Samatrix.io

INTEGRATED MCA COURSE IN UNIVERSITY WITH SAMATRIX

2 Year MCA with Data Science.

Semester	Name of Course	Total Credits
Sem 1	Overview of Data Science, AI, Ethics and Foundation of Data Analysis	3
Sem 2	Data Analysis using Python	3
Sem 2	Probabilistic Modelling and Reasoning with Python	3
Sem 3	Machine Learning and Pattern Recognition	3
Sem 3	R Programming for Data Science and Analysis	3
Sem 4	Project	3

SEMESTER: 1	Credit: 2-0-1
Software Req: MS Office 2013/2016 Version	Hours: 4 per Week

Overview of Data Science,AI, Ethics and Foundation of Data Analysis

Objectives: The objective of this course is to teach students the concepts of current main conceptual frameworks at use in AI

UNIT – I

Introduction to AI: What is AI, Turing test, cognitive modelling approach, law of thoughts, the relational agent approach, the underlying assumptions about intelligence, techniques required to solve AI problems, level of details required to model human intelligence, successfully building an intelligent problem, history of AI

UNIT – II

Introduction to Machine Learning: What is Machine Learning, Learning from Data, History of Machine Learning, Big Data for Machine Learning, Leveraging Machine Learning, Descriptive vs Predictive Analytics, Machine Learning and Statistics, Artificial Intelligence and Machine Learning, Types of Machine Learning – Supervised, Unsupervised, Semi-supervised, Reinforcement Learning, Types of Machine Learning Algorithms, Classification vs Regression Problem, Bayesian, Clustering, Decision Tree, Dimensionality Reduction, Neural Network and Deep Learning, Training machine learning systems

UNIT – III

AI Research Trends: Research trends in machine learning, deep learning, reinforcement learning, robotics, computer vision, natural language processing, collaborative systems, algorithmic game theory, internet of things (IoT), neuromorphic computing

Applications of AI by domain: Transportation, home/service robots, healthcare, education, low- resource communities, public safety and security, employment and workplace, entertainment, finance, baking and insurance

UNIT – IV

Role of Artificial Intelligence in Society: Societal challenges AI presents, Ethical and Societal implications, policy and law for AI, fostering dialogue, sharing of best practices

Malicious Use of AI: Prevention and Mitigation: Security relevant properties of AI, Security domains and scenarios: digital security, physical security, pollical security, factors affecting the equilibrium of AI and security

Explainable AI: Introduction to explainable AI, why explainable AI, interpretability and explain ability, methods of interpretability and explain ability

Introduction to Data Analytics: Working with Formula and Functions, Introduction to Charts, Logical functions using Excel, Analyzing Data with Excel.

DATA ANALYSIS USING PYTHON

SEMESTER: 2	Credit: 2-0-1
Software: Python, NumPy, Pandas, Matplotlib, Seaborn, SciPy	No of Hours : 4 Per Week

Objectives: The objective of this course is to teach students the concepts of Python Programming Language with Libraries

UNIT – I

Python programming Basic: Python interpreter, IPython Basics, Tab completion, Introspection, %run command, magic commands, matplotlib integration, python programming, language semantics, scalar types. Control flow

Data Structure, functions, files: tuple, list, built-in sequence function, dict, set, functions, namescape, scope, local function, returning multiple values, functions are objects, lambda functions, error and exception handling, file and operation systems

UNIT – II

NumPy: Array and vectorized computation: Multidimensional array object. Creating ndarrays, arithmetic with numpy array, basic indexing and slicing, Boolean indexing, transposing array and swapping axes, universal functions, array-oriented programming with arrays, conditional logic as arrays operations, file input and output with array

Pandas: Pandas data structure, series, DataFrame, Index Object, Reindexing, dropping entities from an axis, indexing, selection and filtering, integer indexes, arithmetic and data alignment, function application and mapping, soring and ranking, correlation and covariance, unique values, values controls and membership, reading and writing data in text format

UNIT -III

Visualization with Matplotlib: Figures and subplots, colors, markers, line style, ticks, labels, legends, annotation and drawing on sublots, matplotlib configuration

Plotting with pandas and seaborn: line plots, bar plots, histogram, density plots, scatter and point plots, facet grids and categorical data

PROBABILISTIC MODELLING AND REASONING WITH PYTHON

SEMESTER: 2	Credit: 2-0-1
Software: Python, NumPy, Pandas, Matplotlib, Seaborn, SciPy	No of Hours: 4 per week

Objectives: The objective of this course is to teach students the concepts of Statistics, probability, probability distribution, and other statistical methods to solve various engineering problems

UNIT – I

Introduction to Statistics: Introduction to Statistics. Role of statistics in scientific methods, current applications of statistics.

Scientific data gathering: Sampling techniques, scientific studies, observational studies, data management.

Data description: Displaying data on a single variable (graphical methods, measure of central tendency, measure of spread), displaying relationship between two or more variables, measure of association between two or more variables.

UNIT – II

Probability Theory: Sample space and events, probability, axioms of probability, independent events, conditional probability, Bayes' theorem.

Random Variables: Discrete and continuous random variables. Probability distribution of discrete random variables, binomial distribution, poisson distribution. Probability distribution of continuous random variables, The uniform distribution, normal (gaussian) distribution

UNIT -III

Point Estimations: Methods of finding estimators, method of moments, maximum likelihood estimators, bayes estimators. Methods of evaluating estimators, mean squared error, best unbiased estimator, sufficiency and unbiasedness

Interval Estimations: Confidence interval of means and proportions, Distribution free confidence interval of percentiles

UNIT - IV

Test of Statistical Hypothesis and p-values: Tests about one mean, tests of equality of two means, test about proportions, p-values, likelihood ratio test, Bayesian tests

Univariate Statistics using Python: Mean, Mode. Median, Variance, Standard Deviation, Normal Distribution, t-distribution, interval estimation, Hypothesis Testing, Pearson correlation test, ANOVA F-test

MACHINE LEARNING AND PATTERN RECOGNITION

SEMESTER: 3	Credit: 2-0-1
Software: Python, NumPy, Pandas, Matplotlib, Seaborn, SciPy, Scikit-Learn	No of Hours: 4 per week

Objectives: The objective of this course is to teach students the basic concepts of machine learning, supervised learning, unsupervised learning, and reinforcement learning

UNIT – I

Introduction: Learning systems, real world applications of machine learning, why machine learning, variable types and terminology, function approximation

Types of machine learning: Supervised learning, unsupervised learning, reinforcement learning

Important concepts of machine learning: Parametric vs non-parametric models, the trade-off between prediction accuracy and model interpretability, the curse of dimensionality, measuring the quality of fit, bias-variance trade off, overfitting, model selection, no free lunch theorem

UNIT – II

Linear Regression: Linear regression, estimating the coefficients, accessing the accuracy of coefficient estimates, accessing the accuracy of the model

Classification: Logistic regression, estimating regression coefficients, making predictions, multiple logistic regressions, linear discriminant analysis, bayes' theorem of classification,

UNIT – III

Resampling Methods, Model Selection and Regularization: Cross-validation, leave-one-out cross-validation, k-fold cross-validation, the bootstrap, subset selection, shrinkage methods, ridge and lasso regression, dimension reduction methods, principal components regression

Tree Based Methods: Advantages and disadvantages of trees, regression Trees, classification trees, bagging, random forest, boosting

UNIT – IV

Support Vector Machine: Maximum margin classifier, classification using a separating hyperplane, the maximal margin classifier, support vector classifier, support vector machines, classification with non-linear decision boundaries, support vector machine

Unsupervised Learning: Principle component analysis, what are principal components, clustering methods, k-means clustering, hierarchical clustering,

R PROGRAMMING FOR DATA SCIENCE AND DATA ANALYSIS

SEMESTER: 3	Credit: 2-0-1
Software Req: R Studio	Hours: 4 per Week

Objectives: The objective of this course is to teach students R Programming Language, basic functions in R programming language and critical techniques

UNIT – I

Getting Started with R and R Workspace: Introducing R, R as a programming Language, the need of R, Installing R, RStudio, RStudio's user interface, console, editor, environment pane, history pane, file pane, plots pane, package pane, help and viewer pane

R Workspace, R's working directory, R Project in R Studio, absolute and relative path, Inspecting an Environment, Inspect existing Symbols, View the structure of object, Removing symbols, Modifying Global Options, Modifying warning level, Library of Packages, Getting to know a package, Installing a Package from CRAN, Updating Package from CRAN, Installing package from online repository, Package Function, Masking and name conflicts

UNIT – II

Basic Objects and Basic Expressions: Vectors, Numeric Vectors, Logical Vectors, Character Vectors, subset vectors, Named Vectors, extracting element, converting vector, Arithmetic operators, create Matrix, Naming row and columns, subsetting matrix, matrix operators, creating and subsetting an Array, Creating a List, extracting element from list, subsetting a list, setting value, creating a value of data frame, subsetting a data frame, setting values, factors, useful functions of a data frame, loading and writing data on disk, creating a function, calling a function, dynamic typing, generalizing a function. Assignment Operators, Conditional Expression, using if as expression and statement, using if with vectors, vectorized if: ifelse, using switch, using for loop, nested for loop, while loop

UNIT – III

Working with Basic Objects and Strings: Working with object function, getting data dimensions, reshaping data structures, iterating over one dimension, logical operators, logical functions, dealing with missing values, logical coercion, math function, number rounding functions, trigonometric functions, hyperbolic functions, extreme functions, finding roots, derivatives and integration, Statistical function, sampling from a vector, Working with random distributions, computing summary statistics, covariance and correlation matrix, printing string, concatenating string, transforming text, Formatting text, formatting date and time, formatting date and time to string, finding string pattern, using group to extract data, reading data

UNIT – IV

Working with Data – Visualize and Analyze Data: Reading and Writing Data, importing data using built-in-function, READR package, export a data frame to file, reading and writing Excel worksheets, reading and writing native data files, loading built-in data sets, create scatter plot, bar chart, pie chart, histogram and density plots, box plot, fitting linear model and regression tree