

UNIVERSITY OF DELTA, AGBOR, NIGERIA
COMPUTING
B.Sc. Information and Communication Technology

UNIDEL-ICT 304: Data Warehouse and Data Mining (3 Units; Compulsory; LH=30; PH=45)

Senate-approved Relevance

The training of high-skilled graduates who are conversant with the current and future issues, chances, and solutions using data mining and data warehouse as well be able to use them to performance analysis for businesses which will ultimately lead to good business decision making process by these organizations in Delta State and Nigeria is in tandem with the vision and mission of the University of Delta, Agbor. This ensures that Information and Communication Technology graduates with demonstrable potentials and necessary skill-set to attend to trending issues in data collection, reporting and analysis of businesses using cloud-based technologies. The relevance of this is seeing and producing ICT graduates of the University of Delta, Agbor being versed in using appropriate data mining algorithms and techniques to generate good business insights which could guide business owners and managers in making good business decision and ensure higher profitability and better business management in Nigeria.

Overview

Data warehousing and data mining are technologies that delivers critical and optimally useful information to facilitate performance analysis of business organizations. These technologies are not an emerging trend in information technology but also a burgeoning market in a range of industries as there are rapid changing corporate environment where organizations are turning to cloud-based technologies for convenient data collection, reporting and analysis. Data warehouse and data mining are twin core component of business intelligence that enables businesses to enhance their performance.

This course will cover fundamentals of data mining and data warehousing. Topics will range from statistics to machine learning to database, with a focus on analysis of data sets exhibiting different distributions. The course is to introduce students to the concepts of data mining and data warehousing. Other topics that will be discussed include the data mining problems, application of data mining, commercial tools of data mining, knowledge discovery, architecture of data warehousing, data marts, data warehousing lifecycle, data modelling, building of data warehouse, OLAP, MOLAP, ROLAP, data warehousing and future views. Most recent issues are open for discussion.

Objectives

The objectives of this course are to: (i) provide students with in-depth knowledge, skills and understanding in the areas of Data Mining and Data Warehousing and a range of techniques, conceptual models and tools to develop into professionals in the areas of ‘Data, Information and Knowledge Management’, data mining approaches such as clustering, classification, regression etc. and its applicability in a wide range of application areas.(ii) provide students with high-level operational skills in the use of state-of the art software for KD/DM and DW/DSS, based on understanding of basic principles and the use of real-world case studies(iii) provide students with

independent exploratory and research skills, linked with abilities to synthesize, integrate and critically analyses and compare features of the Knowledge Discovery/Data mining/Business Intelligence/Data Warehousing area (iv) Be familiar with mathematical foundations of data mining tools. (v) Understand and implement classical models and algorithms in data warehouses and data mining (vi) Characterize the kinds of patterns that can be discovered by association rule mining, classification and clustering. (vii) Master data mining techniques in various applications like social, scientific and environmental context. (viii) Develop skill in selecting the appropriate data mining algorithm for solving practical problems

Learning Outcomes

Upon completion of this course, students should be able to: (i) describe the underlying concepts of data warehousing and mining. Understand the Data Mining Process and implement data mining process based solutions. (ii) find valid patterns in test data using data mining experiments with test data (iii) Understand the functionality of the various data mining and data warehousing component Knowledge (iv) Appreciate the strengths and limitations of various data mining and data warehousing models (v) Explain the analyzing techniques of various data (vi) Describe different methodologies used in data mining and data ware housing. (vii) Compare different approaches of data ware housing and data mining with various technologies

Course Contents

Introduction to Data warehouse. Differences between operational database systems and data warehouse. Data warehouse characteristics. Data warehouse architecture and its components. Extraction-Transformation-Loading. Logical (Mult-Dimensional). Data Modeling. Schema Design. Star and snow-Flake Schema. OLAP cube. OLAP Operations. OLAP Server Architecture-ROLAP, MOLAP and HOLAP. Data Mining Definition. KDD Challenges. Data Mining Tasks. Data Preprocessing (Data Cleaning. Missing Data Dimensionality Reduction, Feature Subset Selection. Discretization and Binarization. Data Transformation. Measures of similarity and dissimilarity-Basics). Problem Definition of Association Rules. Frequent Item Set Generation. The APRIORI Principle. Support and Confidence Measures. Association Rule Generation. APRIORI Algorithm. The Partition Algorithms. FP-Growth Algorithms. Compact Representation of Frequent Item Set-Maximal Frequent Item Set. Closed Frequent Item Set. Problem definition of Classification. General Approaches to solving a classification problem. Evaluation of Classifiers. Classification techniques. Decision trees-Decision Tree Construction. Methods for expressing attribute test conditions. Measures for Selecting the Best split. Algorithm for Decision tree Induction. Naïve-Bayes Classifier. Bayesian Belief Networks. K-nearest neighbor classification-Algorithm and characteristics. Problem Definition. Clustering overview. Evaluation of clustering algorithms. Partitioning clustering K-Means Algorithm. K-Means Additional Issues. PAM Algorithm. Hierarchical Clustering-Algorithm- Agglomerative Methods and Divisive Methods. Basic Agglomerative Hierarchical Clustering Algorithm. Specific techniques. Key Issues in Hierarchical Clustering. Strengths and weakness. Outlier Detection

Minimum Academic Standard:

NUC minimum academic standard requirements for facilities.