SEMESTER – I **Hours/week: 4**
PDS1501 **Credits : 4**

INTRODUCTION TO DATA SCIENCE

.

Unit – I: Introduction
Introduction to Data Science – Evolution of Data Science – Data Science Roles – Stages in a Data Science Project – Applications of Data Science in various fields – Data Security Issues.

Unit – II: Data Collection and Data Pre-Processing
Data Collection Strategies – Data Pre-Processing Overview – Data Cleaning – Data Integration and Transformation – Data Reduction – Data Discretization.

Unit – III: Exploratory Data Analytics
Descriptive Statistics – Mean, Standard Deviation, Skewness and Kurtosis – Box Plots – Pivot Table – Heat Map – Correlation Statistics – ANOVA.

Unit – IV: Model Development
Simple and Multiple Regression – Model Evaluation using Visualization – Residual Plot – Distribution Plot – Polynomial Regression and Pipelines – Measures for In-sample Evaluation – Prediction and Decision Making.

Unit – V: Model Evaluation
Generalization Error – Out-of-Sample Evaluation Metrics – Cross Validation – Overfitting – Under Fitting and Model Selection – Prediction by using Ridge Regression – Testing Multiple Parameters by using Grid Search.

REFERENCES:

1. Jojo Moolayil, “Smarter Decisions : The Intersection of IoT and Data Science”, PACKT, 2016.
2. Cathy O’Neil and Rachel Schutt , “Doing Data Science”, O'Reilly, 2015.
3. David Dietrich, Barry Heller, Beibei Yang, “Data Science and Big data Analytics”, EMC 2013
4. Raj, Pethuru, “Handbook of Research on Cloud Infrastructures for Big Data Analytics”, IGI Global.

SEMESTER – I **Hours/week: 4**
PDS1502 **Credits : 4**

STATISTICS FOR DATA SCIENCE

.

Unit – I: Descriptive Statistics
Sampling Techniques – Data Classification – Tabulation – Frequency and graphic Representation – Measures of Central Tendency – Measures of Variation – Quartiles and Percentiles – Moments - Skewness and Kurtosis.

Unit – II: Correlation and Regression
Scatter Diagram – Karl Pearson's Correlation Coefficient – Rank Correlation - Correlation Coefficient for Bivariate Frequency Distribution – Regression Coefficients – Fitting of Regression Lines.

Unit – III: Probability Theory
Random Experiment – Sample Space – Events – Axiomatic Definition of Probability – Addition Theorem – Multiplication Theorem – Baye's Theorem -Applications.

Unit – IV: Distribution Function
Continuous and Discrete Random Variables – Distribution Function of a Random Variable – Probability Mass Functions and Probability Density Functions – Characteristic Functions – Central Limit Theorems.

Unit – V: Probability Distributions
Probability Distributions – Recurrence Relationships – Moment Generating Functions – Cumulant Generating Functions – Continuous Probability Distributions - Rectangular Distribution – Binomial Distribution – Poisson Distribution – Continuous Probability Distributions – Uniform Distribution - Normal Distribution – Exponential Distribution.

REFERENCES:

1. Gupta, S.C. and Kapoor, V.K.: "Fundamentals of Mathematical Statistics", Sultan & Chand & Sons, New Delhi, 11th Ed, 2002.
2. Hastie, Trevor, et al. "The elements of Statistical Learning", Springer, 2009.
3. Practical Statistics for Data Scientists, 2nd Edition, Peter Bruce, Andrew Bruce and Peter Gedeck, May 2020
4. Statistics for Machine Learning, By Pratap Dangeti, July 2017

SEMESTER – I **Hours/week: 5**
PDS1503 **Credits : 5**

PYTHON FOR DATA SCIENCE

.

Unit – I: Data Structures and OOP

Python Program Execution Procedure – Statements – Expressions – Flow of Controls – Functions – Numeric Data Types – Sequences – Strings – Tuples – Lists – Dictionaries.

Class – Constructors – Object Creation – Inheritance – Overloading.

Text Files and Binary Files – Reading and Writing.

Unit – II: Numpy and Pandas Packages

NumPy ndarray - Vectorization Operation - Array Indexing and Slicing - Transposing Array and Swapping Axes - Saving and Loading Array - Universal Functions - Mathematical and Statistical Functions in Numpy .

Series and DataFrame data structures in pandas - Creation of Data Frames – Accessing the columns in a DataFrame - Accessing the rows in a DataFrame - Panda's Index Objects - Reindexing Series and DataFrames - Dropping entries from Series and Data Frames - Indexing, Selection and Filtering in Series and Data Frames - Arithmetic Operations between Data Frames and Series - Function Application and Mapping.

Unit – III: Data Wrangling

Combining and Merging Data Sets – Reshaping and Pivoting – Data Transformation – String manipulations – Regular Expressions.

Unit – IV: Data Aggregation and Group Operations

GroupBy Mechanics – Data Aggregation – GroupWise Operations – Transformations – Pivot Tables – Cross Tabulations – Date and Time data types.

Unit – V: Visualization in Python

Matplotlib and Seaborn Packages – Plotting Graph - Controlling Graphs – Adding Text – More Graph Types – Getting and Setting Values – Patches.

REFERENCES:

1. Gowrishanker and Veena, "Introduction to Python Programming", CRC Press, 2019.
2. Python Crash Course, 2nd Edition, By Eric Matthes, May 2019
3. NumPy Essentials, By Leo Chin and Tanmay Dutta, April 2016
4. Joel Grus, "Data Science from scratch", O'Reilly, 2015.
5. Wes Mc Kinney, "Python for Data Analysis", O'Reilly Media, 2012.
6. *Kenneth A. Lambert, (2011), "The Fundamentals of Python: First Programs", Cengage Learning*
7. Jake Vanderplas. Python Data Science Handbook: Essential Tools for Working with Data 1st Edition.

SEMESTER – I **Hours/week: 4**
PDS1504 **Credits : 4**

PYTHON FOR DATA SCIENCE - LAB

LIST OF EXERCISES:

1. Editing and executing Programs involving Flow Controls.
2. Editing and executing Programs involving Functions.
3. Program in String Manipulations
4. Creating and manipulating a Tuple
5. Creating and manipulating a List
6. Creating and manipulating a Dictionary
7. Object Creation and Usage
8. Program involving Inheritance
9. Program involving Overloading
10. Reading and Writing with Text Files and Binary Files
11. Combining and Merging Data Sets
12. Program involving Regular Expressions
13. Data Aggregation and GroupWise Operations

SEMESTER – I **Hours/week: 4**
PDS1505 **Credits : 4**

RDBMS LAB

1. Creating a database
2. Creating a table
3. Inserting records in a table
4. Altering the table structure.
5. Deleting data from table
6. Updating data from table.
7. Select command
8. Where clause
9. Aggregate functions
10. Numeric functions (Absolute, ceiling, floor, modulo, round off, square, Square Root, power)
11. Constraints
12. Group By, Having
13. Operators (and, or, not between, In , not in, is null, is not null, like, Order By)
14. String Functions (Lower, Upper, Replace, left-trim, right-trim, substring, Length, rename)
15. Drop (table, database)
16. Truncate
17. *Sub Queries, Alias*

SEMESTER – I **Hours/week: 5**
PDS1506 **Credits : 5**

MACHINE LEARNING

.

Unit – I: Introduction

Machine Learning Foundations – Overview – Design of a Learning System – Types of Machine Learning – Supervised Learning and Unsupervised Learning – Mathematical Foundations of Machine Learning – Applications of Machine Learning.

Unit – II: Supervised Learning - I

Simple Linear Regression – Multiple Linear Regression – Polynomial Regression – Ridge Regression – Lasso Regression – Evaluating Regression Models – Model Selection – Bagging – Ensemble Methods.

Unit – III: Supervised Learning - II

Classification – Logistic Regression – Decision Tree Regression and Classification – Random Forest Regression and Classification – Support Vector Machine Regression and Classification - Evaluating Classification Models.

Unit – IV: Unsupervised Learning

Clustering – K-Means Clustering – Density-Based Clustering – Dimensionality Reduction – Collaborative Filtering.

Unit – V: Association Rule Learning and Reinforcement Learning

Association Rule Learning – Apriori – Eclat – Reinforcement Learning – Upper Confidence Bound – Thompson Sampling – Q-Learning.

REFERENCES:

1. Kevin P. Murphy, “Machine Learning: A Probabilistic Perspective”, MIT Press, 2012.
2. Ethem Alpaydin, “Introduction to Machine Learning”, MIT Press, Third Edition, 2014.
3. Tom Mitchell, "Machine Learning", McGraw-Hill, 1997.
4. Sebastian Raschka, Vahid Mirjilili,”Python Machine Learning and deep learning”, 2nd edition, kindle book, 2018
5. Carol Quadros,”Machine Learning with python, scikit-learn and Tensorflow”, Packet Publishing, 2018
6. Gavin Hackeling,” Machine Learning with scikit-learn”, Packet publishing, O’Reily, 2018
7. Stanford Lectures of Prof. Andrew Ng on Machine Learning

SEMESTER – I **Hours/week: 4**
PDS1507 **Credits : 4**

MACHINE LEARNING - LAB

.

1. Write a program to implement the naïve Bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier, considering few test data sets.

2. Assuming a set of documents that need to be classified, use the naïve Bayesian algorithm.

3. Classifier model to perform this task. Built-in Java classes/API can be used to write the program. Calculate the accuracy, precision, and recall for your data set.

4. Write a program to implement k-Nearest Neighbour algorithm to classify the iris. print both correct and wrong predictions. Java/Python ML library classes can be used for this problem.

5. Write a program to implement Logistic Regression algorithm to classify the housing price data set. Print both correct and wrong predictions. Java/Python ML library classes can be used for this problem.

6. Write a program to implement and compare SVM, KNN and Logistic regression algorithm to classify the iPhone purchase records data set. Print both correct and wrong predictions. Java/ Python ML library classes can be used for this problem.

SEMESTER – II **Hours/week: 4**

PDS2501 **Credits : 4**

STATISTICAL INFERENCE

UNIT – I: TESTING OF HYPOTHESIS - PART 1

Testing of Hypothesis - Statistical Hypothesis - Simple and composite hypothesis, Null and Alternative hypothesis - two kinds of errors, level of significance, size and power of a test, most powerful test, Neyman-Pearson lemma with proof.

UNIT – II: TEST OF HYPOTHESIS – PART 2

Simple examples using Neyman-Pearson lemma .Uniformly most powerful tests and unbiased tests based on normal Likelihood ratio test (without proof) and its properties. Application of LR test for single mean.

UNIT – III: TEST OF SIGNIFICANCE FOR LARGE SAMPLES

Test of significance for mean(s), variance(s), proportion(s), correlation coefficient(s) based on Normal distribution.

UNIT – IV: TEST OF SIGNIFICANCE FOR SMALL SAMPLES

Test of significance for mean(s), variance(s), correlation coefficient(s), regression coefficient, based on t, Chi-square and F-distributions. Applications of Chi-square in test of significance (independence of attributes, goodness of fit).

UNIT – V:

Non-parametric tests – Kolmogorov -Smirnov test, Sign test, Wald- Wolfowitz run test, run test for randomness, median test, Wilcoxon test and Wilcoxon – Mann-Whitney U test.

REFERENCE BOOKS

1. Gupta, S.C. and Kapoor, V.K.: “Fundamentals of Mathematical Statistics”, Sultan & Chand & Sons, New Delhi, 11th Ed, 2002.
2. Rohatgi, V.K. : “Statistical Inference”, John Wiley and sons, 1984.
3. Hogg, R.V, Craig. A.T. and Tannis: “Introduction to mathematical statistics”, Prentice Hall, England, 1995.
4. Dudewicz. E.J and Mishra.S.N.: “Modern Mathematical statistics”, John Wiley and sons, 1988.

SEMESTER – II **Hours/week: 4**
PDS2502 **Credits : 4**

BIG DATA ANALYTICS THROUGH SPARK

.

Unit – I: Introduction to Spark

Apache Spark Ecosystem - Setting up the Spark Python Environment – Execution of a PySpark Program – Resilient Distributed Datasets – Spark Architecture – Spark Project Workflow.

Unit – II: Spark Programming with Python

Loading and Storing Data – Transformations – Actions – Key-Value Resilient Distributed Datasets – Local Variables – Broadcast Variables – Accumulators – Partitioning – Persistence.

Unit – III: Spark SQL

Overview of Spark SQL – Spark Session – Data Frames – Schema of a Data Frame – Operations supported by Data Frames – Filter, Join, GroupBy, Agg operations – Nesting the Operations – Temporary Tables – Viewing and Querying Temporary Tables.

Unit – IV: Spark Streaming

Use Cases for Realtime Analytics – Transferring, Summarizing, Analysing Realtime data – Data Sources supported by Spark Streaming – Flat files, TCP/IP – Flume – Kafka – Kinesis – Streaming Context – DStreams – Dstream RDDs – Dstream Processing.

Unit – V: Machine Learning with Spark

Linear Regression – Decision Tree Classification – Principal Component Analysis – Random Forest Classification – Text Pre-processing with TF-IDF – Naïve Bayes Classification – K-Means Clustering – Recommendation Engines.

REFERENCES:

1. Tomasz Drabos, “Learning PySpark”, PACKT, 2017.
2. Padma Priya Chitturi, “Apache Spark for Data Science”, PACKT, 2017.
3. Holden Karau, “ Learning Spark”. PACKT, 2016.
4. Sandy Riza, “Advanced Analytics with Spark”, O’ Reilly, 2016.
5. Romeo Kienzler, “Mastering Apache Spark”, PACKT, 2017.

SEMESTER – II **Hours/week: 4**
PDS2503 **Credits : 4**

BIG DATA ANALYTICS THROUGH SPARK - LAB

LIST OF EXERCISES:

1. Program involving Resilient Distributed Datasets
2. Program involving Transformations and Actions
3. Program involving Key-Value Resilient Distributed Datasets
4. Program involving Local Variables, Broadcast Variables and Accumulators
5. Program involving Filter, Join, GroupBy, Agg operations
6. Viewing and Querying Temporary Tables
7. Transferring, Summarizing and Analysing Twitter data
8. Program involving Flume, Kafka and Kinesis
9. Program involving DStreams and Dstream RDDs
10. Linear Regression
11. Decision Tree Classification
12. Principal Component Analysis
13. Random Forest Classification
14. Text Pre-processing with TF-IDF
15. Naïve Bayes Classification
16. K-Means Clustering

SEMESTER – II **Hours/week: 4**
PDS2504 **Credits : 4**

NOSQL DATABASES

- An Overview of NoSQL (1 hour)
- HDFS (3 hours)
- Apache Hive as an HDFS Data Warehouse (5 hours)
- HBase (5 hours)
- MongoDB (6 hours)
- Cassandra (7 hours)
- Neo4j (3 hours)

Unit – I: NoSQL and HDFS

Unit – II: Hive

Unit – III: HBase

Unit – IV: MongoDB

Introduction – Features - Data types - Mongo DB Query language - CRUD operations – Arrays - Functions: Count – Sort – Limit – Skip – Aggregate - Map Reduce. Cursors – Indexes - Mongo Import – Mongo Export.

Unit – V: Cassendra

Introduction – Features - Data types – CQLSH - Key spaces - CRUD operations – Collections – Counter – TTL - Alter commands - Import and Export - Querying System tables.

SEMESTER – II **Hours/week: 4**
PDS2505 **Credits : 4**

NOSQL DATABASES - LAB

- Exercises on HDFS
- Exercises on Apache Hive as an HDFS Data Warehouse
- Exercises on HBase
- Exercises on MongoDB
- Exercises on Cassandra
- Exercises on Neo4j

SEMESTER – II **Hours/week: 4**
PDS2601 **Credits : 3**

ELECTIVE 1A: FINANCIAL ANALYTICS

Unit: I

Introduction: Meaning-Importance of Financial Analytics uses-Features-Documents used in Financial Analytics: Balance Sheet, Income Statement, Cash flow statement-Elements of Financial Health: Liquidity, Leverage, Profitability. Financial Securities : Bond and Stock investments - Housing and Euro crisis - Securities Datasets and Visualization - Plotting multiple series.

Unit: II

Using Excel to Summarize Data, Slicing and Dicing Financial Data with PivotTables, Excel Charts to Summarize Marketing Data. Excel Functions to Summarize Data, Pricing Analytics, Risk based pricing, Fraud Detection and Prediction, Recovery Management, Loss Risk Forecasting, Risk Profiling, Portfolio Stress Testing.

Unit: III

Descriptive Analytics, Data Exploration, Dimension Reduction and Data Clustering Geographical Mapping Market Basket Analysis. Predictive Analytics Fraud Detection Churn Analysis Crime Mapping, Content Analytics Sentiment Analysis

Unit: IV

Forecasting Analytics Estimating Demand Curves and Optimize Price, Price Bundling, Non Linear Pricing and Price Skimming, Forecasting, Simple Regression and Correlation Multiple Regression to forecast sales. Modelling Trend and Seasonality Ratio to Moving Average Method, Winter's Method

UNIT - V

Analyzing financial data and implement financial models using R. Process of Data analytics using R: obtaining publicly available data, refining such data, implement the models and generate typical output, Prices and individual security returns, Portfolio returns, Risks, Factor Models

TEXTBOOKS

- Analysis of Economic Data, Gary Koop, (4th Edition), Wiley.
- Statistics and Data Analysis for Financial Engineering: with R examples; David Ruppert, David S. Matteson, Springers.

REFERENCE BOOKS

- Analyzing Financial Data and Implementing Financial Models Using 'R', Ang Clifford, Springers.
- Microsoft Excel 2013: Data Analysis and Business Modeling, Wayne L. Winston, Microsoft Publishing

SEMESTER – II **Hours/week: 4**
PDS2602 **Credits : 3**

ELECTIVE 1B: HEALTH ANALYTICS

UNIT I

Introduction

Introduction to Healthcare Data Analytics- Electronic Health Records– Components of EHR-Coding Systems- Benefits of EHR- Barrier to Adopting HER Challenges-Phenotyping Algorithms.

Unit II

Image Analysis

Biomedical Image Analysis- Mining of Sensor Data in Healthcare- Biomedical Signal Analysis- Genomic Data Analysis for Personalized Medicine.

Unit III

Data Analytics

Natural Language Processing and Data Mining for Clinical Text- Mining the Biomedical Social Media Analytics for Healthcare.

Unit IV

Advanced Data Analytics

Advanced Data Analytics for Healthcare– Review of Clinical Prediction Models- Temporal Data Mining for Healthcare Data- Visual Analytics for Healthcare- Predictive 53 Models for Integrating Clinical and Genomic Data- Information Retrieval for Healthcare- Data Publishing Methods in Healthcare.

Unit V

Applications

Applications and Practical Systems for Healthcare– Data Analytics for Pervasive Health-Fraud Detection in Healthcare- Data Analytics for Pharmaceutical Discoveries- Clinical Decision Support Systems- Computer-Assisted Medical Image Analysis Systems- Mobile Imaging and Analytics for Biomedical Data.

TEXT BOOKS

- Chandan K. Reddy and Charu C Aggarwal, “Healthcare data analytics”, Taylor & Francis, 2015.

REFERENCE BOOKS

- Hui Yang and Eva K. Lee, “Healthcare Analytics: From Data to Knowledge to Healthcare Improvement, Wiley, 2016.

SEMESTER – II **Hours/week: 3**
P__2901 **Credits : 2**

CROSS DISCIPLINARY: DATA VISUALIZATION

[TO BE OFFERED TO STUDENTS FROM OTHER SCHOOLS]

UNIT I- Introduction to Tableau (9 Hours)

Introducing real time dashboards – creating real time dashboards with Tableau – build a Tableau dashboard – real time dashboard updates in Tableau – organizing your Tableau dashboard – formatting your Tableau dashboard – interactive Tableau dashboard – Tableau dashboard starters – Tableau dashboard extensions – Tableau dashboards and story points – sharing your Tableau dashboard

UNIT II- Data Visualization Concepts (9 Hours)

Storytelling process – interpreting context – analysis types – who – what – and how of storytelling – Visualization for storytelling – –Graphical tools for data elaboration – storytelling scenarios – storyboarding – Visual selection – slope graphs – bar charts and types of bar charts – clutter and clutter elimination – Gestalt principle – story design best practices – tools for storytelling – Decluttering – crafting visual data – visual design concerns – storytelling with power BI – model visual and Tableau

UNIT III- Data Dashboards using Tablet (9 Hours)

Introducing real time dashboards – creating real time – dashboards with Tableau – build a Tableau dashboard – real time dashboard updates in Tableau – organizing your Tableau dashboard – formatting your Tableau dashboard– interactive Tableau dashboard – Tableau dashboard starters – Tableau dashboard extensions – Tableau dashboards and story points – sharing your tableau dashboard.

UNIT IV- Open Source Data Visualization with Seaborn (9 Hours)

Introduction to Seaborn – install Seaborn – Simple Univariate distributions – configure univariate – distribution plots – Simple Bivariate distributions – explore different types of – Bivariate distributions – analyse multiple variable pairs – Regression plots – themes and – styles in seaborn – searching for patterns in a dataset – configuring plot aesthetics – normal distribution and outliers – distributions within categories-part –distributions within categoriespart – analysing categories with facet grids - part – analysing categories with facet grids-part – introducing colour palettes – using colour palettes

UNIT V- Open Source Data Visualization with Matplotlib, Bokeh And Pygal (9 Hours)

An Introduction To Matplotlib – analysing Data Using NumPy and Pandas – visualizing – Univariate and Bivariate distributions – summary statistics using native – Python functions – Summary Statistics using NumPy – summary statistics using the SciPy library – Correlation and covariance – Z-score – relevance of data visualization for business – libraries for data visualization in python – Python data visualization environment – configuration – matplotlib libraries for visualization – bar chart using ggplot – bokeh and pygal – select visualization libraries – interactive graphs and image files – plot graphs – multiple lines in graphs – using scatter plots – using line graphs – using bar – charts – using box and whisker plots – using histograms – using a bubble plot – chart types – stacked bar plot – animate plots with matplotlib – plotting in Jupyter notebook

TEXTBOOKS:

• Fundamentals of Data Visualization, By Claus O. Wilke, April 2019
• Visual Analytics with Tableau, By Alexander Loth, May 2019

SEMESTER – III **Hours/week: 4**
PDS3501 **Credits : 4**

MULTIVARIATE TECHNIQUES FOR DATA ANALYTICS

.

Unit – I: Introduction to Multivariate Techniques

Measurement Scales(Metric and Non-metric Measurement Scles) – Classification of Multivariate Techniques(Dependence and Inter-dependence Techniques) – Applications of Multivariate Techniques in different disciplines.

Unit – II: Factor Analysis

Introduction to Factor Analysis – Meaning, Objectives and Assumptions – Designing a Factor Analysis Study – Deriving Factors – Assessing Overall Factors – Validation of Factor Analysis.

Unit – III: Cluster Analysis

Introduction to Cluster Analysis – Objectives and Assumptions – Research Design in Cluster Analysis – Hierarchical and Non-hierarchical Methods – Interpretation of Clusters – Validation of Profiling of Clusters.

Unit – IV: Discriminant Analysis

Introduction to Discriminant Analysis – Concepts, Objectives and Applications – Procedure for conducting Discriminant Analysis – Stepwise Discriminant Analysis – Mahalanobis Procedure – Logit Model.

Unit – V: Principal Component Analysis

Dimensionality Reduction – Deriving Orthogonal Projections – Lower Dimensional Subspaces – Characterization through Singular Value Decomposition and Eigenvalue Analysis – Rayleigh Quotient – Kernel PCA – Functional PCA.

REFERENCES:

1. Joseph F Hair, William C Black etal , “Multivariate Data Analysis” , Pearson Education, 7th edition, 2013.
2. T. W. Anderson , “An Introduction to Multivariate Statistical Analysis, 3rd Edition”, Wiley, 2003.
3. William r Dillon, John Wiley & sons, “Multivariate Analysis methods and applications”, Wiley, 1984.
4. Naresh K Malhotra, Satyabhusan Dash, “Marketing Research Anapplied Orientation”, Pearson, 2011.

SEMESTER – III **Hours/week: 4**
PDS3502 **Credits : 4**

DEEP LEARNING

.

Unit – I: Artificial Neural Networks
The Neuron – Activation Function – Gradient Descent – Stochastic Gradient Descent – Back Propagation – Business Problem.

Unit – II: Convolutional Neural Networks
Convolution Operation – ReLU layer – Pooling – Flattening – Full Conversion Layer – Softmax and Cross-Entropy.

Unit – III: Recurrent Neural Networks
RNN intuition – Tackling Vanishing Gradient Problem – Long Short-Term Memory – Building a RNN – Evaluating the RNN – Improving the RNN – Tuning the RNN.

Unit – IV: Boltzmann Machines
Introduction to Boltzmann Machine – Energy-Based Models – Restricted Boltzmann Machine – Contrastive Divergence – Deep Belief Networks – Deep Boltzmann Machine.

Unit – V: Computer Vision
Viola-Jones Algorithm – Haar-like Features – Integral Image – Training Classifiers – Adaptive Boosting – Cascading – Face Detection with Open CV.

REFERENCES:

1. Francois Challot, " Deep learning with Python", Manning, 2017.
2. Deep Learning Illustrated: A Visual, Interactive Guide to Artificial Intelligence,By Jon Krohn, Grant Beyleveld and Aglaé Bassens, September 2019
3. Ian Goodfellow, "Deep Learning", MIT Press, 2017.
4. Josh Patterson, "Deep Learning: A Practitioner's Approach", PACKT, 2017.
5. Dipayan Dev, " Deep Learning with Hadoop", PACKT, 2017.
6. Hugo Larochelle's Video Lectures on Deep Learning

SEMESTER – III **Hours/week: 4**
PDS3503 **Credits : 4**

DEEP LEARNING - LAB

LIST OF EXERCISES:

1. Setting up the Spyder IDE Environment and Executing a Python Program
2. Installing Keras, Tensorflow and Pytorch libraries and making use of them
3. Artificial Neural Networks
4. Convolutional Neural Networks
5. Image Transformations
6. Image Gradients and Edge Detection
7. Image Contours
8. Image Segmentation
9. Harris Corner Detection
10. Face Detection using Haar Cascades
11. Chatbot Creation

SEMESTER – III **Hours/week: 4**

PDS3504 **Credits : 4**

CLOUD COMPUTING

Unit – I: Introduction

Evolution of Cloud Computing –Essential Characteristics of cloud computing – Operational models such as private, dedicated, virtual private, community, hybrid and public cloud – Service models such as IaaS, PaaS and SaaS – Governance and Change Management – Business drivers, metrics and typical use cases. Example cloud vendors – Google cloud platform, Amazon AWS, Microsoft Azure, Pivotal cloud foundry and Open Stack.

Unit – II: Infrastructure Services

Basics of Virtual Machines - Taxonomy of Virtual Machines. Virtualization Architectures. Challenges with Dynamic Infrastructure - Principles of Infrastructure as Code - Considerations for Infrastructure Services and Tools - Monitoring: Alerting, Metrics, and Logging - Service Discovery - Server Provisioning via Templates - Patterns and Practices for Continuous Deployment - Organizing Infrastructure and Testing Infrastructure - Change Management Pipelines for Infrastructure.

Unit – III: Platform Engineering

Cloud Native Design and Microservices– Containerized - Dynamically orchestrated design – Continuous delivery - Support for a variety of client devices – Monolithic vs Microservices Architecture - Characteristics of microservice architecture – 12 factor application design - Considering service granularity – Scalable Services - Sharing dependencies between microservices - Stateless versus Stateful microservices - Service discovery – Service Registry – Performance Considerations.

Unit – IV: Serverless Architecture and DevOps

Function as a Service (FaaS) - Backend as a Service (BaaS) - Advantages of serverless architectures - Taking a hybrid approach to serverless architecture - Function deployment and Function invocation. Introduction to DevOps - The Deployment Pipeline - The Overall Architecture - Building and Testing - Deployment - Crosscutting Concerns such as Monitoring, Scalability, Repeatability, Reliability, Recoverability, Interoperability, Testability, and Modifiability,

Unit – V: Cloud Security

Security Considerations – STRIDE Threat Model - Cloud Security Challenges – Cloud specific Cryptographic Techniques – CIA Triad – Security by Design – Common Security Risks - Risk Management – Security Monitoring – Security Architecture Design – Data Security – Application Security – Virtual Machine Security.

REFERENCES:

1. Dr.AnandNayyar, (2019), “Handbook of Cloud Computing”, BPB
2. Mastering Azure Machine Learning, By Christoph Korner and Kaijisse Waaijer, April 2020
3. Hands-On Machine Learning on Google Cloud Platform,By Giuseppe Ciaburro, V Kishore Ayyadevara and Alexis Perrier, April 2018
4. Learning Path: AWS Certified Machine Learning-Specialty ML, By Noah Gift, April 2019
5. Software Architect's Handbook, by Joseph Ingeno, Published by Packt Publishing, 2018
6. Architecting Cloud Computing Solutions by Scott Goessling, Kevin L. Jackson, Publisher: Packt Publishing, Release Date: May 2018
7. Microservices: Flexible Software Architecture, by Eberhard Wolff, Publisher: Addison-Wesley Professional, Release Date: October 2016

SEMESTER – III **Hours/week: 4**
PDS3601 **Credits : 3**

ELECTIVE 2A: NATURAL LANGUAGE PROCESSING

.

Unit – I: Introduction

Overview: Origins and challenges of NLP- Theory of Language -Features of Indian Languages – Issues in Font –Models and Algorithms- NLP Applications.

UNIT II - MORPHOLOGY AND PARTS-OF-SPEECH

Phonology – Computational Phonology - Words and Morphemes – Segmentation – Categorization and Lemmatisation – Word Form Recognition – Valency - Agreement - Regular Expressions – Finite State Automata – Morphology- Morphological issues of Indian Languages – Transliteration.

UNIT III - PROBABILISTIC MODELS

Probabilistic Models of Pronunciation and Spelling – Weighted Automata – N- Grams – Corpus Analysis – Smoothing – Entropy - Parts-of-Speech – Taggers – Rule based – Hidden Markov Models – Speech Recognition.

UNIT IV - SYNTAX

Basic Concepts of Syntax – Parsing Techniques – General Grammar rules for Indian Languages – Context Free Grammar – Parsing with Context Free Grammars – Top Down Parser – Earley Algorithm – Features and Unification - Lexicalised and Probabilistic Parsing.

UNIT V - SEMANTICS AND PRAGMATICS (6 hours) Representing Meaning – Computational Representation – Meaning Structure of Language – Semantic Analysis – Lexical Semantics – WordNet – Pragmatics – Discourse – Reference Resolution – Text Coherence – Dialogue Conversational Agents.

REFERENCES:

1. Daniel Jurafskey and James H. Martin “Speech and Language Processing”, Prentice Hall, 2009.
2. Christopher D.Manning and Hinrich Schutze, “Foundation of Statistical Natural Language Processing”, MIT Press, 1999.
3. Ronald Hausser, “Foundations of Computational Linguistics”, Springer-Verleg, 1999.
4. James Allen, “Natural Language Understanding”, Benjamin/Cummings Publishing Co. 1995.
5. Applied Natural Language Processing with Python: Implementing Machine Learning and Deep Learning Algorithms for Natural Language Processing,By Taweh Beysolow II, September 2018

SEMESTER – III **Hours/week: 4**
PDS3602 **Credits : 3**

ELECTIVE 2B: COMPUTER VISION

Here is a rough outline of topics and the number of lectures to be spent on each topic:

- Image formation / projective geometry / lighting *(3 lectures)*
- Practical linear algebra *(2 lectures)*
- Image processing / descriptors *(2 lectures)*
- Image warping *(2 lectures)*
- Linear models + optimization *(2 lectures)*
- Neural networks *(3 lectures)*
- Applications of neural networks *(3 lectures)*
- Motion and flow *(2 lectures)*
- Single-view geometry *(2 lectures)*
- Multi-view geometry *(3 lectures)*
- Applications *(3 lectures)*

Textbooks:

- *Computer Vision: Algorithms and Applications* by Richard Szeliski. Available for free online.
- *Computer Vision: A Modern Approach (Second Edition)* by David Forsyth and Jean Ponce. Available for free online.
- *Elements of Statistical Learning* by Trevor Hastie, Robert Tibshirani, and Jerome Friedman. Available for free online (Warning: Direct PDF link).
- *Multiple View Geometry in Computer Vision (Second Edition)* by Richard Hartley and Andrew Zisserman. Available for free online through the UM Library (Login required).