Sub-2:Machine Learning (Supervised & Unsupervised learning)

Feature transformation (one-hot, label encoding), Label Encoding, One-Hot (dummy variable)encoding, Dummy variable trap.

Machine Learning Packages: Scikit-Learn → Custom Transformers, Scikit-Learn → Pipeline.

End-to-End Regression: Linear & Multiple Regression Algorithms, Poisson Regression,Evaluation metrics (R-Squre, Adj R-Squre, MSE, RMSE), Linear Regression Solvers, Normal

Equation, Gradient Descent (Batch, Stochastic and Mini-Batch), Fine Tuning model, PolynomialRegression, Overfitting vs Underfitting (bias-variance tread off), Model Selection

(Train/Validation/Test split, K-Fold Cross Validation), Regularization (Ridge, Lasso and Elastic-Net)

Machine Learning Model Deployment: Simple Flask Web Service Development, Basic

concepts of Git, Code repository in Git, Basic Docker concepts,

End-to-End Classification: Binary class Classification with Logistic Regression Algorithm,

Evaluation metrics, Accuracy score, Confusion matrix, Precision, Recall, Precision – Recall

tread off curve, ROC curve, AUC score, Multi-Class Classification, Logistic Regression Solvers– Gradient Descent, Forward Propagation, Back Propagation

Use Case: Hand written digit recognition (MINST dataset)

Tree Based Algorithms: Regression Trees vs Classification Trees, Entropy, Gini Index,

Information Gain.

Ensemble models: Voting Classifiers (Ensemble Models), Homogeneous Ensemble Models,

Random Forest, Bagging, Pasting

Feature selection

Use Case: Melbourne House price prediction

PROJECT:

Problems related to the following domains /areas

Banking, Stock Market, Medical Domain, Weather, Insurance, etc.

(Examples like Credit Card Fraud detection - Anomaly Detection Algorithm, Deploy ML modelas REST Full web service on AWS EC2 server, Weather Prediction, Stock Market PricePrediction, Medical Domain: Brain Tumor Prediction, etc)

Clustering: K-Means Clustering, DBSCAN – Image segmentation, How to use unsupervised

outcome as support to solve supervised problem

Use Case: Cluster analysis in Image Data

Machine Learning - Natural Language Processing

Web Mining:Textual data sources and formats, social media, web scrapping APIs (example: scrapy).

Introduction to NLP: What is NLP? Tokenization, N-Grams, Stemming & Lemmatization, What is syntactic representation of text data?, Bag Of Words (BOW) representation, TF-IDF representation.

Sentiment Analysis: using Bag Of Words representation, using TF-IDF representationsData Set : IMDB dataset using Scikit-Learn

Introduction to Natural Language Understanding: Parts of Speech Tagging (POS),Dependency Parsing, [named entity recognition (using Spacy module)

TEXT BOOKS:

1. Eric Matthews, 'Python Crash Course'

2. Joe Papa, PyTorch Packet Reference Building and Deploying Deep Learning Models

3. AurelienGeron, Hands On Machine Learning with Scikit-Learn and Tensor Flow Concepts, Tools and Techniques to Build Intelligent Systems

REFERENCE BOOKS:

1. Mark Lutz, Learning Python, 5th Edition

2. Paul Barry, Head-First Python

3. Adnan Aziz, Elements of Programming Interviews in Python: The Insiders'

4. AndriyBurkov, The Hundred-Page Machine Learning Book

5. Drew Conway and John Myles White, Machine Learning for Hackers: Case Studies and Algorithms to Get you Started

6. Nishant Shukla, Machine Learning with TensorFlow