Osmania University

Faculty of Informatics

Diploma in Data Science

Scheme and Syllabi for Semesters – I and II

With Effect from Academic Year 2021 – 2022

Osmania University

Hyderabad

SCHEME OF INSTRUCTION

Diploma in Data Science

SEMESTER- I

SNo	Course Code	Course Title	Hours E-Content		Scheme of Examination Max Marks		Duration (hrs)	No of Credits
THEORY			**L**	**P**	**CIE**	**SEE**	**SEE**	**Cr**
1	DDS101	Data Analytics and Data Management	**40**	**-**	**40**	**60**	**3**	**4**
2	DDS102	Mathematics for Data Science	**40**	**-**	**40**	**60**	**3**	**4**
3	DDS103	Python Programming	**40**	**-**	**40**	**60**	**3**	**4**
PRACTICALS								
4	DDS151	Data Analytics Lab	**-**	**60**	**25**	**50**	**3**	**2**
5	DDS152	Statistical Methods Lab	**-**	**60**	**25**	**50**	**3**	**2**
6	DDS153	Python Lab	**-**	**60**	**25**	**50**	**3**	**2**
			120	**180**	**195**	**330**	**-**	**18**

SEMESTER- II

SNo	Course Code	Course Title	Hours E-Content		Scheme of Examination Max Marks		Duration (hrs)	No of Credits
THEORY			**L**	**P**	**CIE**	**SEE**	**SEE**	**Cr**
1	DDS201	Machine Learning	**40**	**-**	**40**	**60**	**3**	**4**
2	DDS202	Big Data Analytics	**40**	**-**	**40**	**60**	**3**	**4**
PRACTICALS								
4	DDS251	Machine Learning Lab	**-**	**60**	**25**	**50**	**3**	**2**
5	DDS252	Big Data Analytics Lab	**-**	**60**	**25**	**50**	**3**	**2**
6	DDS253	Project	**-**	**100**	**50**	**100**	**3**	**6**
			80	**220**	**180**	**320**	**-**	**18**

DDS101

Python Programming

Credits : 4

Instruction	40 hrs e-content	Duration of SEE	3 hours
CIE	40 marks	Max Marks SEE	60 marks

Course Objectives

- Introduce the student to Python programming fundamentals
- To learn about Python programming language syntax, functions and the runtime environment
- Expose students to application development and prototyping using Python
- To learn and know the concepts of file handling, exception handling

Course Outcomes

- Able to apply fundamental problem solving techniques
- Use modern tools for creation and running a python program
- Develop applications for real time problems by applying object oriented programming concepts

Unit I

Python Data Types, Values, and Identifiers: Introduction, Assigning Values to Variables ,Multiple Assignment, Naming Rules, Understanding Reference Semantics, Python's Standard Data Types, Python Integer (Number),Sequence Types (Lists, Tuples, and Strings), Operations on Sequence Types, Lists and Tuples (Mutability versus Immutability), Dictionary, Data Type Conversion.

Statements, Expressions, and Operators: Introduction, Statements and Expressions, Operators and Operands, Arithmetic Operators, Relational Operators, Assignment Operators, Bitwise Operators, Logical Operators, Membership Operators, Identity Operators, Precedence and Associativity of Python Operators, Python Operator Associativity, Non-associative Operators (Assignment and Comparison Operators).

Unit II

Python's Flow Control Tools: Introduction, if Statement, if...else Statement, if...elif...else Statement, for Loop, range() Method, for/else Statement, while Loop, while/else Statement, Infinite Loops, break and continue Statements, break Statement, continue Statement, pass Statement.

Functions in Python: Introduction, Name, Parameters, Return Type, Standard Mathematical Functions, Time Functions, Random Functions, Reasons to Write Your Own Functions, Functions Basics, Docstring, return Statement ,Flow of Control at Function Call and Return, Recursive Functions.

Unit III

Python Modules: Introduction, import Statement , Import by Renaming , from..import Statement, Importing All Names in a Module, dir() Function, Packages, Importing Modules from Packages.

Python Classes and Objects : Introduction, Introduction to Object-Oriented Programming, Classes, Class Objects, Instance Objects, Method Object, Class and Instance Variables, Inheritance, Multiple Inheritance, Private Variables, Encapsulation, Polymorphism and Method Overriding, Object Class

Unit IV

Files and Input/Output: Introduction, File Objects, Built-in Function (open()), Built-in Methods of File Objects, Built-in Attributes, Standard Files, Command-Line Arguments.

Unit -V

Errors and Exceptions : Introduction, Errors, Exceptions, Exceptions in Python, NameError: Trying to Access an Undeclared Variable, ZeroDivisionError: Division Operation with a Numeric Zero, SyntaxError: A Syntax Error, IndexError: Requesting an Index out of Range on a Sequence Object.

Prescribed Book:

1. Allen B. Downey, ``Think Python: How to Think Like a Computer Scientist‘‘, 2nd edition, Updated for Python 3, Shroff/O‘Reilly Publishers, 2016. (http://greenteapress.com/wp/thinkpython/)

2. Python for Data Science, Dr.Mohd. Abdul Hameed, **WILEY**

3. Guido van Rossum and Fred L. Drake Jr, —An Introduction to Python – Revised and updated for Python 3.2, Network Theory Ltd., 2011.

DDS102 **MATHEMETICS FOR DATA SCIENCE**

Credits : 4

Instruction	40 hrs e-content	Duration of SEE	3 hours
CIE	40 marks	Max Marks SEE	60 marks

Statistical Methods

Course Objectives

By the end of the course, students will be able to

- Understand concepts of basic statistics and their applications.
- understand models in data analysis and applying in different types of real time situations;
- be capable of developing the models and understand the complexity of providing the accurate decisions;
- Understand basics programming concepts and able to apply

Course Outcomes

By the end of this course, the student will be Understand the concepts of P&S, its concepts and applications.

- To study the concepts as to describe the statistics and the flow of data structures with low to high level Knowledge
- To develop the statistical models using regression and interpret the inferences.
- Differentiate between regression and correlation, multiple regressions.
- Able to implement R programs and able to build functions, packages

Unit- I

Introduction to Statistics- Descriptive Statistics, Summary Statistics Basic probability theory, Bayes' Theorem, Central Limit theorem, Data Exploration & preparation, Statistical Concepts (uni-variate and bi-variate sampling, distributions, re-sampling, statistical Inference, prediction error).

Unit- II

Probability Distribution (Discrete and continuous- Bernoulli, Binomial, Poisson, Negative Binomial, Geometric distribution, Normal), Concepts of Correlation, Regression, Covariance, Outliers etc.

Unit- III

Parametric and Non parametric Tests- t-Test, ANOVA, chi-Square, U-Test, Introduction to Graphical Analysis, Using Plots(Box Plots, Scatter plot, Pie Charts, Bar charts, Line Chart), Plotting variables, Designing Special Plots, Simple Linear Regression, Multiple Regression.

Unit- IV

Linear Algebra : to be expanded by mathematics/ statistics board

Unit- V

Forecasting techniques: to be expanded by mathematics/ statistics board

Text Book

1. Fundamentals of Mathematical Statistics Paperback by S.C. Gupta & V.K. Kapoor

References

1. Probability And Statistics 2Nd Edition by Ravish R Singh and Mukul Bhatt, McGraw Hill
2. Probability And Statistical Inference 9Th Edition by Robert V Hogg and Elliot Tanis and Dale Zimmerman, PEARSON INDIA
3. Spiegelhalter, David. The Art of Statistics: How to Learn from Data. Basic Books, 2019.

OU-Diploma in Data Science With effect from the academic year 2021-2022

DDS103 DATA ANALYTICS & DATA MAHAGEMENT

Credits : 4

Instruction	40 hrs e-content	Duration of SEE	3 hours
CIE	40 marks	Max Marks SEE	60 marks

Course Objectives

By the end of the course, students will be able to

1. To learn R programming with respect to machine learning
2. To understand data handling in R
3. To gain insights into linear and logistic regression
4. To comprehend decision tree and time series
5. To study the clustering techniques

Course Outcomes

By the end of the course, the student will be able to

1. Work with R for loading and handling data
2. Explore data in R
3. Describe linear and logistic regression
4. Elaborate decision tree
5. Explain clustering techniques

Data Visualization : information visualization, data analytics life cycle, Analytics Processes and tools Analysis vs Reporting, Modern Data Analytic Tools, Visualization Techniques, Visual encodings, visualization Algorithms, Data Collection and binding, Cognitive issues, interactive visualization, visualization big data structured vs unstructured, Visual Analytics, Geo mapping, Dashboard Desighn

First two units Data Analytics
Third & fourth units Data Management with R
Fifth Unit Data Management with SQL

Text

1. Data Analytics using R by Seema Acharya. McGraw Hill education.

Suggested Reading

1. Practical Data Science with R, Nina Zumel and John Mount, Manning Shelter Island.
2. The R book, Crawley, Michael J. John Wiley & Sons, Ltd

OU-Diploma in Data Science With effect from the academic year 2021-2022

DDS151 **DATA ANALYTICS LAB**

Credits : 2

Instruction	60 hrs e-content	Duration of SEE	3 hours
CIE	25 marks	Max Marks SEE	50 marks

1. R Programming
2. SQL Programming
3. Data Pre- Processing & formatting
4. Data Visualization
5. s

DDS152 **STATISTICAL METHODS LAB**

Credits : 2

Instruction	60 hrs e-content	Duration of SEE	3 hours
CIE	25 marks	Max Marks SEE	50 marks

OU- Diploma in Data Science With effect from the academic year 2021-2022

DDS153 **MACHINE LEARNING LAB**

Credits : 2

Instruction	60 hrs e-content	Duration of SEE	3 hours
CIE	25 marks	Max Marks SEE	50 marks

Course Objectives

1. Learn R programming basics
2. Study descriptive statistics
3. Understand reading and writing datasets
4. Learn correlation, covariance and regression model
5. Comprehend multiple regression model and its use for prediction

Course Outcomes

1. Execute R programming basics
2. Implement descriptive statistics
3. Execute reading and writing datasets
4. Implement correlation, covariance and regression model
5. Execute multiple regression model and its use for predictionE

SNo	Programs
1	**R AS CALCULATOR APPLICATION** a. Using with and without R objects on console b. Using mathematical functions on console c. Write an R script, to create R objects for calculator application and save in a specified location in disk.
2	**DESCRIPTIVE STATISTICS IN R** a. Write an R script to find basic descriptive statistics using summary, str, quartile function on mtcars& cars datasets. b. Write an R script to find subset of dataset by using subset (), aggregate () functions on iris dataset.
3	**READING AND WRITING DIFFERENT TYPES OF DATASETS** a. Reading different types of data sets (.txt, .csv) from Web and disk and writing in file in specific disk location. b. Reading Excel data sheet in R. c. Reading XML dataset in R.
4	**VISUALIZATIONS a.** Find the data distributions using box and scatter plot. b. Find the outliers using plot. c. Plot the histogram, bar chart and pie chart on sample data.

5 **CORRELATION AND COVARIANCE**

a. Find the correlation matrix.

b. Plot the correlation plot on dataset and visualize giving an overview of relationships among data on iris data.

c. Analysis of covariance: variance (ANOVA), if data have categorical variables on iris data.

6 **REGRESSION MODEL** Import a data from web storage. Name the dataset and now do Logistic Regression to find out relation between variables that are affecting the admission of a student in a institute based on his or her GRE score, GPA obtained and rank of the student. Also check the model is fit or not. Require (foreign), require (MASS).

7 **MULTIPLE REGRESSION MODEL**

Apply multiple regressions, if data have a continuous Independent variable. Apply on above dataset.

8 **REGRESSION MODEL FOR PREDICTION** Apply regression Model techniques to predict the data on above dataset.

9 **CLASSIFICATION MODEL**

a. Install relevant package for classification.

b. Choose classifier for classification problem.

c. Evaluate the performance of classifier.

10 **CLUSTERING MODEL**

a. Clustering algorithms for unsupervised classification.

b. Plot the cluster data using R visualizations.

Diploma in Data Science

SEMESTER- II

SNo	Course Code	Course Title	Hours E-Content		Scheme of Examination			No of Credits
					Max Marks		Duration (hrs)	
THEORY			**L**	**P**	**CIE**	**SEE**	**SEE**	**Cr**
1	DDS201	Advanced Machine Learning / Deep learning	**40**	**-**	**40**	**60**	**3**	**4**
2	DDS202	Big Data Analytics	**40**	**-**	**40**	**60**	**3**	**4**
PRACTICALS								
4	DDS251	Lab/?	**-**	**60**	**25**	**50**	**3**	**2**
5	DDS252	lab?	**-**	**60**	**25**	**50**	**3**	**2**
6	DDS253	Project	**-**	**100**	**50**	**100**	**3**	**6**
			80	**220**	**180**	**320**	**-**	**18**

OU- Diploma in Data Science With effect from the academic year 2021-2022

DDS201 **ADVANCED MACHINE LEARNING(to be decided)**

Credits : 4

Instruction	40 hrs e-content	Duration of SEE	3 hours
CIE	40 marks	Max Marks SEE	60 marks

Course Objectives

1. Learn machine learning concepts through keras
2. Study deep learning with keras
3. Determine model evaluation methods
4. Gain insights on improving model performance
5. Learn working with images and text

Course Outcomes

1. Implement machine learning models using keras
2. Execute deep learning models
3. Evaluate models performance
4. Improve models performance
5. Create simple deep learning models for image and text data

Unit I

Introduction to Machine Learning with Keras: Introduction, Data Representation-Tables of Data, Loading Data, Data Preprocessing-Appropriate Representation of the Data , Life Cycle of Model Creation-Machine Learning Libraries, scikit-learn, Keras-Advantages of Keras, Disadvantages of Keras, More than Building Models, Model Training-Classifiers and Regression Models,Classification Tasks, Regression Tasks, Training Datasets and Test Datasets, Model Evaluation Metrics, Model Tuning-Baseline Models, Regularization, Cross-Validation

Machine Learning versus Deep Learning: Introduction-Advantages of ANNs over Traditional Machine Learning Algorithms, Advantages of Traditional Machine Learning Algorithms over ANNs, Hierarchical Data Representation, Linear Transformations-Scalars, Vectors, Matrices, and Tensors, Tensor Addition, Reshaping, Matrix Transposition, Matrix Multiplication, Introduction to Keras-Layer Types, Activation Functions, Model Fitting

Unit - II

Deep Learning with Keras: Introduction, Building Your First Neural Network-Logistic Regression to a Deep Neural Network,Activation Functions,Forward Propagation for Making Predictions, Loss Function, Backpropagation for Computing Derivatives of Loss Function Gradient Descent for Learning Parameters

Model Evaluation-Evaluating a Trained Model with Keras, Splitting Data into Training and Test Sets

Underfitting and Overfitting, Early Stopping

Evaluating Your Model with Cross-Validation Using Keras Wrappers: Introduction, Cross-Validation-Drawbacks of Splitting a Dataset Only Once, K-Fold Cross-Validation, Leave-One-Out Cross-Validation, Comparing the K-Fold and LOO Methods, Cross-Validation for Deep Learning Models-Keras Wrapper with scikit-learn, Cross-Validation with scikit-learn, Cross-Validation Iterators in scikit-learn, Model Selection with Cross-Validation-Cross-Validation for Model Evaluation versus Model Selection

Unit - III

Improving Model Accuracy:Introduction, Regularization-The Need for Regularization, Reducing Overfitting with Regularization, L1 and L2 Regularization-L1 and L2 Regularization Formulation, L1 and L2 Regularization Implementation in Keras, Dropout Regularization- Principles of Dropout Regularization, Reducing Overfitting with Dropout, Other Regularization Methods-Early Stopping, Data Augmentation, Adding Noise, Hyperparameter Tuning with scikit-learn-Grid Search with scikit-learn, Randomized Search with scikit-learn

Model Evaluation: Introduction, Accuracy- Advantages and Limitations of Accuracy, Imbalanced Datasets-Working with Imbalanced Datasets, Confusion Matrix-Metrics Computed from a Confusion Matrix

Unit IV

Computer Vision with Convolutional Neural Networks: Introduction, Computer Vision, Convolutional Neural Networks, The Architecture of a CNN-Input Image, Convolution Layer, The Pooling Layer, Flattening, Image Augmentation-Advantages of Image Augmentation

Transfer Learning and Pre-Trained Models:

Introduction, Pre-Trained Sets and Transfer Learning-Feature Extraction, Fine-Tuning a Pre-Trained Network-The ImageNet Dataset,Some Pre-Trained Networks in Keras

Unit V

Sequential Modeling with Recurrent Neural Networks: Introduction, Sequential Memory and Sequential Modeling, Recurrent Neural Networks (RNNs)- The Vanishing Gradient Problem, A Brief Explanation of the Exploding Gradient Problem, Long Short-Term Memory (LSTM)

Text

1. Matthew Moocarme, Mahla Abdolahnejad, Ritesh Bhagwat, “The Deep Learning with Keras Workshop Second Edition, An Interactive Approach to Understanding Deep Learning with Keras ”, Second edition 2020, Packt Publishing

OU-Diploma in Data Science With effect from the academic year 2021-2022

DDS202

BIG DATA ANALYTICS

Credits : 4

Instruction	40 hrs e-content	Duration of SEE	3 hours
CIE	40 marks	Max Marks SEE	60 marks

COURSE OBJECTIVES :

- Understand the Big Data Platform and its Use cases
- Provide an overview of Apache Hadoop
- Provide HDFS Concepts and Interfacing with HDFS
- Understand Map Reduce Jobs
- Exposure to NOSQL for data processing

COURSE OUTCOMES:

The students will be able to:

- Identify Big Data and its Business Implications
- List the components of Hadoop and Hadoop Eco-System
- Access and Process Data on Distributed File System
- Manage Job Execution in Hadoop Environment
- Develop Big Data Solutions using Hadoop Eco System
- Apply NOSQL for managing data

UNIT - I

Getting overview o f big data: Historical Review of Big Data, The Origin of Big Data, Debates of Big Data Implication, Historical Interpretation of Big Data, Methodology for Defining Big Data, Different Attributes of Definitions, Summary of 7 Types Definitions of Big Data, Motivations Behind the Definitions, Defining Big Data From 3Vs to 32Vs, Data Domain, Business Intelligent (BI) Domain, Statistics Domain, 32 Vs Definition and Big Data Venn Diagram.

Big Data Analytics and Machine Learning: Big Data Analytics, Machine Learning, Big Data Analytics and Cloud Computing, Hadoop, HDFS, MapReduce, Spark, and Flink, Google File System (GFS) and HDFS, MapReduce, The Origin of the Hadoop Project, Spark and Spark Stack, Flink and Other Data Process Engines, Summary of Hadoop and Its Ecosystems

UNIT –II

Resource Management in Big Data Processing Systems:

Types of Resource Management: CPU and Memory Resource Management, Storage Resource Management, 3 Network Resource Management.

Big Data Processing Systems and Platforms,:Hadoop, Dryad, Pregel, Storm, Spark, Summary.

Single-Resource Management in the Cloud**:** Desired Resource Allocation Properties, Problems for Existing Fairness Policies, Long-Term Resource Allocation Policy, Experimental Evaluation.

Multiresource Management in the Cloud: Resource Allocation Model, Multiresource Fair Sharing Issues, Reciprocal Resource Fairness, Experimental Evaluation.

Related Work on Resource Management: Resource Utilization Optimization, Power and Energy Cost Saving Optimization, Monetary Cost Optimization, Fairness Optimization.

Open Problems:SLA Guarantee for Applications, Various Computation Models and Systems, Exploiting Emerging Hardware

UNIT – III

Local Resource Consumption Shaping: A Case for MapReduce:

Motivation: Pitfalls of Fair Resource Sharing.

Local Resource Shaper: Design Philosophy, Splitter, The Interleave MapReduce Scheduler.

Evaluation:Experiments With Hadoop 1.x, Experiments With Hadoop 2.x

System Optimization for Big Data Processing:

Basic Framework of the Hadoop Ecosystem,

Parallel Computation Framework: MapReduce: Improvements of MapReduce Framework, Optimization for Task Scheduling and Load Balancing of MapReduce

Job Scheduling of Hadoop: Built-In Scheduling Algorithms of Hadoop, Improvement of the Hadoop Job Scheduling Algorithm, Improvement of the Hadoop Job Management Framework

Performance Optimization of HDFS: Small File Performance Optimization, HDFS Security Optimization

Performance Optimization of HBase**:** HBase Framework, Storage, and Application Optimization, Load Balancing of HBase, Optimization of HBase Configuration

Performance Enhancement of Hadoop System**:** Efficiency Optimization of Hadoop, Availability Optimization of Hadoop

UNIT – IV

Real-Time Analytics:

Computing Abstractions for Real-Time Analytics,

Characteristics of Real-Time Systems:Low Latency, High Availability, Horizontal Scalability.

Real-Time Processing for Big Data — Concepts and Platforms: Event, Event Processing, Event Stream Processing and Data Stream Processing, Complex Event Processing, Event Type, Event Pattern

Data Stream Processing Platforms**:** Spark, Storm, Kafka, Flume, Amazon Kinesis

Data Stream Analytics Platforms:Query-Based EPSs, Rule-Oriented EPSs, Programmatic EPSs

Data Analysis and Analytic Techniques:Data Analysis in General, Data Analysis for Stream Applications

Finance Domain Requirements and a Case Study: Real-Time Analytics in Finance Domain, Selected Scenarios, CEP Application as a Case Study.

UNIT – V:

Database Techniques for Big Data: Introduction, Background, Navigational Data Models, Relational Data Models

NoSQL Solutions for Big Data Management:

NoSQL Data Models:Key-Value Stores, Column-Based Stores Graph-Based Stores Document-Based Stores

A Case Study in Big Data Analytics: Exploring Twitter Sentiment Analysis and the weather

Big Data System Components**:** System Back-End Architecture, System Front-End Architecture, Software Stack.

Machine-Learning Methodology**:** Tweets Sentiment Analysis, Weather and Emotion Correlation Analysis

System Implementation:Home Page, Sentiment Pages, Weather Pages

Key Findings:Time Series, Analysis with Hourly Weather Data, Analysis with Daily Weather Data, DBSCAN Cluster Algorithm, Straightforward Weather Impact on Emotion

Reference Books:

1 Big Data Analytics From Strategic Planning to Enterprise Integration with Tools,Techniques,NoSQL,and Graph. By **David Loshin**

Suggested Books:

1. Big Data Principals and Paradigms by **Rajukumar Buyya,Rodrigo N. Calheiros,Amir Vahid Dastjerdi**
2. Big Data Black Book New Dream Tech publisher.
3. Big Data by Nathan Mcerz and James varren Dream Tech Publisher.

OU- Diploma in Data Science With effect from the academic year 2021-2022

DDS251 **ADVANCED MACHINE LEARNING**

Credits : 2

Instruction	60 hrs e-content	Duration of SEE	3 hours
CIE	25 marks	Max Marks SEE	50 marks

Course Objectives

1. Learn loading and cleaning data from UCI repository
2. Study creation and evaluation of simple models
3. Determine other aspects of model development
4. Gain insights on various hyperparameters of the model
5. Learn working with images and text using keras

Course Outcomes

1. Implement models by loading and cleaning data from UCI repository
2. Execute deep learning models on anaconda/google colab
3. Evaluate models performance by printing accuracy and confusion matrix
4. Improve models performance by adjusting various hyperparameters
5. Create simple deep learning models for image and text data

Experiments to be conducted either using Anaconda or Google Colab

- Loading a Dataset from the UCI Machine Learning Repository
- Appropriate Representation of the Data
- Cleaning the Data
- Creating a Simple Model
- Determining a Baseline Model
- Adding Regularization to the Model
- Performing Various Operations with Vectors, Matrices, and Tensors
- Matrix Reshaping and Transposition
- Matrix Multiplication
- Tensor Multiplication
- Creating a Logistic Regression Model Using Keras
- Building a Single-Layer Neural Network for Performing Binary Classification
- Neural Network Implementation with Keras
- Advanced Fibrosis Diagnosis with Neural Networks
- Building the Keras Wrapper with scikit-learn for a Regression Problem
- Evaluating Deep Neural Networks with Cross-Validation
- Model Evaluation Using Cross-Validation for an Advanced Fibrosis Diagnosis Classifier
- Writing User-Defined Functions to Implement Deep Learning Models with Cross-Validation
- Model Selection Using Cross-Validation for the Advanced Fibrosis Diagnosis Classifier
- Model Selection Using Cross-validation on a Traffic Volume Dataset

- Weight Regularization on an Avila Pattern Classifier
- Dropout Regularization on the Traffic Volume Dataset
- Dropout Implementation in Keras
- Hyperparameter Tuning on the Avila Pattern Classifier
- Implementing Early Stopping in Keras
- Calculating Null Accuracy on a Pacific Hurricanes Dataset
- Computing Accuracy and Null Accuracy with APS Failure for Scania Trucks Data
- Computing the Accuracy and Null Accuracy of a Neural Network When We Change the Train/Test Split
- Deriving and Computing Metrics Based on a Confusion Matrix
- Calculating the ROC Curve and AUC Score
- Building a CNN and Identifying Images of Cars and Flowers
- Amending Our Model with Multiple Layers and the Use of softmax
- Amending Our Model by Reverting to the Sigmoid Activation Function
- Changing the Optimizer from Adam to SGD
- Classifying a New Image
- Classifying a New Image
- Identifying an Image Using the VGG16 Network
- Using the VGG16 Network to Train a Deep Learning Network to Identify Images
- Classifying Images That Are Not Present in the ImageNet Database
- Fine-Tuning the VGG16 Model
- Image Classification with ResNet
- Image Classification with ResNet
- Predicting the Trend of Alphabet's Stock Price Using an LSTM with 50 Units (Neurons)
- Predicting the Trend of Amazon's Stock Price Using an LSTM with 50 Units (Neurons)
- Predicting the Trend of Alphabet's Stock Price Using an LSTM with 100 units
- Predicting Amazon's Stock Price with Added Regularization
- Predicting the Trend of Amazon's Stock Price Using an LSTM with an Increasing Number of LSTM Neurons (100 Units)

OU- Diploma in Data Science With effect from the academic year 2021-2022

DDS252 **BIG DATA ANALYTICS LAB**

Credits : 2

Instruction	60 hrs e-content	Duration of SEE	3 hours
CIE	25 marks	Max Marks SEE	50 marks

OU- Diploma in Data Science With effect from the academic year 2021-2022

DDS253 **PROJECT**

Credits : 6

Instruction	100 hrs e-content	Duration of SEE	3 hours
CIE	50 marks	Max Marks SEE	100 marks

Project has to be carried out by each student individually. Students should submit a synopsis at the end of 2nd week in consultation with the Project Guide. The synopsis should consist of definition of the problem, scope of the problem and plan of action. After completion of seven weeks students are required to present a Project Seminar on the topic covering the aspects of analysis, design and implementation of the project work.

At the end of the semester the students are required to present themselves for a University Viva-voce examination.

The guide will evaluate the project and award CIE marks.

Each student will be required to:

1. Submit one page of synopsis on the project work for display on notice board.
2. Give a 20 minutes presentation followed by 10 minutes discussion.
3. Submit a technical write-up on the project.

The project seminar presentation should include the following components of theproject:

- Problem definition and specification.
- Literature survey, familiarity with research journals.
- Broad knowledge of available techniques to solve a particular problem.
- Planning of the work, preparation of bar (activity) charts, Presentation both oral and written.