



Syllabus

Honours Programme in

Data Science & Analytics

(Offered by Department of Computer Engineering)

From

Academic Year 2021-22

Revision 2

(Approved in Academic Council meeting dated)



K J Somaiya College of Engineering, Mumbai-77

(A Constituent College of Somaiya Vidyavihar University)

Honours' Degree Programme in Data Science & Analytics

Offered by Department of Computer Engineering

Introduction:

In today's data-driven society, Data Science provides a foundation for problem solving that impacts virtually all areas of the economy, including science, engineering, medicine, banking, finance, sports and the arts. Data science is an interdisciplinary field that focuses on analysing large amounts of data to identify inherent patterns, extract underlying models, and make relevant predictions.

Data processing and analytics converts raw data into format which can be analysed and interpreted for a variety of purposes. It focuses on data processing techniques and algorithms for representation of data in a meaningful way for human intelligence. The focus is also on innovative and intelligent ways of handling data which may be unstructured, high in volume, of different variety and analyzing data for various purposes.

The data science and analytics programme is designed to prepare students in wide disciplines who want to gain practical know-how of data analytics methods as it relates to their field of interest. It is designed to empower them to employ computational thinking and data science tools to solve practical business problems. The coursework consists of courses that cover the spectrum of Data Science to equip the students with knowledge of data analysis techniques and data-centric computation to address problems that require large data.

Objectives:

- Applications of principles of Data Science to the analysis of diverse problems.
- Use software tools and algorithms from statistics, applied mathematics, Computer Science to model and analyze real-world data, communicate findings, and effectively present results using data visualization techniques.
- Deployment of latest tools and technologies to analyze large amounts Data.
- Understand the ethical practices that are importantly and inevitably tied to data-driven decision-making.

Learning Outcomes of the Honours' Degree Programme:

At the successful completion of this programme an engineering graduates will be able to

- Apply principles of Data Science for analytics to diverse problems.
- Demonstrate the use of various tools in the domain of Data Science for data visualization.
- Discuss ethical practices related to data-driven decision-making.

- Implement solutions to data analysis problems using latest tools and technologies.

Assessment Methods: Evaluation is done by a variety of tools including Open book tests, MCQs (multiple choice questions), Study of research papers, Internal Assessment tools and End Semester Examinations etc. Mini-Projects are offered in courses also to encourage project based learning among students.

Acronyms used in syllabus document	
Acronym	Definition
CA	Continuous Assessment
ESE	End Semester Exam
IA	Internal Assessment
O	Oral
P	Practical
P&O	Practical and Oral
TH	Theory
TUT	Tutorial
TW	Term work
ISE	In-semester Examination
CO	Course Outcome

Acronyms used in Course code e.g. 116h54C301

Position of Digit	Acronym	Definition
1	1	First revision SUV KJSCE 2020
2	16	KJSCE
3	h	Honour Degree Program
4	55	Data Science & Analytics
5	C	Core Course
	L	Laboratory Course
	T	Tutorial
	P	Project Based Course
6	1/2/3/4	Semester Number
7	01/02/03--	Course Number

Credit Scheme

Course Code	Course Name	Teaching Scheme (Hrs.) TH – P – TUT	Total (Hrs.)	Credits Assigned TH – P – TUT	Total Credits	Suggested semester of Honours' degree
116h54C301	Data Visualization	3 – 0 – 0	03	3 – 0 – 0	03	III
116h54L301	Data Visualization	0 – 2 – 0	02	0 – 1 – 0	01	III
116h54C401	Applied Data Science	3 – 0 – 0	03	3 – 0 – 0	03	IV
116h54L401	Applied Data Science	0 – 2 – 0	02	0 – 1 – 0	01	IV
116h54C501	Data Analytics	3 – 0 – 0	03	3 – 0 – 0	03	V
116h54L501	Data Analytics	0 – 2 – 0	02	0 – 1 – 0	01	V
116h54C601	Advanced Data Mining	3 – 0 – 0	03	3 – 0 – 0	03	VI
116h54C701	Advanced Machine Learning	3 – 0 – 0	03	3 – 0 – 0	03	VII
116h54P801	Applied Project/ Internship	0 – 4 – 0	04	0 – 2 – 0	02	VIII
	Total	15 – 10 – 0	25	15 – 5 – 0	20	

Somaiya Vidyavihar University
K. J. Somaiya College of Engineering, Mumbai -77
(A Constituent College of Somaiya Vidyavihar University)

Examination Scheme

Course Code	Course Name	Examination Scheme							
		Marks							
		CA		ESE	TW	O*	P	P&O	Total
		ISE	IA						
116h54C301	Data Visualization	30	20	50	-	-	-	-	100
116h54L301	Data Visualization	-	-	-	25	25	-	-	50
116h54C401	Applied Data Science	30	20	50	-	-	-	-	100
116h54L401	Applied Data Science	-	-	-	25	25	-	-	50
116h54C501	Data Analytics	30	20	50	-	-	-	-	100
116h54L501	Data Analytics	-	-	-	25	25	-	-	50
116h54C601	Advanced Data Mining	30	20	50	-	-	-	-	100
116h54C701	Advanced Machine Learning	30	20	50	-	-	-	-	100
116h54P801	Applied Project/ Internship	-	-	-	50	50	-	-	100
Total		150	100	250	125	125			750

Course Code	Course Title							
116h54C301	Data Visualization							
	TH			P	TUT			Total
Teaching Scheme(Hrs.)	03			--	--			03
Credits Assigned	03			--	--			03
Examination Scheme	Marks							
	CA		ESE	TW	O	P	P&O	Total
	ISE	IA						
	30	20	50	--	--	--	--	100

Course prerequisites (if any):

Basics of statistics, database and data analysis

Course Objectives

- Employ best practices in data visualization to develop charts, maps, tables, and other visual representations of data
- Use visualization tools to conduct data analysis, especially exploration of an unfamiliar dataset.
- Create compelling, interactive dashboards to combine several visualizations into a cohesive and functional whole.
- Use data visualizations, dashboards and Stories to support relevant communication for diverse audiences.

Course Outcomes

At the end of successful completion of the course the student will be able to

CO1	Learn how to locate and download datasets, extract insights from that data and present their findings in a variety of different formats
CO2	Detect and understand the stories within datasets and its applications.
CO3	Apply data visualization best practices
CO4	Design static charts, interactive Dashboards and data stories

Module No.	Unit No.	Details	Hrs.	CO
1	Introduction data visualization		10	CO1 CO2
	1.1	Introduction to data visualization and its need. Data analysis lifecycle. A Visual Revolution, Various types of visualization with its best practices. From Visualization to Visual Data Storytelling: An Evolution, From Visual to Story: Bridging the Gap		
	1.2	Data Fundamentals, Collecting data, Preparing Data		
2	Design principles and various charts/plots		06	CO 3
	2.1	Gestalt Design principles		
	2.2	Data cleaning using Excel		
	2.3	Introduction to various chart/ plots and its interpretation The Bar Chart The Line Chart The Pie and Donut Charts The Scatter Plot The Packed Bubble Chart Box plot, The Treemap Plotting of PDF		
3	Storytelling and Multivariate displays		10	CO4
	3.1	Storytelling Multivariate displays The Science of Storytelling The Power of Stories Context in Action Exploratory versus Explanatory Analysis Structuring Stories Audience Analysis for Storytelling Steps to Visual Data Storytelling The Important Role of Feedback		
	3.2	Graphical Perception		
	# Multivariate displays			
4	Geospatial displays		09	CO4
	4.1	The Heat Map Maps Connecting to Geographic Data Assigning Geographic Roles Creating Geographic Hierarchies Proportional Symbol Maps Choropleth Map		
5	Dashboards, interactive and animated displays		10	CO4

	5.1	Visual Design Building Blocks Color Stepped Color Reversed Color Color Effects Opacity Mark Borders Mark Halos The Truth about Red and Green Lines Formatting Grid Lines, Zero Lines, and Drop Lines Formatting Borders Formatting, Shading, and Banding Shapes Shape Marks Card Custom Shapes Timelines Bar-in-Bar Charts Likert Visualizations Lollipop Charts Word Clouds Create dashboard Working with dashboard Publishing through dashboard		
	5.2	Any other tool for visualization		
		#case study, Publishing dashboard over mobile devices		
Total			45	

Recommended Books:

Sr. No.	Name/s of Author/s	Title of Book	Name of Publisher with country	Edition and Year of Publication
1.	Sosulski, K.	<i>Data Visualization Made Simple: Insights into Becoming Visual</i>	New York: Routledge.	First edition, 2018
2.	Lindy Ryan	<i>Visual Data Storytelling with Tableau</i>	Pearson Education	First edition, 2018
3.	Kristen Sosulski	Data Visualization made simple	New York: Routledge	First edition, 2019
4.	Cole Nussbaumer Knaflie	Storytelling with Data	Wiley	First edition, 2015

Course Code	Course Title							
116h54L301	Data Visualization							
	TH		P		TUT		Total	
Teaching Scheme(Hrs.)	-		02		--		02	
Credits Assigned	-		01		--		01	
Examination Scheme	Marks							
	CA		ESE	TW	O	P	P&O	Total
	ISE	IA						
	-	-						

Term-Work:

Term work will consist of experiments/ tutorials covering entire syllabus of the course 'Data Visualization'. Students will be graded based on continuous assessment of their term work.

Course Code	Course Title							
116h54C401	Applied Data Science							
	TH			P	TUT			Total
Teaching Scheme(Hrs.)	03			--	--			03
Credits Assigned	03			--	--			03
Examination Scheme	Marks							
	CA		ESE	TW	O	P	P&O	Total
	ISE	IA						
	30	20	50	--	--	--	--	100

Course prerequisites (if any):

Students are expected to have basic knowledge of algorithms and programming experience.

Course Objectives

- To develop understanding of the applied data science in the real world problems.
- To get the understanding of R programming language with respect to data analysis.
- To understand the application of Machine Learning Algorithms for data modeling.
- To apply various data visualization techniques using real-world data sets and analyze the graphs and charts.
- To understand various analytics metrics, processing unstructured text/data, and the ability to investigate hidden patterns.

Course Outcomes

At the end of successful completion of the course the student will be able to

CO1	Develop an understanding of data science and business analytics.
CO2	Application of Exploratory data analysis (EDA) on Real world problems.
CO3	Understand the basic concept and techniques of Machine Learning regression and classification.
CO4	Understand the basic concept and techniques of Machine Learning clustering.

Module No.	Unit No.	Details	Hrs.	CO
1	Introduction to Applied Data Science and Data Scrapping Process		10	CO1
	1.1	Introduction to Applied Data Science: What is Data Science? - Big Data and Data Science, Datafication - Current landscape of perspectives - Skill sets needed and various application areas. Challenges and skill Sets needed and various applications areas.		
	1.2	Impact of applying Data Science in business scenario, Introduction to need of estimation and validation for added value due to data science		
	1.3	Introduction to the mathematical foundations required for data science. Statistical Inference: Populations and samples, Statistical modeling, Probability distribution, Fitting a model Normal Distribution, Skewness and Kurtosis, Heteroskedasticity, Descriptive Statistics, Higher-Order Moments, Matrices, Maximum-likelihood, Introduction to Brownian Motions, Monte Carlo		
	1.4	Data Scrapping: Introduction, Need, Sources, Web Scrapping, Scrapping of Images, Data Wrangling, ETL Process, Data Munging		
2	Exploratory Data Analysis		08	CO2
	2.1	Exploratory Data Analysis and the Data Science Process, Basic tools (plots, graphs and summary statistics) of EDA, Measuring similarity and dissimilarity		
	2.2	Why Preprocessing? Data Cleaning; Data Integration; Data Reduction: Attribute subset selection, Histograms, Clustering and Sampling; Data Transformation & Data Discretization: Normalization, Binning, Histogram Analysis and Concept hierarchy generation.		
	2.3	The Data Science Process : Case Study on Online E-Commerce Dataset		
3	Machine Learning Algorithms – Regression		05	CO3
	3.1	Introduction to Machine learning, Linear Regression, Building model, Model Diagnostic		
	3.2	Multiple Linear regression, Logistic Regression,		
4	Machine Learning Algorithms: Classification		12	CO3

	4.1	Decision Trees , Naive Bayes theorem, Bayes Classifier, Rule-Based Classification, Bayesian Belief networks, CART, Random Forests, k-Nearest Neighbors (k-NN), Hidden Markov Models, Support Vector Machines		
	4.2	Dimensionality Reduction Techniques: Principal Component Analysis, Independent Component Analysis, Singular value decomposition		
	4.3	Model Evaluation and Selection: Metrics, Confusion Matrix, Precision and Recall, Accuracy, False Positives, Techniques to Improve Classification Accuracy		
5	Machine Learning Algorithms: Clustering		10	CO4
	5.1	Cluster Analysis: Basic Concepts, Partitioning Methods: KMeans, KMediods and hierarchical methods: Agglomerative. Expectation Maximization Algorithm, Radial Basis functions		
	5.2	Cost Function, how to Minimize cost function, coefficients of determination.		
	5.3	Introduction Text processing, plain and simple, Text Classification, Text Summarization.		
		# Self Learning –Mini Project		
Total			45	

Students should prepare all Self Learning topics on their own. Self-learning topics will enable students to gain extended knowledge of the topic. Assessment of these topics may be included in IA and Laboratory Experiments.

Recommended Books:

Sr. No.	Name/s of Author/s	Title of Book	Name of Publisher with country	Edition and Year of Publication
1.	Han, J., Kamber, M., Pei, J.	<i>Data mining concepts and techniques</i>	Morgan Kaufmann	2011
2.	James, G., Witten, D., Hastie, T., Tibshirani, R.	<i>An introduction to statistical learning with applications in R</i>	Springer	2013
3.	Cathy O'Neil and Rachel Schutt	<i>Doing Data Science, Straight Talk From The Frontline</i>	O'Reilly	2014
4.	Kevin P. Murphy	<i>Machine Learning: A Probabilistic Perspective</i>	ISBN 0262018020	2013
5.	Mohammed J. Zaki and Wagner Miera Jr.	<i>Data Mining and Analysis: Fundamental Concepts and Algorithms</i>	Cambridge University Press	2014
6.	Avrim Blum, John Hopcroft, and Ravindran Kannan	<i>Foundations of Data Science</i>	ONLINE	2014
7	C R Kothari	Research Mythology	New Age International Publishers	4 th edition 2019

Course Code	Course Title							
116h54L401	Applied Data Science							
	TH			P	TUT			Total
Teaching Scheme(Hrs.)	-			02	--			02
Credits Assigned	-			01	--			01
Examination Scheme	Marks							
	CA		ESE	TW	O	P	P&O	Total
	ISE	IA						
	-	-	-	25	25	--	--	50

Term-Work:

Term work will consist of experiments/ tutorials covering entire syllabus of the course 'Applied Data Science'. Students will be graded based on continuous assessment of their term work.

Course Code	Course Title							
116h54C501	Data Analytics							
	TH		P		TUT		Total	
Teaching Scheme(Hrs.)	03		--		--		03	
Credits Assigned	03		--		--		02	
Examination Scheme	Marks							
	CA		ESE	TW	O	P	P&O	Total
	ISE	IA						
	30	20	50	--	--	--	--	100

Course prerequisites (if any):

- Concepts of DBMS
- Data Mining
- Knowledge of basic Machine Learning algorithms and programming language (C/C++/Java/ Python).

Course Objectives

Introduction to the fundamental concepts of Data Analytics through real world case studies
Comprehension of data analytics and visualization for decision-making in Geospatial, healthcare and text mining business applications

Course Outcomes

At the end of successful completion of the course the student will be able to

CO1	Understand basic concepts of data analytics to solve real-world problems
CO2	Apply the data analytics in the field of geospatial system
CO3	Perform the social data analytics
CO4	Perform Time series Analytics and forecasting
CO5	Apply the data analytics in the field of Health care.

Module No.	Unit No.	Details	Hrs.	CO
1	Introduction to Data Analytics		05	CO1
	1.1	Introduction to Data Analytics, Different types of data analytics: Descriptive analytics, Diagnostics Analytics, Predictive analytics, Prescriptive analytics		
	1.2	# Self-Learning: LinkedIn analytics, Netflix Analytics, Cricket and FIFA Analytics.		
2	Data Analytics in GIS		10	CO2 CO1
	2.1	Introduction, Definition of GIS, Evolution of GIS , components of GIS		
	2.2	Vector Data Model : Topology, Non topological Vector models, Attribute Data in GIS, Attribute Data Entry, Vector Data Query, Manipulation of Fields and Attribute Data Raster Data Model : Elements of Raster Data Model, Types of Raster Data, Raster Data Structure, Raster Data Query, Data Compression, Data Conversion, Integration of Raster and Vector data		
	2.3	Terrain Analysis: Data for Terrain Mapping and analysis, Terrain Mapping, slope and aspect, Surface curvature, Raster vs TIN, View shed and water shed analysis .		
	2.4	GIS application Case study: A real world problem and its step by step procedure using open source software tools.		
		# Self-learning: QGIS, Hadoop, GeoSpark R PostgreSQL, PostGIS, Python		
3	Graph Analytics		10	CO3
	3.1	Introduction to the Social Network, Clustering of Social-Network Graphs, Direct Discovery of Communities		
	3.2	Partitioning of Graphs, Finding Overlapping Communities, Simrank, Counting Triangles, Neighborhood Properties of Graphs		
	3.3	# Self-learning: GraphX tools of Apache.		
4	Time series Analytics and forecasting		10	CO4
	4.1	Introduction, Finding and Wrangling Time Series Data, Exploratory Data Analysis for Time Series, Simulating Time Series Data, Storing Temporal Data,		
	4.2	Statistical Models for Time Series, State Space Models for Time Series, forecasting methods, Testing for randomness, Regression based trend model, random walk model, moving average forecast, exponential smoothing forecast, seasonal models		
5	Data analytics in Health Care Systems		10	CO5
	5.1	Introduction, Components of HER, Benefits of EHR-		

		Barrier to Adopting HER challenges		
	5.2	Mining Sensor Data in Medical Informatics Challenges in Healthcare Data Analysis Sensor Data Mining Applications		
	5.3	Natural Language Processing and data mining for clinical text data: Mining Information from Clinical Text, Challenges of Processing Clinical Reports, Clinical Applications		
		Self-learning: Introduction to Social media analytics for healthcare,		
Total			45	

Students should prepare all Self Learning topics on their own. Self-learning topics will enable students to gain extended knowledge of the topic. Assessment of these topics may be included in IA and Laboratory Experiments.

Recommended Books:

Sr. No.	Name/s of Author/s	Title of Book	Name of Publisher with country	Edition and Year of Publication
1.	Michael J. de Smith, Michael F. Goodchild and Paul A. Longley	Geospatial Analysis: A Comprehensive Guide to Principles, Techniques, and Software Tools,	Wiley, Second Edition	2019
2.	Anil Maheshwari	<i>Data Analytics</i>	Mc Graw Hill	2017
3.	James, G., Witten, D., Hastie, T., Tibshirani, R.	<i>An introduction to statistical learning with applications in R</i>	Springer	2013
4.	Chandan K. Reddy and Charu C Aggarwal	<i>Healthcare data analytics</i>	Taylor & Francis	2015
5.	Hui Yang and Eva K. Lee	<i>Healthcare Analytics: From Data to Knowledge to Healthcare Improvement</i>	Wiley	2016
6.	Mohammed J. Zaki and Wagner Miera Jr.	<i>Data Mining and Analysis: Fundamental Concepts and Algorithms</i>	Cambridge University Press	2014
7.	U. Dinesh Kumar	<i>Business Analytics</i>	Wiley	2017
8.	Kang-tsung Chang	“Introduction to Geographical Information Systems”,	Tata McGraw Hill	Third Edition, 2003
9	Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman	Mining Massive dataset	Dreamtech Press	2 nd edition 2016
10	Albright and Winston	Bussiness Analytics	Cengage Publication	5 th edition, 2015
11	Aileen Nielsen	Practical Time Series Analysis	O'Reilly Media, Inc.	1 st edition October 2019

Somaiya Vidyavihar University
K. J. Somaiya College of Engineering, Mumbai -77
(A Constituent College of Somaiya Vidyavihar University)

Course Code	Course Title							
116h54L501	Data Analytics							
	TH		P	TUT			Total	
Teaching Scheme(Hrs.)	-		02	--			02	
Credits Assigned	-		01	--			01	
Examination Scheme	Marks							
	CA		ESE	TW	O	P	P&O	Total
	ISE	IA						
	-	-	-	25	25	--	--	50

Term-Work:

Term work will consist of experiments/ tutorials covering entire syllabus of the course ‘Data Analytics’. Students will be graded based on continuous assessment of their term work.

Course Code	Course Title							
116h54C601	Advanced Data mining							
	TH			P	TUT			Total
Teaching Scheme(Hrs.)	03			--	--			03
Credits Assigned	03			--	--			03
Examination Scheme	Marks							
	CA		ESE	TW	O	P	P&O	Total
	ISE	IA						
	30	20	50	--	--	--	--	100

Course prerequisites: Data mining

Course Objectives:

1. To analyze various algorithms and techniques to mine complex data beyond conventional record data to mining complex structure and complex data
2. Extract useful knowledge from massive data sources using distributed computing solutions for data intensive applications
3. To develop the basic skills necessary to pursue research in data mining.

Course Outcomes

On completion of the course students will be expected to

1. Describe the fundamental issues and challenges of mining complex data
2. Analyze patterns in streaming data
3. Derive patterns from complex structures and sequence data
4. Understand the concepts of information retrieval and web search
5. Analyze patterns in multivariate time series data

Module No.	Unit No.	Details	Hrs.	CO
Advanced Data Mining				
1	Data mining Introduction		03	CO1
	Data mining process, different types of data representation, different types of knowledge mined, common data mining tasks, Distributed computing solution for data mining and applications			
2	Incremental Datamining and Stream mining		06	CO2
	Incremental algorithms for mining frequent patterns Characteristics of Streaming Data, Issues and Challenges, Streaming Data Mining Algorithms			
3	Mining complex structures		10	CO3
	Mining trees- Tree Model Guided Framework, TMG framework for mining ordered & unordered subtrees, Tree Mining Applications, Mining maximal and closed frequent trees, Tree mining application Mining Graphs- Approaches to graph mining. Mining social-network graph			
4	Text mining & Web Search		08	CO4
	Text Classification, Vector Space Model, Flat and Hierarchical Clustering Web search: Crawling & Indexing, Hyperlink Analysis, Page Rank algorithm, Web Search and Information Retrieval, Application: Query Recommender System			
5	Sequence mining & Multivariate and Time series mining		08	CO5
	Sequence Mining- Characteristics of Sequence Data, Problem Modeling, Sequential Pattern Discovery , Timing Constraints Applications in Bioinformatics Multivariate and Time series mining- Importance of Multivariate and Time series data, Sources of MVTs data, Mining MVTs data: Sign Language Data, Agro-meteorological Data			

Recommended Books:

Sr. No.	Name/s of Author/s	Title of Book	Name of Publisher with country	Edition and Year of Publication
1.	Hadzic F., Tan H. & Dillon T. S	Mining data with Complex Structures	Springer	2011
2.	Yates R. B. and Neto B. R	Modern Information Retrieval	Pearson Education	2005
3.	Han J. & Kamber M	Data Mining: Concepts and Techniques	Morgan Kaufmann Publishers	Third edition, 2011
4.	Christopher D.M., Prabhakar R. & Hinrich S	Introduction to Information Retrieval”	Cambridge UP	Online edition, 2009
5.	Tan P. N., Steinbach M & Kumar V	Introduction to Data Mining	Pearson Education	2006

Course Code	Course Title							
116h54C701	Advanced Machine Learning							
	TH			P	TUT			Total
Teaching Scheme(Hrs.)	03			2	--			05
Credits Assigned	03			1	--			04
Examination Scheme	Marks							
	CA		ESE	TW	O	P	P&O	Total
	ISE	IA						
	30	20	50	--	--	--	--	100

Course prerequisites: Probability concepts, calculus, linear algebra & basic python programming.

Course Objectives:

6. To analyse various machine learning algorithms and techniques with a modern outlook focusing on recent advances.
7. Explore supervised and unsupervised learning paradigms of machine learning.
8. To understand natural language processing and to learn how to apply basic algorithms in this field.
9. To develop the basic skills necessary to pursue research in machine learning.

Course Outcomes

On completion of the course students will be expected to

1. Describe the fundamental issues and challenges of machine learning: data, model selection, model complexity.
2. Understand & describe mathematical foundation behind deep learning architectures.
3. Design & implement various deep supervised learning architectures for image data.
4. Learn and train various deep learning models in NLP.
5. Apply various deep learning techniques to design efficient algorithms for real-world applications.

Module No.	Unit No.	Details	Hrs.	CO
1	Machine learning foundation:		05	CO1
	What is Machine Learning? Types of learning, applications, Bias, variance, overfitting, underfitting, cross validation and feature engineering, gradient descent learning algorithm and its variations			
2	Deep learning fundamentals:		10	CO2
	Mathematical foundations of deep learning & neural network playground. Learning rates and data normalization, activation functions, Optimizers, Regularization, Dropout, Momentum & Batch Norm. Neural Nets-Deep Vs Shallow Networks- deep multi-layer perceptron. #Self-Learning: forward and backward propagation in neural networks.			
3	Deep Learning Architecture:		10	CO3
	Convolutional Neural Networks: Convolution, pooling operations, basic CNN architecture and image classification using CNN. Recurrent Neural Networks: RNN introduction, difference between feed forward and RNN, forward, backward propagation in RNN & Vanishing/Exploding gradient problem. Long Short Term Memory: LSTM gates, forward, backward propagation in LSTM. #Self-Learning: Auto encoders			
4	Deep learning for NLP:		10	CO4
	NLP introduction, text preprocessing techniques, text representation (one hot encoding, bag of N grams, TFIDF features). Word Embedding – word2vec model. Text generation, Seq2Seq models (Encoder decoder architecture). #Self-Learning: NLTK & Gensim library			
5	Machine learning case studies:		10	CO5
	<ul style="list-style-type: none">Churn Analysis and PredictionSentiment Analysis or Topic Mining from New York TimesCustomer Segmentation and ValueNetflix Movie Recommendation SystemImage & text classification using TensorFlow & PyTorch.Transfer learning with Pre-trained CNN using TensorFlow.			
Total			45	

Recommended Books:

Sr.	Name/s of Author/s	Title of Book	Name of Publisher	Edition and
-----	--------------------	---------------	-------------------	-------------

Somaiya Vidyavihar University
K. J. Somaiya College of Engineering, Mumbai -77
(A Constituent College of Somaiya Vidyavihar University)

No.			with country	Year of Publication
1	Tom M.Mitchell	Machine Learning	McGraw Hill	2017
2	M. Gopal	Applied Machine Learning	McGraw Hill	2018
3	Ian Goodfellow, Yoshua Bengio, Aaron Courville	Deep Learning	An MIT Press book	2016
4	Deng & Yu	Deep Learning: Methods and Applications	Now Publishers	2013
5	Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta and Harshit Surana	Practical Natural Language Processing	O'Reilly Media, Inc.	June 2020
6	Sudharsan Ravichandiran	Hands-On Deep Learning Algorithms with Python	Packt Publishing	July 2019

Course Code	Course Title								
116h54P801	Applied Project/Internship								
	TH			P	TUT			Total	
Teaching Scheme (Hrs./Week)	-			04	-			04	
Credits Assigned	-			02	-			02	
Examination Scheme	Marks								
	CA			ESE	TW	O	P	P&O	Total
	ISE		IA						
	-	-	-	-	50	50	-	-	100

Course prerequisites: Conceptual knowledge of Data Science & Analytics

Course Objectives: The objectives are to address a real-world problem, which includes identify and solve the problem by implementing the solution using the courses learned in earlier semesters. Recognize various hardware and software requirements for solving the problem. It will also inculcate qualities such as working in team, meeting deadlines, making and following work plan. The Project may include some software or techniques not covered in the courses taught to provide solution of the chosen problem.

Course Outcomes:

At the end of successful completion of the course the student will be able to

- CO1. Define the problem statement and scope of problem.
- CO2. Identify various hardware and software requirements for problem solution
- CO3. Describe the design with the help of flowchart/block diagrams or any design Tool.
- CO4. Implement and test the design to meet the desired specifications.
- CO5. Analyze, interpret results and correspondingly modify the designed system to get the desired results.
- CO6. Prepare a technical report and technical paper based on the project.

Term Work and Oral: This is an activity to be undertaken by the group of 2 or 3 students. Each group will be assigned one faculty member as a supervisor. There will be continuous assessment of the project and progress report of the project needs to be maintained by students. The final oral will be a presentation based on a demonstration of the project in front of a committee of examiners. Students are expected to publish technical paper based on the project.