# FUTURE VISION BIE

## One Stop for All Study Materials

## & Lab Programs

Future Vision

By K B Hemanth Raj

Scan the QR Code to Visit the Web Page

Or

**Visit : https://hemanthrajhemu.github.io**

Gain Access to All Study Materials according to VTU,
CSE – Computer Science Engineering,
ISE – Information Science Engineering,
ECE - Electronics and Communication Engineering
& MORE...

**Join Telegram to get Instant Updates: https://bit.ly/VTU_TELEGRAM**

**Contact: MAIL: futurevisionbie@gmail.com**

**INSTAGRAM: www.instagram.com/hemanthraj_hemu/**

**INSTAGRAM: www.instagram.com/futurevisionbie/**

**WHATSAPP SHARE: https://bit.ly/FVBIESHARE**

**DATA MINING AND DATA WAREHOUSING**
**[As per Choice Based Credit System (CBCS) scheme]**
**(Effective from the academic year 2017 - 2018)**
**SEMESTER – VI**

| | | | |
|---|---|---|---|
| Subject Code | 17CS651 | IA Marks | 40 |
| Number of Lecture Hours/Week | 3 | Exam Marks | 60 |
| Total Number of Lecture Hours | 40 | Exam Hours | 03 |

**CREDITS – 03**

| **Module – 1** | **Teaching Hours** |
|---|---|
| **Data Warehousing&modeling:** Basic Concepts: Data Warehousing: A multitier Architecture, Data warehouse models: Enterprise warehouse,Data mart and virtual warehouse, Extraction, Transformation and loading, Data Cube: A multidimensional data model, Stars, Snowflakes and Fact constellations: Schemas for multidimensional Data models, Dimensions: The role of concept Hierarchies, Measures: Their Categorization and computation, Typical OLAP Operations. | **8 Hours** |
| **Module – 2** | |
| **Data warehouse implementation& Data mining:**Efficient Data Cube computation: An overview, Indexing OLAP Data: Bitmap index and join index, Efficient processing of OLAP Queries, OLAP server Architecture ROLAP versus MOLAP Versus HOLAP.**:** Introduction: What is data mining, Challenges, Data Mining Tasks, Data: Types of Data, Data Quality, Data Preprocessing, Measures of Similarity and Dissimilarity, | **8 Hours** |
| **Module – 3** | |
| **Association Analysis:** Association Analysis: Problem Definition, Frequent Item set Generation, Rule generation. Alternative Methods for Generating Frequent Item sets, FP-Growth Algorithm, Evaluation of Association Patterns. | **8 Hours** |
| **Module – 4** | |
| **Classification :**Decision Trees Induction,Method for Comparing Classifiers, Rule Based Classifiers, Nearest Neighbor Classifiers,Bayesian Classifiers. | **8 Hours** |
| **Module – 5** | |
| **Clustering Analysis:** Overview, K-Means, Agglomerative Hierarchical Clustering, DBSCAN, Cluster Evaluation, Density-Based Clustering, Graph-Based Clustering, Scalable Clustering Algorithms. | **8 Hours** |

**Course outcomes:** The students should be able to:

- Understands data mining problems and implement the data warehouse
- Demonstrate the association rules for a given data pattern.
- Discuss between classification and clustering solution.

**Question paper pattern:**
The question paper will have TEN questions.
There will be TWO questions from each module.
Each question will have questions covering all the topics under a module.
The students will have to answer FIVE full questions, selecting ONE full question from each module.

**Text Books:**

1. Pang-Ning Tan, Michael Steinbach, Vipin Kumar: Introduction to Data Mining,

Pearson, First impression,2014.
2. Jiawei Han, MichelineKamber, Jian Pei: Data Mining -Concepts and Techniques, 3rd Edition,Morgan Kaufmann Publisher, 2012.

**Reference Books:**

1. Sam Anahory, Dennis Murray: Data Warehousing in the Real World, Pearson,Tenth Impression,2012.
2. Michael.J.Berry,Gordon.S.Linoff: Mastering Data Mining , Wiley Edition, second edtion,2012.