PGPAIML - Program Outline

The post-graduate program in Artificial Intelligence & Machine Learning (PGPAIML) is a 12-month, online, comprehensive program covering various facets of analytics – including statistical models, machine learning and deep learning – all in an attempt to help businesses make smart decisions, and build more intelligent products.

The program spans over 175 hours of learning sessions over the course of 12 months. Quizzes, online instruction, labs, projects and hackathons are all completed online. The program comprises of the following courses:

Rough Month Wise Split of Topics	
1	Python for Data Science
2	Statistical Learning
3	Supervised Learning
4	Ensemble Techniques
5	Unsupervised Learning
6	Featurization, Model Selection & Tuning
7	Recommendation Systems
8	Neural Networks & Deep Learning
9	Computer Vision
10	Natural Language Processing
11	Visualization using Tensorboard, Reinforcement Learning, GANs
12	Capstone Project
+	Industry Case Sessions
+	Capstone Project
+	Hackathons

Industry Case Sessions

Industry case sessions are led by industry practitioners in Data Science and Analytics at a variety of partner companies. Case studies include application of analytical techniques to financial services, healthcare, retail and e-commerce among other common applications. A typical industry case session walks the students through a real-life use case, and explores the choice of analytical models and concomitant trade-offs.

Capstone Project

Through a comprehensive Capstone project, the students design and develop an end-to-end analytical solution for a stated business need.

The projects, carried out in groups, are developed in three stages.
Stage 1 – Groups of students will work on a real-life problem and draft an appropriate architecture to solve the problem. This mimics what an IT manager or CIO would need to do at corporations – presenting the problem, the proposed solution, cost and time implications, as well as risks.
Stage 2 – Groups implement the proposed solution (after incorporating feedback from their mentors) and document the implementation process as well as the result.
Stage 3 – Groups reflect on the implementation challenges and present their work to their respective industry mentors (practitioners).

Hackathons

With the help of hackathons, students are given a platform to work as team, collaborate and utilize all the skills they have learnt as a part of the program to come up with the best possible solution to solve a real life case study under a time constraint. Hackathons give our students a realistic view of how data science problems are solved in the industry.

Learning model

The program is divided into two parts: Conceptual Learning followed by mentored sessions. Logistically, the program delivers on both these outcomes regularly. A typical week consists of reviewing at least 2 to 3 hours of recorded videos concluded by a dedicated 2-hours mentor session conducted by Industry experts that focuses on clarifying doubts from the students. Quizzes are conducted at the end of each week to reinforce the concepts and projects at the end of every course to emphasize the learning through practice. Submission of projects is done online through Olympus, our Learning Management System.

Attending all mentored sessions is mandatory for all participants. Students are required to score a minimum of 60 out of 100 in every course and successfully completed the Capstone project.

AIML – Course Curriculum

Python for AIML

- Jupyter notebook – Installation & function
- Functions, Packages & libraries like Numpy, Scikitlearn, Seaborn
- Initializing & working with data structures, arrays, vectors & data frames

Statistical Learning

- Population, sample, methods of sampling – basics of statistics
- Measures of distribution – Mean, median, mode, S.D & variance
- Probability & its basics – Mutually exclusive events, independent events
- Laws of probability
- Conditional probability
- Types of distribution – Binomial, Poisson & Normal distribution
- Hypothesis & T - testing
- ANOVA

SUPERVISED LEARNING

- Regression (Linear, Multiple, Logistic)
- Classification (k-NN, naïve Bayes) techniques
- SVM

UNSUPERVISED LEARNING

- K means Clustering
- Hierarchical clustering
- Principal component analysis

ENSEMBLE TECHNIQUES

- Boosting methods
- Bagging techniques
- Decision Trees
- Random Forests

FMST

- Data pre-processing
- Feature extraction
- Model selection
- Tuning the parameters of an estimator
- Model evaluation

RECOMMENDATION SYSTEMS

- Introduction to Recommendation systems
- Market Basket Analysis
- Content based recommendation system
- Popularity based model
- Collaborative filtering (User similarity & Item similarity)
- Hybrid models

NEURAL NETWORKS BASICS

- Gradient Descent
- Introduction to Perceptron & Neural Networks
- Batch Normalization
- Activation and Loss functions
- Hyper parameter tuning
- Deep Neural Networks
- Tensor Flow & Keras for Neural Networks & Deep Learning

COMPUTER VISION

- Introduction to Convolutional Neural Networks
- Forward propagation & Backpropagation for CNNs
- Convolution, Pooling, Padding & its mechanisms
- CNN architectures like AlexNet, VGGNet, InceptionNet & ResNet
- Transfer Learning
- Semantic segmentation
- YOLO
- Siamese Networks
- Object & face recognition using techniques above

NATURAL LANGUAGE PROCESSING

- Introduction to Sequential data
- Bag of Words Model
- TF-IDF
- POS Tagging
- Named Entity Recognition
- Tokenization
- Stop Words
- Word Vectorizer
- RNNs and its mechanisms
- Vanishing & Exploding gradients in RNNs
- LSTMs
- LSTMs with attention mechanism
- GRU
- Case study: Machine Translation

- Case study: Sentiment analysis

VISUALIZATION USING TENSORBOARD

- What is Tensor board?
- Test vs Train set accuracy
- T-SNE
- Occlusion Experiment
- CAM, Saliency and Activation maps
- Visualizing Kernels

REINFORCEMENT LEARNING

- Value-based methods (e.g. Q-learning)
- Policy-based methods

GANs

- Introduction to GANs
- Auto encoders
- Deep Convolutional GANs
- How to train and common challenges in GANs
- Semi-supervised GANs
- Practical Application of GANs

PROJECTS

1. **A Campaign to sell Personal Loans**
   Identify potential customers for a personal loan product for a bank, allowing the bank to design targeted marketing campaigns to increase conversion

2. **Bank Note Analysis**
   The project is about building a classification model to predict the authenticity of a bank note if it is counterfeit or genuine based on the characteristics of the image of the bank notes.

3. **Predict the onset of Parkinson's disease**
   Parkinson's Disease (PD) is a degenerative neurological disorder marked by decreased dopamine levels in the brain. Traditional diagnosis of onset of Parkinson's disease is often difficult, as monitoring the onset requires repeated clinic visits by the patient. The objective of the project is to build a Machine Learning algorithm over a recording dataset to predict the onset of the disease

4. **Customer Sentiment from Amazon Reviews**
   Build a recommendation model that can determine and predict the sentiments of the customer from text reviews on Amazon's website

5. **Street House View Numbers**
   Recognizing multi-digit numbers in photographs captured at street level is an important component of modern-day map making. More broadly, recognizing numbers in photographs is a problem of interest to the optical character recognition community. The objective of this project is to build a python code for image classification from scratch to understand the nuances of building and training a model and further to understand the advantages of neural networks

6. **Face Recognition**
   Facial recognition is a biometric solution that measures unique characteristics about one's face. Given an image or a video capture of a scene with one or more faces, the project is designed to detect and classify each face as one of the persons whose identity is already known or as an unknown face.

7. **Fake news detection**
   Fake news is increasingly becoming a threat to our society. It is typically generated for increasing commercial interests—attract viewers and collect more advertising revenue. In this project, you will build a classifier model which can predict whether a piece of news is fake by using sequential models in Natural Language Processing