



Supporting SME companies in mapping out AI potential: a finnish AI development case

Pouya Jafarzadeh¹ · Tanja Vähämäki¹ · Paavo Nevalainen¹ · Antti Tuomisto² · Jukka Heikkonen¹

Accepted: 4 July 2024
© The Author(s) 2024

Abstract

Products and services relying upon Artificial Intelligence (AI) have moved from mere concepts to reality. However, challenges still exist in applying AI technologies to traditional industrial and service enterprises. Two central problems are a proper understanding of the opportunities AI could bring to the business processes and making the business logic and data sources transparent to AI experts. As small and medium-sized enterprises (SMEs) are considered the economic backbone of many countries, this paper studies how to support SMEs in understanding the potential of AI in their business and how to prepare their data and requirements for a possible AI project. For this purpose, we first proposed the Cross-Industry Standard Process for Data Mining (CRISP-DM) an industry-proven way to apply AI solutions. The weight was in early business and data understanding. Then, we performed data visualization and developed some machine learning methods for 11 SMEs in South-western Finland as case studies to get more ideas for improving their business using AI. Two surveys probed the possible changes in AI practises of companies.

Keywords Artificial intelligence · Business · Development · collaboration · Small and medium sized enterprises

JEL Classification O330

Pouya Jafarzadeh and Tanja Vähämäki contributed equally to this work.

✉ Pouya Jafarzadeh
poujaf@utu.fi

Tanja Vähämäki
tmvaha@utu.fi

Paavo Nevalainen
ptneva@utu.fi

Antti Tuomisto
anttuo@utu.fi

Jukka Heikkonen
jukhei@utu.fi

¹ Department of computing, University of Turku, 20014 Turku, Finland

² Department of Management and Entrepreneurship, University of Turku, 20014 Turku, Finland

1 Introduction

AI and data analytics can help companies better understand customers, develop processes, and produce new services and products. AI is a wide discipline that simulates human skills such as reasoning, learning, and problem-solving. It has gained the attention of developers and providers of various real-time business applications during the last few years. It is a growing trend in many companies such as Google and Facebook (Abadi et al., 2016; Paszke et al., 2017). In these companies, a lot of data has accumulated in various information systems over the decades.

Traditional business analytics based on statistical methods are not necessarily sufficient for making business management decisions when the uncertainties of the business environment increase. In addition, the accumulated data may contain signals that remain unutilized by traditional means. AI research has developed increasingly reliable and better applicable tools and methods that offer the opportunity to model and evaluate the benefits, risks and level of uncertainty of future operations. AI is usually used for decision support (Martínez & Rodríguez, 2015), customer support and relationship management (Bahari & Elayidom, 2015), process acceleration (Nassiri-Mofakham, 2014; Guzella & Caminhas, 2009; Turchi et al., 2009), knowledge management (Nemati et al., 2002) and predictive maintenance (Schlögl et al., 2019). Generally, AI systems can provide a solution in a better and quicker way to significantly improve the overall business. To design these systems, we first need to know the business target of the solution and what kind of opportunities data and AI tools can offer. This is especially challenging for small and medium-sized enterprises (SMEs) with limited resources and skills to understand the potential of data and AI (Schlögl et al., 2019). The European Commission defines SMEs as having less than 250 employees and less than 50 million euros in turnover or 43 million euros in revenue (<https://ec.europa.eu/growth/smes/sme-definition/en/>). SMEs must understand the essential tasks in their business systems to develop suitable AI tools. In addition, they need to integrate AI tools with their workforce to increase productivity effectively. Also, companies must prepare themselves to be able to engage in cooperation with AI experts (either employed or consultants), who may not know the business domain or central concepts of the business sector in question. In Westenberger et al. (2022), they have studied the critical factors that can cause AI project failure. The lack of proper data and machine learning techniques are critical factors in the failure of AI projects. Introducing AI to SMEs is usually not easy due to many issues, such as uncertainty about the benefits, the necessary initial investment, changes in the organization's operating models and business management level, and uncertainty about its effect on the preservation of jobs among employees.

This paper studies how to support SMEs in utilizing data and AI to foster business value. The main reason is that 99% of all those employed in the EU are in an SME; therefore, SMEs are the backbone of the European economy (<https://ec.europa.eu/growth/smes/sme-definition/en/>). This study is a part of a European project called "AI Champion". The project's primary goal is to give SMEs a general view and idea for applying AI tools in their business. A big challenge for introducing AI in companies is seeing the added value and benefit of AI in one's own company and operations (oy, 2019). For this purpose, SMEs' challenges are considered, and suitable solutions to lower these barriers are examined. According to the Cross Industry Standard Process for Data Mining (CRISP-DM), we defined our steps during this project. We selected 11 SMEs in southwestern Finland who had plans to develop their ability to adopt AI in the company. The companies are of different sizes, operate in various sectors, and differ in AI maturity levels. The available data

varies, from sensor data to natural language or data collected from sales registers or customer questionnaires. We arranged several meetings with the AI Champions and representatives of other SMEs. We first discussed the business goals of the company, the available data, and the possibilities of using AI methods. In most cases, examples of data can also be found, and data visualization and machine learning (ML) methods are applied. In addition, we studied several AI methods for different areas, such as clustering analysis, natural language processing and regression models. By participating in the AI Champion project, the company gets support in understanding the possibilities of using its own data and AI methods, insight into starting AI projects, and issues related to the implementation and adoption of AI in the company.

In addition, a tailored AI Canvas (Thi e, 2021) is introduced to the AI Champions of the SMEs to learn what to take into account when starting an AI project and adopting AI in the company's business processes. For example, the SMEs are encouraged to proceed with a proof of concept or individually suggested next steps after this project. This project helps SMEs understand the concept of AI, how AI methods can be used in their business, and essential perspectives to consider when going forward with their ideas of utilizing AI in the company. This paper summarized all the results that are achieved during the project. The major research questions in this paper are mentioned below:

- How to support SMEs to refine data and understand the potential of AI in their business?
- How to help SMEs to address the challenges of adopting AI and to clarify next steps?
- What are the barriers SMEs must overcome to adopt AI solutions?

The paper is organized as follows. Section 2 describes the most relevant works on the challenges of adopting AI in SMEs. The study procedure and environment is defined in Sect. 3. Study results and questionnaire feedback are presented in Sect. 4 followed by conclusions in Sect. 6.

2 Related work

2.1 Challenges of adopting AI in SMEs

AI and ML techniques are used in many areas, including predictive analysis (Gandomi & Haider, 2015), social commerce (Zhou et al., 2013), finance (Moro et al., 2015), and business intelligence (Chen, 2012). However, there are still different challenges to using AI in SMEs, such as lack of data, technology, and budget. These studies (Jung et al., 2020; Schl ogl et al., 2019) have described various influential factors for these challenges. They found that SMEs face more challenges in using AI than big companies because SMEs lack the resources (financial and human) to screen the market and efficiently implement the right tool. To prove this, they (Schl ogl et al., 2019) interviewed 19 European companies to investigate the current use of AI applications.

Mehta and Rajendran (2020) found that the weak financial condition of SMEs, limited awareness of AI benefits, and limited availability of skilled human resources are significant barriers to AI adoption. Bhalerao et al. (2022) also explored these barriers and highlighted the benefits of AI adoption in SMEs.

In Drydakis (2022), the business risks faced by SMEs due to the COVID-19 pandemic were mitigated through the use of AI applications in London, England. Their outcomes proved robust across different specifications such as enterprise size, turnover, and years of operation, indicating that AI applications can help SMEs adapt to unprecedented conditions during the COVID-19 pandemic. Another work (Lu et al., 2022) shows how the deployment of AI technology can help SMEs in the post-pandemic era, including impacts on work, organizations, and performance. A survey of 98 SMEs in north-western Germany revealed that they are concerned that using AI is a very expensive and lengthy process (Christoph Szedlak & Leyendecker, 2020). Instead, SMEs increasingly rely on AI-as-a-service and prefer to use cloud-based solutions. This study shows that poor data quality is a major challenge for the adoption of ML in SMEs (Bauer et al., 2020).

DeStefano et al. (2022) found that firm size and the use of intangible assets are significant determinants of AI adoption in SMEs, along with other digital technologies like big data, cloud computing, and the Internet of Things. However, most small and medium-sized companies do not fall into the general AI adoption category in the literature, which is often biased towards data-intensive businesses such as manufacturing, accounting, and maintenance.

Jöhnk et al. (2021) noted that AI readiness in companies is in turmoil, and while all should adopt AI, including SMEs, the practicalities of development steps and their integration into daily activities are challenging. Respondent E09 described the process of AI adoption as requiring small use cases to demonstrate advantages, build competencies, and spread the technology throughout the organization. Bauer et al. (2020) emphasized the importance of an agile and collaborative approach for the initial steps of AI adoption in SMEs.

In specific lines of business, such as manufacturing and other data-intensive industries like accounting, the essential role of data favors AI adoption and appropriate data analytics tools. These businesses also see a natural emergence of AI service providers and partners. For instance, the market size of AI in the manufacturing sector is expected to exceed USD 2 billion by 2025, with an average annual growth rate of more than 40% from 2019 (Deloitte, 2020; McKinsey, 2022). AI has become pivotal in boosting productivity and innovation in manufacturing, driven by smart production and advanced analytics applications (Pwc's global, 2024).

2.2 Machine learning in SMEs

A qualitative empirical study (Bauer et al., 2020) shows that ML methods are rapidly growing in SMEs by identifying the success factors. Generally, ML is used in many applications for solving different tasks such as forecasting, classification (Susto et al., 2015; Farahnakian & Heikkonen, 2018), and optimization (Zhou et al., 2018) in these companies. In Susto et al. (2015), they used a classification method (decision rule) for predictive maintenance. The effectiveness of their proposed method is demonstrated using a benchmark semiconductor manufacturing maintenance problem. A multiple classifier approach is presented in Susto et al. (2015) for predictive maintenance which allows dynamic decision rules to be adopted for maintenance management. In Huang and Rust (2020), they developed a strategic AI framework for marketing planning. Their framework includes three steps: marketing research to understand the market and competitors, marketing strategy to develop a plan, and marketing action to execute the respective strategy. Clustering approaches such as K-means can group the customer data for targeted customer services.

In Munde and Mishra (2022), they proposed a decision tree and a random forest in the bankruptcy prediction of SMEs. Another ML method, Support Vector Machine (SVM), is widely used in many prediction tasks, such as estimating students' performance in final examinations (Al-Shehri et al., 2017).

Recently, Deep Learning (DL) (Goodfellow et al., 2016) has been widely used in SMEs. However, it requires high computational power, teamwork for development, cooperation among experts from various fields, often a well-orchestrated accumulation of large datasets, and a keen awareness of related solutions and possibilities for transfer learning. In Farahnakian et al. (2021), they show the application of DL for automatically detecting warehouse defects. In Farahnakian and Heikkonen (2018), they show the applicability of a DL model (auto-encoder) for intrusion detection. Their results also show that the DL model can outperform the ML methods for this application. In Zhou et al. (2018), a deep reinforcement learning method is proposed for molecule optimization by combining chemistry domain knowledge with reinforcement learning.

3 Study procedure

The practical level of the study consisted of an intervention with approx. 18 companies at Western Finland over the years 2019–2020. A survey was conducted in Spring 2024 (18 months after the project end) to measure perceived changes in companies and their AI approaches. The following treatise is about the structure and content of the initial intervention. The survey is described in Sect. 4.3.

We proposed the Cross-Industry Standard Process for Data Mining [CRISP-DM (Chapman et al., 2000)] as an industry-proven way to apply AI solutions. The main objective of the AI Champions is to offer a practical learning experience on using the CRISP-DM process and help them to understand the potential of data and AI tools in the company. Although they might have limited knowledge of AI, incomplete available data, and scarce resources. In addition, we provide the required information about further steps to support the SMEs to move towards implementing an AI solution and adopting AI in their business after our project. For this purpose, we first discussed their business objectives, available data, and potential AI tools with the AI Champions. Then, we developed data visualization and ML methods to see if they had the proper data for SMEs' business goals. Otherwise, we will provide information about data collection and preparation for further AI solutions. The development ideas and utilization of data and AI tools are brainstormed further when iterating the steps in the CRISP-DM process. Therefore, the focus of the early steps is to understand business objectives and data. We then applied the steps of data preparation, modeling, and evaluation to show the initial results and the possible AI solution for the AI Champions. The deployment step of CRISP-DM is not applied in SEMs as this project does not have a plan to implement a model in the production environments of the SMEs. In this section, we summarize CRISP-DM steps based on our study in the AI Champion project. Figure 1 details the steps of two example case studies.

3.1 Business understanding

In the first step, we find information about the company's business, development ideas and available data through a questionnaire. Some main questions to clarify the business objectives of the development ideas are:

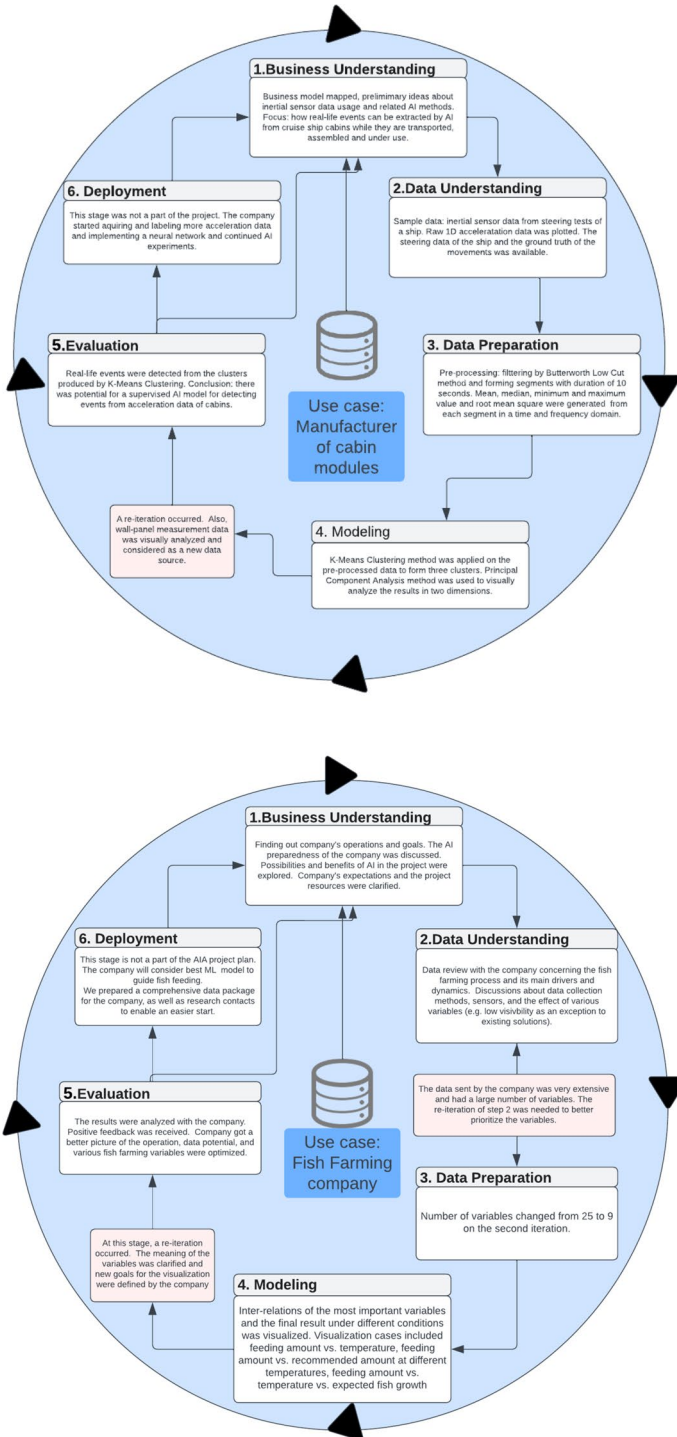


Fig. 1 The CRISP-DM for our case studies in **a** cabin manufacture and **b** fish farming company

1. What outcome(s) do you want to see?
2. What are the main obstacles to achieving these outcomes?
3. How can AI help your business move towards success?
4. How will you measure success?
5. What data do you have today, and what additional data do you need?

The answers to these questions help us to define the business goals of the development ideas and then step towards brainstorming suitable AI solutions for the company. In addition, we gave them essential information about AI methods that could be applied to extract valuable knowledge from their data. Therefore, the AI Champions learn about the potential of their data and AI tools. A project plan is unnecessary because the aim is not to start implementing a solution. It would be required later if the company decides to go forward with the idea and chooses to implement, for example, a Proof of Concept of the AI solution.

3.2 Data understanding

This step is performed if suitable data is available for further processing. We chose a data sample from each company in order to visualize and analyze. We also demonstrate how we can generate valuable information from the data with AI tools. A large quantity of data is not needed for the demonstration purposes. Querying and visualization techniques produce a light exploration analysis of the sample data. It is interpreted with the AI Champion to see the overall picture of the data and find possible solutions. It may be needed to recollect data, select a new sample data, or go back to the previous step to further brainstorm the business development idea. The quality of the data is discussed by asking questions such as:

1. Is the data complete (does it cover enough cases required)?
2. Is it correct, or does it contain errors or noise and if there are errors, how common are they?
3. Are there any missing values in the data?

3.3 Data preparation

We first decide on the sample data that should be used for making demonstrations. Then, we clean the data based on the light exploration report from the previous step. We also select and generate the features which are suitable for implementing the selected AI methods.

3.4 Modeling

This step applies the selected ML method to the prepared sample data to create one or more models. A procedure or mechanism is generated to test the quality and validity of the models. Finally, we revise the parameter settings and tune the models for the next run according to the model assessment until the model is good enough for the demonstrations.

3.5 Evaluation

This step assesses the degree to which the model is sufficient for demonstrating purposes. Based on the assessment and process review results, the project team recommends how to proceed.

In some cases, a recommendation to go back to the first step (business understanding) or step "data understanding" is needed. Therefore, more discussions with the AI Champion are required in order to brainstorm the development ideas further, discuss possibilities for acquiring new data, or clarify or change the content of the sample data.

After the iterations are completed, the final results of the demonstrations are discussed with the AI Champion. The aim is to increase understanding the opportunities of utilizing data and AI and answer the following questions:

1. Is the quality and relevance of the data what was expected?
2. Was there a signal in the data?
3. Was interesting information extracted from the data?
4. Is there potential to use data and AI to proceed with the business case?

3.6 Deployment

This step takes the evaluation results and determines a strategy for continuing after the AI Champion project. The deployment is not a target of the project but a duty of the companies. We make individual recommendations on how to move forward for each company. Suitable ways to increase AI competence and acquire resources are suggested, such as increasing AI competence at a general level, e.g., by training, following the AI field, and networking. With more mature AI potential, cooperating with universities, applying for funding for AI initiatives, finding a suitable AI partner, or hiring a trainee or full-time resources may be recommended.

In addition, we suggest that the SMEs use the AI Champion Canvas, a high-level strategic framework for defining key questions and feasibility challenges regarding the initial adoption of AI. The AI Champion Canvas shown in Fig. 2 is developed in the AI Champion project based on the standard AI Canvases (Thi e, 2021). Detailed questions suitable for addressing both business and technical issues are included. They help SMEs brainstorm potential business cases further and consider important issues when planning to start an AI project and adopt AI in the company.

4 Results

In this project, data visualization or ML modelling tasks are developed for the studied companies depending on their available data and own requirements and wishes. Figure 3 shows the list of companies participating in the AI Champion project and their properties, including business goals, size, number of employees, available data type, current implemented methods or visualization, the project results and follow-up plan.

Table in Fig. 3 is mainly a backdrop for statistics like number of meetings, work hours needed per visualization and data pre-processing, estimated shift in productivity etc. business measure, and the perceived difficulty of the AI problem. Together the characterization of companies and the observed statistics listed above (numerical

AI Champion Canvas		Filler(s) of the form: Company: Information from the form can be shared with other AI ambassadors: Yes / Not yet
<p>Opportunity: Why are we doing this? Description of the need and business case. What would be needed for more efficiency, savings, more income? Or "just" information for decision-making for further measures: what the data can tell about the business, customers, resources, etc.</p> <p>Users: Who needs this? Who uses/needs new information in the company? Whose work does it affect? Who is responsible for decision-making using new information (regarding business and/or further development of an artificial intelligence project)</p> <p>Strategy: Why us? Why would we do this? Why artificial intelligence and data? What competitive/alternative ways are there to develop a competitive advantage?</p>		<p>Solution: What is it? A preliminary idea about acquiring new data or an artificial intelligence solution to be developed/new utilization of data in work and decision-making, and their effects on the workflow</p> <p>Data: What data is entered into the model? What own data is used? What about outside/open? Is the data already ready? How can or should it be grouped (which variables to use, e.g. day of the week, temperature, time, etc.)?</p>
<p>Operating methods/processes: What other changes are needed? Is there enough data and of sufficient quality/accurate? Who produces and monitors it? How will production, services or organization change? What about security, GDPR, data protection and security?</p>	<p>Changes & learning: How do we build it? From where IT/artificial intelligence partner or purchasing service? Where do you get your expertise, time and money? Will there be more work, will something be left out?</p>	<p>Criteria for success: How do we know it works? What are the measures of success in the initial stages of doing it yourself (idea stage, experiments/learning, pilots/Proof-Of-Concepts). What are the testing criteria for preliminary results?</p>

Fig. 2 AI Champion Canvas

ID	Company Field of Business	Size	Employees	Targets	Type Of Available Data	Methods, Models And Visualizations	Results	Follow-up Plan
A	Cruise ship's cabin design and manufacture	Medium	162	Identification of ship's cabin vibration in different situations	Acceleration sensor data from both collected and annotated transport data and non-annotated on-board data	Cabin sensor data is grouped into different group based on similarity by using k-means	Finding various vibrations to the cabin and related actual events in the environment	Identification of environmental events using AI
B	Software development company for develop smart technology (IoT sensors and gadgets) to collect real-time data from devices and environment	Small	11	Status monitoring of district heating wells with IoT sensors for detection or prediction of leak. Using available data for evolution of (leaked) situations under different condition for different types of wells (size, structure, environment, etc.) Providing valuable product / service development information	Temperature, humidity and water surface level	Data preparation for measured values such as segmentation and feature generation (temperature and humidity changes), k-means clustering for grouping different types of wells	Identifying anomalous observations	Predicting or detecting a leak using AI
C	The company produce healthy food for consumers sustainably, effectively and profitably, promoting animal welfare.	Small	33	Optimization of feeding in fish ponds Analyzing the current feeding planning with the collected data	Feeding volume, recommended food amounts, water temperatures in different rearing pools, water oxygen values, location, feeding recommendations, etc.	Visualizations for amount of feeding vs. temperature, amount of feeding vs. recommended amount, differences between temperatures and oxygen levels amount of feeding vs. temperature vs. expected growth of fish	The visualizations can show the effects of various variables (temperature, oxygen, weight, etc.) on the feeding of the fish and their relative growth. Also differences were seen in the performance of different fish ponds compared to the current recommended feeding values	ML models for making better feeding recommendations
E	The company employs a wide range of professionals in kitchens, restaurant basis and support function and is responsible for managing about nine restaurants	Small	120	Forecasting the number of customer based on different days and times	Hourly receipts and sales reports	Random Forest (RF) Regression and Support Vector Machine (SVM) for forecasting sales and receipt volumes	Forecast: future sales using a regression model	Anticipate people's movements and utilize this information more widely in restaurants
F	The Company for restaurants and cafe management, mainly on the campuses of the University of Turku and the University of Applied Sciences.	Medium	88	Predicting customer behavior in shopping	Sold products in the restaurant at different times and customer groups	Visualization of purchase prices vs the customer group and purchase prices by product. Utilize a regression model for predictions of sales	Predict future sales	Real-time restaurant-specific sales forecasting
G	Logistics services such as 4PL logistics.	Small	17	Development of transport routes and time	Transport routes, timetables, departure and arrival points, loaded weights	Visualization map view of freight data, departures and destinations, duration, customer-specific groupings	Visualization provide a better view of the overall situation for available data.	Through visualizations, the company has clearer insights into the utilization of AI in cargo planning
H	Psychotherapy centre	Micro	1	Optimization of reception room	Information on room reservations based on the date, time and Doctor	Visualization and classification based on customer, duration and volume	Get an overview of the overall situation according to the different customer groups, users, schedule times, booking duration and volume	Utilizing AI for optimizing room reservations
I	On of largest designer and manufacturer of aluminium boats	Medium	161	Customer segmentation for target marketing and gather product development ideas.	Consumer survey collected from customers	k-means clustering	Customer survey responses 'grouped into subsets for identifying marketing target groups	Marketing targeting and product development using AI
J	The company is specializes in digital transformation and automation for large and medium-sized companies	Medium	224	New solutions for document (text) analysis and finding appropriate tools for text analysis	Very large amount of financial document data.	FinBERT model as a deep neural network developed by the University of Turku	The university demo gave the capability for further development by the customer utilization of service data	Possibility to integrate FinBERT in their system
K	Software development	Micro	1	Proposing strategic measures based on collected text data	Data collected on the strategic action proposals	Natural language processing (NLP), Natural language AI- Google Cloud	Simple NLP model based on AI-Google Cloud	Integrating NLP in the system
L	The company produces customer service chatbots for hotels	Micro	7	Customer satisfaction for better performing of chatbots and development of chatbots	Data related to the questions asked to the chatbots and the topics they concluded.	Latent Dirichlet Allocation (LDA) method to find topic areas from the questions in chatbots	Finding interesting subject areas Topic areas help in choosing training data for chatbots: are there new topics that current bots don't recognize? let's choose sentences as teaching data for chatbots.	Continuing experiments with the LDA to find topic areas in questions asked to chatbots Adopting the model in automating the selection of teaching data for chatbots Possibilities are automatically suggested to the customer, which topic the question is

Fig. 3 List of the studied companies and their characteristics

presentation not included) help to evaluate possible new projects. Such an evaluation is a rather intuitive process, which requires summaries of the past cases.

4.1 Data visualization cases

Two data visualization cases demonstrate how vastly different each case can be. This means the company's domain knowledge is not easy to communicate nor interpreted by the AI personnel (neither by the ones in this project or possible future contacts of the company).

4.1.1 Fish feeding

The collected data, in this case, represent the amount of feeding, recommended amounts, water temperatures in different rearing ponds, water oxygen levels, etc. Working with this data was challenging due to the scope of the variables, the amount of data, and the company's requirements. Visualizations were obtained in which the most important variables from the company's point of view were visualized, as well as connections and correlations such as amount of feeding vs temperature, amount of feeding vs recommended amount at different temperatures, amount of feeding vs. temperature vs. assumed growth of the fish. Figure 4 illustrates the visualization plots for different variables in this dataset.

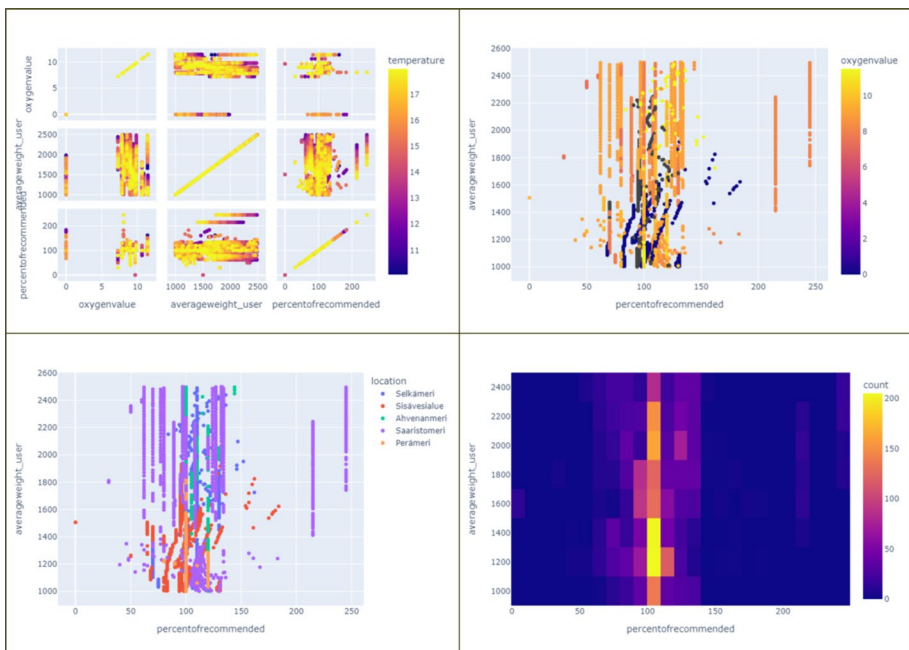


Fig. 4 Visualization of fish feeding data for company C

4.1.2 Logistic

Company G. is an integrated supply chain company specializing in complete logistics services known as 4PL logistics. The company had ready a large amount of data related to logistics, such as departure and destination information, cargo weights, routes, transport methods, durations, etc. The purpose of the visualization was to give the company a complete picture of the current situation in the simplest and most informative way possible. The project was implemented with an interactive map (Fig. 5) where the company could check transport-related issues from different perspectives. In the visualization, other categories were also made so that the results could be viewed based on various customer groups, modes of transport or geolocation data.

4.2 Machine learning cases

4.2.1 Clustering of a cabin's sensor data

The company's goal, a global manufacturer of cabin modules for cruise ships, was to understand the potential of their data and AI in product development. We applied an unsupervised ML method, K-means (Lloyd, 1982), on acceleration data recorded by a sensor that was installed in the cabin during the steering tests of the ship. The aim was to study if it was possible to identify real-life events, such as sharp steering movements of the ship and vibrations they caused to the cabin's structure with this AI method.

The one-dimensional acceleration data was first processed, split into segments of 10 s and filtered with the Butterworth low-pass filtering by 30 Hz to cut noise. Features, such as, a mean, medium, minimum value, maximum value, and root mean square were generated from each segment in both the time and frequency domains. After that, the K-means clustering method was applied to form 3 clusters. Finally, the clusters produced by the K-means method were processed with the Principal Component Analysis (PCA) method to project the generated features into two principal components. It could be seen that the method had produced clearly separated clusters.



Fig. 5 Visualization map view of freight data for company G

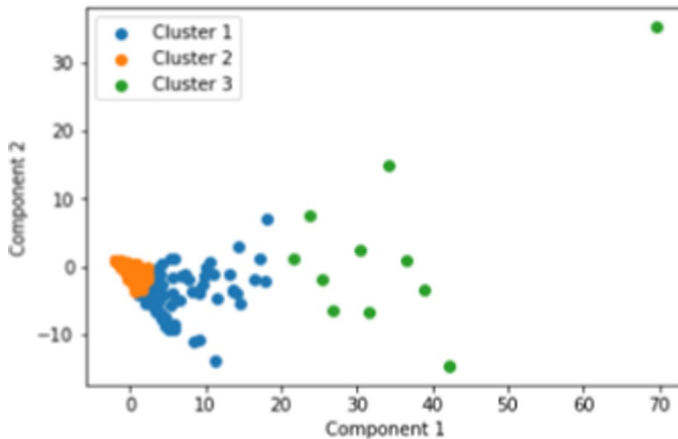


Fig. 6 Result of cabin's data clustering for company A

Figure 6 illustrates this visualization. In addition, steering data of the ship had been recorded, and thus the ground truth of the actual events of the ship was available. By comparing the produced clusters and the real-life events, it could be concluded that helpful information was generated from the acceleration data using the K-means method.

Based on the discussions and the experiment results, the company decided to start collecting and labelling training data in a laboratory and real-life situations. They also implemented a supervised neural network to continue AI experiments on the acceleration data. The plan was to collect more data with sensors installed in various types of cabins both when transporting the manufactured cabins to the harbour and onboard in the ship.

4.2.2 Forecasting restaurant sale

The company managing several restaurants aimed to understand the potential of their data and AI in process optimization and product/service innovation. The goal was to investigate whether it was possible to forecast future sales and customer numbers from the cash register data to adjust resources accordingly. We applied the Random Forest regression model to the available cash register sales data and demonstrated how the model made hourly forecasts for the following weeks.

The sales data of the cash register was first grouped hourly, and features, like hourly sales of the previous day, a week ago and two weeks ago, were generated. Also, information about a weekday was added. The model succeeded in making hourly forecasts with sufficient accuracy for usual days. However, it was clear that more data would be needed, such as holidays and information about significant events arranged near the restaurants, to forecast unusual days with better accuracy. This demonstration is illustrated in Fig. 7. The company suggested that the university would start a research project to anticipate future movements of the people in the area with available data and AI.

This information could more widely benefit the hotel and restaurant industry and allow the companies to build their own AI solutions to develop their business.

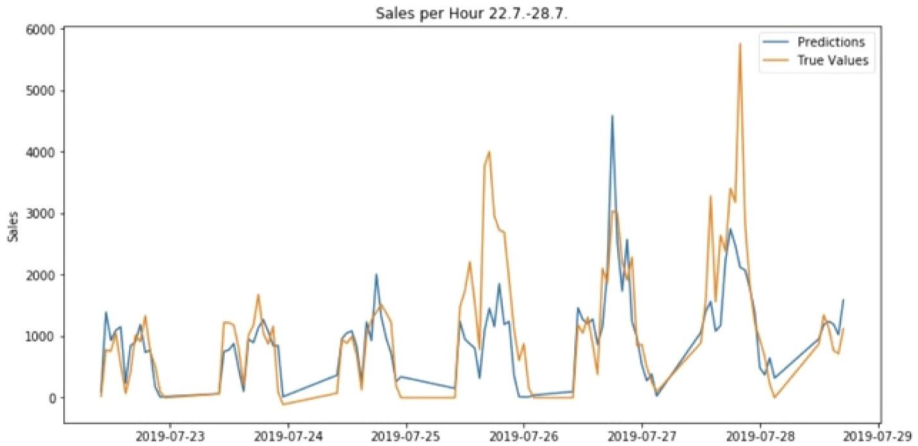


Fig. 7 Result of restaurant sale prediction for company E

4.2.3 Natural Language Processing (NLP)

NLP is a branch of artificial intelligence focusing on understanding written or spoken human language. We used Latent Dirichlet Allocation (LDA) (Blei et al., 2001) as it is widely used in text mining. It can increase chatbot performance for hotel customer service. LDA is a generative statistical model that explains a set of observations through unobserved groups, and each group explains why there is similarity between data. The model succeeded in categorizing the words in the questions related to the same subject areas into the same category. With the help of the method in question, the chatbot can,

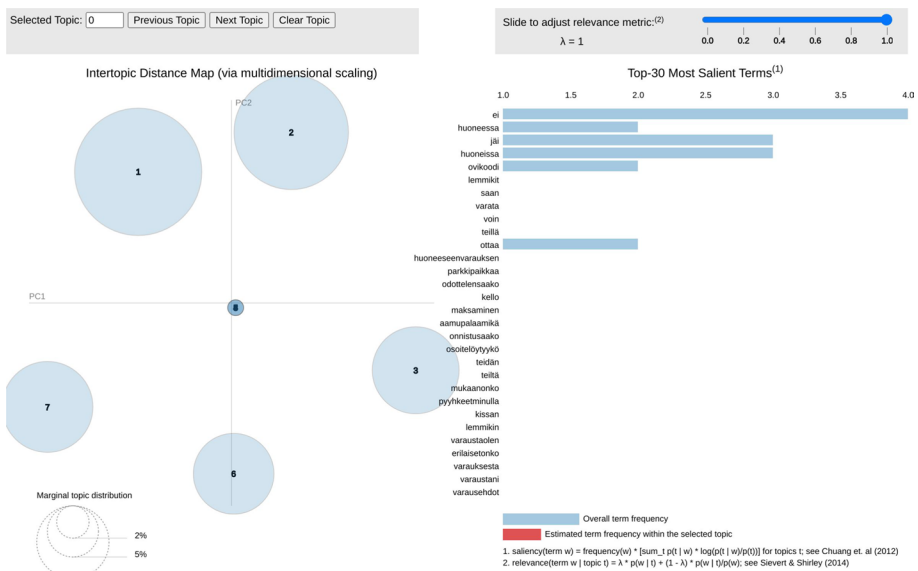


Fig. 8 Result of Latent Dirichlet Allocation (LDA) for company L

Table 1 The positive segment of answers in the feedback survey made 18 months after the project end

Positive gain from the project?	Yes, 57%
Proposed AI solutions were implementable?	Yes, 37%
The cooperation started during the project was fruitful?	Yes, 25%
It was easier to contact AI experts afterwards?	Yes, 25%
Business model have been adapted to AI?	Yes, 13%

Table 2 SMEs' feedback after 2 months from the ending of the project about how well the data understanding improved because of the project

Poorly	0%
Fairly	25%
Well	62%
Excellent	13%

unsurprisingly, learn and improve its skills. Figure 8 shows the result of this model for company L.

4.3 Long-lasting intervention effects

A survey was conducted in the spring of 2024, and companies were asked about the effect of the AI Champions project. From 8 companies giving answers 0,1,...,5 to 11 questions, no single cluster was to be found.

Since the data is small (8 companies \times 11 questions \times 5 options per question), we formed a histogram of 28 possible company pairs and observed nearly uniform distribution of lengths between pairs. As Jain (2008) puts it, there are no interesting subsets in a small sample set like this. The principal component analysis (PCA) points out the same way.

The above is a justification for evaluating answers as a whole, based only on the mean values of answers. The following is a short summary of the observations for the companies with positive indications (Table 1).

There was also a negative or passive subset of companies. The main reasons for not gaining many results were these:

- A concrete solution needs a lot of preliminary conditions fulfilled (sensors, hardware, legal issues)
- Co-operation with a contractor delivering a subsystem did not succeed.
- Company budget does not yield to required experimentation.
- Recent GPT (Generative pre-trained transformer) technology gave a new direction for the company project
- Co-operation with a peer company (with a similar problem) did not continue
- The project sketched in AI Champions did not continue (75%)

Often the answer was divided between the contact person of the project in the company (usually positive response, person learned new skills and acquired new understanding) and the company. The discontinuation of the AI Champions project is not so drastic than it sounds since project topics were only examples, and the focus was more in the new concise process view to AI projects.

5 Discussion

Table 2 shows the SMEs' feedback at the end of the project about this project by asking the central question: "How did your understanding of data and its potential grow during your participation?". The results show that more than 62% gave us good feedback, and they found a general view for data understating and developing AI in their company by providing some examples of data visualization or ML models from their data.

Furthermore, we investigated how the AI Champion project was able to support the SMEs based on the Gartner AI maturity model (<https://www.gartner.com/smarterwithgartner/the-cio-s-guide-to-artificial-intelligence>). We use this model to identify the level in which AI is used in a company in the starting and ending points of the project. We also show how many discussion rounds were arranged with the company and what kind of experiments were made to get value out of the sample data. In some cases the AI maturity level increased from an early AI interest to starting AI experiments in the company. Also, other benefits the companies gained during the project are described in Fig. 9.

The focus of the study was not to cluster companies and provide each cluster with different support due to the relatively small sample set. Instead, we engaged with companies to improve their generic AI readiness, which is again a not-so-easily quantifiable property. Possible quantifiable categories are mentioned in Westenberger et al. (2022), where commercial AI project failures are divided into organizational and technological issues. Our study found that two companies had organizational problems only, while one had technological issues only. Four companies did not have any problems, but some of them still faced

Company	Starting Point	Discussion rounds	Experiments with sample data	Ending Point	AI Maturity Increased	Benefit gained
A	Early AI interest: product innovation with AI	4	Data visualization, K-means clustering	AI Experimentation: started acquiring and labeling training data and implementing a neural network	1	Real life events were detected from the acceleration data. A decision was made to acquire and label new acceleration data and implement a neural network.
B	Early AI interest: product/service innovation with AI	7	Data visualisation, K-means clustering	AI Experimentation: started to analyse data to generate features for AI experiments.	1	A challenge to detect or predict leakages in different kinds of district heating wells was recognized. A plan to first cluster different wells was made.
C	Early AI interest: product/service innovation with AI	5	Data visualization	Early AI interest: an interest to start a research project about acquiring more accurate data about weights of the fish living in cloudy sea water	0	Useful business information was generated from the data. A challenge to forecast growth of the fish without knowing accurate real weights was recognized.
E	Early AI interest: process optimization with AI	5	Data visualization, Random Forest regression	Early AI interest: Interested in attending a research project about forecasting movements of the people in the area.	0	Forecasting future sales was demonstrated with fairly good accuracy for usual days. A need to acquire data related to unusual days, such as big events in the area, was recognized.
F	Early AI interest: process optimization or product/service innovation with AI	4	Data visualization, Random Forest regression	AI Experimentation: started a project to continue experiments with a bigger amount of cash register data	1	A need to acquire more cash register data and data related to unusual days, such as student events in the area, was recognized.
G	Early AI interest: process optimization with AI	4	Data visualization, map views, possibilities to use graphical models, simulations and optimizations of graphical	Early AI interest: interested in continuing data analysis and seeking possibilities to utilize AI.	0	The company got useful information about the overall situation and better understanding of utilizing data and AI in cargo planning
H	Early AI interest: process optimization with AI	4	Data visualization	Early AI interest: interested in learning more about utilising data and AI	0	The company got useful information about the overall situation and better insight of utilizing data in optimising room reservation services.
I	Early AI interest: process optimization or product/service innovation with AI	4	Data visualization, K-means clustering	Early AI interest: interested in learning more about utilising data and AI	0	Useful business information was generated from the customer survey data.
J	AI in production creating value: adding AI services	4	A demonstration how to take a FinBERT language model into use: possibilities to use the FinBERT model and the FinGPT-3 model being trained in University of Turku	AI in production creating value: adding AI services, a plan to take FinBERT language model into use, interested to utilize FinGPT-3 in the future	0	The company got insight how to take the FinBERT language model into use and what kind of NLP tasks it is suitable for. Also, utilising GPT-3 model was clarified
K	Early AI interest: product/service innovation with AI	4	NLP: Google Cloud tools	AI experimentation: started to make NLP experiments, interested in implementing AI services for strategy execution software	1	Understanding how to utilize natural language and NLP tools increased, new skills how to use NLP: Google Cloud tools were achieved.
L	Early AI interest: product/service innovation with AI	6	NLP: Latent Dirichlet Allocation	Early AI interest: a plan to continue AI experiments	0	Useful information was generated. A plan was made to continue experiments and automate training of chatbots.

Fig. 9 Starting and ending points based on the Gartner AI maturity model

challenging issues. Three companies had both types of problems, but one was relatively advanced in AI. Overall, it seems companies can take on more challenging and complex projects as they improve their AI skills and capacity.

One could separate data quality and theoretical ML hardness from the previous two categories (technological and organizational issues) to get a new 4-category division. The data quality relates to the amount of the preliminary processing needed and the problem hardness to the first hands-on performance using the most standard methods.

According to our judgment, it seems that data quality and the difficulty of the AI problem is independent of organizational and technological issues. This means that anyone repeating our project format will need a multi-talented, flexible, and efficient project team to face a variety of business domains, human interaction challenges and AI problems.

Quite consistently, customer behaviour prediction is one of the most complex AI tasks in the business domain. This is because most psychological and social processes are haphazardly understood and sparsely monitored. On the other hand, the most straightforward tasks have commonly known solutions and existing pipelines, such as data collection.

Working with each company takes time, and thus our sample set is limited. Typical feedback from the executives was that after the AI Champion project, they are more ready to involve AI experts in the problem-solving processes and more aware of the possibilities and pitfalls. One intended consequence was an increased readiness of the company staff for planned or unplanned interaction with AI experts, networked companies, and employee candidates.

5.1 Repeatability of the experiment

We divide the repeatability to two components, (1) general instructions to project planner, who intends to repeat our project in a loosely similar form, and (2) remarks on AI technology changes.

(1) General design of the AI regional development Considering our experiences with the selected 11 companies, and attempting to extrapolate to the present moment technologies available 2 years later from the project, we would set the following guidelines for the repetition of a similar project e.g. in developing countries:

- select a heterogeneous project team where creativity, industrial and academic experience, fast prototyping and data analysis skills, high-standard teamwork, networking, communication and documentation co-exist. Motivate AI experts by CV and portfolio benefits.
- prescreen SM companies based on their perceived potential to R & D.
- make sure the CEO directly supports the effort and that the needed personnel (usually 1 or 2 persons) is allocated to act at the project interface with a clear resource mandate.
- identify groups of companies in a non-competitive situation with common needs.
- pronounce clearly the size and quality of the consulting resource and effort available in the project.
- if possible, implement the project in 2 or 3 stages, where companies with the most potential continue and are subjected to more intense interaction.

(2) Developments in AI technology We focus in this short account to the data pre-processing and recent approaches in building the AI pipeline.

Implementing data cleaning has become a routine task, where large language models are capable of proposing both algorithms and software and generating necessary code lines. This phase requires excellent understanding of core AI principles, though, since human validation of approaches chosen is a necessity.

Data understanding can be reached faster with explorative tools like Tableau (Salesforce, 2019) and Power BI (Microsoft, 2015) (Wimmer & Powell, 2016).

Whole the AI pipeline is supported by Amazon SageMaker and AWS Glue, Microsoft Azure Data Factory and Anaconda Enterprise. It is hard to find data analysis researcher who masters all these tools, so a well managed teamwork is a requirement here.

We omit exciting technological developments like transformers and large language models from this exposition and focus on project methodologies, which adapt these new technologies well.

The CRISP-DM used in this research is not outdated, but it has to be augmented by approaches like machine learning operations MLOps (<https://ml-ops.org/>), data-centric AI (Zha et al., 2023) and collaborative AI with differential privacy. For academy (organizers of research like this) and industry (targets of this research), utilizing these new project formats and leading ideologies takes practise. The effort concerns teams, since it is hard for any individual to master all of these large AI project management styles.

For a minimalistic approach, the data-centric AI (and collaborative AI if it is possible by participants) are best additions to CRISP-DM. The data-centric AI reduces failures by guaranteeing the best possible data understanding, and collaborative AI with privacy issues covered makes approaches rational and economical with shared data and code, and helps in adopting large scale technologies like LLM and ready-made image recognition pipelines.

6 Conclusions

We mapped out potentials and first steps in applying AI to the industrial problems of 11 companies. We outlined their business case, AI skills and capabilities, and operating environment and provided solutions ranging from a general strategy recommendation to working singular solutions. The emphasis was on improving participants' AI skills and knowledge and changing many SMEs' current AI mindset and culture towards more collaboration between various stakeholders. Most of our output was preliminary, often focusing on visualizations. Also, an AI Champion Canvas was produced in order to document some significant points of a potential further effort. As a result, we report an increase in both data understanding and readiness to manage AI projects, especially in communicating with AI consultants and company representatives. One Result of the project was an increased skill set of both the core project team and involved members of some companies. We state that this kind of project focusing on the initial phases of potential AI projects may be beneficial and cost-effective for SME companies to start the well-formulated case-specific AI journey to learn, collaborate and share knowledge. This would result in the initial approach of companies being fine-tuned, and some premature attempts can be either avoided or re-focused.

The benefit for the company is better information, knowledge and understanding in making a high-quality decision on the following concrete steps regarding (a) the specific case, (b) AI development more general in the company, (c) continuous learning, and also with partners. One must remember that in IT investments, the decision to postpone, e.g. an AI project, is also valid if resources, skills, costs, or expected benefits are not met. And all this relevant, high-quality business development information with an investment of 10–50

person-hours per company, as was the case in our project. In this sense, the participating companies felt that they, and their company, transferred to a more AI-centric and more skilled workforce to work with AI issues by themselves and with partners and AI service providers.

Our current understanding based with the project is that a similar intervention could help in finding new sectors and domains with AI application potential. It also reduces the threshold of contacting and communicating with AI experts. We know, that as AI field keeps developing, the initial experiments and data-driven prototyping become easier, but perils of AI projects stay the same (balancing expectations, resources and risks).

As a future research, it seems necessary to repeat this kind of project shortly to map out updated AI application strategies for the industry, especially focusing on the impact of several ready-made pipelines in the fields of image and natural language processing.

Acknowledgements This work was supported by tit-for-tat. The views and conclusions in this paper are those of the authors and do not represent any funding agencies.

Funding Open Access funding provided by University of Turku (including Turku University Central Hospital).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D.G., Steiner, B., Tucker, P.A., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., & Zhang, X. (2016). Tensorflow: A system for large-scale machine learning. CoRR [arxiv:abs/1605.08695](https://arxiv.org/abs/1605.08695)
- Al-Shehri, H., Al-Qarni, A., Al-Saati, L., Batoaq, A., Badukhen, H., Alrashed, S., Alhiyafi, J., & Olatunji, S. O. (2017) Student performance prediction using support vector machine and k-nearest neighbor. In *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)* (pp. 1–4). <https://doi.org/10.1109/CCECE.2017.7946847>
- Bahari, T. F., & Elayidom, M. S. (2015). An efficient crm-data mining framework for the prediction of customer behaviour. *Procedia Computer Science*, *46*, 725–731. <https://doi.org/10.1016/j.procs.2015.02.136>
- Bauer, M., Dinther, C., & Kiefer, D. (2020). Machine learning in sme: An empirical study on enablers and success factors. In *26th Americas Conference on Information Systems, AMCIS 2020*.
- Bhalerao, K., Kumar, A., Kumar, A., & Pujari, P. (2022). A study of barriers and benefits of artificial intelligence adoption in small and medium enterprise. *Academy of Marketing Studies Journal*, *26*(1), 1–6.
- Blei, D., Ng, A., & Jordan, M. (2001). *Latent Dirichlet Allocation*, *3*, 601–608.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T.P., Shearer, C., & Wirth, R. (2000). Crisp-dm 1.0: Step-by-step data mining guide.
- Chen, C. (2012). Storey: Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, *36*(4), 1165. <https://doi.org/10.2307/41703503>
- Christoph Szedlak, P. P., & Leyendecker, B. (2020). Application of artificial intelligence in small and medium-sized enterprises. In *Proceedings of the 5th NA International Conference on Industrial Engineering and Operations Management Detroit* (pp. 10–14).
- Deloitte. (2020). Survey on ai adoption in manufacturing. Retrieved from <https://www2.deloitte.com>
- DeStefano, T., Teodorovic, T., Cho, J., Kim, H., & Paik, J. (2022). What determines ai adoption? *Academy of Management Proceedings*. <https://doi.org/10.5465/ambpp.2022.14791abstract>

- Drydakis, N. (2022). Artificial intelligence and reduced SMEs' business risks. A Dynamic Capabilities Analysis During the COVID-19 Pandemic. GLO Discussion Paper Series 1045, Global Labor Organization (GLO) (April). <https://ideas.repec.org/p/zbw/glodps/1045.html>
- EC.: What Is an SME? https://ec.europa.eu/growth/smes/sme-definition_en/
- Farahnakian, F., & Heikkonen, J. (2018). Anomaly-based intrusion detection using deep neural networks. Farahnakian, F., Koivunen, L., Mäkilä, T., & Heikkonen, J. (2021). Towards autonomous industrial warehouse inspection. In *2021 26th International Conference on Automation and Computing (ICAC)* (pp. 1–6). <https://doi.org/10.23919/ICAC50006.2021.9594180>
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- Goodfellow, I. J., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Guzella, T. S., & Caminhas, W. M. (2009). A review of machine learning approaches to spam filtering. *Expert Systems With Applications*, *36*(7), 10206–10222.
- Huang, M.-H., & Rust, R. (2020). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science*. <https://doi.org/10.1007/s11747-020-00749-9>
- Jain, A. K. (2008). Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, *31*, 651–666.
- Jöhnik, J., Weißert, M., & Wyrski, K. (2021). Ready or not, ai comes - an interview study of organizational ai readiness factors. *Business & Information Systems Engineering*, *63*(5), 441–451. <https://doi.org/10.1007/s12599-021-00693-1>
- Jung, W.-K., Kim, D.-R., Lee, H., Lee, T. H., Yang, I., Youn, B. D., Zontar, D., Brockmann, M., Brecher, C., & Ahn, S.-H. (2020). Appropriate smart factory for smes: Concept, application and perspective. *International Journal of Precision Engineering and Manufacturing*. <https://doi.org/10.1007/s12541-020-00445-2>
- Lloyd, S. (1982). Least squares quantization in pcm. *IEEE Transactions on Information Theory*, *28*(2), 129–137. <https://doi.org/10.1109/TIT.1982.1056489>
- Lu, X., Wijayarathna, K., Huang, Y., & Qiu, A. (2022). Ai-enabled opportunities and transformation challenges for smes in the post-pandemic era: A review and research agenda. *Frontiers in Public Health*, *10*, 885067. <https://doi.org/10.3389/fpubh.2022.885067>
- Martínez, D. M., & Rodríguez, J. C. F. (2015). Artificial intelligence applied to project success: A literature review. *International Journal of Interactive Multimedia and Artificial Intelligence*, *3*, 77–84.
- McKinsey. (2022). Adopting ai in manufacturing at speed and scale. Retrieved from <https://www.mckinsey.com>
- Mehta, R., & Rajendran, P. (2020). *The future of artificial intelligence adoption in India*. Retrieved September 14, 2021, from <https://www.financialexpress.com/industry/technology/the-future-of-ai-adoption-in-india/2123897/>
- Moro, S., Cortez, P., & Rita, P. (2015). Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation. *Expert Systems with Applications*, *42*(3), 1314–1324. <https://doi.org/10.1016/j.eswa.2014.09.024>
- Munde, A., & Mishra, N. (2022). Corporate performance: Smes performance prediction using the decision tree and random forest models. *Corporate Ownership and Control*, *20*, 103–113. <https://doi.org/10.22495/cocv20i1art10>
- Nassiri-Mofakham, F. (2014). How does an intelligent agent infer and translate? *Computers in Human Behavior*, *38*, 196–200. <https://doi.org/10.1016/j.chb.2014.05.019>
- Nemati, H. R., Steiger, D. M., Iyer, L. S., & Herschel, R. T. (2002). Knowledge warehouse: An architectural integration of knowledge management, decision support, artificial intelligence and data warehousing. *Decision Support Systems*, *33*(2), 143–161. [https://doi.org/10.1016/S0167-9236\(01\)00141-5](https://doi.org/10.1016/S0167-9236(01)00141-5)
- Oy, T. (2019). Varsinais-suomen tekoälyverokoto - ai-kartoitus 2019, Turku science park oy, 2–32 . <https://turkubusinessregion.com/wp-content/uploads/2019/04/VS-teko%C3%A4lyverokoto-tutki-mus-yhteenveto-20190201.pdf>
- Panetta, K. The CIO's Guide to Artificial Intelligence. <https://www.gartner.com/smarterwithgartner/the-cios-guide-to-artificial-intelligence>
- Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., & Lerer, A. (2017). Automatic differentiation in pytorch. In *NIPS 2017 Workshop on Autodiff*. <https://openreview.net/forum?id=BJJsrmlfCZ>
- PwC. (2024). Pwc's global artificial intelligence study. Retrieved from <https://www.pwc.com>
- Schlögl, S., Postulka, C., Bernsteiner, R., & Ploder, C. (2019). Artificial intelligence tool penetration in business: Adoption, challenges and fears (pp. 259–270). https://doi.org/10.1007/978-3-030-21451-7_22

- Schlögl, S., Postulka, C., Bernsteiner, R., & Ploder, C. (2019). Artificial intelligence tool penetration in business: Adoption, challenges and fears. In L. Uden, I.-H. Ting, & J. M. Corchado (Eds.), *Knowledge management in organizations* (pp. 259–270). Springer.
- Susto, G. A., Schirru, A., Pampuri, S., McLoone, S., & Beghi, A. (2015). Machine learning for predictive maintenance: A multiple classifier approach. *IEEE Transactions on Industrial Informatics*, 11(3), 812–820. <https://doi.org/10.1109/TII.2014.2349359>
- Thi e, L.-W. (2021). A systematic literature review of machine learning canvases. In *INFORMATIK 2021* (pp. 1221–1235). Gesellschaft f ur Informatik. <https://doi.org/10.18420/informatik2021-101>
- Turchi, M., Bie, T.D., & Cristianini, N. (2009). An intelligent agent that autonomously learns how to translate. In *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology* (Vol. 2, pp. 12–19). <https://doi.org/10.1109/WI-IAT.2009.120>
- Westenberger, J., Schuler, K., & Schlegel, D. (2022). Failure of ai projects: Understanding the critical factors. *Procedia Computer Science*, 196, 69–76. <https://doi.org/10.1016/j.procs.2021.11.074>
- Wimmer, H. & Powell, L. M. (2016) A comparison of open source tools for data science. <https://api.semanticscholar.org/CorpusID:64082666>
- Zha, D., Pervaiz Bhat, Z., Lai, K.-H., Yang, F., & Hu, X. (2023) Data-centric AI: Perspectives and Challenges. arXiv e-prints, 2301–04819 <https://doi.org/10.48550/arXiv.2301.04819>
- Zhou, Z., Kearnes, S., Li, L., Zare, R. N., & Riley, P. (2018) Optimization of molecules via deep reinforcement learning. CoRR [arxiv: abs/1810.08678](https://arxiv.org/abs/1810.08678)
- Zhou, L., Zhang, P., & Zimmermann, H.-D. (2013). Social commerce research: An integrated view. *Electronic Commerce Research and Applications*, 12(2), 61–68. <https://doi.org/10.1016/j.elerap.2013.02.003>

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.