

# Center of Excellence in Artificial Intelligence and Innovation

## Online Certification Courses

### LEVEL-A : 3 Month Certificate course in Data Engineering and Business Intelligence

#### Course A1: (Data Engineering) Introduction to Database Management Systems

**Unit 1: Introduction to Database System:** Database System Applications, Purpose of Database, Database Architectures, Database Properties, Database Languages, Views of Data, Instances and Schema.

**Unit 2: Database Design:** Data Modelling: Relational Model, Data Modelling: ER Model, Attributes and their types, Relationships, Cardinalities, Extended ER Diagram, Specialization and Generalization, Aggregation and Attribute Inheritance.

**Unit 3: Relational Databases:** Relational Model, Keys concepts, Integrity Constraints, Introduction to SQL: Data Types and Literals, DDL, DML statements, Views: Creating, Dropping, Updating using Views, Indexes, Handling Nulls. Basic Filtering and Advance Filtering with SQL, Wildcards in SQL, Sorting and Math Operations, Built-in Database Functions: Numeric, Set Operations.

**Unit 4: Advanced SQL:** Aggregate Functions, Grouping Data with SQL, Sorting and Ordering Data with SQL, SQL Join Operations, Sub Queries, SQL Join Operations, Sub Queries, working with text and Strings, working with Date and Time Strings, Windows functions, Accessing database with Python, Analysing Data with Python, Working with Real world Dataset.

#### **Suggested List of Assignments:**

1. Create a schema using SQL DDL Commands and perform basic SQL DML commands on that schema .
2. Write SQL queries using clauses such as Group by, Order by, Having etc. for statements given.
3. Write SQL queries using aggregate functions for given statements on a given schema.
4. Write Nested Subqueries in SQL using Set membership and Set Comparison Commands such for a given statement on a given set of relations.
5. Write SQL queries using join (left, right and full outer join) operations for the statements given.
6. Write SQL queries using Date Functions, String Functions and Math Functions for statements given.
7. Write SQL queries to create a view, drop a view and select from a view for a given statement on a given schema.
8. Create the schema and specify constraints on the given relations using statements given.

9. Demonstrate database connectivity in Python.

10. Draw an E-R diagram and convert entities and relationships to a relation table for a given scenario. Normalize the database up to appropriate normal form.

### Course A2: Foundations of Statistics and Linear Algebra

**Unit 1: Introduction to Statistical Thinking:** Population and Sample, Parameters and statistics, variables and organization of data, Visualizing Data, Frequency Distribution, Histogram, visualizing qualitative and quantitative data.

**Unit 2: Descriptive Statistics:** Measures of Central Tendency, measures of variability, Distributions, Normal Distribution, standard normal distribution, Poisson Distribution and Binomial Distribution, Sampling Distribution, Estimation, Confidence Interval.

**Unit 3: Inferential Statistics:** Hypothesis Testing, Type-I and Type-II errors, Steps in hypothesis testing, one and two-tailed tests, T-Test, one sample t-test, two sample t-test, independent sample t-test, ANOVA, within and between subjects factors, one factor anova, Correlation and Regression, Chi-Square Test, contingency table.

**Unit 4: Linear Algebra:** Intro to Linear Algebra, Scalar, Vectors, Matrix and Tensors, Linear Dependence and Span, Norms, SVD, Eigenvalues and Eigenvectors

### Course A3: Data Visualization with Tableau

**Unit 1: Data Connections:** Connect to Tableau Server, Describe connection options, Connect to different data source types, from single and multiple databases, Prepare Data for Analysis: Blending Metadata Grid, Pivot, Union, Data Interpreter, Explain data extract formats and capabilities, Create extracts with multiple tables Explain performance considerations between blends, joins, and cross-database joins, Use Automatic & Custom Split.

**Unit 2: Organizing & Simplifying Data:** Understand how to: Filter data, Sort data, Build groups, Build hierarchies, Build sets.

**Unit 3: Field & Chart Types:** Difference between measures and dimensions, difference between discrete and continuous fields, Tableau-generated fields, Understand how and when to build: Histograms, Heat maps, Tree maps, Bullet graphs, Combined axis charts, Dual axis charts, Scatter plots, Cross tabs, Bar in bar charts, Box plots, Use titles, captions and tooltips effectively, Edit axes, Use mark labels and annotations.

**Unit 4: Calculations:** Manipulate string and date calculations, Create quick table calculations, Use level of detail (LOD) expressions, Explain different types of LOD expressions, Use Ad-hoc calculations, Work with aggregation options, Build logic statements, Build arithmetic calculations, Build grand totals and sub-totals, Use calculations in join clauses.

**Unit 5: Mapping:** Navigate maps, including: Pan & Zoom, Filtering, Map layering, Custom territories, Geographic search, Modify locations within Tableau, Import and manage custom geocoding, Use a background image map, Connect to spatial files.

**Unit 6: Analytics:** Reference Lines, Reference Bands, Trend Lines, Trend Model, Forecasting, Drag & Drop Analytics, Box Plot, Reference distributions, Statistical summary card, Instant Analytics, Data Highlighter.

**Unit 7: Dashboards:** Build dashboards and stories, Create dashboard actions, Design dashboards for viewing on devices, Utilize visual best practices for viewing on devices, Describe publishing & sharing options.

**Suggested List of Assignments:**

1. Principles of Visualizations using examples from published sources, Install Tableau, Analyze a data and make inferences (from a pre-built report/viz from sources)
2. Connecting to Excel Data Sets, Cleaning Data Sets, Working with Joins, Unions, Relationships, Blending to see how data will differ when we do each one of these method to get the data into tableau
3. Using IPL, Sample SuperStore Data creating and inference charts
4. Using Covid-19, Sample SuperStore Data creating and inference charts
5. Create Basic Dashboards to understand what are containers, how to use action filters and add images, text boxes and home button, Back button
6. Examples on How to use Extract and Data Source Filters, importance of Context filters, Creating Top N using Filters Option
7. Examples on How to use Date Filters, Sets various options in Sets
8. Create Advanced charts using sets, Groups, Hierarchies
9. What is a KPI, how to get KPI Colors Using Parameters, in a chart Change Reference Line dynamically using Parameter
10. Sheet Swap, Column Swap using parameters
11. Working with integer and Text calculations
12. Date Calculations
13. LOD
14. Format a dashboard, add navigation Buttons and global Filters, Create a Story View
15. Publishing dashboards to Server, Adding Users, Enable Subscriptions

=====

## **LEVEL-B : 3 Month Certificate course in Machine Learning and Deep Learning**

### Course B1: Basics of Python programming

**Unit 1: Basic of Python Programming:** Introduction to Python and Jupyter Notebooks, Data Types and Typecasting, ipython magic commands and shell commands Control Structures, User Input, Exception Handling, Operators in python.

**Unit 2: Data Structures and Functions:** Strings, Indexing and Slicing, String Operators, Useful methods for string manipulation and processing, Escape Sequences, String Formatting. List, inbuilt functions with list, in and not in operators, Tuples, difference between tuples and list. Dictionary, dictionary methods, creating dictionary from text, sorting dictionary, inverting dictionary, writing functions in Python, function design recipe, calling functions within another functions.

**Unit 3: List and Dictionary Comprehensions:** List comprehension, List comprehension using for, nested for loops, if, if-else. Dictionary comprehension, Useful Functions: Lambda operator for defining anonymous functions, Map, filter and reduce functions. Enumerate and zip functions.

**Unit 4: Numpy:** Creating ndarray, Array indexing and slicing, Integer Indexing, Boolean Indexing, Broadcasting, Arithmetic and Statistical operations, Any and all conditionals, Merging and splitting, Important built-in methods (np.zero, np.ones, np.where, np.unique etc .....).

**Unit 5: Data Manipulation and Analysis using Panda:** Creating series, Data manipulation with series, Creating Dataframes, Data manipulation with data frames, Data Cleaning, Data Analyzing, Data Visualization, Exploratory Data Analysis using pandas, importing data in python.

**Unit 6: Data Visualization:** Matplotlib architecture, Basic plots (scatter, line, histogram, boxplots, barcharts etc.....), Subplots, Advance plots (heatmap, pairplot, distplot, regplot, Implot etc .....), interactive plots using plotly, geoplots, bubble plots, violin plots etc..

### **Suggested List of Assignments:**

1. Getting started with Jupyter Notebook, Implementing basic programming concepts. (2 hrs)
2. Implementing Data structures and their functionalities. (4 hrs)
3. Applying list comprehension and use lambda operators with map, filter and reduce functions to solve a scenario. (2 hrs)
4. Analyzing Image Data using Numpy. (4 hrs)
5. Perform Data Manipulation on a dataset using Panda. (4 hrs)
6. Perform EDA (Exploratory Data Analysis) on a dataset using Panda. (6 hrs)
7. Perform Data visualization using static and interactive plots on a given dataset. (4 hrs)
8. Connect to a RDBMS and execute SQL queries on a dataset. (4 hrs)

## Course B2: Machine Learning Fundamentals

**Unit 1: Introduction to Machine Learning:** Machine Learning applications, Types of ML: supervised ML, unsupervised ML, semi-supervised ML, reinforcement ML, batch and online ML systems, instance based and model based ML systems, Challenges in ML, Overfitting and underfitting.

**Unit 2: Supervised Learning:** Linear Regression, Best fit regression line using least square and gradient descent method, performance metrics for regression, Logistic Regression, Performance metrics for classification, Thresholding, ROC and AUC curve, Decision Tree, basic tree terminology, pruning trees, hyper parameter tuning, Decision trees for Regression, Decision Tree for Classification, Entropy, Information Gain and Gini Index.

**Unit 3: Unsupervised Learning:** Content based, collaborative and hybrid filtering techniques, Clustering, Hierarchical Clustering, distance metrics, inter-cluster and intra-cluster distances, dendrogram, kmeans clustering, kmeans++ clustering, elbow method for determining number of clusters.

### **Unit 4: Ensemble Methods:**

Bias-Variance Tradeoff, Cross Validation and Bootstrapping, Ensemble Methods (Bagging), Ensemble Methods (Random Forest), Ensemble Methods (Boosting, XGBoost based binary classification problems.

### **Unit 5: Text Analytics:**

Handling text data, steps to pre-process text data for modeling, Naïve Bayes classification algorithm, sentiment analysis.

### **Suggested List of Assignments:**

1. Getting started with scikit learn library and understand various modules for data preprocessing, feature engineering and data modelling. (4 hrs)
2. Implementing Linear Regression to solve a regression problem. (4 hrs)
3. Implementing Logistic Regression to solve a classification problem. (4 hrs)
4. Implementing Decision Trees for Regression problem. (4 hrs)
5. Implementing Decision Trees for Classification problems. (4 hrs)
6. Solving an End to End ML case study for regression problem. (4 hrs)
7. Solving an End to End ML case study for classification problem. (6 hrs)

## Course B3: Deep Learning

**Unit 1: Introduction to AI:** AI vs ML vs DL, what is Deep Learning, Biological Neurons to Artificial Neurons, Perceptron: Training a perceptron, Activation Functions and its need, Gradient Descent and Loss Functions, Derivatives and Computation graph.

**Unit 2: Multilayer Perceptron:** Logistics Regression as NN, Multi Layer Perceptron: Naming conventions, Matrix dimensions and Representation as NN, MLP: Forward Propagation, MLP: Backward Propagation, Vectorizing on training samples and Random Initialization.

**Unit 3: Deep Neural Network:** Deep neural network notations, Matrix dimension and Parameters and Hyperparameters, Forward Propagation and Backward Propagation in Deep NN.

**Unit 4: Introduction to Keras:** Keras features, Keras architecture, Anatomy of neural network, building blocks of deep learning, Keras Sequential model, keras functional API, loss functions and optimizers and compilation of models, Training of model, evaluating and perform predictions on model, Implementing Regression MLP, Implementing Classification MLP.

**Unit 5: Training Deep Neural Networks:** Vanishing /exploding gradients problem, Batch Normalization and its implementation, Reusing pretrained models: Transfer learning with keras.

**Unit 6: Faster Optimizers:** Momentum, NAG, RMSProp, Adam. Learning rate Scheduling, Avoid overfitting using Regularization.

### **Suggested List of Assignments:**

1. Handwriting recognition using CNN on MNIST database
2. Face recognition using DeepFace library
3. Car detection using YOLO for autonomous driving

=====

## **LEVEL-C : 6 month Certificate course in Advanced AI + internship/ project**

### Course C1: R Programming

#### **Contents:**

##### **Unit 1: Introduction to R Programming:**

Getting started with RStudio, workspace in R, packages in R, installing packages in R, Built-in Datasets in R, defining variables, , operators in constants and strings R

##### **Unit 2: Data Structures and Functions:**

Vectors, operations on vectors, generating vectors using seq(), Repeating Vector Constants with rep(), NA and NULL values, arrays and matrices, matrix indexing, matrix operations, List, list functions, data frames, extracting sub-data frames, using the rbind() and cbind() functions and alternatives, factors, common functions used with factors, split() and by()

##### **Unit 3: R Programming Structures:**

Conditions and loops, control functions, apply family, default values for arguments, lazy evaluation in functions, objects and classes, debugging

##### **Unit 4: Data Manipulation in R:**

Importing and exporting data in R, Reading text files, writing and saving data objects to file in R, data Manipulation using dplyr package, exploratory data analysis using R, data cleaning, data cleaning, ggplot2 package, statistics with R, machine learning implementation in R

#### **Books:**

1. Norman Matloff, "The art of R programming: A tour of statistical software design", No Starch Press, ISBN: 9781593273842
2. Andrie De Vries and Joris Meys, "R for Dummies", John Wiley & Sons, ISBN: 9781119962847
3. Sandip Rakshit, "R Programming for Beginners", McGraw Hill India, Edition 1, 2017, ISBN: 9789352604555
4. Hadley Wickham, Garrett Golemund, "R for data science : Import, Tidy, Transform, Visualize, And Model Data", O'Reilly, ISBN: 9781491910399
5. Andy Field, "Discovering Statistics Using R", Sage publications, ISBN: 9781446200469

## Course C2: Time Series Data Analysis

### **Contents:**

**Unit :1 Introduction to Time Series:** Trend Effect, Seasonal effect, Cyclic effect, Components of Time series, Additive Model, Naive/Snaive, Multiplicative Model, Autocorrelation and white noise.

**Unit 2: Basic Forecasting Models:** Naive Approach, Average Method, Simple exponential smoothing, Holt's Linear Trend Model, Holt's Winter Model.

**Unit 3: Single and Multivariate Time Series:** Stationary, Differencing, Auto-regressive (AR), Moving Average (MA), Auto-regressive moving average (ARMA), Auto-regressive integrated moving average (ARIMA), Seasonal ARIMA (SARIMA), Granger's Causality Test, Vector Autoregression (VAR).

**Unit 4: Time Series using Deep Learning:** Prepare Time Series Data for CNNs and LSTMs, MLPs for Time Series Forecasting, CNNs for Time Series Forecasting, Simple LSTM, Stacked RNN and Bidirectional LSTMs, RNN.

### **Books:**

1. Introduction to Time Series and Forecasting, 2nd Edition; Peter J. Brockwell, Richard A. Davis; ISBN 0-387-95351-5; Springer-Verlag New York, Inc.
2. Practical Time Series Analysis, 1st Edition; Aileen Nielsen; ISBN: 978-1-492-04165-8; O'Reilly Media, Inc.
3. Rob J Hyndman and George Athanasopoulos, Forecasting: Principles and Practice, OTexts, ISBN-0987507109.

## Course C3: Advanced Machine Learning

### **Contents:**

#### **Unit 1: Regularization:**

Subset Selection Methods, Regularization, ridge regularization and lasso regularization, variable selection lasso regularization. Applying cross validation for learning rate.

**Unit 2: Dimensionality Reduction:** Dimensionality reduction using PCA,

**Unit 3: Classification Algorithms:** Support Vector Machine, Maximal Margin Classifier, Support Vector Classifier, Kernel Trick in SVM, Naive Bayes, Naive Bayes, KNN

### **Books:**

1. Giuseppe Bonaccorso, "Machine Learning Algorithms", Packt Publishing Limited, ISBN10: 1785889621, ISBN-13: 978-1785889622 2.
2. Josh Patterson, Adam Gibson, "Deep Learning: A Practitioners Approach", O'REILLY, SPD, ISBN: 978-93-5213-604-9, 2017 Edition 1st .
3. Tom Mitchell "Machine Learning" McGraw Hill Publication, ISBN :0070428077 9780070428072



## Course C4: Natural language Processing and Text Analytics

### **Contents:**

**Unit 1: Introduction to NLP:** Basics of Natural Language Processing, NLP applications, Language syntax and structure, Language Semantics, Installing NLTK package.

**Unit 2: Text Preprocessing:** Text tokenization, text normalization, removing stopwords and special characters, expanding contractions, Stemming, Lemmatization. POS Tagging

**Unit 3: Text Classification:** Feature Extraction from text, Bag of words model, tf – idf model, multinomial naïve bayes, support vector machine. Topic modeling.

**Unit 4: Text Clustering:** Information retrieval, Similarity measures, Analyzing term similarity, analyzing document similarity, Kmeans clustering, Agglomerative Hierarchical clustering.

**Unit 5: Semantic and Sentiment Analysis:** Wordnets, Word sense disambiguation, Named Entity Recognition, Sentiment Analysis.

### **Books:**

1. Dipanjan Sarkar, Text analytics with Python, Apress, ISBN-13 (pbk): 978-1-4842-2387-1
2. Steven Bird, Ewan Klein, and Edward Loper, Natural Language Processing with Python, O'Reilly Media, Inc. ISBN: 978-0-596-51649-9
3. Jacob Perkins, Python 3 Text Processing with NLTK 3 Cookbook, Packt Publishing Ltd., ISBN 978-1-78216-785-3

## Course C5: Industry project

