Course Code:	Course Title	Credit
CSC701	Machine Learning	3

Prerequisite: Engineering Mathematics, Data Structures, Algorithms	
Course Objectives:	
1	To introduce the basic concepts and techniques of Machine Learning.
2	To acquire in depth understanding of various supervised and unsupervised algorithms
3	To be able to apply various ensemble techniques for combining ML models.
4	To demonstrate dimensionality reduction techniques.
Course Outcomes:	
1	To acquire fundamental knowledge of developing machine learning models.
2	To select, apply and evaluate an appropriate machine learning model for the given
3	To demonstrate ensemble techniques to combine predictions from different models.
4	To demonstrate the dimensionality reduction techniques.

Module		Content	Hrs
1		Introduction to Machine Learning	04
	1.1	Machine Learning, Types of Machine Learning, Issues in Machine Learning, Application of Machine Learning, Steps in developing a Machine Learning Application.	
	1.2	Training Error, Generalization error, Overfitting, Underfitting, Bias-Variance trade-off.	
2		Learning with Regression and Trees	09
	2.1	Learning with Regression: Linear Regression, Multivariate Linear Regression, Logistic Regression.	
	2.2	Learning with Trees: Decision Trees, Constructing Decision Trees using Gini Index (Regression), Classification and Regression Trees (CART)	
	2.3	Performance Metrics: Confusion Matrix, [Kappa Statistics], Sensitivity, Specificity, Precision, Recall, F-measure, ROC curve	
3		Ensemble Learning	06
	3.1	Understanding Ensembles, K-fold cross validation, Boosting, Stumping, XGBoost	
	3.2	Bagging, Subagging, Random Forest, Comparison with Boosting, Different ways to combine classifiers	
4		Learning with Classification	08
	4.1	Support Vector Machine Constrained Optimization, Optimal decision boundary, Margins and support vectors, SVM as constrained optimization problem, Quadratic	

		Programming, SVM for linear and nonlinear classification, Basics of Kernel trick.	
	4.2	Support Vector Regression, Multiclass Classification	
5		Learning with Clustering	07
	5.1	Introduction to clustering with overview of distance metrics and major clustering approaches.	
	5.2	Graph Based Clustering: Clustering with minimal spanning tree Model based Clustering: Expectation Maximization Algorithm, Density Based Clustering: DBSCAN	
6		Dimensionality Reduction	05
	6.1	Dimensionality Reduction Techniques, Principal Component Analysis, Linear Discriminant Analysis, Singular Valued Decomposition.	
		Total	39

Textbooks:	
1	Peter Harrington, "Machine Learning n Action", DreamTech Press
2	Ethem Alpaydın, "Introduction to Machine Learning", MIT Press
3	Tom M. Mitchell, "Machine Learning" McGraw Hill
4	Stephen Marsland, "Machine Learning An Algorithmic Perspective", CRC Press
References:	
1	Han Kamber, —Data Mining Concepts and Techniques‖, Morgan Kaufmann Publishers
2	Margaret. H. Dunham, —Data Mining Introductory and Advanced Topics, Pearson Education
3	Kevin P. Murphy , Machine Learning — A Probabilistic Perspective‖
4	Samir Roy and Chakraborty, —Introduction to soft computing‖, Pearson Edition.
5	Richard Duda, Peter Hart, David G. Stork, "Pattern Classification", Second Edition, Wiley Publications.
Assessment:	
Internal Assessment:	
Assessment consists of two class tests of 20 marks each. The first class test is to be conducted when approximately 40% syllabus is completed and the second class test when an additional 40% syllabus is completed. Duration of each test shall be one hour.	
End Semester Theory Examination:	
1	Question paper will comprise a total of six questions.
2	All question carries equal marks
3	Questions will be mixed in nature (for example supposed Q.2 has part (a) from module 3 then part (b) will be from any module other than module 3)

Course Code	Course Name	Credit
CSC702	Big Data Analysis	03

Prerequisite: Database, Data mining.	
Course Objectives: The course aims:	
1	To provide an overview of the big data platforms, its use cases and Hadoop ecosystem.
2	To introduce programming skills to build simple solutions using big data technologies such as MapReduce, Scripting for No SQL and R
3	To learn the fundamental techniques and principles in achieving big data analytics with scalability and streaming capability.
4	To enable students to have skills that will help them to solve complex real-world problems for decision support.
Course Outcomes:	
1	Understand the building blocks of Big Data Analytics.
2	Apply fundamental enabling techniques like Hadoop and MapReduce in solving real world problems.
3	Understand different NoSQL systems and how it handles big data.
4	Apply advanced techniques for emerging applications like stream analytics.
5	Achieve adequate perspectives of big data analytics in various applications like recommender systems, social media applications, etc.
6	Apply statistical computing techniques and graphics for analyzing big data.

Module		Detailed Content	Hours
1		**Introduction to Big Data and Hadoop**	2
	1.1	Introduction to Big Data - Big Data characteristics and Types of Big Data	
	1.2	Traditional vs. Big Data business approach	
	1.3	Case Study of Big Data Solutions	
	1.4	Concept of Hadoop, Core Hadoop Components; Hadoop Ecosystem	
2		**Hadoop HDFS and MapReduce**	8
	2.1	Distributed File Systems: Physical Organization of Compute Nodes, Large-Scale File-System Organization.	
	2.2	MapReduce: The Map Tasks, Grouping by Key, The Reduce Tasks, Combiners, Details of MapReduce Execution, Coping With Node Failures.	
	2.3	Algorithms Using MapReduce: Matrix-Vector Multiplication by MapReduce, Relational-Algebra Operations, Computing Selections by MapReduce, Computing Projections by MapReduce, Union ,Intersection,	

		and Difference by MapReduce	
	2.4	Hadoop Limitations	
3		**NoSQL**	10
	3.1	Introduction to NoSQL, NoSQL Business Drivers	
	3.2	NoSQL Data Architecture Patterns: Key-value stores, Graph stores, Column family (Bigtable)stores, Document stores, Variations of NoSQL architectural patterns, NoSQL Case Study	
	3.3	NoSQL solution for big data, Understanding the types of big data problems; Analyzing big data with a shared-nothing architecture; Choosing distribution models: master-slave versus peer-to-peer; NoSQL systems to handle big data problems.	
4		**Mining Data Streams**	11
	4.1	The Stream Data Model: A Data-Stream-Management System, Examples of Stream Sources, Stream Queries, Issues in Stream Processing.	
	4.2	Sampling Data techniques in a Stream	
	4.3	Filtering Streams: Bloom Filter with Analysis.	
	4.4	Counting Distinct Elements in a Stream,Count-Distinct Problem, Flajolet-Martin Algorithm, Combining Estimates, Space Requirements	
	4.5	Counting Ones in a Window: The Cost of Exact Counts,The Datar-Gionis-Indyk-Motwani Algorithm, Query Answering in theDGIM Algorithm, Decaying Windows.	
5		**Real-Time Big Data Models**	4
	5.1	A Model for Recommendation Systems, Content-Based Recommendations, Collaborative Filtering	
	5.2	Case Study: Product Recommendation	
	5.3	Social Networks as Graphs, Clustering of Social-Network Graphs, Direct Discovery of Communities in a social graph	
6		**Data Analytics with R**	4
	6.1	Exploring Basic features of R, Exploring RGUI, Exploring RStudio, Handling Basic Expressions in R, Variables in R, Working with Vectors, Storing and Calculating Values in R, Creating and using Objects, Interacting with users, Handling data in R workspace, Executing Scripts, Creating Plots, Accessing help and documentation in R	
	6.2	Reading datasets and Exporting data from R, Manipulating and Processing Data in R, Using functions instead of script, built-in functions in R	
	6.3	Data Visualization: Types, Applications	

Textbooks:	
1	Cre Anand Rajaraman and Jeff Ullman —Mining of Massive Datasets, Cambridge UniversityPress
2	Alex Holmes —Hadoop in Practice, Manning Press, Dreamtech Press.
3	Dan Mcary and Ann Kelly —Making Sense of NoSQL – A guide for managers and the rest of us, Manning Press.
4	DT Editorial Services, "Big Data Black Book", Dreamtech Press

Course Code	Course Title	Credit
CSDC7012	Cyber Security	3

Prerequisite: Computer Networks, Databases			
Course Objectives:			
1	To understand the need for Cyber Security Awareness.		
2	To understand the Flow and Methodology of an attack		
3	To learn and explore various Static and Web vulnerability analysis tools		
4	To understand the various IPR, Privacy and Security compliances		
Course Outcomes:			
1	Understand the need of Cyber Security and awareness of existing law infrastructure		
2	Illustrate the various tools and techniques used by attackers to launch their attacks		
3	Appraise various mechanisms of conducting system vulnerability analysis		
4	Discuss various web application vulnerability scanning techniques		
5	Identify the various network defense methodologies		
6	Describe the various Privacy and standard compliances		
Module		**Content**	**Hrs**
1		**Introduction to Cyber Crime and Attack Psychology**	**7**
	1.1	**Cyber Crime:** Need for Cyber Security, Cybercrime definition, Types of Cybercrime. Classifications of cybercrime, Hackers, Crackers and cyber criminals Indian IT ACT 2008 and its amendments.	
	1.2	**Attack Psychology:** How cyber criminals plan the attacks, Social Engineering, Cyber stalking, Cyber café and Cybercrimes, Botnets, Attack vector, Attacks on Mobile.	
2		**Threat and Vulnerability Landscape**	**7**
		Goals of Security, Vulnerability vs Threat, What is Privacy, Anonymity and pseudonymity, The Vulnerability Landscape, Threat Modeling and risk Assessment, The Zero Trust model, Spyware, Adware, Scareware , Browser Hijacking, Phishing vs Vishing vs SMShing, Doxing and Spamming,Social Engineering attack.	
3		**System Vulnerability Scanning and Network Reconnaissance**	**7**

		Password Cracking, Key loggers and Spywares, Virus and Worms, Steganography, DoS and DDoS Attacks, SQL Injection, Buffer Overflow, Phishing, Identity Theft (ID Theft),network information gathering, vulnerability scanning, Open Port Identification, Banner Grabbing Techniques, Vulnerability probing.	
4		**Web Application Security**	**8**
		OWASP, Web Security Considerations, Management, Cookies, SSL, HTTPS, SSH, Privacy on Web, Web Browser Attacks, Account Harvesting, Web Bugs, Clickjacking, Cross-Site Request Forgery, Session Hijacking and Management, Phishing and Pharming Techniques, Web Service Security, OAuth 2.0	
5		**Network Security and Defense Tools**	**6**
		Network security: Cyber Hygine, Firewalls vs Packet Filters, Stateless vs Statefull Firewalls, Intrusion Detection System and its types, Honeypots, Demilitarized Zones, Network Address Translation (NAT), Port Forwarding. Virtual Private Networks, Email Security(GPG Encryption), Registry Settings for Mobile Devices	
6		**Information Security Privacy and Standard Compliances**	**4**
		HIPPA, FISMA, PCI DSS, GDPR, Intellectual Property Aspect of Cyber Law, Creative Commons Library, Data Protection Laws in India.	

Textbooks:

1	William Stallings, Computer Security Principles and Practice, , Sixth Edition, Pearson Education
2	Charles P. Pfleeger Security in Computing, , Fifth Edition, Pearson Education
3	The Complete Cyber Security Course -Volume 1- Nathan House
4	Eric Cole, Network Security Bible, Second Edition, Wiley

References:

1	Nina Godbole, Sunit Belapure, Cyber Security, Wiley India, New Delhi
2	The Information technology Act, 2000; Bare Act- Professional Book Publishers,
3	Michael Gregg, "Build your own Security Lab", Wiley India
4	Dieter Gollman, "Computer Security", Third Edition, Wiley

Digital References:

Virtual Penetration Testing Labs- https://pentesterlab.com
OWASP- https://owasp.org/
DVWA- https://dvwa.co.uk
FISMA - https://csrc.nist.gov/projects/risk-management/fisma-background
PCI DS https://www.itgovernance.eu/blog/en/a-guide-to-the-4-pci-dss-compliance-levels

Course Code	Course Name	Credit
CSDC7013	Natural Language Processing	03

Pre-requisite: Theory of Computer Science, System Programming &Compiler Construction	
Course Objectives: The course aims	
1	To define natural language processing and to learn various stages of natural language processing.
2	To describe basic concepts and algorithmic description of the main language levels: Morphology, Syntax, Semantics, and Pragmatics &Discourse analysis.
3	To design and implement various language models and POS tagging techniques.
4	To design and learn NLP applications such as Information Extraction, Question answering.
5	To design and implement applications based on natural language processing.

Course Outcomes: Students will be able	
1	To describe the field of natural language processing.
2	To design language model for word level analysis for text processing.
3	To design various POS tagging techniques and parsers.
4	To design, implement and test algorithms for semantic and pragmatic analysis.
5	To formulate the discourse segmentation and anaphora resolution.
6	To apply NLP techniques to design real world NLP applications.

Module		Detailed Content	Hours
1	1.1	**Introduction to NLP**	3
		Origin & History of NLP; Language, Knowledge and Grammar in language processing; Stages in NLP;Ambiguities and its types in English and Indian Regional Llanguages; Challenges of NLP;Applications of NLP	
	1.2	**Self-Learning topics:** Variety types of tools for regional languages pre-processing and other functionalities	
2	2.1	**Word Level Analysis**	9
		Basic Terms: Tokenization, Stemming, Lemmatization; Survey of English Morphology, Inflectional Morphology, Derivational Morphology; Regular expression with types; Morphological Models: Dictionary lookup, finite state morphology; Morphological parsing with FST (Finite State Transducer);Lexicon free FST Porter Stemmer algorithm; Grams and its variation: Bigram, Trigram; Simple (Unsmoothed) N-grams; N-gram Sensitivity to the Training Corpus; Unknown Words: Open	

		versus closed vocabulary tasks; Evaluating N-grams: Perplexity; Smoothing: Laplace Smoothing, Good-Turing Discounting;	
	2.2	**Self-Learning topics:** Noisy channel models, various edit distance, Advance Issues in Language Modelling	
3	3.1	**Syntax analysis**	10
		Part-Of-Speech tagging(POS); Tag set for English (Upenn Treebank); Difficulties /Challenges in POS tagging; Rule-based, Stochastic and Transformation-based tagging; Generative Model: Hidden Markov Model (HMM Viterbi) for POS tagging; Issues in HMM POS tagging; Discriminative Model: Maximum Entropy model, Conditional random Field (CRF);Parsers: Top down and Bottom up; Modelling constituency; Bottom Up Parser: CYK, PCFG (Probabilistic Context Free Grammar), Shift Reduce Parser; Top Down Parser: Early Parser, Predictive Parser	
	3.2	**Self-Learning topics:** Evaluating parsers, Parsers based language modelling, Regional languages POS tree banks	
4	4.1	**Semantic Analysis**	7
		Introduction, meaning representation; Lexical Semantics; Corpus study; Study of Various language dictionaries like WorldNet, Babelnet; Relations among lexemes & their senses –Homonymy, Polysemy, Synonymy, Hyponymy; Semantic Ambiguity; Word Sense Disambiguation (WSD); Knowledge based approach(Lesk's Algorithm), Supervised (Naïve Bayes, Decision List),Introduction to Semi-supervised method (Yarowsky) Unsupervised (Hyperlex)	
	4.2	**Self-Learning topics:** Dictionaries for regional languages, Distributional Semantics, Topic Models	
5	5.1	**Pragmatic & Discourse Processing**	5
		Discourse: Reference Resolution, Reference Phenomena, Syntactic & Semantic constraint on coherence; Anaphora Resolution using Hobbs and Cantering Algorithm	
	5.2	**Self-Learning topics:** Discourse segmentation, Conference resolution	
6	6.1	**Applications of NLP**	5
		Case studies on (preferable in regional language):Machine translation; Text Summarization; Sentiment analysis; Information retrieval; Question Answering system	
	6.2	**Self-Learning topics:** Applications based on Deep Neural Network with NLP such as LSTM network, Recurrent Neural network etc.	

Textbooks:	
1	Daniel Jurafsky, James H. and Martin, Speech and Language Processing, Second Edition, Prentice Hall, 2008.
2	Christopher D.Manning and HinrichSchutze, Foundations of Statistical Natural Language Processing, MIT Press, 1999.
References:	
1	Siddiqui and Tiwary U.S., Natural Language Processing and Information Retrieval, Oxford University Press, 2008.
2	Daniel M Bikel and ImedZitouni — Multilingual natural language processing applications:

Course Code	Course Name	Credit
CSDC7023	Information Retrieval	03

Prerequisite: Data structures and algorithms	
Course Objectives: The course aims students :	
1	To learn the fundamentals of Information Retrieval
2	To analyze various Information retrieval modeling techniques
3	To understand query processing and its applications
4	To explore the various indexing and scoring techniques
5	To assess the various evaluation methods
6	To analyze various information retrieval for real world application
Course Outcomes: Learner will be able to: -	
1	Define and describe the basic concepts of the Information retrieval system.
2	Design the various modeling techniques for information retrieval systems.
3	Understand the query structure and various query operations
4	Analyzing the indexing and scoring operation in information retrieval systems
5	Perform the evaluation of information retrieval systems
6	Analyze various information retrieval for real world application

Module		Detailed Content	Hours
1		**Introduction to Information Retrieval**	4
	1.1	Introduction to Information Retrieval, Basic Concepts, Information Versus Data, Trends and research issues in information retrieval.	
	1.2	The retrieval process, Information retrieval in the library, web and digital libraries.	
2		**Modeling in Information Retrieval**	8
	2.1	Taxonomy of Information Retrieval models, Classic Information Retrieval, Alternate set: Theoretical model, Alternative Algebraic models, Alternative Probabilistic models	
	2.2	Structured text Retrieval models, Models for browsing	
3		**Query and Operations in Information Retrieval**	8
	3.1	Query structures, Keyboard based querying, Pattern matching, Structured queries	
	3.2	User relevance feedback, Automatic local analysis, Automatic global analysis	
4		**Indexing and Scoring in Information Systems**	8
	4.1	Introduction, Inverted Files, Other Indices for Text, Boolean queries and Introduction to Sequential searching	

	4.2	Scoring, term weighting and the vector space model, Parametric and zone indexes, Weighted zone scoring, Learning weights, The optimal weight, Term frequency and weighting, Inverse document frequency, Tf-idf weighting. The vector space model for scoring, Queries as vectors, Computing vector scores, Efficient scoring and ranking, Inexact top K document retrieval	
5		**Evaluation of Information Retrieval Systems**	6
	5.1	Information retrieval system evaluation, Standard test collections, Evaluation of unranked retrieval sets, Evaluation of ranked retrieval results, Assessing and justifying the concept of relevance	
	5.2	System quality and user utility, System issues, Refining a deployed system	
6.		**Applications of Information Retrieval Systems**	5
	6.1.	Introduction to Multimedia Information Retrieval	
	6.2	Introduction to Distributed Information Retrieval	

Textbooks:	
1	Modern information retrieval, Baeza-Yates, R. and Ribeiro-Neto, B., 1999. ACM press.
2	Introduction to Information Retrieval By Christopher D. Manning and PrabhakarRaghavan, Cambridge University Press
3	Information Storage & Retrieval By Robert Korfhage – John Wiley & Sons
References:	
1	Storage Network Management and Retrieval, VaishaliKhairnar
2	Introduction to Modern Information Retrieval. G.G. Chowdhury. NealSchuman
3	Natural Language Processing and Information Retrieval by Tanveer Siddiqui, U.S Tiwarey

Useful Digital Links	
1	https://web.stanford.edu/class/cs276/
2	https://www.coursera.org/learn/text-retrieval

Assessment:	
Internal Assessment:	
Assessment consists of two class tests of 20 marks each. The first-class test is to be conducted when approx. 40% syllabus is completed and second class test when additional 40% syllabus is completed. Duration of each test shall be one hour.	
End Semester Theory Examination:	
1	Question paper will consist of 6 questions, each carrying 20 marks.
2	The students need to solve a total of 4 questions.
3	Question No.1 will be compulsory and based on the entire syllabus.
4	Remaining question (Q.2 to Q.6) will be selected from all the modules.