Miami Dade College

Course Description

CAP4631C | Machine Learning for Data Analytics | 4.00 Credits

This upper division course is for students majoring in data analytics. Students will learn why machine learning is crucial for data analytics and why regression analysis is a foundation of supervised machine learning. Using Python programming, students will use a variety of packages to create regression models that make predictions. Prerequisites: COP1047C; STA3164 or CAP3330

Course Competencies:

Competency 1: The student will demonstrate an understanding of how Python can be used for regression analysis, data analytics, and machine learning by:

1. Describing the relationship between regression analysis, data analytics, machine learning, and artificial intelligence
2. Describing the three different types of machine learning
3. Installing various Python packages used for data analytics and machine learning such as NumPy, Scikit- learn, Pandas, and Matplotlib
4. Using Python packages and functions to build linear models

Competency 2: The student will use Python and simple linear regression to build supervised machine learning models by:

1. Defining supervised learning and identifying various linear models
2. Describing the measure of linear relationship
3. Implementing linear regression
4. Finding the coefficient of determination and evaluating the significance of coefficients
5. Evaluating fitted values
6. Making predictions with a regression model
7. Implementing a minimizing the cost function

Competency 3: The student will use Python and multiple regression to build supervised machine learning models by:

1. Working with multiple features for linear regression
2. Applying gradient descent
3. Estimating feature importance by inspecting standardized coefficients and comparing models by R- squared
4. Discovering interaction models
5. Implementing Polynomial regression

Competency 4: The student will use Python and logistic regression to build machine learning models for classification by:

1. Identifying a binary classification problem and assessing a classifier's performance
2. Defining a probability-based approach using logit function
3. Implementing multiclass logistic regression

Competency 5: The student will demonstrate an understanding of data preparation for machine learning algorithms by:

1. Implementing numeric feature scaling using Mean centering, Standardization, and Normalization
2. Implementing qualitative feature encoding
3. Implementing numeric feature transformation by observing residuals

4. Finding and inputting missing data
5. Identifying and removing outliers on the data set and among the predictors

Learning Outcomes:

- Use quantitative analytical skills to evaluate and process numerical data
- Solve problems using critical and creative thinking and scientific reasoning
- Use computer and emerging technologies effectively