Stage 1 - Syllabus

Section 1: Foundational Topics

- **Computational and Logical Thinking**
  - Pattern recognition and rule discovery
  - Logical reasoning puzzles
  - Basic understanding of programming constructs (variables, operations, if-then else, for loops)
- **Foundations of Algebra**
  - Functions
  - Roots and properties of Polynomials, Quadratic Equations
  - Arithmetic and Geometric Series; Summations
  - Common inequalities (AM-GM inequality, Jensen's inequality)
- **Probability**
  - Basics of probability (independent and dependent events)
  - Applications of Bayes' Theorem with practical examples
- **Statistics**
  - Descriptive statistics: mean, median, mode, variance, percentile
  - Independent variables
  - Correlation and causation
- **Linear Algebra and Matrices**
  - Basic matrix operations (addition, multiplication, and inverse)
  - Solving systems of linear equations
- **Optimization**
  - Linear optimization with constraints
  - Concept of maxima and minima in simple scenarios
- **Data Representation and Visualization**
  - Base-n coding
  - Understanding tabular data, hierarchies, entity-relationship graphs
  - Interpreting bar graphs, histograms, pie charts, and scatter plots
  - Drawing insights and identifying trends from visualized data

Section 2: ML Theoretical Foundations

Students are NOT expected to know these concepts in depth but should be able to apply these when provided the necessary definitions in the question.

- **Learning Fundamentals**
  - Concept of an ML model: Inputs, outputs, and learning from data.
  - Train-Test Process
  - Purpose of loss functions
- **Evaluation Metrics**
  - **For Regression:** Root Mean Squared Error,, Mean Absolute Error

    - **For Classification:** Accuracy, Precision, Recall, F1-Score, Confusion matrix, ROC Curve and AUC (basic interpretation)

Section 3: Algorithms

Students are NOT expected to have any prior knowledge of these ML algorithms. Questions on the topics below will include all the relevant background information in the problem statement and students are only expected to reason about them and apply them to specific problems.

- **Supervised Learning**
    - Linear Regression
    - Linear Classifier
    - K-Nearest Neighbors (K-NN)
    - Decision Trees
- **Unsupervised Learning**
    - K-Means Clustering
    - Principal Component Analysis (PCA)
    - Hierarchical Clustering
- **Probabilistic Models**
    - Naive Bayes
- **Famous Algorithms**
    - PageRank
    - TF-IDF Ranking

Stage 2 - Syllabus

Section 1: Foundational Topics

- **Computational and Logical Thinking**
    - Pattern recognition and rule discovery
    - Logical reasoning puzzles
- **Foundations of Algebra**
    - Functions
    - Roots and properties of Polynomials, Quadratic Equations
    - Arithmetic and Geometric Series; Summations
    - Common inequalities (AM-GM inequality, Jensen's inequality)
- **Probability**
    - Basics of probability (independent and dependent events)
    - Applications of Bayes' Theorem with practical examples
- **Statistics**
    - Descriptive statistics: mean, median, mode, variance
    - Correlation and causation

- **Linear Algebra and Matrices**
    - Basic matrix operations (addition, multiplication, and inverse)
    - Solving systems of linear equations
    - Determinants and their applications
- **Optimization**
    - Linear optimization with constraints
    - Concept of maxima and minima in simple scenarios
- **Data Representation and Visualization**
    - Base-n coding
    - Understanding tabular data, hierarchies, entity-relationship graphs
    - Interpreting bar graphs, histograms, pie charts, and scatter plots
    - Drawing insights and identifying trends from visualized data

Section 2: ML Theoretical Foundations

Students are expected to have some fluency in these concepts and should be able to apply them when provided with definitions

- **Learning Fundamentals**
    - Concept of an ML model: Inputs, outputs, and learning from data.
    - Train-Test Process
    - Purpose of loss functions
    - Overfitting and underfitting
    - Hyperparameter Tuning
- **Evaluation Metrics**
    - **For Regression:** Root Mean Squared Error,, Mean Absolute Error
    - **For Classification:** Accuracy, Precision, Recall, F1-Score, Confusion matrix, ROC Curve and AUC (basic interpretation)

Section 3: ML Algorithms

Students are not expected to know these algorithms in depth but some familiarity will be helpful. They should be able to apply these algorithms to specific problems/datasets when provided with minimal guidance.

- **Supervised Learning**
    - Linear Regression
    - Linear Classifier
    - K-Nearest Neighbors (K-NN)
    - Decision Trees
- **Unsupervised Learning**
    - K-Means Clustering
    - Principal Component Analysis (PCA)
    - Hierarchical Clustering
- **Probabilistic Models**
    - Naive Bayes

- **Famous Algorithms**
    - PageRank
    - TF-IDF Ranking

Section 4: Python Programming

- **Python Basics**
    - Syntax, loops, and functions
    - File handling and basic input-output operations
- **Data Handling**
    - NumPy and Pandas for efficient data processing
- **Visualization**
    - Matplotlib and Seaborn for creating and interpreting visualizations
- **ML Libraries**
    - Introduction to scikit-learn: classification, clustering, regression, metrics modules
    - Training and testing models with specific algorithms on provided datasets

Section 5: Gen AI Tool Use

- **ChatGPT [free version]**
    - Prompt engineering for data manipulation, code generation,
    - Prompt engineering techniques (chain of thought, in-context learning, instruction-chaining , retrieval-augmentation)