

Advanced Diploma in Data Science

Department of Mathematical Sciences and Computer Applications
Bundelkhand University, Jhansi

2019



Handwritten signatures and initials in blue ink, including a large signature at the top right, a signature in the middle right, and two sets of initials at the bottom.



बुन्देलखण्ड विश्वविद्यालय, झाँसी
BUNDELKHAND UNIVERSITY, JHANSI

Dated:.....

Department of Mathematical Sciences and Computer Applications

Ordinance, Course Structure & Syllabus

Course Name

Advanced Diploma in Data Science (PGDS)

Course Code: PGDS

Admission procedure: Through counseling

Number of seats: 30

Programme start date: 15-07-2019

Preamble: Data Science refers to extraction of knowledge from large volumes of data that are structured or unstructured, which is continuation of data mining and predictive analytics. It involves different categories of analytical approaches for modeling various types of business scenario and arriving at solution and strategies for optimal decision-making in marketing, finance, operations, organizational behavior and other managerial aspects. The aspirant professionals willing to work in data science / data analytics field is expected four types skills, that includes sound mathematical background, statistical thinking, expertise in programming and software operational skills and effective presentation – communication skills. The curriculum of PGDS is designed to incorporate all these optimum skills. As a outcome of this course, it is aimed to accelerate career of budding data scientist or data analytics in data science by mastering concepts of *Data Management, Statistics, Machine Learning and Big Data* from the most influential analytics leaders and academicians of India

Pre-requisite: Minimum basic knowledge of mathematics, computer programs, and statistics up to 12th standard.

Eligibility for admission: Any student with M.Sc. in Statistics, Mathematics, Computer Science, Information Technology, M.C.A., BE/B. Tech. with minimum 60% of marks are eligible for the admission to the course.

Duration of the course: It is a full time *Two Semester* (One Year) course.

- (i) No supplementary examination for any of the semester shall be conducted.
- (ii) English shall be the medium of instructions and examinations.
- (iii) The minimum pass marks for each theory paper would be 35%, while the aggregate of each semester shall be 40%.

- (iv) The minimum pass marks in practical shall be 50% of marks. In case candidates are absent or failed in practical, shall get the opportunity only in the next subsequent exam of their semester.
- (v) The minimum attendance for each paper for appearing the semester examination shall be 75%.

Choice-Based Credit System (CBCS):

CBCS is a flexible system of learning that permits students to:

- Learn at their own pace,
- Choose electives from a wide range of elective courses offered by the University departments,
- Adopt an inter-disciplinary approach in learning, and Make best use of the expertise of available faculty.

(a) **Credits:** Credit is a kind of weightage given to the contact hours to teach the prescribed syllabus, which is in a modular form. Normally one credit is allocated to 15 contact hours.

(b) In each of the courses, credits will be assigned on the basis of the number of lectures / tutorials / laboratory work and other forms of learning required for completing the course contents in maximum 18 week schedule.

- The instructional days as worked out by BU Jhansi for one academic year are 180 working days i.e. 90 days per semester.
- Programmes have minimum five papers and one practical in each semester. It means student has to complete 24 credits in each semester.
- Mechanics of contact hours: As per BU Jhansi standard 42 hours per semester.
- Mechanics of Credit Calculation: As per BU Jhansi standard, **1Credit = 14 hours of lectures.**
- Contact hours will include all the modes of teaching and it includes forms like lectures / tutorials / laboratory work or other forms. In determining the number of hours of instruction required for a course involving laboratory, 2 hours of laboratory is generally considered equivalent to 1 hour of lecture.

(b) **Credit Point, (P):** Credit point is the value obtained by multiplying the grade point (G) by the credit (C): $P = G \times C$. Grade point is an integer indicating the numerical equivalent of the letter grade.

(c) **Semester Grade Point Average (SGPA):** Semester Grade Point Average (SGPA) is the value obtained by dividing the sum of credit points (P) earned by a student in various courses taken in a semester by the total number of credits earned by the student in that semester. SGPA shall be rounded off to two decimal places.

(d) **Cumulative Grade Point Average (CGPA):** Cumulative Grade Point Average' (CGPA) is the value obtained by dividing the sum of credit points in all the courses earned by a student for the entire programme, by the total number of credits. CGPA shall be rounded off to two decimal places. CGPA indicates the comprehensive academic performance of a student in a programme.

An overall letter grade (Cumulative Grade) for the entire programme shall be awarded to a student depending on his/her CGPA.

(e) **Grading System:**

The image shows four distinct handwritten signatures in blue ink, arranged horizontally. The first signature on the left is a cursive 'Sue'. The second is a more complex, stylized signature. The third is a signature that appears to be 'Sue' with a flourish. The fourth is a signature that looks like 'Sue' with a horizontal line extending to the right.

The grade points are the numerical equivalent of letter grade assigned to a student in the 07 points scale as given below.

% Mark Range	Grade	Grade Point
90 and above	A+	10
80-89	A	9
70-79	B+	8
60-69	B	7
50-59	C+	6
40-49	C	5
Below 40	F	0

(f) **Extra Credits:** Extra credits may be awarded to a student for achievements in co-curricular activities carried out outside the regular class hours, as decided by the University. These credits shall not be counted while considering the minimum credits for completing the programme. The University shall frame detailed guidelines for the award of co-curricular credits and grades.

(g) **Computation of (SGPA) and CGPA**

The UGC recommends the following procedure to compute the Semester Grade Point Average (SGPA) and Cumulative Grade Point Average (CGPA):

(h) The SGPA is the ratio of sum of the product of the number of credits with the grade points scored by a student in all the courses taken by a student and the sum of the number of credits of all the courses undergone by a student, i.e

$$SGPA(S_i) = \frac{\sum(C_i \times G_i)}{\sum C_i}$$

where C_i is the number of credits of the i th course.

G_i is the grade point scored by the student in the i th course.

(i) The CGPA is also calculated in the same manner taking into account all the courses undergone by a student over all the semesters of a programme, i.e.

$$CGPA = \frac{\sum(C_i \times S_i)}{\sum C_i}$$

where S_i is the SGPA of the i th semester.

C_i is the total number of credits in that semester.

(j) The SGPA and CGPA shall be rounded off to 2 decimal points and reported in the transcripts.

Illustration of Computation of SGPA and CGPA and Format for Transcripts

Illustration for SGPA

Course	Credit	Grade Letter	Grade Point	Credit Point
Course 1	4	A+	10	10x4=40
Course 2	4	B+	8	8x4=32
Course 3	3	C	5	5x3=15
Course 4	3	B	7	7x3=21
	14			108

Thus, $SGPA = 108/14 = 7.71$

Illustration for CGPA

Semester I	Semester II	Semester III	Semester IV	Semester V	Semester VI
Credit: 26 SGPA: 6.37	Credit: 20 SGPA: 7.55	Credit: 25 SGPA: 6.35	Credit: 22 SGPA: 6.75	Credit: 19 SGPA: 5.98	Credit: 24 SGPA: 7.00

$$\text{Thus, CGPA} = \frac{26 \times 6.37 + 20 \times 7.55 + 25 \times 6.35 + 22 \times 6.75 + 19 \times 5.98 + 24 \times 7.00}{136} = 6.63$$

11. Successful candidate shall be classified on the basis of the combined results of semester I, II, III, IV, V, VI examination as follows:

Candidate securing in aggregate:

75 % and above

60 % to less than 75 %

50 % to less than 60 %

First Division with Distinction

First Division

Second Division

The minimum pass marks for the whole year would be 50%

Evaluation: Each subject in the curriculum (Theory / practical/ semester) is an independent entity and should be evaluated separately. The attendance and marks obtained in each entity should be above a minimum required level. They should not be aggregated.

Reregistering for a subject: For professional courses it is imperative that the course contents are modified at regular intervals. Allowing a candidate to appear for examination long after attending a subject cannot be justified academically. It is therefore necessary for the student at the earliest or some other relevant subject as per the requirement and then appears in the examination.

BOS: It is imperative that the course contents are modified at the regular activities. Allowed a candidate to appear for examination long after attending a subject as per the requirement & then appear in the examination.

Last Date of Registration: 01-07-2019

Curriculum of the course: The two semester course consists of the following curriculum.

Structure of the Syllabus PGDS

PGDS Semester-I

Paper Code	Title of the Paper	Contact hours/ week	Distribution of Marks for Examination			Credits
			Internal	External	Total	
PGDS-101	Statistical Analysis with R	04	30	70	100	04
PGDS-102	Data Visualization - Analysis and Reporting	04	21	54	75	03
PGDS-103	Programming with Python	04	21	54	75	03
PGDS-104	Java with Scala	04	30	70	100	04

PGDS-105	Effective Communication / Elective	02	15	35	50	02
Total		18	117	283	400	16

PGDS Semester-II

Paper Code	Title of the Paper	Contact hours/ week	Distribution of Marks for Examination			Credits
			Internal	External	Total	
PGDS-201	Business Decisions and Analytics	04	30	70	100	04
PGDS-202	Practical Machine Learning	04	30	70	100	04
PGDS-203	Hands on Training on XLSTAT	04	21	54	75	03
PGDS-204	Project	08	-	-	150	04
PGDS-205	Aptitude	02	21	54	75	02
Total		22	102	248	500	16

PGDS - 101: Statistical Analysis with R

(100 Hours)

Probability & Statistics: Introduction to Statistics- Descriptive Statistics, Summary Statistics Basic probability theory, Statistical Concepts (uni-variate and bi-variate sampling, distributions, re-sampling, statistical Inference, prediction error), Probability Distribution (Continuous and discrete- Normal, Bernoulli, Binomial, Negative Binomial, Geometric and Poisson distribution), Bayes' Theorem, Central Limit theorem, Data Exploration & preparation, Concepts of Correlation, Regression, Covariance, Outliers etc.

R Programming: Introduction & Installation of R, R Basics, Finding Help, Code Editors for R, Command Packages, Manipulating and Processing Data in R, Reading and Getting Data into R, Exporting Data from R, Data Objects-Data Types & Data Structure. Viewing Named Objects, Structure of Data Items, Manipulating and Processing Data in R (Creating, Accessing, Sorting data frames, Extracting, Combining, Merging, reshaping data frames), Control Structures, Functions in R (numeric, character, statistical), working with objects, Viewing Objects within Objects, Constructing Data Objects, Building R Packages, Running and Manipulating Packages, Non parametric Tests- ANOVA, chi-Square, t-Test, U-Test, Introduction to Graphical Analysis, Using Plots(Box Plots, Scatter plot, Pie Charts, Bar charts, Line Chart), Plotting variables, Designing Special Plots, Simple Liner Regression, Multiple Regression.

Books for References

1. Peter Dalgaard: Introductory Statistics with R, Second Edition, Springer Publications.
2. Prabhanjan N. Tattar, Suresh Ramaiah, B. L. Manjunath: A course in Statistics with R, Wiley Estern Publication.
3. Crawley M. J. (2006): Statistics-An Introduction Using R, Jhon-Wiley London.
4. Purohit S. G., Gore S. D. and Deshmukh S.R. (2015): Statistics Using R, Second Edition, Narosa Publishing House, New Delhi.

5. Verzani J. (2005): Using R for Introductory Statistics, Chapman and Hall ? CRC Press, New York.
6. Andrea Cirillo(2017): R Data Mining, Packet Publishing.
7. Duis Torgo: Data Mining with R, Learning with Case Studies, CRC Press.
8. Hadley Wickham and Garrett Grolemund : R for Data Science.
9. Seema Acharya: Data Analytics Using R, McGraw Hill Publications
10. Bernard Marr: Data Strategy: How to Profit from a World of Big data, Analytics and Internet of Things, Kogen Page Publications.

PGDS - 102: Data Visualization - Analysis and Reporting

(75 Hours)

Information Visualization, Data analytics Life Cycle, Analytic Processes and Tools, Analysis vs. Reporting, Modern Data Analytic Tools, Visualization Techniques, Visual Encodings, Visualization algorithms, Data collection and binding, Cognitive issues, Interactive visualization. Visualizing big data – structured vs unstructured, Visual Analytics, Geomapping, Dashboard Design.

Books for References

1. Andy Kirk: Data Visualization - A handbook for data driven design.
2. Cole Nussbaumer Knaflic: Storytelling with data: A data visualization Guide for business professionals, Wiley Publications.
3. Sejal Vora: The power of data storytelling, Sage Publications.
4. Edward R. Tufle: The Visual Display of Quantification information, 2nd Edition.
5. Edward R. Tufle: Envisioning Information.
6. Alberto Cairo: the Truthful Arts – data, charts and maps for communications.

PGDS - 103: Programming with Python

(75 Hours)

Introduction to Python, Basic Syntax, Data Types, Variables, Operators, Input/output, Flow of Control (Modules, Branching), If, If- else, Nested if-else, Looping, For, While, Nested loops, Control Structure, Break, Continue, Pass, Strings and Tuples, Accessing Strings, Basic Operations, String slices, Working with Lists, Introduction, Accessing list, Operations, Function and Methods, Files, Modules, Dictionaries, Functions and Functional Programming, Declaring and calling Functions, Declare, assign and retrieve values from Lists, Introducing Tuples, Accessing tuples

Books for References

1. Eric Matters: Python crash course – A Hands on Projects based Introduction to Programming.
2. David Beazley and Brian K Jones: Python Cookbooks Recipes for Mastering Python 3.
3. Krishna Rungta : Learn Python in One day – Complete Python Guide with examples.
4. Sebastian Raschka and Vahid Mirjalili : Python Machine Learning, 2nd Edition, Packt Publications.

Handwritten signatures and initials in blue ink at the bottom of the page, including a signature that appears to be 'Suj' and another that appears to be 'Raj'.

5. Wes Mekinney: Python for Data Analytics.
6. Paul Barry: Head first Python, 2nd Edition
7. Lucaino Ramalho, Fluent Python

Advanced Python: Object Oriented, OOPs concept, Class and object, Attributes, Inheritance, Overloading, Overriding, Data hiding, Operations Exception, Exception Handling, Except clause, Try finally clause, User Defined Exceptions, Python Libraries, Introduction to Machine learning packages like NUMPY, SCIPY, PANDAS etc.

PGDS - 104: Java with Scala

(100 Hours)

Data Types, Operators and Language, Constructs, Inner Classes and Inheritance, Interface and Package, Exceptions, Threads, Introduction, Unified Types, Classes, Traits, Mixin Class Composition, Anonymous Function Syntax, Higher-order Functions, Nested Functions, Currying, Case Classes, Pattern Matching, Singleton Objects, XML Processing, Regular Expression Patterns, Extractor Objects, Sequence Comprehensions, Generic Classes

Books for References

1. Toby Western: Scala for Java developers – A practical primer, Apress Publications.
2. Thomas Alexndere : Scala for Java developers.
3. Diego Pacheco: Building applications with Scala.
4. Vikas Sharma: Learning Scala Programming.
5. Chiranjit Hazarika: Scala Programming in Scala, A Comprehensive step by step guide.
6. Aleksandar Prokopec: Learning Concurrent Programming in Scala.

PGDS - 201: Business Decisions and Analytics

(100 Hours)

Introduction to Business Analytics using some case studies, Making Right Business Decisions based on data, Exploratory Data Analysis - Visualization and Exploring Data, Descriptive Statistical Measures, Probability Distribution and Data, Sampling and Estimation, Statistical Interfaces, Predictive modeling and analysis, Regression Analysis, Forecasting Techniques, Simulation and Risk Analysis, Optimization, Linear, Non linear, Integer, Decision Analysis, Strategy and Analytics. Overview of Factor Analysis, Directional Data Analytics, Functional Data Analysis.

Books for References:

1. S. Christion Albright, Wayne. L. Winston: Business Analytics: Data Analytics & Dicision making, 6th edition.
2. Randy Bartett: A Practitioner's Guide to Business Analytics Using Data Analysis tools to Improve your organizations Decision making & Strategy.



3. Albright: Business Analytics: Data Analysis & Decision making.
4. Dursun Delen: Real World Data Mining Applied Business Analytics & Decision Making.
5. Priyanka Jain, Puneet Sharma: Behind Every Good Decision: How Anyone can Use Business Analytics
6. Katherine Marcoxi, Harold Lehmann: Big Data & Health Analysis.
7. Hardeep Chahd, Jeevan Jyoti, Jochen Wirtz: Understanding the role of Business Analytics: Some Applications.

PGDS - 202: Practical Machine Learning

(100 Hours)

Supervised and Unsupervised Learning , Uses of Machine learning , Clustering, K means, Hierarchical Clustering, Decision Trees, Oblique trees, Classification problems, Bayesian analysis and Naïve bayes classifier, Random forest, Gradient boosting Machines, Association rules learning, Apriori and FP-growth algorithms, Support vector Machines, Linear and Non liner classification, ARIMA, ML in real time, Neural Networks and its application, Neural Net & its applications

Books for References:

1. Ian. H. Witten, Eibe Frank, Mark A. Hall: Data Mining: Practical Machine Learning Tools & Technique.
2. Dipanjan Sarkar, Raghav Bali, Tushar Sharma: Practical ML with Python – A Problem Solvers Guide to building Real world Intelligent Systems.
3. Oliver Theobald: ML for absolutely beginners – A plain English Introduction.
4. Sinan Ozdemir: Principles of Data science.
5. Sebastain Raschka and Vahid Mirjalili : Python ML.
6. Peter Harrington: ML in Action.
7. Ethem Alpaydin: Introduction to ML.
8. Luis Pedro Coelho and Willi Richert: Building ML systems with Python, 2nd Edition.

PGDS - 203: Hands on Training on XLSTAT

(75 Hours)

Given the preferable huge data perform XLSTAT tools like data visualization, association rules, distribution fitting, classification rules, PCA, factor analysis, sensor analysis, text mining, contingency tables, Simulation, regression analysis, clustering, predictive modelling.

Books for References:

1. Barnabas Crist Bal: Xlstat, Cede Publishing.
2. Elaine McDonald – Newman: Excel Manual with XLSTAT Business Statistics.

PGDS - 204: Project

(150 Hours)

Project in maximum TWO students and supervised by a faculty. The topic of the project should be given such that students should learn many of the concepts through it. It must give a scope to programming as well as software use. Focus must be given on the data visualization and interpretation of data analytics.

PGDS - 205: Aptitude

(75 Hours)

- Numerical ability - 25 Hours
- Logical reasoning - 25 Hours
- Technical ability - 25 Hours

Examination: There should be Continuous Internal Assessment (CIA-20%) and End of odd Semester Examination (ESE-80%).

PGDS - 105: Effective Communication or Domain Elective

(50 Hours)

(To be designed at departmental level) (Any one from Elective – 101 to Elective 104)

Elective – 101: BFSI

Acquisition analytics: Understand the component of acquisition strategies & practice hands-on exercise of Data analytics for acquiring the potential customers.

Assignment- Acquisition Analytics Build a response model based on the clients, campaign and economic information provided by the Portuguese Bank.

Engagement Analytics: Now that you have learnt how to acquire customers, learn how to engage them and prevent their attrition **RISK ANALYTICS** Learn about the risk associated with customers who default on their loan or credit, and the analytics related to it **mini capstone project** help credx identify the ideal applicants to provide credit cards to by building an application scorecard

Elective – 102: E-Commerce

MARKET MIX MODELLING Learn how to optimise your marketing spends in order to maximise the ROI **RECOMMENDATION SYSTEMS** Learn about the algorithms that power the recommendation engines of the e-commerce sites **ASSIGNMENT - RECOMMENDATION SYSTEMS** Build a recommendation engine based on beer preferences of users **PRICE OPTIMIZATION** Learn how prices are dynamically optimised on an e-commerce platform **A/B TESTING*** Understand the concept behind A/B tests and also learn how to execute an A/B test in Optimizely **MINI CAPSTONE PROJECT** Model the impact of different marketing levers on the sales figure of ElecKart

Elective – 103: Healthcare

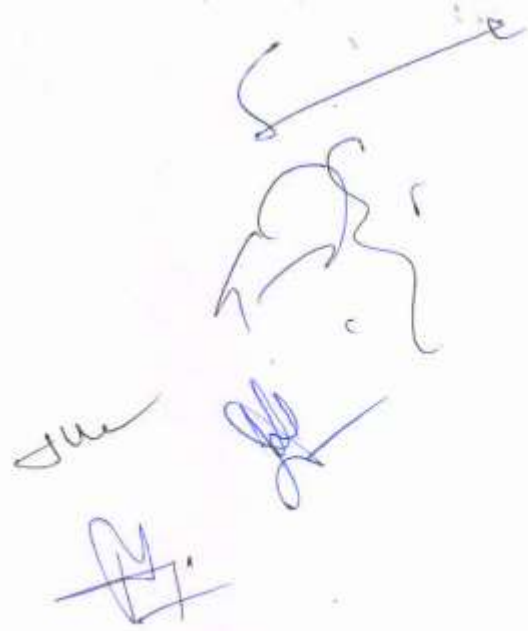
PAYER ANALYTICS In this module, you will explore the different analytics opportunities that exist in the healthcare payer space **ASSIGNMENT- PAYER ANALYTICS** Stratify patients according to the risk of cost they pose to the healthcare payer **PROVIDER ANALYTICS** In this module, you will explore the different analytics opportunities that exist in the healthcare provider space **ANALYTICS IN THE PHARMACEUTICAL INDUSTRIES** Learn how pharmaceutical companies harness the power of data analytics **MINI CAPSTONE PROJECT** Decipher the CMS hospital star rating system using supervised and unsupervised models

Elective – 104: Natural Language Processing

BASICS OF TEXT PROCESSING Get started with the Natural Language Toolkit, learn the basics of text processing in Python **LEXICAL PROCESSING** Learn to extract features from unstructured text and build machine learning models on text data **SYNTAX AND SEMANTICS** Conduct sentiment analysis, learn to parse English sentences and extract meaning from them **OTHER PROBLEMS IN TEXT ANALYTICS** Explore the applications of text analytics in new areas and various business domains

Elective – 105: Deep Learning And Neural Networks:

INFORMATION FLOW IN A NEURAL NETWORK Understand the components and structure of artificial neural networks **TRAINING A NEURAL NETWORK** Learn the cutting-edge techniques used to train highly complex neural networks **Convolutional Neural Networks:** Use CNNs to solve complex image classification problems **RECURRENT NEURAL NETWORKS** Study LSTMs and RNNs applications in text analytics **CREATING AND DEPLOYING NETWORKS USING TENSORFLOW AND KERAS** Build and deploy your own deep neural networks on a website, learn to use tensorflow API and Keras

Handwritten signatures and scribbles in blue ink, including a long horizontal line with an arrowhead at the end, and several illegible signatures.