COURSE TITLE	**WEB INTELLIGENCE AND MINING**
COURSE CODE	**01AI0602**
COURSE CREDITS	**4**

Objective:

1 Student will study how to search and gather the data in the web and process as well as mine that data to extract meaningful information to understand its application to search engines. Students will gain an understanding of the strategic and operational aspects of Web analytics. The course will cover theory as well as practice aspects of a subject through scheduled lectures and labs. Students will also understand the basic concepts behind information retrieval and data mining.

Course Outcomes: After completion of this course, student will be able to:

1 Understand the concepts of web, web mining, web intelligence and analytics. (Understand)

2 Learn and implement web Mining using web intelligent algorithms. (Apply)

3 Learn and apply how to retrieve information from web. (Apply)

4 Learn how to use and deploy web/social analytics platforms such as Google Analytics. (Apply)

Pre-requisite of course: Student will study how to search and gather the data in the web and process as well as mine that data to extract meaningfulinformation to understand its application to search engines. Students will gain an understanding of the strategic and operational aspects of Web analytics. The course will cover theory as well as practice aspects of a subject through scheduled lectures and labs. Students will also understand the basic concepts behind information retrieval and data mining

Teaching and Examination Scheme

Theory Hours	Tutorial Hours	Practical Hours	ESE	IA	CSE	Viva	Term Work
3	0	2	50	30	20	25	25

Contents : Unit	Topics	Contact Hours
1	**Introduction to Web** Introduction to Internet, web, blogs, tweets, wikis, grid, and cloud, Collaborative mapping, Components of typical web, Characteristics and benefits of the Web	
2	**Web Intelligence** Semantic web, Social intelligence, Search engine techniques, Web information retrieval and filtering, Levels of WI, Goal of WI, Characteristics of web intelligence, Challenges and issues of WI, Future of WI	

DR. MADHU SHAMBHU SHUKLA
Digitally signed by (Name of HOD)

DR. RAJENDRASINH BAHADURSINH JADEJA
Digitally signed by (Name of Dean/ Principal)

Contents : Unit	Topics	Contact Hours
3	**Web Information Retrieval:** Managing web data. Web search engines, Google searching, Introduction to web crawler, Architecture of a web crawler, Distributed crawling, Focused spiders/crawlers, Collaborative crawling, Some tools and open source for web crawling, Models of information retrieval	
4	**Web mining** Introduction to data mining, Classification & clustering, Pattern recognition, Introduction to web mining, Evolution, Process, Web content mining, Web usage mining, Web structure mining	
5	**Intelligent Web Algorithms** The intelligent-algorithm lifecycle, Classes of intelligent algorithm, Recommendation engines based on users, items, and content	
6	**Web Analytics** How Web Analytics Works – Basic Concepts, collection of Web Data and other types of data, basic dashboards, Predictive Analytics, Web Analytics Ecosystem and Tools, Data Visualization, Acquisition and Conversions	
Total Hours		

Suggested List of Experiments:

Contents : Unit	Topics	Contact Hours
1	**Experiments** 1. Understanding XML structure and DTD, create files in XML format and view with plugins in any browser., 2. Report on comparative study of web search engines. Use advanced search with google, yahoo, bing and try to learn where personalized results are obtained, 3. Learn about scrapy library and scan any one website of online business, 4. Learn about sklearn library and apply decision tree algorithm on web data., 5. Create a report any 3 browser plugins for web data analysis., 6. Create google analytics account by registering yourself with a blogpost, 7. Create a report for steps or methods to find if a website is fake or not., 8. Create a report on functionalities avalaible in youtube studio analytics., 9. Create python script to find domain name from IP address	
Total Hours		

Textbook :

1 Intelligent Technologies for Web Applications, Priti Srinivas Sajja, Rajendra Akerkar, , CRC Press - Taylor & Francis Group, 2011

References:

1 Intelligent Technologies for Web Applications - Priti Srinivas Sajja, Rajendra Akerkar, CRC Press - Taylor & Francis Group

DR. MADHU SHAMBHU SHUKLA
Digitally signed by (Name of HOD)

DR. RAJENDRASINH BAHADURSINH JADEJA
Digitally signed by (Name of Dean/ Principal)

References:

2 Algorithms of the Intelligent Web, Second Edition - Douglas G. McIlwraith, HaralambosMarmanis, and Dmitry Babenko, Manning Publications

3 Mining the Social Web, Third Edition, Matthew A. Russell and Mikhail Klassen, Published by O'Reilly Media

4 Data Mining Practical Machine Learning Tools and Techniques, Fourth Edition, Ian Witten, Eibe Frank, Mark Hall, Christopher Pal, Elsevier

Suggested Theory Distribution:

The suggested theory distribution as per Bloom's taxonomy is as follows. This distribution serves as guidelines for teachers and students to achieve effective teaching-learning process

Distribution of Theory for course delivery					
Remember / Knowledge	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Higher order Thinking / Creative**
10.00	30.00	35.00	15.00	10.00	0.00

Instructional Method:

1 The course delivery method will depend upon the requirement of content and need of students. The teacher in addition to conventional teaching method by black board, may also use any of tools such as demonstration, role play, Quiz, brainstorming, MOOCs etc

2 The internal evaluation will be done on the basis of continuous evaluation of students in the laboratory and class-room

3 Practical examination will be conducted at the end of semester for evaluation of performance of students in laboratory.

4 Students will use supplementary resources such as online videos, NPTEL videos, e-courses, Virtual Laboratory

Supplementary Resources:

1 https://nptel.ac.in/courses/106/105/106105174/

2 https://nptel.ac.in/courses/106/105/106105078/

DR. MADHU SHAMBHU SHUKLA
Digitally signed by (Name of HOD)

DR. RAJENDRASINH BAHADURSINH JADEJA
Digitally signed by (Name of Dean/ Principal)