



CALIFORNIA STATE SCIENCE FAIR 2014 PROJECT SUMMARY

Name(s) Tanay Tandon	Project Number S1427
Project Title A Machine Learning Model for Automated Semantic Short Essay Assessment through Random Forest Based Ensembles and NLP	
<p style="text-align: center;">Abstract</p> <p>Objectives/Goals This research focused on the development and training of a Machine Learning model for extracting meaning from natural language data. In our world today, it has become of increasing importance for computer models to intelligently understand and analyze textual data in natural language format. The model was taught to classify semantics from text through the use of Treebanks (Parse-Trees), Bag of Words Models, Sentiment Classification, and Structural Subject-Predicate patterns. The research was applied to teach an artificial-intelligence the task of autograding short essay responses by learning from teacher grading patterns.</p> <p>Methods/Materials The research was conducted primarily in Python through the use of the Natural Language Toolkit, scikit-learn RF package, and public training data from the Hewlett Foundation. Ten different essay sets with 17,044 responses ranging from the fields of chemistry to language arts were used for training the algorithm. The research hypothesized that through the use of semantic feature modeling and randomized ensemble learners, the learnt model could achieve a level of accuracy comparable to human grading accuracy # 0.80-0.85 Quadratic Weighted Kappa Inter-Rater agreement. The kappa score represents a level of agreement between a gold standard and a test standard for content evaluation. Development occurred over a period of 5 months, with focus on 4 distinct stages: Preprocessing, Feature Generation and Training, K-Fold Cross Validation, Final Model Packaging.</p> <p>Results In the Model Evaluation and Testing phase the algorithm received a Quadratic Weighted Kappa (QWP) average score of 0.76, which represents the inter-rater agreement between the model and the grades given by the human grader. This was based off of a separate test set of 6,000 elements and averaged over 10 cross validation sets.</p> <p>Conclusions/Discussion Overall, the developed model can be of valuable use for education practitioners and standardized testing in order to automatically grade large swaths of short written responses. The algorithm presents a near human accuracy in the autograding task (0.76 QWK). Furthermore, the research represents a level of text-intelligence and the model can be applied in a variety of military applications such as opinion mining, threat identification, and preemptive terror analysis.</p>	
Summary Statement This research focused on training an artificial intelligence model to extract meaning from text and learn to accomplish the task of autograding short essay-responses through Random Forest ensembles.	
Help Received Parents and teacher provided advice on paper, Richard Socher from Stanford AI Lab provided answers regarding questions over email.	