PGP DATA SCIENCE DETAILED SYLLABUS

IIDST

Orientation and Preparatory

Module : Orientation and Preparatory

Welcome to the Course

- **What's in course**
- **Overview of course**
- **TimeLine of course**

Platform Overview

- **Who To Use LMS**
- **Submission, project in LMS**

Phase 1

Module : Excel

Introduction of Excel

- **About Excel & Microsoft**
- **Uses of Excel**
- **Excel software**
- **Spreadsheet window pane**
- **Title Bar**
- **Menu Bar**
- **Standard Toolbar**
- **Formatting Toolbar**
- **The Ribbon**
- **File Tab and Backstage View**
- **Formula Bar**
- **Workbook Window**
- **Status Bar**
- **Task Pane**
- **Workbook & Sheets**

IIDST

Columns & Rows

- **Selecting Columns & Rows**
- **Changing Column Width & Row Height**
- **Autofitting Columns & Rows**
- **Hiding/Unhiding Columns & Rows**
- **Inserting & Deleting Columns & Rows**
- **Cell**
- **Address of a cell**
- **Value**
- **Formula**
- **Use of paste and paste special**

Functionality Using Ranges

- **Using Ranges**
- **Selecting Ranges**
- **Entering Information Into a Range**
- **Using Autofill**

Creating Formulas

- **Using Formulas**
- **Formula Functions - Sum**
- **Average**
- **If**
- **Count**
- **Max**
- **Min**
- **Proper**
- **Upper**
- **Lower**
- **Using Autosum**

Spreadsheet Tools

- **Moving between Spreadsheets**
- **Selecting Multiple Spreadsheets,**
- **Inserting and Deleting Spreadsheets**
- **Renaming Spreadsheets**
- **Copying and Pasting Data between Spreadsheets**
- **Hiding**
- **Protecting worksheets**

Formatting

- **Working with Font Formatting Commands**
- **Changing the Background Color of a Cell**
- **Adding Borders to Cells**
- **Excel Cell Borders Continued**
- **Formatting Percentages**
- **Using Excel's Format Painter**
- **Autofitting Columns & Rows**
- **Creating Styles to Format Data**
- **Merging and Centering Cells**
- **Using Conditional Formatting**
- **Editing Excel Conditional Formatting**

Nested Functions, Advanced Logical Functions(If, AND, OR)

- **Nested If**
- **Average**
- **Max**
- **Min**
- **Logical Functions**

Module : Advance Excel

Advance Formulas

- **Concatenate**
- **Vlookup**
- **Hlookup**
- **Match**
- **Countif**
- **Text**
- **Trim**

Spreadsheet Charts

- **Creating Charts**
- **Different types of chart**
- **Formatting Chart Objects**
- **Changing the Chart Type**
- **Showing and Hiding the Legend**
- **Showing and Hiding the Data Table**

Data Analysis

- **Sorting**
- **Filter**
- **Text to Column**
- **Data validation**

Pivot Tables

- **Creating Pivot Tables**
- **Manipulating a PivotTable**
- **Using the PivotTable Toolbar**
- **Changing Data Field**
- **Properties**
- **Displaying a PivotChart**
- **Setting PivotTable Options**
- **Adding Subtotals to PivotTables**
- **Filtering PivotTable Data**
- **Filtering with the Slicer Tool**

Power Pivot

- **Excel Power Pivot**
- **Activating the Excel PowerPivot AddIn**
- **Creating Data Models with PowerPivot**
- **Excel Power Pivot Data Model Relationships**
- **Creating PivotTables based on Data Models Excel Power Pivot KPIs**

VBA

- **VBA Concepts**
- **VBA Object Oriented Programming Concepts**
- **Visual Basic Editor (VBE)**
- **Excel VBA Immediate Window**
- **Excel VBA Procedure**
- **Adding Code to a VBA Procedure**
- **Excel VBA Comments**
- **Understanding and Working with Excel VBA Variables**
- **Building Logic with an Excel VBA IF Statement**
- **Including an Else Statement in the VBA IF Statement**
- **Working with an Excel VBA For Next Loop**

Module : Storytelling & Dashboard Creation

Dashboard Creation

- **Dashboard Creation**

PROJECTS

Exams

Phase 2

Module : SQL

Introduction

- **Getting Started with SQL and MySQL**
- **What is database ?**
- **Why Use SQL ?**
- **Importance of MySQL**

SQL Server Languages & Relational Databases

- **SQL Language Statements (DDL, DCL, DML, TCL)**
- **Relational Database Terminology**
- **Relational Database essentials**
- **Primary key**
- **Foreign Key**
- **Unique Key & Null Values**

Basics Of SQL

- **First Steps in SQL**
- **Creating a Database**
- **Introduction to datatypes**
- **Creating a Table**

Constraints of MYSQL

- **Different constraints in MySQL**
- **Primary Key Constraint**
- **Foreign key Constraint**
- **Unique constraint**
- **Default constraint**
- **Not Null**

The Select Statement

- **Load the database**
- **Loading any database**
- **Starting with SELECT statement**
- **Group By and Having Clause**

The SQl Statement

- **Insert statement:(Inserting data INTO table)**
- **Update statement: (Commit and rollback)**
- **Delete statement: (Drop vs Truncate)**

Aggregate Functions

- **Functions**
- **Count()**
- **Sum()**
- **Min() and Max()**
- **Avg()**
- **Round()**

Module : Advanced SQL

SQL JOINS

- **Introduction to Joins**
- **Joins Lecture**
- **Left join Lecture**
- **Right join Lecture**
- **Cross join Lecture**
- **Union and Union all**

SUBQUERIES

- **Working with Subqueries**
- **IN nested inside Where**
- **EXISTS nested inside Where**

Stored procedures and functions

Window Functions

CTE

PROJECTS

Module : Maths

Linear Algebra

- **Introduction to vectors and scalars**
- **Vector addition, subtraction, and scalar multiplication**
- **Vector norms and inner products**
- **Introduction to matrices**
- **Matrix addition, subtraction, and scalar multiplication**
- **Matrix multiplication and properties**
- **Transpose, inverse, and determinant of a matrix**
- **Eigenvalues and Eigenvectors**
- **Understanding eigenvalues and eigenvectors**
- **Eigenvalue decomposition (EVD)**
- **Singular Value decomposition (SVD) a:Eand its applications**
- **Principal Component Analysis (PCA) and dimensionality reduction**

Probability

- **Introduction to Probability**
- **Basic concepts of probability theory, repeated random trials**
- **Sample space, events, and outcomes**
- **Discriminative learning, and independence**
- **Probability Distributions**
- **Discrete probability distributions (Bernoulli, binomial, Poisson)**
- **Continuous probability distributions(uniform, normal, exponential)**

- **Probability density function (PDF) and probability mass function (PMF)**
- **Conditional Probability and Bayes' Theorem**
- **Conditional probability definition and calculation**
- **Bayes' theorem and its application in data science**
- **Naive Bayes classifier**
- **Random Variables and Expected Values**
- **Definition of random variables**
- **Expected value and variance of a random variable**
- **Law of large numbers and the central limit theorem**
- **Hypothesis Testing and Confidence Intervals**
- **Null and alternative hypotheses**
- **Type I and Type II errors**
- **Confidence intervals and significance levels**

Calculus

- **Basic of Limits and Continuit**
- **Differentiation**
  - Definition of Derivative
  - Rules of Differentiation (Product Rule, Quotient Rule, Chain Rule)
  - Higher Order Derivatives
  - Implicit Differentiation
  - Applications of Differentiation (Optimization, Rates of Change)
- **Finding Maximums and Minimums-Use derivatives to find the maximum and minimum values of Multivariable Calculus**
  - Partial Derivatives
  - Gradient and Directional Derivatives
- **Derivatives**
  - Derivative of common functions
  - Product and Chain Rule-Use the product and chain rules to calculate the derivatives of more complicated function
- **Optimization with derivatives**
- **Intro to optimization: Temperature example**
- **Optimizing cost functions in ML: Squared loss**
- **Optimizing cost functions in ML: Log loss**
- **Functions of two or more variables: Gradients and gradient descent**
- **Optimization in Neural Networks and Newton's method**

Calculus

- **Stats types, Descriptive and Infential stats**
- **Describing distributions**
- **Measures of central tendency: mean, median, mode**
- **Expected values**
- **Quantiles and box-plots**
- **Measures of dispersion: variance, standard deviation**
- **Biased vs Unbiased estimates**
- **Maximum likelihood estimation**
- **ML motivation example: Linear Discriminant Analysis**
- **Likelihood**
- **Intuition behind maximum likelihood estimation**
- **MLE: How to get the maximum using calculus**
- **Bayesian statistics**
- **ML motivation example: Naive Bayes**
- **Frequentist vs. Bayesian statistics**
- **A priori/ a posteriori distributions**
- **Bayesian estimators: posterior mean, posterior median, MAP**
- **Interval statistics**
- **Confidence Intervals**
  - Margin of error
  - Interval estimation
  - Confidence Interval for mean of population
  - CI for parameters in linear regression
  - Prediction Invterval
- **Hypothesis Testing**
- **ML Motivation: AB Testing**
- **Two types of errors**
- **Test for proportion and means**
- **Two sample inference for difference between groups**
- **ANOVA**

Exams

Module : POWER BI

Introduction of power BI

- **What is BI?**
- **What is Data Visualization?**
- **Data Visualization Preview, Data Visualization Benefits?**
- **What is Power BI?**
- **System requirements, What is Visualizations?**
- **Reports, Dashboards**

Product Info And Installation

- **Power BI Product suite**
- **Power BI Components**
- **Power BI Desktop**
- **Power BI Pro**
- **Power BI Premium**
- **Power BI Desktop Installation**
- **Desktop UI**

Connecting data

- **Types of Data Connectors**
- **The Power Query Editor**
- **Basic Table Transformations**
- **Storage & Connection Modes**
- **Connecting to a Database**
- **Extracting Data from the Web**
- **Data, Refresh data, Data Source Settings**

Data Shaping

- **Text Tools,Numerical Tools, Date & Time Tools**
- **Change Type with Locale**
- **Index & Conditional Columns**
- **Calculated Column Best Practices Grouping & Aggregating**
- **Pivoting & Unpivoting**
- **Merging, Queries, Appending Queries**

Creating Data Model

- Data Modeling
- Database Normalization
- Primary & Foreign Keys
- Relationships vs. Merged Tables
- Creating Table Relationships
- Managing & Editing Relationships
- Star & Snowflake Schemas
- Active & Inactive Relationships
- Relationship Cardinality
- Connecting Multiple Fact Tables
- Filter Context & Filter Flow
- Bi-Directional Filters & Ambiguity
- Hiding Fields from Report View
- Model Layouts
- Data Formats & Categories
- Creating Hierarchies

Calculation with DAX

- Data Analysis Expressions 101
- DAX vs. M Languages
- Intro to DAX Calculated Columns
- Intro to DAX Measures
- Implicit vs Explicit Measures, Quick Measures
- Calculated Columns vs. Measures,
- Dedicated Measure Tables
- Understanding Filter Context
- DAX Syntax & Operators
- Common DAX Function Categories
- Basic Math & Stats Functions,
- Counting Functions
- Conditional & Logical Functions
- The SWITCH Function
- Common Text Functions
- Basic Date & Time Functions
- Joining Data with RELATE
- The CALCULATE Function

- **DAX Measure Totals**
- **The ALL Function**
- **The FILTER Function**
- **Time Intelligence Patterns**
- **Iterator (X) Functions**

Visualizing Data & Report

- **Dashboard Design Framework**
- **Adding Report Pages & Objects**
- **Naming & Grouping Objects**
- **Cards & Multi-Row Cards**
- **Building & Formatting Charts**
- **Line Charts, Trend Lines & Forecasts**
- **KPI Cards,Bar & Donut Charts**
- **Basic Filtering Options**
- **Table & Matrix Visuals**
- **Conditional formatting**
- **Top N Filtering, Top N Text Cards**
- **Map Visuals, Report Slicers**
- **Gauge Charts, Area Charts**
- **Drill Up & Drill Down**
- **Drillthrough Filters, Editing Report Interactions**
- **Adding Bookmarks**
- **Custom Navigation Buttons**
- **Slicer Panels, Numeric Range Parameters**
- **Fields Parameters**
- **Importing Custom Visuals**
- **Managing & Viewing Roles**
- **Mobile Layouts**
- **Publishing to Power BI Service**
- **Visualizing Data with Reports**

PROJECTS

Exams

Phase 3

Module : Python Programming

Python Basic Building

- Python Keywords and identifiers
- Comments, indentation and statements
- Variables and data types in Python
- Standard Input and Output
- Operators
- Control flow: if else elif
- Control flow: while loop
- Control flow: for loop
- Control flow: break and continue

Python Data Structures

- Strings
- Lists, List Comprehension
- Tuples
- Sets
- Dictionary,Dictionary Comprehension

Python Functions

- Python Built-in Functions
- Python User-defined Functions
- Python Recursion Functions
- Python Lambda Functions

Python Exception Handling, Logging And Debugging

- Exception Handling Using Try Catch Block
- Custom Exception Handling
- Logging With Python
- Debugging With Python

Python OOPs

- Python Objects And Classes
- Python Inheritance
- Abstraction In Python
- Polymorphism in Python
- Encapsulation in Python

Flask

- **Flask Fundamentals**
- **Building Rest API's**

Python Project With Deployment

- **End To End Review Scraper Project With Deployment In Cloud**
- **Weather App- Build A Web app that displays current weather conditions for a specific location using OpenWea**
- **Image web scraper- Build A Image Web Scraper which extracts images of Google**

Module : Python Data Wrangling

Numpy

- **Understand the NumPy arrays, operations, methods**

Pandas

- **DataFrame, series, operations, methods**

Matplotlib

- **Plot types of chart, design and customize chart**

Statistics and Visualization with NumPy and Pandas

- **Descriptive Statistics**
- **Introduction to Matplotlib Through a Scatter Plot**
- **Definition of Statistical Measures**
- **Random Variables**
- **Probability Distribution**
- **Discrete Distributions**
- **Continuous Distributions**

Using NumPy & Pandas to Calculate Basic Descriptive Statistics on the DataFrame

- **Use dataframe for advanced data generation, analysis, and visualization.**

Module : Deep Dive In Wrangling

Flask

- **Subsetting the DataFrame**
- **The unique Function**
- **Conditional Selection and Boolean Filtering**
- **Setting and Resetting the Index**
- **The GroupBy Method**

Detecting Outliers and Handling Missing Values

- **Outlier detection**
- **Missing Values in Pandas**
- **Filling and dropping missing Values in Pandas**
- **Outlier Detection Using a Simple Statistical Test**

Concatenating, Merging, and Joining

- **JOIN queries involving multiple DataFrame objects.**
- **Useful Methods of Pandas**

Work with Differnet datasource

- **Go through various data sources and how they can be imported into pandas DataFrames, thus imbuing wrangling professionals with extremely valuable data ingestion knowledge.**
- **Beautiful Soup 4 and Web Page Parsing for data collect**

Use function for advance wrangling

- **Advanced List Comprehension and the zip Function, data formatting dataset, identify and cleaning outliers**
- **Reading Data from XML**
- **Reading Data from an API**
- **Fundamentals of Regular Expressions (RegEx)**

Use function for advance wrangling

- **Define, insert, manipulate, and retrieve data from the databases**
- **Using an RDBMS (MySQL/PostgreSQL/SQLite)**
- **Connect to a database from Python**

Module : Data Visualization In Python

Basic Vizualization using Pandas

- **Data Generation**
- **Line Plot, More on Line Plot, Bar Plot**
- **Stacked Plot, Histogram, Box Plot**
- **Area and Scatter Plot, Hex and Pie Plot,**
- **Scatter Matrix and Subplots**

Basic Vizualization using Pandas

- **Line Plot, Label, Scatter, Bar, and Hist Plots**
- **Box Plot, Subplot,xlim, ylim, xticks, and yticks**
- **Pie Plot, Pieplot text color**
- **Nested Pie Plot, Labeling a Pie Plot**
- **Bar Chart on Polar Axis, Line Plot on a Polar Axis**
- **Scatter Plot on a Polar Axis**
- **Integral in Calculus Plot as Area Under the Curve**
- **Animation Plot**
- **Time Series plot**
  - Dataset Loading
  - Line and Scatter Plots
  - Subplots
  - Heatmap
  - Histogram and KDE Plots

Seaborn

- **Introduction**
- **Scatter Plot, Hue, Style & Size**
- **Pie Plot, Pieplot text color**
- **Line Plot ,Subplot, sns.lineplot(),**
- **sns.scatterplot()**
- **Cat Plot,Box Plot, Boxen Plot**
- **Violin Plot,**
- **Pair Plot,**
- **Regression Plot,**
- **Pair Plot, Regression Plot,**
- **Point Plot, Joint Plot**
- **Controlling Plotted Figure Aesthetics**

Ploty

- **Installation and Setup**
- **Line Plot, Scatter Plot**
- **Bar Plot, Box Plot and Area Plot**
- **3D plot, Spread Plot and Hist Plot**
- **Bubble Plot and Heatmap**

PROJECT

Assignment

Phase 4

Module : Machine Learning

Overview of AI

- **Introduction of AI**
- **AI Vs ML Vs DL Vs DS**
- **Data science and machine learning**
- **Show the usecase of ML**

Introduction of ML

- **What is Machine Learning?**
- **Types of Machine Learning**
- **Differences between supervised learning, unsupervised learning, Reinforcement learning**
- **ML – Applications**
- **Regression and Classification in all type of ML**

Feature Engineering Basic Before Learn Algorithms

- **Feature Selection**
- **Handling missing values**
- **Handling imbalanced data**
- **Handling outliers**
- **Encoding**
- **Feature Scaling**

Supervised learning

- **Defination overview**
- **Label data**
- **Basic Terminology**
- **Types of SL (Regression and Classification)**

Gradient descent

- **Gradient Descent algorithm and its variants**
- **Stochastic Gradient Descent (SGD)**
- **Mini-Batch Gradient Descent with Python**
- **Optimization Techniques for Gradient Descent**
- **Introduction to Momentum-based Gradient Optimizer**

Regression in Supervised Learing

- **Simple Linear Regression**
  - What is Linear Regression?
  - Implement Simple Linear Regression
  - What is the best fit line?
  - Cost Function for Linear Regression
  - Gradient Descent for Linear Regression
  - Evaluation Metrics for Linear Regression
  - Coefficient of Determination or R-Squared (R2)
  - Root Mean Squared Error
  - Assumptions of Linear Regression
  - Linear Regression (Python Implementation)
  - Univariate Linear Regression in Python
  - Multiple Linear Regression using Python
  - Locally weighted Linear Regression
- **Multiple Linear Regression**
- **Polynomial Regression**

Basic of Some Functions (Use in both regression & classification)

- **Loss Functions or Cost function(predict & actual label during trainnig)**
- **Error function/Evaluation (performance evaluation on validation data or test data or unseen data for generalization of the model performance.)**
- **Optimizer functions(updata parameter to minimize the loss funtion)**
- **Overfitting and underfitting**

Regression in Supervised Learing

- **Reqularization in regression(also for classification)**
- **Lasso,Ridge, ElasticNet**
- **Decision tree regression**
- **Bayesian linear regression**
- **Support vector regressor**
- **Ensemble learning in Regression**

Time Series In ML

- **Data Forcasting**

Basic Project on Linear Regession

- **House Price Predictions**

Classification in Supervised Learning

- **Started with Classification (label data)**
- **Classification Types(Binary classification, Multi-class Classification), mulit-Label classification, Imbalanced Classification**

Classification Algorithms

- **Linear Classification**
  - Logistic regression
  - SVM
  - Single-layer Perceptron
  - SGD Classifier
- **Non-Linear Classification**
  - K-Nearest Neighbours
  - Kernel SVM
  - Naive Bayes
  - Decision Tree Classification
  - Ensemble learning classifiers : Random Forests, AdaBoost, Bagging Classifier, Voting Classifier, ExtraTrees Classifier.
  - Multi-layer Artificial Neural Networks

Evaluation metrics in classification:

- **Classification accuracy**
- **Confusion matrix,Precision and Recall**
- **F1-Score**
- **ROC and AUC curve**
- **Cross-validation**

Start Model creation in classification Logistic Regression (it is classification model in ML) then all

- **How does Classification Machine Learning Work?**
- **Classification process steps:**
  - Understanding the problem
  - Data preparation(collecting and preprocessing the data and splitting it into training, validation, and test sets. In this step, the data is cleaned, preprocessed, and transformed into a format for model train)
  - Feature Extraction
  - Model Selection
  - Model Training
  - Model Evaluation
  - Fine Tuning the model
  - Deploying the model

PROJECTS

- **Binary Classification project**
- **Multi-Class Classification project**

Logistic Regression

- **Understanding Logistic Regression**
- **Why Logistic Regression in Classification?**
- **Logistic Regression using Python**
- **Cost function in Logistic Regression**
- **Logistic Regression using Tensorflow**
- **Naive Bayes Classifiers**

Support Vector Machine

- **Support Vector Machines(SVMs) in Python**
- **SVM Hyperparameter Tuning using GridSearchCV**
- **Using SVM to perform classification on a non-linear dataset**

Decision Tree

- **Decision Tree**
- **Decision Tree Regression using sklearn**
- **Decision tree implementation using Python**

Random Forest

- **Random Forest Regression in Python**
- **Ensemble Classifier**
- **Voting Classifier using Sklearn**
- **Bagging classifier**

PROJECTS

Evaluation and Model Selection

- **Bias Variance Trade-Off**
- **Model evaluation techniques**
- **Importance of Splitting the data into training, validation, and testing**
- **Cross-validation techniques**
- **ML Evaluation Metrics**
- **Classification Evaluation Metrics**
- **Accuracy Score**
- **Precision, recall, and F1 score**
- **Confusion Matrix**
- **ROC curve**
- **Regression Evaluation Metrics**
  - Mean Absolute Error
  - Mean Squared Error
  - Mean Absolute Percentage Error
  - R2 Score
- **Hyperparameter tuning**
  - GridSearchCV
  - RandomizedSearchCV

PROJECTS

PROJECTS

Unsupervised Learning

- **Overview of UL,Basic of clustering, dimensionality reduction, and density estimation**
- **Types of UL (clustering, dimensionality reduction, and density estimation)**

Clustering

- **K-means Clustering**
- **Hierarchical Clustering**
- **DBSCAN (Density-Based Spatial Clustering of Applications with Noise)**

Dimensionality Reduction

- **Principal Component Analysis (PCA)**
- **Singular Value Decomposition (SVD)**
- **t-Distributed Stochastic Neighbor Embedding (t-SNE)**
- **Autoencoders**

Dimensionality Reduction

- **Statistical Methods (e.g., Z-score, Mahalanobis distance)**
- **Density-Based Methods (e.g., Isolation Forest, Local Outlier Factor)**
- **Clustering-Based Methods**
- **Support Vector Machines (SVM)**

Associate Rule Learning

- **Apriori Algorithm**

PROJECTS

Reinforcement Learning

- **Introduction to Reinforcement Learning**
- **Density-Based Methods (e.g., Isolation Forest, Local Outlier Factor)**
- **Q-Learning**
- **Deep Q-Networks (DQN)**

RL Implementation

- **Code**

PROJECTS

Assignment

Module : BIG DATA

Introduction

- **Definition of Big Data**
- **Characteristics of Big Data (Volume, Variety, Velocity, Veracity)**
- **Importance and Challenges of Big Data**

Big Data Technologies Overview

- **Apache Spark**
- **MongoDB**
- **Hadoop**
- **Basic of (Apache Hbase, Apache Hive, Apache Kafka, Apache Cassandra, Apache Flink, Apache Storm, Elasticsearch,Scala)**

Introduction To Big Data & Hadoop

- **Types of Digital Data**
- **Introduction to Big Data**
- **Big Data Analytics**
- **History of Hadoop**
- **Apache Hadoop**
- **Analysing Data with Unix tools**
- **Analysing Data with Hadoop**
- **Hadoop Streaming**
- **Hadoop Echo System,**
- **IBM Big Data Strategy**
- **Introduction to Infosphere BigInsights and Big Sheets.**

HDFS(Hadoop Distributed File System)

- **The Design of HDFS**
- **HDFS Concepts**
- **Command Line Interface**
- **Hadoop file system interfaces**
- **Data Flow**
- **DataIngest with Flume and Scoop & Hadoop archives**
- **Serialization**
- **Avro & File Based Data Structures**

Map Reduce

- **Anatomy of a Map Reduce Job Run**
- **Job Scheduling**
- **Shuffle and Sort**
- **Task Execution**
- **Map Reduce Types and Formats**
- **Map Reduce Features.**

Hadoop Eco System

- Pig : Introduction to PIG
- Execution Modes of Pig
- Comparison of Pig with Databases
- Grunt, Pig Latin
- User Defined Functions
- Data Processing operators
- Hive : Hive Shell, Hive Services
- Hive Metastore
- Comparison with Traditional Databases
- HiveQL, Tables
- Querying Data and User Defined Functions
- Hbase : HBasics
- Concepts, Clients, Example
- Hbase Versus RDBMS
- Big SQL : Introduction

Apache Spark

- Introduction to Spark
- Spark Basics
- Working with RDDs in Spark
- Aggregating Data with Pair RDDs
- Writing and Deploying Spark Applications
- Parallel Processing
- Spark RDD Persistence
- Spark MLlib
- Integrating Apache Flume and Apache Kafka
- Spark Streaming
- Improving Spark Performance
- Spark SQL and Data Frames
- Scheduling/Partitioning

MONGO DB

- Introduction to NoSQL and MongoDB
- MongoDB Installation
- Importance of NoSQL
- CRUD Operations
- Data Modeling and Schema Design
- Data Management and Administration

- **Data Indexing and Aggregation**
- **MongoDB Security**
- **Working with Unstructured Data**

Apache Kafka

- **What is Kafka – An Introduction**
- **Multi-Broker Kafka Implementation**
- **Multi Node Cluster Setup**
- **Integrate Flume with Kafka**
- **Kafka API**
- **Producers & Consumers**

CLOUDS

- **AWS, AZURE, GCP**
- **Azure Started**
- **Introduction to Microsoft Azure**
- **Introduction to ARM & Azure Storage**
- **Introduction to Azure storage**
- **Azure Virtual Machines**
- **Azure App and Container services**
- **Azure Networking - I**
- **Azure networking - II**
- **Authentication and Authorization in Azure using RBAC**
- **Microsoft Azure Active Directory**
- **Azure Monitoring**

Case Studies

PROJECT

- **Big Data in Machine Learning**
- **ML Algorithm K-means using Map Reduce for Big Data Analytics**
- **Parallel K-means using Map Reduce on Big Data Cluster Analysis**

PROJECT

- **Big Data Analytics**
- **Decision Trees for Big Data Analytics**
- **Big Data Predictive Analytics**

PROJECT

ASSIGNMENTS

Module : ML OPs

- **Evolution of ML Ops**
- **Key Concepts and Components**
- **Fundamental of Mlops,Stages of MLOps**
- **Why DevOps alone is not Suitable for Machine Learning ?**
- **Why need the new field of ML ops**

Devops for Data Scientists

- **What is SDLC & Why its Important**
- **Types of SDLC**
- **Waterfall Vs Agile Vs DevOps**
- **DevOps Lifecycle & Tools**
- **MLOps Lifecycle & Tools, Devops vs MLops**
- **Basic of Linux, Git & Github, YAML Basic, Docker, Mlflow, CICD**

Basic Terms

- **Model versioning, Auto-ML**
- **Low-code MLOps**
- **Conainerized ML workflow(docker)**
- **Orchestrate ,MLOps step**
- **Version Control Systems**
- **Data Versioning**
- **Model Explainability, Auditability, and Interpretable machine learning**
- **Model Packaging and Serialization**
- **Model Metadata Management**
- **Model Governance Policies**
- **Model Experimentation and Tracking**
- **Monitoring and Logging**

Continuous Integration & Continuous Deployment GIT & GITHUB

- **Overview of CI/CD Pipelines**
- **Automated Testing**
- **Version Control Systems (Git, SVN)**
- **CI/CD Tools (Jenkins, GitLab CI/CD, CircleCI)**

Model Development and Training

- **Data Collection and Preprocessing**
- **Model Selection and Evaluation**
- **Hyperparameter Tuning**
- **Model Training Techniques**

Packaging the ML models

- **Typical Experimentation with Dataset**
- **Model fit and generate Predictions**
- **Challenges in Working inside the Jupyter Notebook**
- **Understanding the Modular Programming**
- **Creating Folder Hierarchy for ML Project**
- **Create Config Module**
- **Data Handling Module**
- **Data Preprocessing part 1**
- **Data Preprocessing part 2**
- **Sklearn pipeline**
- **Training Pipeline**
- **Prediction Pipeline**
- **Perform Training and Predictions**
- **Requirements txt file**
- **Testing the New Virtual Environments**
- **Create Python tests**
- **Running Pytest**
- **Create Manifest file**
- **Create Version File**
- **Create setup.py**
- **Packagiing the ML Model & testing**
- **Summary**

Docker

- **Docker for Machine Learning**
- **Introduction to Docker**
- **Installation of Docker Desktop**
- **Working with Docker**

- **Running the Docker Container**
- **Working with Dockerfile**
- **Push the Docker Image to DockerHub**
- **Dockerize the ML Model**
- **Kubernetes Basics**
- **Generating a container for an ML API with Docker**
- **Docker to generate a container of a web application from Flask, Fastapi, HTML**

Automating the ML model Cycle

- **Packaging the training code in Docker Environment & Summary**

Model versioning and registration with MLflow

- **AutoML Basics**
- **Solution Design**
- **Building a model from start to finish with Pycaret**
- **EDA and Advanced Preprocessing with Pycaret**
- **Development of advanced models (XGBoost, CatBoost, LightGBM) with Pycaret)**
- **Production deployment with Pycaret**
- **Model registry and versioning with MLFlow**
- **Registering a Scikit-Learn model with MLFlow**
- **Registering a Pycaret model with MLflow**
- **Pycaret and Dagshub integration**

Versioning data with DVC

- **Hands on laboratory of registering a model and dataset with Pycaret and DagsHub**
- **Introduction to DVC**
- **DVC commands and process**
- **Hands-on lab with DVC**

Code repository with DagsHub, DVC, Git & MLFlow

- **DVC Pipelines**
- **Introduction to DagsHub for the code repository**
- **EDA and data preprocessing**
- **Training and evaluation of the prototype of the ML model**
- **DagsHub account creation**
- **Creating the Python environment and dataset**

- Deployment of the model in DagsHub
- Training and versioning the ML model
- Improving the model for a production environment
- Using DVC to version data and models
- Sending code, data and models to DagsHub
- Experimentation and registration of experiments in DagsHub

Model Interpretablity

- Using DagsHub to analyze and compare experiments and models
- Basics of interpretability with SHAP
- Interpreting Scikit Learn models with SHAP
- Interpreting models with SHAP in Pycaret

Model serving Through APIs

- Then putting model into production
- Fundamentals of APIs and FastAPI
- Functions, methods and parameters in FastAPI
- POST Method, Swagger and Pydantic in FastAPI
- API development for Scikit-learn model with FastAPI
- API Security
- Automated API development with Pycaret

Deploy to Cloud (Azure)

- Introduction to Machine Learning in Cloud
- Putting the ML application into production in Azure Container with Docker
- SDKs and Azure Blob Storage for model deployment to Azure
- Model training and production deployment in Azure Blob Storage
- Download the Azure Blob Storage model and get predictions
- Run the Model

MLFlow Tools

- Introduction MLFlow
- MIflow Tracking component
- MIflow Loging functions
- Launch multiple Experiments and Runs
- Autologging in MIflow
- Tracking Server of MIflow
- MIflow Model component
- Handling Customized model in MIflow
- MIflow Model evaluation

- **MIflow Registry component**
- **MIflow Project component**
- **MIflow client**
- **MIflow CLI commands**
- **Cloud integration with MIflow**

End To End Project 1

End To End Project 2

PROJECT

ASSIGNMENT

Phase 5

Module : Deep Learning

Basic Learning Of Deep Learning ANN

- **Artifical Neural Network Working**
- **The Neuron**
- **The Activation Function**
- **How do Neural Networks learn?**
- **Back Propogation In ANN**
- **Gradient Descent, Stochastic Gradient Descent**
- **Chain Rule Of Derivatives**
- **Vanishing Gradient Problem**
- **Exploding Gradint Problem**
- **Building an ANN**
- **Neurons, perceptron**
- **Input layer, Output layer**
- **Weights, Bias**

Multi-Layer perceptron Overview

- **Forward propagation**
- **Backword Propagation**
- **Activation function**
- **loss function**
- **Optimizers**

Deep Learning Frameworks

- **Working With Tensorflow, Keras**
- **Working With Pytorch**

Forward Propagation

Backward Propagation

Activation Function

- **Tanh**
- **Relu**
- **Step function**
- **Sigmoid function**
- **ELU**
- **Leaky Relu and Parametric Relu**
- **Softmax**
- **Which activation function used to when**

Loss Function or Cost function

- **LF for Classification and regression**
- **ANN- regression- MSE, MAE, Humber loss, RMSE**
- **Classification -> Cross Entropy--> Binary CE,Catagorical CE, Sparse categorical**
- **IN Neural Network use the combination of all**
- **SGD**

Optimizers

- **Gradient Descent**
- **Mini batch SGD**
- **SGD with Momentum**
- **Adagrad and RMSPROP**
- **Adam Optimizers**

Weight initialization Techniques

- **Exploring Gradient problem --> Weight initialization**
- **Uniform Distribution**
- **Xavier/Glorot initialization**
- **Kaiming he inutialization**

DropOut Layer

CNN

- **Overview ,usecase , pro, cons**
- **Use for images (RGB, GrayScale)**
- **What You'll Need for CNN**
- **Convolution Operation**
- **What are convolutional neural networks?**
- **Step 1(b) - ReLU Layer**
- **Step 2 - Pooling**
- **Step 3 - Flattening**
- **Summary**
- **Softmax & Cross-Entropy**
- **Building an CNN**

RNN Recurrent Neural Networks

- **What You'll Need for RNN**
- **The idea behind RNN**
- **The idea behind Recurrent Neural Networks**
- **The Vanishing Gradient Problem**
- **LSTMs**
- **Practical intuition**
- **EXTRA: LSTM Variations**
- **Building a RNN**
- **Evaluation and Improving RNN**

Extra

- **Self organizing Maps(SOM)**
- **Boltzman machine**
- **AutoEncoders**

PROJECT RNN , LSTM

- **Do Classification and Regression Project**

PROJECT CNN

- **Do Classification and Regression Project**

PROJECT CNN

Exams

Module : Deep Learning

Basic of COPUTER VISION

- **Overview of CV, basic terms use**
- **CNN Fundamentals**
- **Opencv library**
- **Image PIL library**

OPENCV

- **Image operations**
- **Image segmentation**
- **Haar Cascade Classifiers**
- **Image analysis and transformation**
- **Motion and object tracking**
- **Facial Landmark Detection & Face Swaps**
- **Working with Video**
- **Face detection, face recognition, object tracking**

Neural Networl for Image Classification

- **Learn with doing Project**

CNN for Image Classification

- **Learn with doing Project**

Transfer Learing and fine tuning

- **Learn with doing Project**

NN for classification of emotions

- **Learn with doing Project**

Autoencodes

- **Learn with doing Project**

Obejct detection with YOLO

- **Learn with doing Project**

GANs(Generative adversarial networks)

- **GANs & Autoencoders - Generate Digits**
- **Anime Characters**
- **Transform Styles and implement Super Resolution**

Image segmentation

- **Learn with doing Project**

Extra

- **OCR, all major Object Detection Frameworks from YOLOv8**
- **R-CNNs, Detectron2**
- **SSDs, EfficientDetect**
- **Image Classification & Transfer Learning**
- **Googlenet With Research Paper And Practical**
- **Vggnet With Research Paper And Practical**
- **Resnet With Research Paper And Practical**

DO PROJECT

- **Making a Computer Vision API and Web App using Flask**

DO PROJECT

DO PROJECT

Module : NLP

Basics Overview

- **What is NLP, WHY NLP ?, Usecase, Industry Use**
- **Natural Language Processing - Problems and perspectives**
- **Introduction/Recall to/of probability calculus**
- **N-grams and Language Models**
- **Markov Models**
- **Introduction to Machine Learning and Deep Learning**
- **Recurrent Neural Network Language Models**
- **The evaluation of NLP applications**
- **Practical Usecases Of NLP**

NLP BAsic

- **Tokenization Basic**
- **Stemming & Lemmatization**
- **Stop Words**

- **Vocabulary and Matching Part - 1**
- **Vocabulary and Matching Part - 2 (Rule Based)**
- **Vocabulary and Matching Part - 3 (Phrase Based)**
- **Parts of Speech Tagging**
- **Named Entity Recognition**
- **Sentence Segmentation**
- **Vizualizing pos, visualizing NER**

Tools or Libraries

- **NLTK, Spacy**

Python text basic

- **Introduction to Python Text Basics**
- **Working with Text Files with Python**
- **Working with PDFs**
- **Regular Expressions**

Text Preprocessing

- **Tokenization, stop words,regex**
- **Stemming**
- **Lemmatization**
- **Find out how to prepare your text data for most NLP tasks**

Language Parsing

- **Apply regular expressions (regex) and other natural language parsing tactics to find meaning and insights in the texts**

Language Quantification

- **One-hot-encoding, Bag of word**
- **TF,TF-IDF**
- **Word-Embedding**
- **Bag-of-word (ngram)**
- **CBOW and Skipgram**
- **Word-to-vec**
- **Avg wor2Vac**

Word Embeddings

- **Introduction**
- **Train the model for Embedding**

- **Embeding with pretrained models**
- **Attachments of this section - code reference**
- **Introduction to Word Embeddings**
- **Intuition of Vector Representation**
- **Hands On Word Embeddings - Usage of Pre-trained models**
- **Skip-gram Word Embeddings - Understanding Data Preperation**
- **Skip Gram Model Architecture**
- **Skip Gram Hands On - Deep Dive**
- **CBOW Model Architecture & Hands On**
- **Hyperparameters - Negative Sampling and Sub Sampling**
- **Practical Difference between CBOW and Skip-gram**
- **Bonus : How does a Network is trained - Back-propagation**

End to End Pipeline for Text Classification

- **General Pipeline for Classification**
- **Pipeline of nlp**
- **Approaches to Classification**
- **Loading the Dataset**
- **Exploratory Data Analysis & Text Preprocessing**
- **Remove Low Frequency Words**
- **Remove Stop Words with Stemming & Lemmatisation**
- **Application of Model**
- **TfIDF Approach (Text vectorization)**
- **Challenges of NLP & N-grams**
- **Information Extraction (NCR, understanding CRF)**

Markov Models

- **Markov Models Section Introduction**
- **The Markov Property**
- **The Markov Model**
- **Probability Smoothing and Log-Probabilities**
- **Building a Text Classifier (Theory)**
- **Building a Text Classifier (Code)**
- **Language Model (Theory)**
- **Language Model (Exercise Prompt)**
- **Language Model (Code)**

LSTM RNN and NLP

- BackPropogation In Recurrent Neural Network And NLP Application
- Word Embedding Layer And LSTM Practical Implementation In NLP Application
- LSTM Practical Implementation In NLP Application
- Advance NLP Series-Bidirectional LSTM Intuition And Implementation Deep Learning

Transformer NLP Architecture

- BERT model
- Hugging Face model
- Attention models

DO PROJECT

- Latent Semantic Analysis - What does it do?
- SVD - The underlying math behind LSA
- Latent Semantic Analysis in Python
- What is Latent Semantic Analysis Used For?
- Extending LSA

DO PROJECT

- Tweet Sentiment Analysis

DO PROJECT

- Build ChatBot

ASSIGNMENTS

PROJECT

ASSIGNMENT

Module : GEN AI

Foundation

- **Python**
- **Basic Of Statistics**
- **Basic Of MAchine Learning**
- **Deep Learning**
- **NLP**
- **ML**
- **One hot Encoding, Bag Of Words**
- **TFIDF**
- **Word2vec,AvgWord2vec**
- **Tokenization**
- **Basic Deep Learning Concepts**
- **ANN - Working Of MultiLayered Neural Network**
- **Forward Propogation, Backward Propogation**
- **Activation Functions, Loss Functions**
- **Optimizers**

NLP

- **Advanced NLP Concepts**
  - RNN, LSTM RNN, GRU RNN
  - Bidirection LSTM Encoder Decoder
  - Seq to Seq
  - Transformers

NLP

- **Perform prompt engineering with Python**

Understanding GPT(Generative Pre-trained Transformer)

- **Overview**

Generative AI

- **Basic of AI**
- **Types Of Generative AI Models**
  - Transfer Learning
  - Word embeddings
  - Diffusion Models
  - Image genration
  - Generative adversarial networks
  - Transformers & Attention Mechanisms

Generative AI Frameworks

- **LangChain**
- **Vector DataBase (ChromaDB, FAISS)**
- **APIs**
- **Retrieval Augmented Generation(RAG)**
- **OpenAI API**
- **Prompt design**
- **Sematic Search**

Generative AI Tools

- **Hugging face**
- **Chatgpt**
- **Vertex AI (gemini)**
- **Clouds (GCP, AWS,AZURE)**
- **Data science GenAI tools**

Use Generative AI In Data Science

- **Generative AI in Data Science Lifecycle**
- **Gen AI For Data Preparation & Data Querying**
- **Generative AI for Data Insights**
- **Generative AI for Data Visualization**
- **Generative AI for Understanding Data and Model Building**

Introduction to Ethics and Responsibilities in GenAI

- **Understanding the ethical implications of generative models**
- **Addressing bias and fairness in generative AI systems**
- **Ensuring responsible use and deployment of generative models**

Product Development Deploy WebApps with Flask

- **Do a Project**

Module : Git & GitHub

Overview Of Git

- **What is Git ?**
- **Vizualising Git ?**
- **History Of Git**

- **Version control System, features, types**
- **Who use git, Git distributed VCS**
- **Git vs github: what's the difference?**

Installation git & configure shell

- **Install in Macos, Windows, Linux**
- **Install Github Desktop**
- **Setup the git, Git file life-cycle**

Basic Shell Commands

- **Echo, pwd, cwd, cd, dir**
- **Rename, delete, rm , open , edit**

GIT

- **Git Setup: Your Name & Email**
- **Create a New Local Git Repository (Initialize Repository)**
- **Stage & Commit Files**

Working with Branch

- **What Really Matters In This Section**
- **Introducing Branches**
- **The Master Branch (Or Is It Main?)**
- **What On Earth Is HEAD?**
- **Viewing All Branches With Git Branch**
- **Creating & Switching Branches**
- **More Practice With Branching**
- **Another Option: Git Checkout Vs. Git Switch**
- **Switching Branches With Unstaged Changes?**
- **Deleting & Renaming Branches**

Merging Branches

- **Basic of merge**
- **Vizualizing merges**
- **Generate merge commit**
- **Using VSCode to resolve Conflits**

Comparing changes with Git Diff

- **Git diff command**
- **Viewing Unstaged Changes**

- **Vewing working Directory changes**
- **Viewing staged changes**
- **Diffing Specific Files**
- **Comparing Changes Across Branches**
- **Comparing Changes Across Commits**
- **Visualizing Diffs With GUIs**

GitHub Tour

- **Basic of github**
- **Signup**
- **Account configure or setup**
- **What Does Github Do For Us?**
- **Why You Should Use Github!**
- **Cloning Github Repos With Git Clone**
- **Cloning Non-Github Repos**
- **Github Setup SSH Config**
- **Creating Our First Github Repo!**
- **A Crash Course on Git Remotes**
- **Introducing Git Push**
- **Touring A Github Repo**
- **Practice With Git Push**
- **A Closer Look At Git Push**
- **What does "git push -u" mean?**
- **Another Github Workflow Cloning First**
- **Main & Mastithub Default Branches**
- **GitHub: Push to a Remote Repository**
- **GitHub: Pull From a Remote Repository**
- **GitHub: Clone (Download) a Remote Repository**

Fetching & Pulling

- **Remote Tracking Branches: What Are They?**
- **Checking Out Remote Tracking Branches**
- **Working With Remote Branches**
- **Git Fetch: The Basics**
- **Demonstrating Git Fetch**
- **Git Pull: The Basics**
- **Git Pull & Merge Conflicts**
- **A Shorter Syntax For Git Pull?**

Git basic files configuration

- **What Really Matters In This Section**
- **Github Repo Visibility: Public Vs. Private**
- **Adding Github Collaborators**
- **Github Collaboration Demo**
- **What are READMEs?**
- **A Markdown Crash Course**
- **Adding a README To A Project**
- **Creating Github Gists**
- **Introducing Github Pages**
- **gitignore files**
- **Licences**

Git Collaboration Workflows

- **The Pitfalls Of A Centralized Workflow**
- **Centralized Workflow Demonstration**
- **The All-Important Feature Branch Workflow**
- **Merging Feature Branches**
- **Introducing Pull Requests**
- **Making Our First Pull Request**
- **Merging Pull Requests With Conflicts**
- **Configuring Branch Protection Rules**
- **Introducing Forking**
- **Forking Demonstration**
- **The Fork & Clone Workflow**
- **Fork & Clone Workflow Demonstration**

PRACTICE

PROJECT

Exams