Project 1 for w.3KIA

Task description for End-to-End Machine Learning project

Your goal is to build and deploy a machine learning application that uses either a regression or classification model. You will need to combine data from at least two different sources, merge the datasets, and apply feature engineering techniques to extract relevant features for training.

The project will involve training at least two different models, such as Random Forest and Linear Regression, and comparing their performance. After evaluating the models, you will select the best-performing one and deploy it to Huggingface. Note that multiple iterations of data collection, preparation, and model training may be required throughout the process.

Documentation: As you work on the project, be sure to document your process. The following GitHub page (https://github.com/bkuehnis/ai-applications-fs25/blob/main/end-of-module-exam-doc/end-to-end-application/) provides an example of how to structure the documentation, using the Housing Price Prediction use case as a reference. This will help you create clear and organized documentation for your project.

Project Submission: You will need to submit a link to a GitHub repository containing all the files used to train the model. If the dataset is too large, you can upload it to SwitchDrive or another cloud service and provide a link to it. The GitHub repository must also include a README file with project documentation. Be sure to add the user bkuehnis to the repository.

Deadline: You must submit you project 48 hours before the end-of-semester oral exam.

Note: The performance of the model will not directly affect your grade. What is most important is that you follow a logical process in building your model, clearly explain the results you obtain, and describe the steps you took to test your hypothesis.