

Syllabus Honours Programme in Data Science & Analytics

(Offered by Department of Computer Engineering)

From Academic Year 2021-22

(Approved in Academic Council meeting dated _____)



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

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.

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

• 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				
0	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. 116N54C301

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

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

Honours Programme in Data Science & Analytics

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
116N54C301	Data Visualization	3 - 2 - 0	05	3 - 1 - 0	04	III
116N54C401	Applied Data Science	3-2-0	05	3-1-0	04	IV
116N54C501	Data Analytics	3-2-0	05	3-1-0	04	V
116N54C601	Advanced Data Mining	3-0-0	03	3-0-0	03	VI
116N54C701	Advanced Machine Learning	3-0-0	03	3-0-0	03	VII
116N54P801	Applied Project/ Internship	0 - 4 - 0	04	0-2-0	02	VIII
	Total	15 - 10 - 0	25	15 – 5 – 0	20	

Examination Scheme

		Examination Scheme									
Course	Course Name	Marks									
Code	Course Name	CA		ESE	TW	\mathbf{o}^*	P	P&O	Total		
		ISE	IA	ESE	1 11			140	Total		
116N54C301	Data Visualization	30	20	50	25	25	-	-	150		
116N54C401	Applied Data Science	30	20	50	25	25	-	-	150		
116N54C501	Data Analytics	30	20	50	25	25	-	-	150		
116N54C601	Advanced Data Mining	30	20	50	-	-	-	-	100		
116N54C701	Advanced Machine Learning	30	20	50	-	-	-	-	100		
116N54P801	Applied Project/ Internship	-	_	_	50	50	-	-	100		
	Total	150	100	250	125	125			750		

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

Course Code		Course Title							
116N54C301		Data Visualization							
	ŗ	TH			P		TUT	Total	
Teaching Scheme(Hrs.)			02				05		
Credits Assigned		03		01				04	
		Marks							
Examination	CA	CA		TEXX 7		D	De O	Total	
Scheme	ISE	IA	ESE	TW	0	P	P&O	1 Otal	
	30	20	50	25	25			150	

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	Unit	Details	Hrs.	CO
No.	No.			
1	Introd	duction data visualization		
	1.1			
		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	10	CO1
	1.2	Data Fundamentals, Collecting data, Preparing Data	10	CO2
	1.3	Introduction to basic Visualization and its need.		
		The Bar Chart		
		The Line Chart		
		The Pie and Donut Charts		
		The Scatter Plot		
		The Packed Bubble Chart		
	#self-l	Learning Topic: Box plot, The Treemap, Plotting of PDF		
2	Design			
	2.1	Design principles Categorical, time series, and statistical	06	CO 3
		data graphics		
3	Story	telling and Multivariate displays		
	3.1			
		The Science of Storytelling		
		The Power of Stories		
		Context in Action		
		Exploratory versus Explanatory Analysis	10	CO4
		Structuring Stories		
		Audience Analysis for Storytelling		
		Steps to Visual Data Storytelling		
		The Important Role of Feedback		
	3.2	Graphical Perception		
	# Mul	tivariate displays		
4	Geosp	patial 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	Dashb	poards, interactive and animated displays		
	5.1	Visual Design Building Blocks	10	CO4
		Color		

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

	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		
#case	study, Publishing dashboard over mobile devices		
	Total	45	

Term work:

Term-Work consists of problems / programming assignments covering entire syllabus. Students will be graded based on continuous assessment of their term work.

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

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

Course Code	Course Title							
116N54C401	Applied Data Science							
	TH			P		TUT		Total
Teaching Scheme(Hrs.)			02				05	
Credits Assigned		03		01				04
	Marks							
Examination	CA		ESE	TW	0	P	De O	Total
Scheme	ISE	IA	ESE	1 44	U	P	P&O	Total
	30	20	50	25	25	-		150

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	Implementation of basic statistical modeling and analysis using R Programming.
CO3	Application of Exploratory data analysis (EDA) on Real world problems.
CO4	Understand the basic concept and techniques of Machine Learning.
CO5	Describe the Data Science Process and how its components interact.

Module	Unit	Details	Hrs.	CO		
No.	No.					
1	Introd	10	CO1			
	Proce	ss	10	COI		
	1.1	Introduction to Applied Data Science: What is Data				
		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				
	1.4	Data Scrapping: Introduction, Need, Sources, Web				
		Scrapping, Scrapping of Images, Data Wrangling, ETL				
		Process				
2	Open	Source: Modelling in R	10	CO2		
	2.1	System Commands, Loading Data, Matrices, Descriptive				
		Statistics, Higher-Order Moments, Quick Introduction to				
		Brownian Motions with R, Estimation using maximum-				
		likelihood, GARCH/ARCH Models.				
	2.2	Introduction to Monte Carlo, Portfolio Computations in				
		R, Finding the Optimal Portfolio, Root Solving,				
		Regression, Heteroskedasticity, Auto-regressive models,				
		Vector Auto-Regression, Logit, Probit, Solving				
		Nonlinear Equations, Web-Enabling R Functions				
3	Explo	10	CO3			
	3.1	Exploratory Data Analysis and the Data Science Process,				
		Basic tools (plots, graphs and summary statistics) of				
		EDA, Data Pre-processing, Features Identification, Data				
		Munging, Normal Distribution, Skewness and Kurtosis				
	3.2	The Data Science Process: Case Study on Online E-				
4	Introd	03	CO4			
	for data modelling					
	4.1	Introduction to the first level Machine level algorithms				
		useful for data science: Linear Regression, Logistic				
		Regression, k-Nearest Neighbors (k-NN), k-means,				
		Naive Bayes for predictive modelling				

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

5 Dat	a Communication and Information Visualization	12	C05
5.1	Data Communication: cost Function, how to Minimize		
	cost function, coefficients of determination.		
	Information visualization: effective information		
	visualization, visual Encodings, perception of visual		
	cues, data scales, visualizing time series data, data		
	journalism, dashboards, Feature Selection algorithms -		
	Filters, Wrappers, Decision Trees, Random Forests.		
5.2	Using Twitter, Using Facebook, Text processing, plain		
	and simple, A Multipurpose Function to Extract Text,		
	Text Classification, Bayes Classifier, Support Vector		
	Machines.		
5.3	Metrics, Confusion Matrix, Precision and Recall,		
	Accuracy, False Positives, Sentiment Error,		
	Disagreement, Correlations, Aggregation Performance,		
	Phase-Lag Metrics, Economic Significance, Grading		
	Text, 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.

Term work:

Term-Work consists of problems / programming assignments covering entire syllabus. Students will be graded based on continuous assessment of their term work.

Sr.	Name/s of Author/s	Title of Book	Name of	Edition and
No.			Publisher with	Year of
			country	Publication
1.	Han, J., Kamber, M.,	Data mining concepts and	Morgan	2011
	Pei, J.	techniques	Kaufmann	
2.	James, G., Witten,	An introduction	Springer	2013
	D., Hastie, T.,	to statistical		
	Tibshirani, R.	learning with applications		
		in R		
3.	Cathy O'Neil and	Doing Data Science,	O'Reilly	2014
	Rachel Schutt	Straight Talk From The		
		Frontline		
4.	Kevin P. Murphy	Machine Learning: A	ISBN	2013
		Probabilistic Perspective	0262018020	
5.	Mohammed J. Zaki	Data Mining and Analysis:	Cambridg	2014
	and Wagner Miera Jr.	Fundamental	e	
		Concept	Universit	
		S	у	
		and Algorithms	Press	
6.	Avrim Blum, John	Foundations of	ONLINE	2014
	Hopcroft, and	Data		
	RavindranKannan	Science		

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

Course Code		Course Title						
116N54C501			Data	a Analy	tics			
	П	Н		I		,	TUT	Total
Teaching Scheme(Hrs.)	03			02				05
Credits Assigned		03		01				04
	Marks							
Examination	CA	CA		TW	0	D	P&O	Total
Scheme	ISE	IA	ESE	1 //	U	P	180	1 Otal
	30	20	50	25	25			150

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	Experiment using advanced software techniques and tools to conduct thorough
	and insightful analysis
CO3	Apply software processing techniques to prepare statistical modeling for
	geospatial healthcare and text data
CO4	Synthesize the results for control and to draw inferences

Module	Unit	Details	Hrs.	CO
No.	No.			
1		duction to Data Analytics	08	CO1
	1.1	Introduction to Data Analytics, Importance of Data		
		analytics, Impact of data analytics in business applications		
	1.2	Different types of data analytics: Descriptive analytics,		
	1.2	Diagnostics Analytics, Predictive analytics,		
		Prescriptive analytics		
	1.3	Introduction, Terminologies used in data Analytics,		
		types of Data, Quantitative data, Qualitative data.		
	- .	Normal Distribution of data		000
2		ductions to Basic Steps of Data Analytics	07	CO2
	2.1	Introduction to basic steps of spatial data analytics, healthcare data analytics, text mining data analytics		
	2.2	Introduction to analytical software tools and		
	2.2	technologies used by Google and Amazon		
		applications.		
		# Self-Learning: LinkedIn analytics, Netflix Analytics,		
3	Data A	Analytics in GIS	12	
	3.1	Perspectives of spatial data science: business,		
		technology, and data. DBMS for GIS and knowledge		CO2
		base creation using big data, Technologies for GIS		
	2.0	data analytics applications		
	3.2	Introduction to open source software tools - QGIS, Hadoop, GeoSpark R		CO3
	3.3	GIS application Case study: A real world problem and		
		its step by step procedure using open source software		CO3
		tools.		
		# Self-learning: PostgreSQL, PostGIS, Python		
4		analytics in Health Care Systems	10	CO3
	4.1	Components of EHR- Coding Systems- Benefits of		
		EHR- Barrier to Adopting HER challenges		
	4.2	Phenotyping Algorithms. Collection and preprocessing of sensor data for		
	7.2	healthcare applications		
		Natural Language Processing and data mining for		
		clinical text data, biomedical data		
		Introduction to Social media analytics for healthcare		
		solutions		
	4.3	Healthcare application- Case study: A real world		
		problem and its step by step procedure using open		
		source software tools.		
5		h Analytics	08	C04
	5.1	Introduction to the Social Network. Mining Social-		
		Network Graphs.		

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

	5.2	Graph Algorithms and real time application. GraphX tools of Apache.			
		To	otal	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.

Term work:

Term-Work consists of problems / programming assignments covering entire syllabus. Students will be graded based on continuous assessment of their term work.

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	Data Analytics	Mc Graw Hill	2017
3.	James, G., Witten, D., Hastie, T., Tibshirani, R.	An introductionto statistical learning with applications in R	Springer	2013
4.	Chandan K. Reddy and Charu C Aggarwal	Healthcare data analytics	Taylor & Francis	2015
5.	Hui Yang and Eva K. Lee	Healthcare Analytics: From Data to Knowledge to Healthcare Improvement	Wiley	2016
6.	Mohammed J. Zaki and Wagner Miera Jr.	Data Mining and Analysis: Fundamental Concepts and Algorithms	Cambridge University Press	2014
7.	U. Dinesh Kumar	Business Analytics	Wiley	2017

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

Course Code	Course Title							
116N54C601		Advanced Data mining						
	Г	TH		P	•	,	TUT	Total
Teaching Scheme(Hrs.)	03							03
Credits Assigned	03							03
	Marks							
Examination	CA	CA		FENNY.	0	P	P&O	Total
Scheme	ISE	IA	ESE	TW	U	P	rau	1 otai
	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	Unit	Details	Hrs.	CO
No.	No.			
		Advanced Data Mining		
1	Data 1	nining Introduction		
	differe	mining process, different types of data representation, ent types of knowledge mined, common data mining tasks, buted computing solution for data mining and applications	03	CO1
2		nental Datamining and Stream mining		
	Chara	nental algorithms for mining frequent patterns acteristics of Streaming Data, Issues and Challenges, ning Data Mining Algorithms	06	CO2
3	Minin	g complex structures		
	for mining mining Mining network	g trees- Tree Model Guided Framework, TMG framework ning ordered & unordered subtrees, Tree Mining rations, Mining maximal and closed frequent trees, Tree g application g Graphs- Approaches to graph mining. Mining social-rk graph	10	CO3
4		nining & Web Search		
	Cluste Web s Rank	Classification, Vector Space Model, Flat and Hierarchical ring search: Crawling & Indexing, Hyperlink Analysis, Page algorithm, Web Search and Information Retrieval, eation: Query Recommender System	08	CO4
5		nce mining & Multivariate and Time series mining		
	Model Applic Multiv and Ti	ing, Sequential Pattern Discovery, Timing Constraints eations in Bioinformatics variate and Time series mining- Importance of Multivariate me series data, Sources of MVTS data, Mining MVTS Sign Language Data, Agro-meteorological Data	08	CO5

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

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

Course Code	Course Title							
116N54C701		Advanced Machine Learning						
	r	TH			P		TUT	Total
Teaching Scheme(Hrs.)	03							03
Credits Assigned		03						03
	Marks							
Examination	CA	CA		TW		Ъ	D.C.O.	Total
Scheme	ISE	IA	ESE	1 W	O	P	P&O	1 otai
	30	20	50					100

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

Course Objectives:

- 1. To analyse various machine learning algorithms and techniques with a modern outlook focusing on recent advances.
- 2. Explore supervised and unsupervised learning paradigms of machine learning.
- 3. To understand natural language processing and to learn how to apply basic algorithms in this field.
- 4. 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	Unit Details	Hrs.	CO
No.	No.		
1	Machine learning foundation: What is Machine Learning? Types of learning, applications, Bias, variance, overfitting, underfitting, cross validation and feature engineering, gradient descent learning algorithm and its	05	CO1
2	variations	<u> </u>	
2	Deep learning fundamentals: 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.	10	CO2
3	Deep Learning Architecture:		
	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	10	CO3
4	Deep learning for NLP: 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	10	CO4
5	Machine learning case studies:		
	 Churn Analysis and Prediction Sentiment Analysis or Topic Mining from New York Times Customer Segmentation and Value Netflix Movie Recommendation System Image & text classification using TenserFlow & PyTorch. Transfer learning with Pre-trained CNN using TenserFlow. 	10	CO5
	Total	45	

Sr. No.	Name/s of Author/s	Title of Book	Name of Publisher with country	Edition and Year of
				Publication
6.	Tom M.Mitchell	Machine Learning McGraw Hill		2017
7.	M. Gopal	Applied Machine McGraw Hill Learning		2018
8.	Ian Goodfellow, Yoshua Bengio, Aaron Courville	Deep Learning	An MIT Press book	2016
9.	Deng & Yu	Deep Learning: Methods and Applications	Now Publishers	2013
10.	Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta and Harshit Surana	Practical Natural Language Processing	O'Reilly Media, Inc.	June 2020
11.	Sudharsan Ravichandiran	Hands-On Deep Learning Algorithms with Python	Packt Publishing	July 2019

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

Course Code	Course Title									
116N54P801	Applied Project/Internship									
	ТН			P		TUT		Total		
Teaching Scheme (Hrs./Week)	-			04			-		04	
Credits Assigned		-			02		-		02	
	Marks									
Examination	CA			ECE	TW		P	P&O	Total	
Scheme]	ISE IA		ESE	1 44	O	Г	rau	Total	
	-	-	-	-	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.