**Mukesh Patel School of Technology Management and Engineering**

**Computer Engineering Department**

**Course Policy**

|  |  |  |
| --- | --- | --- |
| **Program/Branch/Semester** | **:** | B. Tech (Computer Science and Business Systems) /COMP/ Sem VII |
| **Academic Year** | **:** | 2022-23 |
| **Course Code & Name** | **:** | Advanced Social, Text and Media Analytics (Elective IV) |
| **Credit Details** | **:** | |  |  |  |  | | --- | --- | --- | --- | | L | T | P | C | | 3 | 0 | 2 | 3 | |
| **Course Faculty 1** | **:** | Prof. Abhishek Vichare  022-49330576  [abhishek.vichare@nmims.edu](mailto:abhishek.vichare@nmims.edu) /  Office: 2nd Floor, SBMP Building |
| **Course Faculty 2** |  | Prof. Chandravadan Prajapati  chandravadanp@live.com |
| **Course Faculty 3** |  | Ms. A. Jain  Avani.Joshi@nmims.edu |
| **Student Contact Hrs** | **:** | 1pm-2 pm (Monday,Wed,Thursday) |
|  |  |  |
| ***Queries by Emails are encouraged.*** | | |
| **Course link** | **:** | Portal Link |

# Introduction to the Course

## Importance of the course

Social media analytics refers to the approach of collecting data from social media sites and evaluating that data to make business decisions. This process consists the usual monitoring , a basic analysis of sentiments to develop an idea of the social consumer. Social media not only provides a means of communicating with their customers, but also a way to better understand their customers. Social media analytics entails collecting statistics from media platforms and evaluating the data gotten from the statistics. These data could be collected using the recommended social media analytics tools to inform marketing assessments.

## Objective of the Course

## The objective of this course is to impart knowledge of Application Programming Interface (API) services to collect data and analyze structures or unstructured data - primarily textual comments - for sentiments expressed in them. To Use different tools for collecting, analyzing, and exploring social media data for research and development purposes.

1.3 Pre-requisite

* Data Mining, Machine Learning, Python Programming

# Course Outcomes (CO) and mapping with Program Outcomes (PO)

## Course Outcomes

After successful completion of the course, a student will be able to-

## 1. Understand various techniques for Text Mining and carry out Pattern Discovery, Predictive Modelling

## 2. Explore the use of social network analysis to understand the growing connectivity and complexity in the world around us on different scales – ranging from small groups to the World Wide Web

## 3. Develop social network analysis model to identify important social actors, subgroups (i.e., clusters), and network properties in social media sites such as Twitter, Facebook, and YouTube

## CO-PO Mapping

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | PO1 | PO2 | PO3 | PO4 | PO5 | PO6 | PO7 | PO8 | PO9 | PO10 | PO11 | PO12 |
| CO1 | M | H |  |  | M | H |  |  |  |  |  |  |
| CO2 | H |  |  |  |  |  |  |  |  |  |  |  |
| CO3 | M |  |  |  | H |  |  |  |  |  |  |  |

***Green- medium mapping Blue- high mapping***

***3.*** Syllabus, Pre-class activity and References

## Teaching and evaluation scheme

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Teaching Scheme** | | | | **Evaluation Scheme** | |
| **Lecture**  **Hours per week** | **Practical**  **Hours per week** | **Tutorial**  **Hours per week** | **Credit** | **Internal Continuous Assessment**  **(ICA)**  **(Marks -50)** | **Term End Examinations (TEE)**  **(Marks -100 in Question Paper)** |
| 3 | 2 | 0 | 4 | Marks Scaled to 50 | Marks Scaled to 50 |

## 3.2 Syllabus

|  |  |  |
| --- | --- | --- |
| **Unit** | **Description** | **Duration** |
| **1** | Text Mining:  Introduction, Core text mining operations, Pre-processing techniques, Categorization, Clustering, Information extraction, Probabilistic models for information extraction, Text mining applications | **9** |
| Methods & Approaches:  Content Analysis; Natural Language Processing; Clustering & Topic Detection; Simple Predictive Modeling; Sentiment Analysis; Sentiment Prediction | **12** |
| **2** | Web Analytics:  Web analytics tools, Clickstream analysis, A/B testing, online surveys; Web search and retrieval, Search engine optimization, Web crawling and Indexing, Ranking algorithms, Web traffic models | **12** |
| **3** | Social Media Analytics:  Social network and web data and methods. Graphs and Matrices. Basic measures for individuals and networks. Information visualization; Making connections: Link analysis. Random graphs and network evolution. Social contexts: Affiliation and identity; Social network analysis | **12** |
|  | **Total hours** | **45** |

## Pre-class activity

Outline for preliminary study to be done for each unit will be provided prior to commencement of each unit. Preliminary study material (video links, presentation, notes etc) will be made available on the student portal. Students are expected to go through this material before attending the upcoming session. It is expected that the students put in at least two hours of self-study for every one hour of classroom teaching. During the lecture session, more emphasis will be given on in-depth topics, practical applications and doubt solving.

## References

|  |
| --- |
| **Text Books:**   1. Ronen Feldman and James Sanger, “The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data”, Cambridge University Press, 2006. 2. Hansen, Derek, Ben Sheiderman, Marc Smith. 2011 Analyzing Social Media Networks with NodeXL: Insights from a Connected World, Morgan Kaufmann, 304 3. Avinash Kaushik. 2009. Web Analytics 2.0: The Art of Online Accountability.   Hanneman, Robert and Mark Riddle. 2005. Introduction to Social Network Method |
| **Reference Books:**   1. Wasserman, S. & Faust, K. (1994). Social network analysis: Methods and applications. New York: Cambridge University Press. 2. Monge, P. R. & Contractor, N. S. (2003). Theories of communication networks. New York: Oxford University Press. http://nosh.northwestern.edu/vita.htm |

# Laboratory details

Knowledge of Python, Java programming for laboratory exercise is a prerequisite. Students are expected to recall the fundamental theory concepts relevant to the exercise to be performed in the upcoming laboratory.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sr. No.** | **Title** | **Prerequisite** | **CO mapping** | **Deadline** |
| 1 | Introduction to social media, text analysis | Knowledge of media and text analysis | CO1 | 2nd week |
| 2 | To explore /practice NLTK tool kit | Machine Learning, Python Programming | CO1 | 3rd week |
| 3 | To implement information extraction | Machine Learning, Python Programming | CO1 | 4th Week |
| 4 | To implement data tokenization, data stemming using NLP | Machine Learning, Python Programming | CO1 | 5th Week |
| 5 | To implement data lemmatization, part of speech using NLP | Machine Learning, Python Programming | CO1 | 6th Week |
| 6 | To implement sentiment analysis using NLP | Machine Learning, Python Programming | CO1 | 7th Week |
| 7 | To explore/practice anyone of the web analytics tool for given scenario | Machine Learning, Python Programming | CO2 | 8th Week |
| 8 | To implement social media content extraction using API | Machine Learning, Python Programming | CO2 | 9th Week |
| 9 | To write program to measure social media network properties | Machine Learning, Python Programming | CO3 | 10th week |
| 10 | To install and explore network analysis tool | Machine Learning, Python Programming | CO3 | 11th week |
| 11 | Presentation |  |  | 12th week |
| 12 | Presentation |  |  | 13th week |
| 13 | Viva |  |  | 14th week |
| 14 | Doubt solving session |  |  | 15th week |
| 15 | Submission |  |  | 15th Week |

# Assessment Policy

## Component wise Continuous Evaluation Internal Continuous Assessment (ICA) and Term End Examination (TEE)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Assessment Component** | **ICA (100 Marks)**  **(Marks scaled to 50)** | | | | | | **TEE (100 marks)**  **(Marks scaled to 50)** |
|  | **Lab Performance** | **Viva** | **Presentation (**on recent trends in subject ) | **Class Assignment** | **Class Test1 and Class Test 2** | **Class Participation** |  |
| **Weightage** | 10% | 5% | 5% | 5% | 20% | 5% | 50% |
| **Marks** | 20 | 10 | 10 | 10 | 20+20 | 10 | 100 |

## Assessment Policy for Internal Continuous Assessment (ICA)

Assessment of ICA comprises of the following components.

1. **Class test 1 and 2 (20+20 Marks)**
   1. Two class tests will be conducted as per the academic calendar.
   2. It may be conducted online/ offline for 20 marks each

1. **Lab experiments (20 marks)**
   * 1. Continuous assessment for laboratory experiments will be conducted. There are 10 practical, each carrying weightage of 10 marks. At the end of the course, average of total marks will be taken to obtain marks out of 10.
     2. Discussion of your work with your peers is allowed. However each student is expected to submit his/her original work. Submissions which are very similar will be marked zero. Assessment of the lab work will be carried out based on parameters like timely completion of lab work file, understanding of the experiment performed, originality in the work, involvement of the student, regularity, discipline etc. during the session. There is a 30% penalty on late submission.
2. **Viva (10 marks) –** Viva will be conducted for 10 marks.
3. **Class Participation (10 marks)-** The faculty will ask some questions in every class based on the content being taught. The question could be asked to a chosen student or a student group (which is formed at the beginning of the semester). One mark can be given to the correct answer. The idea is to encourage students to pay attention in class and actively participate. These marks will be added in ICA class participation component.
4. **Presentation (10 marks):** Students need to select the topic (after faculty approval) of presentation w.r.t current trends in subject.
5. **Assignments (10 marks):** Minimum 2 assignments will be given in the class.

Lab and Assignment Rubrics are tabulated below:

|  |  |  |
| --- | --- | --- |
| 1. Preparedness and Efforts | | |
| 3: Well prepared and puts efforts | 2: Not prepared but puts efforts or prepared but doesn't put efforts | 1: Neither prepared nor puts efforts |
| 2. Knowledge of concepts and tools | | |
| 3: Good knowledge | 2: Partial knowledge | 1: Minimal knowledge |
| 3. Debugging and results | | |
| 3: Gets accurate results based on calculations | 2: Calculates but doesn't get proper result or gets result but with the help of faculty in-charge | 1: Doesn’t get results at all |
| 4. Punctuality | | |
| 3: Gets all the experiments checked in-time and is always in-time to the lab sessions | 2: Some time delays the experiment checking or is late to the lab sessions for few times | 1: Most of the time delays experiment checking and / or comes late for lab sessions |
| 5. Lab Ethics | | |
| 3: Follows proper lab ethics by keeping the lab clean, properly handling the components & devices and placing things back at their right place | 2: Sometimes doesn't follow the lab ethics | 1: Most of the times makes the lab untidy and keeps things at wrong place |

|  |  |  |
| --- | --- | --- |
| **1. Timely submission** | | |
| 3: If submitted on-time | 2-delayed by not more than 3 days | 1-delayed by more than 3 days (consider medical reason if any accordingly) |
| **2. Originality of the material** | | |
| 3: No copy material from other classmates. All is self written | 2: Some copy from other classmates | 1: Most of the material is copied from others |
| **3. Neatness** | | |
| 3: Very neat | 2: Just Neat enough | 1: Untidy |
| **4. Attempting all questions satisfactorily** | | |
| 3: attempted all questions satisfactorily | 2: Majority of questions attempted satisfactorily | 1: Attempted few questions satisfactorily. |
| **5. Innovative solution** | | |
| 3: Innovative solutions to problems | 2: Slight innovation solutions | 1: no innovative solutions |

## Assessment Policy for Term End Examination (TEE)

A written examination of 100 marks for 3 Hours duration will be conducted for the course as per the academic calendar.

# 7. Lesson Plan

|  |  |  |  |
| --- | --- | --- | --- |
| **Session**  **No.** | **Topics** | **Mapped CO** | **Reference** |
| 1 | Text Mining:  Introduction, | CO1 | TB1, RB1 |
| 2 | Core text mining operations, | CO1 | TB1, RB1 |
| 3 | Pre-processing techniques, | CO1 | TB1, RB1 |
| 4 | Categorization, | CO1 | TB1, RB1 |
| 5 | Clustering, | CO1 | TB1, RB1 |
| 6 | Information extraction, | CO1 | TB1, RB1 |
| 7 | Probabilistic models for information extraction, | CO1 | TB1, RB1 |
| 8 | Text mining applications | CO1 | TB1, RB1 |
| 9 | Text mining applications | CO1 | TB1, RB1 |
| 10 | Methods & Approaches:  Content Analysis; | CO1 | TB1, RB1 |
| 11 | Natural Language Processing; | CO1 | TB1, RB1 |
| 12 | Natural Language Processing; | CO1 | TB1, RB1 |
| 13 | Clustering & Topic Detection | CO1 | TB1, RB1 |
| 14 | Test1 |  |  |
| 15 | Clustering & Topic Detection | CO1 | TB1, RB1 |
| 16 | Clustering & Topic Detection | CO1 | TB1, RB1 |
| 17 | Simple Predictive Modeling; | CO1 | TB1, RB1 |
| 18 | Simple Predictive Modeling; | CO1 | TB1, RB1 |
| 19 | Simple Predictive Modeling; | CO1 | TB1, RB1 |
| 20 | Sentiment Analysis; | CO1 | TB1, RB1 |
| 21 | Sentiment Prediction | CO1 | TB2, RB1 |
| 22 | Web Analytics:  Web analytics tools, | CO2 | TB2, RB1 |
| 23 | Web analytics tools, | CO2 | TB2, RB1 |
| 24 | Clickstream analysis, | CO2 | TB2, RB1 |
| 25 | A/B testing, | CO2 | TB2, RB1 |
| 26 | online surveys; | CO2 | TB2, RB1 |
| 27 | Test2 |  |  |
| 28 | Web search and retrieval, | CO2 | TB2, RB1 |
| 29 | Search engine optimization, | CO2 | TB2, RB1 |
| 30 | Web crawling and Indexing, | CO2 | TB2, RB1 |
| 31 | Ranking algorithms, | CO2 | TB2, RB1 |
| 32 | Web traffic models | CO2 | TB2, RB1 |
| 33 | Web traffic models | CO2 | TB2, RB1 |
| 34 | Social Media Analytics:  Social network and web data and methods. | CO3 | TB2, RB1 |
| 35 | Social network and web data and methods. | CO3 | TB2, RB1 |
| 36 | Graphs and Matrices | CO3 | TB2, RB1 |
| 37 | Graphs and Matrices | CO3 | TB2, RB1 |
| 38 | Basic measures for individuals and networks. | CO3 | TB2, RB1 |
| 39 | Information visualization; | CO3 | TB2, RB1 |
| 40 | Making connections: | CO3 | TB2, RB1 |
| 41 | Link analysis. | CO3 | TB2, RB1 |
| 42 | Random graphs and network evolution. | CO3 | TB2, RB1 |
| 43 | Social contexts: Affiliation and identity; | CO3 | TB2, RB1 |
| 44 | Social network analysis | CO3 | TB2, RB1 |
| 45 | Revision |  |  |
|  |  |  |  |

# Teaching-learning methodology

Faculty will make a group of 2-3 students for any group based activity such as class participation, project, presentation etc. Lecture and laboratory session will be conducted as follows-

1. **Lectures:** 
   * Outline for preliminary study to be done for each unit will be provided prior to commencement of each unit.
   * Deeper concepts and applications will be explained through Presentation and Video Lectures.
2. **Laboratory:**
   * Lab manual consisting of theory and algorithm to support the lab experiment will be uploaded on student portal.
   * Regular lab assessment and grading will be done. Students will be marked based on parameters like completion of lab assignment, originality, logic developed, interaction during the lab, submission, punctuality and discipline

**10. Active learning techniques**

Active learning is a method of learning in which students are actively or experientially involved in the learning process. Following active learning techniques will be adopted for the course.

1. **Blended Learning:** Students will be introduced to the topic at home while the in-depth topics, applications and numerical problems will be discussed by the faculty in the lecture session. Outline for preliminary study to be done for each unit will be provided prior to commencement of each unit. Preliminary study material (video links, presentation, notes etc) will be made available on the student portal.
2. **Brainstorming: S**tudents will be asked to generate ideas on a certain topic, category or question while the faculty will facilitate and record the answers on the blackboard/whiteboard.

**11. Course Material**

Following course material is uploaded on the student portal: [(give](https://sites.google.com/a/nirmauni.ac.in/2cs101-computer-programming/) student portal link)

* Course Policy
* Lecture Videos
* Lecture Presentations
* Books / Reference Books / NPTEL video lectures link
* Assignments
* Lab Manuals, Test images database link

**12. Course Outcome Attainment**

Following means will be used to assess attainment of course learning outcomes.

* Use of formal evaluation components of continuous evaluation, assignments, laboratory work, semester end examination
* Informal feedback during course conduction

**13. Academic Integrity Statement**

Students are expected to carry out assigned work under Internal Continuous Assessment (ICA) independently. Copying in any form is not acceptable and will invite strict disciplinary action. Evaluation of corresponding component will be affected proportionately in such cases. Plagiarism detection software will be used to check plagiarism wherever applicable. Academic integrity is expected from students in all components of course assessment.