DATA SCIENCE

Lesson plan:

- **The objective of this course is to give students**
  - basic knowledge about the key algorithms and theory that form the foundation of machine learning, natural language processing, recommendations , deep learning and computational intelligence ·
  - A practical knowledge of data science algorithms and methods so that they will be able to explain the principles, advantages, limitations such as over fitting and possible applications of machine learning ·
  - Identify and apply the appropriate data science technique to classification,decision,optimization, pattern recognition and recommendation problems.

- **TEACHING METHODOLOGY**
  - Teaching Methodology consists of theory and practical with interactive discussions on every topic.. After class assignments on every topics for practicing.

- PRE-**REQUISITES**
  - Fundamentals of programming and algorithms, statistics and linear algebra

Lesson Plan:

Unit 1. (1.5 hours)

- **Introduction**
  - Data Science

- Data and Big data
- Predictive Analysis
- Machine Learning
- Data Mining
- Neural Network
- Deep learning

- **Introduction to Machine learning:** Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed.

  - Machine Learning Basics
  - Why machines learn
  - Machine learning in practice
  - Types of Machine Learning
    - Supervised
    - Unsupervised
    - Reinforcement

Unit 2.

- **Regression**

  - Linear Regression (3 hours) : Linear regression predicts a real-valued output based on an input value. We discuss the application of linear regression to housing price prediction, present the notion of a cost function, and introduce the gradient descent method for learning. Also we will cover practical use-case with Python.

- Cost Function
- Gradient Descent
- Univariate and Multivariate
- Polynomial regression

- Support Vector Regression - SVR (3 hours) : SVR is a machine learning algorithm for regression.we introduce the idea and intuitions behind SVR and discuss how to use it in practice.
  - Intuition
  - Margin
  - Kernel

- Regression Trees (3 hours): Regression trees are supervised machine learning model used to predict a target by learning decision rules from features. As the name suggests, we can think of this model as breaking down our data by making a decision based on asking a series of questions.
  - Decision Tree Regression
  - Random Forest Regression

- Model Evaluation (1 hours):Model Evaluation is an integral part of the model development process. It helps to find the best model that represents our data and how well the chosen model will work in the future.

Unit 3:

- **Classification**
  - Logistic Regression (3 hours) : Logistic regression is a method for classifying data into discrete outcomes. For example, we might use logistic regression to classify an email as spam or not spam. In this module, we introduce the notion of classification, the cost function for logistic regression, and the application of logistic regression to multi-class classification.

- Why Regression?
- Sigmoid Fucntion
- ROC Curve

- K-Nearest Neighbours (KNN) (3 hours): KNN is used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether k-NN is used for classification or regression
- Support Vector Machines (SVM) (3 hours): SVM is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples.
- Naive Bayes (3 hours): NB is a classification technique based on Bayes' Theorem with an assumption of independence among predictors.
- Classification Trees: This are same as Regression trees used to predict a target by learning decision rules from features.
  - Decision Tree Classification (3 hours)
  - Random Forest Classification (3 hours)

- Model Evaluation (2 hours)

Unit 4:

- **Unsupervised Learning** (3 hours): We use unsupervised learning to build models that help us understand our data better.We discuss the k-Means algorithm for clustering that enable us to learn groupings of unlabeled data points.

  - Clustering
    - Hierarchical
    - K-means

Unit 5:

- **Dimensionality Reduction (3 hours):** In this module, we introduce PCA & LDA, and show how it can be used for data compression to speed up learning algorithms as well as for visualizations of complex datasets.

  - Principal Component Analysis - PCA
  - Linear Discriminant Analysis - LDA

Unit 6:

- **Artificial Neural Networks (3 hours):**In this module, we introduce the backpropagation algorithm that is used to help learn parameters for a neural network. At the end of this module, you will be implementing your own neural network for digit recognition.

  - Neural Network Architecture
  - Hidden Layers
  - Neurons
  - Back-Propogation

Unit 7:

- **Deep Neural Networks (DNN) (3 hours) :**Understand the key computations underlying deep learning, use them to build and train deep neural networks, and apply it to computer vision.

  - Convolutional Neural Network (CNN)
  - Recurrent Neural Network (RNN)
    - LSTM

Unit 8:

- **Natural Language Processing (3 hours):** In this module, you will learn how to go from raw texts to predicted classes with traditional methods. It could be news flows classification, sentiment analysis, spam filtering, etc.

    - Text Processing
    - NLTK/SPACY
    - Word Vectors
    - Count Vectors
    - TF-IDF
    - Similarity Matrix

Unit 9:

- **Recommendation Engines (3 hours):** Define what recommendation systems are, review the different types of recommendation systems and discuss common problems that arise when developing recommendation systems.

    - Association Rules
        - Apriori Algorithm
    - Recommendation Types
        - Collaboritive
        - Content-Based