

Opponent review of the Master Thesis

Title: "Text classification of customer inquiries via E-Mail"

Author: Armin Weigold (Mat. 00629413)

Opponent: Nicki Bodenschatz, Technische Universität Deggendorf

The goal of the thesis is the establishment of a workflow for the classification of incoming customer inquiries by mail into different categories. To achieve the goal different text classification methods, like Word2vec are used. The work was conducted in cooperation with "CHECK24 Vergleichsportal Energie GmbH", which provides their customers an online comparisons for electricity and gas tariffs. The company provided mails from their real production environment. The main questions of the work are (I) to find the best approach to transform the incoming mails into a machine readable format and (II) to find the best classification method based on machine learning.

In chapter 1 the author presents a brief introduction for his work. Chapter 2 shows the related work. In chapter 3 the author establishes a theoretical background for the different methods of text classification based on the related work. Chapter 4 presents the experimental setup for the workflow to classify the mails. The workflow is split into three parts. The preprocessing of the data, the vectorization and the classification of the mails. Chapter 5 provides an evaluation for the different methods. The evaluation uses on the one hand, calculated values for the precision of the model and on the other hand, a confusion matrix for the discussed approach. The chapter is concluded with a brief comparison of the different methods. Last, in chapter 6 the work is discussed very shortly.

For the work a literature review was performed. The literature review consists of relevant work in the research field of text classification, but it also includes many sources of websites. However, text classification is a very big research topic with many publications. Considering this, the literature research is sufficient but very superficial.

The language of the work is good, but could be more precise in the evaluation of the presented figures and results. The structure of the work is clear and follows scientific standards. The key results of the work are clearly arranged in figures and tables.

Mr. Weigold compares multiple machine learning approaches. The work shows a good understanding of these approaches. However, there is a big imbalance in the labeling of the data set. Mr. Weigold does not really tackle that issue. A better handling of this problem could improve the results of the work significantly.

Overall, the thesis is a sufficient work and I recommend a fair (befriedigend) rating of the work.

Deggendorf, 09. August 2023

Nicki Bodenschatz