ICIS 2024 Proceedings

International Conference on Information Systems (ICIS)

December 2024

Designing for Effective Human-Guided Machine Learning Feasibility Analysis

Jonas Gunklach
Karlsruhe Institute of Technology, jonas.gunklach@kit.edu

Mario Nadj
University of Duisburg-Essen, mario.nadj@ris.uni-due.de

Merlin Knaeble
Karlsruhe Institute of Technology (KIT), merlin.knaeble@kit.edu

Isabela Bragaglia
Aioneers AG, isabela.bragaglia@aioneers.com

Alexander Mädche
Karlsruhe Institute of Technology (KIT), alexander.maedche@kit.edu

Designing for Effective Human-Guided Machine Learning Feasibility Analysis

Completed Research Paper

Jonas Gunklach
Karlsruhe Institute of Technology
jonas.gunklach@kit.edu

Mario Nadj
University of Duisburg-Essen
mario.nadj@ris.uni-due.de

Merlin Knaeble
Karlsruhe Institute of Technology
merlin.knaeble@kit.edu

Isabela Bragaglia
aioneers Technologies GmbH
isabela.bragaglia@aioneers.com

Alexander Maedche
Karlsruhe Institute of Technology
alexander.maedche@kit.edu

Abstract

Machine learning (ML) holds immense potential for enterprise data use cases, but a lack of skilled data scientists hinders its utilization. Automated ML (AutoML) aims to empower business users but often falls short, especially when domain knowledge influences model selection. It remains unclear how human-guided ML (HGML) systems can effectively empower business users. To address this, we establish a design science research project. Drawing on the theory of effective use and interviews with seven business users, we present three design principles that we instantiated in our prototype, MLFeasi. Our formative evaluation with business users and data scientists revealed the impracticality of completely replacing data scientists. Instead, a collaborative approach involving data scientists is advocated when ML use cases are deemed feasible - a process we refer to as HGML feasibility analysis. The summative evaluation, including a small-scale experiment and real-world use cases, demonstrates MLFeasi's effectiveness in improving HGML feasibility analysis.

Keywords: *Machine Learning, Guidance, Feasibility Analysis, Theory of Effective Use, Business Users, Design Science Research*

Introduction

Machine Learning (ML) has emerged as a transformative technology that holds the potential to revolutionize business and decision-making processes thus offering the potential to solve a wide range of business challenges (Abbasi et al. 2016). However, despite the growing demand for ML applications, there exists a significant hurdle that impedes progress – a critical shortage of skilled data scientists. Gartner (2016) estimates revealed that a staggering 85 percent of data science use cases were left unrealized due to the scarcity of professionals possessing the required analytical skills. Moreover, data scientists are often overloaded with analytical tasks and must wade through various ML use cases that many times cannot be implemented (Sun et al. 2023). In such a challenging landscape, business users who possess invaluable domain-specific knowledge but lack ML expertise can bridge the gap (Michalczyk et al. 2021). Specifically, these professionals have a deep understanding of the business domain and can translate complex business challenges into ML use cases (Sambasivan and Veeraraghavan 2022). Enabling business users to harness ML themselves will reduce the workload of data scientists and promote more efficient task allocations to them (Pinhanez 2019). Therefore, one promising solution lies in the development of accessible, interactive systems designed to guide business users through the entire data science lifecycle (Microsoft 2023b). This would enable them to assess the feasibility of ML use cases without the need for extensive technical skills (Crisan and Fiore-Gartland 2021). Currently, there is much to configure when training ML models and business users may not know how to select and parameterize models in order to explore the solution space or interpret the results (Gil et al. 2019). In addition, the handover between business users and data scientists is also a challenge, as the process of implementing a use case is not linear, but often requires multiple iterations between the two groups.

Automated Machine Learning (AutoML) targets these issues by creating end-to-end ML pipelines, from data loading, data cleaning, and feature engineering to model fitting and selection, and hyperparameter tuning (Crisan and Fiore-Gartland 2021). With AutoML, business users can potentially be enabled to apply ML on their own without having to rely on input from data scientists (Zöller and Huber 2021). Commercial solutions such as Google's Cloud AutoML and Microsoft's Azure Machine Learning Studio also support the automated creation of ML models. Lately, researchers have proposed new approaches for AutoML with promising performance outcomes (Falkner et al. 2018; Feurer et al. 2015). However, the high degree of automation of current AutoML systems, leading to a lack of human control, may be inadequate in cases where the user's domain knowledge influences model selection, or in exploratory situations where the problem is not well-defined (Gil et al. 2019; Lee et al. 2020). Furthermore, current AutoML systems have shown to be insufficient to support business users (Santos et al. 2019). To enable ML for business users, easy-to-use interactive AutoML systems with an appropriate level of user control are needed to help users develop an ML pipeline on their own and guide them through the entire model-building process (Crisan and Fiore-Gartland 2021).

In this context, guidance becomes paramount in supporting business users with limited analytical skills to successfully apply ML (Gil et al. 2019; Gunklach and Nadj 2023; Silver 1991). Recognizing this importance, Gil et al. (2019) coined the term human-guided machine learning (HGML) to describe an emerging research area dedicated to guiding users to leverage domain expertise towards optimal ML algorithm selection and multistep problem solving through system guidance. Exploiting domain knowledge is particularly important for enterprise ML use cases to ensure that the developed models align closely with the specific intricacies and requirements of the business domain (Michalczyk et al. 2021). Santos et al. (2019) also made an attempt to develop a system for business users with limited ML knowledge, however in the evaluation with business users they identified that the system does not provide significant assistance to users lacking analytical skills - underscoring the ongoing difficulties in effectively bridging the communication gap between ML concepts and business users. Moreover, there remains a dearth of comprehensive end-to-end AutoML systems covering the entire data science lifecycle (Crisan and Fiore-Gartland 2021), with many systems tailored only to specific use cases (e.g., predictive maintenance (Bourcevet et al. 2019), demand forecasting (Sun et al. 2020), time series analysis (Bogl et al. 2013). Consequently, Gerhart et al. (2023) concludes that the effective use of HGML for business users remains a challenge. According to Burton-Jones and Grange (2013), effective use of information systems (IS) involves three core elements: transparent interaction, representational fidelity, and informed action. Business users need to unimpededly interact with the HGML system to obtain faithful representations of their business domain, which ultimately enables them to take informed actions (e.g., create the ML model). Thus, HGML systems need to be designed to facilitate transparent interaction and representational fidelity. Hence, we articulate the following research question: *How should HGML systems be designed to faciliate effective use for business users?*

Agsinst this backdrop, we establish a design science research (DSR) project and present the first design cycle in this article. In particular, we gather concrete requirements of business users with initial ML experience and theorize meta-requirements (MR) for the effective use of HGML feasibility analysis. Subsequently, we propose design principles (DP) that address those MRs. Furthermore, we develop *MLFeasi* - a system that enables business users to effectively participate in HGML by enabling them to train ML models and analyze the feasibility of their ML use cases. *MLFeasi* supports the stages of the data science lifecycle from data understanding to modeling and deployment (Microsoft 2023b). We do not specifically address the 'business understanding' phase, as business users already possess a profound understanding of the business domain. In a formative evaluation with three additional business users and four data scientists, we found that a complete replacement of data scientists is not practical - however, a feasibility analysis performed by business users and involving data scientists only when the use case has potential seems promising. Following Gil et al. (2019) and Berkin et al. (2023), we call this concept HGML feasibility analysis. On this basis, we refined *MLFeasi* and conducted a summative evaluation including (1) a small-scale experiment with 16 business users aimed to assess how business users perceive the guidance provided by *MLFeasi* compared to a benchmark and (2) three additional business users and real-world industry data. The results of our system can be provided to data scientists so that they can optimize the model and implement the ML use case more efficiently and effectively.

Our paper contributes to the body of design knowledge by demonstrating how HGML can be applied to increase transparent interaction and representation in order to achieve effective use of HGML feasibility analysis. In particular, we contribute with three DPs for effective HGML feasibility analysis that are grounded in the theory of effective use (TEU) and interviews with business users. Hereby, our work represents an improvement in the DSR knowledge contribution framework according to Gregor and Hevner (2013), as it represents a more efficient and effective solution for a known problem. For practitioners, we explore the collaboration dynamics between business users and data scientists to enhance the implementation of ML use cases, while also examining the role of goal conflicts between them (Gregor and Jones 2007).

Foundations

Human-Guided Machine Learning

AutoML aims to democratize ML by automating key processes such as model selection, hyperparameter tuning, and feature engineering (Zöller et al. 2023). However, despite its promise, AutoML systems face several limitations, particularly in their ability to handle complex real-world applications that require nuanced human expertise and domain-specific adjustments (Sun et al. 2023). For example, many current AutoML systems function as black-box models, providing limited transparency and explainability, which can hinder users' trust and their ability to understand or refine models based on domain knowledge (Barbudo et al. 2023). Moreover, while AutoML can simplify the development of ML models for those with limited technical expertise, it often fails to adequately support users in contexts where domain knowledge is crucial for the ML process (Gil et al. 2019). For instance, systems like XAutoML highlight that the overwhelming amount of visual information and lack of real-time interaction can make the process cumbersome for users without in-depth ML knowledge (Zöller et al. 2023) This aligns with findings that the high degree of automation in current AutoML systems can lead to inadequate user control, particularly in exploratory data analysis or when specific domain knowledge needs to be integrated into the ML model (Lee et al. 2020).

In response to these challenges, human-guided machine learning (HGML) has been proposed as a more interactive alternative, integrating human expertise directly into the ML process (Mosqueira-Rey et al. 2023). Recent implementations of HGML systems using AutoML are Visus (Santos et al. 2019), Snowcat (Cashman et al. 2019), and Two Ravens (Gil et al. 2019). All systems guide the user during initial exploratory data analysis and specify an AutoML optimization procedure. They allow for basic data use specification, such as setting which features to use in order to solve a problem and comparing the generated ML models using traditional metrics and visualizations (e.g., accuracy, confusion matrix). Visus provides functionalities for problem specification, where the user can create a new problem or select an existing one (Santos et al. 2019). Options are available for data exploration through summaries of the input dataset for feature selection, and for examining and comparing output models to their performance and prediction results. In addition, Visus supports data augmentation, which is not supported by the other two tools (Santos et al. 2019). Snowcat provides the additional functionality of model exploration, which is the exploration of what types of models can be created from a data source (Cashman et al. 2019). For example, a data summary is displayed before the problem specification. After data exploration, Snowcat displays an automatically generated list of possible problem specifications (Cashman et al. 2019). However, these systems do not enable business users to perform ML feasibility analysis; they primarily focus on specific ML steps.

Effective Use Theory

The benefit that organizations obtain from an information system is determined by how effectively it is used by its intended users (Trieu et al. 2022). Effective use is defined as "using a system in a way that helps attain the goals for using the system" (Burton-Jones and Grange 2013, p. 634). Their conceptualization of effective use is rooted in representation theory and posits that an information system's purpose is to faithfully represent a real-world domain, relying on deep, surface, and physical structures (Burton-Jones and Straub 2006). Effective use is shaped by three dimensions: transparent interaction, representational fidelity, and informed action (Burton-Jones and Grange 2013). Transparent interaction enhances access to the system's representation, leading to a more faithful representation of the domain. Higher representational fidelity, in turn, improves the ability to take informed action by enhancing user knowledge and capabilities. In HGML systems, the system guides users effectively in selecting ML algorithms and optimizing model performance based on their domain knowledge through transparent interaction. Representational fidelity ensures that the system accurately represents the domain, giving users insights into the implications of their choices. This ensures users can make informed decisions and effectively utilize the system's capabilities for ML feasibility analysis (Trieu et al. 2022).

Design of MLFeasi

To develop *MLFeasi* and support business users in ML feasibility analysis, we adopted a DSR approach. DSR is particularly suited to prescribing how to design and improve information systems for specific situations (Gregor and Hevner 2013). It's recommended for investigating complex, non-decomposable research and business problems, as it facilitates understanding and changing generative events, and emphasizes knowledge creation through rigorous validations (Hevner et al. 2004). DSR scholars begin by identifying the relevant business problem and theorizing attributes of the future system, known as MRs. These MRs reflect the generic requirements the future system should meet. To enhance scientific contribution, DSR should build on justificatory knowledge (or kernel theory) to inform the design theoretically (Möller et al. 2022). Next, the system needs to be designed to fulfill the identified MRs, proposing DPs describing how the new system should be built followed by an evaluation (Kuechler and Vaishnavi 2008). Following this, a rigorous evaluation becomes essential to validate the system's adherence to these principles and its efficacy in addressing the business problem at hand (Peffers et al. 2018).

Therefore, we first conducted interviews with business users in the domains of purchasing and controlling at our industry partner, a large multinational engineering and technology company. On this basis, we collected their requirements for HGML feasibility analysis. Grounded in TEU as our kernel theory, we then synthesize eight MRs and three DPs for HGML feasibility analysis. Building on the commercial KNIME Analytics Platform (Berthold et al. 2008), we then implemented *MLFeasi* as a web application. We chose KNIME due to its seamless integration capabilities with enterprise data sources and other tools, which allowed us to efficiently leverage existing resources and create a user-friendly system for business users to train ML models and analyze the feasibility of their ML use cases (Fillbrunn et al. 2017). To evaluate our implementation of the DPs and demonstrate its impact, we conducted a series of eight formative think-aloud sessions followed by a summative evaluation. The summative evaluation included a small-scale experiment with 16 business users to assess how they perceived the guidance provided by *MLFeasi* compared to a benchmark, as well as validation with three additional business users using real-world industry data.

Interviews and Think-aloud Sessions with Business Users

We conducted seven semi-structured user interviews including think-aloud sessions with business users to receive an initial understanding of the needs and requirements (R) for HGML for feasibility analysis in enterprises. The interviewees were business users who were all potential users of the system due to limited skills in ML. Participation was voluntary and no compensation was offered. We provide information regarding the participant's role and business domain in Table 1. The interview guideline consists of 29 questions and each interview lasted on average 54 minutes. We structured the interviews in four parts: (1) general questions about the participant's role in the enterprise and daily tasks, (2) proficiency in ML relying on the simple ML practice questions from Workera, (3) requirements for a system supporting business users in ML feasibility analysis, and (4) think-aloud-session with Microsoft's Azure Machine Learning Studio. For the think-aloud session, users were tasked to train a simple ML model on a dataset of their choice and vocalize all thoughts loud. We conducted a think-aloud-session to collect business users problems when using Microsoft's Azure Machine Learning Studio and gather further requirements for the ML feasibility analysis system. We relied on Microsoft's Azure Machine Learning Studio as it promises easy-to-use ML (Microsoft 2023a) and is currently used at the industry partner. We recorded all interviews on audio and transcribed them.

Interview	Role	Domain	Proficiency in ML	Duration
BU1	Project Manager	Purchasing	21%	00:50
BU2	Project Manager	Purchasing	15%	00:35
BU3	Controller	Controlling	13%	01:18
BU4	Project Manager	Purchasing	21%	00:56
BU5	Purchasing Buyer	Purchasing	9%	00:52
BU6	Controller	Controlling	13%	01:02
BU7	Product Owner	Purchasing	9%	00:47

Table 1. Semi-Structured Interviews with Business Users

Requirements

During our requirement elicitation process, we identified several key challenges faced by business users when applying ML. Business users frequently encounter issues related to poor data quality, such as missing or inconsistent data, which they find difficult to manage and which significantly hampers the successful application of ML. Additionally, they face challenges in integrating ML tools with their company's existing data infrastructure, leading to difficulties in accessing and preparing data efficiently. These users often lack the necessary technical skills and require comprehensive guidance throughout the ML process, from configuring models to interpreting results, to ensure they can effectively engage with ML tasks and derive meaningful insights. Finally, there is often no standard procedure or clear point of contact within departments for evaluating and implementing ML use cases, further complicating the process for business users.

Following Crisan and Fiore-Gartland (2021), we relied on a set of four higher-order processes for data science to carry out a selective coding of our interview transcripts. Selective coding is a stage in grounded theory research that serves to organize the analysis around a core set of variables (Charmaz and Bryant 2010). Using a selective coding process allowed us to scaffold our analysis around what features business users need in a system for ML feasibility analysis. The set of four higher-order processes were: **Preparation** (i.e., defining needs, data gathering, data creation, profiling, and data wrangling), **analysis** (i.e., experimentation, exploration, modeling, verification, and interpretation), **deployment** (i.e., evaluation and refinement), communication (i.e., dissemination and documentation). Some participants made explicit statements, for instance "For data processing, I want to see the correlation of features". Other statements were rather implicit. "I need a simple way to clean data". We grouped such statements with other implicit and explicit statements into the respective higher-order process step (in this case: preparation). Please note that participants have not specified requirements for the higher-order process step communication. We finally added one category to group **general system requirements**. In the following, we present the identified requirements along the selective coding processes and the two main dimensions of effective use: (1) transparent interaction and (2) representational fidelity.

Data Preparation

Business users articulated various requirements aimed at facilitating data gathering, wrangling, and preparation. **Transparent interaction.** During the interviews, we identified that business users have difficulty determining what type of ML method is appropriate for their user case. For instance, during the think-aloud sessions, participants often wrongly chose the algorithm suggested by the system. Participants requested that the system should allow the user to specify the problem. *"I personally can tell whether it's supervised or unsupervised learning. I know that much. Let's say I want to do a regression, which regression is linear, non-linear, or not? I don't know that"* (BU1). Hereby the system should not do this automatically, but rather suggest which ML task is appropriate for the target variable but still present the other options, including explanations of possible ML tasks so that the user can make an informed decision (**R1**). To explore and understand the data, participants requested that the system should show a data preview and also visualize univariate and bivariate statistics. Therefore, the system should visualize data and statistics that allow the user to determine the quality of the data, the relationships between variables, and the relationships between explanatory and target variables (**R2**). Further, the system should automatically pre-select potential features, either based on correlation to the target variable or feature importance values. *"For me, it would be best if the system chooses the best features for me so I have fewer options to break it or* *do it wrong"* (BU5). For that, the system should suggest a suitable number of features to be selected to train the model (**R3**). **Representational fidelity.** This step further includes data collection as well as data loading into the analysis tool (Hu et al. 2018). Various data sources were mentioned in the interviews as possible and required sources for the system, including relational databases, multidimensional model databases, and data from CSV or Excel files. *"My data is already preprocessed in a PowerBI Cube, I would like to use this data"* (BU2). Hence, the system should provide the user with the option to upload data or get data directly from the enterprise's data infrastructure (**R4**). Moreover, it is necessary to either select a clean dataset or the data needs to be cleaned to achieve the desired result (Gil et al. 2019). However, business users mentioned that they are often unsure about which data cleaning steps to perform and how to handle incorrect data. To this end, the system should propose data cleaning steps along with treatment options (e.g., for outliers or missing values), and guide them through data cleaning to improve the model (**R5**). Another important task is feature engineering. *"I often ask myself what features correlate with each other and with featured should I choose"* (BU2). To assist the user in performing this task, participants requested that the system should recognize and display correlations between features and compute and display feature importance values (**R6**).

Analysis Requirements

Most modeling techniques have a variety of parameters or settings that can be adjusted to control the ML process (Cao 2018). **Transparent interaction.** Crisan and Fiore-Gartland (2021) suggests that the data preparation and modeling phases form an iterative loop so that data preparation steps can be adjusted to iteratively improve the model's performance (**R7**). Participants also confirmed this requirement in the interviews, *"the possibility of training loops would be great [...] only after training I know if the data quality is bad and what needs to be improved"* (BU1). **Representational fidelity.** Further, since AutoML can take several hours to run, the system should also notify the user when the training is complete (**R8**). Due to the limited analytical skills of business users, participants requested a selection of different ML tasks - the system should then find the best model for the task and apply various model optimization techniques to improve the model (**R9**). An issue often mentioned by the participants after the think-aloud session was that they were not sure about how long they should train the model to get accurate results. *"It would be great to have information on which training time is appropriate"* (BU5). Thus, they requested that the system should recommend an appropriate training time based on the data and the ML task (**R10**).

Deployment Requirements

Business users delineated requirements aimed at monitoring, refining, and deploying the ML model. **Transparent interaction.** Participants requested that the system should output the best model and display the feasibility of the use case in a simple form (**R11**), for instance with a green or red light. *"I need a simple way to show me if the use case is feasible. Green could mean the use case is feasible, go to a data scientist - and red means try it again or let it be"* (BU3). **Representational fidelity.** Moreover, to assist the user in interpreting the results, the system should graphically display the results of the model (**R12**). Graphical representations should include feature importance values or coefficients to explain the learned model, a confusion matrix with colored diagonal for better visualization, and a brief preview of the data including the column with predicted values and probabilities.

General System Requirements

Lastly, some of the requirements mentioned by participants are not specific to a particular phase of the ML process but rather pertain to the entire system. **Transparent interaction.** Since there are several commercial tools for AutoML that support different ML process phases, we asked in the interviews about the option to switch between tools for different tasks. However, we found that users would prefer to stay in a single environment or, if the solution includes multiple tools, they should be able to interact with each other (**R13**). Further, business users mentioned that they would like to have a process pipeline that they can orient on when using the systems (**R14**). *"It would be great if the system provides me an ML process that guides me through the use case"* (BU6). **Representational fidelity.** The primary target users for our proposed system are business users. Participants concluded that targeting the business user is critical for ML feasibility analysis as 🗨 *"you need a lot of process knowledge of the enterprise to work on ML use cases"* (BU4). Further, all participants had differing ML knowledge and varying skills in data analysis. The participants themselves often stated that the system should provide different functions depending on the skill level of the user. 🗨 *"I would let the user decide if he wants to have a basic view or more possibilities"* (BU5). To account for and distinguish between the abilities of the users, the system should offer a basic view and an expert view that provides more functionalities (**R15**).In addition, the system should briefly explain ML concepts in both views to enhance user learning (**R16**).

Design Principles

Overall, we theorize six MRs. These are developed based on the requirements presented (see Table 2). To address the identified MRs, we propose three DPs for effective HGML feasibility analysis. It is important to note that there is an m:n-relationship between MRs and DPs. That is, each MR may be addressed by multiple DPs and each DP may address multiple MRs.

MR	MR Description
1: TI	Enable business users to specify ML tasks, providing explanations of options to support model creation.
2: TI	Display data previews and visualizations to aid business users in exploring the dataset and model.
3: TI	Tailor the interface to the preferences of the business user (e.g., basic and advanced views)
4: TI	Recognize and display correlations between features, and compute feature importance values.
5: RF	Automatically pre-select potential features and suggest suitable numbers for feature selection.
6: RF	Propose data cleaning steps and guide business users through data preparation.
7: RF	Provide options for uploading data or accessing it from enterprise infrastructure.
8: RF	Offer suggestions to help business users increase the model's performance iteratively.

Table 2. Meta requirements (TI = Transparent Interaction, RF = Representational Fidelity)

DP1: User-Tailored HGML. The system should be designed with a strong emphasis on user-centricity (Lee et al. 2020), aiming to enable business users to specify ML tasks with ease and clarity, providing explanations of options to support informed model creation (**MR1**). It should also incorporate a user-friendly interface with data previews and visualizations to aid business users in exploring and understanding the dataset (**MR2**). Additionally, the system should automatically pre-selects potential features and suggests suitable numbers for feature selection, streamlining the process for users (**MR5**). To accommodate business users with varying expertise levels, the interface allows for seamless switching between basic and advanced views, tailoring the interface to their preferences and requirements (**MR3**). This adaptability ensures usability throughout ML tasks. Therefore, we define our first DP as follows: To improve **transparent interaction** and **representational fidelity**, provide a **user-tailored interface** to empower business users with customizable views and seamless tool integration, fostering usability and efficiency in HGML feasibility analysis.

DP2: Transparent Feasibiliy Evaluation. For HGML feasibility analysis, ensuring transparent and interpretable model evaluation is crucial (Meza Martínez et al. 2023). The system should prioritize transparency in model evaluation to enable business users to recognize the value and feasibility of the use case and iteratively improve it to increase its potential. It should present model outputs in a visually intuitive manner, utilizing graphical representations such as feature importance values and confusion matrices (**MR4**). This enhances transparency interaction, empowering business users to easily assess and understand model performance (**MR1**, **MR2**). Transparent feasibility evaluation not only aids business users in understanding the model but also facilitates collaboration with data scientists for productive implementation. Therefore, the system fosters a seamless transition from feasibility analysis to productive implementation, enhancing the overall efficiency of the ML process. We define our second DP as follows: To improve **transparent interaction**, ensure **transparent feasibility evaluation** by presenting model results in an interpretable and visually intuitive manner, enabling business users to assess and understand model performance effectively.

DP3. Educational Support. This principle aligns with the need to facilitate user learning throughout the ML process (Gunklach and Nadj 2023). By offering brief explanations of ML concepts and providing informative tooltips or documentation, the system enhances user understanding and confidence in utilizing its functionalities (**MR5, MR8**). These explanations are designed to help business users learn ML concepts over time and improve step by step while using the system. Additionally, the inclusion of process pipelines guides users through the steps involved in ML feasibility analysis, further supporting their learning (**MR6**). Thus, we define as our third DP: To increase **representational fidelity**, offer **educational support** throughout the ML process, equipping business users with comprehensive resources and guidance to enhance their understanding and proficiency in ML.

User Interface

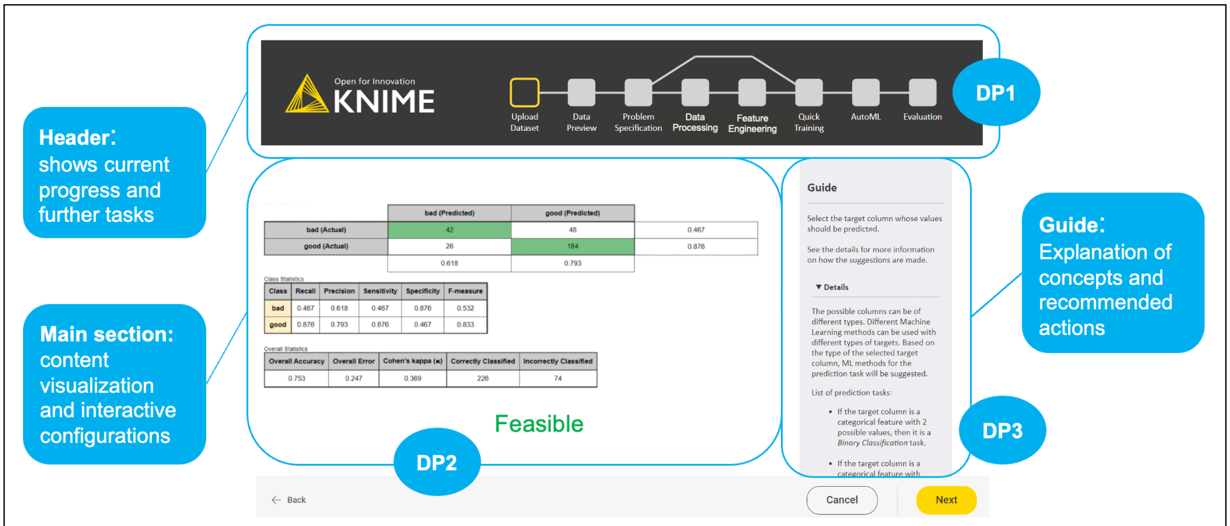

Figure 1. Composition of User Interface

Based on the identified requirements, we implemented *MLFeasi* as a simple web application using KNIME where business users can either upload or query the data, are guided through ML feasibility analysis, and at the end receive the results from AutoML. We chose KNIME due to its seamless integration capabilities with enterprise data sources and other tools, which allowed us to efficiently leverage existing resources and create a user-friendly system for business users to train ML models and analyze the feasibility of their ML use cases (Fillbrunn et al. 2017). The user interface can be divided into three components (see Figure 1): (1) On the top is the **header** showing all steps covered by the system. For easier use, we support business users (**DP1**) in eight steps: upload data, data preview, problem specification, data preprocessing, feature selection, training loop, model training, and model evaluation (see Figure 2). (2) The **main section** provides users with an interactive configuration for each task of the corresponding ML step. For each configuration option, we provide pre-calculated suggestions that the user can choose. Ultimately, the objective of this entire process is for business users to create a ML model for the given use case and then determine its feasibility (**DP2**) (3) The **right sidebar** shows a guide that explains the current ML step including various ML concepts that are relevant to that step, aligning with **DP3**. In the following, we explain each step in detail.

In the **data upload** step (**step 1**), the user can select a data source and upload a CSV file or query data from the enterprise's infrastructure. *MLFeasi* checks the data dimensions to ensure that the data contains at least two columns, the target column, and at least one feature, and also if the CSV was read incorrectly, e.g. due to a wrong separator. Finally, a preview of the data is displayed so that the user can verify whether the data has been read correctly before continuing with the workflow.For **data preview** (**step 2**), users can inspect the input data and perform simple type conversions. To perform data type conversions interactively, the user must specify the columns to be converted and the format patterns. Possible conversions are from number to string, from string to number, and from string to date. *MLFeasi* offers three different ML tasks to cover most use cases in enterprises (Iftikhar and Nordbjerg 2022): Binary classification, multi-class classification, and regression. To **specify** the **problem** (**step 3**), *MLFeasi* requires two inputs from the user: the target column and the ML task. Depending on the type of the selected target column, a suggestion of an appropriate ML task is displayed with brief explanations of the ML method along with some examples. Further, *MLFeasi* plots the distribution of the target column to assist the user in this process.

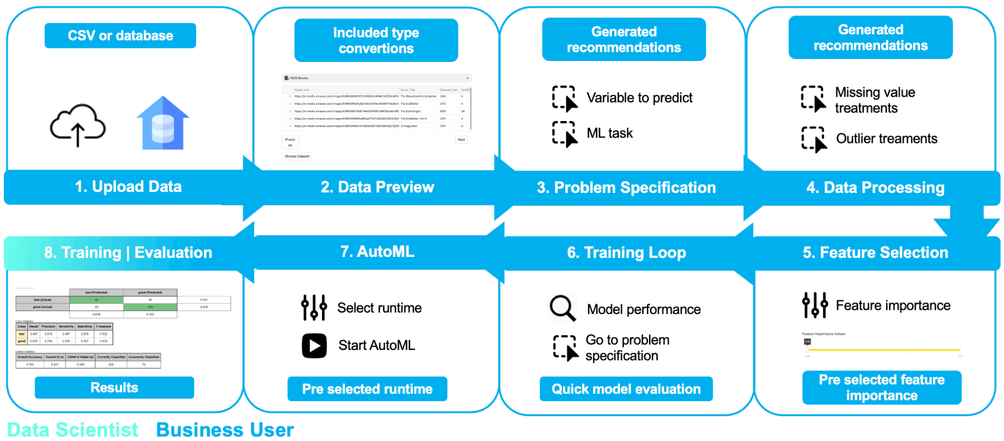

Figure 2. MLFeasi Usage Flow

Data preprocessing (**step 4**) consists of two steps. First, columns are removed that negatively impact model performance or have poor data quality. Second, data quality issues are dealt with, such as missing values or outliers. To remove columns, *MLFeasi* calculates the correlation between features, the percentage of missing values, and the standard deviation. To help the user decide which column to remove, the results are presented in the form of a pre-selected table view. The pre-selection provides a suggestion of which columns to remove and is based on a pre-defined threshold for each criterion. The user can then change the selection to include or exclude columns. Next, the system allows the user to interactively apply different data cleaning steps. The user can decide how to handle missing values and outliers. The available treatment options are for missing values: do nothing, remove the row, or replace it with the mean, median, or a fixed value, and for outliers: do nothing, replace the outlier values with the closest allowed values, or remove the outlier rows. Further, a data explorer displays data statistics for numeric and nominal columns in an interactive view. *MLFeasi* includes the data explorer on this page to allow the user to review the data and statistics to make more informed decisions about the data cleaning steps. The system allows the user to interactively **select** a subset of **features** (**step 5**) based on the feature importance value. To this end, an interactive slider is displayed that allows the user to set a threshold for minimum feature importance. This threshold is then applied to the data and features that do not meet this condition are selected to be excluded from the data. Finally, *MLFeasi* filters the columns according to the user's selection and append the target column back to the data. Moreover, users can manually exclude features. *MLFeasi* displays a data explorer that allows the selection of columns to be excluded. The final set of features will be the unique set resulting from both options. The purpose of the **training loop** (**step 6**) is to provide an initial estimate of the performance of the model based on an example set so that the user can review the data preparation steps before running AutoML. For model evaluation, *MLFeasi* displays a comparison with baseline models, overall statistics, and feature importance values or coefficients. In addition, for regression tasks, residuals are also visualized. The user then has the choice to continue with AutoML or go back and adjust the data preparation settings. For example, the user can test different subsets of features. After each training loop iteration, *MLFeasi* checks if the performance of the model has improved. The number of training loops is currently capped at ten rounds. This limitation is based on our observations that, typically, after ten rounds, further iterations yield no significant improvement and can be considered a waste of time. The system allows the user to set the maximum **model training** (**step 7**) time with a slider that ranges from 15 minutes to a maximum of three hours. The three-hour limit is imposed to manage server resource usage and associated costs efficiently. In addition, the user must enter an email address to be notified when the AutoML run is complete. Based on the training data and user settings, *MlFeasi* leverages H2O AutoML (LeDell and

Poirier 2020) to learn specific model types and obtain the best model. Included modeling techniques are generalized linear models, gradient boosting machines, and stacked ensemble. For **model evaluation** (**step 8**), the idea is to visualize the results in a simple and clear form as well as to support the user in interpreting the results with short explanations. Based on the model evaluation overview, the goal is for the user to assess the feasibility of the ML use case (i.e., to determine whether it is feasible to solve the problem satisfactorily using ML with the available data). If the user determines that the machine learning use case is feasible, there is a handover from the business user to the data scientist (indicated with the green color in Figure 2). At this point, the data scientist takes over for the more technical tasks of model training and further evaluation. To facilitate this transition, *MLFeasi* provides a comprehensive overview of the model performance and feature importance values. Model performance includes the accuracy of the model and a bar plot comparing the performance to baseline classifiers. This information ensures that both the business user and the data scientist are aligned on the feasibility and expected outcomes before moving forward with further refinement or model deployment.

Formative Evaluation

Throughout the development process, we carried out a two-phase evaluation approach. Firstly, a formative evaluation was conducted to gain insights into the interactions between business users, data scientists and *MLFeasi*. After conducting the formative evaluation and gaining valuable insights, we improved *MLFeasi* based on the requirements identified during this phase. Subsequently, we proceeded with a summative evaluation involving business users who dealt with real-world enterprise ML use cases to assess the practicality of *MLFeasi*. In both phases of evaluation, we actively engaged data scientists to capture their valuable perspectives on *MLFeasi*.

Formative Evaluation			
Interview	**Role**	**Business Domain**	**Duration**
BU8	Business user	Purchasing	00:35
BU9	Business user	Logistics	00:39
BU10	Business user	Logistics	00:32
DS1	Data Scientist	Logistics	01:05
DS2	Data Scientist	Manufacturing	00:38
DS3	Data Scientist	Logistics	00:43
DS4	Data Scientist	Logistics	00:37

Table 3. Participants of the Formative Evaluation

Formative evaluations aim at collecting information to improve an artifact (Ritchie et al. 2013). We therefore conducted think-aloud sessions with three business users and four data scientists to evaluate how *MLFeasi*'s design supports business users in assessing the feasibility of their ML use case. For the think-aloud session, we provided participants with an artificial fraud detection use case and tasked them to train a simple ML model to classify the data as fraud and not fraud. We relied on the fraud detection use case because it is simple and generally understandable for non-experts. In our formative evaluation, we aimed to evaluate the functionality and usability of our tool, to improve it in preparation for our final (see Figure 3), summative evaluation round. In the following, we describe the requirements raised.

Findings

The response from **business users** was largely positive, with participants expressing optimism about the system's potential for ML use cases. One participant noted, *"The tool seems to be quite promising for assessing the feasibility of my ML use cases"* (BU9). However, interviews highlighted a need for more explanations and support, as articulated by another participant, *"I need way more explanations of ML concepts on the right side"* (DE10). Users sought comprehensive documentation detailing steps, tasks, and contact information for data scientists. Concerns were raised about CSV file reading, prompting requests for explicit format descriptions and the ability for type conversions. For data preprocessing, participants found the system's guidance useful but suggested pre-selecting constant and ID columns for exclusion. Despite a clear data-cleaning step, concerns emerged about potential data reduction during duplicate removal that it *"could result in an extreme reduction in data if configured incorrectly"* (BU8). Feature selection received positive feedback on relationship information but lacked explicit ML task suggestions. Model evaluation was criticized for its absence of information on feature importance.

Data scientists initially doubted business users' ability to assess the feasibility of ML use cases but after a demonstration they recognized the system's utility. One data scientist remarked, *"I think for a business user's first attempt, it's quite sufficient. [...] I think it can be a useful tool"* (DS1). The system's output was deemed helpful for data scientists to evaluate use cases, as expressed by another participant, *"I have to say quite honestly if a business user comes with the printout of this system, yes, it would help me a lot"* (DS3). Model Evaluation was considered helpful and important, as data scientists could compare *"the results of the system quickly with a baseline model"* (DS2). Positive feedback on data quality assessment was tempered by concerns about false results and system complexity. Participants stressed the importance of upfront business value assessment, with one participant cautioning against *"never-ending stories that cost a lot of money and never lead to a productive applied model" (DS3)*. For that, participants highlighted the importance of assessing the business value of the use case *"Just because it is possible, does not mean that it can be accomplished in a way that will help the company economically in the foreseeable future"* (DS1). As an example, consider a scenario where a company invests heavily in developing an intricate recommendation system for a niche product category, only to realize that the market demand for that category is too small to justify the investment. Assessing the business value upfront could have prevented such a costly endeavor. Data scientists noted the absence of a standard process for evaluating ML use cases, suggesting checklists and predefined questions. Use case owners with process knowledge and some data analysis skills were identified as primary users, aligning with the system's suitability for business users. Data scientists suggested adding additional information on model details to improve the reproducibility of the model.

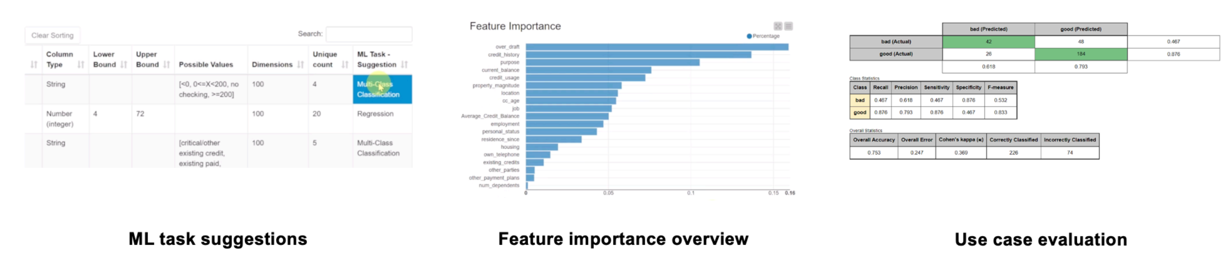

Figure 3. MLFeasi Improvements after Formative Evaluation

Summative Evaluation

A summative evaluation of an intervention or artifact is concerned with its impact and resulting outcomes (Ritchie et al. 2013). As such, we evaluated *MLFeasi*'s effectiveness against the existing implementation of Microsoft's Azure Machine Learning Studio at the industry partner, hereafter referred to as the benchmark. Hereby, we used the improved version of *MLFeasi* (see Figure 3). To ensure a comprehensive evaluation, we conducted two different summative evaluations. The first aimed to measure how business users perceived the guidance provided by *MLFeasi* compared to the benchmark. The second evaluation targeted the effectiveness of *MLFeasi* in supporting HGML feasibility analysis. This involved engaging three business users from different business units, each representing a unique ML use case.

Evaluation of MLFeasi's Guidance for Business Users

In this evaluation, our objective was to determine how business users perceive the guidance provided by *MLFeasi* as opposed to the benchmark tool. For that, we recruited 16 business users, whose ages ranged from 25 to 51 years (with a mean age of 31), from a variety of business domains. We designed the study as a small-scale experimental study, utilizing a 2x2 factorial design to create four distinct groups. Each group was assigned a combination of tasks (fraud detection and price predication) and tools: Group 1 utilized *MLFeasi* for a fraud detection model and then the benchmark for a price prediction task; Group 2 began with *MLFeasi* for price prediction before moving to the benchmark for fraud detection; Group 3 started with the benchmark for fraud detection, followed by *MLFeasi* for price prediction; and Group 4 used the benchmark for price prediction, then *MLFeasi* for fraud detection. This design enabled a comprehensive comparison of the two systems across two distinct ML tasks. After an initial demographic survey, participants were briefed that they would use two different systems to create two ML models. They participated in a tutorial, watching a video and then practicing by constructing a simple ML model, before performing the assigned tasks with the specified system. Post-treatment, they evaluated the system's transparent interaction and representational fidelity using a 1 to 7 Likert scale, based on the items developed by Trieu et al. (2022). On average, participants spent 29 minutes completing the study. The results were analyzed using the Mann-Whitney U test, appropriate for the non-normally distributed data of transparent interaction and representational fidelity. Our analysis (see Figure 4) revealed that participants perceived the transparent interaction with *MLFeasi* (mean = 4.9375) significantly higher ($p = 0.0006$, $U = 231.5$) than the Benchmark (mean = 3.5). Participants also rated rated the representation fidelity of *MLFeasi* (mean = 5.125) significantly higher ($p = 0.0047$, $U = 229.0$) than the benchmark (mean = 3.9375).

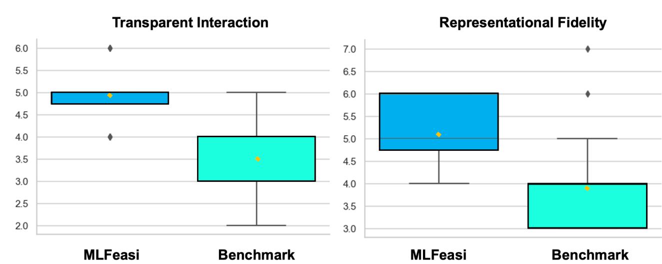

Figure 4. Analysis of First Summative Evaluation

Evaluation of MLFeasi with three real world use cases

In the second summative evaluation, we assessed the effectiveness of *MLFeasi* for HGML feasibility analysis as a more direct measurement of informed action (Burton-Jones and Grange 2013) by recruiting three business users (BU11, BU12, BU13) from different business domains, each encompassing a unique enterprise ML use case. Hereby we choose a two-step approach. Firstly we have business users use *MLFeasi* in think-aloud sessions much like our formative evaluation, now with real-world data and business problems. Afterward, we invited a resident data scientist (DS5) at the industry partner, that would have otherwise been tasked with checking the ML feasibility of these use cases. For the think-aloud session, we relied on real enterprise ML use cases such as (1) predicting savings, (2) predicting prices, and (3) classifying reference prices. We searched for potential ML use cases at our industry partner and asked participants if they could collect and bring the data to the evaluation sessions.

Findings: Business Users

In the **first use case**, the participant focused on analyzing planned versus ad-hoc purchasing data, using regression to predict savings. They found the data quality assessment easy, stating, *"This feature is great, it usually costs me a lot of time to do this"* (BU11). Adjusting the date column and utilizing features for missing values and correlation, the participant appreciated the tool's suggestions. *"It is nice that it does it all for me and then I can make decisions based on the suggestions"* (BU11). For problem specification, the participant read the instructions and guides again and selected the right target column (savings). Then they selected the appropriate ML task without using the suggestions or visualizations provided by the tool. *"This is also a good feature but I already knew that I wanted to train a regression model"* (BU11). Outliers were removed, and the participant noted, *"Playing with outliers is very important for me"* (DE11). Feature selection suggestions were found helpful, with the participant stating, *"I think it's good, I just needed to understand it better"* (BU11). Due to time constraints, the training loop was skipped. Challenges in interpreting feature importance were noted, with the participant expressing, *"Well, the point is, I still do not know which features are most important, but the model seems promising"* (BU11). Restricting the modeling technique to GLM for better insights was recommended, but overall *MLFeasi* confirmed the feasibility of the use case. In the **second use case**, the participant aimed to study the characteristics of different raw materials driving price values through regression. They suggested usability enhancements, emphasizing the need for a *”convert button under each conversion table to enhance usability”* (BU12). Appreciating the system's ML task suggestion, the participant found it helpful, stating, *”I like the suggestion because I would not have known which ML task to choose by just looking at the options”* (BU12). Continuing with data processing, they noticed that some columns had missing values, therefore they followed the system's suggestion and selected columns to be removed that did not exceed the threshold of 70%. Further the participant expressed interest in retaining high-correlation columns. They found the expert view inconsistent and questioned the need for the training loop. *”I wonder if I need this, I think I would rather restart the whole process”* (BU12). Interpretation of model performance led to suggestions for user-friendly explanations. Despite feeling unprepared, the participant found the model promising and deemed the use case feasible, describing the tool as *”very clear, with intuitive steps and helpful explanations”* (BU12). Also, they emphasized that the tool was suitable for self-learning. They suggested more detailed explanations for users with little ML knowledge, including terms like outliers and residuals, and using simplified language. In the **third use case**, the participant aimed to classify reference prices. After uploading the data, they quickly scanned the data and verified that the data was loaded correctly. Next, they identified that the data formats of some columns were wrong, so they converted them from numeric to strings and continued to problem specification. They directly selected the target column to be predicted and the corresponding ML task, skipping the system's instructions and suggestions since they already knew what they intended to predict. However they noted, *”the suggestions are very helpful for users who do not know exactly what they want to predict”* (BU13). Altering the system's pre-selection for suspect columns based on correlation, they found the pre-selection helpful. In data cleaning, they suggested allowing users to define treatment options for each column individually. Despite feasibility concerns with existing data, the participant expressed excitement about the tool's simplicity and guidance, stating, *”I must honestly say I am really excited about the tool, [...] it is kept very simple. You are well instructed and you are well guided through the process of data mining”* (BU13). Positive feedback included a clear training results overview but suggested more explanations for users with less ML knowledge, particularly regarding precision, sensitivity, and specificity calculations, stating, *”describe for some users how the precision, sensitivity, and specificity are calculated”* (BU13).

Findings: Data Scientist

In addition to evaluating how our system supported business users in performing ML feasibility analysis, we aimed to evaluate the results of the system and the support provided to data scientists in analyzing the feasibility of the use case. Therefore, we conducted a semi-structured interview with a data scientist. The data scientist reviewed the results of all three use cases described in the previous subsections and briefly discussed his/her assessment of the use cases' feasibility. Table 4 provides an overview how *MLFeasi*, the benchmark and the data scientist perceived the feasibility of the corresponding use case.

Use Case	Model	MLFeasi	Benchmark	DS5
Predict Savings	Reg.	✓	✗	✓
Predict Prices	Reg.	✓	✓	?
Classify prices	Cla.	✗	✓	✗

Table 4. Feasibility of Use Cases

The data scientist reviewed the results of all three use cases described in the previous subsections and briefly discussed his assessment of the use cases' feasibility. Regarding the **first use case**, the data scientist found it suitable for ML application. They identified areas for improvement in the model, noting the R^2 value's inadequacy in explaining data variance. While acknowledging the feasibility of the use case, they recommended enhancing performance by incorporating time as a categorical variable in future iterations. Despite considering the model useful for prediction, they proposed a shift to binary classification to assess savings presence. For the **second use case** (price driver's analysis), the data scientist deemed it suitable but complex, with potential impact from numerous parameters and data quality. They identified poor model performance due to issues like currency representation and missing values in crucial features. Despite recognizing potential after addressing these issues, they determined the current use case as not feasible based on the available data and results. Concerning the **third use case** (reference price classification), the data scientist expressed uncertainty about its suitability for ML. They highlighted the challenge of interpreting results in a highly unbalanced dataset, suggesting up-sampling methods for better evaluation. However, acknowledging system limitations, they concluded that the use case was not feasible due to difficulties in determining feasibility given the data imbalance and organizational factors influencing decisions.

Discussion

Theoretical Contribution

In DSR, a theoretical contribution is usually considered to take the form of prescribing how a particular solution should be designed to solve a relevant real-world problem - such prescriptive recommendations are typically called DPs (Kuechler and Vaishnavi 2008) and guide the implementation of specific instantiations. Combined with a description of the problem a DP is intended to address, each DP represents a theoretical contribution because it prescribes how to mitigate or solve an associated problem (Peffers et al. 2018). In other words, each combination of MR and DP represents a theoretical relationship - and thus a theoretical contribution - that could be empirically tested through either confirmatory studies or action research (Hevner et al. 2004). This type of knowledge contribution is typically referred to as a prescriptive knowledge contribution (Gregor 2006). In our study, we are drawing on the TEU (Burton-Jones and Grange 2013) as justificatory knowledge and contribute to the body of design knowledge by presenting three DPs for *MLFeasi* that are closely aligned with the foundational aspects of HGML (Gil et al. 2019).

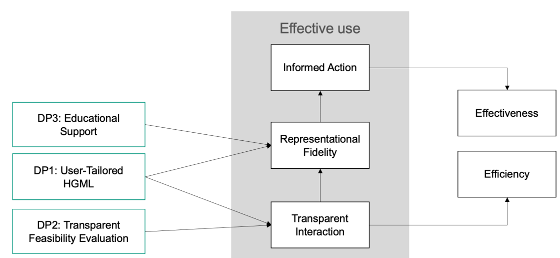

Figure 5. Design Principles in Relation to the Theory of Effective Use

The principle of User-Tailored HGML (**DP1**) ensures that business users can intuitively engage with *MLFeasi*, reducing cognitive load by fostering transparent interaction and enabling focused utilization of domain knowledge in the ML process, thereby improving both *transparent interaction* and *representational fidelity*. Transparent Feasibility Evaluation (**DP2**) ensures that business users can accurately assess model performance, enhancing *transparent interaction*. Also, Educational Support (**DP3**), enhances *representational fidelity* by providing explanations and process pipelines, enabling business users to deepen their understanding of ML concepts and make more informed decisions during the feasibility analysis process. Through a careful research and development process, we conducted a formative evaluation with three business users and four data scientists, followed by a comprehensive summative evaluation that included two distinct components: a small-scale experiment with 16 business users, and the application of *MLFeasi* to three real-world enterprise ML use cases. We demonstrated in the first summative evaluation that by implementing these DPs, business users reported significantly higher transparent interaction and representational fidelity than using the benchmark. Moreover, the effectiveness of our DPs was further validated in the real-world enterprise settings, where *MLFeasi* not only demonstrated robustness and adaptability but also empowered business users with limited ML knowledge to actively participate and contribute to the ML process.

Novelty and Practical Relevance

During our interviews, we identified that due to varying levels of analytical skills of business users (Gil et al. 2019), they had different requirements for *MLFeasi* and needed different support for specific ML tasks. We identified that there is a trade-off between providing a high degree of guidance and limiting autonomy for savvy users versus a low degree of guidance with more advanced features, making the system less usable for less-skilled users (Lee et al. 2020). We received feedback from many participants that the explanations and guidance provided by the system are helpful in learning the subject of ML. An implication from the evaluation of *MLFeasi* is that business users are able to evaluate their use cases with the system, but still have difficulties understanding ML concepts (Santos et al. 2019). In order for business users to take full advantage of the features provided, we suggest offering a short training session for the system and an introduction to the ML process first. We also found that for data scientists to implement the use case with an output of *MLFeasi*, information about the objective of the use case and the current process is relevant. Hence, we recommend collaboration between business users and data scientists, with the data scientist taking an advisory role.

We also found that the interplay between business users and data scientists often reveals a deep conflict of goals. Business users seek actionable insights and prefer simple, interpretable models that can be used immediately for decision support. In contrast, data scientists prioritize model accuracy, sometimes at the expense of accessibility or immediate business applicability. This inherent conflict underscores a significant challenge: while business users prioritize models that facilitate quick and informed decisions, data scientists prioritize technical excellence and optimization, sometimes at the expense of practical business needs (Gerhart et al. 2023). Our interviews with data scientists revealed their commitment to rigorous model performance standards, following the concept of agents acting in ways they perceive as beneficial within fair and reciprocal norms (Bosse and Phillips 2016). This discrepancy is often evident during collaborative project phases, particularly in the business and data understanding and deployment phases, where close collaboration is essential. Interestingly, we found that most conflicts arise during the analysis phase, despite its technical complexity. Addressing this conflict requires enhanced communication between business users and data scientists to foster mutual understanding of their respective goals and constraints. Cultivating empathy and a shared commitment to aligning technical and business goals ensures that the models developed not only exhibit technical rigor, but also serve as practical tools for strategic decision-making, bridging the gap between ML model excellence and business utility.

Limitations and Future Research

Our study has several limitations. First, *MLFeasi* is designed specifically for business users, who have varied analytical skills and often present conflicting requirements. To address this, we opted to include more advanced features in an expert dropdown list that users can toggle, allowing for customization while keeping the interface uncluttered. Second, following Iftikhar and Nordbjerg (2022), our focus was solely on supervised learning, which covers over 80% of enterprise ML use cases as indicated by business users. However, our summative evaluation revealed the relevance of unsupervised learning methods like clustering for gaining insights, especially for users less familiar with the data. This suggests potential future research to support unsupervised learning. Third, while business users found the model performance information sufficient, they expressed a need for deeper understanding of how models make decisions. This highlights an opportunity to incorporate explainable AI techniques, such as visual model explanations, to enhance system transparency and build user trust (Meza Martínez et al. 2023). Additionally, while our study covered domains such as purchasing, manufacturing, logistics, and controlling—areas that are integral to an organization's operations and often involve complex machine learning use cases—we did not explicitly examine domains like marketing or recruiting. As the design principles of *MLFeasi* are grounded in the Effective Use Theory and intended to be generalizable, future research could explore the applicability and effectiveness of these principles in other business domains to further validate and extend our findings. Lastly, while we used the feasibility of the use case as a proxy for informed action (Burton-Jones and Grange 2013) in the second summative evaluation (see Table 4), this was not quantitatively measured in our first summative evaluation, which we note as a minor limitation.

Conclusion

In this paper, we presented *MLFeasi*, a system that allows business users to evaluate the feasibility of their ML use cases and relieves data scientists by helping them implement use cases more efficiently and effectively. We structured *MLFeasi* along eight steps: upload data, data preview, problem specification, data preprocessing, feature selection, training loop, model training, and evaluation. By offering three DPs (1) user-tailored HGML, (2) transparent feasibility evaluation, and (3) educational support, we contribute to the effective use of HGML feasibility analysis by empowering business users with limited ML experience, to contribute their valuable domain knowledge. Our findings illustrate that *MLFeasi* proves to be a valuable tool for conducting ML feasibility analysis, further fostering collaborations between business users and data scientists. Further, our practical implications underscore *MLFeasi*'s potential to democratize ML tasks, improve project efficiency, and drive innovation in data-driven decision-making across organizations.

References

Abbasi, A. et al. 2016. "Big Data Research in Information Systems: Toward an Inclusive Research Agenda," *Journal of the Association for Information Systems* (17:2) 2016, pp. I–XXXII.

Barbudo, R., Ventura, S., and Romero, J. R. 2023. "Eight years of AutoML: categorisation, review and trends," *Knowledge and Information Systems* (65:12) 2023, pp. 5097–5149.

Berkin, A., Aerts, W., and Van Caneghem, T. 2023. "Feasibility analysis of machine learning for performance-related attributional statements," *International Journal of Accounting Information Systems* (48) 2023, p. 100597.

Berthold, M. R. et al. 2008. "KNIME - the Konstanz information miner," (11:1).

Bogl, M. et al. 2013. "Visual Analytics for Model Selection in Time Series Analysis," *IEEE Transactions on Visualization and Computer Graphics* (19:12) 2013, pp. 2237–2246.

Bosse, D. A. and Phillips, R. A. 2016. "Agency Theory and Bounded Self-Interest," *Academy of Management Review* (41:2) 2016, pp. 276–297.

Bourcevet, A. et al. 2019. "Guided Machine Learning for Business Users," in: *Humanizing Technology for a Sustainable Society,* Univresity of Maribor Press, pp. 257–270.

Burton-Jones, A. and Grange, C. 2013. "From Use to Effective Use: A Representation Theory Perspective," *Information Systems Research* (24:3) 2013, pp. 632–658.

Burton-Jones, A. and Straub, D. W. 2006. "Reconceptualizing System Usage: An Approach and Empirical Test," *Information Systems Research* (17:3) 2006, pp. 228–246.

Cao, L. 2018. "Data Science: A Comprehensive Overview," *ACM Computing Surveys* (50:3) 2018, pp. 1–42.

Cashman, D. et al. 2019. "A User-based Visual Analytics Workflow for Exploratory Model Analysis," *Computer Graphics Forum* (38:3) 2019, pp. 185–199.

Charmaz, K. and Bryant, A. 2010. "The SAGE handbook of grounded theory: Paperback edition," *The Sage handbook of grounded theory* (). Publisher: Sage, pp. 1–656.

Crisan, A. and Fiore-Gartland, B. 2021. "Fits and Starts: Enterprise Use of AutoML and the Role of Humans in the Loop," in: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems,* ACM, 2021, pp. 1–15.

Falkner, S., Klein, A., and Hutter, F. 2018. "BOHB: Robust and Efficient Hyperparameter Optimization at Scale," *International Conference on Machine Learning* ().

Feurer, M. et al. 2015. "Efficient and Robust Automated Machine Learning," *Advances in neural information processing systems 28* ().

Fillbrunn, A. et al. 2017. "KNIME for reproducible cross-domain analysis of life science data," *Journal of Biotechnology* (261) 2017, pp. 149–156.

Gartner 2016. *Gartner Says Worldwide Business Intelligence and Analytics Market to Reach $16.9 Billion in 2016*.

Gerhart, N., Torres, R., and Giddens, L. 2023. "Challenges in the Model Development Process: Discussions with Data Scientists," *Communications of the Association for Information Systems* (53:1), pp. 591–611.

Gil, Y. et al. 2019. "Towards human-guided machine learning," in: *Proceedings of the 24th International Conference on Intelligent User Interfaces,* Marina del Ray California: ACM, 2019, pp. 614–624.

Gregor 2006. "The Nature of Theory in Information Systems," *MIS Quarterly* (30:3), p. 611.

Gregor, S. and Hevner, A. R. 2013. "Positioning and Presenting Design Science Research for Maximum Impact," *MIS Quarterly* (37:2) 2013, pp. 337–355.
Gregor, S. and Jones, D. 2007. "The anatomy of a design theory," *Journal of the Association for Information Systems* (8:5), p. 1.
Gunklach, J. and Nadj, M. 2023. "Guidance in Business Intelligence & Analytics Systems: A Review and Research Agenda," in: *ECIS 2023 Research Papers. 329.*
Hevner et al. 2004. "Design Science in Information Systems Research," *MIS Quarterly* (28:1), p. 75.
Hu, K., Orghian, D., and Hidalgo, C. 2018. "DIVE: A Mixed-Initiative System Supporting Integrated Data Exploration Workflows," in: *Proceedings of the Workshop on Human-In-the-Loop Data Analytics,* ACM, 2018, pp. 1–7.
Iftikhar, N. and Nordbjerg, F. E. 2022. "Implementing Machine Learning in Small and Medium-Sized Manufacturing Enterprises," in: *Towards Sustainable Customization: Bridging Smart Products and Manufacturing Systems,* pp. 448–456.
Kuechler, B. and Vaishnavi, V. 2008. "On theory development in design science research: anatomy of a research project," *European Journal of Information Systems* (17:5) 2008, pp. 489–504.
LeDell, E. and Poirier, S. 2020. "H2O AutoML: Scalable Automatic Machine Learning," *7th ICML Workshop on Automated Machine Learning* ().
Lee, D. J.-L. et al. 2020. "A Human-in-the-loop Perspective on AutoML: Milestones and the Road Ahead," *IEEE Data Engineering Bulletin (2020)* (), p. 12.
Meza Martínez, M. A. et al. 2023. "Does this Explanation Help? Designing Local Model-agnostic Explanation Representations and an Experimental Evaluation Using Eye-tracking Technology," *ACM Transactions on Interactive Intelligent Systems* (13:4) 2023, pp. 1–47.
Michalczyk, S. et al. 2021. "Demystifying Job Roles in Data Science: A Text Mining Approach," *ECIS 2021 Research Papers* () 2021.
Microsoft 2023a. *Azure Machine Learning: Machine-Learning-as-a-Service | Microsoft Azure.*
Microsoft 2023b. *What is the Team Data Science Process?*
Möller, F. et al. 2022. "Unveiling the Cloak: Kernel Theory Use in Design Science Research," ().
Mosqueira-Rey, E. et al. 2023. "Human-in-the-loop machine learning: a state of the art," *Artificial Intelligence Review* (56:4) 2023, pp. 3005–3054.
Peffers, K., Tuunanen, T., and Niehaves, B. 2018. "Design science research genres: introduction to the special issue on exemplars and criteria for applicable design science research," *European Journal of Information Systems* (27:2) 2018, pp. 129–139.
Pinhanez, C. 2019. *Machine Teaching by Domain Experts: Towards More Humane,Inclusive, and Intelligent Machine Learning Systems.* arXiv:1908.08931 [cs]. 2019.
Ritchie, J. et al. 2013. *Qualitative research practice: A guide for social science students and researchers,* sage.
Sambasivan, N. and Veeraraghavan, R. 2022. "The Deskilling of Domain Expertise in AI Development," in: *CHI Conference on Human Factors in Computing Systems,* New Orleans LA USA: ACM, 2022, pp. 1–14.
Santos, A. et al. 2019. "Visus: An Interactive System for Automatic Machine Learning Model Building and Curation," in: *Proceedings of the Workshop on Human-In-the-Loop Data Analytics - HILDA'19,* ACM Press, pp. 1–7.
Silver, M. S. 1991. "Decisional Guidance for Computer-Based Decision Support," *MIS Quarterly* (15:1) 1991, p. 105.
Sun, D. et al. 2020. "DFSeer: A Visual Analytics Approach to Facilitate Model Selection for Demand Forecasting," in: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems,* ACM, 2020, pp. 1–13.
Sun, Y. et al. 2023. "AutoML in The Wild: Obstacles, Workarounds, and Expectations," in: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems,* 2023, pp. 1–15.
Trieu, V.-H. et al. 2022. "Applying and Extending the Theory of Effective Use in a Business Intelligence Context," *MIS Quarterly* (46:1) 2022, pp. 645–678.
Zöller, M.-A. and Huber, M. F. 2021. "Benchmark and Survey of Automated Machine Learning Frameworks," *Journal of Artificial Intelligence Research* (70) 2021, pp. 409–472.
Zöller, M.-A. et al. 2023. "XAutoML: A Visual Analytics Tool for Understanding and Validating Automated Machine Learning," *ACM Transactions on Interactive Intelligent Systems* (13:4) 2023, pp. 1–39.