6. CONCLUSION AND FUTURE WORK

The aim of this work was to develop an NLP solution capable of predicting sentiment and topic of online customer feedback on Twitter with a goal of improving TransferWise's understanding of public sentiment.

The development process started with analysis of the peculiarities of the problem and proposal of model architecture. For each step of the processing pipeline possible implementation options were considered and compared, followed by selection of the most suitable choice. Once the architecture was finalized, a number of experiments were conducted in order to evaluate the performance of the solution and put it in comparison with alternatives.

The model proposed makes use of supervised techniques, performing sentiment classification on multi-point integer scale and topic classification with 63 labels. Sentiment labelling is achieved by performing TF-IDF vectorization of preprocessed documents, combined with Logistic Regression and oversampling for improved performance on imbalanced datasets. Topic classification features TF-IDF vectorization with a somewhat different preprocessing procedure, where classification is done by Support Vector Machine with linear kernel using Classifier Chain as a problem transformation method. The solution was implemented using the Python programming language and a number of open-source data science libraries.

Model presented was evaluated over three datasets and has shown significant improvement in the results. Sentiment classification achieved balanced accuracy of 94,45% on one of sets of documents, outperforming alternative solutions by an average of 17,5% over all three datasets. Topic classification with 63 labels gained mean $F1^M$ score of 85,72%, incorrectly predicting as little as 1,12% of labels. Overall, the performance of the model exceeded that of the existing models in the task of Twitter classification.

As a result of this work, the following objectives have been achieved:

1. Sentiment classification of TransferWise product feedback on Twitter platform has been implemented
2. The insights from sentiment analysis have been supplemented with topic classification
3. A machine learning model for customer experience analysis by employing natural text processing techniques has been developed