

# TABLE OF CONTENTS

| Module No. | Title  | Page No.  |
|------------|--|-----------|
| <b>1</b>   | <b>FOUNDATIONS OF AI AND ML IN INDUSTRY</b>  | <b>1</b>  |
|            | 1.1 Introduction to AI, ML, and Deep Learning: Definitions, Types, and History                                 | 1         |
|            | 1.2 The ML Pipeline: Data Acquisition, Preprocessing, Model Building, Evaluation, and Deployment               | 21        |
|            | 1.3 Industry Applications of AI/ML: Examples from different sectors (e.g., Finance, Healthcare, Manufacturing) | 25        |
|            | 1.4 Introduction to Python for Data Science  | 34        |
|            | 1.5 Setting up the Development Environment (Jupyter Notebooks, Anaconda)                                       | 35        |
| <b>2</b>   | <b>DATA UNDERSTANDING AND PREPROCESSING</b>  | <b>38</b> |
|            | 2.1 Data Acquisition Techniques: APIs, Databases, Web Scraping   | 38        |
|            | 2.2 Data Cleaning: Handling Missing Values, Outliers, and Inconsistent Data                                    | 41        |
|            | 2.3 Data Transformation: Scaling, Normalization, Encoding Categorical Variables                                | 45        |
|            | 2.4 Exploratory Data Analysis (EDA): Visualizing Data Distributions and Relationships                          | 56        |
| <b>3</b>   | <b>SUPERVISED LEARNING: REGRESSION AND CLASSIFICATION</b>  | <b>61</b> |
|            | 3.1 Linear Regression: Simple and Multiple Linear Regression   | 61        |
|            | 3.2 Polynomial Regression  | 63        |
|            | 3.3 Regularization Techniques (Ridge, Lasso)   | 66        |
|            | 3.4 Model Evaluation Metrics: R-squared, Mean Squared Error (MSE)  | 69        |
|            | 3.5 Logistic Regression  | 72        |
|            | 3.6 Support Vector Machines (SVM)  | 74        |
|            | 3.7 Decision Trees and Ensemble Methods (Random Forest)  | 76        |
|            | 3.8 Model Evaluation Metrics: Accuracy, Precision, Recall, F1-score, AUC-ROC Cancellation institution.         | 79        |
|            | 3.9 Predicting equipment failures in a manufacturing plant.  | 83        |
|            | 3.10 Analyzing customer sentiment from product reviews.  | 92        |

|          |   |            |
|----------|---|------------|
| <b>4</b> | <b>UNSUPERVISED LEARNING: CLUSTERING AND<br/>DIMENSIONALITY REDUCTION</b>                                 | <b>96</b>  |
|          | 4.1 Clustering Techniques   | 96         |
|          | 4.2 K-Means Clustering  | 96         |
|          | 4.3 Hierarchical Clustering   | 99         |
|          | 4.4 Evaluation Metrics for Clustering<br>(Silhouette Score, Davies-Bouldin Index)                         | 101        |
|          | 4.5 Dimensionality Reduction  | 104        |
|          | 4.6 Principal Component Analysis (PCA)  | 107        |
| <b>5</b> | <b>MODEL EVALUATION AND OPTIMIZATION</b>  | <b>110</b> |
|          | 5.1 Model Selection: Bias-Variance Tradeoff, Cross-Validation   | 110        |
|          | 5.2 Hyperparameter Tuning: Grid Search, Randomized Search   | 111        |
|          | 5.3 Regularization Techniques (L1, L2) to Prevent Overfitting   | 113        |
|          | 5.4 Model Evaluation Metrics (Revisited): Selecting the<br>appropriate metrics for specific problem types | 115        |
|          | 5.5 Model Persistence: Saving and Loading Trained Models  | 117        |