QUICKSTEP COMPUTER CENTER

National Accreditation Board of Education Training.
(NABET)- Quality council of India) An ISO 9001:2008

DATA SCIENCE - AI, ML, DL

1. Introduction to Data Science

    a. **What is data science?**
- ✓ How is data science different from BI and Reporting?

    b. **What is difference between AI, Data Science, Machine Learning, Deep Learning**

    c. **Job Land scape and Preparation Time**

    d. **Who are data scientists?**
- ✓ What skillsets are required?

    e. **What is day to day job of Data Scientist**
- ✓ What kind of projects they work on?

    f. **End to End Data Science Project Life Cycle**

    g. **Data Science roles – functions, pay across domains, experience**

2. Business Statistics

    a. **Data types**
- ✓ Continuous variables
- ✓ Ordinal Variables
- ✓ Categorical variables
- ✓ Time Series
- ✓ Miscellaneous
- ✓ Common Data Science Terminology

    b. **Descriptive statistics**
- ✓ Basics concepts of probability
- ✓ Frequentist versus Bayesian Probability
- ✓ Axioms of probability theory,
- ✓ Permutations and combinations
- ✓ Conditional and marginal probability
- ✓ Joint Probability
- ✓ Bayes Theorem
- ✓ Probability Mass Function and Probability Density Function
- ✓ Cumulative Mass Function and Cumulative Density Function

    c. **Central Tendencies**
- ✓ Mean
- ✓ Median
- ✓ Mode
- ✓ Spread
- ✓ Variance
- ✓ Standard Deviation
- ✓ Effects on central tendencies after transformations
- ✓ Quartile Analysis
- ✓ Implementation of central tendencies using python
- ✓ Box Plots for outlier identification

✓ **Drawing Box plots using python**

d. **Sampling**
- ✓ **Need for Sampling?**
- ✓ **Different types of Sampling**
- ✓ **Simple random sampling**
- ✓ **Systematic sampling**
- ✓ **Stratified Sampling**
- ✓ **Implementation of sampling techniques using python**

e. **Data distributions**
- ✓ **Normal Distribution**
- ✓ **Binomial Distribution**
- ✓ **Binomial Approximated to Normal**
- ✓ **Implementation of distributions using python**

f. **Inferential statistics**
- ✓ **Why inferential statistics?**
- ✓ **Z score calculation**
- ✓ **Defining p value and implementations using python**
- ✓ **Inferring from sample to population**
- ✓ **Sampling distribution of sample means**

g. **Hypothesis testing**
- ✓ **Confidence Interval**
- ✓ **Testing the hypothesis**
- ✓ **Type I error**
- ✓ **Type II error**
- ✓ **Null and alternate hypothesis**
- ✓ **Reject or acceptance criterion**

3. Introduction to R

a. **A Primer to R programming**
b. **What is R? Similarities to OOP and SQL**
c. **Types of objects in R – lists, matrices, arrays, data.frames etc.**
d. **Creating new variables or updating existing variables**
e. **If statements and conditional loops - For, while etc.**
f. **String manipulations**
g. **Sub setting data from matrices and data.frames**
h. **Casting and melting data to long and wide format.**
i. **Merging datasets**

4. Python for Data Science

a. **Understanding the reason of Python's popularity**
b. **Basics of Python: Operations, loops, functions, dictionaries**
c. **Numpy – creating arrays, reading, writing, manipulation techniques**
d. **Ground-up for Deep-Learning**

5. Exploratory Data Analysis with Python

a. Getting to understand structure of Matplotlib
b. Configuring grid, ticks.
c. text, color map, markers, widths with Matplotlib
d. configuring axes, grid,
e. hist, scatterplots
f. bar charts
g. multiple plots
h. 3D plots
i. Correlation matrix plotting

6. Data Munging with Python
a. Introduction to pandas
b. Data loading with Pandas
c. Data types with python
d. Descriptive Statistics with Pandas
e. Quartile analysis with Pandas
f. Sort, Merge, join with Pandas
g. Indexing and Slicing with pandas
h. Pivot table, Aggregate and cross tab with pandas
i. Apply function for parallel processing with Python
j. Cleaning Data with python
k. Determining correlation
l. Handling missing values
m. Plotting with Pandas
n. Time series with Pandas

7. Introduction to Artificial Intelligence
a. Dealing Prediction problem
b. Forecasting for industry
c. Optimization in logistics
d. Segmentation in customer analytics
e. Supervised learning
f. Unsupervised Learning
g. Optimization
h. Types of AI : Statistical Modelling, Machine Learning, Deep Learning, Optimization, Natural Language Processing, Computer vision, Speech Processing, Robotics

8. Artificial Intelligence I - Statistical Modelling
a. Linear Regression
   ✓ Assumptions
   ✓ Model development and interpretation
   ✓ Sum of least squares

- ✓ **Model validation – tests to validate assumptions**
- ✓ **Multiple linear regression**
- ✓ **Disadvantages of linear models**

b. Logistic Regression
- ✓ **Need for logistic regression**
- ✓ **Logit link function**
- ✓ **Maximum likelihood estimation**
- ✓ **Model development and interpretation**
- ✓ **Confusion Matrix – error measurement**
- ✓ **ROC curve**
- ✓ **Measuring sensitivity and specificity**
- ✓ **Advantages and disadvantages of logistic regression models**

c. Time series analysis - Forecasting
1. **Simple moving averages**
2. **Exponential smoothing**
3. **Time series decomposition**
4. **ARIMA**

d. Model validation and deployment
- ✓ **RMSE – Root Mean squared error**
- ✓ **MAPE – Mean Average Percentage Error**
- ✓ **Confusion matrix and Misclassification rate**
- ✓ **Area under the curve (AUC) , ROC curve**

9. Artificial Intelligence II – Machine Learning

a. Supervised Learning
- ✓ **Decision trees and Random Forest**
1. **C5.0**
2. **Classification and Regression trees(CART)**
3. **Process of tree building**
4. **Entropy and Gini Index**
5. **Problem of over fitting**
6. **Pruning a tree back**
7. **Trees for Prediction (Linear) – example**
8. **Tress for classification models – example**
9. **Advantages of tree-based models?**
- ✓ **Association Rule Mining**
1. **Rules generation from decision trees,**
2. **Apriori algorithm**
3. **Support, confidence and lift measures**
- ✓ **Support Vector Machines**
1. **Linear learning machines**
2. **SVM case for linearly separable data**
3. **Kernel space**
- ✓ **Neural Networks**

1. Motivation for Neural Networks
2. Perceptron and Single Layer Neural Network
3. Back Propagation algorithm
4. Feed Forward Neural Net
5. Sigmoid parameters
6. Weights initialization,
7. Decay of weights
8. Learning rate
9. Momentum
   - ✓ Ensemble Techniques
1. Bagging
2. Boosting
3. Stacking
4. Gradient Boosting Machines

b. **Unsupervised Learning**

   - ✓ Clustering Techniques
1. Hierarchical clustering
2. K-Means clustering
3. Distance measures
4. Applications of cluster analysis – Customer Segmentation
   - ✓ Collaborative Filtering, PCA

10. **Artificial Intelligence III – Natural Language Processing**
    a. NLP I - Text Preprocessing
       - ✓ Tokenization
       - ✓ Stemming
       - ✓ Lemmatization
    b. **NLP II – Text Modelling**
       - ✓ POS tagging
       - ✓ TFIDF and classification

11. **Artificial Intelligence IV - Deep Learning**
    a. ReLU
    b. Sigmoid, Depth vs Width tradeoffs
    c. Convolutional networks
    d. Concepts of filters
    e. Sliding
    f. Pooling and Padding
    g. Comparison between DL and ML performances over the MNIST dataset

12. **Practical use cases of AI and best practices in AI**
    a. Business problem to an analytical problem
    b. Guidelines in model development

13.   Big Data, Azure for AI, Data Science applications

   a. Big data and analytics?
      ✓   Leverage Big data platforms for Data Science
   b. Introduction to evolving tools
      ✓   Machine learning with Spark
   c. Creation of R-Server clusters
   d. Computation of Big-Data ML algorithms over the Azure cloud

14.   Analytical Visualisation with Tableau

   a. Why is it important for Data-Analyst
   b. Tableau workbook walkthrough
   c. Instruction of creation of your own workbooks
   d. Demo of few more workbooks