The methodology we will be adopting will seek to predict student performance within course assessments by examining engagement patterns. First, we will define student engagement by conducting a comprehensive literature review. Based on the literature, a model for predicting student performance based on the sum of the clicks that were observed by students will be designed. Other constant variables will also be included to determine the effect on course performance by student. Cluster analysis will also be conducted to determine if student groups exist within the course data and if variation of effects is different between and across any identified groups. Fixed effects models will identify the relationship between predictors and outcomes within each week as opposed to exploring the variation across groups.

We will conduct our student level analysis on one course sample from the Online University dataset. If possible, a second analysis will be conducted on the full sample of courses (7 total VLEs). Regressions will be used on assessment outcomes using the sum of clicks as a main variable. Persistence will also be examined as the difference in the sum of clicks between VLE modules and the vector of fixed effects on assessment scores. Specific methods of analysis will utilize Decision Trees and Support Vector Machines.

Decision trees (DT) are one of the most understandable and easy to interpret prediction

methods. The decision tree construction procedure starts with splitting the entire dataset into several subsets, which contain more or fewer homogeneous states of dependent variable (Turban et al., 2010). At each split in the tree, the impacts of all predictor variables on the dependent variable are evaluated. This procedure takes place successively, until a decision tree gets to a stable state. Decision trees are powerful classification tools used in data mining, also offer a simplistic visual that can be explained easily to administrators and executives (Turban et al., 2010). Among the decision tree models, “classification tree” and “regression tree” types will be used for creating two different prediction model of the student data. Support Vector Machines, acting as a supervised learning algorithm, will perform regression or classification using the student interaction data as feature engineered groups (adapted from Oztekin, 2015).

Asil Oztekin , (2016),"A Hybrid Data Analytic Approach to Predict College Graduation Status and its Determinative Factors", *Industrial Management & Data Systems*, Vol. 116 Iss 8 pp