

**MADHYA PRADESH BHOJ (OPEN) UNIVERSITY,  
Raja Bhoj Marg Kolar Road, BHOPAL (M.P.)**



**DIPLOMA IN  
DATA SCIENCE  
(from Calendar Year 2023-24)**

**MADHYA PRADESH BHOJ (OPEN) UNIVERSITY,  
Raja Bhoj Marg Kolar Road, BHOPAL (M.P.)**

---

*Ahda*

**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 its analysis, Various graphical representation of data for analysis, Concept of Probability and its 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.**

*Adarsh B.*

## 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. Spiegelhalter, 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.

Arde B.

## 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 ObjectOriented 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, ControllingLoops 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

*Aravind B.*

## 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 Userinterface:** Building a Basic GUI, Models, Views, and Controllers, Customizing the VisualStyle 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

*Arav* *J.*

## 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.

*Ardes B.*

## 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 Zolt Nagy
4. Data Mining Concepts and Techniques, Jiawei Han and Micheline Kamber, Else

Ahmed B.

## DDS-7/8 Project

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

*Arde B.*



**DIPLOMA IN  
DATA SCIENCE  
(from Calendar Year 2023-24)**

**Course of Study and Scheme of Examinations**

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	-	6	70	30	100
2	DDS-2	Python for Data Science	2	-	6	70	30	100
3	DDS-3	AI and Machine Learning	2	-	6	70	30	100
4	DDS-7	Project	-	2	6	70	30	100
		Total	6	2	24	280	120	400

*Ardes*  
04/09/2023