ABSTRACT

The theme title is «Information system for tonality analysis of the text».

Masterr's thesis: 99 p., 15 fig., 26 tabl., 2 appendices and 37 sources.

Object of research - semantic orientation of news.

Subject of research - the Naive Bayes method and the convolutional neural network.

The main goal - studying the semantic orientation of the text, using different approaches to constructing a list of features and different methods of assessing the orientation (positive, negative). A review of existing models used to build a list of features and evaluate semantic orientation was carried out, and optimal ones were selected.

Methods of investigation - neural networks, methods for text processing.

The main result of this study is the development of algorithms for the classification of news tones. For this purpose two methods are studied. First, it is the Naive Bayes algorithm, which uses a representative group for classification. The second one is a convolutional neural network, which showed very good results when solving image processing tasks, but can also be used to search semantic text orientation. Experiments and studies of the effectiveness of two different algorithms that show positive and negative colors of the text were carried out. In addition, it is necessary to determine the algorithm that gives better results; It is still important to study how accuracy of algorithms can affect the pre-processing of data, the choice of attributes and data.

The main sources of data are the Twitter platform, the expert data on the impact of news on the US economy, news headlines from the Australian news source ABC. The analysis was conducted using the Python programming language and such libraries for data analysis like pandas, sklearn, tensorflow etc.

SEMANTIC ORIENTATION OF THE TEXT, NAIVE BAEYS, CONVOLUTIONAL NEURAL NETWORKS, FEATURE EXTRACTION, CLASSIFICATION ALGORITHM