Artificial Intelligence and Machine Learning: Sem VIII								
Course Code	**Course Name**	**Teaching Scheme (Contact Hours)**			**Credits Assigned**			
		Theory	**Practical**	**Tutorial**	**Theory**	**Practical**	**Tutorial**	**Total**
HAIMLC801	**Text, Web and Social Media Analytics**	**04**	**--**	**--**	**04**	**--**	**--**	**04**

Course Code	**Course Name**	**Examination Scheme**							
		Theory Marks				**Exam Duration**	**Term Work**	**Practical and Oral**	**Total**
		Internal Assessment			**End Sem. Exam.**				
		Test1	**Test2**	**Avg.**					
HAIMLC801	**Text, Web and Social Media Analytics**	**20**	**20**	**20**	**80**	**03**	**--**	**--**	**100**

Course Prerequisites:	
Python, Data Mining	
Course Objectives: The course aims	
1	To have a strong foundation on text, web and social media analytics.
2	To understand the complexities of extracting the text from different data sources and analysing it.
3	To enable students to solve complex real-world problems using sentiment analysis and Recommendation systems.
Course Outcomes:	
After successful completion of the course, the student will be able to:	
1	Extract Information from the text and perform data pre-processing
2	Apply clustering and classification algorithms on textual data and perform prediction.
3	Apply various web mining techniques to perform mining, searching and spamming of web data.
4	Provide solutions to the emerging problems with social media using behaviour analytics and Recommendation systems.
5	Apply machine learning techniques to perform Sentiment Analysis on data from social media.

Module		**Topics**	**Hours.**
1.0		**Introduction**	**06**
	1.1	**Introduction to Text Mining:** Introduction, Algorithms for Text Mining, Future Directions	
	1.2	**Information Extraction from Text**: Named Entity Recognition, Relation Extraction, Unsupervised Information Extraction	
	1.3	**Text Representation:** tokenization, stemming, stop words, NER, N-gram modelling	
2.0		Clustering and Classification	10

	2.1	**Text Clustering**: Feature Selection and Transformation Methods, distance based Clustering Algorithms, Word and Phrase based Clustering, Probabilistic document Clustering	
	2.2	**Text Classification**: Feature Selection, Decision tree Classifiers, Rule-based Classifiers, Probabilistic based Classifiers, Proximity based Classifiers.	
	2.3	**Text Modelling:** Bayesian Networks, Hidden Markovian Models, Markov random Fields, Conditional Random Fields	
3.0		**Web-Mining:**	05
	3.1	**Introduction to Web-Mining:** Inverted indices and Compression, Latent Semantic Indexing, Web Search,	
	3.2	**Meta Search:** Using Similarity Scores, Rank Positons	
	3.3	**Web Spamming**: Content Spamming, Link Spamming, hiding Techniques, and Combating Spam	
4.0		**Web Usage Mining:**	**05**
	4.1	Data Collection and Pre-processing, Sources and types of Data, Data Modelling, Session and Visitor Analysis, Cluster Analysis and Visitor segmentation, Association and Correlation Analysis, Analysis of Sequential and Navigational Patterns, Classification and Prediction based on Web User Transactions.	
5.0		**Social Media Mining**:	**05**
	5.1	Introduction, Challenges, Types of social Network Graphs	
	5.2	Mining Social Media: Influence and Homophily, Behaviour Analytics, Recommendation in Social Media: Challenges, Classical recommendation Algorithms, Recommendation using Social Context, Evaluating recommendations.	
6.0		**Opinion Mining and Sentiment Analysis**:	**08**
	6.1	The problem of opinion mining,	
	6.2	**Document Sentiment Classification**: Supervised, Unsupervised	
	6.3	**Opinion Lexicon Expansion:** Dictionary based, Corpus based	
	6.4	**Opinion Spam Detection**: Supervised Learning, Abnormal Behaviours, Group Spam Detection.	
		Total	**48**

Textbooks:

1 Daniel Jurafsky and James H. Martin, “Speech and Language Processing,” 3rd edition, 2020

2 Charu. C. Aggarwal, Cheng Xiang Zhai, Mining Text Data, Springer Science and Business Media, 2012.

3 BingLiu, “Web Data Mining-Exploring Hyperlinks, Contents, and Usage Data”, Springer, Second Edition, 2011.

4 Reza Zafarani, Mohammad Ali Abbasiand Huan Liu, “Social Media Mining- An Introduction”, Cambridge University Press, 2014

Assessment:

Internal Assessment: (20)

1 Assessment consists of two class tests of 20 marks each.
2 The first-class test is to be conducted when approx. 40% syllabus is completed and second-class test when additional 40% syllabus is completed.
3 Duration of each test shall be one hour.

End Semester Theory Examination: (80)

1 Question paper will comprise of **total 06** questions, each carrying **20 marks**.
2 **Question No: 01** will be **compulsory** and based on the entire syllabus wherein 4 to 5 sub-questions will be asked.
3 Remaining questions will be mixed in nature and randomly selected from all the modules.
4 Weightage of each module will be proportional to number of respective lecture hours as mentioned in the syllabus.
5 **Total 04 questions** need to be solved.