

DATA MINING

FROM THE PERSPECTIVE OF BIG DATA ANALYSIS

COURSE FLYER – FALL 2013 – Graduate Level

Course Description:

This course focuses on the concepts, techniques and applications of the most prominent concepts and techniques of Data Mining. Data Mining, and Big Data Analysis in general, is currently the most practical technology in diverse corporate sectors across the world. It focuses on the extraction of interesting patterns from large amount of data (big data), which are implicit, previously unknown, and practically useful. This course aims to train the students in pattern-extraction skills, for different types of data mining techniques and algorithms. This is done within the cadre of Big Data Analysis and Predictive Analytics, the two cutting-edge data mining technologies in force today. It also aims to train the students in research-based activities, specifically through implementing the course project – students already involved in the corporate sector will be able to implement research-based projects on their corporate data. Besides data mining, the course will also focus on Data Warehousing and Business Intelligence, which are very critical for data analysis in any corporate environment.

Course Outline

A tentative list of topics is given below (subject to modification later on):

Topic	Description	Lectures
1	Introduction to Big Data Analysis and Data Mining –Background & Applications	1
2	Mining Associations and Correlations – Linking data through rules – Apriori, FP Growth	3
3	Predictive Analytics and Time Series Forecasting – Dividing the data into different classes, making predictions, predicting future values etc.	4
4	Cluster Analysis – Making sense of the data without any supervision; learning to divide data into useful categories	3
5	Mining Data Streams	1
6	Data Warehousing Seminar Series: Data Preprocessing – Preparing data for mining, OLAP – Mining analytical knowledge from large data repositories, Dashboard Reporting tools – Jasperforge	3

Nuts and Bolts (Algorithms and Tools):

The students will be learning about Predictive Analytics, the latest cutting-edge technology to make predictions for optimizing business objectives and acquiring corporate goals efficiently and effectively. It involves algorithms like Decision Trees, Naïve Bayes, and Neural Networks. Their use will be made to look simple through the Rapid Miner tool. We will also focus on two cluster analysis algorithms, e.g., CLOPE and K-means, which assist in acquiring Customer Relation Management (CRM) and product marketing objectives. Time series forecasting algorithms will include ARIMA and Exponential Smoothing, and will be demonstrated using IBM SPSS tool – they are helpful in predicting sales, cash flows, financial fraud, purchase behavior etc. Finally, Data warehousing will be demonstrated through MS SQL Server and dashboards will be shown on Jasper's iReport.

Administrative Info:

Instructor: Dr. Tariq Mahmood – Associate Professor – PAF-KIET | Office Hours: TBA

Instructor Web Site: <https://sites.google.com/a/nu.edu.pk/tariq-mahmood/>

Primary Textbooks: Data Mining: Concepts and Techniques (3rd Edition), by Jiawei Han, Micheline Kamber, and Jian Pei - **Reference Books:** Wikipedia entries (the most reliable information source on the Internet), or any text book on Data Mining.

Assignments and Project

The students would be required to divide themselves into groups. Both the assignments and the project are to be submitted collectively by the whole group.

About the Instructor

Dr. Tariq Mahmood was previously an Assistant Professor at the National University of Computer & Emerging Sciences (FAST-NU), Karachi. He has around 30 conference publications and five recognized journal publications, and was the head of the Data Mining and Business Intelligence Laboratory at FAST-NU. He has around ten years of research experience in the domains of Machine Learning and Data Mining technologies. He also has five years of both research and practical experience in Big Data Analysis, Predictive Analytics and Customer Relationship Management, which are the current hallmarks of Business Intelligence applications within corporate organizations across the world. Notably, Dr. Tariq has currently applied these technologies to the domains of Healthcare, Finance, Banking, Sales, Telecommunication Call Detail Records, Social Networks (Twitter), and Sports records (Cricket and Football) in order to extract useful and interesting patterns for the top-level management. He has also offered the very first course on Big Data Analysis in Pakistan. He has been an invited speaker at various conferences and summits, notably the first CIO summit held in Karachi in May 2013.