Abstract

The aim of this paper is to present an approach for prediction of customer opinion, using supervised machine learning approach and Decision tree method for classification of online hotel reviews as positive or negative. The preliminary extraction and preparation of the data used in the research are described. Three classification models are generated for three different data sets - balanced and unbalanced training sets with two schemes of filtering frequent and infrequent words in the attribute list. The results from the classifier evaluation are compared and discussed. The three classification models are also applied on new unseen data for predicting opinion of hotel guests. The achieved results reveal that the most accurate prediction is achieved when applying the model generated from the balanced training set with filtering rare words.

References

1. Andreas M. Kaplan, Michael Haenlein, 2010. Users of the world, unite! The challenges


Index Terms

Computer Science

Information Sciences

Keywords

Sentiment classification, supervised machine learning, decision tree