|  |  |  |  |
| --- | --- | --- | --- |
|  | **KENYATTA UNIVERSITY**  **QUALITY MNAGEMENT SYSTEM** | Ref. : | *KU/ACAD/SOP/8.5-3* |
| Ver. : | 1.0 |
| **Title** | **Course Outline** | **Date :** |  |

1. **SIT 409: DATA MINING AND KNOWLEDGE DISCOVERY**
2. **Purpose**

This course aims at exposing the student to concepts of data mining and the various predictive analytics techniques.

1. **Expected Learning Outcomes**

On completion of this unit the learners will be able to:

* Explain the need for data mining
* Outline the data mining application
* Describe the data preprocessing activities
* Describe the data ware house
* Use Apriori Algorithm to determine frequent item sets.
* Illustrate data classification method
* Apply various data clustering method in predictive analytics
* Describe the outlier detection method

1. **Course Content**

Overview of Knowledge Discovery and Data Mining, Overview of Knowledge Discovery and Data Mining, Data Preprocessing, Classification and Prediction, Classification and Prediction, Mining Association Rules, Mining Association Rules, Cluster Analysis, Cluster Analysis, Text Mining, Web Mining ,Knowledge evaluation  Case Studies

|  |  |  |
| --- | --- | --- |
|  | **TOPIC** | **SUBTOPIC** |
| Week 1 | **DATA MINING CONCEPTS** | Introduction to data mining  , data mining application , kinds of data that can be mined , types of pattern/ knowledge to be mined, data mining methodologies |
| Week 2 | **DATA PRE-PROCESSING** | Data pre-processing , data cleaning & data integration data reduction , data transformation & discretization |
| Week 3 | **DATA WAREHOUSING AND OLAP** | Data warehouse: basic concepts, merits and demerits, architecture of data warehouse, online analytical processing, data mining issues, mining methodology, performance and data types issues, data mining engine , data mining systems |
| Week 4 | **INTRODUCTION TO PYTHON** | Python environment, basic syntax, variable types, python modules, statistical functions, visualization |
| Week 5 | **MINING FREQUENT PATTERNS – ASSOCIATION** | Frequent item sets – association , support and confidence, Apriori algorithm, CARMA |
| Week 6 | **CLASSIFICATION** | Basic concepts , regression analysis, decision trees, |
| Week 7 | Cat one | |
| Week 8 | **CLASSIFICATION** | Naïve Bayes classification method, Decision tree, KNN, Support vector machines, rule based classification |
| Week 9 | **CLUSTER ANALYSIS** | Basic concepts, partitioning method , hierarchical method , K-means, density-based and grid-based |
| Week 10 | **SYSTEM EVALUATION** | metrics to measure the quality of data mining algorithms. Precission, accuracy, recall, F1score |
| Week 10 | **OUTLIER DETECTION** | Outliers and outlier analysis, types of outlier , challenges of outlier detection method , outlier detection method |
| Week 11 | **MINING TEXT DATA AND WORLD WIDE WEB** | MINING TEXT DATA , Basic Measures for Text Retrieval, Mining World Wide Web |
| Week 11 | **DATA MINING APPLICATIONS** | Financial Data Analysis and Retail Industry , Telecommunication Industry  , Biological Data Analysis  , Other Scientific Applications |
| Week 12 | CAT TWO | |
| Week 13 | **DATA MINING AND SOCIETY** | Ubiquitous and Invisible Data mining , Privacy and Security in Data Mining, Social Impacts of Data Mining |
| Week 14 | REVISION | |
| WEEK 15& 16 | EXAMINATIONS | |

1. **Mode of delivery**

* Discussions and working out problems
* Lectures 2 hours per week
* Hands-on laboratory exercises 3 hours per week
* E-learning

1. **Instruction Materials and/or equipment**

* Computers
* White board
* Projector

1. **Course Assessment**

* Assignments, tutorials and practical exercises (30 %)
* Written examination (70%)

1. **Core Reading Materials for the Course**

* Kantardzic, M. (2011). *Data mining: concepts, models, methods, and algorithms*. John Wiley & Sons.Jennifer Niederst Robbins (2012)

1. **Recommended Reference Materials**

* Daniel, T. (2006). Data mining methods and models.
* Džeroski, S. (2009). Relational data mining. In *Data Mining and Knowledge Discovery Handbook* (pp. 887-911). Springer US.
* Kantardzic, M. (2011). *Data mining: concepts, models, methods, and algorithms*. John Wiley & Sons.Jennifer Niederst Robbins (2012)
* Sharda, R., Delen, D., & Turban, E. (2013). *Business intelligence: a managerial perspective on analytics*. Prentice Hall Press.
* <http://www.tutorialspoint.com/data_mining/dm_quick_guide.htm>