

The focus of this thesis is short text classification. Short text is the prevailing form of text on e-commerce and review platforms, such as Yelp, Tripadvisor or Heureka. As the popularity of the online communication is increasing, it is becoming infeasible for users to filter information manually. It is therefore becoming more and more important to recognise the relevant information in text. Classification of reviews is especially challenging, because they have limited structure, use informal language, contain a high number of errors and rely heavily on context and common knowledge. One of the possible applications of machine learning is to automatically filter data and show users only relevant pieces of information. We work with restaurant reviews from Yelp and aim to predict their usefulness. Most restaurants have relatively many reviews, yet only few are truly useful. Our objective is to compare machine learning methods for predicting usefulness.