MSET 5930.007 (Research Problem in lieu of thesis)

Instructor: Zhenhua Huang
Office: F115M
Office Hours:
Phone: 8148
Email: Zhenhua.huang@unt.edu

Course Description:

Open to graduate students in the project option. Three (3) hours credit is required. Independent, applied research that addresses a significant problem in engineering technology supervised by a member of the engineering technology graduate faculty
Pre-requisite: Approval of a project proposal by the major professor.

Course Objectives:

Objective of this course is to provide an opportunity to non-thesis MS students gain skills in defining engineering technology problem, proposing and implementing a solution, and documenting findings to completing their degree requirement by presenting their findings to Major Professor as a formal report.

The project is an opportunity for students to practice effective research strategies. Projects must be related to engineering or engineering technology and approved by the student's Major Professor. The formulation of the project plan is a collaborative, mentored experience. Successful project plans are those in which students can do the following:

- State a research problem in a chosen area of study and demonstrate the value of the solution to the research problem;
- Apply sound research methods/tools to problems in an area of study and describe the methods/tools effectively;
- Analyze/interpret research data;
- Draw valid conclusions from data and make a convincing case for the contribution of those conclusions in advancing knowledge within that area; and communicate their research clearly and professionally in both written and oral forms appropriate to the field.

Sources of Projects:

Grading: Major Professor assigns a final grade to this project course. No incomplete grade is allowed unless unforeseen situations occur that was beyond student's control. Department chair's approval is needed before grade of "I" is assigned. **Final grade will be based on bi-weekly presentation and a final presentation.**

Project Title:

Machine Learning in Civil Engineering

Project Description (confirmed by Major Professor):

The machine learning models for the decision making will be applied in different civil engineering fields, which include Logistic/Linear Regression, k-Nearest Neighbor (kNN), Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Gradient Boosting. Those models will be tuned to obtain the optimum prediction model for the proposed target variables.

Logistic and Linear regression: As one of the most popular and simplest techniques for machine learning predictive modeling, a multiple linear regression analysis is used to build a relationship between the numerical target and input variables using a mathematical algorithm. Using the linear regression analysis, the relationship between output target variables and input known as features and numeric value categorized as predictors can be mapped. If target variables are not continuous, the traditional linear regression models' response variable output is not appropriate due to the nominal nature of classifiers. Therefore, a multinomial logistic regression model is used for nominal responses. Although logistic regression is similar to linear regression, it is required to use the logit link function for solving the problem of the non-normal distribution of categorical target variables. A link function is simply a function of the mean of the target variable Y that is used as the target instead of Y itself. All that means is when Y is categorical, the logit of Y is used as the target in a regression equation instead of just Y. The logit function is the natural log of the odds that Y equals one of the categories. If P is a probability, then $P/(1 - P)$ is the corresponding odds; the logit of the probability is the logarithm of the odds.

K-Nearest Neighbor (kNN): The essential theory of k-Nearest Neighbor (kNN) is a process for finding a group of k samples that are nearest to unknown samples based on distance functions in the calibration dataset. From these k samples, the label (class) of unknown samples are determined by calculating the average of the response variables, namely, the class attributes of the k-nearest neighbor. As a result, for this classifier, the k plays an important role in the performance of the kNN, which is the key tuning parameter of kNN. Using a bootstrap procedure, the parameter k was estimated. In this proposed study, k values will be examined from 1 to 20 to identify the optimal k value for all sample sets.

Support Vector Machine (SVM): The Support Vector Machine (SVM) can handle non-separable data, which generalizes the optimal separating hyperplane as the solution to minimize a cost function that expresses a combination of two criteria: margin maximization and error minimization to penalize the wrongly classified samples.

Artificial Neural Network: Artificial Neural Network (ANN) is one of the advanced machine learning algorithms based on the model of a human neuron, which works like the way the human brain processes information. As one of the most well-known and widely adopted machine learning methods, the ANN model enables learning from a training dataset and store the pattern of the data simulating connections of neurons. After training, when new data is applied to the ANN algorithm, it recognizes the pattern from the data and classifies it. Finally, the algorithm gives results quickly and accurately. The ANN models consist of three layers: input, hidden, and output. The input layer represents the input variables while the output layer shows target variables. The hidden layer, where the data are processed, is presented between the input and output layers. Moreover, the hidden layer is essential for nonlinear data. Each layer includes an array of artificial neurons. Therefore, each of the neurons is connected with the succeeding or proceeding layers. This research project will adopt a Multi-layer Perceptron (MLP) architecture of the ANNs with a weighted linear combination function and the hyperbolic tangent activation function. The most popular back-propagation optimization method will be adopted.

Gradient Boosting: As one sub-technique of decision trees, Gradient boosting models are a sequential assembly of different decision trees. According to several subtrees, prediction makes these models more robust. Gradient boosting involves three key elements. A loss function needs to be optimized. A weak learner (decision tree) is used to make predictions. An additive model for adding weak learners to minimize the loss function. The high complexity of gradient boosting models provides good prediction power, which is very useful for variable selections. However, the interpretability of the models is reduced because of the high complexity.