CALIFORNIA STATE SCIENCE FAIR
2017 PROJECT SUMMARY

Name(s)

Anjali Singh

Project Number

J0814

Project Title

Predicting Formality of Written Texts Using Machine Learning Algorithms

Abstract

Objectives/Goals

My project goal was to create a program that could automatically classify a given text document with >80% accuracy into three levels of Formality: Formal, Casual and Semi-Formal.

Methods/Materials

Python program, Natural Language Tool Kit (NLTK) Library, 8 kinds of machine learning algorithms, 2 kinds of feature extraction methods, and 130 pieces of documents as labeled data. 10 experiments were conducted. The classifiers were trained using 100 labeled documents. The model was tested using 30 documents and accuracy was measured against the correct labels.

Results

5 of the 8 classifiers were able to successfully predict the levels of formality with an average accuracy of 88.6%. Bag-of-Words feature extraction method was found more accurate (92.28%) than Part-of-Speech feature extraction method (84.5%).

Conclusions/Discussion

In conclusion, the project achieved its goal and successfully demonstrated a working program that used machine learning algorithms to predict levels of formality of any given text with >80% accuracy.

Summary Statement

My project goal was to analyze the levels of formality of a given text with >80% accuracy using machine learning algorithms and natural language processing.

Help Received

I created this program by myself with guidance from my parents.