Miami Dade College

Course Description

CAP3770 |Predictive Analytics Algorithms | 4.00 credits

This course is for students majoring in Data Analytics. Students will learn the fundamental algorithms used in data mining and analysis. Students will learn various methods and techniques used in data mining, clustering and classification. Prerequisite: STA2023.

Course Competencies:

Competency 1: The student will demonstrate data modeling skills, basic clustering and classification algorithms by:

1. Representing data in "n x d" data matrix
2. Identifying attributes and their data types
3. Treating data instances and attributes as vectors and the entire data set as a matrix
4. Applying basic statistical methods and functions to data including random variable, bivariate, and multivariate variables

Competency 2: The student will demonstrate the application of basic statistical algorithms for exploratory data analysis of numeric attributes and categorical attributes by:

1. Performing univariate analysis on single attributes of a data matrix
2. Performing bivariate analysis on two attributes of a data matrix at the same time to determine the empirical relationship between them
3. Performing multivariate analysis of the attributes of the entire data matrix at the same time or across multiple dimensions
4. Normalizing the values of the attributes in cases where the values of the attributes are vastly different in scale.
5. Computing univariate and multivariate normal distribution

Competency 3: The student will demonstrate the use of sequence mining algorithms to discover patterns across time or positions in each dataset by:

1. Using sequence and item mining algorithms to discover all frequent patterns occurring in a dataset
2. Using algorithms that deal with single sequences to find dispersed fixed length and maximal length repeats, tandem repeats, unique subsequences, and missing (un- spelled) subsequences
3. Explaining how projection position -based sequential pattern algorithms can be used to locate projected sequence position for mining local frequency
4. Employing algorithms to find frequent item sets that are used to discover regularities between frequently co-occurring items in large transactions

Competency 4: The student will demonstrate how to assess the significance of mined frequent patterns, as well as the association rules derived from them by:

1. Using rule and pattern assessment measures that aim to quantify different properties of the mined results
2. Employing algorithms to test for the statistical significance of rules and patterns
3. Employing algorithms to obtain confidence intervals for a given assessment measure

Competency 5: The student will employ representative -based cluster analysis methods to group a set of objects in such a way that statistically similar objects are placed into groups (clusters) by:

1. Using the K-means clustering algorithm to partition many observations into k clusters in which each observation belongs to the cluster with the nearest mean serving as a prototype of the cluster
2. Using the expectation–maximization (EM) algorithm to estimate problems for approximately obtaining the maximum likelihood estimates of parameters when some of the data is missing

Competency 6: The student will demonstrate the ability to employ hierarchical cluster analysis methods to create a sequence of nested partitions by:

1. Using agglomerative algorithms or a "bottom-up" approach to clustering data
2. Using the divisive algorithms or "top-down" approach to clustering data

Competency 7: The student will demonstrate how to employ density-based cluster analysis methods/algorithms to group a set of statistically similar objects into clusters by:

1. Using Density -based algorithms, including the DBSCAN (Density Based Spatial Clustering of Applications with Noise) algorithm, to discover clusters of arbitrary shape
2. Using kernel density estimation algorithms to determine the unknown probability density function

Competency 8: The student will apply objective assessment techniques to assess the effectiveness of the clustering analysis by:

1. Performing clustering evaluation to assess the goodness or quality of the clustering
2. Performing clustering stability to understand the sensitivity of the clustering result to various algorithmic parameters
3. Performing clustering tendency to assess the suitability of applying clustering

Competency 9: The student will utilize cluster validation techniques to assess clustering results by:

1. Employing external validation measures that use criteria that are not inherent to the dataset
2. Using internal validation measures which use criteria that are derived from the data itself
3. Applying relative validation measures that aim to compare different clustering directly

Competency 10: The student will employ probabilistic classification algorithms to predict probability distribution over a set of classes by:

1. Using the Bayes classifier algorithm predicts the class that maximizes the posterior probability
2. Using the multilayer perceptron's algorithm to classify patterns and distinguish data that are not linearly separable
3. Using the naïve Bayes classifier algorithm method to categorize objects into classes

Competency 11: The student will employ decision tree classifier algorithms to predict, given a sample input, the item's target value or classification by:

1. Using a classification algorithm to create a model that classifies/predicts the value of a target variable based on several input variables
2. Using the regression tree algorithm for problems where the predicted outcome is an actual number

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