DVS DATA SCIENCE with PYTHON Content

Introduction: **This course providestechniques for datacleaning, visualizingthe data, predictive modeling, machine learning and deep learning by using python, tensor flow and keras**

Duration: **2.5 Months long instructor- led pure class room training**

Features:

- Main focus on hands-on training
- Guidance in Resume Preparation
- Contains real world business problems and examples.
- Rich material and handouts for reference
- FAQs, Real Time Scenarios, Real Time Interview Questions

After Completion of this training:

- Build predictive models using linear, logistic regression and decision trees
- Build machine learning models using Decision trees Neural nets, SVM and Random forest
- Build reinforcement models using the Markov chain models
- Build deep learning models using CNN, RNN and DNN

Statstics concepts:

1. Statstics terminology

- Basics Statstics:
- Descriptive statstics and inferential statstics
- Measure of central tendency -Mean, Median and Mode
- Measure of Dispersion-Range, Variance, standard deviation and coefficient of variation
- Frequency distribution
- Introduction to Probablity
- Practice Session & Assignments

2. Python Introduction

- What is Python & History?
- Installing Python & Python Environment
- Basic commands in Python
- Data Types & Operators
- Data Structures in python- List, tuples, dictionary and sets
- Python packages – math, Numpy, pandas, Matplotlib, seaborn, scikit learn Loops- for loop do while
- User Defined Functions
- My First Python Program
- Conditional statements- IF, IF-Else, nested IF, nested IF-Else conditional
- Lambda, map, filter and reduce
- LAB Session

3. Data Handling and Data manipulation

- Data importing
- Working with datasets
- Manipulating the data sets
- Subset the data
- Sort the data
- Creating new variables
- Bins creation
- Identifying & removing duplicates
- Exporting the datasets into external files
- Data Merging
- Pivot table analysis
- Data visualization through matplotlib, seaborn
- Histogram
- Bar Plot
- Pie Chart
- Scatter Matrix Pandas
- Scatter matrix Violin
- Plots
- Line Graphs

4. Python Basic Statistics

- Taking a random sample from data
- Descriptive statistics
- Central Tendency
- Variance
- Quartiles
- Percentiles
- Box Plots
- Graphs
- Visualization case study with poke man data

5. Pyspark

- Install Spark with Python
- Loading the data into spark
- RDD
- RDD to Data frame
- Introduction to Spark-MLIB

6. Probability distribution

Discreate distribution:

- Bi-nominal distribution
- Poisson distribution
- Hyper geometric probability distribution
- Multinomial distribution

Continuous distribution:

- Normal distribution
- Uniform normal distribution
- T-student distribution
- Exponential distribution
- Chi- squaredistribution
- F- distribution
- Box cox distribution

7. Sampling Techniques:

1. **Random sampling:**
    - Sample with replacement
    - Sample without replacement
    - Training, testing and hold out dataset
2. **Stratified sampling**
3. **Sequential or systematical sampling**
4. **Clustering sampling techniques**

8. Hypothesis testing:

- What is Hypothesis testing
- Need of hypothesis testing
- Null hypothesistesting
- Alternative hypothesis testing
- Use case to solve the hypothesis testing

9. Data preprocessing

- Data sanity checks
- Anomalies detection
- Missing Value detections & treatments

- **Outliers detection and outlier's treatment**
    - Boxplot
    - QQ-plot
    - IQR method
    - Variable transformation techniques
    - Exploratory Data Analysis
    - Uni-variate analysis
    - Bi-variate analysis
    - Multi-variate analysis
    - EDA Analysis with HR domain

10. Python Data Handling and EDA analysis by using Telecom data project

- Project on Data handling
- Data exploration
- Data validation
- Missing values identification
- Outliers Identification
- Data Cleaning

- Basic Descriptive statistics
- EDA analysis
- Generating the insights
- Presentating the insights

11. Varaible Reduction Techniques

- Correlation
    - Pearson correlation
    - Rank Correlation
- VIF/Multi collinearity
- PCA
- Factor analysis
- Chi-Square Technique
- Information value
- Cluster based method
- Tree based method
- Lasso regression method
- Stepwise regression method

Machine learning

Supervised learning -Regression

- Linear Regression
- Multiple linear Regression
- Rigid Regression
- Lasso Regression
- Elastic Net Regression
- Polynomial Regression
- Time series Analysis:
- Need of time series
- Moving average method
- Holt-winter method
- ARIMA method
- Model Evolution metrics
- Use case with Regression models-Project and Assignments

Supervised learning -Classification

- Logistic Regression
- Decision Tree
- Decision Tree Regressors

- Decision Tree Classifier
- Naive Bayes
- KNN
- KNN-Regressors
- KNN-Classifiers- Binary labels and multi labels
- Support Vector Machines
- Support vectors-Regressors
- Support vectors-Classifiers
- Ensemble learning
- Bagging
- Boosting
- Random Forest
- Random Forest -Regressor
- Random Forest-Classifier
- Extra Tree Network
- Model Elevation metrics

Un supervised learning

a. Clustering Analysis
   - Hierarchical Clustering
     Agglomerative Clustering
   - Non-Hierarchical Clustering
     K-Means

b. Market Basket Analysis

Reinforcement learning

- Markov chain- Loan prediction
- Value prediction problems for stock market data

Model Selection and Cross validation

- How to validate a model?
- What is a best model?
- Types of data
- Types of errors
- The problem of over fitting
- The problem of under fitting
- Bias Variance Tradeoff
- Cross validation
- Boot strapping

Neural Networks

- Neural Networks Introduction
- Neural network Intuition
- Neural network and vocabulary
- Neural network algorithm
- Math behind neural network algorithm
- Building the neural networks
- Validating the neural network model
- Neural network applications
- Image recognition using neural networks

NLP-Natural Language Processing:

- What is Text mining
- Corpus
- Tokenizer
- POS
- Named Entry recognizers
- Lemmatization
- NLTK
- Text cleaning
- Words Cleaning
- Stop words
- Cleaning Twitter Data
- Sentimental Analysis
- Text blob
- Word2Vec
- Spelling correction
- TFIDF
- Use Case with Text mining Analysis

Deep learning

1. **Overview of Deep learning by using keras and Tensorflow**
2. **Tensorflow**
    - Introduction to Tensorflow
    - Constant
    - Place holders
    - Varaibles

3. **Multi layers Neural Networks**
    - Neurons
    - Weights
    - Activations
    - Networks of Neurons
    - Training Networks
    - Back propagation
    - Gradient Descent
4. **RNN**
    - Language modeling and generating text
    - Machine translation
    - Speech recognition
    - Generating image description
5. **CNN**
    - Feature learning
        1. convolution
        2. pooling
    - Classification learning
    - Flatten
    - Fully Connnected
    - SoftMax
6. **DNN**
7. **Digit recognizer Classification**
8. **Image processing using OpenCV**
9. **Noisy erodes using open cv**
10. **Image text analysis by using Open CV**
11. **Recommendation engine**
    - Item based
    - Collaberative filtering method

Final Project-Python

- Business understanding-Credit cards and Telecom
- Data requirement
- Data cleaning
- EDA and insight generation
- Variable creation
- Varaible reduction
- Model Building
- Validation Building
- Recommendation to clients