

**CHENNAI – PONDICHERRY**

**Identifying Student Difficulty in a Digital Learning Environment**

**Abstract**

This paper discusses the development of TutorAlert, a natural language processing system similar to those used in sentiment analysis, but applied to the data generated by students in a digital online learning environment in order to detect confused or frustrated students. A number of machine learning algorithms were tested in the development process, including Support Vector Machines (SVM), Naive Bayes, and Random Forest classifiers. As well, an array of natural language preparation techniques were employed to determine the optimum preprocessing configuration to produce relevant results. We found that detecting potential student frustration or confusion was most successful using a Sequential Minimal Optimization algorithm (SMO), along with the Stanford Part-Of-Speech Tagger (POS Tagger), the iterated version of the Lovins stemmer, and a custom dictionary to help determine relevance probability. This model produced a promising initial F1 score of 0.79 and an accuracy of 0.83. Further, agreement values of 88% were achieved during inter-rater reliability testing between the classifier and human judges.