

National University of Sciences & Technology
School of Electrical Engineering and Computer Science
Department of Computing

CS 423: Data Warehousing and Data Mining (3+1): BESE-10/BSCS-9 Fall 2022

Course Project	
Maximum Marks: 80 (50 component+30 lab component)	Instructor: Dr. Rabia Irfan
Announcement Date: Week 5; Fall 2022	Due Date: Week 17; Fall 2022

Instructions:

--You can have a group of 3-4 members.

Project Board:

1. Media conspiracy of spreading fake news against Pakistan
2. Crop prediction based on soil quality (mainly focusing on satellite data)-For Pakistan Region
3. Price prediction in real estate's mainly focusing on Pakistan\
4. Text mining applications-Research Assistant
5. Flood Prediction using different types of data such as satellite images, meteorological etc.

Choose any of the above ideas¹, the focus should be on the following steps:

1. Domain Understanding/Gap Analysis/Application in Real-life
2. Data Collection and Preparation (Avoid using data from Kaggle or any such open links)
3. Detailed Exploratory Data Analysis and Pre-processing/Data Modelling/System Architecture Designing
4. Application of Data Mining Techniques (e.e. Clustering, Classification, Association Rule Mining etc.)
5. Dashboard Creation²/Prediction from the created Model

Project Deliverables Revisions:

1. **Project Report-Part 1:** Develop a short report presenting your analysis of the real-world problem that you can be able to address via your project solution. Express your motivation and gap analysis clearly. (PLO 7)-[5 pts]
2. **Project Report-Part 2:** Develop a comprehensive report for demonstrating your data model/architectural model/software model for the solution you are intending to develop and also give the timeline and work break for different phases. (PLO 9)-[10 pts]
3. **Project Report-Part 3:** Develop a comprehensive report for demonstrating your project solution and its implementation details. (PLO 12)-[10 pts]
4. **Project Presentation:** Prepare an 8-10 slides presentation demonstrating your complete project. All team members are required to participate in that. (PLO 9)-[15 pts]

¹ Ideas can be moulded as per your desired data but broadly it should follow the theme.

² Tableau can be used for dashboard creation. Help material is available on LMS for self learning.

5. **Project Dataset:** Submit the dataset developed for the purpose of your project implementation. (PLO 7)-[10 pts]
6. **Project Code/Final Lab:** Submit the complete code for your project. (PLO 12)-[10 pts]
7. **Project Code Demo/Final Lab:** Demonstrate your project and also prepare for the viva as part of your final lab. (PLO 12)-[20 pts]

Note:

- Plagiarism will be checked so please be careful.
- Data Mining, Machine learning, AI and Deep Learning are interconnected fields. So you may find them overlapping in terms of the project execution. Guidance will be provided as required; You are open to use tools and technology of your choice during implementation, however, Python is preferred.