Savitribai Phule Pune University
(Formerly University of Pune)

Department of Technology
STRUCTURE OF ONE YEAR FULL TIME POST GRADUATE DIPLOMA IN ARTIFICIAL INTELLIGANCE & MACHINE LEARNING (PGD-AIML)

Intake- 60

Eligibility Criteria – Anyone who has completed under graduation in Engineering, Science, Commerce, Art, Management, Medicine, Pharma, Economics, Agriculture etc.

Semester - I

Sr. No.	Course Code	Course Name	Teaching Scheme			Credits
			L	**T**	**P**	
1.	**PGDAIML1**	Fundamentals of AI-ML & Data Science	2	0	0	2
2.	**PGDAIML2**	Data Engineering	2	1	1	4
3.	**PGDAIML3**	Python and R Programming	2	1	2	4
4.	**PGDAIML4**	Predictive Analytics (Machine Learning)	3	1	2	4
5.	**PGDAIML5**	Visual Analytics	2	1	1	4
6.	**PGDAIML6**	Research Methodology	4	0	0	4
		Total Credits				**22**

Semester - II

Sr. No.	Course Code	Course Name	Teaching Scheme			Credits
			L	T	P	
1.	**PGDAIML7**	Business Forecasting with Time series analysis	2	1	1	4
2.	**PGDAIML8**	Text Mining (Natural Language Processing)	2	1	1	4
3.	**PGDAIML9**	Deep Learning and Computer Vision	2	1	1	4
4.	**PGDAIML10**	IOT and Edge Analytics	2	1	1	4
5.	**PGDAIML11**	Major Project	0	0	0	6
		Total Credits				**22**

PGDAIML1: Fundamentals of AI – ML and Data Science

Unit – I: Data Science Foundation

Data - What is Data, Types of Data, Types of Data sources, Common data storage units ofmeasurement, Characteristics of data.

Introduction to Artificial Intelligence and Data Science - Need of AI-ML in business, Applications of DataScience in Business, Feature Engineering

Types of analytics - Descriptive, Prescriptive, Predictive, Diagnostic, Data Driven Decision Making

Unit – II: Data Science Process and Platform

Data Acquisition, Data Preparation, Data Modelling, Data visualization, Data Science in DecisionMaking, Data Security, Governance and Administration

Introduction to Rubiscape Unified Data Science Platform and Toolkits

Introduction to Agile Data Science and Rubiwise Agile Practice Guide with Rubiscape

Unit – III: Statistics in Data Science

Introduction to Statistics, Harnessing the Data, Exploratory Data Analysis, Distributions, Correlationand Regression, Introduction to Statistical workbench in Rubistudio.

Unit – IV: Decision Making Fundamentals

The Decision-Making Process, Data Exploration, Model Selection, Performance Parameters of Models,Introduction to Model lifecycle: Train-Test-Validate-Publish, Visual Data Discovery with Rubistudio

PGDAIML2: Data Engineering

Unit – I: Introduction to Data Engineering

Introduction, Definition, and Overview of Data Engineering, Introduction to Rubiflow – Data Integrator

Unit – II: Source System and Data Ingestion

Data Lake, Data Warehouse, Source Systems, Replication of Source Data, Batch Processing, Streaming

Unit – III: Data Wrangling and Validation

Data Quality of Source Systems, Statistical Validation, Rule Based Validation

Unit – IV: Data Modelling, Transformation, and Visualization

Normalization, Dimensional Modelling, Creating Tables, Schema Migration, Building the Data Warehouse, Transforming Source Data in Source Model, Building Data Products (Data Mart), Introduction to Rubiscape Metadata Manager

PGDAIML3: Python and R Programming

Unit – I: Languages for Data Science

Introduction to programming languages, Introduction to Python, Application of Python, Python Installation and Setup, Jupyter Notebook, Basic Syntax of Python, Python data types and variables.

About R Language, History of R, Overview of R and S Language, Features advantages and limitations Introduction to Rubiscape Pro-code: Embedded Juptyer Environment

Unit – II: Python Programming

Arrays, Operators, Conditional Statements and Loops, Methods/ Functions, Objects and Class, Modules and Packages, Libraries (Pandas, Numpy, Scikit), Getting Started with RubiPython

Unit – III: R Programming

Data types and Variables, Operators, Loops and Statement, Functions, Object Oriented Programming in R, Modules and Packages, Matrices, Libraries, Getting Started with RubiR

Unit – IV: Capstone Project using Rubiwise Agile Practice Guide

Capstone Project -1 (Based
on Python) Capstone Project
-2 (Based on R)

PGDAIML4: Predictive Analytics (Machine Learning)

Unit – I: Introduction to Machine Learning with Rubistudio

What is Machine Learning, Need for Machine Learning, Types of Machine Learning, Applications, andchallenges in Machine Learning.

Unit – II: Supervised Machine Learning

Regression: Gradient Descent, Linear Regression, Performance Matric Calculations, LogisticRegression, SVM etc

Classification: Introduction, Logistic Regression, Confusion Matrix, Accuracy, F-score, Sensitivity, Specificity, ROC, AUC, Introduction to Decision Tree, Entropy, Information Gain, Gini index, Pruning

Unit – III: Ensemble Learning

Introduction, Bagging, Boosting, Overfitting and Underfitting, Regularization, Normalization, Random Forest, Hyperparameter Tuning, Advanced Algorithms on Bagging and Boosting

Unit – IV: Unsupervised Machine Learning

Introduction, Clustering, Distance Based Clustering, Centroid Based Clustering, Hierarchical Clustering,Incremental Clustering.

PGDAIML5: Visual Analytics

Unit – I: Introduction to Visual Analytics with Rubisight

Introduction to Data Visualization, Role of Visualization in Data Analysis, Challenges

Related to Data Visualizations, Visual Analytics Process, Applications of Data Visualization, Introduction to Data Visualization Tools

Unit – II: Visualization Fundamentals
What are Dashboards, Dashboard Classification, Dashboard Design Process, Design Thinking in Data Visualization - Who, What, Why and How, Visualization Best Practices - For Data Representation (Timeseries, Social Data, Machine Data, Score Cards)

Unit – III: Visualization Tools and Techniques
Data Acquisition, Metadata Management, Data Preparation, and Information Modelling with Rubistudio, Data Visualization with Rubisight

Unit – IV: Dashboard Creation on Sample Industrial Data
Development of Visual Analytics Simulators with Rubisight

PGDAIML6: Research Methodology

Unit I: Research Preparation and Planning:
Objectives of research, Critical thinking. Topic selection and justification. Techniques involved in designing a questionnaire – Methods of scientific enquiry – formulation of hypotheses and testing of the same – Development of a research proposal – Theoretical and Experimental Processes.

Unit II: Research Resources:
Literature search. World Wide Web, Online databases – search tools. Literature review – Case studies, review articles and Meta-analysis. Ethical and Moral Issues in Research, Plagiarism, tools to avoid plagiarism – Intellectual Property Rights – Copy right laws – Patent rights.

Unit III: Academic Writing and Presentation:
Research proposal submission for funding agencies, Organization of proposals, Basic knowledge of research funding agencies, Research report writing, Submission of research articles for Publication to Reputed journals, Thesis writing, and Research report writing. Elements of excellent presentation.

Unit IV: Data Collection, Analysis, and Inference using Rubiscape:
Basic Statistical Distributions and their applications, Sample size determination & sampling techniques, Large Sample Tests and Small Sample Tests and their applications in research studies. Correlation and Regression Analysis-Time series analysis: Forecasting methods. Factor analysis, Cluster Analysis and Discriminant Analysis. Principles of Experimentation, Basic Experimental Designs. Precision and error analysis

PGDAIML7: Business Forecasting & Time Series Analysis
Unit – I: Introduction to Forecasting with Rubistudio
Introduction, Types of Forecasting, Applications of Forecasting,

Unit – II: Time-Series Data
Components of Time Series Data, Trend Seasonality, Cyclicity, Irregularity, ACF, PACF, Process of Forecasting

Unit – III: Forecasting Algorithms and Evaluation Methods
Moving Average, Exponential Smoothing, Simple Exponential Smoothing, Holt's exponential smoothing, Holt's Winter exponential smoothing, Automated exponential

smoothing, Introduction toevaluation metrics, Error, MAE, MAPE, RMSE, AIC, BIC

Unit – IV: Forecasting with Regression
Introduction to Regression, Types of Regression, Forecasting with Multiple Regression, Forecasting with Multiple Regression - Lurking Variable, Dummy Variable, Near Multicollinearity Variable.

PGDAIML8: Text Mining (Natural Language Processing)

Unit – I: Introduction to Natural Language Processing
Introduction to Natural Language Processing, Applications of NLP,Clustering – Centroid, Connectivity, Density, Incremental Classification – Textual Classification, AdaBoost, SVM, Naïve Bayes, Performance Matrix

Unit – II: Data Preparation for Text Analytics
Text Pre-processing, Tokenization, Case Converter, Spelling Corrector, Punctuation Remover, Frequent Word Remover, Custom Words Remover, Stemming, Lemmatization, N-Grams, Word Frequency, Advance Entity Extraction. Count Vectorization, TD-IDF.

Unit – III: Sentiment Analysis and Clustering
Sentiment Analysis: Sentiment Score, Algorithms for Sentiment Analysis Clustering – Centroid based, Connectivity based, Density based, and Incremental.

Unit – IV: Classification and Topic Modelling
Textual Classification, AdaBoost, SVM, Naïve Bayes, Performance Matrix, Unsupervised MachineLearning Models for Topic/ Entity Extraction

PGDAIML9: Deep Learning and Computer Vision

Unit – I: Introduction to Deep Learning with Rubistudio
Introduction to Deep Learning, Application of Deep Learning, Introduction to Neural Networks, Perceptron, Multilayer Perceptron, Supervised Learning, Logistic Regression, Back Propagation, Forward Propagation

Unit – II: Activation Functions and Optimization Functions
Activation Function: Introduction, Sigmoid, Tanh, ReLu, PReLu, and Softmax
Optimization Functions: Cost Function, Gradient Descent, Derivatives with computation Graph, ADAM

Unit – III: RNN and CNN
RNN: Introduction, Architecture, Applications, LSTM, and GRU
CNN: Introduction, Architecture, Stride and Padding, Feature Maps, Pulling Layer, Transfer Learning

Unit – IV: Computer Vision
Image Formation and sensing, Image Analysis, Pre-Processing, Edge Detection, Segmentation, FeatureExtraction, Pattern Analysis, Clustering, and Classification

PGDAIML10: IOT and Edge Analytics

Unit – I: Introduction to IoT with Rubithings

    Introduction to IoT, Applications of IoT, Physical and Logical Deign of IoT, IoT Enabling Technologies,Machine to Machine Communication, Cyber Physical Form, Web of Things

Unit – II: Architectures in IoT

    IoT architecture, Functional Components of IoT, Layers in IoT: Sensing layer, Network Layer and Communication Protocols, Analytics Layer, Application Layer, Functionality based IoT Protocols

Unit – III: Programming for Edge Devices and Data Protocols

    Introduction to Raspberry Pi, Arduino, Node-MCU, ESP32, Jetson, Data Protocols, Introduction to Broker, Types of Brokers, Function of Broker, Broker Architectures

Unit – IV: Data Analytics Storage and Security

    Data and Types of Data, Security and Validation of Data, Types of Data Validation, Encryption, Client Authentication: Public and Private keys, Data Handling, Data Analytics, Data Analytics Challenges

PGDAIML11: Major Project

    AI-ML solution development with Rubiwise Agile Practice Guide.

    Development of use case for functional AI-ML solution development using industrial datasets. Desirable: Student will write a research paper and publish in the conference / journal of the majorproject.