IPS Academy, Institute of Engineering & Science

(A UGC Autonomous Institute, Affiliated to RGPV, Bhopal)
Scheme Based on AICTE Flexible Curriculum

Department of Computer Science & Engineering

Bachelor of Technology (B.Tech.)

[Computer Science & Engineering – Data Science]

Semester VI CSE-DS

S.No.	Course Type	Course Code	Course Title	Hrs./Week			Credits
				L	T	P	
1	PCC	DS13	Big Data and Hadoop	2	1	-	3
2	PCC	DS14	Machine Learning-II	2	1	-	3
3	PCC	DS15	Natural Language Processing	2	1	-	3
4	PEC	DS02	Professional Elective Course -II	2	-	-	2
5	IOC	—---	Interdisciplinary Open Course-I	2	1	-	3
6	HSMC	HS05	Humanities and Social Sciences Open Courses - I	2	-		2
7	LC	DS13(P)	Big Data and Hadoop Lab	-	-	2	1
8	LC	DS14(P)	Machine Learning-II Lab	-	-	2	1
9	PROJ	DS02	Minor Project	-	-	4	2
10	PROJ	DS03	Evaluation of Internship-I	-	-	4	2
11	LLC	LLC03	Liberal Learning Course -III	-	-	2	1
			Internship-II	Credit to be added in Seventh Semester			
12	MLC	MLC04	Intellectual Property Rights	1	-	-	Audit
			Total Credits	12	4	14	23
				28			

Professional Elective Course(PEC) –II DS02(Any One Course)	Interdisciplinary Open Course(IOC)-I,MA01(Any One Course)
(A) Cyber Security	(A) Scientific Aptitude
	(B) Green Technology
(B) Compiler Design*	(C) Operations Research
(C) Information Storage Management	(D) Fundamental of Fire & Safety

Humanities and Social Sciences Open Courses – I, HS05 (Any One Course)
(A) English Language Proficiency
(B) German Language
(C) French Language
(D) Japanese Language
(E) Soft Skills and Interpersonal Communication

IPS Academy, Institute of Engineering & Science

(A UGC Autonomous Institute, Affiliated to RGPV, Bhopal)
Scheme Based on AICTE Flexible Curriculum

Department of Computer Science & Engineering

Bachelor of Technology (B.Tech.)

[Computer Science & Engineering – Data Science-VI Sem]

PCC-DS13	**Big Data and Hadoop**	**2L: 1T(4hrs.)**	**Credits: 3**

Course Objective:

To study the basic technologies that forms the foundations of Big Data.

Module 1: **(07hrs.)**

Introduction to BigData Platform, Traits of Big data, Challenges of Conventional Systems, Web Data, Evolution of Analytic Scalability, Analysis vs Reporting, Statistical Concepts: Sampling Distributions, Re-Sampling, Statistical Inference, Prediction Error.

Module2: **(08 hrs.)**

Need of Hadoop, Data centers and Hadoop Cluster overview, Overview of Hadoop Daemons, Hadoop Cluster and Racks, Learning Linux required for Hadoop, Hadoop ecosystem tools overview, Big data Hadoop opportunities

Module3: **(08hrs.)**

HDFS Daemons – Namenode, Datanode, Secondary Namenode, Hadoop FS and Processing Environment's UIs, Fault Tolerant, High Availability, Block Replication, Hadoop Processing Framework: YARN Daemons – Resource Manager, NodeManager, Job assignment & Execution flow, MapReduce Architecture, MapReduce life cycle, Word Count Example(or) Election Vote Count

Module4: **(08 hrs.)**

Introducing Hadoop Hive, Detailed architecture of Hive, Comparing Hive with Pig and RDBMS, Working with Hive Query Language, Creation of a database, table, group by and other clauses, Various types of Hive tables, HCatalog, Storing the Hive Results, Hive partitioning, and Buckets.

Module5: **(06hrs.)**

Introduction to Hadoop Framework: Spark and Scala, Apache Pig: Advantage of Pig over MapReduce, Pig vs Hive Use case, Introduction to HBASE, Fundamentals of HBase, SQL vs. NOSQL, Application of Sqoop, Flume, Oozie

IPS Academy, Institute of Engineering & Science

(A UGC Autonomous Institute, Affiliated to RGPV, Bhopal)
Scheme Based on AICTE Flexible Curriculum

Department of Computer Science & Engineering

Bachelor of Technology (B.Tech.)

[Computer Science & Engineering – Data Science-VI Sem]

.

Course Outcome:

1: Understand the foundational concepts and challenges of big data and statistical analysis.
2: Recognize the importance of Hadoop and its ecosystem in managing big data.
4: Learn to use Hadoop Hive for data warehousing and SQL-like querying.
5: Explore advanced Hadoop tools and frameworks like Spark, Pig, HBase, and more.

List of Text / Reference Books:

1.Michael Minelli, Michelle Chambers, and Ambiga Dhiraj, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses", Wiley

2.DT Editorial Services, Big-Data Black Book, Wiley

3.Dirk deRoos, Chris Eaton, George Lapis, Paul Zikopoulos, Tom Deutsch, “Understanding Big Data Analytics for Enterprise Class Hadoop and Streaming Data”, McGraw Hill.

4.Thomas Erl, Wajid Khattak, Paul Buhler, “Big Data Fundamentals: Concepts, Drivers and Techniques”, Prentice Hall.

5.Bart Baesens “Analytics in a Big Data World: The Essential Guide to Data Science and its Applications (WILEY Big Data Series)”, John Wiley & Sons

6.ArshdeepBahga, Vijay Madisetti, “Big Data Science & Analytics: A HandsOn Approach “, VPT

7.nand Rajaraman and Jeffrey David Ullman, “Mining of Massive Datasets”, CUP

8.Tom White, "Hadoop: The Definitive Guide", O'Reilly.

IPS Academy, Institute of Engineering & Science

(A UGC Autonomous Institute, Affiliated to RGPV, Bhopal)
Scheme Based on AICTE Flexible Curriculum

Department of Computer Science & Engineering

Bachelor of Technology (B.Tech.)

[Computer Science & Engineering – Data Science-VI Sem]

PCC-DS14	Machine Learning-II	2L: 1T(4hrs.)	Credits: 3

Course Objective:

Students understand issues and challenges of Machine Learning. And Data,model selection, model Complexity with Understanding of the strengths and weaknesses of many popular machine learning approaches.

Course Contents:

Module 1: **(06hrs.)**

Support Vector Machines- What are SVMs, Linear & Non- Linearly separable problems, kernel functions, digit recognition example, Advantages & Disadvantages of SVMs, SVM vs Other algorithms. VC Dimensions in SVM

Module2: **(08 hrs.)**

Decision Tree: Basic entropy, gini index, information gain, pros & cons, mathematical formulation of decision trees, Cancer Identification Example. Random Forest: Definition & relation with decision trees, Pruning & Bagging in Random Forests, Mathematical Formulation of Random Forests, OCR parameter tuning example.

Module3: **(10hrs.)**

Neural Network Theory: Inspiration for Feed forward neural networks, Types of neural networks, layers, artificial neural networks, activation functions, bias nodes, error of the networks, backpropagation explained, Optimisation with gradient descent

Module4: **(08 hrs.)**

Deep Neural Networks: Activation functions revisited, loss functions, stochastic gradient descent, hyperparameters, implementation of Deep Neural Networks, Multiclass Classification, Evaluating & Testing the model

Module5: **(06hrs.)**

Convolutional Neural Networks: Basics, feature selection, kernels, pooling, flattening, illustration, CNN visualisation tools. Appendices: Some methods of optimization like Genetic algorithms, Simulated Annealing, Hilbert Space

IPS Academy, Institute of Engineering & Science

(A UGC Autonomous Institute, Affiliated to RGPV, Bhopal)
Scheme Based on AICTE Flexible Curriculum

Department of Computer Science & Engineering

Bachelor of Technology (B.Tech.)

[Computer Science & Engineering – Data Science-VI Sem]

Course Outcomes:

1. Learning about Support Vector Machines
2. Understanding the concepts of Decision Trees & Random forests.
3. Getting to know the basics of Neural Networks
4. Building Deep Neural Networks
5. Learning about Convolutional Neural Networks & auxiliary features of Neural Networks.

List of Text/Reference Books:

1. Introduction to Machine learning, Nils J.Nilsson

2. Machine learning for dummies, IBM Limited ed, by Judith Hurwitz and Daniel Kirsch

3. Introduction to Machine Learning with Python A guide for data scientists, Andreas, C. Muller & Sarah Guido, O'Reilly

4. [2022] Machine Learning and Deep Learning Bootcamp in Python: Course from Udemy

IPS Academy, Institute of Engineering & Science

(A UGC Autonomous Institute, Affiliated to RGPV, Bhopal)
Scheme Based on AICTE Flexible Curriculum

Department of Computer Science & Engineering

Bachelor of Technology (B.Tech.)

[Computer Science & Engineering – Data Science-VI Sem]

PCC-DS15	Natural Language Processing	2L: 1T(4hrs.)	Credits: 3

Course Objective:

To gain the knowledge for developing advanced technology of computer systems like speech recognition and machine translation.

Course Contents:

Module 1: **(08hrs.)**

Introduction to NLP: Machine Learning & NLP, Argmax computation, WSD: WordNet, application in query expansion, wikitionary, semantic relatedness, Measures of WordNet similarity, Rensick's work on WordNet similarity

(08 hrs.)

Module 2:

Algorithms: Parsing algorithms, Evidence for deeper structure; Top Down Parsing algorithms; Noun structure, Non-noun structure and Parsing algorithms, probabilistic parsing: sequence labelling, PCFG, training issues, inside outside probabilities.

(08hrs.)

Module 3:

Speech & Phonetics: Arguments & Adjuncts, Hidden Markov Models, Morphology, Graphical Models of sequence labelling in NLP, Phonetics, Vowels & Consonents

(08 hrs.)

Module 4:

Forward Backward Probability, Viterbi algorithm, phonology, sentiment analysis & opinions on the web, Machine Translation & MT Tools – GIZA++ & Moses, Text Entailment, POS Tagging, Phonology: ASR & speech synthesis

(08hrs.)

IPS Academy, Institute of Engineering & Science

(A UGC Autonomous Institute, Affiliated to RGPV, Bhopal)
Scheme Based on AICTE Flexible Curriculum

Department of Computer Science & Engineering

Bachelor of Technology (B.Tech.)

[Computer Science & Engineering – Data Science-VI Sem]

Module 5:

Hidden Markov Models & Viterbi, Precision, Recall, F-score and map, Semantic relations, Universal Networking Language, Towards dependency parsing, Semantic Role Extraction, Baum Welch algorithm & Hidden Markov Models Training.

Course Outcomes:

1. Learning about the basics of NLP
2. Working with Parsing algorithms
3. Getting acquainted with Speech & Phonetics & Morphology.
4. Understanding the concepts of Forward Backward Probability
5. Learning about Hidden Markov Models & other advanced concepts.

List of Text/Reference Books:

1.D. Jurafsky and J.H. Martin, "Speech and Language Processing; Processing",Prentice Hall,2000.

2.C. Manning and H. Schutze,"Foundations of Statistical Natural Language Processing", MIT Press

3.James Allen."Natural Language Understanding", Addison Wesley,1994.

4.Richard M Reese, "Natural Language Processing with Javall", OReilly Media,2015.

5.Tanveer Siddiqui, U.S. Tiwary, "Natural Language Processing and InformationRetrieval", Oxford University Press,2008

6.NPTEL Course: https://nptel.ac.in/courses/106101007

IPS Academy, Institute of Engineering & Science

(A UGC Autonomous Institute, Affiliated to RGPV, Bhopal)
Scheme Based on AICTE Flexible Curriculum

Department of Computer Science & Engineering

Bachelor of Technology (B.Tech.)

[Computer Science & Engineering – Data Science-VI Sem]

PEC-DS02(A)	**Cyber Security**	**3L: 0T(3hrs.)**	**Credits: 2**

Course Objective:

Analyze and resolve security issues in an organization to secure an IT infrastructure.

Course Contents:

Module 1: **(06hrs.)**

Introduction of Cyber Crime, Challenges of cyber crime, Classifications of Cybercrimes: E- Mail Spoofing, Spamming, Internet Time Theft, Salami attack/Salami Technique.

Module2: **(10 hrs.)**

Web jacking, Online Frauds, Software Piracy, Computer Network Intrusions, Password Sniffing, Identity Theft, cyber terrorism, Virtual Crime, Perception of cyber criminals: hackers, insurgents and extremist group etc. Web servers were hacking, session hijacking..

Module3: **(10hrs.)**

Cyber Crime and Criminal justice: Concept of Cyber Crime and the IT Act, 2000, Hacking, Teenage Web Vandals, Cyber Fraud and Cheating, Defamation, Harassment and E- mail Abuse, Other ITAct Offences, Monetary Penalties, jurisdiction and Cyber Crimes, Nature of Criminality, Strategies totackle Cyber Crime and Trends.

Module4: **(08 hrs.)**

The Indian Evidence Act of 1872 v. Information Technology Act, 2000: Status of Electronic Records as Evidence, Proof and Management of Electronic Records; Relevancy, Admissibility and Probative Value of E-Evidence, Proving Digital Signatures, Proof of Electronic Agreements, Proving Electronic Messages.

Module5: **(06hrs.)**

Tools and Methods in Cybercrime: Proxy Servers and Anonymizers , Password Cracking, Key loggers and Spyware, virus and worms, Trojan Horses, Backdoors, DoS and DDoS Attacks, Buffer and Overflow, Attack on Wireless Networks, Phishing : Method of Phishing, Phishing Techniques. Introduction to KALI Linux.

IPS Academy, Institute of Engineering & Science

(A UGC Autonomous Institute, Affiliated to RGPV, Bhopal)
Scheme Based on AICTE Flexible Curriculum

Department of Computer Science & Engineering

Bachelor of Technology (B.Tech.)

[Computer Science & Engineering – Data Science-VI Sem]

Course Outcomes:

1. Define and explain the concepts of cybercrime and its classification.
2. Delineate the components online frauds, intrusions, virtual crimes and hacking.
3. Knowledge of different act's in cybersecurity
4. List the various parts of IT act related to electronic records.
5. Knowledge of different Cyber Security tools.

List of Text/Reference Books:

1. Jonathan Clough, "Principles of Cyber crime", Cambridge University Press, 2ndEdition, 2015.
2. John R. Vacca, "Computer Forensics: Computer Crime Scene Investigation", Charles River Media, 2nd Edition,2005.
3. Vivek Sood"Cyber Law Simplified",TMH,2001.
4. Nina Godbole, SunitBelapure, "Cyber Security",Wiley-India
5.William Hutchinson, Mathew Warren, "Information Warfare: Corporate attack and defense in digital world" , Elsevier, Reed International and Professional Publishing Ltd,2001
6. Harish Chander , "Cyber Laws and IT Protection", Prentice Hall India Learning PrivateLimited,2012

IPS Academy, Institute of Engineering & Science

(A UGC Autonomous Institute, Affiliated to RGPV, Bhopal)
Scheme Based on AICTE Flexible Curriculum

Department of Computer Science & Engineering

Bachelor of Technology (B.Tech.)

[Computer Science & Engineering – Data Science-VI Sem]

PEC-DS02(B)	**Information Storage & Management**	**3L: 0T(3hrs.)**	**Credits: 2**

Course Objective:

To introduce solutions available for data storage, Core elements of a data center infrastructure, role of each element in supporting business activities

Course Contents:

Module 1: **(06hrs.)**

Introduction to Storage Technology: Data proliferation, evolution of various storage technologies, Overview of storage infrastructure components, Information Lifecycle Management, Data categorization.

Module2: **(10 hrs.)**

Storage Systems Architecture: Intelligent disk subsystems overview, Contrast of integrated vs. modular arrays, Component architecture of intelligent disk subsystems, Disk physical structure components, properties, performance, and specifications, RAID levels & parity algorithms, hot sparing, Front end to host storage provisioning, mapping and operation.

(10hrs.)

Module3:

Introduction to Networked Storage: JBOD, DAS, NAS, SAN & CAS evolution and comparison. Applications, Elements, connectivity, standards, management, security and limitations of DAS, NAS, CAS & SAN.

Module4: **(08 hrs.)**

Hybrid Storage solutions; Virtualization: Memory, network, server, storage & appliances. Data center concepts & requirements, Backup & Disaster Recovery: Principles Managing & Monitoring: Industry management standards (SNMP, SMI-S, CIM), standard framework applications, Key management metrics (Thresholds, availability, capacity, security, performance)

(06hrs.)

IPS Academy, Institute of Engineering & Science

(A UGC Autonomous Institute, Affiliated to RGPV, Bhopal)
Scheme Based on AICTE Flexible Curriculum

Department of Computer Science & Engineering

Bachelor of Technology (B.Tech.)

[Computer Science & Engineering – Data Science-VI Sem]

Module5:

Information storage on cloud :Concept of Cloud, Cloud Computing, storage on Cloud, Cloud Vocabulary, Architectural Framework, Cloud benefits, Cloud computing Evolution, Applications & services on cloud, Cloud service providers and Models, Essential characteristics of cloud computing, Cloud Security and integration

Course Outcomes:

1. To Understand the Concept of Information Storage and Data centre Environment.
2. To Understand about Data Protection.
3. To Understand Fiber Channel SAN.
4. To describe the different backup and recovery topologies and their role in providing disaster recovery and business continuity capabilities.
5. To Understand Cloud Computing.

List of Text/Reference Books:

1. G. Somasundaram & Alok Shrivastava (EMC Education Services) editors, "Information Storage and Management: Storing, Managing, and Protecting Digital Information",Wiley India,2009.
2. Ulf Troppens, Wolfgang Mueller-Friedt, Rainer Erkens, Rainer Wolafka, Nils Haustein, "Storage Network explained: Basic and application of fiber channels, SAN, NAS, iSESI, INFINIBAND and FCOE", Wiley India.
3. John W. Rittinghouse and James F. Ransome, "Cloud Computing: Implementation, Management andSecurity", CRC Press, Taylor Frances Pub.1stEdition,2017
4. Nick Antonopoulos, Lee Gillam, "Cloud Computing: Principles, System &Application", Springer.
5. Anthony T. Velete, Toby J.Velk, and Robert Eltenpeter, "Cloud Computing: A practical Approach" ,McGraw-Hill Education (India) Pvt. Limited,2009
6. Dr. Kumar Saurabh, "Cloud Computing: Insight into New Era I", Wiley India Pvt. Limited, 2011.

IPS Academy, Institute of Engineering & Science

(A UGC Autonomous Institute, Affiliated to RGPV, Bhopal)
Scheme Based on AICTE Flexible Curriculum

Department of Computer Science & Engineering

Bachelor of Technology (B.Tech.)

[Computer Science & Engineering – Data Science-VI Sem]

PEC-DS02(C)	**Compiler Design**	**3L: 0T(3hrs.)**	**Credits: 2**

Prerequisite: Theory of Computation

Course Objective:

To explain the different stages in the process of compilation

Course Contents:

Module 1: **(06hrs.)**

Introduction to compiling & Lexical Analysis Introduction of Compiler, Major data Structure in compiler, types of Compiler, Front-end and Backend of compiler, Compiler structure: analysis-synthesis model of compilation, various phases of a compiler, Single & Multipass Compiler, Lexical analysis: Input buffering, Specification & Recognition of Tokens, Design of a Lexical Analyzer Generator, LEX

Module2: **(10 hrs.)**

Syntax Analysis &Syntax Directed Translation Syntax analysis: CFGs, Top down parsing, Brute force approach, recursive descent parsing, transformation on the grammars, predictive parsing, bottom up parsing, operator precedence parsing, LR parsers (SLR,LALR, LR),Parser generation. Syntax directed definitions: Construction of Syntax trees, Bottom up evaluation of S-attributed definition, L attribute definition, Top down translation, Bottom Up evaluation of inherited attributes Recursive Evaluation, Analysis of Syntax directed definition.

Module3: **(10hrs.)**

Type checking: type system, specification of simple type checker, equivalence of expression, types, type conversion, overloading of functions and operations, polymorphic functions. Runtime Environment: storage organization, Storage allocation strategies, Parameter passing, dynamic storage allocation, Symbol table, Error Detection & Recovery..

Module4: **(08 hrs.)**

Intermediate code generation: Declarations, Assignment statements, Boolean expressions, Case statements, back patching, Procedure calls Code Generation: Issues in the design of code generator, Basic block and flow graphs, Register allocation and assignment, DAG representation of basic blocks, peephole optimization, and generating code from DAG

IPS Academy, Institute of Engineering & Science

(A UGC Autonomous Institute, Affiliated to RGPV, Bhopal)
Scheme Based on AICTE Flexible Curriculum

Department of Computer Science & Engineering

Bachelor of Technology (B.Tech.)

[Computer Science & Engineering – Data Science-VI Sem]

Module5: **(06hrs.)**

Introduction to Code optimization: sources of optimization of basic blocks, loops in flow graphs, dead code elimination, loop optimization, Introduction to global data flow analysis, Code Improving transformations ,Data flow analysis of structure flow graph Symbolic debugging of optimized code.

Course Outcomes:

1. Understand the overview of phase of compiler and Lexical analysis.
2. Design and implement various parsing techniques of compiler.
3. Apply type checking for semantic analysis and analyze Run time environment.
4. Design and implement different intermediate code generation techniques.
5. Analyze various code optimization techniques

.

List of Text / Reference Books:

1.A.V. Aho, R. Sethi, and J.D. Ullman. “Compilers: Principles, Techniques and Tools” , Pearson Education,2nd Edition ,2007.

2.V Raghavan, “Principals of Compiler Design”, TMH Pub.,2017

3.Louden. “Compiler Construction: Principles and Practice”, CengageLearning,1997

4.A. C. Holub. “Compiler Design in C” , Prentice-Hall Inc., 1993.

5.Ronald Mak, “Writing compiler & Interpreters”, Willey Pub.,3rd Edition,2009

IPS Academy, Institute of Engineering & Science

(A UGC Autonomous Institute, Affiliated to RGPV, Bhopal)
Scheme Based on AICTE Flexible Curriculum

Department of Computer Science & Engineering

Bachelor of Technology (B.Tech.)

[Computer Science & Engineering – Data Science-VI Sem]

LC- DS13(P)	Big Data and Hadoop Lab	02P(4hrs.)	Credits: 2

Course Objective:

To introduce solutions available for data storage, Core elements of a data center infrastructure, role ofeach element in supporting business activities

Course Contents:

Module 1:

- Introduction to Hadoop Ecosystem
- Hadoop Architecture: HDFS, YARN, and MapReduce
- Introduction to Hadoop Distributed File System (HDFS)
- Basic HDFS Commands: put, get, ls, rm, etc
- Concept of Serialization and Deserialization
- Working with integer data types

Module2:

- Overview of MapReduce Architecture
- Key Concepts: Mapper, Reducer, Combiner
- Writing and executing basic MapReduce program
- Implementing a Grep Program
- Understanding MapReduce Paradigm through word count and execution time analysis

Module3:

- Overview of NoSQL Databases
- Introduction to MongoDB and its features
- Installation and SetupCreate, Read, Update, Delete (CRUD) operations
- Working with collections and data types
- Storing student information using various collection types (Map)

Module4:

- Overview of Spark Framework
- Key Features and Components of Spark
- Introduction to Spark RDDs and DataFrames
- Basic Transformations and Actions in Spark
- Writing and executing Spark applications
- Integration with Hadoop ecosystem

IPS Academy, Institute of Engineering & Science

(A UGC Autonomous Institute, Affiliated to RGPV, Bhopal)
Scheme Based on AICTE Flexible Curriculum

Department of Computer Science & Engineering

Bachelor of Technology (B.Tech.)

[Computer Science & Engineering – Data Science-VI Sem]

Module5:

- Overview of Hive and its architecture
- Hive Metastore and HiveQL (HQL)
- Creating and managing databases and tables
- Loading, inserting, and querying data in Hive

Course Outcome

- Understand and Explain Big Data and Hadoop Ecosystem
- Implement and Analyze MapReduce Programs
- Perform Data Management Operations Using NoSQL Databases:
- Utilize Apache Spark for Big Data Processing:
- Execute Data Queries Using Hive and HQL

List of Experiments:

1. To Study of Big Data Analytics and Hadoop Architecture.
2. To Understand Overall programming architecture of Map reduce API. Implement MapReduce Programming.
3. To Study HDFS Commands.
4. To Study serializes and deserializes data of integer type in Hadoop.
5. To run a basic Word Count MapReduce program to understand MapReduce Paradigm.
6. Basic CRUD operations in MongoDB.
7. Store the basic information about students such as roll no and name using various collection types Map.
8. To run a Grep program on Hadoop to understand Mapreduce Paradigm: To count words in a given file, To view the output file, and To calculate execution time.
9. Installation of SPARK framework with or without Hadoop framework.
10. To Study about the Hive commands using HQL (DDL and DML).

IPS Academy, Institute of Engineering & Science

(A UGC Autonomous Institute, Affiliated to RGPV, Bhopal)
Scheme Based on AICTE Flexible Curriculum

Department of Computer Science & Engineering

Bachelor of Technology (B.Tech.)

[Computer Science & Engineering – Data Science-VI Sem]

LC- DS14(P)	Machine Learning-II Lab	02P (4hrs.)	Credits: 2

Course Objective:

Gain a comprehensive understanding of machine learning concepts, develop proficiency in various algorithms, and enhance practical skills in Python for real-world applications.

Course Contents:

Module 1:

- Linear SVM classifier, visualize the decision boundary, and identify support vectors.
- SVM with linear, polynomial, and RBF kernels on a dataset. Compare and visualize performance.
- SVM on the MNIST dataset for digit recognition, assess model accuracy and performance.

Module2:

- Compare SVM with other classification algorithms.
- Implement and compare SVM, Logistic Regression, and k-NN on a dataset.
- Discuss performance metrics.
- The concept of VC dimensions.

Module3:

- Build and evaluate a basic neural network.
- Train and evaluate the network.
- Exploring Activation Functions
- Understand and compare different activation functions.
- Implement ReLU, Sigmoid, and Tanh activation functions in a neural network.
- Compare their performance on a dataset.
- Optimize a neural network using backpropagation.

Module4:

- Implement and train a deep neural network.
- Use a deep learning framework (e.g., TensorFlow, Keras) to build and train a deep neural network on a complex dataset.
- Use grid search or random search for optimization and observe effects on model

performance.

Module5:

- Understand the fundamentals of CNNs.
- CNN for image classification, visualize learned filters and feature maps.
- Advanced CNN Techniques
- Implement pooling and dropout techniques in a CNN. Use visualization tools to understand learned features.

Course Outcome		
1: Master SVM classifiers with various kernels, visualize decision boundaries, and apply SVM to digit recognition. **2:** Compare SVM with other classifiers, implement SVM, Logistic Regression, and k-NN, and discuss VC dimensions and performance metrics. **3:** Build, train, and optimize neural networks, exploring and comparing different activation functions. **4:** Implement and optimize deep neural networks using frameworks like TensorFlow or Keras. **5:** Understand and apply CNNs for image classification, including advanced techniques like pooling and dropout. •		
List of Experiments:		
1. Implement a linear SVM classifier using a simple dataset. Visualize the decision boundary and support vectors. 2. Implement an SVM with different kernel functions (linear, polynomial, RBF) on a dataset. Compare the performance and visualizations. 3. Implement SVM for digit recognition using the MNIST dataset. Evaluate the model's accuracy and performance. 4. Compare SVM with other classification algorithms (e.g., Logistic Regression, k-NN) on a dataset. Discuss the results and performance metrics . 5. Implement examples to illustrate the concept of VC dimensions and its implications on model		

complexity and capacity.

6. Implement a simple neural network from scratch for a classification problem. Train and evaluate the network.

7. Implement various activation functions (ReLU, Sigmoid, Tanh) in a neural network. Compare their performance on a dataset.

8. Implement the backpropagation algorithm and use gradient descent to optimize the neural network. Visualize the error reduction over iterations.

9. Implement a deep neural network using a deep learning framework (e.g., TensorFlow, Keras). Train the network on a complex dataset and evaluate its performance
.
10. Experiment with different hyperparameters (learning rate, batch size, number of layers) and observe their effects on model performance. Use grid search or random search for optimization

11. Implement a basic CNN for image classification. Visualize the learned filters and feature maps.

12. Implement techniques like pooling and dropout in a CNN. Use visualization tools to understand what the CNN has learned