**TIMESPO PROFESSIONAL LEARNING**

**POST GRADUATE DIPLOMA IN DATA SCIENCE**

**COURSE STRUCTURE**

**COURSE STRUCTURE SUMMARY**

|  |  |  |
| --- | --- | --- |
| Session | Subjects | No of Hours |
| 1 | Introduction to Data science | 5 |
| 2 | R Programming | 15 |
| 3 | Data Base Management | 20 |
| 4 | Data visualization | 20 |
| 4 | Python programming | 30 |
| 5 | Applied Business statistics | 40 |
| 6 | Machine Learning | 75 |
| 6 | Deep Learning | 20 |
| 7 | Advanced Business analytics | 20 |
| 8 | Career Orientation | 5 |
| Total Hours | | 250 |
| CAPSTONE PROJECT | | 30 |

**Subject: INTRODUCTION TO DATA SCIENCE Total no. of Hours: 5**

|  |  |  |
| --- | --- | --- |
| Session | Topics to be Covered | No of Hours |
| 1 | Definition to DS, The HYPE, Job opportunities, Basic understanding on DS, Identifying the data problems, Introduction to Programming languages associated. | 3 |
| 2 | DS business application, DS relations with BI, statistics, DBMS, Visualization and machine learning, Discussion on related areas like Stochastic/Statistics, Machine Learning, Databases, Distributed Systems, Big data, Networking, Cloud Computing, Natural Language Processing, Visualization etc. | 2 |

**Subject: R Programming Language Total no. of Hours: 15**

|  |  |  |
| --- | --- | --- |
| Session | Topics to be Covered | No of Hours |
| 1 | R basics: Introduction, Installing R and R studios Math, Variables, and Strings | 2 |
| 2 | Data structures in R: Vectors and Facto, Vector operations Arrays & Matrices, Lists, Data frame | 3 |
| 3 | R programming fundamentals: Conditions and loops, Functions in R, Objects and Classes, Debugging | 2 |
| 4 | Working with data in R: Reading CSV and Excel Files, Reading text files, Writing and saving data objects to file in R | 3 |
| 5 | Strings and Dates in R: String operations in R, Regular Expressions and Dates in R | 2 |
| 6 | Cast study and exercises | 3 |

**Subject: DATA BASE MANAGEMENT Total no. of Hours: 20**

|  |  |  |
| --- | --- | --- |
| Session | Topics to be Covered | No of Hours |
| 1 | Introduction to Database Management System | 2 |
| 2 | Data Collection and challenges | 3 |
| 3 | MySQL | 10 |
| 4 | Big Data: Hadoop and Spark over view | 5 |

**Subject: Python Programming Language Total no. of Hours: 30**

|  |  |  |
| --- | --- | --- |
| Session | Topics to be Covered | No of Hours |
| 1 | Introduction to Python: Python Development Environments, Data types, Data Structure, Control Structures, Loop and iterations and Function | 3 |
| 2 | Jupyter notebooks: Setting up and using Jupyter notebooks, Using the browser efficiently | 2 |
| 3 | Numpy : Why Numpy (demonstrate less memory, fast and convenient), Arithmetic functions in Numpy, Numpy Aggregation functions, Sorting arrays, Comparison operators, Broadcasting, Array manipulations(attributes, indexing, slicing, reshaping, joining and splitting of arrays),Using array for different data problems and Vectorization | 3 |
| 4 | One dimesional data structure of Pandas: How to import the Series from a DataFrame, Sorting values, also sorting on index values, Inplace : for overwriting the existing Series ,'in' keyword in SERIES, Extracting the values based on the Index position, Maths methos on the SERIES, idxmax(index max) and idxmin(index min) method, value\_ counts on Series, Using apply method over a Seriesmap for the Series | 5 |
| 5 | Importing using Pandas: How to import a csv file in python. Checking the shape and the appearence of the dataset? How to import an excel file in python, also extracting data from different tabs?  We do have a dataset and column names for the dataset. Unfortunately they are not together, then how to align them?  How to export the dataset in form of a csv file?  What are the other readers functions of Pandas commonly used?  How to import the Series from a DataFrame | 5 |
| 6 | Handling the missing values using Pandas   * Handling missing value data using fillng as 0? * Filling NAN with different values for different columns? * Using ffill and bfill to copy the before and after value within a column * Replacing the NAN values vertically * Droping the NA values on the basis of the conditionality * When we want to replace the outliers(or other values) with NAN. * Replacing different values (column wise) not just the NAN ones, this time dictionary will come in rescue. * Giving numerical value to the categorical data for the problem solving purposes | 5 |
| 7 | Two dimensional Datastructure of Pandas   * How to set a meaningful column to be the index? * How to reindex a column from default one for a meaningful outcome. * How to check the data types of the imported file. If the type is wrong then how to transform it to the desired type? * Renaming a column? * Adding a new column to an existing dataframe * Deleting rows or columns from a dataframe * Sorting dataframes on the basis of columns * Retaining the subset of the dataframe on basis of one condition * Applying multiple conditions on the datasets * Usage of isin for multliple conditions in a single column * Using the duplicate method over a dataframe * Cheking the unique and nunique methods * The copy() method * loc and iloc * apply, map and applymap * Applying function to a Series in Pandas(value\_counts). * Finding out unique values in the dataset and replacing the column values * How can I find and remove duplicate rows? * Introduction to Groupby * Dealing with multiindex * Renaming a multindex * How to use the transpose * Swaping (swaplevel)the multindex * Stacking * Getting the random sample out of the DataFrame * concatenating * merging * reshaping (pivoting) | 5 |
| 8 | Case study 1 | 3 |
| 9 | Case study 2 | 2 |
| 10 | Case study 3 | 2 |

**Subject: Data visualization Total no. of Hours: 20**

|  |  |  |
| --- | --- | --- |
| Session | Topics to be Covered | No of Hours |
| 1 | MS EXCEL: pivot table and pivot chart | 2 |
| 2 | MS BI and Power BI | 3 |
| 2 | Tableau Introduction: Grammar of Graphics, controlling parameter | 3 |
| 3 | Visualizing unstructured information | 5 |
| 4 | Talking to visuals/Dash board | 5 |
| 5 | Art of story telling | 2 |

**Subject: Applied Business statistics Total no. of Hours: 40**

|  |  |  |
| --- | --- | --- |
| Session | Topics to be Covered | No of Hours |
| 1 | Statistics Introduction | 2 |
| 2 | Visualizing the Data | 3 |
| 3 | Describing Data Through Statistics | 3 |
| 4 | Distributions of the Sample Mean, Proportion and Sampling Techniques | 2 |
| 5 | Testing Hypotheses about Single Population Parameters | 3 |
| 6 | Analysis of Variance | 2 |
| 7 | Introduction to Regression Analysis and Correlation | 2 |
| 8 | Basic Multiple Regression Analysis | 3 |
| 9 | Advanced Multiple Regression Analysis | 3 |
| 10 | Logistic Regression | 3 |
| 11 | Count Data Regression | 2 |
| 12 | Going Beyond Linear Regression | 2 |
| 13 | Case study-1 | 3 |
| 14 | Case study-2 | 3 |
| 15 | Case study-3 | 2 |
| 16 | Case study-4 | 2 |

**Subject: Machine Learning Total no. of Hours: 75**

|  |  |  |
| --- | --- | --- |
| Session | Topics to be Covered | No of Hours |
| 1 | Introduction to Data Science and Algorithms, Examples: Difference between statistical models and ML models. What is supervised and unsupervised? \* Algorithms. \* A brief process on how to use algorithms and how to evaluate them. \* Math behind machine learning and basic concepts that people should be aware. | 2 |
| 2 | Recommendation Engines | 4 |
| 3 | Supervised Learning: OLS Regression \* Checking Assumptions of OLS Regression \* Interpretation of OLS Regression | 4 |
| 4 | Supervised Learning: Regularized Regression \* LASSO \* Ridge \* Elastic Net | 3 |
| 5 | Supervised Learning: Logistic Regression \* Theory and Math Behind Logistic \* Application using Programs \* Use case Interpretation | 4 |
| 6 | Supervised Learning: Decision Tree \* Decision Tree for Classification \* Math Behind Decision Tree \* Programming on Decision Tree \* Bias-Variance Trade-off \* Model Selection in Decision Tree | 4 |
| 7 | Supervised Learning: Ensemble Methods \* Math Behind Ensemble Models \* Bagging Models \* Boosting Models \* Stacking Models \* Model Evaluation \* Hyper-parameter Training | 4 |
| 8 | Supervised Learning: Support Vector Machine Theory \* SVM for Classification \* SVM for Regression \* Model Interpretation | 4 |
| 9 | Supervised Learning: Stochastic Gradient Descent Algorithm \* SGD for regression \* SGD for Classification \* SGD for Sparse data | 3 |
| 10 | Supervised Learning: \* SGD for regression \* SGD for Classification \* SGD for Sparse data | 3 |
| 11 | Supervised Learning: Nearest Neighbors \* Knn for classification \* Knn for Regression \* model evaluation | 3 |
| 12 | Supervised Learning: Naïve Bayes \* Gaussian NB \* Multinomial NB \* Bernoulli NB | 2 |
| 13 | Supervised Learning: Neural Network Basics \* NN for Classification \* NN for Regression \* NN for Autoencoder | 2 |
| 14 | Un supervised learning: Clustering, PCA, Classification, Extreme gradient boosting, Feature engineering,cross validation, Modelling approximation, Market basket reasearch | 7 |
| 15 | Model Selection and Evaluation: Cross-validation and evaluating estimator performance \* GRID search \* Random Search \* Model Scoring Parameter \* Model Persistence | 4 |
| 16 | Feature Creation and Seelction: \* Standardization, or mean removal and variance scaling \* Scaling features to a range \* Scaling sparse data \* Scaling data with outliers \* Centering kernel matrices \* Non-linear transformation \* Normalization \* Binarization \* Feature binarization \* Encoding categorical features \* Imputation of missing values \* Generating polynomial features \* Custom transformers | 4 |
| 17 | NLP: Installation and Get Started \* NLP Concept Introduction \* NLP- Text Classification \* NLP- Sentiment Analysis | 5 |
| 18 | Bigdata | 5 |
| 19 | SPARK/PySpark: Spark Eco-System \* Getting Started with Spark \* Installation and Set Up \* Spark and R-Studio \* Running Algorithms \* Python and Spark \* PySpark- data import \* data management \* feature creation \* model building \* model storing | 5 |
| 20 | CASE STUDY | 3 |

**Subject: Deep Learning Total no. of Hours: 20**

|  |  |  |
| --- | --- | --- |
| Session | Topics to be Covered | No of Hours |
| 1 | Deep Learning Theory and Tools Introduction | 2 |
| 2 | What is deep learning, Deep Learning Architectures, CNN, DNN, RNN, what are the tools necessary for DL | 3 |
| 3 | Introduction to Neural Network Basics, Basic concepts like, Neuron, MLP, Activation functions, Optimization Algorithm | 5 |
| 4 | concept of back propagation, single layer neural network, multi-layer neural network, types of neural network architecture | 5 |
| 5 | NVIDEA TRAINING | 5 |

**Subject: Analytics Applications in Business Functions - case studies No of Hours: 20**

|  |  |  |
| --- | --- | --- |
| Session | Topics to be Covered | No of Hours |
| 1 | Human Resource Analytics | 2 |
| 2 | Marketing Analytics | 3 |
| 3 | Health Care Analytics | 5 |
| 4 | Retail Analytics | 5 |
| 5 | BFSI | 5 |
| 6 | TELECOM | 5 |