MACHINE LEARNING
WITH PYTHON

CURRICULUM

Machine Learning Basics
- Converting business problems to data problems
- Understanding supervised and unsupervised learning with examples
- Understanding biases associated with any machine learning algorithm
- Ways of reducing bias and increasing generalisation capabilites
- Drivers of machine learning algorithms
- Cost functions
- Brief introduction to gradient descent
- Importance of model validation
- Methods of model validation
- Cross validation & average error

Generalised Linear Models in Python
- Linear Regression
- Regularisation of Generalised Linear Models
- Ridge and Lasso Regression
- Logistic Regression
- Methods of threshold determination and performance measures for classification score models

Tree Models using Python
- Introduction to decision trees
- Tuning tree size with cross validation
- Introduction to bagging algorithm
- Random Forests
- Grid search and randomized grid search
- ExtraTrees (Extremely Randomised Trees)
- Partial dependence plots

Boosting Algorithms using Python
- Concept of weak learners
- Introduction to boosting algorithms
- Adaptive Boosting
- Extreme Gradient Boosting (XGBoost)

Support Vector Machines (SVM) & kNN in Python
- Introduction to idea of observation based learning
- Distances and similarities
- k Nearest Neighbours (kNN) for classification
- Brief mathematical background on SVM/li>
- Regression with kNN & SVM

Unsupervised learning in Python
- Need for dimensionality reduction
- Principal Component Analysis (PCA)

- Difference between PCAs and Latent Factors
- Factor Analysis
- Hierarchical, K-means & DBSCAN Clustering

Artificial Neural Networks in Python
- Introduction to Neural Networks
- Single layer neural network
- Multiple layer Neural network
- Back propagation Algorithm
- Neural Networks Implementation in Python

Text Mining in Python
- Gathering text data using web scraping with urllib
- Processing raw web data with BeautifulSoup
- Interacting with Google search using urllib with custom user agent
- Collecting twitter data with Twitter API
- Naive Bayes Algorithm
- Feature Engineering with text data
- Sentiment analysis

Version Control using Git and Interactive Data Products
- Need and Importance of Version Control
- Setting up git and github accounts on local machine
- Creating and uploading GitHub Repos
- Push and pull requests with GitHub App
- Merging and forking projects

Open CV
- Basic of Computer Vision & Open CV
- Images Manipulations
- Image Segmentation
- Object Detection
- Face, People and Car Detection
- Face Analysis and Fulters
- Machine Learning in Computer Vision
- Motion Analysis & Object Tracking

Case Study
We start with implementing machine learning algorithms in this module. We also get exposed to some important concepts related to regression and classification which we will be using in the later modules as well. Also this is where we get introduced to scikit-learn, the legendary python library famous for its machine learning prowess.

- **Automate lender & borrower matching through prediction of loan interest rates-** In this case study, we try to automate the process of lender and borrower matching for a fintech company by predicting interest rates offered.

- **Classify customers based on revenue potential for a wealth management firm-** In this classification case study, we help a financial institution to predict which one of their customers are going to fall in high revenue grid so that they can be given selective discounts for customer acquisition in a highly competitive industry of wealth management.

- **Capture risks associated with micro loans:** In the 1st exercise you will work on micro loans. Its inherently risky to hand out micro loans because of lack of checks in the natural process of micro loans. and in this case study we try to capture risk associated with these micro loans.

- **How do the tech specifications of a vehicle impact its emissions?** In the 2nd case study we find out effect of technical design specification of a vehicle on average emission and thus its environmental impact.

- **Save lives by predicting health issues in diabetics:** A health care system in a state is struggling with poor detection of severity of health issues in diabetic people. This results in need for re-hospitalisation and many unfortunately not in time. Find out if boosting algos can save lives!

- **Predicting annual income based on census data:** In the take home exercise, find out whether someone is going to have annual income higher than a certain amount just by simple census data and thus identifying potential fraud cases when it comes to filing their taxes.

- **Understanding impact of cash assistance programs in New York:** To understand PCA, we take up data of cash assistance programs in New York. This has more than 60 variables. We'll see how can we reduce the size of the data.

- **Car Survey Data:** We take up car survey data which contains technical & price detail of vehicles through 11 numeric variables. We'll see if these 11 variables represent any hidden factors representing different properties of a vehicle.

- **Pricing wines based on chemical properties:** For K-Means we take data containing chemical properties of 4000+ white wines and examine whether we can find segments of wines based on their chemical compositions.

- **Customer spend data at a retail chain:** For DBSCAN we see how DBSCAN can be used for anomaly detection using expense data of customers from a retail chain

- **Predicting annual income based on census data :** find out whether someone is going to have annual income higher than a certain amount just by simple census data and thus identifying potential fraud cases when it comes to filing their taxes.

- Live demonstrations of web scraping and data cleaning
- Making a portfolio tracking tool using Yahoo finance with Python
Tagging an SMS as SPAM or NON-SPAM based on its content algorithmically with Naive Bayes