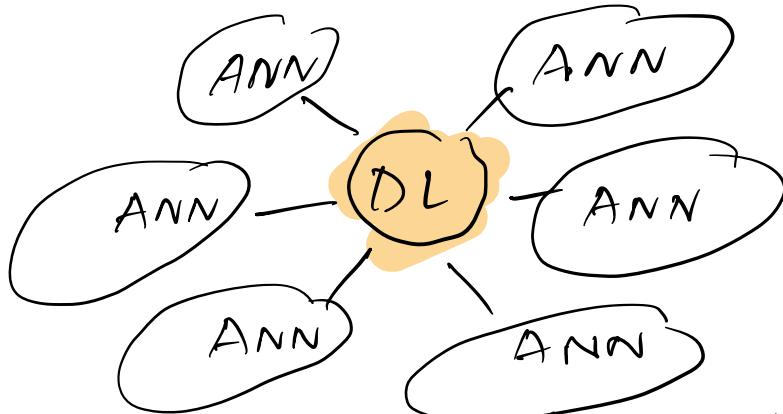
 $(\text{parent } E) - w * (\text{child } E)$

↳ Quality of split

Q5. **AI** = Developing intelligence for machines mimicing Human Intelligence

- ↳
- ① Generalized / strong
 - ② Applied Ai

DEEP LEARNING → Concept of (unstructured data) Neural Networks



- Large amounts of unlabeled data
- Highly complex computation power required

Q6. Chat GPT;

↳ is an AI chat Bot developed by open AI

↳ Language Models in addition to fine tuned supervised and reinforcement learning Techniques.

Q7. AUC

Area under curve. = (0, 1)

⇒ Ability of model to distinguish between positive and negative classes

$AUC = 1$ = perfect

Q8. ROC

Receiver operating characteristic curve

↳ performance of classification at all classification thresholds

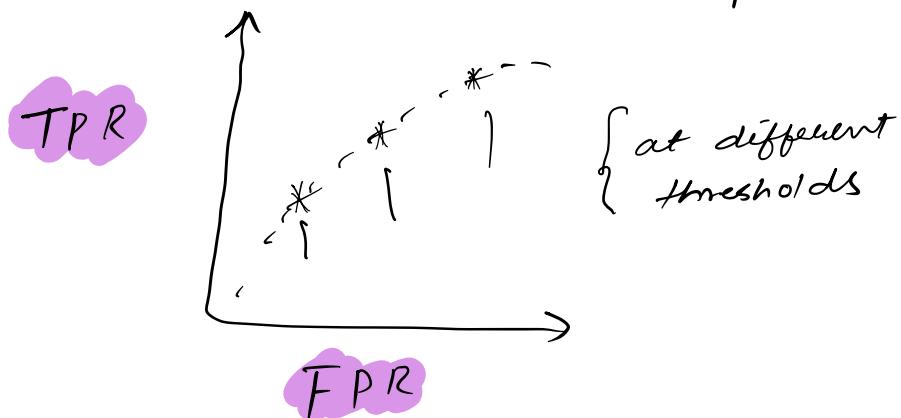
{ Two parameters used;

$$TPR = \frac{TP}{TP + FN}$$

True positive Rate

$$FPR = \frac{FP}{FP + TN}$$

False positive Rate



NOTE; Lowering classification thresholds classifies more items as +ve;
Thus TP and FP increases

Q9. Machine Learning Projects;

- Regression
- classification

Q10. Explain

- Univariate
 - Bivariate
 - multi-variate
- } Analysis

Q11. How does Random Forest avoid Overfitting

→ Decision Trees training
on different Data Samples

"
)
OOB → Out of Bag
() low variance

⑤ SQL

1. How to delete primary and foreign key tables
2. Group by clauses

3. char & varchar
4. difference between primary
& Secondary key.
5. Joins → Important
6. write Queries

All
THE
BEST!