Data Mining and Business Intelligence

ITA5007

L T P J C

3 0 0 4 4

PROF. E.P.EPHZIBAH

Course Objectives:

1. To learn and apply appropriate data pre-processing techniques.

2. To learn data mining algorithms and significance.

3. To learn to apply appropriate predictive and descriptive mining algorithms for business intelligence

Expected Course Outcomes:

1. Understand the distribution of data and its type to proceed with the data pre-processing and mining.
2. Apply data summarization and appropriate pre-processing techniques as per the requirement of the data mining task.
3. Understand and incorporate the statistical models behind the prediction process.
4. Apply various representations of classification models and evaluate the performance.
5. Identify the appropriate association data mining techniques to improvise business applications.
6. Implement the clustering techniques and apply them in real-time business applications.
7. Use previously observed values to evaluate and interpret future results.

Module:1 Introduction

Data Mining(DM)–origin–rapid growth

Core Ideas in Data Mining

Supervised and Unsupervised Learning

Steps in Data Mining

Data Warehousing

Business Intelligence(BI)

Role of the mathematical model

Business Intelligent Architecture

Development of business intelligent system.

Module:2 Dimension Reduction

Data Summaries

Correlation Analysis

Reducing the Number of Categories in Categorical Variables

Converting a Categorical Variable to a Numerical Variable

Principal Component Analysis -PCA

Module:3 Performance Evaluation and prediction

Evaluating Classification and Predictive Performance

Introduction

Judging Classification Performance

Evaluating Predictive Performance

Prediction - Multiple linear regression

Explanatory vs predictive modeling

Estimating the regression equation and prediction

Variable selection in linear regression.

Module:4 Classifications

Classification methods

Naïve Bayes

K-Nearest-Neighbors

Classification and regression trees

Logistic regression models

Evaluating classification performance

Evaluating the Goodness of fit

Logistic regression for more than two classes

Module:5 Discriminant Analysis and Association Rules

Discriminant analysis

Classification performance of discriminant analysis

Prior probabilities

Unequal classification costs

Classifying more than two classes

Association Rules: Introduction

Discovering Association Rules in Transaction Databases

Generating Candidate Rules

Selecting Strong Rules.

Module:6 Cluster Analysis

Cluster analysis –Introduction

Distance between two records

Measuring distance between two clusters

Hierarchical clustering

Non-hierarchical clustering

k-means algorithm

Module:7 Forecasting Time Series

Introduction to time series

Explanatory versus Predictive Modelling

Popular Forecasting Methods in Business

Time Series Components

Data Partitioning

Regression-Based Forecasting

Model with Trend

Model with Seasonality

Model with Trend and Seasonality

Autocorrelation and ARIMA Models

Smoothing Methods

Books

- Text Books

GalitShmueli, Peter C. Bruce, Nitin R. Patel. Data Mining for Business Analytics: Concepts, Techniques and Applications in XL Miner, 2010, 2ndEdition,Wiley Publications

https://doc.lagout.org/Others/Data%20Mining/Data%20Mining%20for%20Business%20Intelligence_%20Concepts%2C%20Techniques%2C%20and%20Applications%20in%20Microsoft%20Office%20Excel%20with%20XLMiner%20%5BShmueli%2C%20Patel%20%26%20Bruce%202010-10-26%5D.pdf

Carlo Vercellis, Business Intelligence: Data Mining and Optimization for Decision Making, 2009,1 st Edition, Wiley Publications.

http://www.biomedicahelp.altervista.org/Magistrale/Clinics/BIC_PrimoAnno/IdentificazioneModelliDataMining/Business%20Intelligence%20-%20Carlo%20Vercellis.pdf

- Reference Books

JiaweiHan, Micheline, Jian Pei. Data Mining: Concepts and Techniques, 2011, 3rd Edition, The Morgan Kaufmann Series.

http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf

Margaret. H. Dunham, Data Mining: Introductory and Advanced Topics, 2006, 1 st Edition, Pearson Education.

https://theswissbay.ch/pdf/Gentoomen%20Library/Data%20Mining/Dunham%20-%20Data%20Mining.pdf