Teaching Machine Learning and Data Literacy to Students of Logistics using Jupyter Notebooks

Marvin Kastner[1], Janna Franzkeit[2] and Anna Lainé[3]

Abstract: Teaching machine learning in fields outside of computer sciences can be challenging when the students do not have a solid code knowledge. In this work, the requirements for teaching data literacy and code literacy to students of logistics are explored. Specifically, the use of Jupyter Notebooks in a machine learning course for students in logistics is evaluated, using "Teaching and Learning with Jupyter" written by Barba et al. in 2019 that lists several teaching patterns for Jupyter Notebooks.

Keywords: Jupyter Notebooks, Code Literacy, Data Literacy, Machine Learning, Data Science, Logistics, Supply Chain

1   Introduction

In the course "Machine Learning in Logistics", students from logistics Master's programs learn about data science and machine learning with a focus on their future workspace. A common definition of logistics is the accumulation of seven R (7-Rs): How to transport the right good in the right quantity and in the right quality at the right time to the right place in the right condition while this service can be offered at the right price [Co09]. During the lecture, concepts are taught while in the exercises practical skills are in focus, such as how to read in, explore, analyze, and visualize temporal and spatiotemporal data sets as well as large image data sets. Here, Jupyter Notebooks with Python code are used. Students enter the course with a very heterogeneous knowledge about programming partly because coding abilities are not a prerequisite for the course. Therefore, creating exercises that are appealing for both beginners and skilled programmers was a challenge.

A data scientist in logistics is expected to be both literate in code and data. Code literacy describes the ability to understand the underlying concepts of technology that surrounds us. In the scope of the presented course, no prerequisite in coding is required, for code is mainly seen as a tool to teach data literacy. Therefore, the students are mainly taught to read and modify existing code. Data literacy, on the other hand, represents the ability to understand and make decisions from data in their context. In the course, three main axes of this topic are explored with the use of coding exercises: collection of data from data

---

[1] Hamburg University of Technology, Institute of Maritime Logistics, Am Schwarzenberg-Campus 4, 21073 Hamburg, Germany, marvin.kastner@tuhh.de

[2] Hamburg University of Technology, Institute of Maritime Logistics, Am Schwarzenberg-Campus 4, 21073 Hamburg, Germany, janna.franzkeit@tuhh.de

[3] Hamburg University of Technology, Institute for Software Systems, Am Schwarzenberg-Campus 3, 21073 Hamburg, Germany, anna.laine@tuhh.de

sources, performance of basic statistical analyses including problem solving with machine learning, and creation of visuals for dataset characteristics and experiment results.

Over the last two years, students had been continuously invited to provide feedback on the course, both orally and in anonymous, written form. Since it is a highly specialized course in the Master's programs, the number of participants at the end of the semester was not high enough to obtain statistically significant results from questionnaires. Nevertheless, the open questions lead to the conclusion that the general concept was addressing the expectations of the students. However, they did not comment on the style of the questions. In the exercise sheets, whenever students were asked to write code, interpret statistics, or describe graphs, this was done with Jupyter Notebooks. The exercise sheets were created without common guidelines by several research assistants. They displayed large individual differences in types and frequencies of specific patterns. This led to the idea of a detailed evaluation. Therefore, an overview that lists 22 different types of pedagogical patterns that can be used for teaching with Jupyter Notebooks [Ba19] was found. The patterns were analyzed regarding their ability to teach data literacy and 14 patterns were identified as useful in that regard. It was evaluated, that seven of them were used already, at least in parts, in the scope of the exercise. Furthermore, six patterns were identified, that can enrich the exercises in future and are thus planned to be applied newly or in a larger scope. To gain better insights, as well as evaluate the impact of the new patterns, more course iterations are required.

Jupyter Notebook for teaching code and data literacy to logistics students proved to be successful in that it allows a comfortable environment for the students to try out coding without the constraints of having to write a piece of code from the beginning to the end. In addition, visualizations and error messages are displayed close to the code that evoked it. This supports better the use of code as a tool for focusing on data literacy. This use of Jupyter Notebook allows beginners to stay in a safe environment when coding, without restricting experts from writing complex code.

Bibliography

[Ba19]   Barba, L. A. et al.: Teaching and Learning with Jupyter. https://jupyter4edu.github.io/jupyter-edu-book/, accessed 21 Feb 2020.

[Co09]   Coyle, J. J.: Supply chain management. A logistics perspective. South-Western, Mason, Ohio, 2009.