# COMS W4995 Applied Machine Learning

**Instructor:** Vijay Pappu
**Email Address:** vp2472@columbia.edu
**Phone #:** (859)-806-0368

## Course synopsis:

The course introduces students to the applied aspects of Machine Learning, where they gain a better understanding of ML by working with real-world datasets. While the course introduces different ML concepts, the focus would be on gaining hands-on-experience using Python programming. The first half of the course focuses on exploratory data analysis, followed by introduction to supervised and unsupervised learning techniques. The second half focuses on more advanced topics like learning with data (sparse, imbalance & text), neural networks, embeddings and recommender systems. The topics are complemented with programming assignments and a final project that focuses on helping students gain a deeper understanding.

## Prerequisites:

1. Familiarity with ML methods from a course like W4721
2. Introductory-level courses in linear algebra and calculus
3. Familiarity with Python programming and working with packages like numpy, scipy, pandas and matplotlib

## Textbooks:

There is no required textbook for the course, however the lecture material would reference the following books throughout the course:

1. Introduction to machine learning with Python - Mueller, Guido
2. Applied predictive modeling - Kuhn, Johnson
3. Deep Learning - Goodfellow, Bengio & Courville
4. Learning from data - Abu-Mostafa, Magdon-Ismail & Lin

## Grading:

The course consists of 5 programming assignments, 1 midterm and 1 final project, with the following distribution:

- Assignments (5) - **50%**
- Midterm (1) - **20%**

- Project (1) - **30%**

## Project:

The details about the project and submission schedule will be posted at a later date.

## Schedule with tentative topics:

| | Week | Topics | Assignments | By the end of class |
|---|---|---|---|---|
| 1 | 01/17 | 1. Introduction<br>2. Exploratory Data Analysis & Visualization | | Students would be familiar with basic data exploration |
| 2 | 01/24 | 1. Introduction to supervised learning<br>2. Preprocessing | | |
| 3 | 01/31 | 1. Linear models for regression<br>2. Linear models for classification<br>3. Support Vector Machines (SVMs) | | |
| 4 | 02/07 | 1. Trees, Forests & Ensembles<br>2. Gradient Boosting, Calibration | | |
| 5 | 02/14 | 1. Model evaluation<br>2. Calibration<br>3. Automatic machine learning | | Students would be familiar with training & evaluation of linear and ensemble models |
| 6 | 02/21 | 1. Model Interpretation & Feature Selection<br>2. Linear & non-linear dimensionality reduction<br>3. Clustering & mixture models | | |
| 7 | 02/28 | 1. Learning with imbalanced data<br>2. Learning with sparse data | | |
| | 03/07 | Midterm | | |

| | | | | |
|---|---|---|---|---|
| | 03/14 | Spring break | | |
| 8 | 03/21 | 1. Deep Neural Networks (DNN)<br>2. Convolutional Neural Networks | | |
| 9 | 03/28 | 1. Advanced neural networks | | Students would be familiar with applying neural networks to different tasks |
| 10 | 04/04 | 1. Working with text data<br>2. Topic models for text data<br>3. Word & document embeddings | | Students would be familiar working with text data |
| 11 | 04/11 | 1. Content-based recommendations<br>2. Collaborative filtering & matrix factorization | | |
| 12 | 04/18 | 1. Recommendations using DNNs | | Students would be familiar training and evaluating recommender systems |
| 13 | 04/25 | 1. ML in production<br>2. Course Recap | | Students would be familiar with understanding how ML systems are deployed into production |