



# Specialization/Minor in Data Science

EFFECTIVE FOR 2021-22 BATCH

2<sup>ND</sup> YEAR TO 4<sup>TH</sup> YEAR

## Eligible Branches to adopt as Specialization

1. B.Tech.- Computer Science & Engineering
2. B.Tech.- Electronics and Communication Engineering
3. B.Tech.- Electronics Engineering



**VEER MADHO SINGH BHANDARI  
UTTARAKHAND TECHNICAL UNIVERSITY DEHRADUN**

**Evaluation Schemes for Specializations/Minor in B.Tech**

<b>Specialization in Data Science</b>										
S.N	Code	Sem	Subject	Periods			Evaluation Scheme		Total Marks	Credits
				L	T	P	Internal	External		
1.	SDS301	3 <sup>rd</sup>	INFORMATION MANAGEMENT	3	0	0	50	100	150	3
2.	SDS401	4 <sup>th</sup>	Scalable Data Science	3	0	0	50	100	150	3
3.	SDS501	5 <sup>th</sup>	Data Science for Engineers	3	0	0	50	100	150	3
4.	SDS601	6 <sup>th</sup>	Business Analytics and data mining Modeling using R	3	0	0	50	100	150	3
5.	SDS701	7 <sup>th</sup>	DATA-VISUALIZATION	3	0	0	50	100	150	3
6.	SDS801	8 <sup>th</sup>	Big Data Analysis	3	0	0	50	100	150	3
<b>Total</b>				<b>18</b>	<b>0</b>	<b>0</b>	<b>300</b>	<b>600</b>	<b>900</b>	<b>18</b>



**VEER MADHO SINGH BHANDARI  
UTTARAKHAND TECHNICAL UNIVERSITY DEHRADUN**

<b>SDS301</b>	<b>INFORMATION MANAGEMENT</b>	<b>L</b>	<b>T</b>	<b>P</b>	<b>C</b>
		3	0	0	3

<b>Contents</b>		<b>Hours</b>
<b>Unit 1</b>	DATABASE MODELLING, MANAGEMENT AND DEVELOPMENT Database design and modelling - Business Rules and Relationship; Java database Connectivity (JDBC), Database connection Manager, Stored Procedures. Trends in Big Data systems including NoSQL - Hadoop HDFS, MapReduce, Hive, and enhancements.	8
<b>Unit 2</b>	DATA SECURITY AND PRIVACY Program Security, Malicious code and controls against threats; OS level protection; Security – Firewalls, Network Security Intrusion detection systems. Data Privacy principles. Data Privacy Laws and compliance.	8
<b>Unit 3</b>	INFORMATION GOVERNANCE Master Data Management (MDM) – Overview, Need for MDM, Privacy, regulatory requirements and compliance. Data Governance – Synchronization and data quality management.	8
<b>Unit 4</b>	INFORMATION ARCHITECTURE Principles of Information architecture and framework, Organizing information, Navigation systems and Labelling systems, Conceptual design, Granularity of Content.	8
<b>Unit 5</b>	INFORMATION LIFECYCLE MANAGEMENT Data retention policies; Confidential and Sensitive data handling, lifecycle management costs. Archive data using Hadoop; Testing and delivering big data applications for performance and functionality; Challenges with data administration.	

**Suggested Readings :**

1. Data Science For Cyber-security, by Adams Niall M, Heard Nicholas A, Rubin-delanchy Patrick, Turcotte Mellisa
2. Research Methods for Cyber Security, by Thomas W. Edgar, David O. Manz
3. Cybersecurity: The Beginner's Guide: A comprehensive guide to getting. by [Erdal Ozkaya](#)



**VEER MADHO SINGH BHANDARI  
UTTARAKHAND TECHNICAL UNIVERSITY DEHRADUN**

<b>SDS401</b>	<b>SCALABLE DATA SCIENCE</b>	<b>L</b>	<b>T</b>	<b>P</b>	<b>C</b>
		<b>3</b>	<b>0</b>	<b>0</b>	<b>3</b>

<b>Contents</b>		<b>Hours</b>
<b>Unit 1</b>	Background: Introduction Probability: Concentration inequalities Linear algebra: PCA, SVD Optimization: Basics, Convex, GD Machine Learning: Supervised, generalization, feature learning, clustering.	8
<b>Unit 2</b>	Memory-efficient data structures: Hash functions, universal / perfect hash families Bloom Filters Sketches for distinct count Misra-Gries sketch Statistical Mechanics an overview.	8
<b>Unit 3</b>	Memory-efficient data structures (contd.): Count Sketch, Count-Min Sketch   Approximate near neighbors search: Introduction, kd-trees etc LSH families, MinHash for Jaccard, SimHash for L2	8
<b>Unit 4</b>	Randomized Numerical Linear Algebra CUR Decomposition Sparse RP, Subspace RP, Kitchen Sink.	8
<b>Unit 5</b>	Map-reduce and related paradigms Map reduce - Programming examples - (page rank, k-means, matrix Multiplication) Big data: computation goes to data. + Hadoop ecosystem	

**Suggested Readings:**

1. **Data Science from Scratch: First Principles with Python, By Joel Grus.**
2. **Python for Data Science For Dummies, By John Paul Mueller, Luca Massaron**
3. **Data Analytics , by Anil Maheshwari**



**VEER MADHO SINGH BHANDARI  
UTTARAKHAND TECHNICAL UNIVERSITY DEHRADUN**

<b>SDS501</b>	<b>DATA SCIENCE FOR ENGINEERS</b>	<b>L</b>	<b>T</b>	<b>P</b>	<b>C</b>
		3	0	0	3

<b>Contents</b>		<b>Hours</b>
<b>Unit 1</b>	Linear algebra for data science (algebraic view - vectors, matrices, product of matrix & vector, rank, null space, solution of over-determined set of equations and pseudo-inverse)	8
<b>Unit 2</b>	Linear algebra for data science (geometric view - vectors, distance, projections, eigenvalue decomposition).	10
<b>Unit 3</b>	Statistics (descriptive statistics, notion of probability, distributions, mean, variance, covariance, covariance matrix).	8
<b>Unit 4</b>	Optimization; Typology of data Science problems and a solution framework, Univariate and multivariate linear regression Model assessment (including cross validation).	10
<b>Unit 5</b>	Verifying assumptions used in linear regression, assessing importance of different variables, subset selection, Introduction to classification and classification using logistics regression, Classification using various clustering techniques	9

**Suggested Readings:**

- 1. Data Science and Big Data Analytics: ACM-WIR 2018 (Lecture Notes on Data Engineering and Communications Technologies) , by Durgesh Kumar Mishra, Xin-She Yang, et al.**
- 2. Introducing Data Science: Big Data, Machine Learning, and More, Using Python Tools , by Davy Cielen, Arno D.B. Meysman,**
- 3. Data Science and Big Data Analytics: ACM-WIR , by Durgesh Kumar Mishra, Xin-She Yang.**



**VEER MADHO SINGH BHANDARI  
UTTARAKHAND TECHNICAL UNIVERSITY DEHRADUN**

<b>SDS601</b>	<b>BUSINESS ANALYTICS AND DATA MINING MODELING USING R</b>	<b>L</b>	<b>T</b>	<b>P</b>	<b>C</b>
		3	0	0	3

<b>Contents</b>		<b>Hours</b>
<b>Unit 1</b>	General Overview of Data Mining and its Components Introduction and Data Mining Process Introduction to R Basic Statistical Techniques. Data Preparation and Exploration Visualization Techniques.	8
<b>Unit 2</b>	Data Preparation and Exploration Visualization Techniques Dimension Reduction Techniques Principal Component Analysis, Performance Metrics and Assessment Performance Metrics for Prediction and Classification.	12
<b>Unit 3</b>	Supervised Learning Methods Multiple Linear Regression, Supervised Learning Methods Naïve Bayes, Supervised Learning Methods Classification & Regression Trees, Supervised Learning Methods Logistic Regression	8
<b>Unit 4</b>	Supervised Learning Methods Logistic Regression Artificial Neural Networks. Supervised Learning Methods and Wrap Up Artificial Neural Networks. Discriminate Analysis Conclusion	8

**Suggested Readings:**

1. Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data , by EMC Education Services.
2. Practical Data Science with R Paperback, by Nina Zumel
3. Introducing Data Science: Big Data, Machine Learning, and More, Using Python Tools , by Davy Cielen, Arno D.B. Meysman.



# VEER MADHO SINGH BHANDARI UTTARAKHAND TECHNICAL UNIVERSITY DEHRADUN

<b>SDS701</b>	<b>DATA-VISUALIZATION</b>	<b>L</b>	<b>T</b>	<b>P</b>	<b>C</b>
		3	0	0	3

<b>Contents</b>		<b>Hours</b>
<b>Unit 1</b>	Overview of Data Visualization, Introduction to Web Technologies Why Visualize Data, Introduction to SVG and CSS, Introduction to JavaScript, Introduction to VizHub, Making a Face with D3.	10
<b>Unit 2</b>	The Shapes of Data, Marks and Channels Input for Visualization: Data and Tasks, Loading and Parsing Data with D3.js, Encoding Data with Marks and Channels, Rendering Marks and Channels with D3.js and SVG, Introduction to D3 Scales, Creating a Scatter Plot with D3.	12
<b>Unit 3</b>	Common Visualization Idioms and Visualization of Spatial Data, Networks, and Trees Reusable Dynamic Components using the General Update Pattern: Reusable Scatter Plot, Common Visualization Idioms with D3.js, Bar Chart, Vertical & Horizontal, Pie Chart and Coxcomb Plot, Line Chart, Area Chart.	10
<b>Unit 4</b>	Using Color and Size in Visualization Encoding Data using Color, Encoding Data using Size, Stacked & Grouped Bar Chart, Stacked Area Chart & Stream graph, Line Chart with Multiple Lines.	8
<b>Unit 5</b>	Interaction Techniques and Multiple Linked Views Adding interaction with Unidirectional Data Flow, Using UI elements to control a scatter plot, Panning and Zooming on a Globe, Adding tooltips, Small Multiples, Linked Highlighting with Brushing, Linked Navigation: Bird's Eye Map.	

**Suggested Readings :**

1. **Data Science and Big Data Analytics: ACM-WIR** , by Durgesh Kumar Mishra, Xin-She Yang.
2. **Practical Data Science with R Paperback**, by Nina Zumel
3. **Data Science from Scratch: First Principles with Python**, By Joel Grus.



# VEER MADHO SINGH BHANDARI UTTARAKHAND TECHNICAL UNIVERSITY DEHRADUN

<b>SDS801</b>	<b>BIG DATA ANALYSIS</b>	<b>L</b>	<b>T</b>	<b>P</b>	<b>C</b>
		3	1	0	4

<b>Course Contents</b>		<b>Hours</b>
<b>UNIT I:</b>	<b>Introduction to Big Data:</b> Big Data and its Importance – Four V’s of Big Data – Drivers for Big Data –Introduction to Big Data Analytics – Big Data Analytics applications.	<b>8hrs</b>
<b>UNIT II:</b>	<b>Big Data Technologies:</b> Hadoop’s Parallel World – Data discovery – Open source technology for Big Data Analytics – cloud and Big Data –Predictive Analytics – Mobile Business Intelligence and Big Data – Crowd Sourcing Analytics – Inter- and Trans-Firewall Analytics - Information Management.	<b>8hrs</b>
<b>UNIT III:</b>	<b>Processing Big Data:</b> Integrating Disparate Data Stores - Mapping Data To The Programming Framework- Connecting And Extracting Data From Storage - Transforming Data For Processing - Subdividing Data In Preparation For Hadoop Map Reduce.	<b>8hrs</b>
<b>UNIT IV:</b>	<b>Hadoop Map reduce:</b> Employing Hadoop Map Reduce - Creating the components of Hadoop Map Reduce jobs - Distributing data processing across server farms -Executing Hadoop Map Reduce jobs - Monitoring the progress of job flows - The Building Blocks of Hadoop Map Reduce - Distinguishing Hadoop daemons -Investigating the Hadoop Distributed File System Selecting appropriate execution modes: local, pseudo-distributed, fully distributed.	<b>12hrs</b>
<b>UNIT V:</b>	<b>Advanced Analytics Platform:</b> Real-Time Architecture – Orchestration and Synthesis Using Analytics Engines– Discovery using Data at Rest – Implementation of Big Data Analytics – Big Data Convergence – Analytics Business Maturity Model. Hadoop Eco-System: Pig – Installing and Running , Comparison with Databases – Pig Latin – User-Define Functions – Data Processing Operators – Installing and Running Hive– Hive QL – Tables – Querying Data – User-Defined Functions. Fundamentals of H Base and Zoo Keeper - IBM Info Sphere Big Insights and Streams. Visualizations - Visual data analysis techniques, Interaction techniques; Systems and applications.	<b>12hrs</b>

**Text Books:**

1. Michael Minelli, Michehe Chambers, “Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today’s Business”, 1st Edition, Ambiga Dhiraj, Wiely CIO Series, 2013.
2. Arvind Sathi, “Big Data Analytics: Disruptive Technologies for Changing the Game” 1<sup>st</sup> Edition, IBM Corporation, 2012.
3. Bill Franks, “Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data. Streams with Advanced Analytics”, 1st Edition, Wiley and SAS Business Series, 2012.