CS1AC16 – Application of Computer Science – Spring Term

Roadmap of Data Analytics Parts

1. **Introduction** (Week 7)
    - Definition: Data Analytics, Data Mining, and Data Science
    - Applications of Data Analytics
    - Tools of Data Analytics
    - Exercises 1
2. **Data** (Week 7)
    - Properties of Data
        - Type of Features/Attributes
        - Type of Values of Features (Nominal, Ordinal, Interval, Ratio)
    - Type of Data
        - Structured data (excel, database, and graphs)
        - Unstructured data (Image and text)
    - Quality of Data
        - Data dimensionality, sparsity, and resolution
        - Data cleaning, noise, missing values, outlier and duplicate values
    - Exercises 2
3. **Data Pre-processing** (Week 8)
    - Simple pre-processing
        - Processing missing values (mean, mode, and linear interpolation)
        - Aggregation and Sampling
        - Dimensionality reduction (Principal component analysis)
        - Exercises 3
    - Advanced pre-processing
        - Feature selection and feature engineering
        - Attribute Transformation
            - Discretization and binarization
            - Normalization and standardization
        - Principle of data Modelling, validation and Scoring
        - Exercises 4
4. **Data Modelling** (Week 9)
    - Type of Learning (Supervised and Unsupervised)
    - Linear Regression
    - Metric for evaluation of models (mean squared error and R2 score)
    - Problems of model and solutions
        - Model overfitting and generalization
        - Solutions to model overfitting
    - Exercises 5
5. **Lab exercises** (Week 9 and 10)
    - KNIME tool – data loading, filtering, modelling, and visualization Exercises 1 – 5