Shivaji University, Kolhapur

Department of Computer Science

Post Graduate Diploma in Data Science (PGDDS)

(Under faculty of Science and Technology)

1. Introduction

The Post Graduate Diploma in Data Science (PGDDS) aims to prepare the student for a career as a data scientist, in the corporate sector, industries, for entrepreneurship, public policy or even academia. While focusing on the core statistical, quantitative and computing skills required in these careers, this Data Science course also arms students with domain knowledge in allied verticals so they can add value as data scientists.

The PGDDS intends to provide broad exposure to key concepts and tools viz. Python, Machine Learning and Deep Learning as well as hands-on laboratory and project work in Data Science. With successful completion, students can start their career as a Data Analyst, Data Scientist, Data Engineer, Product Analyst, Machine Learning Engineer, Decision Scientist and so on.

Learning Goals:

Learning Goals for the PGDDS are:

- Students will develop relevant programming abilities.
- Students will demonstrate proficiency with statistical analysis of data.
- Students will develop the ability to build and assess data-based models.
- Students will execute statistical analyses with professional statistical software.
- Students will demonstrate skill in data management.
- Students will apply data science concepts and methods to solve problems in real-world contexts and will communicate these solutions effectively

2. Duration of the Course:

The Post Graduate Diploma in Data Science (PGDDS) will be one year programme.

Pattern of examination will be Annual.

Intake capacity: **40**

Fees: **40,000/-**

3. Medium of Instruction:

The medium of Instruction will be English only.

4. Admission Procedure

1. Eligibility: Bachelor's Degree with minimum 50% or equivalent passing marks. Mathematics at 12^{th} Standard is compulsory.
2. Reservation of Seats As per rules of Government of Maharashtra.
3. Admission will be through entrance examination.

You are well-suited to pursue the Graduate Diploma in Data Science if:

- You have a strong quantitative undergraduate degree, such as in Statistics, Mathematics, Computing, Economics, the physical sciences, or engineering, to name a few
- You enjoy working with numbers to glean trends and patterns from them
- You want to pursue a rigorous Data Science programme, with applications in social, political, economic, legal, business and marketing fields.

5. Course Structure:

Lectures and Practical shall be conducted as per the scheme of lectures and practical indicated in the course structure. The program will be conducted in the morning session from 7.30 am to 11.30 am to suit the working professionals.

Teaching and Practical Scheme

1. Each contact session for teaching or practical shall be of 60 minutes each.
2. One Practical Batch shall be of 20 students.
3. Practical and project evaluation shall be conducted before the commencement of annual examination.

Project Work:

4. Project work may be done individually or in groups in case of bigger projects. However if project is done in groups, each student must be given a responsibility for a distinct module and care should be taken to see the progress of individual modules are independent of others.
5. Students should take guidance from assigned guide and prepare a Project Report on "Project Work".
6. The project report should be prepared in a format prescribed by the University, which also specifies the contents and methods of presentation. IEEE Computer Society templates are recommended in this regard.
7. The external viva shall be conducted by a panel of minimum two examiners out of which one will be external and other will be internal examiner.

OR

The student shall be allowed to formulate a proposal for startup and the same shall be rated equivalent to project. A detailed problem statement showing innovation along with marketability, business plan and cash flow shall be part of the evaluation criteria.

8. Assessment:

1) For each theory paper, 50% marks will be based on CIE and 50% marks for university Examination.
2) The project will be evaluated by the university appointed examiners both internal as well as external.

1. The final practical examination will be conducted by the university appointed examiners both internal as well as external at the end of year for each laboratory course and marks will be submitted to the university by the panel. The pattern of final Practical Examination will be as follows;

1	Programming and Execution of Program	60 Marks
2	Viva-voce	20 Marks
3	Journal(Internal)	20 Marks
4	Total	100 Marks

2. The final Examinations shall be conducted at the end of the year.
3. Nature of question paper:

Nature of question paper is as follows for University end year examination

a. **Theory Examination:**

1. There will be seven (7) questions of 10 Marks and out of which four to be attempted from question no 2 to7.
2. Question No.1 is compulsory and is of multiple choice questions. There will be 5 multiple choice question each carries 2 marks

b. **Practical Examination:**

1. Duration of Practical Examination: 3 Hrs
2. Nature of Question paper: There will be three questions out of which any two questions to be attempted and each question carries 30 Marks.

9. Standard of Passing:

Internal as well as external examination will be held at the end of the year. The candidate must score 40% marks in each head of internal as well as external Examination

Post Graduate Diploma in Data Science (PGDDS)

(Under Faculty of Science and Technology)

To be implemented from the academic year 2021-22

Sr. No	Course code	Course title	Theory contact hours per week	Practical hours per week	Credits	University exam	Internal continuous assessment	Total
1	DDS-1	Foundations Of Data Science	2	-	4	50	50	100
2	DDS-2	Python for Data Science	2	-	4	50	50	100
3	DDS-3	AI and Machine Learning	2	-	4	50	50	100
4	DDS-4	Deep Learning	2	-	4	50	50	100
5	DDS-5	Lab I(Based on DDS-2)	-	5	4	80	20	100
6	DDS-6	Lab II(Based on DDS-3 and DDS-4)	-	5	4	80	20	100
7	DDS-7	Project	-	2	4	80	20	100
		Total	8	12	28	440	260	700

Shivaji University, Kolhapur
Post Graduate Diploma in Data Science (PGDDS)
(Under Faculty of Science & Technology)

DDS-1: Foundations of Data Science

Course Outcome

1. Organize, manage and present data. Analyze statistical data graphically using frequency distributions and cumulative frequency distributions.
2. Analyze statistical data using measures of central tendency, dispersion and location. Data analysis using Excel.
3. Translate real-world problems into probability models. Derive the probability density function of transformation of random variables.
4. Calculate probabilities, and derive the marginal and conditional distributions of bivariate random variables and perform hypothesis testing.
5. Calculate simple linear regression model, extend model up to multiple linear regression. Perform Hot Deck and KNN imputation
6. Calculate a simple Calculate a simple linear regression mode linear regression mode.
7. Ability to integrate machine learning libraries and mathematical and statistical tools with modern technologies like hadoop , map reduce, Hive, Weka, Tableau.

Unit 1 **(15 hrs)**

Data science: Introduction, overview, Components, Model building process , Data types and its measures, Random Variables ,its applications with exercise, Tools ,Applications, Sources of data.

Unit 2 **(15 hrs)**

Basics of statistics: Measures of central tendency (Mean, Median, Mode),Measures of dispersion(Variance, standard deviation, range),Measures of Skewness and Kurtosis – Graphical representation and its application, Measures of correlation coefficient and it analysis, Various graphical representation of data for analysis, Concept of Probability and it's Applications, probability distributions, Central limit theorem for sampling variations, Confidence Interval –computation and analysis. Basic concepts related to hypothesis testing.

Unit 3 **(15 hrs)**

Regression Analysis: Regression model using ordinary least squares, Coefficient of determination as strength of model, Prediction interval and confidence interval, Prerequisites to regression, Model building using regression, Measures of accuracy, Model improvement techniques, listwise, pairwise deletion, Imputation techniques: Regression Imputation ,Hot Deck, KNN imputation.

Unit 4 **(15 hrs)**

Data analysis using Excel, Overview of tools SQL, Hadoop, Hive, Weka, Tableau, R, Python, Introduction of Hadoop Ecosystem(Pig ,Hive and Hbase), Big data analysis.

References

1. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, p. 18). New York: springer.
2. Friedman, J., Hastie, T., & Tibshirani, R. (2001). The elements of statistical learning (Vol. 1, No. 10). New York: Springer series in statistics.
3. Spiegel halter, D. (2020). Introducing The Art of Statistics: How to Learn from Data. *Numeracy*, *13*(1), 7.
4. Kuhlman, D. (2009). *A python book: Beginning python, advanced python, and python exercises* (pp. 1-227). Lutz: Dave Kuhlman.
5. Grus, J. (2019). *Data science from scratch: first principles with python*. O'Reilly Media.
6. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (Vol. 821). John Wiley & Sons.

Shivaji University, Kolhapur
Post Graduate Diploma in Data Science (PGDDS)
(Under Faculty of Science & Technology)

DDS-2: Python for Data Science

Course Outcome

1. Learn about Python fundamentals, Python data structures, and working with data in Python
2. Become familiar with key Python functions, objects, and classes.
3. After a brief overview of the Scientific Python ecosystem, we dive into techniques for numeric data processing, including efficiently manipulating and processing large data sets using NumPy arrays and data visualization with 2D plots using Matplotlib.
4. Next up is an introduction to Pandas to efficiently load, clean, normalize, aggregate, transform, and visualize data.
5. Accessing data from common data file types and databases using Pandas
6. Cleansing and normalizing data with Pandas

Unit-I **(15 hrs)**

Python Introduction: Installing Python, Simple program using Python, Expressions and Values, Variables and Computer Memory, error detection, multiple line statements, Designing and using functions, functions provided by Python, Tracing function calls in memory model, omitting return statement. Working with Text: Creating Strings of Characters, Using Special Characters in Strings, Creating a Multiline String , Printing Information, Getting Information from the Keyboard

Unit-II **(15 hrs)**

A Boolean Type, Choosing Statements to Execute, Nested If Statements, Remembering the Results of a Boolean Expression Evaluation, A Modular Approach to Program Organization, Importing Modules , Defining Modules, Testing Code Semi automatically Grouping Functions Using Methods: Modules, Classes, and Methods, Calling Methods the Object Oriented Way, Exploring String Methods, Underscores.

Unit-III **(15hrs)**

Storing Collections of Data Using Lists: Storing and Accessing Data in Lists, Modifying Lists, Operations on Lists, Slicing Lists, Aliasing, List Methods, Working with a List of Lists. Repeating Code Using Loops: Processing Items in a List, Processing Characters in Strings, Looping Over a Range of Numbers, Processing Lists Using Indices, Nesting Loops in Loops, Looping Until a Condition Is Reached, Repetition Based on User Input, Controlling Loops Using Break and Continue Reading and Writing Files: Kinds of files, Opening a File, Techniques for Reading Files, Files over the Internet, Writing Files, Writing Algorithms that use the File- Reading Techniques, Multiline Record

Unit-IV **(15hrs)**

Storing Data Using Other Collection Types: Storing Data Using Sets, Storing Data Using Tuples, Storing Data Using Dictionaries, Inverting a Dictionary Creating Graphical User interface: Building a Basic GUI, Models, Views, and Controllers, Customizing the Visual Style Widgets, Object-Oriented GUIs, Regular expressions Databases: Overview, Creating and Populating, Retrieving Data, Updating and Deleting, Using NULL for Missing Data, Using Joins to Combine Tables, Keys and Constraints.

References:

1. Practical Programming: An introduction to Computer Science Using Python, second edition, Paul Gries, Jennifer Campbell, Jason Montojo, The Pragmatic Bookshelf.
2. Python for Informatics: Exploring Information, Charles Severance
3. Learning Python, Fourth Edition, Mark Lutz, O'Reilly publication
4. Introduction to Python for Computational Science and Engineering (A beginner's guide), Hans Fangohr
5. John V Guttag. "Introduction to Computation and Programming Using Python", Prentice Hall of India
6. R. Nageswara Rao, "Core Python Programming", Dreamtech

Shivaji University, Kolhapur
Post Graduate Diploma in Data Science (PGDDS)
(Under Faculty of Science & Technology)

DDS-3 AI and Machine Learning

Course Outcome

1. Provide an introduction to the basic principles, techniques, and applications of Artificial Intelligence.
2. Ability to select and implement machine learning techniques and AI computing environment that are suitable for the applications under consideration
3. Ability to understand and apply scaling up machine learning techniques and associated computing techniques and technologies.
4. Ability to recognize and implement various ways of selecting suitable model parameters for different machine learning techniques
5. Use different machine learning techniques to design AI machine and enveloping applications for real world problems.

Unit-I **(15 hrs)**

Introduction to Artificial Intelligence and Machine learning, Essential concepts in Artificial Intelligence and Machine learning. Machine learning basics: Key terminology, Key tasks of machine learning, choosing the right algorithm, Steps in developing a machine learning application.

Unit-II **(15 hrs)**

Supervised Learning The k-Nearest Neighbours classification algorithm, Parsing and importing data from a text file, Creating scatter plots with Matplotlib, Normalizing numeric values. **Decision tree**, Tree construction, plotting trees in Python, Testing and storing the classifier,

Unit-III **(15 hrs)**

Naïve Bayesian decision theory, Conditional probability, classifying with conditional probabilities, Document classification with naïve Bayes, classifying text with python, Case study: classifying spam email with naïve Bayes.

Unsupervised learning: Clustering, Grouping unlabelled data using K-Means clustering, K-means algorithm.

Unit-IV **(15 hrs)**

Recommender System: Introduction, Understanding Recommendation Systems, Content Based Filtering, User Based Collaborative Filtering, Item Based Collaborative Filtering, Methods and tricks of the trade, Issues in Recommendation Systems, Recommender System in Python.

References:

1. Machine Learning and Artificial Intelligence, Ameet V. Joshi, Springer, Cham
2. Machine Learning in Action,Peter Harrington, April 2012 , Manning publications. https://livebook.manning.com/book/machine-learning-in-action/about-this-book/
3. Artificial Intelligence and Machine Learning Fundamentals by Zsolt Nagy
4. Data Mining Concepts and Techniques, Jiawei Han and Micheline Kamber, Else

Shivaji University, Kolhapur
Post Graduate Diploma in Data Science (PGDDS)
(Under Faculty of Science & Technology)

DDS-4 Deep Learning

Course Outcome

1. To introduce the fundamental techniques and principles of Neural Networks
2. To study the different models in ANN and their applications
3. To familiarize deep learning concepts with Convolution Neural Network case studies
4. Apply deep learning mechanisms to various learning problems.
5. Know the open issues in deep learning, and have a grasp of the current research directions

UNIT-I **(15 hrs)**

Neural Networks - Introduction, Basic Concepts of Neural Networks, Model of an Artificial Neuron, Activation Functions, Feed forward Network, Recurrent Network

UNIT-II **(15 hrs)**

Introduction to deep learning: Defination, Need, Relationship between Artificial intelligence, machine learning, and deep learning, Deep learning Process.
Deep Learning Network: Convolutional neural networks (CNN), Deep learning applications, Advantages and Limitations of deep learning.
Deep learning Libraries /Frameworks: Keras, Tensor Flow, PyTorch

UNIT-III **(15 hrs)**

Deep Learning with Keras / PyTorch: Setting up Project, Starting Jupyter, Importing Libraries, and Creating Deep Learning Model.

UNIT-IV **(15 hrs)**

Introduction to convents: Foundations of Convolutional Neural Networks, Training a convnet from scratch on a small dataset, using a pretrained convent, visualizing what convnets learn. Deep Convolutional Models: Case Studies

References:

1. Deep Learning with Python, FRANÇOIS CHOLLET
2. Dive into Deep Learning, Aston Zhang, Zachary C. Lipton, Mu Li, and Alexander J. Smola
3. Deep Learning, Ian Goodfellow, Yoshua Bengio, Aaron Courville
4. Neural Networks and Deep Learning, Michael Nielsen's

Shivaji University, Kolhapur
Post Graduate Diploma in Data Science (PGDDS)
(Under Faculty of Science & Technology)

DDS-5 Lab I (Based on DDS-2)

Practical assignments based on Python for Data Science (DDS-2)

Shivaji University, Kolhapur
Post Graduate Diploma in Data Science (PGDDS)
(Under Faculty of Science & Technology)

DDS-6 Lab II (Based on DDS-3 and DDS-4)

Practical assignments based on AI and Machine Learning & Deep Learning (DDS-3 and DDS-4)

Shivaji University, Kolhapur
Post Graduate Diploma in Data Science (PGDDS)
(Under Faculty of Science & Technology)

DDS-7 Project

Project work should be carried out with application of data science over the period of 8 months.