Pandit Deendayal Energy University, Gandhinagar

School of Technology
Computer Science & Engineering

# M. TECH (Data Science)

Amendment to the Curriculum

# Proposal to BoS

**(Academic Year 2022-24)**

# M.Tech.
# (Data Science)

# I$^{st}$ Semester

**PANDIT DEENDAYAL ENERGY UNIVERSITY GANDHINAGAR**

**SCHOOL OF TECHNOLOGY**

| COURSE STRUCTURE FOR M.TECH - DATA SCIENCE | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Semester I | | | M. Tech. - Data Science | | | | | | | | | | |
| Sr. No. | Course/ Lab Code | Course/ Lab Name | Teaching Scheme | | | | | Examination Scheme | | | | | |
| | | | L | T | P | C | Hrs/ Wk | Theory | | | Practical | | Total |
| | | | | | | | | CE | MS | ES | LW | LE/ Viva | Marks |
| 1 | 20MA503T | Mathematics for Data Science | 3 | 0 | 0 | 3 | 3 | 25 | 25 | 50 | | | 100 |
| 2 | 20DS501T | Foundation of Data Science | 2 | 0 | 0 | 2 | 2 | 25 | 25 | 50 | | | 100 |
| 3 | 20DS502T | Probability & Statistics for Data Science | 3 | 0 | 0 | 3 | 3 | 25 | 25 | 50 | | | 100 |
| 4 | New Code | Pattern Recognition & Machine Learning | 3 | 0 | 0 | 3 | 3 | 25 | 25 | 50 | | | 100 |
| 5 | New Code | Pattern Recognition & Machine Learning Lab | 0 | 0 | 2 | 1 | 2 | | | | 50 | 50 | 100 |
| 6 | 20DS516T | Big Data Analytics | 3 | 0 | 0 | 3 | 3 | 25 | 25 | 50 | | | 100 |
| 7 | 20DS516P | Big Data Analytics Lab | 0 | 0 | 2 | 1 | 2 | | | | 50 | 50 | 100 |
| 8 | 20DS505P | Data Science Lab | 0 | 0 | 4 | 2 | 4 | | | | 50 | 50 | 100 |
| 9 | 20DS506P | Colloquium/Technical Paper presentation | 0 | 0 | 4 | 2 | 4 | | | | 50 | 50 | 100 |
| | | TOTAL | 14 | 0 | 12 | 20 | 26 | 100 | 100 | 200 | 250 | 250 | 900 |

CE- Continuous Evaluation, MS-Mid Semester; ES – End Semester Exam

| 20MA503T | | | | | Mathematics for Data Science | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Teaching Scheme | | | | | Examination Scheme | | | | | |
| L | T | P | C | Hrs/Week | Theory | | | Practical | | Total Marks |
| | | | | | MS | ES | IA | LW | LE/Viva | |
| 3 | 0 | 0 | 3 | 3 | 25 | 50 | 25 | -- | -- | 100 |

**COURSE OBJECTIVES**

- To be acquainted with computational techniques required for performing operations in Data Science.
- To gain advanced algebraic skills essential for Data Science.
- To acquire mathematical understanding of linear systems.
- To formulate and solve problems and present solutions for Data Science applications.

**Unit 1 MATRICES AND BASIC OPERATIONS 10 Hrs.**
Matrices and Basic Operations, interpretation of matrices as linear mappings, and some examples, Properties of determinants, singular and non-singular matrices, examples, finding an inverse matrix, The Range and the Null space of a Matrix, Characteristic Polynomial, Definition of Left/right Eigenvalues and Eigenvectors, Interpretation of eigenvalues/vectors, Caley-Hamilton theorem, Quadratic forms.

**Unit 2 NORMED SPACES, VECTOR SPACES AND MATRIX TRANSFORMATIONS 10 Hrs.**
Definition of complete normed and vector spaces and some examples. Matrix norms and properties - Definition and basic properties, Orthogonality, Orthogonal transformations, Gram-Schmidt algorithm, Singular Value Decomposition: Principal Component Analysis, Gaussian elimination, LU and QR factorization, Definition of positive-definiteness and the role of the eigenvalues, Eigenvalue problems in dimensionality reduction.

**Unit 3 LINEAR SYSTEMS 10 Hrs.**
Definition, applications, solving linear systems, linear inequalities, linear programming, Real-valued functions of two or more variables, Analysis elements: Distance, Limits, continuity, differentiability, the gradient and the Hessian.
.

**Unit 4 OPTIMIZATION PROBLEMS 10 Hrs.**
Motivation, the role of the Hessian, maxima and minima and related extrema conditions, Elements of Convex Optimization Functions of n variables. Convex sets, convex functions, convex problems, and their basic properties. Examples of convex problems, convexity versus non-convexity, Why We Need Gradient Descent, Convergence of Gradient Descent, The Divergence Problem, Bivariate Optimization, Multivariate Optimization

**COURSE OUTCOMES**
On completion of the course, student will be able to
CO1: Interpret existence and uniqueness of solutions using matrix algebra.
CO2: Apply equivalent forms to identify matrices and solve linear systems.
CO3: Apply basic properties of subspaces and vector spaces.
CO4: Compute the orthogonal projection of a vector onto a subspace, given a basis for the subspace.
CO5: Critically analyze and construct mathematical arguments that relate to the study of introductory linear algebra.
CO6: Apply optimization methods and algorithms developed for solving various types of optimization problem.

**TEXT/REFERENCE BOOKS**


1. Lloyd N. Trefethen and David Bau, “Numerical Linear Algebra” III, SIAM, Philadelphia, ISBN 0-89871-361-7
2. Charu C. Agarwal, Linear Algebra & Optimization for Machine Leaning, Springer, 2020.
3. Gilbert Strang, Linear Algebra and Its Applications, Thomson/Brooks Cole
4. Stephen Boyd, Lieven Vandenberghe , Introduction to Applied Linear Algebra, Cambridge University press, 2018.


**END SEMESTER EXAMINATION QUESTION PAPER PATTERN**

**Max. Marks: 100 Exam Duration: 3 Hrs**

Part A: 10 Questions of 2 marks each-No choice 20 Marks
Part B: 2 Questions from each unit with internal choice, each carrying 20 marks 80 Marks

| 20DS501T | | | | | Foundation of Data Science | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Teaching Scheme** | | | | | **Examination Scheme** | | | | | |
| **L** | **T** | **P** | **C** | **Hrs/Week** | **Theory** | | | **Practical** | | **Total Marks** |
| | | | | | **MS** | **ES** | **IA** | **LW** | **LE/Viva** | |
| **2** | **0** | **0** | **2** | **2** | **25** | **50** | **25** | **--** | **--** | **100** |

**COURSE OBJECTIVES**

- To demonstrate proficiency with statistical analysis of data and data management.
- To apply data science concepts and methods to solve problems in real-world contexts and communicate these solutions effectively

**UNIT 1 Data Representation:** **6 Hrs**

Data definition, variety of data, Data representation: Array, Matrix, List tuple, dictionary, stack, queue. Data preprocessing: data cleaning, missing values assimilation, data inconsistency handling, SMOTE algorithm.

**UNIT 2 Data Visualization and Applications of Data science** **6 Hrs**

Introduction, Types of data visualization, Data for visualization: Data types, Data encodings, Retinal variables, mapping variables to encodings, Visual encodings, explore tools libraries: Matplotlib, seaborn.

**UNIT 3 Data analysis:** **5 Hrs**

Introduction, Terminology and concepts, Central tendencies and distributions, Variance, Distribution properties and arithmetic, Maximum likelihood, Bayes estimators, Minimum mean squared error (MMSE), classification, support vector machine.

**UNIT 4 Time Series Data Analysis** **9 Hrs**

Introduction: Examples, simple descriptive techniques, trend, seasonality, the correlogram. Probability models for time series: stationarity. Moving average (MA), Autoregressive (AR), ARMA and ARIMA models. Estimating the autocorrelation function and fitting ARIMA models. Forecasting: Exponential smoothing, Forecasting from ARIMA models. Stationary processes in the frequency domain: The spectral density function, the periodogram, spectral analysis. State-space models: Dynamic linear models and the Kalman filter.

**COURSE OUTCOMES**

On completion of the course, student will be able to
CO1- Explore the fundamental concepts of data science
CO2- Understand data analysis techniques for applications handling large data
CO3- Visualize and present the inference using various tools.
CO4- Learn to think through the ethics surrounding privacy, data sharing and algorithmic decision-making
CO5- Apply the application of data driven solutions in the real world problem
CO6- Explain time series with different structures.

**TEXT/REFERENCE BOOKS**

1. Hector Garcia-Molina, Jennifer Widom, and Jeffrey Ullman. Database Systems: The Complete Handbook, Second edition.
2. Elsmasri and Navathe, Fundamentals of database systems
3. Xun (Brian) Wu, Sudarshan Kadambi, Devram Kandhare, Aaron Ploetz, Seven NoSQL Databases in a Week: Get up and running with the fundamentals, Packt Publishers
4. Raghu Ramakrishnan and Johannes Gehrke, Database management systems

**END SEMESTER EXAMINATION QUESTION PAPER PATTERN**

**Max. Marks: 100** **Exam Duration: 3 Hrs**

Part A: 10 Questions of 2 marks each-No choice 20 Marks
Part B: 2 Questions from each unit with internal choice, each carrying 20 marks 80 Marks

| 20DS502T | | | | | Probability & Statistics for Data Science | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Teaching Scheme | | | | | Examination Scheme | | | | | |
| L | T | P | C | Hrs/Week | Theory | | | Practical | | Total Marks |
| | | | | | MS | ES | IA | LW | LE/Viva | |
| 3 | 0 | 0 | 3 | 3 | 25 | 50 | 25 | | | 100 |

**COURSE OBJECTIVES**

- Identify the independent and dependent variables in a research problem.
- To equip students with consequently requisite quantitative skills that they can employ and build on in flexible ways

**Unit 1 PROBABILITY THEORY & RANDOM VARIABLES** **10 Hrs.**

Calculating probability, Conditional probability, Law of total probability,
**Random variables:** PMF, pdf, cdf, Discrete RVs: Bernoulli, Binomial, Geometric, Indicator, Uniform (a, b), Exponential(λ), Normal (μ, σ2), and its several properties

**10 Hrs.**

**Unit 2: PROBABILITY DISTRIBUTIONS & MARKOV CHAINS**

Joint distributions & conditioning, Joint probability distribution, Linearity and product of expectation, Conditional expectation, Probability Inequalities: Weak Law of Large Numbers, Central Limit Theorem,

**UNIT 3 PARAMETRIC & NON-PARAMETRIC INFERENCES** **10 Hrs.**

Basics of inference, Empirical PMF, Sample mean, Bias, MSE, Empirical Distribution Function (or eCDF)
Kernel Density Estimation (KDE), Statistical Functional, Plug-in estimator, Confidence intervals-Percentiles, quantiles, Normal-based confidence intervals, DKW inequality, Parametric inference

**UNIT 4 HYPOTHESES TESTING & REGRESSION** **9 Hrs.**

Basics of hypothesis testing , Wald Test, Type I and Type II errors , Z-test, t-test, ANOVA, Kolmogorov-Smirnov test (KS test), p-values, permutation test, Pearson correlation coefficient, Chi-square test for independence, Bayesian reasoning & inference,

**Max. 39 Hrs.**

**COURSE OUTCOMES**

On completion of the course, student will be able to
CO1. Understand theoretical foundations of probability theory and mathematical statistics
CO2 Understand the concepts of various parameter estimation methods, like method of moments, maximum likelihood estimation and confidence intervals.
CO3 Apply the central limit theorem to sampling distribution.
CO4. Identify the appropriate hypothesis testing procedure based on type of outcome variable and number of samples
CO5. Analyze hypotheses tests of means, proportions and variances using both one-and two-sample data sets.
CO6. Implement basic simulation methods using statistical software to investigate sampling distributions.

**TEXT/REFERENCE BOOKS**

1. Wasserman, Larry, “All of Statistics: A Concise Course in Statistical Inference” Springer, 2004.
2. S.M. Ross, Introduction to Probability Models, Academic Press
3. Miller & Freund’ Probability and statistics for engineers, ninth edition, Richard a. Johnson, Pearson.
4. Devore. J.L., “Probability and Statistics for Engineering and the Sciences”, Cengage Learning, New Delhi, 8th Edition, 2012.
5. S.M. Ross, Stochastic Processes, Wiley

**END SEMESTER EXAMINATION QUESTION PAPER PATTERN**

**Max. Marks: 100** **Exam Duration: 3 Hrs**

Part A: 5 Questions of 4 marks each-No choice 20 Marks
Part B: 2 Questions from each unit with internal choice, each carrying 20 marks 80 Marks

| New Code | | | | | **Pattern Recognition & Machine Learning** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Teaching Scheme** | | | | | **Examination Scheme** | | | | | |
| **L** | **T** | **P** | **C** | **Hrs/Week** | **Theory** | | | **Practical** | | **Total Marks** |
| | | | | | **MS** | **ES** | **IA** | **LW** | **LE/Viva** | |
| **3** | **0** | **0** | **3** | **3** | **25** | **50** | **25** | | | **100** |

**COURSE OBJECTIVES**

- To introduce the basic concepts and techniques of Machine Learning
- To develop skills using recent machine learning software for solving practical problems

**UNIT 1 INTRODUCTION** **10 Hrs.**
Introduction: Feature Selection and Decision Surfaces, Linear Models for Regression,
**UNIT 2 Linear Models for Classifications** **10 Hrs.**
Discriminant Functions, Probabilistic Models, Graphical Models-Bayesian Decision Theory, Kernel Methods
**UNIT 3 UNSUPERVISED MACHINE LEARNING** **10 Hrs.**
Mixture Models, EM algorithm, Sampling Methods, Continuous Latent Variable-PCA
**UNIT 4 ENSEMBLE METHODS** **9 Hrs.**
The rationale for ensemble method, methods for constructing an Ensemble classifier, Bias-Variance decomposition, Bagging, Boosting, Random forests, Empirical comparison among Ensemble methods.

**Max. 39 Hr**

**COURSE OUTCOMES**
On completion of the course, students will be able to
CO1- Understand Key concepts, tools, and methods for feature engineering
CO2- Understand the theory of statistics in building mathematical models of machine learning
CO3- Evaluate linear models of classification algorithms along with their strengths & weaknesses
CO4- Formulate predictive models corresponding to different applications.
CO5- Apply probabilistic, graphical , and ensemble machine learning models based on their accuracy.
CO6- Develop machine learning-based solutions to the real-world problem, optimize the models learned

**TEXT/REFERENCE BOOKS**

1. Christopher M. Bishop, "Pattern Recognition and Machine Learning", by Springer, 2007
2. Ethem Alpaydin, “Introduction to Machine Learning” MIT Press, 2019
3. Amanda Casari, Alice Zheng, “Feature Engineering for Machine Learning”, O’Reilly, 2018.
4. Andreas Muller, "Introduction to Machine Learning with Python: A Guide for Data Scientists", Shroff/O'Reilly; First edition (2016)

**END SEMESTER EXAMINATION QUESTION PAPER PATTERN**

**Max. Marks: 100** **Exam Duration: 3 Hrs**
Part A: 10 Questions of 2 marks each-No choice 20 Marks
Part B: 2 Questions from each unit with internal choice, each carrying 20 marks 80 Marks

<table>
<tr><td colspan="5">New Code</td><td colspan="6">Pattern Recognition & Machine Learning LAB</td></tr>
<tr><td colspan="5">Teaching Scheme</td><td colspan="6">Examination Scheme</td></tr>
<tr><td rowspan="2">L</td><td rowspan="2">T</td><td rowspan="2">P</td><td rowspan="2">C</td><td rowspan="2">Hrs/Week</td><td colspan="3">Theory</td><td colspan="2">Practical</td><td rowspan="2">Total Marks</td></tr>
<tr><td>MS</td><td>ES</td><td>IA</td><td>LW</td><td>LE/Viva</td></tr>
<tr><td>0</td><td>0</td><td>2</td><td>1</td><td>2</td><td>-</td><td>-</td><td>-</td><td>50</td><td>50</td><td>100</td></tr>
</table>

**COURSE OBJECTIVES**

- To develop skills of using recent machine learning software for solving practical problems
- To gain experience of data analysis and prediction

**LIST OF EXPERIMENT**

Practical list should be prepared by Course Instructor based on the content of the subject. Data sets can be taken from standard repositories (https://archive.ics.uci.edu/ml/datasets.html) or constructed by the students.

**Preferred Programming Language & Platform:** MATLAB, R and Scientific Python (SciPy, NumPy)

1. **Getting Started**
   Local Setup and Development Environment, Python Programming, Flow Control, SciPy Stack, NumPy, Pandas and matplotlib, Statistics, Probability, Calculus and Linear Algebra operations

2. **Statistical Inference**
   Event Space, Probability, Distributions and Hypothesis Testing, Descriptive Statistics, Univariate and Multivariate Exploratory Data Analysis, Data Visualization, Learning & Fitting, Principal Component Analysis, Singular Value Decomposition,

3. **Predictive Modelling**
   Regression, Classification, Data Pre-processing, Model Evaluation and Ensembles, Dimensionality Reduction
   Clustering, Association Rules, Anomaly Detection, Pattern Discovery

4. **Simulations & Course Project**

**COURSE OUTCOMES**

On completion of the course, the students will be able to
CO1-Apply tools and methods for feature engineering
CO2- Apply the theory of statistics in building mathematical models of machine learning
CO3- Evaluate linear models of classification algorithms along with their strengths & weaknesses
CO4- Formulate predictive models corresponding to different applications.
CO5- Apply probabilistic, graphical,and ensemble machine learning models based on their accuracy.
CO6- Develop machine learning-based solutions to the real-world problem, optimize the models learned

**TEXT/REFERENCE BOOKS**

1. Andreas Muller, "Introduction to Machine Learning with Python: A Guide for Data Scientists", First edition (2016) Shroff/O'Reilly
2. Andrew NG's online Course

**END SEMESTER EXAMINATION QUESTION PAPER PATTERN**

**Max. Marks: 100** **Exam Duration: 2 Hrs**

Part A: Continuous Evaluation based on lab records and course project. 50 Marks
Part B: 2 Experiment conducted and Viva at final exam. 50 Marks

| 20DS516T | | | | | Big Data Analytics | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Teaching Scheme | | | | | Examination Scheme | | | | | |
| L | T | P | C | Hrs/Week | Theory | | | Practical | | Total Marks |
| | | | | | MS | ES | IA | LW | LE/Viva | |
| 3 | 0 | 0 | 3 | 3 | 25 | 50 | 25 | - | - | 100 |

**COURSE OBJECTIVES**

- To understand the concept of Big data related tools and techniques.
- Optimize business decisions and create competitive advantage with Big data analytics
- Ability to build and maintain reliable, scalable, distributed systems with Hadoop.
- Evaluate advanced analytics platform and analytics business maturity model

**UNIT 1 INTRODUCTION TO BIG DATA** **10 Hrs.**

Big Data and its Importance – Four V's of Big Data – Drivers for Big Data –Big Data Analytics – Big Data Analytics applications. BIG DATA TECHNOLOGIES: Hadoop's Parallel World – Data discovery – Open source technology for Big Data Analytics – cloud and Big Data

**10 Hrs.**

**UNIT 2 PROCESSING BIG DATA**

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. HADOOP MAPREDUCE, The Building Blocks of Hadoop Map Reduce - Distinguishing Hadoop daemons - Investigating the Hadoop Distributed File System, SPARK based data processing like SQL, ML Lib, Graph Processing.

**UNIT 3 BIG DATA TOOLS AND TECHNIQUES** **10 Hrs.**

Installing and Running Pig – Comparison with Databases – Pig Latin – User Define Functions – Data Processing Operators – Installing and Running Hive – Hive QL – Tables – Querying Data – User-Defined Functions – Oracle Big Data, , SPARK based stream processing, Demonstrating Case Studies.

**9 Hrs.**

**UNIT 4 ADVANCED ANALYTICS PLATFORM**

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. Data Visualization tools: TableU, PowerBI, etc.

**Max. 39 Hrs.**

**COURSE OUTCOMES**

On completion of the course, student will be able to
CO1- Describe Big Data and its importance with its applications.
CO2- Differentiate various big data technologies like Hadoop MapReduce, Pig, Hive, Hbase and No-SQL.
CO3- Apply tools and techniques to analyse Big Data.
CO4- Understand Map Reduce paradigm and the Hadoop system and identify its applicability in real life problems.
CO5- Demonstrate advanced analytics platform for Business.
CO6- Design a solution for a given problem using suitable Big Data Techniques

**TEXT/REFERENCE BOOKS**

1. Michael Minelli, Michehe Chambers, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Business", 1stEdition, AmbigaDhiraj, Wiely CIO Series, 2013.
2. ArvindSathi, "Big Data Analytics: Disruptive Technologies for Changing the Game", 1st 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.
4. Tom White, "Hadoop: The Definitive Guide", 3rd Edition, O'reilly, 2012

**END SEMESTER EXAMINATION QUESTION PAPER PATTERN**

**Max. Marks: 100** **Exam Duration: 3 Hrs**

Part A: 10 Questions of 2 marks each-No choice 20 Marks
Part B: 2 Questions from each unit with internal choice, each carrying 20 marks 80 Marks

<table>
<tr><td colspan="5">20DS516P</td><td colspan="6">Big Data Analytics LAB</td></tr>
<tr><td colspan="5">Teaching Scheme</td><td colspan="6">Examination Scheme</td></tr>
<tr><td rowspan="2">L</td><td rowspan="2">T</td><td rowspan="2">P</td><td rowspan="2">C</td><td rowspan="2">Hrs/Week</td><td colspan="3">Theory</td><td colspan="2">Practical</td><td rowspan="2">Total Marks</td></tr>
<tr><td>MS</td><td>ES</td><td>IA</td><td>LW</td><td>LE/Viva</td></tr>
<tr><td>0</td><td>0</td><td>2</td><td>1</td><td>2</td><td>--</td><td>--</td><td>--</td><td>50</td><td>50</td><td>100</td></tr>
</table>

**COURSE OBJECTIVES**

- Identify the challenges of Big Data Management
- Recognize the key concepts of Hadoop framework, MapReduce and SPARK.
- Apply the tools, techniques and algorithms for big data analysis.

**LIST OF EXPERIMENT**

1. To draw and explain Hadoop Architecture and Ecosystem with the help of a case study using WorkCount example. To define and install Hadoop.
2. To implement the following file management tasks in Hadoop System (HDFS): Adding files and directories, Retrieving files, Deleting files.
3. To run a basic Word Count MapReduce program to understand MapReduce Paradigm: To count words in a given file, To view the output file, and To calculate execution time.
4. To implement Stock count Map reduce program.
5. Write a Map Reduce program that mines weather data. Data available at: https://github.com/tomwhite/hadoopbook/tree/master/input/ncdc/all.
6. Install and Run Hive then use Hive to create, alter, and drop databases, tables, views, functions, and indexes.
7. Install, Deploy & configure Apache Spark Cluster. Run apache spark applications using Scala.
8. Data analytics using Apache Spark on Amazon food dataset, find all the pairs of items frequently reviewed together.
9. Use case - Demonstration
10. Project work: Research article to be submitted as part of LAB manual.

**COURSE OUTCOMES**

On completion of the course, student will be able to
CO1 - Understand the fundamental concepts of Big Data management and analytics
CO2 – Implement the distributed systems with Apache Hadoop.
CO3 - Deploy Hadoop ecosystem components.
CO4 - Apply Map Reduce paradigm for Big Data Analysis.
CO5 – Understand the working of tools (SPARK) and techniques to analyze Big Data
CO5 - Build a solution for a given problem using suitable Big Data Techniques

**TEXT/REFERENCE BOOKS**


1. Chris Eaton et al., *Understanding Big Data*, McGraw Hill, 2011
2. Tom White, *HADOOP: The definitive Guide*, O Reilly, 2009
3. Boris lublinsky et al., *Professional Hadoop Solutions*, Wiley, 2013
4. Donald Miner et al., *MapReduce Design Patterns*, O'Reilly Media, 2012
5. Bill Chambers et al., *Spark: The Definitive Guide*, O'Reilly Media, 2018


**END SEMESTER EXAMINATION QUESTION PAPER PATTERN**

**Max. Marks: 100** **Exam Duration: 2 Hrs**

| | |
|---|---|
| Part A: Evaluation Based on the class performance and Laboratory book | 50 Marks |
| Part B: Viva Examination based conducted experiments | 50 Marks |

| *20DS505P* | | | | | **Data Science LAB** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| ***Teaching Scheme*** | | | | | ***Examination Scheme*** | | | | | |
| ***L*** | ***T*** | ***P*** | ***C*** | ***Hrs/Week*** | ***Theory*** | | | ***Practical*** | | ***Total Marks*** |
| | | | | | ***MS*** | ***ES*** | ***IA*** | ***LW*** | ***LE/Viva*** | |
| ***0*** | ***0*** | ***4*** | ***2*** | ***4*** | ***--*** | ***--*** | ***--*** | ***50*** | ***50*** | ***100*** |

**COURSE OBJECTIVES**

- Implement and apply machine learning algorithms to solve problems
- Select appropriate algorithms for solving an of real-world problems
- Use machine learning techniques in a high-performance computing environment to solve real-world problems.

**LIST OF EXPERIMENT**

**Practi**cal list should be prepared based on the content of the subject. Data sets can be taken from standard repositories (https://archive.ics.uci.edu/ml/datasets.html) or constructed by the students.

**Preferred Programming Language & Platform**: Python/R, Tensorflow/ Matlab, Tableau/ PowerBI,

- Introduction to Exploratory Data Analysis and Visualization
- Overview of the exploratory aspect of data analysis
- Data acquisition from online data sources and preprocessing technique, Graphical Visualization- Visualizing Clusters, Visualization Data Distributions, Multivariate Visualization Graph Data Visualization
- Exploratory Data Analysis for Different Applications: Dimensionality Reduction – Linear and Non-Linear Models, Clustering and Classification, Smoothing Scatterplots and Regression
- ***Course Project: Students are required to pick up one project where they can perform Data pre-processing, feature engineering, suitable ML/statistical technical and demonstrate the results through appropriate data visualization tools***.

**Course Outcome:**

On completion of the course, the student will be able to:

CO1: Evaluate the data analysis techniques for applications handling large data

CO2: Demonstrate the various machine learning algorithms used in data science process

CO3: Understand the ethical practices of data science

CO4: Visualize and present the inference using various tools

CO5: Learn to think through the ethics surrounding privacy, data sharing and algorithmic decision-making

CO6: Implement numerical programming, data handling and visualization.

**TEXT/REFERENCE BOOKS**

1. Joel Grus, Data Science from Scratch: First Principles with Python, O'Reilly, 1st edition, 2015
2. Cathy O'Neil, Rachel Schutt, Doing Data Science, Straight Talk from the Frontline, O' Reilly, 1st edition, 2013
3. Davy Cielen, Arno D. B. Meysman, Mohamed Ali, Introducing Data Science, Manning Publications Co., 1st edition, 2016

**END SEMESTER EXAMINATION QUESTION PAPER PATTERN**

**Max. Marks: 100** **Exam Duration: 2 Hrs**

Part A: Continuous Evaluation based on lab records and performance. 50 Marks

Part B: 2 Experiment conducted and Viva at final exam. 50 Marks

Pandit Deendayal Energy University School of Technology

## COURSE OBJECTIVES

| 20DS506P | | | | | Colloquium/Technical Seminar | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Teaching Scheme | | | | | Examination Scheme | | | | | | |
| L | T | P | C | Hrs/Week | Theory | | | Practical | | Total Marks |
| | | | | | MS | ES | IA | LW | LE/Viva | |
| 0 | 0 | 4 | 2 | 4 | | | | 50 | 50 | 100 |

| 20DS516T | | | | | Big Data Analytics | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Teaching Scheme | | | | | Examination Scheme | | | | | |
| L | T | P | C | Hrs/Week | Theory | | | Practical | | Total Marks |
| | | | | | MS | ES | IA | LW | LE/Viva | |
| 3 | 0 | 0 | 3 | 3 | 25 | 50 | 25 | - | - | 100 |

- Students will develop persuasive speech, present information in a compelling, well-structured, and logical sequence, show depth of knowledge of complex subjects, and develop their ability to synthesize, evaluate and reflect on information.
- Students will be able to show competence in working with a methodology, structuring their oral work, and synthesizing information

## COURSE OUTCOMES

On completion of the course, student will be able to

CO1: Show competence in identifying relevant information, defining, and explaining topics under discussion.
CO2: Use appropriate registers and vocabulary, and will demonstrate command of voice modulation, voice projection, and pacing.
CO3: Demonstrate their understanding of discussions and spark further discussion.
CO4: Apply theories, methods, and knowledge bases from multiple fields to a single question or problem.
CO5: Demonstrate problem-solving skills and apply theoretical knowledge.
CO6: Develop the ability to build and assess data-based models and file systems.

### END SEMESTER EXAMINATION QUESTION PAPER PATTERN

**Max. Marks: 100** **Exam Duration:-**

| | |
|---|---|
| Continuous Internal Evaluation | 50 Marks |
| End Semester Viva Evaluation | 50 Marks |

## COURSE OBJECTIVES

- To understand the concept of Big data related tools and techniques.
- Optimize business decisions and create competitive advantage with Big data analytics
- Ability to build and maintain reliable, scalable, distributed systems with Hadoop.
- Evaluate advanced analytics platform and analytics business maturity model

| 20DS516P | | | | | Big Data Analytics LAB | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Teaching Scheme | | | | | Examination Scheme | | | | | |
| L | T | P | C | Hrs/Week | Theory | | | Practical | | Total Marks |
| | | | | | MS | ES | IA | LW | LE/Viva | |
| 0 | 0 | 2 | 1 | 2 | -- | -- | -- | 50 | 50 | 100 |

**UNIT 1 INTRODUCTION TO BIG DATA** **10 Hrs.**

Big Data and its Importance – Four V's of Big Data – Drivers for Big Data –Big Data Analytics – Big Data Analytics applications. BIG DATA TECHNOLOGIES: Hadoop's Parallel World – Data discovery – Open source technology for Big Data Analytics – cloud and Big Data

**10 Hrs.**

**UNIT 2 PROCESSING BIG DATA**

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.

HADOOP MAPREDUCE, The Building Blocks of Hadoop Map Reduce - Distinguishing Hadoop daemons - Investigating the Hadoop Distributed File System

**UNIT 3 BIG DATA TOOLS AND TECHNIQUES** **10 Hrs.**

Installing and Running Pig – Comparison with Databases – Pig Latin – User Define Functions – Data Processing Operators – Installing and Running Hive – Hive QL – Tables – Querying Data – User-Defined Functions – Oracle Big Data

**9 Hrs.**

**UNIT 4 ADVANCED ANALYTICS PLATFORM**

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.

**Max. 39 Hrs.**

**COURSE OUTCOMES**

On completion of the course, student will be able to

CO1- Describe Big Data and its importance with its applications.

CO2- Differentiate various big data technologies like Hadoop MapReduce, Pig, Hive, Hbase and No-SQL.

CO3- Apply tools and techniques to analyse Big Data.

CO4- Understand Map Reduce paradigm and the Hadoop system and identify its applicability in real life problems.

CO5- Demonstrate advanced analytics platform for Business.

CO6- Design a solution for a given problem using suitable Big Data Techniques

**TEXT/REFERENCE BOOKS**

5. Michael Minelli, Michehe Chambers, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Business", 1stEdition, AmbigaDhiraj, Wiely CIO Series, 2013.
6. ArvindSathi, "Big Data Analytics: Disruptive Technologies for Changing the Game", 1st Edition, IBM Corporation, 2012.
7. 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.
8. Tom White, "Hadoop: The Definitive Guide", 3rd Edition, O'reilly, 2012

**END SEMESTER EXAMINATION QUESTION PAPER PATTERN**

**Max. Marks: 100** **Exam Duration: 3 Hrs**

Part A: 10 Questions of 2 marks each-No choice 20 Marks

Part B: 2 Questions from each unit with internal choice, each carrying 20 marks 80 Marks

**COURSE OBJECTIVES**

- Identify the challenges of Big Data Management
- Recognize the key concepts of Hadoop framework, MapReduce and SPARK.
- Apply the tools, techniques and algorithms for big data analysis.

**LIST OF EXPERIMENT**

1. To draw and explain Hadoop Architecture and Ecosystem with the help of a case study using WorkCount example. To define and install Hadoop.

2. To implement the following file management tasks in Hadoop System (HDFS): Adding files and directories, Retrieving files, Deleting files.
3. To run a basic Word Count MapReduce program to understand MapReduce Paradigm: To count words in a given file, To view the output file, and To calculate execution time.
4. To implement Stock count Map reduce program.
5. Write a Map Reduce program that mines weather data. Data available at: https://github.com/tomwhite/hadoopbook/tree/master/input/ncdc/all.
6. Install and Run Hive then use Hive to create, alter, and drop databases, tables, views, functions, and indexes.
7. Install, Deploy & configure Apache Spark Cluster. Run apache spark applications using Scala.
8. Data analytics using Apache Spark on Amazon food dataset, find all the pairs of items frequently reviewed together.

**COURSE OUTCOMES**

On completion of the course, student will be able to

CO1 - Understand the fundamental concepts of Big Data management and analytics

CO2 – Implement the distributed systems with Apache Hadoop.

CO3 - Deploy Hadoop ecosystem components.

CO4 - Apply Map Reduce paradigm for Big Data Analysis.

CO5 – Understand the working of tools (SPARK) and techniques to analyze Big Data

CO5 - Build a solution for a given problem using suitable Big Data Techniques

**TEXT/REFERENCE BOOKS**


6. Chris Eaton et al., *Understanding Big Data*, McGraw Hill, 2011
7. Tom White, *HADOOP: The definitive Guide*, O Reilly, 2009
8. Boris lublinsky et al., *Professional Hadoop Solutions*, Wiley, 2013
9. Donald Miner et al., *MapReduce Design Patterns*, O'Reilly Media, 2012
10. Bill Chambers et al., *Spark: The Definitive Guide*, O'Reilly Media, 2018


**END SEMESTER EXAMINATION QUESTION PAPER PATTERN**

| **Max. Marks: 100** | **Exam Duration: 2 Hrs** |
| --- | --- |
| Part A: Evaluation Based on the class performance and Laboratory book | 50 Marks |
| Part B: Viva Examination based conducted experiments | 50 Marks |

# II$^{nd}$ Semester

PANDIT DEENDAYAL ENERGY UNIVERSITY GANDHINAGAR

SCHOOL OF TECHNOLOGY

| COURSE STRUCTURE FOR M. TECH - DATA SCIENCE | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Semester II | | | M. Tech. - Data Science | | | | | | | | | | | |
| Sr. No. | Course/ Lab Code | Course/ Lab Name | Teaching Scheme | | | | | Examination Scheme | | | | | | |
| | | | L | T | P | C | Hrs/ Week | Theory | | | Practical | | Total | |
| | | | | | | | | CE | MS | ES | LW | LE/ Viva | Marks | |
| 1 | 20DS507T | Neural Network & Deep Learning | 3 | 0 | 0 | 3 | 3 | 25 | 50 | 25 | | | 100 | |
| 2 | 20DS507P | Neural Network & Deep Learning LAB | 0 | 0 | 2 | 1 | 2 | | | | 50 | 50 | 100 | |
| 3 | New Code | Optimization in Machine Learning | 3 | 0 | 0 | 3 | 3 | 25 | 50 | 25 | | | 100 | |
| 4 | 20DSEXXXT | Department Elective-A | 3 | 0 | 0 | 3 | 3 | 25 | 50 | 25 | | | 100 | |
| 5 | 20DSEXXXP | Department Elective-B LAB | 0 | 0 | 2 | 1 | 1 | | | | | | | |
| 6 | 20DSEXXXT | Department Elective-B | 3 | 0 | 0 | 3 | 3 | 25 | 50 | 25 | | | 100 | |
| 7 | 20DSEXXXP | Department Elective-B LAB | 0 | 0 | 2 | 1 | 2 | | | | 50 | 50 | 100 | |
| 8 | 20DS509P | High Performance Computing LAB | 0 | 0 | 4 | 2 | 4 | | | | 50 | 50 | 100 | |
| 9 | 20DS510P | *Research Project / Capstone Project | 0 | 0 | 6 | 3 | 6 | | | | 50 | 50 | 100 | |
| 10 | 17CE527T | Successful Research and Development Program | 2 | 0 | 0 | 2 | 2 | | | | | | NP/PP | |
| | | TOTAL | 14 | 1 | 14 | 22 | 29 | 100 | 200 | 100 | 200 | 200 | 800 | |

MS = Mid Semester, ES = End Semester; IA = Internal assessment (like Test/quizzes, assignments etc.)

LW = Laboratory work; LE = Laboratory Exam

**Department Elective-A**

| | Subject Code | Subject |
|---|---|---|
| **Elective – A (3-0-2)** | 20CS515T | Machine Learning in Cyber Security |
| | 20CS515P | Machine Learning in Cyber Security lab |
| | New Code | Artificial Intelligence & Reinforcement learning |
| | New Code | Artificial Intelligence & Reinforcement learning Lab |
| | 20DS512T | Time Series Analysis & Forecasting |
| | New Code | Time Series Analysis & Forecasting lab |

**Department Elective-B**

| | Subject Code | Subject |
|---|---|---|
| **Elective-B (3-0-2)** | 20DS515T | Computer Vision |
| | 20DS515P | Computer Vision Lab |
| | 20DS517T | Social Network Analysis |
| | 20DS517P | Social Network Analysis Lab |
| | 20DS518T | Natural Language & Text Mining |
| | 20DS518P | Natural Language & Text Mining Lab |

| 20DS507T | | | | | Neural Network & Deep Learning | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Teaching Scheme | | | | | Examination Scheme | | | | | |
| L | T | P | C | Hrs/Week | Theory | | | Practical | | Total Marks |
| | | | | | MS | ES | IA | LW | LE/Viva | |
| 3 | 0 | 0 | 3 | 3 | 25 | 50 | 25 | | | 100 |

**COURSE OBJECTIVES**

- Ability to learn key concepts in neural network and deep learning, including tools, approaches, and application.
- Solve real-world data- problems using deep learning and neural network.

**UNIT 1 INTRODUCTION TO NEURAL NETWORKS & LEARNING PROCESS** **10 Hrs.**

Neural Network, Human Brain, Models of Neuron, NN as directed graphs, Biological Neural Network, Artificial neuron, ANN architecture, ANN learning, analysis and applications, **Learning Processes**: Introduction, Error correction learning, Memory-based learning, Hebbian learning, Competitive learning, Boltzmann learning, credit assignment problem, learning with and without teacher, learning tasks, Memory and Adaptation.

**UNIT 2 SINGLE LAYER & MULTI-LAYER PERCEPTRON NETWORK** **10 Hrs.**

**Single Layer Perception:** Introduction, Pattern Recognition, Linear classifier, Simple perception, Perception learning algorithm, Modified Perception learning algorithm, Adaptive linear combiner, Continuous perception, Learning in continuous perception. Limitation of Perception **Multi-Layer Perceptron Networks:** Introduction, MLP with 2 hidden layers, Simple layer of a MLP, Delta learning rule of the output layer, Multilayer feed forward neural network with continuous perceptions, Generalized delta learning rule, Back propagation algorithm

**UNIT 3 INTRODUCTION TO DEEP LEARNING** **10 Hrs.**

Neuro architectures as necessary building blocks for the DL techniques, Deep Learning & Neocognitron, Deep Convolutional Neural Networks, Recurrent Neural Networks (RNN)

**UNIT 4 DEEP NEURAL NETWORK** **9 Hrs.**

Feature extraction, Deep Belief Networks, Restricted Boltzman Machines, Autoencoders, Training of Deep neural Networks, Applications and examples (Google, image/speech recognition), Deep Learning Tools: Tensorflow, Caffe, Theano, Torch.

**Max. 39 Hrs.**

**COURSE OUTCOMES**

On completion of the course, student will be able to

CO1- Understand the context of neural networks and deep learning.
CO2- Implement and analyse Neural Network and its applications.
CO3- Develop different single layer/multiple layer Perception learning algorithms.
CO4-Design another class of layered networks using deep learning principles.
CO5- Identify the deep learning algorithms which are more appropriate for various types of learning tasks.
CO6- Implement deep learning algorithms and solve real-world problems.

**TEXT/REFERENCE BOOKS**


1. Simon Haykins, Neural Network- A Comprehensive Foundation, 2nd Edition, 1999, Pearson Prentice Hall, ISBN-13: 978-0-13-147139-9.
2. Ian Goodfellow, Yoshua Bengio and Aaron Courville, Deep Learning, MIT Press, 2016.
3. Christopher M. Bishop. Neural Networks for Pattern Recognition. Oxford University Press, 1996
4. Zurada and Jacek M, Introduction to Artificial Neural Systems, 1992, West Publishing Company, ISBN: 9780534954604


**END SEMESTER EXAMINATION QUESTION PAPER PATTERN**

**Max. Marks: 100** **Exam Duration: 3 Hrs**

Part A: 10 Questions of 2 marks each-No choice 20 Marks
Part B: 2 Questions from each unit with internal choice, each carrying 20 marks 80 Marks

| 20DS507P | | | | | Neural Network & Deep Learning LAB | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Teaching Scheme | | | | | Examination Scheme | | | | | |
| L | T | P | C | Hrs/Week | Theory | | | Practical | | Total Marks |
| | | | | | MS | ES | IA | LW | LE/Viva | |
| 0 | 0 | 2 | 1 | 2 | -- | -- | -- | 50 | 50 | 100 |

**COURSE OBJECTIVES**

- Achieve deeper knowledge of techniques which can be used to train deep networks, and apply them in practice
- Able to interpret one of the most widely used types of deep network: deep convolutional networks

**LIST OF EXPERIMENT**

1. Building the Simplest Neural Network in Simple Python, Use NumPy to Build Neural Networks
2. Extending Neural Network to Use Multiple Samples
3. Understanding Back Propagation
4. Multiple Layers and Back Propagation
5. Parameters Affecting Deep Learning
6. Introduction to Linear Keras
7. Using DL for Vision – Convolution Neural Networks
8. Implementation of a widely used project on their own ('Feedforward Networks for Handwritten Digit Recognition' OR 'Sequence Labelling with Deep Recurrent Networks' OR 'Image Classification with Deep Convolutional Networks')

**COURSE OUTCOMES**

On completion of the course, student will be able to:
CO1- Apply knowledge of statistical theory and methods particularly common problems in economical social sciences especially economics.
CO2 - Explain different network architectures and how these are used in current applications
CO3 - Implement, train, and evaluate neural networks using existing software libraries
CO4 - Implement a problem for CNN and their applications
CO5 - Relate the concepts and techniques introduced in the course to your own research
CO6 - Plan and carry out a research project on neural networks within given time limits

**TEXT/REFERENCE BOOKS**

1. Christopher M. Bishop. Neural Networks for Pattern Recognition. Oxford University Press, 1996.
2. Yoav Goldberg. Neural Network Models in Natural Language Processing. Morgan & Claypool, 2017.
3. Simon O. Haykin. Neural Networks and Learning Machines. Third edition. Prentice Hall, 2008.

## END SEMESTER EXAMINATION QUESTION PAPER PATTERN

**Max. Marks: 100** **Exam Duration: 2 Hrs**

Part A: Evaluation Based on the class performance and Laboratory book 50 Marks
Part B: Viva Examination based conducted experiments 50 Marks

| <Course Code> | | | | | **Optimization in Machine Learning** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Teaching Scheme** | | | | | **Examination Scheme** | | | | | |
| | | | | | **Theory** | | | **Practical** | | **Total Marks** |
| **L** | **T** | **P** | **C** | **Hrs/Week** | **MS** | **ES** | **IA** | **LW** | **LE/Viva** | |
| **3** | **0** | **0** | **3** | **3** | | | | | | **100** |
| **3** | **0** | **0** | **3** | **3** | | | | | | **100** |

**COURSE OBJECTIVES**

- Provides an overview of modern optimization methods, for applications in machine learning and data science.
- Determine the most suitable optimization algorithm for any given task and then apply it to the problem.

| | |
|---|---|
| **UNIT 1 Convex Optimization**<br>Introduction to convex optimization models in data science. Classical examples. Convexity and nonsmooth calculus tools for optimization. Rates of convergence. convex sets, convexity-preserving operations, examples of convex programs (linear programming (LP), second-order cone programming (SOCP), semidefinite programming (SDP)), convex relaxation, KKT conditions, duality | **10 Hrs.** |
| **UNIT 2 Gradient Descent Method**<br>stochastic gradient methods and its variants: stochastic gradient descent (SGD), stochastic variance reduced gradient descent (SVRG), perturbed gradient descent, escape from stationary points (saddle point, local minima), stochastic gradient descent Langevin dynamics (SGLD);<br>**Adaptive methods:** SGD with momentum, AdaGrad, RMSProp, AdaDelta, Adam; | **10 Hrs.** |
| **UNIT 3 Non-convex Optimization & Second Order**<br>Limits and errors of learning. Introduction to (nonconvex) optimization models in supervised machine learning. recent progresses on large-scale nonconvex optimization, Bayesian Optimization<br>**Optimization using second order derivatives:** Newton's method, quasi-Newton's method; dual methods: dual coordinate descent/ascent, Stochastic Dual Coordinate Ascent; | **10 Hrs.** |
| **UNIT 4 Advanced Topics**<br>Optimization with constraints: Subgradient projection method, Franke-Wolfe method, Alternating Direction Method of Multipliers;<br>Optimization with discontinuous objective function: proximal operators and proximal algorithms | **9 Hrs.** |
| | **Max. 39 Hrs.** |

**COURSE OUTCOMES**

On completion of the course, student will be able to
CO1- Evaluate the most important algorithms, function classes, and algorithm convergence guarantees
CO2- Compose existing theoretical analysis with new aspects and algorithm variants.
CO3- Formulate scalable and accurate implementations of the most important optimization algorithms for machine learning applications
CO4- Characterize trade-offs between time, data and accuracy, for machine learning methods
CO5- Apply optimization techniques for the given problems.

**TEXT/REFERENCE BOOKS**

1. A. Beck, First-Order Methods in Optimization, MOS-SIAM Series on Optimization, 2017.
2. S. Bubeck, Convex Optimization: Algorithms and Complexity, Foundations and Trends in Optimization, 2015.
3. Fletcher R., Practical Methods of Optimization, John Wiley, 2000.
4. Research papers

**END SEMESTER EXAMINATION QUESTION PAPER PATTERN**

**Max. Marks: 100** **Exam Duration: 3 Hrs**

Part A: 10 Questions of 2 marks each-No choice 20 Marks
Part B: 2 Questions from each unit with internal choice, each carrying 20 marks 80 Marks

| 17CE527T | | | | | Successful Research and Development Program | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Teaching Scheme | | | | | Examination Scheme | | | | | |
| L | T | P | C | Hrs/Week | Theory | | | Practical | | Total Marks |
| | | | | | MS | ES | IA | LW | LE/Viva | |
| 2 | 0 | 0 | 2 | 2 | 25 | 50 | 25 | | | 100 |

**COURSE OBJECTIVES**

- To develop understanding of the basic framework of research process
- To develop an understanding of various research designs and techniques.
- To identify various sources of information for literature review and data collection.
- To develop an understanding of the ethical dimensions of conducting applied research
- Appreciate the components of scholarly writing and evaluate its quality

**UNIT 1 RESEARCH ORGANIZATION** **9 Hrs.**

Objectives & Goals of a Research Organization, Components of a research organization,
Sponsors & Funding Agencies: Funding Agencies − Types**,** Types of Interface with Funding & Sponsor Agencies**,** Call for Proposals & Opportunity Tracking**,** Types of Proposals & Grants**,** Contracting Vehicles & Arrangements**,** Deliverables, Interim & Final Reviews**,** Cost & Performance Audits**,** Contract Laws

**UNIT 2 DEVELOPMENT OF PROPOSAL WRITING** **9 Hrs.**

Proposals for Research Program Funding: Center & Consortia Proposals, Individual Principal Investigator Proposals, Continuation & Renewal Proposals, Prime Subcontractor Relationships & Contracting, Cost Accounting, Laws and Regulations. intellectual Property & Patent Laws**,** Writing a Successful Research Proposal: Technical Proposal, Management Proposal, Cost Proposal, Technology Proposal, Statement of Work & Deliverables, Case Studies

**UNIT 3 DEVELOPMENT OF RESEARCH METHODOLOGY** **9 Hrs.**

**The Research Process – I**: Steps in development of successful research program, Quality and Cost consideration, Laboratories and infrastructure setup, Staffing & Support Models, Peer−Review, Independent Verification & Validation,

**UNIT 4 ETHICS & REGULATORY LAWS** **9 Hrs.**

Internal & External Review processes, Ethics & Regulatory Laws & Guidelines, Case Studies.

**COURSE OUTCOMES**

On completion of the course, student will be able to

**CO1** - I**dentify** the overall process of designing a research study from its inception >

**CO2** - U**nderstand** the characteristics of various kinds of research (quantitative and qualitative).>

**CO3** - A**pply** the knowledge of a forward chronological, backward chronological and manual search methods in framing the literature review for a scholarly educational study>

**CO4** - A**nalyze** with conducting scholarly educational study: a. The steps in the overall process. b. The types of databases often searched. c. The criteria for evaluating the quality of a study. d. The ways of organizing the material found. e. The different types of literature reviews>

**CO5** - E**xercise** on various Ethical issues in conducting research>

**CO6** - **Develop** research designs and project proposals in achieving project deliverables in stipulated period of time and cost>

**TEXT/REFERENCE BOOKS**

1. <Research Methodology (Methods and Techniques) book by CR Kothari New age Publications 3rd edition>
2. <Research Methodology book by Ranjith Kumar, Sage Publications 3rd edition (Softcopy Available)>
3. Nptel Lectures: Introduction to Research,Prof. Prathap Haridoss, Department of Metallurgical and Materials Engineering, Indian Institute of Technology, Madras
4.

**END SEMESTER EXAMINATION QUESTION PAPER PATTERN**

**Max. Marks: 100** **Exam Duration: 3 Hrs**

Part A/Question1: <identifying overall research process> <> Marks
Part A/Question2: <relation between quantitative and qualitative> <> Marks
Part A/Question3: <literature review process> <> Marks
Part A/Question4: <hypothesizing and concept building> <> Marks
Part A/Question5: <Ethical issues in conducting research> <> Marks

| 20DS509P | | | | | High Performance Computing LAB | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Teaching Scheme** | | | | | **Examination Scheme** | | | | | |
| **L** | **T** | **P** | **C** | **Hrs/Week** | **Theory** | | | **Practical** | | **Total Marks** |
| | | | | | **MS** | **ES** | **IA** | **LW** | **LE/Viva** | |
| **0** | **0** | **4** | **2** | **4** | **-** | **-** | **-** | **50** | **50** | **100** |

**COURSE OBJECTIVES**

- Understand the working of high-performance computing with the graphics processing units and many integrated cores using their architectures and corresponding programming environments.
- Implement parallel algorithms through the GPU and XEON Phi programming environments.

**LIST OF EXPERIMENT**

Practical list should be prepared by Course Instructor based on the content of the subject.
Preferred Programming Language & Platform**:** CUDA, Xeon Phi, OpenMP, and MPI programming.

| Sl. No. | Title | Contents |
|---|---|---|
| 1. | **GPU Programming** | Device Query, Vector Addition, Matrix Multiplication, Tiled Matrix Multiplication, Picture Scaling, Image Blur, Image Grayscaling. 1D, 2D, and 3D Stencil Operations. Histogramming, Convolution, Scan, Reduction. |
| 2. | **Xeon Phi Programming** | Vector Addition, Matrix Multiplication, Tiled Matrix Multiplication, Picture Scaling, Image Blur, Image Grayscaling. 1D, 2D, and 3D Stencil Operations. Histogramming, Convolution, Scan, Reduction. |
| 3. | **OpenMP programming** | Matrix Multiply, Calculation of pi using worksharing and reduction, Producer consumer problem, |
| 4. | **MPI programming** | DAXPY, Calculation of π - MPI Bcast and MPI Reduce, Ocean Kernel, Reduction example, Collective Communication - Scatter – Gather, MPI Derived Datatypes, Matrix Multiplication on a Cartesian Grid (2D Mesh) using Cannon’s Algorithm, Martix Multiplication using Cannon’s Algorithm for Large Matrices. |

**COURSE OUTCOMES**

On completion of the course, student will be able to
CO1- Formulate high performance versions of standard single threaded algorithms
CO2- Demonstrate the architectural features in the GPU and MIC hardware accelarators.
CO3- Design programs to extract maximum performance in a multicore, shared memory execution environment processor.
CO4- Deploy large scale parallel programs on tightly coupled parallel systems using the message passing paradigm.
CO5- Compare performance metrics from the perspectives of Programming, Memory, Computational, Processor Architecture.
CO6- Deploy Components -off-the-shelf (COTS) to enable High performance computing environment.

**TEXT/REFERENCE BOOKS**

1. Rezaur Rahman, Intel Xeon Phi Coprocessor Architecture and Tools, Apress Open, 2013
2. Wen-Mei W Hwu, David B Kirk, Programming Massively Parallel Processors A Hands-on Approach, Morgann Kaufmann, 3e

**END SEMESTER EXAMINATION QUESTION PAPER PATTERN**

**Max. Marks: 100** **Exam Duration: 2 Hrs**
Part A: Continuous Evaluation based on lab records and course project. 50 Marks
Part B: 2 Experiment conducted and Viva at final exam. 50 Marks

# Departmental Elective -A

# (3-0-2)

| 20DS512T | | | | | Time Series Analysis & Forecasting | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Teaching Scheme** | | | | | **Examination Scheme** | | | | | |
| **L** | **T** | **P** | **C** | **Hrs/Week** | **Theory** | | | **Practical** | | **Total Marks** |
| | | | | | **MS** | **ES** | **IA** | **LW** | **LE/Viva** | |
| **3** | **0** | **0** | **3** | **3** | **25** | **50** | **25** | | | **100** |

**COURSE OBJECTIVES**

- Define and explain concept of time series, including trend, seasonal effects, and cyclical effects.
- Recognize the role of transformations for time series and identify possible transformations to address certain non-stationary features of a series, such as non-constant variance and multiplicative seasonal effects.

**UNIT 1 AN INTRODUCTION TO FORECASTING & TIME SERIES REGRESSION** **10 Hrs.**

Forecasting and Data. Forecasting Methods. Errors in Forecasting. Choosing a Forecasting Technique. An Overview of Quantitative Forecasting Techniques. Multiple Linear Regressions, Model Building and Residual Analysis **Time Series Regression:** Modelling Trend by Using Polynomial Functions. Detecting Autocorrelation.

**UNIT 2 ARMA & EXPONENTIAL SMOOTHING** **10 Hrs.**

Autoregressive-moving average models ARMA (p,q) Decomposition Methods: Multiplicative Decomposition. Additive Decomposition. The X-12-ARIMA Seasonal Adjustment Method. Exponential Smoothing: Simple Exponential Smoothing. Tracking Signals. Holt's Trend Corrected Exponential Smoothing. Holt-Winters Methods. Damped Trends and Other Exponential

**UNIT 3 NON-SEASONAL BOX-JENKINS MODELLING** **10 Hrs.**

Stationary and Nonstationary Time Series. The Sample Autocorrelation and Partial Autocorrelation Functions: The SAC and SPAC. An Introduction to Non-seasonal Modelling and Forecasting. Tentative Identification of Non-seasonal Box-Jenkins Models. Estimation, Diagnostic Checking, and Forecasting for Non-seasonal Box-Jenkins Models

**UNIT 4 BOX-JENKINS SEASONAL MODELLING & CAUSALITY IN TIME-SERIES** **9 Hrs.**

Box-Jenkins Seasonal Modelling: Transforming a Seasonal Time Series into a Stationary Time Series. Examples Box-Jenkins Error Term Models in Time Series Regression, Advanced Box-Jenkins Modelling, Causality in time series

**Max. 39 Hrs.**

**COURSE OUTCOMES**

On completion of the course, student will be able to
CO1- Describe the fundamental advantage and necessity of forecasting in various situations.
CO2- Identify how to choose an appropriate forecasting method in a environment.
CO3- Apply various forecasting methods, which include obtaining the relevant data and carrying out the necessary computation using suitable statistical software.
CO4- Improve forecast with better statistical models based on statistical analysis
CO5- Describe the behaviour of the correlogram for series that alternate, have a trend, or show seasonal fluctuations.
CO6- Make a prediction to real time data provided by problem in a time series context.

**TEXT/REFERENCE BOOKS**

1. Bruce L. Bowerman, Richard O'Connell, Anne Koehler, "Forecasting, Time Series, and Regression, 4th Edition", Cengage Unlimited Publishers
2. Enders W. Applied Econometric Time Series. John Wiley & Sons, Inc., 1995
3. Mills, T.C. The Econometric Modelling of Financial Time Series. Cambridge University Press, 1999
4. P. J. Brockwell, R. A. Davis, Introduction to Time Series and Forecasting. Springer, 1996

**END SEMESTER EXAMINATION QUESTION PAPER PATTERN**

**Max. Marks: 100** **Exam Duration: 3 Hrs**

Part A: 10 Questions of 2 marks each-No choice 20 Marks
Part B: 2 Questions from each unit with internal choice, each carrying 20 marks 80 Marks

<table>
<tr><td colspan="5"><Course Code></td><td colspan="6">Time Series Analysis & Forecasting lab</td></tr>
<tr><td colspan="5">Teaching Scheme</td><td colspan="6">Examination Scheme</td></tr>
<tr><td rowspan="2">L</td><td rowspan="2">T</td><td rowspan="2">P</td><td rowspan="2">C</td><td rowspan="2">Hrs/Week</td><td colspan="3">Theory</td><td colspan="2">Practical</td><td rowspan="2">Total Marks</td></tr>
<tr><td>MS</td><td>ES</td><td>IA</td><td>LW</td><td>LE/Viva</td></tr>
<tr><td>0</td><td>0</td><td>0</td><td>1</td><td>2</td><td></td><td></td><td></td><td>50</td><td>50</td><td>100</td></tr>
</table>

**COURSE OBJECTIVES**

- Apply the concept of time series, including trend, seasonal effects, and cyclical effects onto the dataset.
- Apply transformations for time series and identify possible transformations to address certain non-stationary features of a series, such as non-constant variance and multiplicative seasonal effects.

**LIST OF EXPERIMENT**

**Practi**cal list should be prepared based on the content of the subject. Data sets can be taken from standard repositories (https://archive.ics.uci.edu/ml/datasets.html) or constructed by the students.

**Preferred Programming Language & Platform**: Python/R, Tensorflow/ Matlab,

- Setting Feature Derivation Window (FDW) for feature engineering
- Evaluating with Accuracy over Time, Stability, Forecasting Accuracy
- Working with time series feature lists
- Feature selection with Feature Impact
- Causality Analysis
- Seasonality analysis
- Build multivariate time series models to forecast unemployment and learn to iterate and improve on your initial results.
-

**COURSE OUTCOMES**

On completion of the course, student will be able to

CO1- Apply Feature engineering for Time series data.

CO2- Identify how to choose an appropriate forecasting method in an environment.

CO3- Apply various forecasting methods, which include obtaining the relevant data and carrying out the necessary computation using suitable statistical software.

CO4- Improve forecast with better statistical models based on statistical analysis

CO5- Describe the behaviorr of the correlogram for series that alternate, have a trend, or show seasonal fluctuations.

CO6- Make a prediction to real time data provided by problem in a time series context.

**TEXT/REFERENCE BOOKS**

1. Bruce L. Bowerman, Richard O'Connell, Anne Koehler, “Forecasting, Time Series, and Regression, 4th Edition”, Cengage Unlimited Publishers
2. Enders W. Applied Econometric Time Series. John Wiley & Sons, Inc., 1995
3. Mills, T.C. The Econometric Modelling of Financial Time Series. Cambridge University Press, 1999
4. P. J. Brockwell, R. A. Davis, Introduction to Time Series and Forecasting. Springer, 1996

| **Max. Marks: 100** | **Exam Duration: 2 Hrs** |
|---|---|
| Part A: Evaluation Based on the class performance and Laboratory book | 50 Marks |
| Part B: Viva Examination based conducted experiments | 50 Marks |

**END SEMESTER EXAMINATION QUESTION PAPER PATTERN**

Pandit Deendayal Energy University — School of Technology

| New Code | | | | | **Artificial Intelligence & Machine Learning** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Teaching Scheme** | | | | | **Examination Scheme** | | | | | |
| **L** | **T** | **P** | **C** | **Hrs/Week** | **Theory** | | | **Practical** | | **Total Marks** |
| | | | | | **MS** | **ES** | **IA** | **LW** | **LE/Viva** | |
| **3** | **0** | **0** | **3** | **3** | **25** | **50** | **25** | | | **100** |

**COURSE OBJECTIVES**

- Identify the problems where AI is required and the different methods available to solve using AI based techniques
- Learn how to define RL tasks and the core principals behind the RL, including policies, value functions, deriving Bellman equations.
- Recognize current advanced techniques and applications in RL

| | |
|---|---|
| **UNIT 1**<br>Introduction to AI: AI Problems, Intelligent Agents, Problem Formulation, Basic Problem Solving Methods. Searching, Adversarial Search, Simulated Annealing, Measure of performance and analysis of search algorithms. Constraint Satisfaction Problem | **9 Hrs.** |
| **UNIT 2: Knowledge, Reasoning & Planning**<br>Logical Agents, First-order Logic, Inference in First-order Logic, Knowledge Representation<br>Uncertain Knowledge & reasoning, Learning from Examples | **10 Hrs.** |
| **UNIT 3 Foundations & Tabular methods and Q-networks**<br>Introduction and Basics of RL, Defining RL Framework and Markov Decision Process, Policies, Value Functions and Bellman Equations, Exploration vs. Exploitation<br>Planning through the use of Dynamic Programming and Monte Carlo, Temporal-Difference learning methods (TD(0), SARSA, Q-Learning), Deep Q-networks | **10 Hrs.** |
| **UNIT 4 Policy optimization & Recent Advances**<br>Introduction to policy-based methods, Vanilla Policy Gradient, REINFORCE algorithm and stochastic policy search, Actor-critic methods (A2C, A3C), Advanced policy gradient (PPO, TRPO, DDPG), Model-based RL approach, Meta-learning, Multi-Agent Reinforcement Learning, Partially Observable Markov Decision Process, Ethics in RL , Applying RL for real-world problems | **10 Hrs.** |
| | **Max. 39 Hrs.** |

**COURSE OUTCOMES**

Upon completion of the course, the students will be able to

CO1- Identify the AI-based problems and apply techniques to solve the AI problems
CO2- Define learning and explain various knowledge representation & learning techniques
CO3- Learn how to define RL tasks and the core principles behind the RL, including policies, value functions, and deriving Bellman equations.
CO4- Learn the policy gradient methods from vanilla to more complex cases.
CO5- Explore imitation learning tasks and RL-based solutions for real-world problems.

**TEXT/REFERENCE BOOKS**


1. Russell, S.J. and Norvig, P., Artificial Intelligence: A Modern Approach, Pearson Education.
2. Richard S. Sutton and Andrew G. Barto, "Reinforcement learning: An introduction", Second Edition, MIT Press, 2019.
3. Li, Yuxi. "Deep reinforcement learning." arXiv preprint arXiv:1810.06339 (2018)..
4. Wiering, Marco, and Martijn Van Otterlo. "Reinforcement learning." Adaptation, learning, and optimization 12 (2012): 3.
5. Russell, Stuart J., and Peter Norvig. "Artificial intelligence: a modern approach."Pearson Education Limited, 2016.


**END SEMESTER EXAMINATION QUESTION PAPER PATTERN**

**Max. Marks: 100** — **Exam Duration: 3 Hrs**

Part A: 10 Questions of 2 marks each-No choice — 20 Marks
Part B: 2 Questions from each unit with internal choice, each carrying 20 marks — 80 Marks

| <Course Code> | | | | | **Artificial Intelligence & Reinforcement Learning Lab** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Teaching Scheme** | | | | | **Examination Scheme** | | | | | |
| **L** | **T** | **P** | **C** | **Hrs/Week** | **Theory** | | | **Practical** | | **Total Marks** |
| | | | | | **MS** | **ES** | **IA** | **LW** | **LE/Viva** | |
| **0** | **0** | **2** | **1** | **2** | | | | **50** | **50** | **100** |

**COURSE OBJECTIVES**

- To understand tool and technologies related to AI based techniques
- To apply RL based solution for real world problems

**List of Experiments:**

1. Program for search based algorithm
2. Program to generate the output for A* algorithm
3. Program using Heuristic functions
4. Write a program for expert system using Forward Chaining
7. Hands-on on Matlab/Python for AI related problems like Neural Network, Genetic Algorithm, etc.
8. Project work as decided by Tutor. (all tools related to AI can be explored)
9. Approximate solutions to optimal-control problems that are too large or too ill-defined for classical solution methods such as dynamic programming
10. Reinforcement Learning, Deep Learning, Statistical Learning Theory, Multi-agent Systems, Game Theory and Mechanism Design

**COURSE OUTCOMES**

Upon completion of the course, the students will be able to

CO1- Apply search techniques to solve the AI problems
CO2- Apply knowledge representation & learning techniques
CO3- Apply RL based learning for moderate complexity problems
CO4- Implement the policy gradient methods from vanilla to more complex cases.
CO5- Analyze imitation learning tasks and RL-based solutions for real-world problems.

**TEXT/REFERENCE BOOKS**


1. Russell, S.J. and Norvig, P., Artificial Intelligence: A Modern Approach, Pearson Education.
2. Richard S. Sutton and Andrew G. Barto, "Reinforcement learning: An introduction", Second Edition, MIT Press, 2019.
3. Li, Yuxi. "Deep reinforcement learning." arXiv preprint arXiv:1810.06339 (2018)..
4. Wiering, Marco, and Martijn Van Otterlo. "Reinforcement learning." Adaptation, learning, and optimization 12 (2012): 3.
5. Russell, Stuart J., and Peter Norvig. "Artificial intelligence: a modern approach."Pearson Education Limited, 2016.


**END SEMESTER EXAMINATION QUESTION PAPER PATTERN**

**Max. Marks: 100** **Exam Duration: 2 Hrs**

Part A: Evaluation Based on the class performance and Laboratory book 50 Marks

Part B: Viva Examination based conducted experiments 50 Marks

<table>
<tr><td colspan="5">20CS515T</td><td colspan="6">Machine Learning in Cyber Security</td></tr>
<tr><td colspan="5">Teaching Scheme</td><td colspan="6">Examination Scheme</td></tr>
<tr><td rowspan="2">L</td><td rowspan="2">T</td><td rowspan="2">P</td><td rowspan="2">C</td><td rowspan="2">Hrs/Week</td><td colspan="3">Theory</td><td colspan="2">Practical</td><td rowspan="2">Total Marks</td></tr>
<tr><td>MS</td><td>ES</td><td>IA</td><td>LW</td><td>LE/Viva</td></tr>
<tr><td>3</td><td>0</td><td>0</td><td>3</td><td>3</td><td>25</td><td>50</td><td>25</td><td></td><td></td><td>100</td></tr>
</table>

**COURSE OBJECTIVES**

- To understand the concepts of machine learning for cyber security.
- To learn how machine learning can be used to solve various security issues.
- To learn how to implement machine learning algorithms for cyber security.

**UNIT 1 INTRODUCTION** **10 Hrs.**

Why Machine Learning (ML) and Security: Cyber Threat Landscape, The Cyber Attacker's Economy, What Is ML, Real-World Uses of ML in Security, Spam Fighting: An Iterative Approach, Limitations of ML in Security , Classifying and Clustering: ML: Problems and Approaches, Training Algorithms to Learn, Supervised Classification Algorithms, Practical Considerations in Classification, Clustering

**10 Hrs.**

**UNIT 2 ANOMALY DETECTION WITH ML**

**Anomaly Detection**: Anomaly Detection Versus Supervised Learning, Intrusion Detection with Heuristics, Data-Driven Methods, Feature Engineering for Anomaly Detection, Anomaly Detection with Data and Algorithms, Challenges of Using Machine Learning in Anomaly Detection, Response and Mitigation

**UNIT 3 ANALYSIS WITH ML** **10 Hrs.**

**Malware Analysis**: Understanding Malware, Feature Generation, From Features to Classification

**Network Traffic Analysis**: Theory of Network Defense, Machine Learning and Network Security, Building a Predictive Model to Classify Network Attacks

**9 Hrs.**

**UNIT 4 PRODUCTION SYSTEMS AND ADVERSARIAL ML**

**Production Systems**: Defining Machine Learning System Maturity and Scalability, Data Quality, Model Quality, Performance, Maintainability, Monitoring and Alerting, Security and Reliability, Feedback and Usability

**Adversarial Machine Learning**: Terminology, The Importance of Adversarial ML,Security Vulnerabilities in Machine Learning Algorithms

**Max. 39 Hrs.**

**COURSE OUTCOMES**

On completion of the course, student will be able to

CO1- Understand and explain the concepts of machine learning for cyber security.

CO2- Learn how machine learning can be used to solve various security issues

CO3- Compare performance of machine learning algorithms for a security problem

CO4- Determine selection of a machine learning algorithm for anomaly detection.

CO5- Analyze malware and network traffic.

CO6- Design machine learning solutions for cyber security.

**TEXT/REFERENCE BOOKS**

1. Machine Learning & Security: Protecting Systems with Data and Algorithms, by Clarence Chio & David Freeman (O'Reilly).
2. Applications of Data Mining in Computer Security, Daniel Barbará, Sushil Jajodia, (Springer).

**END SEMESTER EXAMINATION QUESTION PAPER PATTERN**

**Max. Marks: 100** **Exam Duration: 3 Hrs**

Part A: 10 Questions of 2 marks each-No choice 20 Marks

Part B: 2 Questions from each unit with internal choice, each carrying 20 marks 80 Marks

| 20CS515T | | | | | Machine Learning in Cyber Security Lab | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Teaching Scheme | | | | | Examination Scheme | | | | | |
| L | T | P | C | Hrs/Week | Theory | | | Practical | | Total Marks |
| | | | | | MS | ES | IA | LW | LE/Viva | |
| 0 | 0 | 2 | 1 | 2 | - | - | - | 50 | 50 | 100 |

**COURSE OBJECTIVES**

- To understand the concepts of machine learning for cyber security.
- To learn how machine learning can be used to solve various security issues.
- To learn how to implement machine learning algorithms for cyber security.

**LIST OF EXPERIMENT**

Practical list should be prepared based on the content of the subject and following guidelines should be useful. Experiment Sessions using Programming would be based on following topics:

1. ML in Cyber Security.
2. Anomaly detection.
3. Malware analysis.
4. Network traffic analysis.
5. Making ML solution reliable.
6. Other experiments.

**COURSE OUTCOMES**

On completion of the course, student will be able to

CO1- Understand and explain the concepts of machine learning for cyber security.
CO2- Learn how machine learning can be used to solve various security issues
CO3- Compare performance of machine learning algorithms for a security problem
CO4- Determine selection of a machine learning algorithm for anomaly detection.
CO5- Analyze malware and network traffic.
CO6- Design machine learning solutions for cyber security.

**END SEMESTER EXAMINATION QUESTION PAPER PATTERN**

**Max. Marks: 100** **Exam Duration: 2 Hrs**

Part A: Continuous Evaluation based on lab records and course project. 50 Marks
Part B: 2 Experiment conducted and Viva at final exam. 50 Marks

# Departmental Elective -B
# (3-0-2)

| 20DS515T | | | | | Computer Vision | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Teaching Scheme** | | | | | **Examination Scheme** | | | | | |
| **L** | **T** | **P** | **C** | **Hrs/Week** | **Theory** | | | **Practical** | | **Total Marks** |
| | | | | | **MS** | **ES** | **IA** | **LW** | **LE/Viva** | |
| **3** | **0** | **0** | **3** | **3** | **25** | **50** | **25** | | | **100** |

**COURSE OBJECTIVES**

- To introduce students the fundamentals of image formation; major ideas, methods, and techniques of computer vision
- To develop an appreciation for various issues in the design of computer vision and object recognition systems; and
- To provide the student with programming experience from implementing computer vision and object recognition applications

**UNIT 1 DIGITAL IMAGE FORMATION AND LOW-LEVEL PROCESSING** **10 Hrs.**

Overview and State-of-the-art, Fundamentals of Image Formation, Transformation: Orthogonal, Euclidean, Affine, Projective, etc; Fourier Transform, Convolution and Filtering, Image Enhancement, Restoration, Histogram Processing. Depth estimation and Multi-camera views: Perspective, Binocular Stereopsis: Camera and Epipolar Geometry; Homography, Rectification, DLT, RANSAC, 3-D reconstruction framework; Auto-calibration

**UNIT 2 FEATURE EXTRACTION** **10 Hrs.**

Edges - Canny, LOG, DOG; Line detectors (Hough Transform), Corners - Harris and Hessian Affine, Orientation Histogram, SIFT, SURF, HOG, GLOH, Scale-Space Analysis- Image Pyramids and Gaussian derivative filters, Gabor Filters and DWT.

**UNIT 3 IMAGE SEGMENTATION** **10 Hrs.**

Region Growing, Edge Based approaches to segmentation, Graph-Cut, Mean-Shift, MRFs, Texture Segmentation; Object detection, **Pattern Analysis**: Clustering: K-Means, K-Medoids, Mixture of Gaussians, Classification: Discriminant Function, Supervised, Un-supervised, Semi-supervised; Classifiers: Bayes, KNN, ANN models; Dimensionality Reduction: PCA, LDA, ICA; Non-parametric methods.

**UNIT 4 MOTION ANALYSIS** **9 Hrs.**

Background Subtraction and Modelling, Optical Flow, KLT, Spatio-Temporal Analysis, Dynamic Stereo; Motion parameter estimation. **Shape from X:** Light at Surfaces; Phong Model; Reflectance Map; Albedo estimation; Photometric Stereo; Use of Surface Smoothness Constraint; Shape from Texture, color, motion and edges.

**Max. 39 Hrs.**

**COURSE OUTCOMES**

On completion of the course, student will be able to

CO1- Identify basic concepts, terminology, theories, models, and methods in the field of computer vision.

CO2- Describe basic methods of computer vision related to multi-scale representation, edge detection and detection of other primitives, stereo, motion, and object recognition.

CO3-Choose appropriate image processing methods for image filtering, image restoration, image reconstruction, segmentation, classification, and representation.

CO4- Improve forecast with better statistical models based on statistical analysis

CO5- Assess which methods to use for solving a given problem, and analyse the accuracy of the methods

CO6- Develop and apply computer vision techniques for solving practical problems.

**TEXT/REFERENCE BOOKS**

1. Richard Szeliski, Computer Vision: Algorithms and Applications, Springer-Verlag London Limited 2011.
2. D. A. Forsyth, J. Ponce, Computer Vision: A Modern Approach, Pearson Education, 2003.
3. Richard Hartley and Andrew Zisserman, Multiple View Geometry in Computer Vision, Second Edition, Cambridge University Press, March 2004
4. K. Fukunaga; Introduction to Statistical Pattern Recognition, Second Edition, Academic Press, Morgan Kaufmann, 1990.

**END SEMESTER EXAMINATION QUESTION PAPER PATTERN**

**Max. Marks: 100** **Exam Duration: 3 Hrs**

Part A: 10 Questions of 2 marks each-No choice 20 Marks

Part B: 2 Questions from each unit with internal choice, each carrying 20 marks 80 Marks

| 20DS515P | | | | | Computer Vision LAB | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Teaching Scheme** | | | | | **Examination Scheme** | | | | | |
| **L** | **T** | **P** | **C** | **Hrs/Week** | **Theory** | | | **Practical** | | **Total Marks** |
| | | | | | **MS** | **ES** | **IA** | **LW** | **LE/Viva** | |
| **0** | **0** | **2** | **1** | **2** | **-** | **-** | **-** | **50** | **50** | **100** |

**COURSE OBJECTIVES**

- Insight into image and video formation design, modelling and analysis.
- Ability to work with features above the pixel level.
- Develop ability to understand the difference in theory and practice of Computer Vision.

**LIST OF EXPERIMENTS**

1. Digital Video Stabilization through curve warping techniques
2. Automatic Target Detection and tracking for thermal image sequences
3. Human Activity analysis based on pose detection
4. Action Recognition in Videos
5. Multiple objects tracking using multiple cameras
6. Camera placement and network surveillance
7. Analysis and annotation of cricket videos
8. Foreground extraction and object tracking, Human activity representation, analysis, and recognition, Multi Camera Pan-Tilt Surveillance Networks, Unsupervised Object Categorization from Surveillance Videos, Visual Recognition of Hand Gestures

**COURSE OUTCOMES**

On completion of the course, student will be able to
CO1- Define low level to high level vision
CO2- Explain use of computer vision in real time applications
CO3- Implement classification, semantic segmentation, tracking, person identification.
CO4- Apply computer vision fundamentals for object tracking and human activity representation.
CO5- Choose appropriate computer vision method for a given problem statement
CO6- Create models based on deep neural networks.

**TEXT/REFERENCE BOOKS**


1. Forsyth and Ponce, Computer Vision: A Modern Approach, , Pearson Education
2. Simon Prince, Computer Vision: Models, Learning, and Interface, Cambridge University Press.
3. Rajalingappaa Shanmugamani, Deep learning for Computer Vision, PACKT publishers.
4. Suetens, P. Fundamentals of Medical Imaging, Cambridge University Press


**END SEMESTER EXAMINATION QUESTION PAPER PATTERN**

**Max. Marks: 100** **Exam Duration: 2 Hrs**
Part A: Evaluation Based on the class performance and Laboratory book 50 Marks
Part B: Viva Examination based conducted experiments 50 Marks

| 20DS517T | | | | | Social Network Analysis | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Teaching Scheme** | | | | | **Examination Scheme** | | | | | |
| **L** | **T** | **P** | **C** | **Hrs/Week** | **Theory** | | | **Practical** | | **Total Marks** |
| | | | | | **MS** | **ES** | **IA** | **LW** | **LE/Viva** | |
| **3** | **0** | **0** | **3** | **3** | **25** | **50** | **25** | | | **100** |

**COURSE OBJECTIVES**

- To know basic notation and terminology used in network science
- To Understand basic principles behind social network analysis algorithms
- To develop skills of using social network analysis software on real world data
- Be capable of analysing real work networks

**UNIT 1 INTRODUCTION** **8 Hrs.**

Social Networks: An Introduction; Types of Networks: General Random Networks, Small World Networks, Scale-Free Networks; Examples of Information Networks; Network Centrality Measures; Strong and Weak ties; Homophily, Walks: Random walk-based proximity measures, Other graph-based proximity measures. Clustering with random-walk based measures

**UNIT 2 COMMUNITY DETECTION** **10 Hrs.**

Community Detection Algorithms: The Kernighan-Lin algorithm, Agglomerative/Divisive algorithms, Spectral Algorithms, Multi-level Graph partitioning, Markov Clustering; Community Discovery in Directed Networks , Community Discovery in Dynamic Networks, Community Discovery in Heterogeneous Networks, Evolution of Community.

**12 Hrs.**

**UNIT 3 LINK PREDICTION**

Feature based Link Prediction, Bayesian Probabilistic Models, Probabilistic Relational Models, Linear Algebraic Methods: Network Evolution based Probabilistic Model, Hierarchical Probabilistic Model, Relational Bayesian Network. Relational Markov Network.

**9 Hrs.**

**UNIT 4 EVENT DETECTION & INFLUENCCE ANALYSIS**

Event Detection: Classification of Text Streams, Event Detection and Tracking: Bag of Words, Temporal, location, ontology based algorithms. Evolution Analysis in Text Streams, Sentiment analysis, SNA in real world: FB/VK and Twitter analysis, Social Influence Analysis: Influence measures, Social Similarity - Measuring Influence, Influencing actions and interactions. Influence maximization.

**Max. 39 Hrs.**

**COURSE OUTCOMES**

On completion of the course, student will be able to
CO1- Understand Key concepts of social network, types of network and Walks.
CO2- Describe community detection algorithms in dynamic, directed and heterogeneous network.
CO3- Apply linear algebraic methods for link prediction.
CO4- Perform sentiment analysis on text Streams.
CO5- Compare the social influence measures based on actions and interactions.
CO6- Analyse real network to solve real world problem

**TEXT/REFERENCE BOOKS**

1. David Easley, Jon Kleinberg: Networks, Crowds and Markets: Reasoning about a highly connected world, Cambridge Univ Press 2010
2. S.Wasserman, K.Faust: Social Network Analysis: Methods and Applications, Cambridge Univ

**END SEMESTER EXAMINATION QUESTION PAPER PATTERN**

**Max. Marks: 100** **Exam Duration: 3 Hrs**

Part A: 10 Questions of 2 marks each-No choice 20 Marks
Part B: 2 Questions from each unit with internal choice, each carrying 20 marks 80 Marks

| 20DS517P | | | | | **Social Network Analysis LAB** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Teaching Scheme** | | | | | **Examination Scheme** | | | | | |
| **L** | **T** | **P** | **C** | **Hrs/Week** | **Theory** | | | **Practical** | | **Total Marks** |
| | | | | | **MS** | **ES** | **IA** | **LW** | **LE/Viva** | |
| **0** | **0** | **2** | **1** | **2** | **--** | **--** | **--** | **50** | **50** | **100** |

**COURSE OBJECTIVES**

- To apply theoretical concepts behind social network analysis algorithms
- To develop skills of using social network analysis software on real world data and be capable of analysing real work networks

**LIST OF EXPERIMENT**

Faculty must prepare the content as per the syllabus given. **Preferred Programming Language: R**

1 Basics of R programming, igraph package, Our first network
2 Basic Cohesion, metrics of density, reciprocity, reach, path distance, and transitivity. In addition, triadic analysis and a measure of ego-network heterogeneity, Data Formats for Networks
3 Plotting Basics
4 Measuring Networks Part 1: Centrality and Global Measures
5 Measuring Networks Part 2: Community structure and Assortment
6 Testing Your Network: Permutations and Randomizations
7 Peer Influence and QAP Regression
Intro to Network Regression (MRQAP)
8 Diffusion in Networks
9 Random Graphs
10 Simulating Network
11 Small world and scale free network

**COURSE OUTCOMES**

On completion of the course, student will be able to:
CO1- identify Key concepts of social network, types of network and Walks.
CO2- Implement community detection algorithms in dynamic, directed and heterogeneous network.
CO3- Apply linear algebraic methods for link prediction.
CO4- Perform sentiment analysis on text Streams.
CO5- Evaluate the social influence measures based on actions and interactions.
CO6- Design real network to solve any real-world problem

**TEXT/REFERENCE BOOKS**

1. David Easley, Jon Kleinberg: Networks, Crowds and Markets: Reasoning about a highly connected world, Cambridge Univ Press 2010
2. S.Wasserman, K.Faust: Social Network Analysis: Methods and Applications, Cambridge Univ

**END SEMESTER EXAMINATION QUESTION PAPER PATTERN**

**Max. Marks: 100** **Exam Duration: 2 Hrs**

Part A: Evaluation Based on the class performance and Laboratory book 50 Marks
Part B: Viva Examination based conducted experiments 50 Marks

| 20DS518T | | | | | Natural Language & Text Mining | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Teaching Scheme | | | | | Examination Scheme | | | | | |
| L | T | P | C | Hrs/Week | Theory | | | Practical | | Total Marks |
| | | | | | MS | ES | IA | LW | LE/Viva | |
| 3 | 0 | 0 | 3 | 3 | 25 | 50 | 25 | | | 100 |

**COURSE OBJECTIVES**

- Gain experience with both the theoretical and practical aspects of text mining.
- Learn how to build and evaluate computer programs that generate new knowledge from natural language text.

**UNIT 1 INTRODUCTION TO NATURAL LANGUAGE PROCESSING** **10 Hrs.**

Basic techniques in natural language processing, including tokenization, part-of-speech tagging, chunking, syntax parsing and named entity recognition

**10 Hrs.**

**UNIT 2 DOCUMENT REPRESENTATION & TEXT CATEGORIZATION**

Unstructured text documents with appropriate format and structure to support later automated text mining algorithms, Basic supervised text categorization algorithms, including Naive Bayes, k Nearest Neighbour (kNN) and Logistic Regression, SVM, Decision Tree

**UNIT 3 TEXT CLUSTERING & TOPIC MODELLING** **10 Hrs.**

Identifying the clustering structure of a corpus of text documents, Assigning documents to the identified cluster, connectivity-based clustering (a.k.a., hierarchical clustering) and centroid-based clustering (e.g., k-means clustering), General idea of topic modelling, Probabilistic Latent Semantic Indexing (pLSI) and Latent Dirichlet Allocation (LDA), and their variants for different application scenarios, including classification, imagine annotation, collaborative filtering, and hierarchical topical structure modelling

**9 Hrs.**

**UNIT 4 DOCUMENT SUMMARIZATION & TEXT VISULALIZATION**

Extraction based summarization methods, visual representations of abstract data to reinforce human cognition. Tools: Introduce some mathematical and programming tools for text visualization

**Max. 39 Hrs.**

**COURSE OUTCOMES**

On completion of the course, student will be able to
CO1- Use basic methods for information extraction and retrieval of textual data.
CO2- Understand quantitative analysis of text, with special focus on applying machine learning methods to text documents
CO3- Apply text processing techniques to prepare documents for statistical modelling
CO4- Apply relevant machine learning models for analysing textual data and correctly interpreting the results.
CO5- Evaluate the performance of machine learning models for textual data.
CO6- Demonstrate and present solution for a given problem using text analytics approach.

**TEXT/REFERENCE BOOKS**

1. Charu C. Aggarwal and Cheng Xiang Zhai, Mining Text Data. Springer, 2012.
2. Dan Jurafsky and James H Martin, Speech & Language Processing. Pearson Education India, 2000.
3. Christopher D. Manning, Prabhakar Raghavan, and Hinrichs Schuetze, Introduction to Information Retrieval. Cambridge University Press, 2007.

**END SEMESTER EXAMINATION QUESTION PAPER PATTERN**

**Max. Marks: 100** **Exam Duration: 3 Hrs**

Part A: 10 Questions of 2 marks each-No choice 20 Marks
Part B: 2 Questions from each unit with internal choice, each carrying 20 marks 80 Marks

| 20DS518P | | | | | **Natural Language & Text Mining LAB** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Teaching Scheme** | | | | | **Examination Scheme** | | | | | |
| **L** | **T** | **P** | **C** | **Hrs/Week** | **Theory** | | | **Practical** | | **Total Marks** |
| | | | | | **MS** | **ES** | **IA** | **LW** | **LE/Viva** | |
| **0** | **0** | **2** | **1** | **2** | **-** | **-** | **-** | **50** | **50** | **100** |

**COURSE OBJECTIVES**

- To understand the structure and basic operations of Natural Language Processing
- To understand the concepts of linguistic rules and machine learning approaches for classification
- To understand the syntax of Natural languages for grouping local words for parsing
- To study the various applications of NLP- machine translation, sentiment analysis, etc.

**LIST OF EXPERIMENT**

Practical list should be prepared based on the content of the subject and following guidelines should be useful.

**Experiment Sessions using Programming would be based on following topics:**

Basic stages of NLP such as tokenization, POS tagging, parsing, etc., Applications of NLP such as Sentiment Analysis, text summarizer, etc.

Following list gives some programming examples. Faculty can prepare their own list in same manner keeping above guidelines and syllabus in mind.

1. Implementation of simple tokenizer using NLTK, TextBlob, Regular Expression
2. Implement Porter Stemmer
3. Implement Lemmatization
4. Implement POS Tagger
5. Implement Parser
6. Implement Sentiment Analyser system
7. Implement Text Summarizer System

**Design based Problems (DP)/Open Ended Problem:**

1. Design Machine translation system for low resourced language
2. Design healthcare system using NLP

**COURSE OUTCOMES**

On completion of the course, student will be able to

CO1- Analyse the natural language text and speech

CO2- Process the Natural Language based on structure.

CO3- Apply the Bayes theorem to design language model for different language

CO4- Apply information retrieval techniques to build search engine, question answering system

CO5- Develop POS tagger, parsers and shallow parser for different languages

CO6- Design machine translation, text summarization, sentiment analysis

**END SEMESTER EXAMINATION QUESTION PAPER PATTERN**

**Max. Marks: 100** **Exam Duration: 2 Hrs**

Part A: Continuous Evaluation based on lab records and course project. 50 Marks

Part B: 2 Experiment conducted and Viva at final exam. 50 Marks

| 20DS510P | | | | | Capstone Course | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Teaching Scheme** | | | | | **Examination Scheme** | | | | | |
| **L** | **T** | **P** | **C** | **Hrs/Week** | **Theory** | | | **Practical** | | **Total Marks** |
| | | | | | **MS** | **ES** | **IA** | **LW** | **LE/Viva** | |
| **0** | **0** | **6** | **3** | **6** | **--** | **--** | **--** | **50** | **50** | **100** |

**COURSE OBJECTIVES**

- Expose students to a holistic review of data science as a discipline, reviewing the broader themes that link the various subfields together
- Allow students to reflect on their knowledge in the course work.
- Ability to solve real world applications of the data science area

**FEATURES**

Capstone experience typically involves:

- Integrating and extending knowledge, skills, perspectives gained through coursework, thus demonstrating program's outcomes concretely
- Reflecting on the social context, the body of literature, or the conceptual framework to which the student's capstone work poses a contribution.
- Bridging coursework with students' careers after graduation
- Preparing students for life-long learning
- The outcome is a tangible product to be presented to the public (written work, oral presentation, multimedia productions in various forms such as websites, CDs, DVDs)
- Topics are selected by students and approved by faculty from the given domains:

**Industry Domain Track-Capstone Course (Any-One)**

| Sl. No. | Domain | Sl. No. | Domain |
|---|---|---|---|
| 1. | Information Security | 4. | Agriculture |
| 2. | Real time streaming analysis | 5. | Healthcare |
| 3. | Banking & Financial | 6. | E-Commerce |

**COURSE OUTCOMES**

On completion of the course, student will be able to

CO1 - Integrate the knowledge learned in the general education and major/minor coursework

CO2 – Apply the gained knowledge to solve real world problems.

CO3 – Assess the existing solutions in given problem area.

CO4 – Adapt to the newly evolved technology and tools in data science.

CO5 – Prepare for life-long learning, future personal, academic and/or professional pursuits, and their roles as members of various communities.

CO5 - Communicate how the Capstone project contributes to a sense of closure, accomplishment, purpose, and agency.

**END SEMESTER EXAMINATION QUESTION PAPER PATTERN**

**Max. Marks: 100** **Exam Duration: 2 Hrs**

Part A: Evaluation Based on the class performance and Laboratory book 50 Marks

Part B: Viva Examination based conducted experiments 50 Marks

# III$^{rd}$ Semester

**PANDIT DEENDAYAL ENERGY UNIVERSITY GANDHINAGAR**

**SCHOOL OF TECHNOLOGY**

**COURSE STRUCTURE FOR M. TECH - DATA SCIENCE**

| Semester III | | | M. Tech. - Data Science | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sr. No. | Course/Lab Code | Course/Lab Name | Teaching Scheme | | | | | Examination Scheme | | | | | |
| | | | L | T | P | C | Hrs./Week | Theory | | | Practical | | Total |
| | | | | | | | | CE | MS | ES | CE | ES | Marks |
| 1 | 20DS611 | Seminar | | | | 5 | | 40 | 60 | -- | | | 100 |
| 2 | 20DS612 | Project | | | | 14 | | 40 | 60 | -- | | | 100 |
| | | Industrial Training | | | | | | | | | | | NP/PP |
| | | TOTAL | | | | 19 | 0 | | | | | | 200 |

CE- Continuous Evaluation, MS-Mid Semester; ES – End Semester Exam

# IV$^{th}$ Semester

**PANDIT DEENDAYAL ENERGY UNIVERSITY GANDHINAGAR**

**SCHOOL OF TECHNOLOGY**

| COURSE STRUCTURE FOR M.TECH - DATA SCIENCE | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Semester IV** | | | **M. Tech. - Data Science** | | | | | | | | | | | |
| **Sr. No.** | **Course/Lab Code** | **Course/Lab Name** | **Teaching Scheme** | | | | | **Examination Scheme** | | | | | | |
| | | | **L** | **T** | **P** | **C** | **Hrs./Week** | **Theory** | | | **Practical** | | **Total** |
| | | | | | | | | **CE** | **MS** | **ES** | **CE** | **ES** | **Marks** |
| 1 | 20DS621 | Seminar | | | | 5 | | | | | 60 | 40 | 100 |
| 2 | 20DS622 | Project and Dissertation | | | | 24 | | | | | 60 | 40 | 100 |
| | | TOTAL | | | | 29 | | | | | 120 | 80 | 200 |

CE- Continuous Evaluation, MS-Mid Semester; ES – End Semester Exam

For Seminar, Project and Dissertation work there would be Three reviews of 30, 30 and 40 marks.