Master Thesis Review

Faculty of Mathematics and Physics, Charles University

Thesis author	Mitchell Borchers
Thesis title	Active learning in E-Commerce Merchant Classification using Website Information
Submitted	2023
Program	Artificial Intelligence **Specialization** IUIPA

Review author	Mgr. Marta Vomlelová Ph.D. **Role** Advisor
Position	Department of Theoretical Computer Science and Mathematical Logic

Review text:

The topic of the E-commerce merchant classification using website information was proposed by the student's business partner. The thesis idea is to test whether the active learning approach can reduce the human workload of the annotator.

The thesis is well structured, the text is accompanied by illustrative graphs and tables. The core of the work is in the data collection and experiments. The code for these experiments is made public on GitHub.

The first three sections review the active learning methodology, different algorithms and the xPAL algorithm used in experiments and related works in e-commerce merchant classification.

The data collection task is time consuming. During the work appeared the necessity to translate the non-English data. Even after this, there were many categories with only a few data samples. This leads to collection of an additional dataset, and the experiments are described on both datasets. Even in the extended dataset, reducing the categories to those with a high number of samples significantly improved the classification results.

The experiments cover different ML models (with linear SVM as the winner) and different active learning approaches (xPAL the best). The integration of general ML models with xPAL methodology appeared to be more difficult than expected.

The student has shown his ability to capture a research topic, collect and pre-process data, and evaluate several experiments.

I recommend the thesis for defense.

I suggest to not consider the thesis for the annual award.

June 2^{nd}, 2023

Signature: