



**Masters In Machine Learning
With
Natural Language Processing**

Masters In Machine Learning and Natural Language Processing

Contents:

1. Python:

- (a) Python –Core**
- (b) Python-Advance**

2. Data Analysis and Visualization

3. Machine Learning and Natural Language Processing

4. Visualization Tools

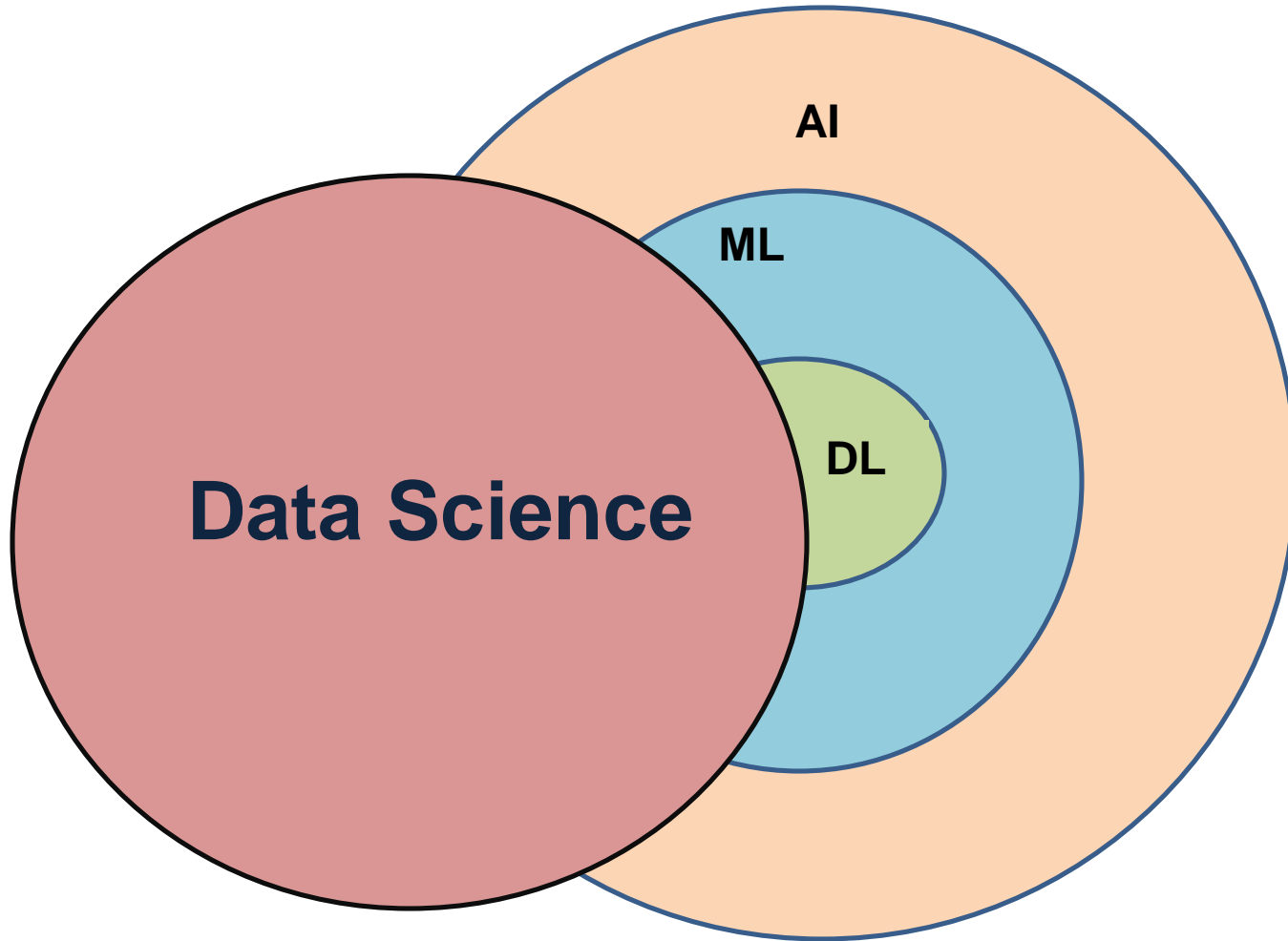
- Tableau**
- Qlik view**

5. Big Data Tools:

- Hadoop**
- Apache Spark**
- SQL**
- Scala**



Road Map For Data Science



Mile Stone 1.

Python For Data Science :

(a) Python Core :

- **Language Fundamentals**
- **Pattern program**
- **Data Types**

(b) Python Advance:

- **Object Oriented Programming**
- **Exception Handling**
- **Working With Files and File Handling**
- **Data Encoding and Processing Digits**
- **Functions**
- **Regular Expressions**
- **Web Crawling**

Mile Stone 2.

Data Analytics and Visualization

1. Statistics

- **Maths**
- **Introductions to Statistics**
- **Sampling Techniques**
- **Descriptive Statistics**
- **Probability and Limitations**
- **Inferential Statistics**

2. Python libraries:

- **Numpy**
- **Pandas**

3.Exploratory Data Analysis

- **Analyse the Data**
- **Feature Engineering and Feature Selections**
- **Data Preprocessing**

4.Visualization

- **Matplotlib**
- **Seaborn**
- **Plotly**



Mile Stone 3.

Machine Learning content :

- 1. Supervised Learning**
 - Regression Techniques**
 - Classification Techniques**
 - Ensemble Methods**
 - DISTANCE BASED MODULES**
 - Support Vector Machines**
- 2. Unsupervised Learning**
 - Principal Components Analysis(PCA)**
 - DBSCAN**
 - K-Means**
 - Hierarchical clustering**
 - Association Rules**
 - Apriori**
- 3. Hyper Parameter Tuning**
- 4. Natural Language Processing**



Python core

Language Fundamentals

- **Introduction To Python**
- **Features of Python**
- **Python Identifiers**
- **Reserved Words and Data Types**
- **Data Types**
- **Type Casting**
- **Fundamental Data Types ,Mutability and Immutability**
- **Numbers, Dates, and Times**
- **Operators**
- **Input and Output Statements**
- **Input Streams**
- **Command Line Arguments**
- **Output Streams**

Pattern Program

- **Flow Control Conditional Statements**
- **Flow Control Iterative Statements**
- **Flow Control Transfer Statements**

Data Type

1. String

- **Pass Statement and Str Data Type**
- **Str Slice Operator**
- **mathematical operators on string**
- **Str Functions Part**

2. List

- **List Data Structure**
- **List Functions**
- **Nested List**
- **List Slice**

3. Tuple and Set Data Types

- **Tuple Data Structure**
- **Set Data Structure**
- **Tuple Data Type**
- **Set Data Type**

4. Dictionary Data Type

- **Dictionary Data Structure**
- **Dictionary Data Type Functions**

Advance Python

Object Oriented Programming (OOPs)

- Introduction
- Class Object Reference variable
- Constructor Vs Method
- Types of variables
- Accessing Of Static Variables
- Types Of Methods
- passing members of one class to another class
- Inner Classes
- Garbage Collection
- Inheritance
- Types of Inheritances
- super() method
- Polymorphism
- Operator overloading
- Method Overloading N Overriding
- Abstract Method N Abstract Class
- Interface Vs Abstract Class Vs Concrete Class
- str() Vs repr()

Exception Handling:

- **Exception Handling and Types**
- **Logging Exceptions**
- **Assertions In Exception**
- **Using Standard Module**
- **Creating new modules**
- **Exceptions Handling with Try-except**
- **Creating, inserting and retrieving Table**
- **Updating and deleting the data.**

working with Files and File Handling I/O

- **Working with files**
- **Reading and writing files**
- **Buffered read and write**
- **Pickling and Unpickling**
- **Other File methods**

Data Encoding and Processing Digits

Functions

- **Functions**
- **Recursive Function**
- **Functions basics**
- **Parameter passing**
- **Iterators**
- **Decorators**
- **Generator functions**
- **Lambda functions**
- **Map, Reduce, filter functions**

Regular Expression:

- **Regular Expression basics**
- **Regular Expression including Web Scraping**

Web Crawling

- **Learning Web Crawling**

Descriptive Statistics

- **What is Univariate and Bi Variate Analysis?**
- **Measures of Central Tendencies**
- **Measures of Dispersion**
- **Skewness and Kurtosis**
- **Box Plots and Outliers detection**
- **Covariance and Correlation**

Probability and Limitations

- **Probability Distribution**
- **Discrete Probability Distributions**
- **Bernoulli, Binomial Distribution, Poisson Distribution**
- **Continuous Probability Distributions**
- **Normal Distribution, Standard Normal Distribution**
- **t- distribution**

Inferential Statistics

- **Sampling variability and Central Limit Theorem**
- **Confidence Intervals**
- **Hypothesis Testing**
- **Z-test, t-test**
- **Chi Square Test**
- **F -Test and ANOVA**

Python Numpy :

- **Ndarray Object**
- **Basic Operations on an Array**
- **Applying Universal functions on an array**
- **Type Conversion**
- **Data Types**
- **Array Attributes**
- **Array Creation Routines**
- **Array from Existing Data**
- **Array From Numerical Ranges**
- **Indexing & Slicing**
- **Advanced Indexing**
- **Broadcasting**
- **Iterating Over Array**
- **Array Manipulation**
- **Binary Operators**
- **String Functions**
- **Mathematical Functions**
- **Arithmetic Operations**
- **Statistical Functions**
- **Sort, Search & Counting Functions**
- **Byte Swapping**
- **Copies & Views**
- **Matrix Library**
- **Linear Algebra (`arange()`,`linspace()`)etc...**



Python Pandas

- **Series**
- **DataFrame**
- **Basic Functionality**
- **Descriptive Statistics**
- **Function Application**
- **Reindexing**
- **Iteration**
- **Sorting**
- **Working with Text Data**
- **Options & Customization**
- **Indexing & Selecting Data**
- **Statistical Functions**
- **Window Functions**
- **Date Functionality**

Python Pandas continue...

- **Timedelta**
- **Categorical Data**
- **Visualization**
- **Reading data from csv,excel etc. into a DataFrame & writing df into csv,excel**
- **Selection and Indexing**
- **Conditional Selection**
- **Groupby**
- **Pivot Table**
- **Merging , Joining, Concatenation**
- **Missing Value Treatment**
- **Data Wrangling Techniques**

Exploratory Data Analysis (EDA)

1. Analyse the data and Understand the data

2. Feature Engineering and Feature Selection

- **Exploratory Data Analysis**
- **Handling Missing Values**
- **Handling Outliers**
- **Categorical Encoding**
- **Normalization and Standardization**
- **Correlation**
- **Forward Elimination**
- **Backward Elimination**
- **Univariate Selection**
- **Random Forest Importance**
- **Feature Selection with Decision Trees**

Data Preprocessing

- **What is Data Preprocessing?**
- **Checking for Null Values**
- **Correlated Feature Check**
- **Data Molding(Encoding)**
- **Data Splitting**
- **Impute Missing Values**
- **Scaling**
- **Label Encoder**
- **One-Hot Encoder**

Visualization

Libraries:

Pandas Builtin Visualisation Library

1. Matplotlib
2. Seaborn
3. plotly

Types:

- Line Plots
- Scatter Plots
- Pair Plots
- Histograms
- Heat Maps
- Bar Plots
- Count Plots
- Factor Plots
- Box Plots
- Violin Plots
- Swarm Plots
- Strip Plots



Machine Learning Terms

Supervised learning :

Regression :

- **Simple Linear Regression: Concept**
- **Minimizing Cost Function**
- **Ordinary Least Square(OLS)**
- **Gradient Descent**
- **Measuring Regression Model Performance: R^2 (R - Square)**
- **Simple Linear Regression**
- **Assumptions of Linear Regression**
- **Dummy Variable**
- **Multiple Linear Regression**
- **Polynomial Linear Regression**
- **Linear Regressions Comparisons**

Classification:

- **Logistic Regression**
- **Confusion Matrix: Measuring Performance of Classification Model**
- **Logistic Regression**
- **K - Nearest Neighbours Algorithm**
- **Naive Bayes**

Unsupervised Learning :

Clustering :

- **K-Means Algorithm**
- **Random Initialization Trap**
- **Elbow Method: Choosing optimum no of clusters**
- **Principal Component Analysis(PCA)**
- **Hierarchical - Agglomerative Algorithm**
- **Agglomerative - Dendrogram**
- **Density Based Clustering – DBSCAN**
- **Association Rules**
- **Apriori**
- **Measuring UnSupervised Clusters Performance**

Machine Learning :

1. Supervised Learning

(a) REGRESSION TECHNIQUES :

Linear Regression

- **Simple Linear Regression:**
- **Estimating the Coefficients**
- **Assessing the Coefficient Estimates**
- **R Squared and Adjusted R Squared**
- **MSE and RMSE**

Multiple Linear Regression

- **Estimating the Regression Coefficients**
- **OLS Assumptions**
- **Multi-Collinearity**
- **Feature Selection**

Evaluating the Metrics of Regression Techniques

- **Homoscedasticity and Heteroscedasticity of error terms**
- **Residual Analysis**
- **Q-Q Plot**
- **Cook's distance and Shapiro-Wilk Test**
- **Identifying the line of best fit**
- **Other Considerations in the Regression Model**
- **Qualitative Predictors**
- **Interaction Terms**
- **Non-linear Transformations of the Predictors**

Polynomial Regression

- **Why Polynomial Regression**
- **Creating polynomial linear regression**
- **Evaluating the metrics**

Time Series (Forecasting)

- **What is Times Series Data?**
- **Stationarity in Time Series Data and Augmented Dickey Fuller Test**
- **The AR Process**
- **ACF & PACF**
- **Decomposition of Times Series Trend, Seasonality and Cyclic**
- **Moving Average, EWMA**
- **Exponential Smoothing**
- **ARIMA**

Regularization Techniques

- **Lasso Regularization**
- **Ridge Regularization**
- **ElasticNet Regularization**

CLASSIFICATION TECHNIQUES :

- **An Overview of Classification**
- **Difference Between Regression and classification Models.**
- **Why Not Linear Regression?**

Logistic Regression:

- **The Logistic Model**
- **Estimating the Regression Coefficients and Making Predictions**
- **Logit and Sigmoid functions**
- **Setting the threshold and understanding decision boundary**
- **Logistic Regression for >2 Response Classes**
- **Evaluation Metrics for Classification Models:**
- **Confusion Matrix**
- **Accuracy and Error rate**
- **TPR and FPR**
- **Precision and Recall, F1 Score**
- **AUC – ROC**
- **Kappa Score**

Naive Bayes

- **Principle of Naive Bayes Classifier**
- **Bayes Theorem**
- **Terminology in Naive Bayes**
- **Posterior probability**
- **Prior probability of class**
- **Likelihood**
- **Types of Naive Bayes Classifier**
- **Multinomial Naive Bayes**
- **Bernoulli Naive Bayes and Gaussian Naive Bayes**

Decision Trees (Rule Based Learning):

- **Basic Terminology in Decision Tree**
- **Root Node and Terminal Node**
- **Regression Trees and Classification Trees**
- **Trees Versus Linear Models**
- **Advantages and Disadvantages of Trees**
- **Gini Index**
- **Overfitting and Pruning**
- **Stopping Criteria**
- **Accuracy Estimation using Decision Trees**

Resampling Methods:

- **Cross-Validation**
- **The Validation Set Approach Leave-One-Out Cross-Validation**
- **k-Fold Cross-Validation**
- **Bias-Variance Trade-Off for k-Fold Cross-Validation**

Ensemble Methods in Tree Based Models

- **What is Ensemble Learning?**
- **What is Bootstrap Aggregation Classifiers and how does it work?**

Random Forest

- **What is it and how does it work?**
- **Variable selection using Random Forest**

Boosting: 1.AdaBoost 2.Gradient Boosting

- **What is it and how does it work?**
- **Hyper parameter and Pro's and Con's**

DISTANCE BASED MODULES :

K Nearest Neighbors

- **K-Nearest Neighbor Algorithm**
- **Eager Vs Lazy learners**
- **How does the KNN algorithm work?**
- **How do you decide the number of neighbors in KNN?**
- **Curse of Dimensionality**
- **Pros and Cons of KNN**
- **How to improve KNN performance**

Support Vector Machines

- **The Maximal Margin Classifier**
- **Hyperplane**
- **Support Vector Classifiers and Support Vector Machines**
- **Hard and Soft Margin Classification**
- **Classification with Non-linear Decision Boundaries**
- **Kernel Trick**
- **Polynomial and Radial**
- **Tuning Hyper parameters for SVM**
- **Gamma, Cost and Epsilon**
- **SVMs with More than Two Classes**

Unsupervised Learning

- **Why Unsupervised Learning**
- **How it Different from Supervised Learning**
- **The Challenges of Unsupervised Learning**

Principal Components Analysis

- **Introduction to Dimensionality Reduction and it's necessity**
- **What Are Principal Components?**
- **Demonstration of 2D PCA and 3D PCA**
- **EigenValues, EigenVectors and Orthogonality**
- **Transforming Eigen values into a new data set**
- **Proportion of variance explained in PCA**

DBSCAN Clustering

- **Why do we need DBSCAN Clustering?**
- **What Exactly is DBSCAN Clustering?**
- **Reachability and Connectivity**
- **Parameter Selection**
- **Implementing DBSCAN Clustering in Python**
- **Applying Clustering Algorithms**

K-Means Clustering

- **Centroids and Medoids**
- **Deciding optimal value of 'k' using Elbow Method**
- **Linkage Methods**
- **Hierarchical Clustering**
- **Divisive and Agglomerative Clustering**
- **Dendrograms and their interpretation**
- **Applications of Clustering**
- **Practical Issues in Clustering**

Hierarchical clustering

- **Divisive and Agglomerative Clustering**
- **Dendrograms and their interpretation**
- **Applications of Clustering Practical Issues in Clustering**

Association Rules

Apriori

- **Metric Support-Confidence-Lift**
- **Improving Supervised Learning algorithms with clustering**

Hyper Parameter Tuning

- **Overfitting**
- **GridSearchCV**
- **RandomizedSearch**
- **Scikit Optimize**
- **Hyperopt**
- **Optuna**
- **Automated Hyperparameter Tuning**
- **Bayesian Optimization**
- **Genetic Algorithms**

NATURAL LANGUAGE PROCESSING (NLP):

- **What is Text Mining?**
- **Libraries**
- **NLTK**
- **Structured and Unstructured Data**
- **Extracting Unstructured text from files and websites**
- **Text Pre processing**
- **Regular Expressions for Pattern Matching**
- **Text Normalization**
- **Text Tokenization**
- **Sentence Tokenization**
- **StopWords**
- **Word Tokenization**
- **Text Segmentation**
- **Stemming**
- **Lemmatization**

NLP :

- **Natural Language Understanding Terms**
- **Bag of Words**
- **Word Vectorizer**
- **TFIDF**
- **Automatic Tagging**
- **Unigrams**
- **Bigrams**
- **N-grams Tagging**
- **Transformation based Tagging**
- **POS Tagging**
- **Cosine Similarity**
- **Named Entity Recognition**
- **Genism**
- **Word2vec**
- **AvgWord2vec**

SQL

Introduction to Databases

1. Basics of SQL:

DML, DDL, DCL and Data Types

Common SQL commands using SELECT, FROM and WHERE

Logical Operators in SQL

2. SQL Joins

INNER and OUTER joins to combine data from multiple tables

RIGHT, LEFT joins to combine data from multiple tables

3. Filtering and Sorting

Advanced filtering using IN, OR and NOT

Sorting with GROUPBY and ORDER BY

4. SQL Aggregations

Common Aggregations including COUNT, SUM, MIN and MAX

CASE and DATE functions as well as work with NULL values

5. Subqueries and Temp Tables

Subqueries to run multiple queries together

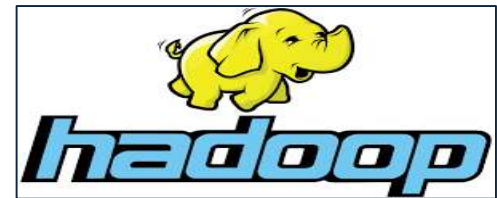
Temp tables to access a table with more than one query

SQL Data Cleaning Perform Data Cleaning using SQL



Big Data for Data Analytics and Machine Learning:

- Big Data and Hadoop Framework Introduction
- HDFS and Exploring Map Reduce
- YARN
- PIG
- Hive
- HBase
- Oozie and Sqoop
- Flume
- Scala
- Apache Spark
- RDD introduction
- Data Frames
- Spark SQL
- Spark Streaming
- MLib and GraphX
- Strom
- MongoDB
- IMPALA



Tableau

- **Tableau Basics**
- **Time series, Aggregations and Filters**
- **Maps, Scatterplots**
- **Joining, Blending and Relationships**
- **Table calculations Advance Dashboards, Storytelling**
- **Advanced Data Preparation**
- **Clusters, Territories, Design Features**



Qlik View

- **Qlikview Basics**
- **Coding Joins in the Script**
- **Script the ETL**
- **KPI**
- **Explore and Develop Chart**
- **Magic Of Variables**
- **Calculated Dimension and Conditional Objects**
- **Comparative analysis**
- **More Qlikview Features**



Spark with Data Analytics and Machine Learning

- Spark overview**
- Spark RDD**
- Spark DataFrame.**
- Spark Architecture.**
- Spark ML lib.**
- Spark Nlp**
- Spark linear regression.**
- Spark logistic regression**
- Spark Decision Tree.**
- Spark Naive Baiyes**
- Spark xg boost**
- Spark time series.**

Projects :

- 5 projects on python
- 15 projects on Data Analytics and Visualization
- 10 Projects on Machine Learning
- 5 projects on Natural Language Processing
- 10 projects on Data Analytics and Machine Learning with Big Data Tools

❖ Complete 45 Projects in Masters in Data Science which will add value to your Resume .