**Applied Machine Learning in Python**

**In Module One,** you will be introduced to basic machine learning concepts, tasks, and workflow using an example classification problem based on the K-nearest neighbors method, and implemented using the scikit-learn library. This week’s assignment has you work through the process of loading and examining a dataset, training a k-nearest neighbors classifier on the dataset, and then evaluating the accuracy of the classifier and using it to classify new data.

**Key Concepts**

* Understand basic machine learning concepts and workflow
* Distinguish between different types of machine learning tasks, based on examples of how they are applied to real-world problems
* Understand how a basic classification algorithm (k-nearest neighbors) learns and makes predictions
* Build and evaluate a basic k-nearest neighbors classifier on an example dataset using Python and scikit-learn

**In Module Two,** you will delve into a wider variety of supervised learning methods for both classification and regression, learning about the connection between model complexity and generalization performance, the importance of proper feature scaling, and how to control model complexity by applying techniques like regularization to avoid overfitting. In addition to k-nearest neighbors, this week covers linear regression (least-squares, ridge, lasso, and polynomial regression), logistic regression, support vector machines, decision trees, and the use of cross-validation for model evaluation. For this week’s assignment, you’ll explore the relationship between model complexity and generalization performance, by looking at the effect of key parameters on the accuracy of different classification and regression models.

**Key Concepts**

* Understand how different supervised learning algorithms - in particular, those based on linear models - estimate their own parameters from data to make new predictions.
* Understand the strengths and weaknesses of particular supervised learning methods in order to apply the right algorithm for a given task.
* Apply specific supervised machine learning algorithms in Python with scikit-learn.
* Recognize general principles of supervised machine learning that are common across algorithms, such as the connection between model complexity and generalization performance.
* Apply techniques like regularization, feature scaling, and cross-validation to avoid common pitfalls like under- and overfitting.

**In Module Three,** you will cover evaluation and model selection methods that you can use to help understand and optimize the performance of your machine learning models. For this week’s assignment, you will train a classifier to detect fraudulent financial transactions, analyze its performance with different evaluation metrics, and then optimize the classifier’s performance based on different evaluation metrics, depending on the goals of the detection task (e.g. to minimize false positives vs false negatives).

**Key Concepts**

* Understand why accuracy alone can be an inadequate metric for getting a more complete picture of a classifier's performance
* Understand the motivation and definition of a variety of important evaluation metrics in machine learning and how to interpret the results of using a given evaluation metric
* Optimize a machine learning algorithm using a specific evaluation metric appropriate for a given task

In **Module Four**, you will cover more advanced supervised learning methods that include ensembles of trees (random forests, gradient boosted trees), and neural networks (with an optional summary on deep learning). You will also learn about the critical problem of data leakage in machine learning and how to detect and avoid it. The final assignment brings everything together: you will design features for, and build your own classifier on, a prediction problem on a complex real-world dataset.

**Key Concepts**

* Understand how specific supervised learning algorithms - in particular, those based on decision trees and neural networks - estimate their own parameters from data to make new predictions.
* Apply the right algorithm for a given task by understanding the strengths and weaknesses of additional supervised learning methods.
* Apply additional types of supervised machine learning algorithms in Python with scikit-learn.
* Recognize and avoid instances of data leakage

Summary

* Describe how machine learning is different than descriptive statistics.
* Supervised and Unsupervised learning.
* Create and evaluate data clusters.
* Explain different approaches for creating predictive models.
* Build features that meet analysis needs.