**University of Engineering & Technology Lahore**

**Narowal campus**

**Department of computer sci. & engg**

| Title | Introduction to Data Science | | |
| --- | --- | --- | --- |
| **Code** | CS- | | |
| **Credit Hours** | 3.0 | | |
| **Prerequisite** |  | | |
| **Instructor** | Sadia Tariq, sadiatariq@uet.edu.pk | | |
| **Aims and Objectives** | This course provides an introduction about data science, its application and needs in today’s world. This course also introduces the students with cutting-edge tools used by data scientists around the world. These tools include Python, Numpy, Pandas, Jupyter Notebooks, Scikit Learn and a bunch more. The basic aspects to understand and learn will be exploration, visualization, prediction, confirmation and evaluation. Also, to introduce and practically apply basic machine learning techniques such as classificaiton and regression used for Data Analysis. | | |
| **Text Books** | 1. Data Science from Scratch by Joel Grus, 1st ed. 2. Python for Data Analysis by Wes McKinney 1st ed. | | |
| **Reference Books** | 1. Data Mining, Concepts and Techniques by Kamber and Han | | |
| **Assessment Criteria** | Sessional 25% | Mid 25% | Final 50% |
|  | Quizzes 10% Assignments 5%  Project/Presentation 10% |  |  |

**Sixteen-week lecture plan**

| **Week wise distribution** | **Topics Covered** |
| --- | --- |
| **Week 1** | Introduction to the subject of Data Science  Data Science Life Cycle and Process, Types of Data and Data Pre-processing stages |
| **Week 2** | Introduction to Python Data Science Stack (jupyter notebooks, pandas, some Basic Statistics, matplotlib (viz) and numpy. |
| **Week 3** | Exploratory Data Analysis |
| **Week 4** | Effective visualization of data |
| **Week 5** | Introduction to Prediction and Inference (supervised & unsupervised) |
| **Week 6** | knn regression,Linear and Polynomial Regression, |
| **Week 7** | Multiple Regression,Model Selection, |
| **Week 8** | Logistic Regression (linear and polynomial),Multiple Log- Regression |
| **Week 9** | Classification with decision trees, Classification Random Forrest, |
| **Week 10** | Bagging, Boosting, Stacking, Support Vector Machine |
| **Week 11** | Bias-Variance Tradeoff, Model Evaluation |
| **Week 12** | Performance Metrics (Accuracy, Contingency Matrix, Precision-Recall, F-1 Score, Lift, etc.), |
| **Week 13** | Introduction to Map-Reduce paradigm. |
| **Week 14** | Dimensionality reduction-Principal component analysis |
| **Week 15** | Clustering |
| **Week 16** | Revision and Future directions |