|  |  |
| --- | --- |
| MONO | **DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING****NATIONAL INSTITUTE OF TECHNOLOGY PATNA** Ashok Raj Path, PATNA 800 005 (Bihar), India |
|  Phone No.: 0612 – 2372715, 2370419, 2370843, 2371929, 2371930, 2371715 Fax – 0612- 2670631 Website: [www.nitp.ac.in](http://www.nitp.ac.in/) |

***CSX456 Advanced Data Mining***

**L-T-P-Cr:3-0-0-3**

**Pre-requisite:** Fundamental knowledge on Frequent Pattern Mining,Classification and clustering algorithms.

**Objectives/Overview:**

* Advanced concepts and technologies of mining association rules, cluster analysis, stream data mining, time series data mining, sequence pattern mining, text mining and web mining.

**Course Outcomes:**

At the end of the course, a student should able to:

|  |  |  |
| --- | --- | --- |
| **S.No** | **Course Outcome** |  **Mapping to POs** |
| 1 | Analyze Algorithms for sequential patterns | PO1, PO2, PO3 |
| 2 | Determine patterns from time series data | PO1, PO2, PO3 |
| 3 | Understand Data Stream Mining models and methods  | PO1, PO2, PO3 |
| 4 | Distinguish computing frameworks for Big Data analytics. | PO1, PO6, PO12 |
| 5 | Apply Graph mining algorithms to Web Mining | PO1, PO2, PO3, PO6, PO12 |

**UNIT I: Sequential Pattern Mining Lectures: 10**

Concepts, primitives, scalable methods for mining sequential patterns – GSP, SPADE, PrefixSpan, Mining Closed Sequential Patterns, Mining Multidimensional and Multilevel sequential Patterns.

**UNIT II: Mining Time series Data Lectures: 04**

Periodicity Analysis for Time-Related Sequence Data, Trend analysis, Similarity search in Time-series analysis;

**UNIT III: Mining Data Streams Lectures: 8**

Methodologies for stream data processing and stream data systems, Frequent pattern mining in stream data, Sequential Pattern Mining in Data Streams, Classification of dynamic data streams, Class Imbalance Problem;

**UNIT IV: Graph Mining Lectures: 8**

Mining frequent subgraphs, finding clusters, hub and outliers in large graphs, Graph Partitioning; Web Mining, Mining the web page layout structure, mining web link structure, mining multimedia data on the web, Automatic classification of web documents and web usage mining.

**UNIT V: Distributed Data Mining Lectures: 8**

Distribute data mining framework, Distributed data source, Distributed data mining techniques, Distributed classifier learning, distributed clustering, distributed association rule mining and Challenges of distributed data mining;

**Text/Reference Books:**

1. Data Mining Concepts and Techniques.2nd Edition, Elsevier, 2011. J Han and M Kamber.
2. Introduction to Data Mining. Addision Wesley,2006, Pang Ning Tan, M Steinbach, Vipin Kumar,
3. Sequence Data Mining. Springer, 2007, G Dong and J Pei