SUBJECT: Data Science

Semester: 7th   **SUB CODE: CIE 405T**

Assignment 1

Introduction to Data Science and Data Pre-processing

1. What is Data Science, and how is it related to AI and Machine Learning?

2. How does Data Science differ from AI and Machine Learning? Provide real-world examples.

3. Explain the role of Python in Data Science. Why is it a popular choice for data analysis?

4. What is Google Colab, and how does it support data science tasks?

5. List and explain the importance of three popular dataset repositories.

6. What are the steps involved in data pre-processing, and why is it essential in data analysis?

7. Define and differentiate between data scales (e.g., nominal, ordinal, interval, ratio).

8. What are similarity and dissimilarity measures, and how are they useful in data analysis?

9. Explain the significance of data transformation and merging in data pre-processing.

10. What is PCA (Principal Component Analysis), and how does it help in dimensionality reduction?

Assignment 2

Regression Analysis

1. What is regression analysis, and how is it used in predictive modeling?

2. Explain the difference between linear regression and generalized regression.

3. What is regularized regression, and why is it used? Give examples of common regularization techniques.

4. Define Ridge regression. How does it differ from regular linear regression?

5. What is the purpose of cross-validation in model evaluation?

6. How do you split data into training and testing sets, and why is this important?

7. What is nonlinear regression, and in which scenarios would you use it?

8. Explain the concept of latent variables and their role in data modeling.

9. What is Structural Equation Modeling (SEM), and how does it differ from traditional regression?

10. What are the advantages and limitations of regression models in machine learning?

Assignment 3

Time Series Analysis and Forecasting

1. What is time series data, and how is it different from other types of data?

2. Explain the concept of stationarity in time series data. Why is it important?

3. What are the common methods to test for stationarity in time series data?

4. Define seasonality in time series analysis. How does it affect forecasting models?

5. What are autoregressive (AR) models, and how do they work in time series forecasting?

6. What is the role of recurrent models (e.g., RNN, LSTM) in time series forecasting?

7. How do you decompose a time series into trend, seasonality, and residual components?

8. Explain the importance of lag in time series analysis. How does it affect forecasting?

9. What are the key metrics used to evaluate the performance of time series forecasting models?

10. What are the challenges associated with time series forecasting, and how can they be addressed?

Assignment 4

Classification and Clustering Techniques

1. What is the difference between classification and clustering in data analysis?

2. Explain Linear Discriminant Analysis (LDA) and its use in classification problems.

3. How do Support Vector Machines (SVM) work, and what are their key features?

4. What are decision trees, and how do they classify data?

5. How do you evaluate the performance of a classification model? Discuss at least three evaluation metrics.

6. What is clustering, and how does it differ from classification?

7. Explain the K-Means clustering algorithm. What are its strengths and limitations?

8. What is hierarchical clustering, and how does it differ from K-Means?

9. Describe the DBSCAN clustering algorithm and its advantage over K-Means.

10. What are the key criteria for evaluating clustering performance? Explain the silhouette score.