Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore

Shri Vaishnav Institute of Information Technology

Choice Based Credit System (CBCS) in the light of NEP-2020

B.Tech (CSE-Data Science IBM)

SEMESTER-V(2023-2027)

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
BTCS501N	DCC	Theory of Computation	60	20	20	0	0	3	1	0	4

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

COURSE OBJECTIVES:

The student will have ability to:

1. To introduce concepts in automata theory and theory of computation.
2. To identify different formal language classes and their relationships.
3. To design grammars and recognizers for different formal languages.

COURSE OUTCOMES:

Upon completion of the subject, students will be able to:

1. Ability to relate practical problems to languages, automata, and computability.
2. Ability to demonstrate an increased level of mathematical sophistication.
3. Ability to apply mathematical and formal techniques for solving problems.

SYLLABUS

UNIT I — 10 HOURS

Introduction: Alphabets, Strings and Languages; Automata and Grammars, Deterministic finite Automata (DFA)-Formal Definition, Simplified notation: State transition graph, Transition table, Language of DFA, Nondeterministic finite Automata (NFA), NFA with epsilon transition, Language of NFA, Equivalence of NFA and DFA, Minimization of Finite Automata, Distinguishing one string from other, Myhill-Nerode Theorem.

UNIT II — 9 HOURS

Regular Expression (RE): Definition, Operators of regular expression and their precedence, Algebraic laws for Regular expressions, Kleen's Theorem, Regular expression to FA, DFA to Regular expression, Arden's Theorem, Non-Regular Languages, Pumping Lemma for regular Languages. Application of Pumping Lemma, Closure properties of Regular Languages, Decision properties of Regular Languages, FA with output: Moore and Mealy machine, Equivalence of Moore and Mealy Machine, Applications and Limitation of FA.

UNIT III — 8 HOURS

Context Free Grammar (CFG) and Context Free Languages (CFL): Definition, Examples, Derivation, Derivation trees, Ambiguity in Grammar, Inherent ambiguity, Ambiguous to Unambiguous CFG, Useless symbols, Simplification of CFGs, Normal forms for CFGs: CNF and GNF, Closure properties of CFLs, Decision Properties of CFLs: Emptiness, Finiteness and Membership, Pumping lemma for CFLs.

UNIT IV — 8 HOURS

Push Down Automata (PDA): Description and definition, Instantaneous Description, Language of PDA, Acceptance by Final state, Acceptance by empty stack, Deterministic PDA, Equivalence of PDA and CFG, CFG to

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore
Shri Vaishnav Institute of Information Technology
Choice Based Credit System (CBCS) in the light of NEP-2020
B.Tech (CSE-Data Science IBM)
SEMESTER-V(2023-2027)

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
BTCS501N	**DCC**	**Theory of Computation**	**60**	**20**	**20**	**0**	**0**	**3**	**1**	**0**	**4**

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

PDA and PDA to CFG.

UNIT V **10 HOURS**

Turing machines (TM): Basic model, definition and representation, Instantaneous Description, Language acceptance by TM, Variants of Turing Machine, TM as Computer of Integer functions, Universal TM, Church's Thesis, Recursive and recursively enumerable languages, Halting problem, Introduction to undecidability, undecidable problems about TM, NP hard and NP complete problem, Post correspondence problem (PCP), Modified PCP, Introduction to recursive function theory.

TEXTBOOKS:

1. J. E. Hopcraft, R. Motwani and J. D. Ullman, Introduction to Automata Theory, Languages, and Computation, 3rd Ed., Pearson, 2013.
2. P. Linz, S. H. Rodger, An Introduction to Formal Languages and Automata, 7th Ed., Jones & Bartlett Learning, 2023.

REFERENCE:

1. J. C. Martin, Introduction to Languages and Theory of Computations, 4th Ed., Tata McGraw Hill, 2010.
2. C. Papadimitriou, and C. L. Lewis, Elements of the Theory of Computation, PHI, 1997.
3. Michael Sipser, Introduction to Theory of Computation, 3th Ed., Cengage Learning, 2013.
4. K. L. P Mishra & N. Chandrasekaran, Theory of Computer Science, 3th Ed., PHI Learning, 2006

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore

Shri Vaishnav Institute of Information Technology

Choice Based Credit System (CBCS) in the light of NEP-2020

B.Tech (CSE-Data Science IBM)

SEMESTER-V(2023-2027)

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
BTIBM507N	DCC	Fundamentals of Data Science	60	20	20	30	20	3	0	2	4

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

COURSE OBJECTIVES:

The student will have ability to:

1. Fundamental principles and methodologies in data science.
2. Understand the core concepts of databases, and learn about Python Libraries.
3. Acquire skills for data collection, cleaning, analysis, interpretation and gain proficiency in visualizing data to extract meaningful insights.
4. Understand various machine learning algorithms and their applications.
5. Apply data science techniques to solve practical problems and make data-driven decisions.

COURSE OUTCOMES:

After the successful completion of this course students will be able to:

1. Acquire data from various sources, including databases, APIs, and web scraping
2. Clean and preprocess raw data to remove errors, outliers, and inconsistencies.
3. Perform EDA techniques to understand data distributions, relationships, and patterns.
4. Apply statistical methods to analyze data distributions, correlations, and make data-driven decisions.
5. Implement machine learning algorithms for regression, classification, and predictive modeling.

SYLLABUS

UNIT I — 7 HOURS

Overview of data science and its applications, Historical context and evolution of data science, Tools for Data Science, Data Science Methodology: CRISP-DM Importance of data-driven decision-making

UNIT II — 8 HOURS

Introduction to Python, Apply Python programming logic Variables, Data Structures, Python libraries such as Pandas & Numpy, and develop code using Jupyter Notebooks Databases and SQL for Data Science with Python: Analyze data within a database using SQL and Python, Create a relational database and work with multiple tables using DDL commands Construct basic to intermediate level SQL queries using DML commands

UNIT III — 8 HOURS

Exploratory Data Analysis (EDA) techniques, Descriptive statistics (mean, median, mode, variance, etc.), Data visualization libraries and tools (Matplotlib, Seaborn), Visualization techniques (scatter plots, histograms, box plots, pairplot,heatmap,violinplot) Data Preprocessing and Feature Engineering: Data preprocessing techniques (scaling, encoding categorical variables, feature selection),Feature engineering methods (creating new features, handling datetime data)

UNIT IV — 7 HOURS

Machine Learning Fundamentals:

Describe the various types of Machine Learning algorithms and when to use them, Compare and contrast linear

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
BTIBM507N	DCC	Fundamentals of Data Science	60	20	20	30	20	3	0	2	4

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

classification methods including multiclass prediction, support vector machines, and logistic regression Write Python code that implements various classification techniques including K-Nearest neighbors (KNN), decision trees, and regression trees Evaluate the results from simple linear, non-linear, and multiple regression on a data set using evaluation metrics

UNIT V **8 HOURS**

Exploring Cutting-Edge Machine Learning: Advanced Topics and Techniques:

Dimensionality reduction techniques (PCA, t-SNE), Ensemble learning methods (random forests, bagging, boosting,), Introduction to deep learning (neural networks, convolutional neural networks, recurrent neural networks) Capstone Project: Application of data science techniques to solve a real-world problem
Project planning, execution, and presentation

TEXTBOOKS AND REFERENCE:

1. Jake VanderPlas, “Python Data Science Handbook”.2012
2. Wes McKinney,"Python for Data Analysis".2012
3. John Paul Mueller, Luca Massaron, “ Python for Data Science For Dummies”.2015
4. Aurélien Géron,"Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow".2017

LIST OF PRACTICALS

1. Write Python programs to demonstrate the use of variables, lists, tuples, dictionaries, and loops for data manipulation.
2. To get hands-on experience with core Python libraries for data science
3. To implement and understand different types of SQL commands such as DDL for database structure creation and DML for data manipulation
4. To create a small dataset and perform the data cleaning process by handling missing values, removing inconsistencies, and preparing the data for analysis.
5. Perform EDA using Matplotlib and Seaborn to visualize the data and identify trends, correlations, and outliers by using different types of visualization tools.
6. Perform data preprocessing techniques on the datasets, such as scaling (Min-Max, Standard), encoding (Label, One-Hot), and feature selection.

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore

Shri Vaishnav Institute of Information Technology

Choice Based Credit System (CBCS) in the light of NEP-2020

B.Tech (CSE-Data Science IBM)

SEMESTER-V(2023-2027)

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
BTIBM507N	**DCC**	**Fundamentals of Data Science**	**60**	**20**	**20**	**30**	**20**	**3**	**0**	**2**	**4**

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

7. To Implement and Evaluate Machine Learning Algorithms, such as Linear Regression and Logistic Regression, on a Sample Dataset.
8. Implement supervised learning algorithms, such as Linear Regression, Logistic Regression, KNN, and Decision Tree, on a sample dataset, and evaluate the accuracy of the models.
9. Apply dimensionality reduction (PCA) and ensemble learning techniques, such as Random Forest and Gradient Boosting, on a dataset.
10. Develop a project covering the entire workflow, including data collection, preprocessing, analysis, model building, and visualization of results.

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore

Shri Vaishnav Institute of Information Technology

Choice Based Credit System (CBCS) in the light of NEP-2020

B.Tech (CSE-Data Science IBM)

SEMESTER-V(2023-2027)

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
BTIBM504N	**DCC**	**Reactive Architecture**	**60**	**20**	**20**	**30**	**20**	**3**	**0**	**2**	**4**

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

COURSE OBJECTIVES:

The student will have ability to:

1. To learn the fundamentals of reactive architecture
2. To understand where and why reactive systems are applicable
3. To gain knowledge about Akka ToolKit

COURSE OUTCOMES:

Upon completion of the subject, students will be able to

1. To describe the detailed structure of reactive systems.
2. To describe reactive manifesto and reactive principles
3. To understand Actor model, props in Akka tool kit.
4. To implement reactor pattern in node js.

SYLLABUS

UNIT I — 7 HOURS

Why Reactive : What is the problem that Reactive Architecture is attempting to solve,How does unresponsive software impact its users,What is the goal of Reactive Architecture. Reactive Principles,An introduction to the Reactive Manifesto, An explanation of the Reactive Principles.

UNIT II — 8 HOURS

Reactive Toolbox : MultiThreading,The Reactor Pattern,The MultiReacter Pattern,Actor Model,Introduction to Akka Tool Kit,Akka Actor System,Props,Child Actor,Send Actor,Stop Actor,Reply Messages ,Forward Messages.

UNIT III — 6 HOURS

Reactive Systems vs Reactive Programming- What are Reactive Systems, What is Reactive Programming, How are Reactive Systems related to Reactive Programming, The Actor Model and its relationship to Reactive Systems.

UNIT IV — 9 HOURS

Putting Your Reactive Toolbox to Work-

Going from Services to Systems: Being Message Driven
Distributed Infrastructure
Orchestrated Cloud Infrastructure
Reactive Meets Machine Learning

Chairperson	Chairperson	Controller of Examination	Registrar
Board of Studies, Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore	Faculty of Studies, Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore	Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore	Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore

Shri Vaishnav Institute of Information Technology

Choice Based Credit System (CBCS) in the light of NEP-2020

B.Tech (CSE-Data Science IBM)

SEMESTER-V(2023-2027)

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
BTIBM504N	**DCC**	**Reactive Architecture**	**60**	**20**	**20**	**30**	**20**	**3**	**0**	**2**	**4**

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

UNIT V **8 HOURS**

Apache Kafta in Reactive Architecture: Asynchronous messaging backbone, Message retention and data persistence, Decoupling, Backpressure, Backpressure in Kafka Consumers, Backpressure in Alpakka Kafka Connector, Backpressure in MicroProfile Reactive Messaging

TEXTBOOKS:

1. IBM Coursware.

REFERENCE:

1. IBM Coursware.

LIST OF PRACTICALS:

1. Create a reactor pattern in node js.
2. Create multireactor pattern in node js.
3. Implement Actor model.
4. Implementations of Akka tell method.
5. Implementation of Akka Ask method.
6. Implementation of stopping top level actor in Akka using stop method.
7. Implement of stopping child actor in Akka.
8. Implementation of stopping Actor Sytem.
9. Implement props in Akka using akka.actor.Props.
10. Write program using Akka toolkit to forward message from one actor to another

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore

Shri Vaishnav Institute of Information Technology

Choice Based Credit System (CBCS) in the light of NEP-2020

B.Tech (CSE-Data Science IBM)

SEMESTER-V(2023-2027)

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
BTIBM511N	DSC	Big Data Technologies	60	20	20	30	20	3	0	2	4

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

COURSE OBJECTIVES:

The student will have ability to:

1. Big Data and its importance in business world
2. Focused on conceptualization and summarization of big data trivial data versus big data.
3. Big data computing technologies, Watson studio
4. Understand the challenges posed by distributed applications and how ZooKeeper is designed to Handle.

COURSE OUTCOMES:

After the successful completion of this course students will be able to:

1. Develop an understanding of the complete open-source Hadoop ecosystem and its near term future direction.
2. Understand the functions and features of HDP.
3. Understand the Map Reduce model v1 and review java code.

SYLLABUS

UNIT I — 10 HOURS

Introduction to Big Data: Develop an understanding of the complete open-source Hadoop ecosystem and its near-term future directions, compare and evaluate the major Hadoop distributions and their ecosystem components both their strengths and their limitations, hands-on experience with key components of various big data ecosystem components and roles in building a complete big data, solution to common business problems.

UNIT II — 9 HOURS

Hadoop and HDFS: The basic need for a big data strategy in terms of parallel reading of large data files and internode network speed in a cluster, Hadoop Distributed File System (HDFS), function of the NameNode and DataNodes in a Hadoop cluster, files are stored and blocks ("splits") are replicated. Hive, Sqoop.

UNIT III — 11 HOURS

Introduction to Hortonworks and its components

Apache Ambari: The purpose of Apache Ambari in the HDP stack, the overall architecture of Ambari and Ambari' relation to other services and components of a Hadoop cluster, the functions of the main components of Ambari, initiating start and stop services from Ambari Web Console.

Overview about Hortonworks Data Platform – HDP: The functions and features of HDP, the IBM value-add components, what IBM Watson Studio is, a brief description of the purpose of each of the value-add components.

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
BTIBM511N	DSC	Big Data Technologies	60	20	20	30	20	3	0	2	4

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

UNIT IV 7 HOURS

Data Processing and Management

MapReduce and YARN: MapReduce model v1, the limitations of Hadoop 1 and MapReduce, review the Java code required to handle the Mapper class, Reducer class and the program driver needed to access MapReduce, the YARN model, compare Hadoop 2/YARN with Hadoop 1

UNIT V 8 HOURS

ZooKeeper, Slider, and Knox: The challenges posed by distributed applications and how ZooKeeper is designed to handle them, the role of ZooKeeper within the Apache Hadoop infrastructure and the realm of Big Data management, the generic use cases and some real-world scenarios for ZooKeeper, the ZooKeeper services that are used to manage distributed systems, use the ZooKeeper CLI to interact with ZooKeeper services.

TEXTBOOKS:

1. Introduction to Infosphere BigInsights, IBM Career Education
2. Changing Business with Data Insight, IBM Career Education
3. Big Insights Analytics for Business Analysts, IBM Career Education
4. Tom White," Hadoop: The Definitive Guide Paperback – 2015" Shroff Publishers & Distributers Private Limited - Mumbai; Fourth edition (2015).
5. V. K. Jain (Author)," Big Data and Hadoop" Khanna Publishers; 1 edition (1 June 2015) .

REFERENCE:

1. Big Data: A Revolution That Will Transform How We Live, Work, and Think; Kenneth Cukier, Viktor Mayer-Schönberger; Mariner Books; Edition (2014).
2. Big Data: Using Smart Big Data, Analytics and Metrics to Make Better; Bernard Marr; Wiley; Edition 1st (2015).
3. Hadoop For Dummies, Dirk deRoos, For Dummies, 2014
4. Cohen et al."MAD Skills: New Analysis Practices for Big Data", 2009
5. Ullman, Rajaraman, Mining of Massive Datasets, Chapter 2.
6. Stonebraker et al., "MapReduce and Parallel DBMS's: Friends or Foes?", Communications of the ACM, January 2010
7. Dean and Ghemawat, "MapReduce: A Flexible Data Processing Tool", Communications of the ACM, January 2010.

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore

Shri Vaishnav Institute of Information Technology

Choice Based Credit System (CBCS) in the light of NEP-2020

B.Tech (CSE-Data Science IBM)

SEMESTER-V(2023-2027)

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
BTIBM511N	**DSC**	**Big Data Technologies**	**60**	**20**	**20**	**30**	**20**	**3**	**0**	**2**	**4**

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

LIST OF PRACTICALS

1. Installing Hadoop, configure HDFS, Install Zookeeper , Pig Installation, Sqoop Installation, Hbase Installation
2. Configuring Hadoop
3. Running jobs on Hadoop
4. Working on HDFS
5. Hadoop streaming
6. Creating Mapper function using python.
7. Creating Reducer function using python
8. Python iterator and generators
9. Twitter data sentimental analysis using Flume and Hive
10. Business insights of User usage records of data cards.
11. Wiki page ranking with hadoop
12. Health care Data Management using Apache Hadoop ecosystem

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore
Shri Vaishnav Institute of Information Technology
Choice Based Credit System (CBCS) in the light of NEP-2020
B.Tech (CSE-Data Science IBM)
SEMESTER-V(2023-2027)

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
BTIBM502M	**DCC**	**Fundamentals of Predictive Analytics**	**60**	**20**	**20**	**30**	**20**	**2**	**0**	**2**	**3**

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

COURSE OBJECTIVES:
The student will have ability to:
1. Understand the evolution and relevance of Analytics in the world today
2. Explore end-to-end analytics industry use cases using the data analytics lifecycle
3. Understand the scientific method for analytics, use cases, and the analytics team key roles
4. Acquire technical expertise using popular open source analytics frameworks including Jupyter notebooks and Python.
5. Gain a competitive edge using low-code cloud-based platform for Analytics using IBM Watson Studio.
6. Data engineering and data modeling practices using machine learning.
7. Explore data science industry case studies: transportation, automotive, human resources, aerospace, banking and healthcare
8. Experience teamwork agile industry practices using design thinking.
9. Engage in role-playing challenge-based scenarios to propose real-world solutions.

COURSE OUTCOMES:
After the successful completion of this course students will be able to:
1. Illustrate the interaction of multi-faceted fields like data mining, statistics and mathematics in the development of Predictive Analytics.
2. Acquaint the student with the concepts of Ordinary Least Squares & Generalized Least Squares.
3. Explain data clustering and dimension reduction techniques.

SYLLABUS
UNIT I — 10 HOURS
Introduction to Analytics with: Use Cases Analytics Overview, Domains, Roles, Data Analytics in Practice, Methodologies, Methods, Integrated Environment for Analytics projects, Cloud Based Analytics Lifecycle, Analytics capabilities on the cloud.

UNIT II — 8 HOURS
Explore and Prepare Data: Business Understanding, Explore Data, Prepare Data, Understanding Data, Statistics and Representation Techniques, Data Transformation, Represent and transform Unstructured Data, Data Transformation Tools.

UNIT III — 8 HOURS
Data Visualization and Presentation: Decision-centered visualization, Fundamentals of Visualization, Common graphs, Common tools.

UNIT IV — 9 HOURS

COURSE CODE	CATEGORY	COURSE NAME	TEACHING & EVALUATION SCHEME					L	T	P	CREDITS
			THEORY			PRACTICAL					
			END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*				
BTIBM502M	DCC	Fundamentals of Predictive Analytics	60	20	20	30	20	2	0	2	3

Legends: **L** - Lecture; **T** - Tutorial/Teacher Guided Student Activity; **P** – Practical; **C** - Credit; ***Teacher Assessment** shall be based following components: Quiz/Assignment/ Project/Participation in Class, given that no component shall exceed more than 10 marks.

Data Modeling and Machine Learning Algorithms: Overview of modeling techniques, Machine Learning techniques, Accuracy Precision & recall, Model Deployment.

UNIT V 8 HOURS

Machine Learning Algorithms: About Machine Learning, From Regression to neural nets, Decision tree classifier, Machine learning Framework.

REFERENCE:

1. https://developer.ibm.com/articles/cc-beginner-guide-machine-learning-ai-cognitive.
2. http://bigdatauniversity.com/bdu-wp/ bdu-course/data-science-methodology.
3. Wikipedia, "Cross Industry Standard Process for Data Mining," http:// en.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_ Mining, http://the-modeling-agency.com/crisp-dm.pdf.
4. https://www.ibm.com/blogs/business-analytics/descriptive-analytics-101-what-happened.
5. https://www.weforum.org/agenda/2015/02/a-brief-history-of-big-data-everyone-should-read.
6. https://medium.com/ibm-watson/introducing-ibm-watson-studio-e93638f0bb47.
7. https://keyskill-clms.mylearnerportal.com/mod/lesson/view.php%3Fid=2808.
8. https://www.ibm.com/design/language/experience/data-visualization

LIST OF PRACTICALS

1. Accessing IBM Cloud: Create an IBM account and Navigate to Catalog.
2. Implementing Data Assets from files from your local system, including structured, unstructured data and Images.
3. Implementing Data Cleaning, analyzing and reshaping of data.
4. Visualize preliminary data wrangling result.
5. Implement below hypothesis
   Hypothesis1: Loss Claim After Expired Policy
   Hypothesis2: Loss Claim After Expired License
   Hypothesis3: Excessive (Over $10,000) Claim Amount
6. Hands on to implement Data Refinery Visualization using Claim Datasets.
7. Hands on Lab for building and deploy models using AutoAI.
8. Hands on Lab of Auto Insurance Fraud Analyzed using Jupyter Notebook.
9. Hands on Lab to Analyze Bank Datasets and Hands on Hidden Facebook Usage.
10. Implementing Prediction of wildfire Intensity.

Chairperson	Chairperson	Controller of Examination	Registrar
Board of Studies, Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore	Faculty of Studies, Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore	Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore	Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore