## LIS - 525: Data Mining and Knowledge Discovery

**(4 Credit/100 marks)**

**Course Description:**

This course aims to provide search tools for data requirements with precision. Students will be able to analyse various sources of information and derive new content through data mining processes.

**Course Outcome:**

Through this course the students will utilise various tools in searching information from different textual sources which will lead to generation of new concepts and information.

**Course Objectives:**

To introduce the fundamental processes of text mining, data warehousing and data mining.

To impart knowledge on various data mining concepts and techniques that can be applied to text mining, web mining etc.

To develop the knowledge for application of data mining for information retrieval from the web.

**Course Outline**

**Unit I: Text Mining** **20 Hours**

Text Mining: Definitions, Process, Techniques and Issues, Text Mining Approaches

Document classification (text classification, document standardisation), Information retrieval (keyword search / querying and indexing), Document clustering (phrase clustering), Natural Language Processing (Spelling correction, lemmatization, grammatical parsing, and word sense disambiguation), Text Summarization, Information extraction (relationship extraction / link analysis), and Web mining (web link analysis)

Applications: Digital Libraries, Academic and Research Field, Life Science, Social media, Business Intelligence.

**Unit II: Data Mining** **20 Hours**

Data Mining overview, Architecture, Process, Classification of Data Mining Systems, Issues with Data Mining

Data Warehouse, Data Warehouse Models, Metadata Repository, Data Pre-processing – Data Integration and Transformation, Data Reduction, Data Mining, Methodologies of Data Mining, Data Mining Applications, Data Mining and Society.

Web Mining: Concepts,Web Content Mining, Web Usage Mining, Web Structure Mining, Mining Tools, Applications.

**Unit III: Big Data** **20 Hours**

Big Data: History of Big Data, Its Phases, Characteristics of Big Data, Big Data Tools

Big Data challenges and Issues, Types of Big Data- Structured Data, Unstructured Data. Semi-Structured Data.
Knowledge Discovery in Databases (KDD): Knowledge Discovery - Introduction, Concepts. Process of Knowledge Discovery, KDD Research Opportunities, Challenges and Trends.

Tools and Techniques in Knowledge Discovery in Databases.

**References/ Readings:**


1. Acharya, S. C. (2019). *Big Data and Analytics.* New Delhi: Wiley.
2. Agarwal, C. (May 2015). *Data Mining: The Textbook.* Springer Nature.
3. Bhatia, P. (2019). *Data Mining and Data Warehousing: Principles and Practical Techniques.* New Delhi: Cambridge University Press.
4. Erl, T., Khattak, W., & Buhler, P. (2016). *Big Data Fundamentals: Concepts Drivers: Concepts, Drivers and Techniques.* Noida Uttar Pradesh: Pearson Education India.
5. Han, J. Kamber , M., & Pei, J. (2012). *Data Mining: Concepts and Techniques.* Morgan Kaufmann.
6. Kamal, R., & Saxena, P. (2019). *Big Data Analytics, Introduction to Hadoop, Spark, and Machine-Learning.* New Delhi: McGraw Hill Education.
7. Liu, B. (2011). *Web Data Mining.* Berlin: Springer.
8. Russell, M. A., & Klassen, M. (2019). *Mining the Social Web* (3rd. ed.). India: O'Reilly Media, Inc.
9. Tan, P. N., Steinbach, Michael, & Kumar, V. (2016). *Introduction to Data Mining.* Noida: Pearson India Pvt. Ltd.
10. Taneja, A. (2012). *Knowledge Discovery in Databases.* New Delhi: Galgotia Publications.