**Intermediate Data Science** 

Duration: 3 Days (optional 5-day version with larger project)

Course Code: NTVPY - 240

Description

The objective of this course is to build upon the foundational knowledge of data science and Python programming. Participants will learn more advanced techniques for data manipulation, analysis, and visualization, as well as delve into machine learning algorithms and model evaluation.

Objectives

* Apply advanced data manipulation techniques with pandas.
* Utilize advanced data visualization methods with Matplotlib and Seaborn.
* Implement regression and regularization techniques in machine learning.
* Develop classification models and evaluate their performance.
* Apply unsupervised learning and dimensionality reduction techniques.

Prerequisites

* Six months or more of Python programming experience is recommended
* Familiarity with data science concepts and libraries (pandas, NumPy, Matplotlib, scikit-learn)
* Completion of an introductory data science course or equivalent experience

Outline

* Chapter One: Advanced Data Manipulation with Pandas
* Multi-indexing
* Reshaping data
* Merging and joining data
* Time series data manipulation
* Handling missing data
* Chapter 2: Advanced Data Visualization with Matplotlib and Seaborn
* Customizing plots with Matplotlib
* Plotting geographic data
* Visualizing distributions and relationships with Seaborn
* Interactive visualizations with Plotly
* Chapter 3: Machine Learning: Regression and Regularization
* Linear regression
* Polynomial regression
* Regularization techniques (L1 and L2 regularization)
* Model evaluation metrics for regression (R-squared, RMSE)
* Chapter 4: Machine Learning: Classification and Model Evaluation
* Logistic regression
* Decision trees and ensemble methods (Random Forest, Gradient Boosting)
* Model evaluation metrics for classification (accuracy, precision, recall, F1-score)
* Cross-validation and hyperparameter tuning
* Chapter 5: Unsupervised Learning and Dimensionality Reduction
* Principal Component Analysis (PCA)
* Clustering algorithms (KMeans, DBSCAN)
* Evaluating clustering performance
* Applications of unsupervised learning
* Chapter 6: Text Processing and Natural Language Processing (NLP)
* Text preprocessing (tokenization, stemming, lemmatization)
* Bag-of-words model
* Sentiment analysis
* Topic modeling with Latent Dirichlet Allocation (LDA)
* Chapter 7: Introduction to Deep Learning with TensorFlow
* Basics of neural networks
* Building and training a neural network using TensorFlow
* Convolutional Neural Networks (CNNs) for image classification
* Recurrent Neural Networks (RNNs) for sequence data

OPTIONAL - Final Project (2 days)

Participants will work on a data science project using the concepts learned throughout the course