ABSTRACT

Along with wide application of internet made email become popular way for sending unsolicited advertising to many people, this kind of email called spam email. Spam email become a serious problem for e-mail users. Many spare e-mails flood into people's email inboxes and bring catastrophe to their study and work.

Spam filtering help us to recognize weather email is a spam or not. Using text classification method, spam filtering can be built by done the data preprocessing first.

The purpose of this final project is implement Neural Network Learning Vector Quantization as Text Classification and Information Gain (IG) as Feature Extraction based spam filtering and get the accuracy measures.

From the testing result known that Learning Vector Quantization can be use as text classification for spam filtering it can be show from the accuracy 98% with 1001 learning data set and 100 testing data set.

Keywords : spam e-mail, spam filtering, text classification, data preprocessing, information gain, learning vector quantization.