**UNIT I:**

INTRODUCTION Outline: Types of ML, ML Process Data exploration (review), ML Process Example K-NN and Accuracy, Feature Normalization, Supervised learning concepts. Regression 4.What data leakage is and how to detect it. 5.Use model selection methods such as cross-validation to tune the choice of model and key parameters. 1.Describe the most common types of machine learning problems, 2.Account for why it is important to have informative data and features for the success of machine learning systems 3.explain on a high level how different machine learning models generalize from training examples. 4.Apply a machine learning toolkit in an application relevant to the data science area 5.write the code to implement some machine learning algorithms Classification k-NN Regression Linear regression, polynomial feature expansion, measuring error: RSS error, k-fold cross validation, Sci-kit learn datasets Overfitting and underfitting.

**UNIT II:**

SUPERVISED LEARNING AND REGULARIZATION Logistic regression, measuring accuracy: ROC, confusion matrix, dealing with categorical and missing data, Regularization: lasso, ridge. Robust regression, Hyper-parameter search, Support vector machines (linear and kernelized): RBF kernels, Multi-class classification, data imputation, data leakage, Decision trees for classification and regression, entropy Boosting, Random forests, gradient boosted decision trees, XGBboost, AdaBoost, feature importance, SVM paper on detecting fraudulent reviews, Naive Bayes,pipelines.

**UNIT III:**

UNSUPERVISED LEARNING Unsupervised learning: density estimation, Unsupervised learning: clustering. Agglomerative/tree-based clustering. K-means and variants, Gradient Descent and EM, dimensionality reduction (PCA, multi-dimensional scaling, t-SNE), Evaluation of unsupervised methods, Midterm Examination (tentative).

**UNIT IV**:

DEEP LEARNING Deep learning, Neural networks, Convolutional NN, Embeddings, Visualizing ConvNets, Sequence problems: Recurrent NN. UNIT V: IMPLICATION OF PRIVACY: Generative Adversarial networks (GANs), FAT-ML: bias in training and data collection, implications of privacy, Final project presentations (or catch-up), Incentives and Learning, adversarial ML.