SAKI SS 2021 Homework 1

Author: Stephanie Mehltretter

Program code: https://github.com/defaultUser3214/saki\_homework\_21

# Summary

**Introduction:**

The fully automated classification of financial transactions may enable an analysis on the buying preferences of the account owner and thus also indirectly assign the account owner to certain marketing target groups. This report elaborates a way to classify the kind of financial transactions using a machine learning model based on a Naïve Bayes classifier. The data set for this task consisting of 209 labelled transaction information was provided by adorsys GmbH & CO KG and the chair for Open Source Software at FAU.

**Development process:**

The development process implements the first four process steps of the CRISP-ML standard [1]:

Business & Data Understanding: It was evaluated, which columns of the data set provide meaningful information for the training of the machine learning (ML) model:

|  |  |  |
| --- | --- | --- |
| **Column** | **Description** | **Relevance for model building** |
| Waehrung | Contains only one value | Not relevant |
| Unamed 0 | Increasing enumeration of the rows | Not relevant |
| Auftragskonto | Contains *missing values* and two values *89990210.0* and *89990201.0* | *89990210.0* shows a significant correlation the label *leisure* 🡪 relevant |
| Buchungstag & Valutadatum | Are different only in one single row. Column *Buchungstag* has a correlation with the label, e.g., transactions with label *leisure* are more frequently booked on Mondays or Tuesdays | Sufficient to use only one column like *Buchungstag* |
|  |  |  |

Data Preparation:

Modeling:

Evaluation:

# Evaluation

An evaluation of your results; includes a metric you chose, its definition and use.

# Screenshot

A screenshot of your Jupyter notebook after final run showing chosen evaluation metric.

Sources:

[1] Studer, S., Bui, T. B., Drescher, C., Hanuschkin, A., Winkler, L., Peters, S., & Müller, K. R. (2021). Towards CRISP-ML (Q): a machine learning

process model with quality assurance methodology. *Machine Learning and Knowledge Extraction*, *3*(2), 392-413.