Deccan Education Society's
FERGUSSON COLLEGE (AUTONOMOUS), PUNE

Syllabus

For

S. Y. M. Sc. (Data Science)

[Pattern 2019]

(S.Y. M.Sc. Semester-III and Semester-IV)

From Academic Year

2020-21

Deccan Education Society's
Fergusson College (Autonomous), Pune

S.Y.M.Sc. Data Science (Pattern 2019)

From academic year 2020-21

Particulars	Name of Paper	Paper Code	Title of Paper	No. of Credits
S.Y. M.Sc. Semester III	Paper - 1	CSD5301	Optimization Techniques	4
	Paper - 2	CSD5302	Emerging Trends in Data Science	4
	Paper - 3	CSD5303	Deep Learning	4
	Paper - 4	CSD5304	Data Science Case Studies OR	4
		CSD5305	Artificial Intelligence OR	
		CSD5306	MOOC-II	
	Paper - 5	CSD5307	Data Science Practical – V(Deep Learning)	4
	Paper - 6	CSD5308	Data Science Practical – VI (Project)	4
S.Y. M.Sc. Semester IV	Paper - 1	CSD5401	Industrial Training	8

SY. M.Sc. Semester III		
Title of the Course and Course Code	**Optimization Techniques (CSD5301)**	**Number of Credits : 04**
Course Outcome (COs) **On completion of the course, the students will be able to:**		
CO1	Identify the role of linear programming problem solving skills in real life business models.	
CO2	Distinguish between transportation problems and assignment problems and identify the importance of their applications.	
CO3	Apply various tools to implement different optimization methods.	
CO4	Analyze and appreciate a variety of performance measures for various optimization problems.	
CO5	Validate mathematical minima/maxima problems into optimization framework.	
CO6	Create efficient computational procedures to solve optimization problems.	

Unit. No.	Title of Unit and Contents	No. Of. Lectures
I	**Classical optimization techniques** 1.1 Maxima,Minima, critical points of singlevariable functions and multivariable functions 1.2 Single variable optimization with and without constraints, multi –variable optimization with and without constraints, method of Penalty methods, Lagrange multipliers 1.3 Kuhn-Tucker conditions	**10**
II	**Linear programming** 2.1 Application of Simplex method 2.2 Two-phase method 2.3 Big-M method, Duality 2.4 Integer linear Programming 2.5 Sensitivity analysis. Assignment problem: Hungarian's Algorithm, Degeneracy, applications, Unbalanced problem. 2.6 Traveling salesman problem	**15**
III	**CPM/PERT** 3.1 Simulation of CPM/PERT network 3.2 Analysis of an activity network 3.3 Simulation of inventory system and manufacturing System	**8**
IV	**Hyperparameter optimization** 4.1 Gradient of a function. 4.2 Steepest descent method 4.3 Nelder Mead's Simplex search method	**15**

	4.4 Newton's method.	

References:

1. Frederick Hillier and Mark Hillier, Introduction to Management Science, McGraw-Hill, 6th Edition, 2018
2. Eric Walter, Numerical Methods and Optimization: A Consumer Guide, Springer Cham,2014
3. Taha, H.A., Operations Research: An Introduction, Prentice Hall of India, 9th Edition, 2010
4. L.S.Srinath, PERT and CPM Principles and Applications, Affiliated East-West Press (Pvt.) Ltd, 3rd edition, 2001
5. S.S. Rao, Optimization Theory and Applications, Wiley Eastern Ltd, 2nd Edition, 1984

Web References

1. Mathematical Foundation of Data Analysis. J .Phillips –
   Download link: http://www.cs.utah.edu/~jeffp/M4D/M4D-v0.6.pdf

Title of the Course and Course Code	**Emerging Trends in Data Science (CSD5302)**	**Number of Credits : 04**
Course Outcome (COs) **On completion of the course, the students will be able to:**		
CO1	Outline different terms and concepts in Data science.	
CO2	Explain the importance of different steps in data processing to get the desired result.	
CO3	Implement different models and concepts in Data Science.	
CO4	Analyze different visualizations to display the result.	
CO5	Measure and test the performance of different models in different domains.	
CO6	Develop and deploy models to solve different problems.	

Unit. No.	Title of Unit and Contents	No. of. Lectures
I	**Introduction** 1.1 Data science Concept 1.2 Different Methodology 1.3 Cloud and Data Science 1.4 Data Preparation, Data Transformation 1.5 Data visual representation 1.6 Machine learning Concept and algorithm	**15**

II	**Technology Used in Data Science** 2.1 Technology Implementation method with data 2.2 Exploring and Preparing auto data 2.3 Validating automotive data 2.4 Visualize preliminary data wrangling results 2.5 Run summary statistics on the results 2.6 Exploring visualization tool for data 2.7 Implementation of ML concept	**18**
III	**Various Domain based Case study implementation with technology**	**15**

References:

1. AI Sweigart, Automate the Boaring Stuff with Python, November 2019
2. Jeff Leek, The Elements of Data Analytic Style, Leanpub publications, published on 2015-03-02
3. Roger D. Peng and Elizabeth Matsui, The Art of Data Science, Leanpub publications. 2015 - 2016
4. Advice and Insights from 25 Amazing Data Scientists,The Data Science Handbook, Leanpub publications. 2015 – 2016
5. Shai Shalev-Shwartz and Shai Ben-David, Understanding Machine Learning: From Theory to Algorithms c 2014.

Web References

1. https://www.packtpub.com/in/data/complete-data-wrangling-and-data-visualization-in-r-video
2. https://www.edureka.co/blog/importance-data-science-cloud-computing/#:~:text=Data%20science%20and%20cloud%20computing,a%20need%20for%20Data%20Scientists.

Title of the Course and Course Code	**Deep Learning (CSD5303)**	**Number of Credits : 04**
Course Outcome (COs) **On completion of the course, the students will be able to:**		
CO1	Identify different techniques in deep learning and describe their importance and application ranges.	
CO2	Discuss the deep learning techniques.	
CO3	Implement the algorithms in deep learning and apply the knowledge in various domains.	
CO4	Analyze the different types of data and map suitable techniques that suit the data.	
CO5	Assess and criticize the model developed on the basis of different performance measures.	
CO6	Build an optimized and efficient model to solve the problem statement.	

Unit	Details	Lectures
I	**Foundations of Neural Network and Deep Learning** 1.1 Introduction to deep learning 1.2 Neural Network Basics 1.3 Artificial neural network and its layers 1.4 Necessary activation functions 1.5 Notion of Partial and total Derivatives and functions 1.6 Weights and weight sharing, loss function, partial ordering	**[5]**
II	**Deep Learning Networks** 2.1 Working with Tensorflow, Keras ,Pytorch 2.2 Backpropagation in Feed-forward Networks 2.3 Feed Forward for Classification and Regression	**[7]**
III	**Improving the Deep Neural Networks** 3.1 Regularization: L1, L2, Dropout 3.2 Early Stopping 3.3 Optimization for Training Deep Models 3.4 Hyperparameter Tuning	**[8]**
IV	**Convolutional Neural Networks** 4.1 Introduction to Convolutional Networks 4.2 Understanding Convolution and Pooling 4.3 Different Classic CNN Architectures 4.4 CNN with MNIST Dataset and also with other datasets 4.5 Application of CNN in Computer Vision: Image	**[8]**

	Classification, Object detection	
V	**Recurrent Neural Networks** 5.1 The Sequential Problem 5.2 The RNN Model 5.3 The LSTM Model 5.4 Application of RNN in NLP and Time Series Forecasting	**[8]**
VI	**Restricted Boltzmann Machines (RBM)** 6.1 Introduction to RBMs 6.2 Training RBMs 6.3 RBM MNIST 6.4 Collaborative Filtering With RBM 6.5 Application of RBM in Recommender Systems	**[8]**
VII	**Autoencoders** 7.1 Introduction to Autoencoders 7.2 Application of Autoencoders in Recommender Systems and Image Processing	**[4]**

References:

1. Nikhil Baruma, Fundamentals of Deep Learning, O'Reilly publication, 2019.
2. Seth Weidman, Deep Leaning from Scratch-Building with Python from First Principal, O'Reilly publication, 2019.
3. Bharat Ramsundar, Peter Eastman, Patrik Walters, Vijay Pnade, Deep Learning for the Life Sciences, O'Reilly publication, 2019.
4. SudharsanRavichandiran, Hands-on Deep Learning Algorithms with Python, Packt Publication 2019.
5. Christopher Bishop, Pattern Recognition and Machine Learning, Springer. 2006. [CB-2006].
6. Duda, R.O., Hart, P.E., and Stork, D.G. Pattern Classi□cation. Wiley-Interscience. 2nd Edition. 2001.
7. Hastie, T., Tibshirani, R. and Friedman, J. The Elements of Statistical Learning. Springer. 2001.
8. Bishop, C. M. Neural Networks for Pattern Recognition. Oxford University Press. 1995.
9. Ian Goodfelllow, YoshuaBenjio, Aaron Courville, Deep Learning, The MIT Press
10. Aggarwal, Charu C., "Neural Networks and Deep Learning", Springer International Publishing, 1st Edition.

Web references

1. www.deeplearning.ai
2. www.tensorflow.org

Title of the Course and Course Code	**Data Science Case Studies (CSD5304)**	**Number of Credits : 04**
Course Outcome (COs) **On completion of the course, the students will be able to:**		
CO1	Define steps for solving case studies of different domains.	
CO2	Illustrate analytical tool's features for problem solving, filtering data and visualization.	
CO3	Apply the data manipulation and transformation techniques to prepare data ready for processing.	
CO4	Analyze data with the help of visualization and predict results.	
CO5	Evaluate model performance and communicate results.	
CO6	Build a comprehensive end to end project.	

Unit. No.	Title of Unit and Contents	No. of. Lectures
I	**Brief Introduction to Data Science** 1.1 What is Data Science 1.2 Why Now? - The importance of Data Science in today's business environment. 1.3 Difference between Data Science, Business Intelligence and Data Analysis 1.4 Real World Applications of Data Science 1.5 Popular Tools and Languages to Apply Data Science 1.6 A Typical Data Science Team 1.7 Team Structures	**4**
II	**Brushing Up on your MS Excel Skills** 2.1 Basic Functions 2.2 Advanced Functions 2.3 Sorting, Filtering and Pivot tables 2.4 Charts and other visualization tools 2.5 Handling Date formats for time series analysis 2.6 Short Cut Keys 2.7 Making simple Dashboards	**6**

III	**Making Data Work for You!** 3.1 What can Data do? 3.2 Types of Data 3.3 Data Exploration 3.4 Data Sources and Risks 3.5 Data Cleaning 3.6 Manipulating Time Series Data 3.7 Data Storage and Retrieval	**4**
IV	**Analysis, Prediction and Visualization** 4.1 Generating Charts and plots to better understand the 4.2 Output 4.3 Interpretation 4.4 Prediction 4.5 Conclusion	**4**
V	**Communicate Results** 5.1 Creating Dashboards in Tableau 5.2 Creating a Story in Tableau 5.3 Creating Dashboards in MS Excel 5.4 Generate High End Presentations in MS PowerPoint	**4**
VI	**The Data Science Workflow** 6.1 Understand the steps in the life cycle of a Data Science project 6.2 Theoretical Vs Applied Data Science	**2**
VII	**Case Study** 7.1 Problem Statement – Domain and scope of the study 7.2 Data Source 7.3 Data Preparation 7.4 Choice of Model 7.5 Model Building 7.6 Type of tools to use 7.7 Analysis 7.8 Presentation	**24**

References:

1. Microsoft Excel Data Analysis and Business Modeling, By Wayne L. Winston, Published by Microsoft Press 2019
1. Ryan Sleeper, Practical Tableau, O'Reilly publication, 2019.
2. Python Data Science Hand Book - https://tanthiamhuat.files.wordpress.com/2018/04/pythondatasciencehandbook.pdf

Title of the Course and Course Code	**Artificial Intelligence (CSD5305)**	**Number of Credits : 04**
Course Outcome (COs) **On completion of the course, the students will be able to:**		
CO1	Describe an overview of artificial intelligence (AI) principles and approaches.	
CO2	Discuss the various building blocks of AI.	
CO3	Apply AI techniques to different real world problems and games.	
CO4	Analyze, categorize the given search problem and write an algorithm for it.	
CO5	Compare various methods for knowledge representation and select the optimized one.	
CO6	Build the capability to represent various real life problem domains using logic based techniques to perform inference or planning.	

Unit. No.	Title of Unit and Contents	No. of. Lectures
I	**Introduction to Artificial Intelligence** 1.1 What is AI? Early work in AI 1.2 AI and related fields 1.3 AI problems and Techniques	**4**
II	**Problems, Problem Spaces and Search** 2.1 Defining AI problems as a State Space Search: examples 2.2 Production Systems 2.3 Search and Control Strategies 2.4 Problem Characteristics 2.5 Issues in Design of Search Programs 2.6 Additional Problems	**8**
III	**Heuristic Search Techniques** 3.1 Generate-and-test 3.2 Hill Climbing 3.3 Best First Search 3.4 Problem Reduction 3.5 Constraint Satisfaction 3.6 Mean-Ends Analysis	**8**
IV	**Knowledge Representation** 4.1 Representations and Mappings 4.2 Approaches to Knowledge Representation 4.3 Knowledge representation method 4.4 Propositional Logic 4.5 Predicate logic	**8**

	4.6 Representing Simple facts in Logic 4.7 Representing Instances and Is-a relationships 4.8 Computable Functions and Predicates 4.9 Resolution 4.10 Forward and backward chaining	
V	**Slot – and – Filler Structures** 5.1 Weak Structures 5.2 Semantic Networks 5.3 Frames 5.4 Strong Structures 5.5 Conceptual Dependencies	**6**
VI	**Game Playing** 6.1 Minimax Search Procedures 6.2 Adding alpha-beta cutoffs 6.3 Uncertainty Reasoning: Basic Probability Axioms, Baye's Rule, Baysian Classification, Certainty Factor Theory, Dempster Shafar Theory	**6**
VII	**Learning** 7.1 What is learning? 7.2 Rote Learning 7.3 Learning by taking advice 7.4 Learning in problem solving 7.5 Learning from examples 7.6 Explanation based learning	**6**

References:

1. Elaine Rich and Kevin Knight, Artificial Intelligence, Tata McGraw Hill, 3rd Edition, 2008
2. Dan Patterson, Introduction to Artificial Intelligence and Expert System, Prentice Hall of India Pvt. Ltd., 2nd Edition, 1990

Title of the Course and Course Code	**Data Science Practical – V(Deep Learning) (CSD5307)**	**Number of Credits : 04**
Course Outcome (COs) **On completion of the course, the students will be able to:**		
CO1	Describe concepts of deep learning.	
CO2	Discuss and compare the functionalities of various frameworks e.g. Pytorch and Tensorflow.	
CO3	Apply neural networks (e.g. CNN, RNN etc.) in problem solving.	
CO4	Analyze different unstructured data sets and solve them using classification techniques.	
CO5	Evaluate model performance based on accuracy, tuning and hyper parameters.	
CO6	Construct Deep learning models to solve the real life problems.	

List of practical (Compulsory 10 + 2 Activity)

1	Deep Learning using Keras
2	Deep Learning using Pytorch
3	Deep Learning using Tensorflow
4	Application of CNN in Computer Vision
5	CNN in Image Classification
6	CNN in Object Detection
7	Application of RNN in NLP
8	RNN in Time Series Forecasting
9	Application of RBM in Recommender Systems
10	Case Study

References:

1. SudharsanRavichandiran, Hands-on Deep Learning Algorithms with Python, Packt Publication, 2019
2. Ian Goodfelllow, YoshuaBenjio, Aaron Courville, Deep Learning, The MIT Press

3. **Web Reference**

1. www.deeplearning.ai
2. www.tensorflow.org

Title of the Course and Course Code	**Data Science Practical - VI (Project) (CSD5308)**	**Number of Credits : 04**
Course Outcome (COs) **On completion of the course, the students will be able to:**		
CO1	Identify and define problem statements.	
CO2	Exemplify the software development life cycle of the project.	
CO3	Solve problem statements by using different tools and techniques.	
CO4	Analyze the gap between specifications defined, developed functionality and code quality.	
CO5	Evaluate accuracy and efficiency of the model.	
CO6	Build a comprehensive end to end project.	

Objective:

The Objective of project is to make the students understand life cycle of data science project. Students should be able to define a problem statement, collect data and process it and explore the avenues in data science modeling (e.g. Predictive).

The Project can be platform, language and technology independent. Project will be evaluated by project guide. Assessment will be done weekly in the respective batch. Evaluation will be on the basis of weekly progress of project work, progress report, oral, results and documentation and demonstration.

You should fill your status of the project work on the progress report and get the Signature of project guide regularly. Progress report should sharply focus how much time you have spent on specific task. (The format of progress report is given as follow.) You should keep all signed progress report. Project will not be accepted if progress report is not submitted and all responsibility remains with student.

Project Progress Report

Roll No and Name of the Student	
Title of the project	
Project guide name	

Sr. No	From Date	To Date	Details of Project work	Project guide sign (with date)

SY. M.Sc. Semester IV		
Title of the Course and Course Code	**Industrial Training (CSD5401)**	**Number of Credits : 04**
Course Outcome (COs) **On completion of the course, the students will be able to:**		
CO1	Identify and describe the building blocks for solving problem statement.	
CO2	Explain the organizational structure of the company and summarize the tools and technologies used in the real world.	
CO3	Solve problem statements by applying acquired knowledge.	
CO4	Analyze different ways to approach a problem and design solution. Illustrate the actual work culture of the I.T. industry.	
CO5	Evaluate the various approaches based on accuracy and efficiency.	
CO6	Build a comprehensive end to end project.	

Duration: Minimum 3 months

Objectives:
1. To make the students get an insight into the actual work culture of the I.T. industry.
2. To make the students aware of the skills and technologies needed to work in the I.T. industry.
3. Understanding the Organizational Structure of the company.

The necessary details for Industrial Training course are as follows:
A student can complete Industrial Training Project (ITP) in any I.T. industry / academic institute / with a research project of a teacher / an expert funded by any funding agency for a minimum period of three months.

1. There will be a teacher coordinator for a group of 10 students. A teacher coordinator is responsible to:
- Maintain a weekly status / progress report of the student.
- Keep in touch with the reporting authorities from industry for each student.
- Help the students to solve their difficulties.
- Arrange the meeting and presentations as per requirement.
- Guide each student for preparing final project report.
- Keep complete documentation record for each student separately.
- Internal assessment of each student for 100 marks

The workload for this teacher coordinator is proposed as four hours per week.
The workload for a teacher coordinator who is guiding 3 students doing their ITP in Fergusson College (Autonomous) Pune (no mentor from industry) is proposed as four hours per week.

2. Guidelines for submitting the final project report

The student must include the project completion certificate issued by the respective industry/research institute/educational institute in the report. A student will submit two hard bound copies and one CD: Student Copy, Department copy, CoE copy of the work carried out during ITP (CD to be given by students).

3. Scheme of Assessment:

- **Continuous Internal Assessment**

Evaluation for internal 100 Marks will be done by the Internal Teacher Coordinator.

Description	Marks
Weekly Reports (Minimum 12)	40
Project Report writing	20
Internal Presentation Demo	30
Weekly Attendance	10

- **End Semester Assessment**

Evaluation for external 100 Marks will be done by a panel of three consisting of One Industrial Expert, One Academic Expert (External from other college) and One Internal Examiner. Each examiner is expected to assess each student for 100 marks independently and average of the three scores is to be considered as the final ESE score (out of 100)

Description	Marks
Knowledge and Execution of the System	20
Final Project Report	20
Presentation	30
Viva	30

- The internal examiner(s) will be responsible for submitting the total marks out of 200 to examination section.

- The final grade (to be printed on the mark list) is to be calculated on the basis of UGC 10 point scale.

Marks	Grade	Grade Point
180 – 200	**O: Outstanding**	**10**
160 – 179	**A+: Excellent**	**9**
141 – 159	**A: Very Good**	**8**
131 – 140	**B+: Good**	**7**
121 – 130	**B: Above Average**	**6**
111 – 120	**C+: Average**	**5**
101 – 110	**C: Below Average**	**4**
91 – 100	**D: Satisfactory**	**3**
80 – 90	**E: Pass**	**2**
0 – 79	**F: Fail**	**0**
	Absent	**0**

Note :- A student who has obtained Grade F will have to carry out this project once again for a complete semester (minimum three months).