TCS 509

MACHINE LEARNING

Unit 1:

Review of Statistical Concepts:

- Mean
- Median
- Mode
- Outliers
- Range
- Average Deviation
- Absolute Deviation
- Squared Deviation
- Standard Deviation
- Total Sum of Squares

Introduction to Machine Learning:

- What is Machine Learning,
- Introduction to ML's three approaches:
    - Supervised
    - Unsupervised
    - Reinforcement Learning

Introduction to Python:

- Basic Operations
- Lists
- Tuples
- Dictionaries
- Flow Control
- Strings
- File handling
- Numpy
- Scikit-learn

Unit 2:

Introduction to Exploratory Data Analysis

- Introduction to Exploratory Data Analysis (EDA)
    - Steps in EDA
- Data Types:
    - Numerical Data
    - Discrete data
    - continuous data
    - Categorical data

Data Transformation

- Transformation Techniques:
    - Performing data deduplication
    - replacing values
    - Discretization and binning
- Introduction to Missing data
- handling missing data
- Data Visualization using Matplotlib, Seaborn

Unit 3:

Supervised Learning Algorithms

- Linear Regression
- Logistic Regression
- Decision Trees
- Random Forest
- Support Vector Machine
- K- Nearest Neighbours
- CN2 Algorithm
- Naive Bayes

Unit 4:

Clustering:

- K-means,
- Silhoutte Scores,
- Hierarchical Clustering,
- Fuzzy c-means,
- DBScan

Dimensionality Reduction:

- Low Variance Filter,
- High Correlation Filter,
- Backward Feature Elimination,
- Forward Feature Selection,
- Principle Component Analysis,
- Projection Methods

Unit 5:

Model Evaluation and Selection:

- Cross-validation
- model evaluation metrics
- model selection
- hyperparameter tuning

Hyperparameter Optimization Techniques

- Manual Search
- Random Search
- Grid Search
- Case study in Python for Hyperparameter Tuning