Tribhuvan University

Institute of Science and Technology

Bachelor in Data Science (BDS)
Curriculum 2024

SCHOOL OF MATHEMATICAL SCIENCES

1. Introduction

Data Science is a field that is growing significantly across all industries that advancing with time and technology. Data science program has become a fundamental discipline in most of the universities as applied research in Statistics and Computer Science. Data Science experts are needed in most government agencies. Businesses depend on big data to better serve their customers. Data Science careers are in high demand and this trend will grow exponentially.

Data Science program is a multi-disciplinary subject that includes the use of **Mathematics, Statistics, and Computer Science** and **Information Technology** to study and evaluate data. The key objective of Data Science is to extract valuable information for use in strategic decision making, product development, trend analysis, and forecasting.

The BDS curriculum is designed to provide the breadth and depth of knowledge needed for a successful career in data science. It emphasizes practical proficiency in applying the relevant skills through courses in statistical modelling, data management, artificial intelligence, machine learning, data visualization, and other related areas to uncover actionable insights hidden in an organization's data. The accelerating volume of data sources, and subsequently data, has made data science is one of the fastest growing field across every industry.

2. Objectives

Upon completion of this program, students will be able to:

- Create data science enabled solutions to different data science related real-world problems applicable to the needs of different sectors.
- Collect, clean, and organize large amounts of data for data analysis.
- Create knowledge and skills in both computer science, statistical modelling, and mathematics for data intensive problem solving.
- Engage in continuous learning understanding ethical, professional, and social issues related to data science.

3. Duration and Nature of the Course

Bachelors in Data Science a full-time program having eight Semesters in four years duration. This program basically comprises of some foundational courses consisting of fundamentals of Mathematics, Statistics, and Computer Science and Technology including algorithms, artificial intelligence, machine learning, computer and statistical programming, data analysis,

database systems, and web development and other data science related subjects. Each course with 3 credit hours except seminar, project work, and internship will be taught 48 hours in each semester.

Total credit hours: 123 Credit hours

Nature of courses: Theoretical, Practical, Project, Seminar, Internship

4. Eligibility Criteria

The candidate applying for admission to the BDS program must have completed 10 + 2 or equivalent examinations in any stream with minimum second division (securing 45% and above) or minimum 'C' grade in all subjects of grade 11 and 12.

5. Evaluation

All the courses except seminar, project work and internship should have internal weightage of 40% and external weightage of 60%. A student should secure minimum of 40% in each category to pass a course. Final examination will be conducted at the end of each semester for each course except seminar, project work and internship. The weightage of this final examination is 60% of the overall weightage. The final grade and grade point in each course will be the sum of overall weightage of in all categories.

For the courses having laboratory work, 50% of the overall internal weightage is assigned for practical examination. The school may invite a subject expert for the evaluation of practical examination.

Seminar, project work and internship are evaluated by different evaluators. To pass these courses, students should secure at least 40% marks in the evaluation of each evaluator and the final grade and grade point will be the sum of all the evaluations. For the evaluation of final presentation, an external examiner will be assigned.

6. Grading System

The grade awarded to each student in each course is based on his/her overall performance through internal and external evaluations. Several evaluation criteria are used for the continuous internal evaluation. External evaluation is solely based on final examination conducted by the University. The grade in each course is assigned using a letter grade that indicates the overall performance of each student in each course. The chart below represents letters with its corresponding grading scale, grade point, and performance remarks.

Letter Grade	Grading Scale	Grade Point	Performance Remarks
A	90 – 100	4	Outstanding
A-	80 – less than 90	3.7	Excellent
B+	70 – less than 80	3.3	Very Good
B	60 – less than 70	3	Good
B-	50 – less than 60	2.7	Satisfactory
C	40 – less than 50	2.3	Pass*
F	0 – less than 40	0	Fail

*Pass refers to acceptable

The performance of each student in each semester shall be evaluated in terms of Semester Grade Point Average (SGPA) which is the grade point average for the semester. SGPA is calculated as

$$SGPA = \frac{Total\ Grade\ Points\ earned\ in\ a\ semester}{Total\ number\ of\ credits\ earned\ in\ the\ semester}$$

The cumulative grade point average (CGPA) is the grade point average for all completed semesters. CGPA is calculated as

$$CGPA = \frac{Total\ Grade\ Points\ earned}{Total\ number\ of\ credits\ completed}$$

7. Teaching Pedagogy

The general teaching pedagogy of BDS includes class lectures, group discussions, case studies, guest lectures, research work, project work (individual and group), assignments (theoretical and practical), and term papers. The teaching faculty will determine the choice of teaching pedagogy as per the need of the course. The concerned faculty shall develop a detailed course outline and work plan at the beginning of each semester.

8. Course Structure

First Semester		
Course Code	**Course Title**	**Credit Hours**
BDS101	Introduction to Data Science	3
BDS102	Basic Computer Organization	3
BDS103	Programming in C	3
BDS104	Statistics for Data science	3
BDS105	Calculus I	3
Total Credit Hours		**15**

Second Semester		
Course Code	**Course Title**	**Credit Hours**
BDS151	Python Programming	3
BDS152	Database Management System	3
BDS153	Probability Distribution	3
BDS154	Calculus II	3
BDS155	Linear Algebra	3
Total Credit Hours		**15**

Third Semester		
Course Code	**Course Title**	**Credit Hours**
BDS201	Data Structure and Algorithms	3
BDS202	Operating System	3
BDS203	R Programming	3
BDS204	Inferential Statistics	3
BDS205	Differential Equations	3
BDS206	Seminar	1
Total Credit Hours		**16**

Fourth Semester		
Course Code	**Course Title**	**Credit Hours**
BDS251	Artificial Intelligence	3
BDS252	Web Development	3
BDS253	Data Communications and Computer Networking	3
BDS254	Discrete Mathematics	3
BDS255	Technical Writing	3
Total Credit Hours		**15**

Fifth Semester		
Course Code	**Course Title**	**Credit Hours**
BDS301	Machine Learning	3
BDS302	Software Design and Development	3
BDS303	Data Visualization	3
BDS304	Numerical Methods	3
BDS305	Economics	3
BDS306	Project I	2
Total Credit Hours		**17**

Sixth Semester		
Course Code	**Course Title**	**Credit Hours**
BDS351	Data Warehousing and Data Mining	3
BDS352	Artificial Neural Network	3
BDS353	Computer Graphics and Image Processing	3
BDS354	Research Methodology	3
BDS355	Principles of Management	3
Total Credit Hours		**15**

Seventh Semester		
Course Code	**Course Title**	**Credit Hours**
BDS401	Object Relational and NoSQL Databases	3
BDS402	Simulation and Modeling	3
BDS403	Data Security	3
BDS404	Project II	3
Elective I (Chose any one)		
BDS421	Cloud Computing	3
BDS422	Deep Learning	3
BDS423	Mobile Application Development	3
BDS424	Blockchain Technology	3
BDS425	Exploratory Data Analysis	3
Total Credit Hours		**15**

Eighth Semester		
Course Code	**Course Title**	**Credit Hours**
BDS451	Information Retrieval	3
BDS452	Big Data Analytics with Hadoop	3
BDS453	Natural Language Processing	3
BDS454	Internship	3
Elective II (Chose any one)		
BDS471	Social Network Analysis	3
BDS472	Forecasting Analysis	3
BDS473	Digital Marketing	3
BDS474	Business Intelligence	3
BDS475	Internet of Things	3
Total Credit Hours		**15**

Tribhuvan University
Institute of Science and Technology

SCHOOL OF MATHEMATICAL SCIENCES

Course of Study

Program: Bachelor in Data Science
Paper: **Introduction to Data Science**
Nature: Theory
Semester: I

Full Marks: 45+30
Pass Mark: 18+12
Course *Code*.: **BDS101**
Credit Hours: 3

Course Description:

This is an introductory course to teach the basics of data science, it's applications, and commonly used tools and techniques. The course is designed to introduce key ideas and methodologies used in the domain of data science. The goal of this course is to help understand the fundamental building blocks of data science.

Course Objectives:

Upon the conclusion of the course, students should be able to

- Describe Data Science, skill sets needed to be a data scientist and be familiar with common tools used for data science.
- Understand the importance of data quality and familiarize with common data munging techniques
- Understand and apply commonly used data analysis and machine learning techniques in data science
- Identify the challenges in handling big data, and gain a general understanding of ecosystem of big data
- Reason around ethical and privacy issues in data science and understand the common biases affecting data science

Course Contents:

Unit 1 Introduction to Data Science **6 hrs**

Introduction to data science; Data Science Hype; Data, Data Science, Engineering and Data-Driven Decision Making; Statistics and Data Science; Data Science Process; Data Science Profile; Data Analytics vs Data Science; Roles and Responsibilities of Data Scientist; Data Science Lifecycles (OSEMN, CRISP-DM, TDSP); Tools and Technologies; Limitations of data science; Applications of Data Science

Unit 2 Big Data **8 hrs**

Structured, Semi-Structured and Unstructured data; Understanding Database, Data Warehouse and Data Lake; Characteristics of Data warehouse, ETL(Extract-transform-load) Techniques

Introduction to Big Data; Need for Big Data; Challenges of handling big data; Characteristics of big data; Map-Reduce programming paradigm, and its differences from conventional programming models; Solving Word Count Problem through Map Reduce Paradigm; Hadoop and its components; Hadoop Ecosystem

Unit 3 Data Wrangling and Feature Engineering 12 hrs

Commonly used data formats; Collecting and Importing data; Exploratory Data Analysis; Data Quality; Common issues with real world data; Data Cleaning Techniques; Data Enrichment; Data Validation; Data Publishing

Feature Engineering: Introduction to Feature Engineer; Feature Selection; Feature Selection Techniques: Filters, Wrappers and Embedded Methods; Feature Scaling and Standardization; Feature Extraction

Unit 4 Machine Learning 12 hrs

Understanding Predictive analytics and Machine Learning; Artificial Intelligence vs Machine Learning; their practical applications; Machine Learning Techniques: Supervised, Unsupervised, Semi-supervised and Reinforcement learning and their types.

Regression Techniques: Linear Regression, Polynomial Regression; Classification Techniques: Logistic Regression, KNN, Decision Tree, Naïve Bayes; Clustering Techniques: K Means, K Medoids; and their pros and cons.

Model Evaluation: Root Mean Squared Error, Mean Absolute Error, Mean Percentage Error; Confusion Matrix; Accuracy, Precision and Recall

Unit 5 Data Visualization and Story Telling 6 hrs

Introduction to Data Visualization, Exploratory vs Explanatory data visualization; Common Data Visualization Techniques and their usage: Table, Pivot Table, Histogram, Bar Chart, Line Chart, Scatter plot, Pie Chart, Box Plot etc.

Data Story Telling: Introduction, Need for Data Story Telling; Components of Data Storytelling; Benefits of Data Story Telling; Communicating data insights.

Unit 6 Ethical Issues in Data Science 4 hrs

Ethics for Data Scientist; Case Study of Facebook and Cambridge Analytica; Common issues with privacy and data ethics

Introduction of biasness and fairness; Issues with fairness and bias in data science; Common Cognitive biases: Anchoring Bias, Sampling Bias, In group favoritism and out-group negativity, Fundamental attribution error, Negativity bias, Stereotyping, Bandwagon effect, Bias blind spot; Addressing Cognitive biases: Group unaware selection, Adjusted group thresholds, Demographic parity, Equal opportunity, Precision parity

Reference Books:

1. O'Neil, Cathy and Schutt, Rachel (2013), *Doing Data Science***,** *Straight Talk From TheFrontline*, O'Reilly Media
2. Skiena, Steven (2017), *The Data Science Design Manual*, Springer
3. Provost, Foster and Fawcett, Tom (2013). *Data Science for Business***:** *What You Need to Know about Data Mining and Data-analytic Thinking*, O'Reilly Media

Tribhuvan University
Institute of Science and Technology
SCHOOL OF MATHEMATICAL SCIENCES

Course of Study

Program: Bachelor in Data Science
Paper: **Basic Computer Organization**
Nature: Theory + Practical
Semester: I

Full Marks: 45+30
Pass Mark: 18+12
Course *Code*.: **BDS102**
Credit Hours: 3

Course Description:

This course covers fundamental principles of computer systems, instructions sets and its design, hierarchy of computer organization, from the digital logic level to the operating system level. It focuses on a modular approach to understanding computer systems and clarifies complex concepts through a systematic presentation.

Course Objectives:

The main objective of this course is to provide comprehensive understanding of how computer systems are structured and function.

Course Contents:

Unit 1 Introduction **8 hrs**

Structured Computer Organization: Languages, Levels, and Virtual Machines; Contemporary Multilevel Machines; Evolution of Multilevel Machines; Milestones in Computer Architecture: Different Generation of Computers; The Computer Zoo: Different Technologies and Different Types of Computers

Unit 2 Computer Systems **10 hrs**

Processors: CPU Organization, Instruction Execution, RISC versus CISC, Design Principles for Modern Computers, Instruction-Level Parallelism, Processor-Level Parallelism; Primary Memory: Bits, Memory Addresses, Byte Ordering, Error-Correcting Codes, Cache Memory, Memory Packaging and Types; Secondary Memory: Memory Hierarchies, Magnetic Disks, IDE Disks, SCSI Disks, RAID, Solid-State Disks, CD-ROMs; Input/Output: Buses, Terminals, Mice, Game Controllers, Printers

Unit 3 Digital Logic Level **12 hrs**

Gates and Boolean Algebra: Gates, Boolean Algebra, Implementation of Boolean Functions, Circuit Equivalence; Basic Digital Logic Circuits: Integrated Circuits, Combinational Circuits, Arithmetic Circuits, Clocks; Memory: Latches, Flip-Flops, Registers, Memory Organization, Memory Chips, RAMs and ROMs; CPU Chips and Buses: CPU chips, Computer Buses, Bus Width, Bus Clocking, Bus Arbitration, Bus Operations

Unit 4 The Microarchitecture Level **4 hrs**

The Data Path, Microinstructions and Notation; Introduction to pipelining

Unit 5 The Instruction Set **10 hrs**

Brief Introduction to ISA Level and its properties; Memory Models; Registers; Instructions, Instruction Formats and Design Criteria for Instruction Formats; Different Addressing Modes

Unit 6 The Operating System **4 hrs**

Virtual Memory; Paging; Segmentation

Laboratory Works:

Students should implement different combinational and sequential circuits in a standard simulator. They must be able to realize pipelining and different addressing modes.

Reference Books:

1. Andrew S. Tanenbaum and Todd Austin, "Structured Computer Organization", Pearson Education, Sixth Edition
2. William Stallings, "Computer Organization and Architecture – Designing for Performance", Pearson Education, Ninth Edition,
3. David A. Patterson and John L. Hennessy, "Computer Organization and Design: The Hardware/Software interface", Elsevier, Third Edition

Tribhuvan University
Institute of Science and Technology
SCHOOL OF MATHEMATICAL SCIENCES

Course of Study

Program: Bachelor in Data Science
Paper: **Programming in C**
Nature: Theory+ Practical
Semester: I

Full Marks: 45+30
Pass Mark: 18+12
Course *Code*.: **BDS103**
Credit Hours: 3

Course Description:

This course introduces the both theoretical and practical concepts of C programming language including basic concepts, data types, operators, handling input and output, control statements, arrays, functions, pointers, structures, unions, and file handling.

Course Objectives:

The main objective of this course is to familiarize students both theoretical and practical concepts of C programming language.

Course Contents:

Unit 1 Introduction **3 hrs**

Program and programming language, Types of programming languages, Program design tools (Algorithm, Flowchart, and Pseudocode), History of C programming, Structure of C program, Compiling and executing C programs, Debugging

Unit 2 Basic Elements of C **4 hrs**

C standards, C character set, C tokens, Escape sequence, Delimiters, Variables, Data types, Constants/Literals, Symbolic constant, Expressions, Statements, Writing comments, Library functions and Pre-processor directives

Unit 3 Data Input and Output **3 hrs**

Input/output operations, Conversion specifications, Formatted I/O and unformatted I/O

Unit 4 Operators and Expression **5 hrs**

Unary and binary operators, Arithmetic operator, Relational operator, Boolean operator, Assignment operator, Ternary operator, Bitwise operator, Increment or decrement operator, Conditional operator, Special operators (size of operator), Evaluation of expression, Operator precedence and associativity, Type conversion

Unit 5 Control Statements **6 hrs**

Branching statements (if and switch), Looping statements (for, while, and do-while), Nested control structures, Break and continue statements, Exit function

Unit 6 Arrays and Strings **6 hrs**

Introduction to array, Types of array (Single dimensional and multidimensional), Declaration and memory representation of array, Initialization of array, Character array and strings, Reading and writing strings, Null character, String library functions

Unit 7 Functions **6 hrs**

User defined functions, Library functions vs. User defined functions, Function prototype, Function call and Function definition, Nested and recursive function, Function arguments and return types, Passing arrays to function, Passing strings to function, Passing arguments by value, Passing arguments by address, Local and global variable, Scope visibility and lifetime of a variable, Macros

Unit 8 Pointers **6 hrs**

Introduction to pointers, Advantages and disadvantages of pointer, The & and * operator, Declaration of pointer, Pointer to pointers (Chain of Pointers), Pointer arithmetic, Pointers and arrays, Pointers and character strings, Array of pointers, Pointers as function arguments, Function returning pointers, Dynamic memory allocation

Unit 9 Structure and Union **5 hrs**

Introduction, Declaration, Initialization, Array of structure, Passing structure to function, Passing array of structure to function, Nested structure, Pointers and structures, Introduction to union, Structure vs union

Unit 10 File Handling in C **4 hrs**

Concept of file, Types of file (Text and binary files), Modes of file, Opening and closing of file, Input output operations in file, Random access in file

Laboratory Works:

Laboratory work includes writing C programs to implement all the concepts of C programming studied in each unit of the course.

Reference Books:

1. Byron Gottfried: "Programming with C," Fourth Edition, McGraw Hill Education.
2. Brian W. Keringhan, Dennis M. Ritchie, The C programming Language, Second Edition, PHI Publication.
3. Al Kelley, Ira Pohl: "A Book on C", Fourth Edition, Pearson Education.
4. YeshvantKanetkar, "Let Us C", 17 th Edition, BPB publication, 2020.
5. Herbert Schildt, C Complete Reference, Fourth Edition, Osborne/McGraw- Hill Publication.
6. K.N. KING: C Programming: A Modern Approach, Second Edition
7. E. Balagurusamy, Programming in ANSI C, Eigth Edition, TMH publication, 2019

Tribhuvan University
Institute of Science and Technology
SCHOOL OF MATHEMATICAL SCIENCES

Course of Study

Program: Bachelor in Data Science
Paper: **Statistics for Data Science**
Nature: Theory+ Practical
Semester: I

Full Marks: 45+30
Pass Mark: 18+12
Course *Code*.: **BDS104**
Credit Hours: 3

Course Description:

This course focuses on fundamental statistical techniques that are of particular relevance to Data Science. The course covers basic statistics with exploratory data analysis, moments, skewness and kurtosis, correlation and regression analysis, probability, index number, time series analysis and forecasting.

Course Objectives:

After successful completion of the course, students will be able to

- Describe the nature of data, summarize and explore data
- Understand and apply descriptive statistical measures relevant to Data Science
- Perform correlational analysis and build linear regression model with model adequacy tests
- Compute index numbers for economic analysis
- Perform time series analysis and develop forecasting models for time series data
- Understand about basic principles and properties of probability, calculate marginal, joint and conditional probabilities using addition, multiplication, and Bayes' rule

Course Contents:

Unit 1 Introduction **2 hrs**

Definition, scope and importance of Statistics; Population and sample data; Variable and its types: discrete and continuous; Qualitative and quantitative Sources of data: primary and secondary; Types of data: cross-sectional, temporal and panel; Scale of measurement of data: nominal, ordinal, interval and ratio; Application of Statistics in Data Science

Unit 2 Data summarization **10 hrs**

Definition and need of data summarization; Methods of data summarization: tabular and graphical methods, frequency tables, cross-tabulations, bar diagrams, pie chart, histogram, frequency curve, ogives, frequency polygon; Measures of central tendency, partition values, dispersion, asymmetry (skewness) and kurtosis and their computational methods; Moments: raw and central moments, relation between raw and central moments, properties of moments; Exploratory data analysis: stem-and-leaf display, five number summary, box-plot;Solving problems relevant to Data Science

Unit 3 Correlation and regression analysis **12 hrs**

Correlation and its types, assumptions, scatterplot, Karl Pearson's correlation coefficient: definition, computation, properties; Spearman's and Kendall's rank correlation coefficients; Simple and multiple linear regressions, assumptions of linear regressions model, lines of regressions, fitting linear regression using ordinary least square method, interpretation of intercept and slope, relation between correlation and regression parameters, goodness of fit, coefficient of determination, residual analysis; Solving problems relevant to Data Science

Unit 4 Index numbers **8 hrs**

Introduction, simple and weighted index numbers, methods of construction of index numbers for price, quantity and value indices using unweighted and weighted aggregate methods: Laspeyers, Paasche, Fisher, fixed-weight aggregate methods, average of relative method; Consumer price index, purchasing power of money, real wages, base shifting, inflation and deflation; Time reversal and factor reversal tests; Issues in constructing and uses of index numbers; Applications of different index number techniques in monetary and finance related data

.

Unit 5 Time series analysis and forecasting **10 hrs**

Introduction, components of a time series- secular trend, cyclic variation, seasonal variation, irregular variation; Time series analysis for forecasting using trend projection method, moving average method, curvilinear methods; Seasonal variation, determining the seasonal index, de-seasonalizing data; Test for autocorrelation; Forecasting methods- naive methods, linear methods, moving average methods, and simple exponential smoothing method; Measuring accuracy of forecasting models using mean square error (MSE), mean absolute deviation (MAD) and mean absolute percentage error (MAPE); Solving problems relevant to Data Science

Unit 6 Probability **6 hrs**

Review of fundamental concepts of probability, counting rules, types of probability, laws of probability, conditional probability, pair-wise and mutually independence; Bayes' theorem and its application, prior and posterior probability. Solving problems relevant to Data Science

Laboratory Works:

The laboratory work includes using any statistical software such as Microsoft Excel, SPSS, R etc.

Practical Problems

S. No.	Title of the Practical Problems	No. of practical Problems
1	Mean, median, mode, partition values, measures of dispersion and coefficient of variation.	2
2	Measure of skewness and kurtosis, five number summary and Box and whisker plot.	2
3	Scatter diagram, Karl Pesrso's correlation coefficient and Spearman's Rank correlation coefficient (ungrouped data) and interpretation. Compute manually and check with computer output.	2
4	Fitting of simple regression model, calculation of error term, standard error of the estimate, coefficient of determination and interpretation. Residuals plot	2
5	Fitting of multiple regression model, calculation of residuals, standard error of the estimate, coefficient of determination and interpretation.	2
6	Computation of probability in different conditions and Bayes' theorem.	1
7	Calculation of unweighted and weighted index number.	2
8	Linear and quadratic models, percent of trend, relative cyclical residual and interpretation of results.	1
9	Forecasting models	1

References Books:

1. Bruce Peter and Bruce Andrew (2017). *Practical Statistics for Data Scientists,* O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472, U.S.A.
2. Sheldon M. Ross (2010). *Introductory Statistics*, Elsevier Publication, U.S.A.
3. Neil Weiss (2010): *Introductory Statistics*, 5th edition: Pearson Publishers
4. Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers and Keying YE, *Probability and Statistics for Engineers and Scientists*, Pearson, 2012.
5. Anderson David R., Williams Thomas A., Cochran James J., Sweeney Dennis, Camm Jeffrey D., Ohlmann Jeffrey W. *Statistics for Business & Economics*, 14th Edition, Cengage Learning Inc., U.S.A.

Tribhuvan University
Institute of Science and Technology

SCHOOL OF MATHEMATICAL SCIENCES

Course of Study

Program: Bachelor in Data Science | *Full Marks*: 45+30
Paper: **Calculus I** | Pass Mark: 18+12
Nature: Theory | Course *Code*.: **BDS105**
Semester: I | *Credit Hours*: 3

Course Description:

This course includes limits, continuity, differentiation and integration of algebraic and transcendental functions, and their applications.

Learning Objectives:

After successful completion of this course, student will be able to

- Solve the problems of limit, continuity, differentiation and integration
- Apply the techniques of calculus to solve real life problems

Course Contents:

Unit 1 Limits and Continuity **8 hrs**

Limit of a function; Calculating limits using the limit laws; Computing limits of algebraic functions by definition; Limits at infinity; Asymptotes: horizontal, vertical and oblique, Continuity at a point

Unit 2 Derivatives **7 hrs**

Derivatives and their rules; Linear approximations and differentials; Indeterminate forms and L'Hospital's rule; Higher order derivatives

Unit 3 Application of Derivative **10 hrs**

Rolle's Theorem and Mean value theorem; Monotonic functions; Optimization problems; Marginal analysis; Sketching curve using calculus; Newton's method

Unit 4 Integrals **13 hrs**

Definite integral; Fundamental theorems of calculus; Indefinite integrals and their properties; Properties of definite integrals; Reduction formulae; Improper integrals; Beta and Gamma functions; Numerical methods of integration.

Unit 5 Application of Integration **10 hrs**

Areas between curves; Volumes; Volumes by cylindrical shells; Average value of a function; Arc length; Area of a surface of revolution; Applications in economics

Reference Books:

1. Stewart J., Calculus: Early Transcendental Functions, 7th edition, Thomson Brooks/Cole
2. Larson, et al, Calculus: Early Transcendental Functions, Houghton Mifflin, 2011
3. Robert A. Adams, Christopher Essex, Calculus: a complete course, Pearson, 2010

Tribhuvan University
Institute of Science and Technology
SCHOOL OF MATHEMATICAL SCIENCES

Course of Study

Program: Bachelor in Data Science
Paper: **Python Programming**
Nature: Theory+ Practical
Semester: II

Full Marks: 45+30
Pass Mark: 18+12
Course *Code*.: **BDS151**
Credit Hours: 3

Course Description:

This course covers different concepts of Python programming including basic language concepts, control statements, functions, comprehensions, generators, decorators, iterators, object-oriented programming concepts, exception handling, file handling, modules and packages, and some common Python libraries.

Course Objectives:

The main objective of this course to provide knowledge to student with different concepts of Python programming language python programming and some common Python libraries such as Numpy, Pandas, and Matplotlib.

Course Contents:

Unit 1 Introduction **3 hrs**

Python introduction; Why Python; Installing Python; Using interactive shell from console; Running Python scripts from console; Using interactive shell from IDLE; Using IDEs; Installing third party libraries; Writing comments; Indentation; Variables; Operators

Unit 2 Built-In Data Types **7 hrs**

Numbers – integers, Booleans, Real numbers, Complex numbers, Fractions and decimals; Immutable sequences – strings, bytes, and tuples; Mutable sequences – lists, bytearrays; Set types; Mapping types – dictionaries; Dates and times; Collections module; Small value cashing; Indexing and slicing.

Unit 3 Conditionals and Iterations **5 hrs**

Introduction; The if statement; The for loop; The while loop; The break and continue statements; The else clause after for or while loops; The walrus operator

Unit 4 Functions **4 hrs**

Introduction; Why use functions? Defining functions; passing arguments; Return values and returning multiple values; Recursive function; Anonymous function; Built-in functions

Unit 5 Comprehensions and Generators **3 hrs**

The map, zip, and filter functions; Comprehensions – list, Dictionary and set comprehensions; Generators

Unit 6 Object-oriented Programming, Decorators, Iterators, and Modules **14 hrs**

Decorators; Object-oriented programming concepts; Creating class and object; Inheritance; Static methods and class methods; Private properties; The property decorator; Operator overloading; Writing custom iterator; Modules and packages.

Unit 7 Exception and File Handling **4 hrs**

Exceptions; Handling exceptions; Defining your own exception; Reading and writing files; Reading and writing CSV files.

Unit 8 Basic Data Processing and Analysis Using Python Libraries 6 hrs

Numpy: Introduction array creating; Dimensions; Data types, Array attributes, Indexing and slicing; Array copy and view; Creating array from numerical range; Array broadcasting; Iterating over array; Sorting and Searching; Statistical Functions

Pandas: Series and data frames; Creating data frames; The head and tail functions; Attributes; Working with missing data; Indexing, slicing, and subsetting; Merging and joining data frames; Working with CSV data.

Unit 9 Data Visualization with Matplotlib 2 hrs

Introduction; Marker; Line; Color; Label; Grid lines; Subplot; Scatter plot; Bra graph; Histogram, pie chart and box plot.

Laboratory Works:

The laboratory work includes writing computer programs using Python programming language covering all the concepts studied in each unit of the course.

References Books:

1. Fabrizio Romano and Heinrich Kruger, *Learn Python Programming – An in-depth Introduction to the Fundamentals of Python,* 3rd Edition, Packt Publishing, 2021
2. Kenneth A Lampart, *Fundamental of Python,* Cengage Learning Publishing.
3. Cody Jackson, *Learn Programming in Python with Cody Jackson*, Packt Publishing, Wesley, 2018
4. Mark Summerfield, *Programming in Python 3: A Complete Introduction to the Python Language*, Addison-Wesley Professional.

Tribhuvan University
Institute of Science and Technology
SCHOOL OF MATHEMATICAL SCIENCES

Course of Study

Program: Bachelor in Data Science
Paper **Database Management System**
Nature: Theory+ Practical
Semester: II

Full Marks: 45+30
Pass Mark: 18+12
Course *Code*.: **BDS152**
Credit Hours: 3

Course Description:
A course in database management systems provides an in-depth understanding of how to design, implement, and maintain a database. The course covers the fundamentals of database management system, data modeling, relational database, structured query language, normalization, transaction, concurrency control, and recovery.

Course Objectives:

The main objective of this course is to provide fundamental concepts of database management system including data modeling, database design, SQL, transaction, concurrency control and database recovery.

Course Contents:

Unit 1 Introduction to Database and Database Users **2 hrs**

Introduction to Database, Database Management System; Database System Environment; Characteristics of Database Approach; Actors in Scene; Workers behind Scene; Advantages of DBMS

Unit 2 Database System Concepts **4 hrs**

Data Models, Schemas, Instances; Three Schema Architecture, Data Independence; DBMS Languages, DBMS Interfaces; Database System Environment; Centralized and Client/Server Architectures for DBMS; Classification of DBMS

Unit 3 Data Modelling using ER Model **7 hrs**

Data Models, Hierarchical Model, Network Model, Relational Model, Object Base Data Model, Entity Relationship Model; Entity Types, Entity Sets, Attributes, Keys, Relationship Types, Relationship Sets, Roles, Structural Constraints; Weak Entity Types; ER Diagrams, Naming Conventions, Design Issues; Relationship Types: Binary, Ternary, N-ary; Extended Entity Relationship, Subclasses, Superclasses, and Inheritance; Specialization and Generalization; Constraints and Characteristics of Specialization and Generalization, Union Types

Unit 4 Relational Data Model and Constraints **5 hrs**

Relational Model Concepts, Domains, Attributes, Tuples, Relations, Characteristics of Relations, Relational Model Notation; Relational Model Constraints and Relational Database Schemas, Domain Constraints, Key Constraints, Null Value Constraints, Relational Databases, Relational Database Schemas; Entity Integrity, Referential Integrity, Foreign Keys; Update Operations and Dealing with Constraint Violations, Insert, Delete and Update Operations, ER to Relational Mapping

Unit 5 Structured Query Language (SQL) **12 hrs**

Data Definition and Data Types, CREATE Schema, CREATE Table, Attribute Data Types and Domains; Specifying Constraints, Attribute Constrains and Defaults, Specifying Keys

and Referential Integrity Constraints, Naming Constraints, Constraint using CHECK; Basic Retrieval Queries, SELECT-FROM-WHERE, Ambiguous Attribute Name, Aliasing, Renaming, Unspecified WHERE Clause, Using Asterisk, Pattern Matching and Arithmetic Operators, ORDER BY; Complex Retrieval Queries, Dealing with NULL Values, Nested Queries, Correlated Nested Queries, EXISTS and UNIQUE Functions, Inner Join and Outer Joins, Aggregate Functions, GROUP BY and HAVING Clauses ; INSERT, DELETE, and UPDATE Statements; Views, DROP and ALTER Commands

Unit 6 Functional Dependencies and Normalization for Relational Databases 6 hrs

Informal Design Guidelines for Relational Schemas Insertion, Deletion, Modification Anomalies, Spurious Tuples; Functional Dependencies; Normalization, First, Second and Third Normal Forms; Boyce-Codd Normal Form; Multivalued Dependency and Fourth Normal Form; Properties of Relational Decomposition, Dependency Preservation, Nonadditive(lossless) Join Properties

Unit 7 Concepts of Transaction Processing 7 hrs

Introduction to Transaction Processing, Transactions, Database Items, Read Write Operations, Need of Concurrency Control, Need of Recovery; Transaction and System Concepts, Transaction States and Operations, Log, Commit Point; Properties of Transactions; Schedules, Conflicting Operations, Characterizing Schedules Based on Recoverability; Characterizing Schedules Based on Serializability, Serial, Non-serial and Conflict-Serializable Schedule, Conflict Equivalence, Serializable Schedules, Precedence Graph for Serializability Testing, Using Serializability for Concurrency Control

Unit 8 Concurrency Control and Database Recovery Techniques 5 hrs

Two-Phase Locking Techniques, Binary Locks, Shared/Exclusive Locks, Basic, Conservative, Strict, and Rigorous Two-Phase Locking; Deadlock and Starvation Timestamp Ordering, Timestamp, Timestamp Ordering Algorithm, for Concurrency Control Recovery Concepts, Immediate and Differed Updates, Buffering, Write Ahead Logging, Steal/No-Steal, Force/No-Force, Checkpoints, Cascading Rollback, NO-UNDO/REDO Recovery Based on Deferred Update; Recovery Technique Based on Immediate Update; Shadow Paging; Database Backup

Laboratory Works:

The laboratory work includes designing conceptual ER diagrams using some CASE tools. The laboratory work further should include writing database queries in SQL for creating databases, schemas, and tables, inserting values, updating, manipulating, altering and deleting data as well as databases, tables and schemas.

Reference Books:

1. Ramez Elmasri, Shamkant B. Navathe, Fundamentals of Database Systems; Seventh Edition, Pearson Education.
2. Avi Silberschatz, Henry F Korth, S Sudarshan, Database System Concepts, Seventh Edition, McGraw-Hill.
3. Raghu Ramakrishnan, Johannes Gehrke, Database Management Systems, Third Edition;; McGraw-Hill.
4. C. J. Date, An Introduction to Database Management Syste, 8th Edition, Pearson
5. Jaffrey D. Ullman, Jennifer Widom, A First Course in Database Systems, Third Edition, Pearson Education Limited.

Tribhuvan University
Institute of Science and Technology

SCHOOL OF MATHEMATICAL SCIENCES

Course of Study

Program: Bachelor in Data Science
Paper: **Probability Distribution**
Nature: Theory+ Practical
Semester: II

Full Marks: 45+30
Pass Mark: 18+12
Course *Code*.: **BDS153**
Credit Hours: 3

Course Description:
This course focuses on probability distribution related to fundamental statistical techniques that are of particular relevance to Data Science. The course covers random variables, mathematical expectation, discrete and continuous probability distributions, bivariate distributions and their applications.

Course Objectives:
After successful completion of the course, students will be able to

- Describe about probability distribution along with essential characteristics and properties
- Calculate and describe mathematical expectation and its properties
- Describe the characteristics features of and compute probabilities using discrete and continuous probability distributions
- Describe the bivariate distributions and calculate moments of bivariate distributions, conditional expectations, conditional variance
- Use Jacobian of transformations for bivariate distributions
- To apply these techniques in different scenario

Course Contents:

Unit 1 Random Variables **8 hrs**

Random variables and their properties, types of random variables: qualitative (Categorical) and quantitative, discrete and continuous random variables; Probability distribution of a random variable, probability mass function, probability density function and its properties, functions of random variable; Transformation of random variable; Joint probability mass function, joint probability distribution function, joint probability density function, conditional probability mass function and conditional probability density function. Solving Data Science related numerical problems

Unit 2 Mathematical Expectation **8 hrs**

Mathematical expectation of a random variable, properties of mathematical expectation, addition and multiplicative theorems of expectation, covariance and correlation, conditional expectation, conditional variance, variance of linear combination of random variables; Moments of random variables, raw and central moments; Generating functions: moment generating function, characteristic function, probability generating function, cumulant generating function with their properties. Solving Data Science related numerical problems**.**

Unit 3 Discrete Probability Distribution **10 hrs**

Binomial distribution, Poisson distribution, negative binomial distribution, geometric distribution, hyper-geometric distribution, negative hyper-geometric distribution: their mass functions, distribution functions, moments, moment generating functions, characteristic functions and properties, distribution fittings. Solving Data Science related numerical problems.

Unit 4 Continuous Probability Distribution **12 hrs**

Uniform distribution, Normal distribution, Lognormal distribution, Exponential distribution, Beta distribution, Gamma distribution, Weibull distribution; their density functions, distribution functions, moments, moment generating functions, characteristic functions, properties and uses with distribution fittings. Solving Data Science related numerical problems.

Unit 5 Bivariate Distribution **10 hrs**

Bivariate distribution for discrete and continuous variables: joint, marginal, and conditional distributions, independence of random variables; Transformations of random variables: Jacobian of transformations, distributions of sum, product and ratio of random variables. Solving Data Science related numerical problems.

Laboratory Works:

The laboratory work includes using any statistical software such as Microsoft Excel, SPSS, R etc.

Practical Problems

S. No.	Title of the Practical Problems	No. of practical Problems
1	Random number generation, expectation, variance, covariance, raw and central moments.	3
2	Calculation of probabilities using Binomial distribution, Poisson distribution, negative binomial distribution and fitting of distributions	5
3	Calculation of probabilities using normal distribution, lognormal distribution, exponential distribution, and fitting of distributions	5
4	Calculation of descriptive statistics of probability distribution.	2

References Books:

1. Blitzstein Joseph K. & Hwang Jessica (2015). *Introduction to Probability*, CRC Press
2. Baron Michael (2014). Probability and Statistics for Computer Scientists, CRC Press
3. Gupta S. C. and Kapoor V. K. (2007). *Fundamentals of Mathematical Statistics*, Sultan Chand and Sons
4. Rohatgi V. K. and Ehsanes Saleh, A. K. MD (2005). *An Introduction to Probability and Statistics*, John Wiley & Sons
5. Bruce Peter and Bruce Andrew (2017). *Practical Statistics for Data Scientists,* O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472, U.S.A.

Tribhuvan University
Institute of Science and Technology

SCHOOL OF MATHEMATICAL SCIENCES

Course of Study

Program: Bachelor in Data Science
Paper: **Calculus II**
Nature: Theory
Semester: II

Full Marks: 45+30
Pass Mark: 18+12
Course *Code*.: **BDS154**
Credit Hours: 3

Course Description:

This course is a continuation in Calculus I. It covers parametric equations and polar coordinates, infinite sequences and series, Fourier analysis. It also focuses on the comprehensive treatment of partial derivatives, optimization problems and multiple integrals.

Course Objectives:

After successful completion of this course the student will be able to

- solve problems of infinite sequences and series
- obtain the elementary knowledge of Fourier series and Fourier transform
- to calculate derivatives of multivariable functions
- to solve the problems of double and triple integrals
- apply the techniques of calculus to solve real life problems.

Course Contents:

Unit 1 Parametric Equations and Polar Coordinates **8 hrs**

Curves defined by parametric equations, Calculus with parametric curves, Polar coordinates, Areas and lengths in polar coordinates, Conic sections, Conic sections in polar coordinates.

Unit 2 Infinite Sequences and Series **12 hrs**

Sequences, Series, Comparison tests, Alternating series, Absolute convergence, Ratio test, Root test, Integral test and estimates of sums, Power series, Taylor and Maclaurin series, Taylor polynomials and their applications.

Unit 3 Fourier Analysis **8 hrs**

Fourier series, Periodic functions, Odd and even functions, Fourier series for arbitrary range, Half range Fourier series, Fourier integral theorem, Fourier sine and cosine integral; Complex form of Fourier integral, Fourier transform, Fourier sine transform, Fourier cosine transform and their properties.

Unit 4 Partial Derivatives **11 hrs**

Functions of two and three variables, Limits and continuity, Partial derivatives, Higher order partial derivatives, Tangent planes and linear approximations, Directional derivatives and gradient vector, Total differentials, Euler's theorem of two and three variables, Maximum and minimum values, Lagrange multipliers.

Unit 5 Multiple Integrals **9 hrs**

Double integrals over rectangle, Iterated integrals, Double integrals over general regions, Double integrals in polar coordinates, Applications of double integrals, Triple integrals, Change of variables in multiple integral.

Reference Books:

1. Stewart J., Calculus: Early Transcendental Functions, 7th edition, Thomson Brooks/Cole.
2. Larson, et al, Calculus: Early Transcendental Functions, Houghton Mifflin, 2011.
3. Robert A. Adams, Christopher Essex, Calculus: a complete course, Pearson, 2010.
4. Kreyszig, E., Advance Engineering Mathematic, Tenth edition, Wiley, New York.

Tribhuvan University
Institute of Science and Technology

SCHOOL OF MATHEMATICAL SCIENCES

Course of Study

Program: Bachelor in Data Science
Paper: **Linear Algebra**
Nature: Theory
Semester: II

Full Marks: 45+30
Pass Mark: 18+12
Course *Code*.: **BDS155**
Credit Hours: 3

Course Description:

This course emphasizing topics useful in other disciplines covers fundamental algebraic tools involving matrices and vectors to study linear systems of equations and linear transformations, eigenvalues and eigenvectors and their wide range of applications.

Course Objectives:

After successful completion of this course the student will be able to

- use matrix and determinants to solve various mathematical and real life problems.
- acquire knowledge of vectors spaces and linear transformations.
- apply eigenvalues and eigenvectors in solving various problems.

Course Contents:

Unit 1 Matrix and Determinants **10 hrs**

Algebra of matrices, Determinants and its properties, Application of determinants, Complex matrices, Rank of matrices, System of linear equations and its matrix form, Row reduction and echelon forms, LU factorization.

Unit 2 Vectors Spaces **12 hrs**

Points in n-space, Algebra of points in n-space, Scalar and dot product, Norm and its properties, Distance, Angle between two vectors, Orthogonality, Scalar and vector projections, Cosines of lines, Projections, Vector spaces and subspaces, Linear combination, dependence and independence, Span, basis and dimension.

Unit 3 Linear Transformations **8 hrs**

Linear transformations, Kernel and image, Algebra of linear transformations, Matrix representation of a linear transformations, Four fundamental subspaces, Applications to difference equations, Applications to Markov chains.

Unit 4 Orthogonality **8 hrs**

Orthogonal vectors and sets, Orthogonal bases and Gram-Schmidt Process, QR-factorization, Least squares method.

Unit 5 Eigenvalues and Eigenvectors **10 hrs**

Eigenvalues and Eigenvectors, Cayley- Hamilton theorem and its application, Eigenvalue decomposition, Diagonalization of a matrix, Difference equations and powers A^k, Singular value decomposition.

Reference Books:

1. David C. Lay, *Linear Algebra and its Applications*, Pearson Education, 2012.
2. Gilbert Strang, *Introduction to Linear Algebra,* 4th Edition, Wellesley- Cambridge Press.
3. Howard Anton, Chris Rorres, *Elementary Linear Algebra*: *Applications Version,* Wiley, 2014.