VTT Technical Research Centre of Finland

InnoSale – Innovating Sales and Planning of Complex Industrial Products Exploiting Artificial Intelligence

Järvinen, Sari; Lämsä, Arttu; Sormunen, Tuomas; Peltola, Johannes; Jurvansuu, Marko; Mäkelä, Satu-Marja; Hyvärinen, Santeri; Siira, Erkki

Published: 03/11/2025

Please cite the original version:
Järvinen, S., Lämsä, A., Sormunen, T., Peltola, J., Jurvansuu, M., Mäkelä, S.-M., Hyvärinen, S., & Siira, E. (2025). InnoSale – Innovating Sales and Planning of Complex Industrial Products Exploiting Artificial Intelligence: Final report. VTT Technical Research Centre of Finland.

RESEARCH REPORT

VTT-R-00502-25

InnoSale – Innovating Sales and Planning of Complex Industrial Products Exploiting Artificial Intelligence: Final report

Authors: Sari Järvinen, Arttu Lämsä, Tuomas Sormunen, Johannes Peltola, Marko Jurvansuu, Satu-Marja Mäkelä, Santeri Hyvärinen, Erkki Siira

Confidentiality: VTT Public

Version: 29.10.2025

beyond the obvious

Report's title
InnoSale – Innovating Sales and Planning of Complex Industrial Products Exploiting Artificial Intelligence: Final report

Project name	**Short name**
Innovating Sales and Planning of Complex Industrial Products Exploiting Artificial Intelligence	InnoSale
Author(s)	**Pages**
Sari Järvinen, Arttu Lämsä, Tuomas Sormunen, Johannes Peltola, Marko Jurvansuu, Satu-Marja Mäkelä, Erkki Siira	25
Keywords	**Report identification code**
Business-to-business sales, artificial intelligence, machine learning, sales processes	VTT-R-00502-25

Summary

The InnoSale project, running from 2021 to 2025, set out to transform the sales processes for complex industrial products in B2B environments by harnessing the power of artificial intelligence. The project brought together a diverse international consortium of industry leaders and research organizations, all united by the goal of making sales processes more efficient, accurate, and customer-centric. At the heart of InnoSale was the recognition that traditional B2B sales, especially for highly configurable products, relied heavily on manual expertise, fragmented data, and intuition, which often led to inefficiencies and inconsistent outcomes.

To address these challenges, the project focused on developing and piloting AI-driven tools that could automate routine sales tasks, provide data-driven insights, and support decision-making throughout the sales funnel. VTT and its partners co-created use cases that reflected real-world needs, such as semantic search tools to help sales and support teams quickly find relevant information from historical support tickets, machine learning models to predict the outcomes of sales calls and offers, and intelligent recommendation systems to streamline the configuration of complex products. These solutions were built on a foundation of integrating data from multiple enterprise systems (ERP, CRM, support tickets, and pricing tools) highlighting the critical importance of data quality, completeness, and integration.

Experiments conducted during the project demonstrated that AI models could achieve promising accuracy in tasks like offer outcome prediction and customer segmentation, even when working with anonymized or B2C datasets due to the scarcity of real B2B data. The project also underscored the importance of trust and user acceptance: while management tended to be optimistic about AI's potential, experienced sales professionals were often skeptical, emphasizing the need for transparent, explainable models and user-centered design. Feedback from these users will be invaluable in refining the tools and identifying areas for further development.

Ultimately, InnoSale showed that while advanced AI can deliver significant value, even basic automation and digitalization can lead to substantial improvements in sales processes. The project concluded that the most successful AI solutions are those that augment human expertise, automate repetitive work, and provide actionable insights freeing up sales experts to focus on higher-value activities. Looking ahead, the report recommends further work to secure richer B2B datasets, pilot AI tools in real business settings, and invest in robust data governance and user training. InnoSale thus lays a strong foundation for the continued evolution of AI-driven sales in complex industrial domains, balancing technological innovation with practical, user-focused implementation.

Confidentiality	Public

Oulu 3.11.2025

Written by	**Reviewed and approved by**
Signed by: Sari Järvinen 1E435D28474B4A3...	DocuSigned by: Marko Jurmu 2A9E33F580E84F0...
Sari Järvinen Senior Scientist, IPMA C Certified Project Manager	Marko Jurmu Research Team Leader

VTT's contact address
VTT Technical Research Centre of Finland Ltd, P.O. Box 1000, FI-02044 VTT, Finland

Distribution (customer and VTT)
Public

Preface

This final report done in InnoSale – Innovating Sales and Planning of Complex Industrial Products Exploiting Artificial Intelligence ITEA project, partially funded by Business Finland 2021-2025 gives an overview of the research activities performed by VTT in the project. The report first describes shortly the state-of-the-art in the domain of B2B sales of complex industrial products in the beginning of the project. Next the use cases, which were defined in collaboration with the Finnish project partners, provide the need for development of artificial intelligence (AI) solutions and define the context for the research work. The external data sources identified collect and provide data necessary for the implementation of AI algorithms using authentic business data. Finally, the key findings give an overview on challenges and outcomes of the InnoSale project. The authors are grateful for the valuable contributions and comments by the project partners during the project execution and on the report.

Oulu 3.11.2025

Sari Järvinen

Contents

1. Introduction

The ITEA InnoSale project, titled "Innovating Sales and Planning of Complex Industrial Products Exploiting Artificial Intelligence," (2021-2025) aimed to renew the business-to-business (B2B) sales and planning processes of complex industrial products through the application of advanced artificial intelligence (AI) techniques. The primary objectives of the project included enhancing the efficiency and accuracy of sales predictions, optimizing planning strategies, and improving customer engagement by leveraging AI-driven insights. In the current ways of working the preparing a B2B offer requires a lot of manual expert work and involves usage of hidden knowledge which is difficult or impossible to share in the organization.

The InnoSale project was supported by an international consortium comprising leading industry partners, research organizations and technology providers from various countries. The consortium included partners such as Software AG, Demag, Natif, Ermetal, Dakik, Konecranes, Molok, LeadDesk and Wapice. This collaborative effort ensured a comprehensive approach to addressing the challenges and opportunities in the industrial sales domain.

Among the Finnish partners, VTT Technical Research Centre of Finland Ltd played a pivotal role in the project. VTT's expertise in AI and industrial applications was instrumental in experimenting with innovative solutions and methodologies. VTT's contributions included co-creating the use cases with the industrial partners, building and experimenting with AI solutions, and facilitating collaboration among the consortium members.

2. Use of artificial intelligence to support B2B sales

AI-driven tools are transforming B2B sales processes by automating tasks and increasing interconnectivity of business operations, often referred to as the fourth industrial revolution [1]. Key drivers include new buyer behaviour, rising customer requirements, new information and communication technologies, and globalization [2].

These tools analyse large amounts of heterogeneous data to provide valuable insights and support decision-making in the sales process. They enhance customer experience, increase efficiency, and reduce costs. Feedback provided by AI tools helps sales personnel improve adaptability and performance [3].

To benefit from these digital tools, sales culture needs to change, and existing routines must be unlearned [4]. Education and support from management are essential for the successful implementation of AI technologies [5]. AI tools are particularly valuable in automating tedious and repetitive sales tasks, learning from sales activities, and performing actions autonomously [10].

Customer Relationship Management (CRM) systems are widely used to track interactions and collect information from customers. AI can enhance CRM systems by supporting data-driven decision-making and marketing strategies. AI has a future impact on decision-making strategies, user behaviour response, innovation strategies, sales forecasting, social media use, and customer orientation [6].

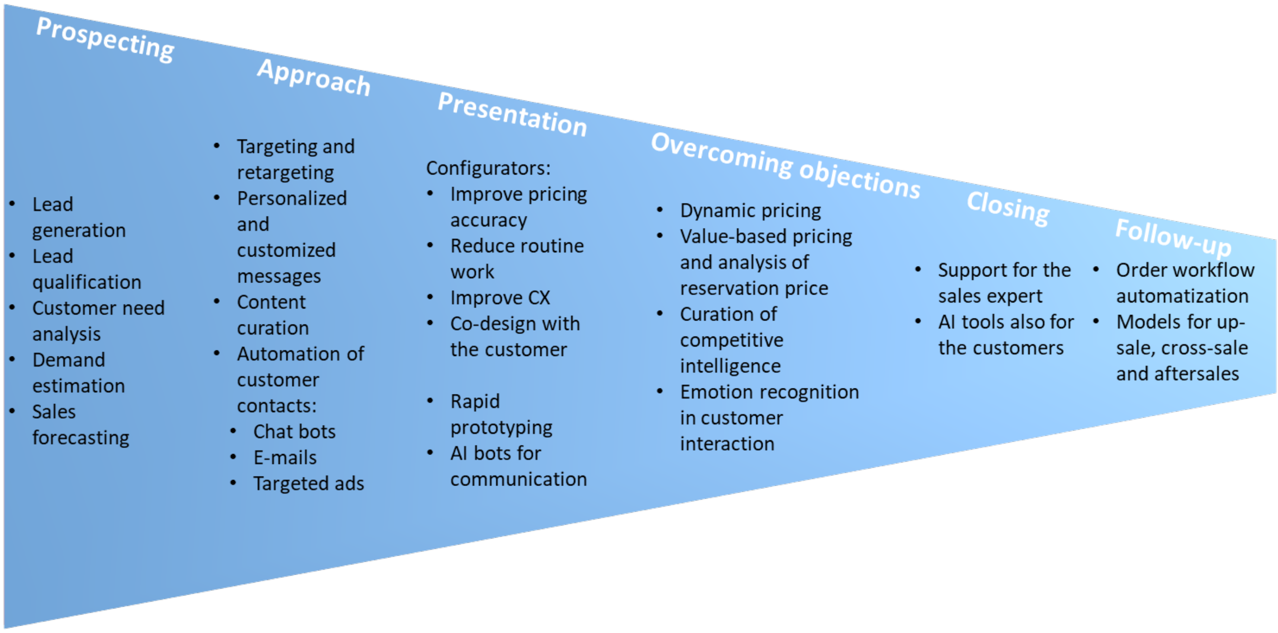

Figure 1. AI-driven tools assist the sales expert along the sales funnel.

Examples and visions from current research show how AI can be utilized along the sales funnel (Figure 1). AI tools can help in prospecting by utilizing social media and other internet sources to harvest potential new customers [1]. In lead generation, AI tools can automatically find potential customers and build rich prospect profiles [1][5]. Predictive lead qualification and demand estimation can be done using natural language processing (NLP) methods [7].

AI can also support marketing strategies, loyalty, and long-lasting customer relationships along the CRM [6]. Digital agents, such as chatbots, can automate the initial contact process with prospects. AI-assisted tools provide value for ad targeting, personalized communication, and content curation. In the future, chatbots with well-developed NLP features and predictive analytics for buying signals will further automate the process.

AI tools can analyse the needs of prospects [10], estimate demand, and forecast sales. They can also provide dynamic pricing, value-based pricing, and estimation of reservation prices. AI can curate competitive intelligence to support a company's value proposition and beat competitors.

Configurators and CPQ (Configure, Price, Quote) systems are considered as sales force automation (SFA) tools for the salesperson to fulfil the customer needs better [8]. Sales configurators are often used with complex products for fulfilling customization needs. Configurators facilitate the salespersons' communication with the customer and help understand the needs of the customer better.

There are many known benefits to configurators. They enable shorter lead times and improved product specifications. Product knowledge is also systematically saved when configurators are used. They can improve pricing accuracy and profitability with better specification, reduce routine work in the design and documentation phase, and generate better customer experience and satisfaction [9].

Configurator usage is moving toward co-design with the customer and more emphasis is given for the role of the customer as a user. In that case, it has been shown that the ease of use and adaptability are important factors in perceived effectiveness, and eventually impact the perceived usefulness of sales configurators. The user experience of the configurator cannot be underestimated as the perceived enjoyment has the most significant effect on perceived usefulness.[11]

Configurators are also seen as a bigger part of the business models. The success of the configurators depends on many things, including "*how tightly the configurator is integrated into the broader business model, user experience and whether companies leverage behavioral data to further innovate and improve their value propositions" [12].*

The more integrated use of product configurators and utilisation of the data they produce in business operations and in sales funnel can provide new opportunities to improve operations. Providing better functionalities even for the customer to use will leverage the opportunity for an improved (and even automatic) customer understanding. This can lead to improved customer service and may lead to a higher and faster pace of closing deals.

There is also potential with new AI-enabled (rapid) prototyping tools for creating the satisfying product with the customer [1] [7]. In the presentation phase, it is very important to show the customer as realistic a view of the end product as possible. Technologies such as augmented reality (AR) [1] [10], virtual reality (VR) and 360-degree video provide the salesperson tools to present on-site demos that are close to reality.

AI tools can automate workflows in various phases of the sales funnel, introduce post-orders and follow-up services, and predict spare part sales [5][7][13]. They can evaluate old customers, analyze future needs, and create models for upsell, cross-sell, and aftersales opportunities. The future of AI tools in sales may lead to the disappearance of some sales tasks, with the sales profession shifting towards consulting [10]. However, human interaction remains important, and AI tools are seen as assistants rather than decision-maker [5].

Such transformative technologies also present opportunities for experimentation and innovation. In the context of InnoSale project, partners exploring the use of AI in B2B sales were delving into areas such as semantic search, which allows for enhanced retrieval of contextual and relevant data, and predictive analytics related to customer segmentation to anticipate sales call outcomes or offer success rates. These experiments serve to refine processes and optimize sales strategies further.

In business-to-business (B2B) sales the customer segmentation differs significantly from business-to-customer (B2C) segmentation. Rather than relying on demographic and behavioural information at the individual level, B2B companies must gather information at the organizational level. Generally, B2B customer segmentation approaches are categorized into five types [1,2,3,4]:

- Firmographics, the business equivalent of demographics
- Needs-based
- Behaviour-based
- Profitability-based segmentation
- Customer sophistication.

[1] https://www.leadspace.com/blog/popular-methods-of-segmentation-for-b2b/
[2] https://sopro.io/added-value/blog/b2b-market-segmentation-guide/
[3] https://peertopeermarketing.co/b2b-segmentation/
[4] https://surveysparrow.com/blog/customer-segmentation-examples/

The objectives of customer segmentation can vary, and include for example:

- Predicting the win/loss outcome of sales opportunities
- Evaluating customer loyalty and churn rates
- Ranking the potential of customer leads
- Predicting customer profitability.

As illustrated, customer segmentation can be implemented at various stages of the sales funnel and with various data sources. Indeed, data serves as the primary enabler for segmentation and is often available internally within a company, such as from CRM systems, transaction logs, or website analytics. External data sources, including social media or public registers, are also valuable. This type of external data is considered "indirect" since it is not explicitly generated by the customers themselves. Alternatively, "direct" data can be gathered through surveys or customer questionnaires. The experiments in InnoSale project primarily focus on indirect data.

Semantic search has been summarized well in previous research: “In a nutshell, semantic search is ‘search with a meaning’. This ‘meaning’ can refer to various parts of the search process: understanding the query (instead of just finding matches of its components in the data), understanding the data (instead of just searching it for such matches), or representing knowledge in a way suitable for meaningful retrieval.“ [14]. Semantic search has the capability of finding semantically similar texts, instead of matching exact words and letters. This lends itself to various usages, such as search engines. Indeed, sifting through a large knowledge base to find documents related to search phrase is paramount for successful company operations.

3. Use cases

The AI solutions developed in InnoSale are built based on use cases deriving from the needs of potential users of AI-driven tools to support the B2B sales of complex industrial products.

3.1 Smart search to support usage of hidden information

Today: managing back-office support for Configured to Order (CTO) B2B sales involves significant manual effort and reliance on back-office teams. The process is often inefficient and time-consuming due to the complexity of the products and the need for customized configurations. Here's how it typically works:

1. Ticket creation: When a sales representative encounters an issue or requires support, they create a ticket in the back-office support system. This ticket includes details about the customer's requirements, the product configuration, and the specific issue or question.
2. Manual review: Back-office support teams manually review the tickets, searching through historical data and previous tickets to find relevant information and solutions. This process is labor-intensive and can lead to delays.
3. Solution provision: Based on their findings, back-office teams provide the necessary support to the sales representatives. This may involve answering questions, providing configuration details, or resolving issues.
4. Follow-Up: Sales representatives may need to follow up with back-office teams multiple times to get the required support, leading to further delays and inefficiencies.

Future: the process of managing back-office support for CTO B2B sales will be significantly improved by leveraging hidden information in support ticket history. Advanced data analytics and machine learning technologies will enable a more efficient and automated approach. Here's how the future process will look:

1. Automated data extraction: Advanced algorithms will automatically extract and analyze hidden information from historical support tickets. This includes identifying patterns, common issues, and frequently used solutions.
2. Predictive analytics: Machine learning models will predict potential issues and provide proactive recommendations based on historical data. These models will continuously learn and improve from new ticket data.
3. Real-Time support: Sales representatives will have access to real-time support through an intelligent system that leverages the analysed ticket history. This system will provide instant answers to common questions, configuration details, and solutions to frequently encountered issues.
4. Reduced back-office dependency: With automated and real-time support, the need for manual intervention by back-office teams will be minimized. Sales representatives will be able to resolve many issues on their own, leading to faster response times and increased efficiency.
5. Continuous improvement: The system will continuously update and refine its knowledge base by learning from new tickets and support interactions. This will ensure that the support provided remains relevant and up-to-date.

By utilizing hidden information in support ticket history, the future process will enable sales teams to operate more independently, reduce the reliance on back-office support, and improve the overall efficiency of managing CTO B2B sales. This transformation will lead to faster issue resolution, better customer satisfaction, and optimized sales processes.

3.2 Offer or sales call outcome prediction

Today: predicting the outcome of offers, sales calls and other customer contacts in the B2B sales of complex industrial products is largely a manual and experience-driven process. Sales representatives rely heavily on their intuition, past experiences, and personal judgment to forecast the likelihood of closing a deal. This process involves:

1. Data collection: Sales reps gather information from various sources, including customer interactions, emails, and CRM systems. This data is often fragmented and not standardized.
2. Analysis: The gathered data is manually analyzed to identify patterns and insights. Sales reps look for indicators such as customer engagement levels, historical purchase behavior, and market trends.
3. Prediction: Based on their analysis, sales reps make educated guesses about the potential outcome of offers and sales calls. These predictions are subjective and can vary significantly between different sales reps.
4. Decision making: Sales strategies and follow-up actions are planned based on these predictions. The lack of standardized and data-driven predictions can lead to inconsistent decision-making and missed opportunities.

Future: the process of predicting the outcome of offers and sales calls will be transformed by advanced data analytics and machine learning technologies. This will enable a more accurate, data-driven, and standardized approach. The future process will involve:

1. Automated data collection: Advanced CRM systems and integrated data platforms will automatically collect and consolidate data from various sources, including customer interactions, emails, social media, and market data. This data will be standardized and stored in a centralized database.
2. Machine learning models: Sophisticated machine learning algorithms will analyze the collected data to identify patterns and correlations that are indicative of successful sales outcomes. These models will be trained on historical data to continuously improve their accuracy.
3. Real-time predictions: Sales reps will have access to real-time predictions generated by the machine learning models. These predictions will provide insights into the likelihood of closing a deal, potential obstacles, and recommended actions to improve the chances of success.
4. Data-driven decision making: With accurate and standardized predictions, sales strategies and follow-up actions will be based on data-driven insights. This will lead to more consistent decision-making, optimized sales processes, and increased sales efficiency.

By leveraging advanced data analytics and machine learning, the future process will enable sales teams to make more informed decisions, reduce the reliance on subjective judgment, and ultimately improve the success rate of offers and sales calls in the B2B sales of complex industrial products.

3.3 Intelligent configuration recommendation

Today: Preparing a product configuration for a B2B offer in the sales of complex industrial products is a manual and time-consuming process. Sales representatives rely on their expertise, customer requirements, and historical data to create customized configurations. Typical process steps include:

1. Requirement gathering: Sales reps collect detailed requirements from the customer, including specifications, budget constraints, and desired outcomes.
2. Manual analysis: Based on these requirements, sales reps manually analyze various product options and configurations. This involves consulting product catalogues, past configurations, and technical documentation.
3. Configuration proposal: Sales reps propose a product configuration that they believe best meets the customer's needs. This proposal is often based on their judgment and experience.
4. Customer feedback: The proposed configuration is presented to the customer, who may request modifications or additional features. This can lead to multiple iterations and adjustments.
5. Final offer: After several rounds of feedback and adjustments, the final product configuration is agreed upon and included in the B2B offer.

Future: the process of recommending product configurations for B2B offers will be significantly enhanced by an AI-driven tool. This tool will leverage advanced data analytics and machine learning to provide accurate and optimized product configurations. In the future, the process steps include:

1. Automated requirement analysis: The AI-driven tool will automatically analyze customer requirements, extracting key specifications, budget constraints, and desired outcomes from various sources such as emails, CRM entries, and direct inputs from sales reps.
2. Data-driven recommendations: Using historical data, product catalogues, and technical documentation, the AI tool will generate data-driven recommendations for product configurations. These recommendations will be based on patterns and insights derived from past successful configurations.

3. Real-time configuration proposals: Sales reps will receive real-time configuration proposals from the AI tool. These proposals will be optimized to meet the customer's requirements and constraints, reducing the need for manual analysis.
4. Interactive feedback loop: The AI tool will facilitate an interactive feedback loop with the customer. Sales reps can quickly adjust configurations based on customer feedback, and the AI tool will provide updated recommendations in real-time.
5. Optimized final offer: The final product configuration will be optimized and agreed upon more efficiently, with fewer iterations and adjustments. This will lead to faster proposal times and increased customer satisfaction.

By leveraging an AI-driven tool, the future process will enable sales teams to provide more accurate, optimized, and timely product configurations for B2B offers. This transformation will reduce the reliance on manual analysis, enhance the efficiency of the sales process, and improve the overall customer experience.

4. Internal and external data sources

AI-driven tools require a source for representative and high-quality data. The main data sources identified during InnoSale project are shortly described below.

4.1 Enterprise Resource Planning Systems

Enterprise Resource Planning (ERP) systems are comprehensive software platforms that integrate and manage all core business processes within an organization. These processes include finance, human resources, manufacturing, supply chain, sales, and procurement. ERP systems provide a unified view of business activities, ensuring data consistency and real-time insights.

ERP systems may handle a wide range of data types, including:

- Financial data: General ledger, accounts payable, accounts receivable, and financial reporting.
- Human resources data: Employee records, payroll, and performance management.
- Manufacturing data: Production planning, inventory management, and quality control.
- Supply chain data: Procurement, supplier management, and logistics.
- Sales and customer data: Sales orders, customer relationship management (CRM), and customer service.

ERP systems help organizations streamline operations, improve efficiency, and make informed decisions.

4.2 Project Repository

In the context of manufacturing complex industrial products, a project repository serves as a centralised storage and management system for all project-related data and documents. This repository is crucial for maintaining an organised and accessible overview of ongoing and past projects, ensuring that all relevant information is readily available for stakeholders.

The project repository stores various types of data and documents, including technical drawings, calculations, commercial documents, offers, and correspondence with customers. This centralisation helps in maintaining a comprehensive record of all project activities. It provides a user interface and services to manage documents efficiently. This includes version control, access permissions, and easy retrieval of documents, ensuring that the latest and most accurate information is always available. The repository maintains an overview of currently ongoing projects and serves as an archive for completed projects. This helps in tracking project progress, milestones, and deliverables, facilitating better project management and decision-making.

The project repository can be integrated with other enterprise systems, such as ERP and CRM, to streamline data flow and enhance collaboration across different departments. This integration ensures that all project-related data is synchronised and up-to-date.

It supports various data and document formats, accommodating the diverse needs of complex industrial projects. This includes handling large files, such as technical drawings and detailed calculations, which are essential for manufacturing processes.

4.3 Back-office support ticket systems

The back-office support ticket system is a comprehensive platform used for managing and tracking support requests and communications within an organisation. It is particularly useful in contexts such as manufacturing complex industrial products, where back-office support is often needed for tasks like offer creation. The support ticket data may contain information on customer requirements, product configurations, issues to be addressed etc.

4.4 Pricing tools

The pricing tools are used to manage and store detailed information about product pricing and configurations. They serve as a crucial tool for offer management, particularly in the context of complex industrial products. The pricing tools help sales teams access accurate and up-to-date pricing information, ensuring that offers are tailored to meet customer requirements effectively. By leveraging historical data and predefined pricing models, they streamline the process of creating customised offers, reducing the need for extensive manual calculations and back-office support

4.5 CPQ Systems

CPQ (Configure, Price, Quote) systems are software tools designed to help sales teams generate accurate quotes for complex and configurable products quickly and efficiently. These systems streamline the entire quote-to-cash process by automating product configuration, pricing calculations, and quote generation:

- Product configuration: Allows users to select and configure products based on customer requirements, including bundling, upselling, and cross-selling options.
- Pricing automation: Automates pricing calculations, taking into account discounts, custom pricing, and other pricing rules.
- Quote generation: Generates professional, branded quote documents that are accurate and consistent, and can be sent directly to customers.

The CPQ systems are used to reduce the time sales experts spend on generating offers, allowing them to focus more on selling. These systems contribute to minimizing errors in pricing and configuration, ensuring that quotes are accurate and consistent. In consequence, the sales cycle from quote to closing is accelerated by automating quoting, approvals, and contracting.

4.6 CRM systems

Customer Relationship Management (CRM) systems are tools for managing a company's interactions with current and potential customers. These systems help streamline processes, improve customer relationships, and enhance profitability. CRM systems integrate various functions such as sales, marketing, customer service, and support into a single platform, providing a unified view of customer interaction:

- Centralised customer data: CRM systems store all customer information in one place, making it easily accessible for all departments.
- Sales management: Track sales activities, manage pipelines, and forecast sales more accurately.
- Marketing automation: Automate marketing campaigns, segment customers, and track campaign performance.
- Customer service: Manage customer support tickets, track service interactions, and improve response times.
- Analytics and reporting: Generate reports and insights to make data-driven decisions.

5. AI experiments using authentic business data

5.1 Smart search to support offer creation

Section 3.1 provides an in-depth analysis of the requirements and methodology of the use case in question, which was provided by Konecranes. In practice, when a sales expert receives a customer inquiry for equipment that cannot be configured, a product support ticket is generated. When a sales expert creates a product support ticket with offer details, including most technical features already specified, the product support personnel utilize a semantic search tool to locate historical support tickets with similar content. The objective of this tool is to comprehend the content of the support ticket to identify analogous tickets and facilitate “search with meaning.” The search engine converts the text into a word-embedding format, compares it with historical data (all historical documents transformed into the same word-embedding space), and identifies relevant tickets using clustering algorithms. The product support team receives matching historical tickets as output and can use the information in them to manually prepare the offer with the required custom options and returns it to the sales expert. The semantic search tool aims to minimize lead time in these scenarios.

The foundation of the technical implementation is based on SentenceBERT[5]. SentenceBERT is a sophisticated language model built on deep neural networks, designed to convert a sequence of text into a numerical vector representation. A notable advantage of this vector representation is that it encapsulates the semantic information of the text. The semantic similarity between two pieces of

[5]Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.

text can be assessed by calculating the cosine similarity of their respective transformed vectors. This capability enables the use of SentenceBERT for semantic search applications.

The semantic search solution developed by VTT leverages the SentenceTransformers framework to deploy SentenceBERT. The semantic search components were implemented using the Python programming language.

At a high level, the solution includes two primary functions:

- Converting text into vector representations using SentenceBERT
- Searching for the most similar texts using cosine similarity from a set of transformed texts

Since the goal of the semantic search component is to identify the closest matching text from a large set of historical texts, the historical data needs to be converted into vector representations in advance. This is because transforming a large amount of data requires computation time, making it impractical to perform this task while the user is searching. The vectors produced during the transformation process (using SentenceBERT) are saved in a file, which is then utilized by the search functionality.

When the user executes the search, they provide a search phrase text as an input. This input is transformed into a similar vector representation as the texts in the historical data. This vector is then used to find the closest matches from the historical data by calculating the cosine similarity between the query vector and the vectors of the historical data. The text(s) corresponding to the most similar vectors are then returned as a result and displayed to the user.

The public demonstrator illustrated in Figure 2 is developed utilizing a consumer complaint dataset[6]. The user interface is designed exclusively for demonstration purposes; however, the underlying mechanism employs the SentenceBERT vector transformation and cosine similarity-based search as detailed in the preceding section.

[6] https://www.consumerfinance.gov/data-research/consumer-complaints/

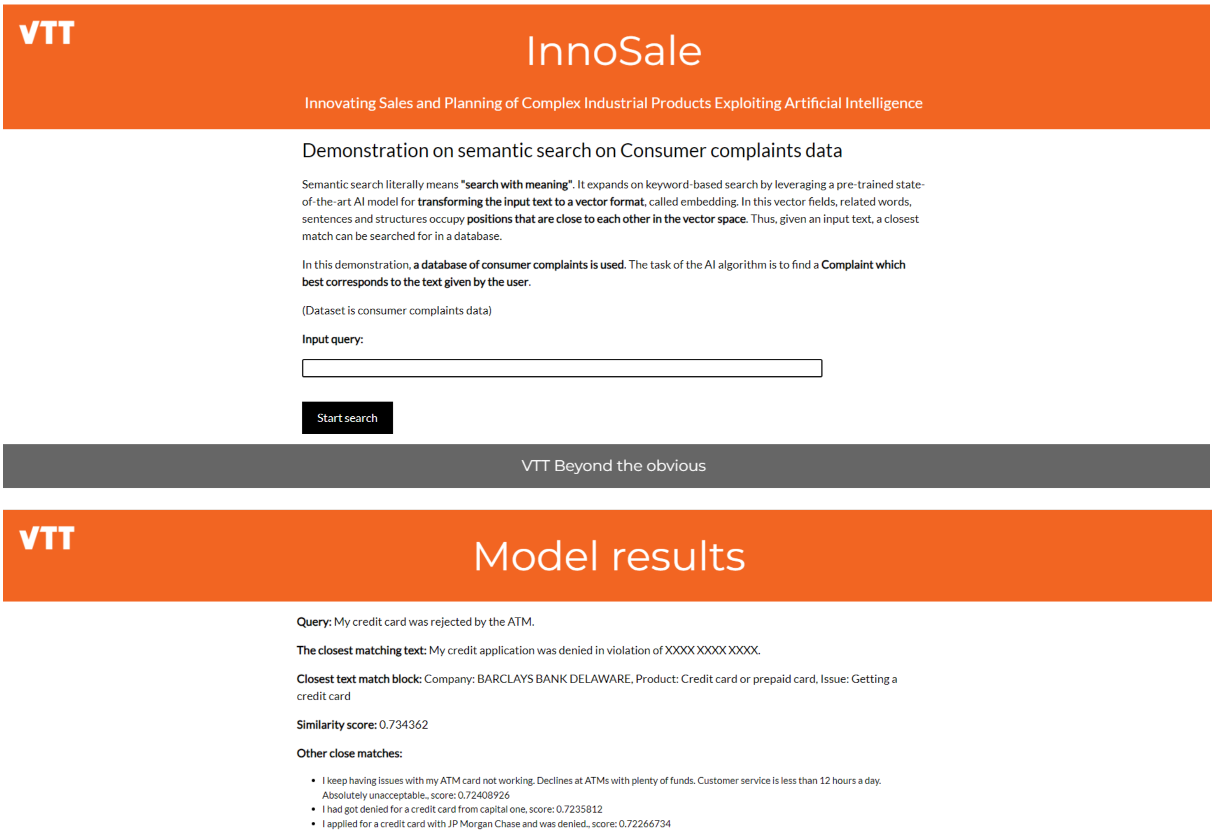

Figure 2. Screenshots of a demonstrator of the VTT semantic search component.

In addition to this also two other datasets have been experimented with to support the development of semantic search component; movie synopsis dataset[7] and arXiv dataset (dataset containing e.g. abstracts and titles of scientific articles published in arXiv.org[8]).

Utilizing the arXiv dataset, functionality for training a customized SentenceBERT model has been developed. This model serves as the foundation for semantic search, allowing for the adaptation of the semantic search component to various contexts and specialized vocabularies.

5.1.1 Semantic search on back-office support tickets

The goal of this task was to find relevant support tickets based on the user provided queries using SentenceBERT. Konecranes provided us with an anonymised dataset of support tickets and related metadata and predefined queries that could be used for confirming model performance. Additionally, Konecranes provided us with evaluation feedback on obtained results.

Ticket dataset consisted of multiple different features, of which mostly relevant were in three separate textual features: One consisting of short keywords, related to specialities of potential product, one providing a description of needed specialities and one providing additional context to said descriptions.

Because SentenceBERT could only process one-dimensional data, proper way to combine features was necessary. The first strategy was to encode each feature into separate corpus embeddings, and concatenate embeddings into single larger embedding. The second strategy was like the first, but

---

[7] https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset

[8] https://www.kaggle.com/datasets/Cornell-University/arxiv

embeddings were averaged instead. In the third strategy instead of encoding features separately, textual features were combined into single larger text sample and was encoded as it is. Additionally, to these cleaning of data was tried by removing unnecessary stopwords and special characters from the dataset and queries.

Based on the evaluation performed by the experts at Konecranes and their feedback on the search results concatenation of embeddings and averaging seemed to be more favoured strategies compared to fusing textual features together. This consensus was true also after cleaning of the data and queries were performed.

5.2 Prediction of sales activity outcome

Section 3.2 provides an overview of the use case in question. The aim of this experiment was to provide tools to support the prediction of sales calls and other customer contacts using historical sales data. An open dataset was used for experimenting with customer segmentation as a proof-of-concept. Due to the unavailability of B2B datasets, a dataset from the B2C domain was utilized.

The dataset used is sourced from the UC Irvine Machine Learning Repository and pertains to a banking institution in Portugal. It includes data from their direct marketing campaigns via phone calls to consumers, collected between 2008 and 2013, comprising a total of 52,944 records. The campaign's objective was to encourage subscriptions to a bank term deposit. The dataset features demographic information, prior customer behavior, temporal characteristics of the phone calls, details of previous contacts, and economic information of the time, such as the customer price index. A comprehensive description of the features is available in the repository. The dataset used in the original article was larger than the one uploaded in the repository, containing more features and samples. Consequently, the results in the paper are not directly comparable to the present study.

A standardized pipeline for data processing was developed to adapt to various open datasets within the B2C domain. A flow chart illustrating this process is provided in Figure 3. The pre-processing phase indexes the file based on user input and handles categorical and numerical values separately. Irrelevant columns are removed, and users have the option to specify rows with certain values (e.g., not-a-numbers or unknowns) for deletion. Additionally, the column "age" is segmented into categories to reduce data volume according to user specifications.

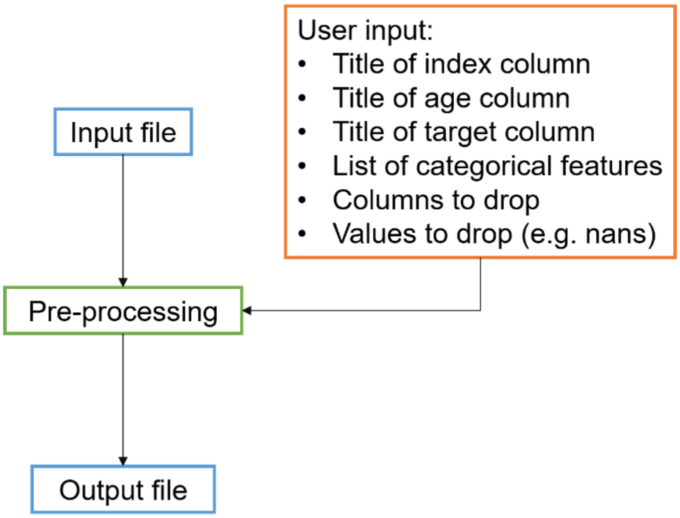

Figure 3. Common pipeline for data processing.

The dataset is split into train (90%) and test sets using a user-defined random seed for accurate evaluation. Stratification is based on the target value, which the user specifies. XGBoost is the base model used. The tool also performs modular customer segmentation via clustering and adds segmentation labels to the data for enhanced prediction.

The pipeline trains and evaluates two models:

- XGBoost
- Dummy classifier with random guessing

The baseline classifier ensures the model outperforms random chance. Models are evaluated using balanced accuracy to account for class imbalance. SHAP values interpret which features affect the model output.

Clustering identifies customer segments via UMAP for dimensionality reduction and spectral clustering. The Davis-Bouldin Index determines the optimal number of clusters. SHAP helps identify key differences between segments by training an XGBoost on cluster labels. New customers can be mapped to these segments, with feature weighting allowing relevance specification for customer similarity.

Table 1. Classification results on the test set using different random splits.

	Balanced accuracy [%]	
Random split	XGBoost	Dummy
1	63.39	46.51
2	64.71	46.21
3	62.23	50.22
4	64.29	50.07
5	63.29	50.07
mean	**63.58**	**48.62**

The XGBoost model achieved a mean balanced accuracy of 63.6% (Table 1). Figure 4 shows the absolute mean SHAP value for each feature, indicating that "euribor3m" and "nr.employed" significantly impact the model. Figure 5 reveals that when these feature values are high (red), the model's output is negative; when low (blue), the impact is positive. This makes sense: "euribor3m" represents the 3-month Euribor during data collection, and "nr.employed" is the number of bank employees—both reflecting the economic situation. Thus, a poor economy correlates with fewer customers subscribing for deposits. The impact of other features is less clear. The "contact-//-cellular" feature (1 for cell phone contact, 0 otherwise) slightly benefits the model. Conversely, the "campaign" feature (1 for associated with a campaign, 0 otherwise) has a slight negative effect.

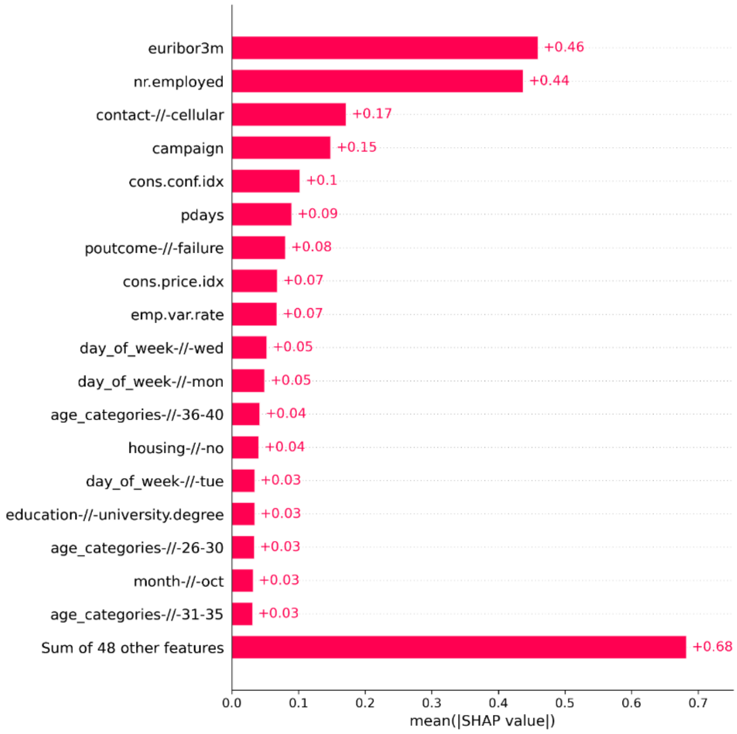

Figure 4. Global SHAP feature importance plot.

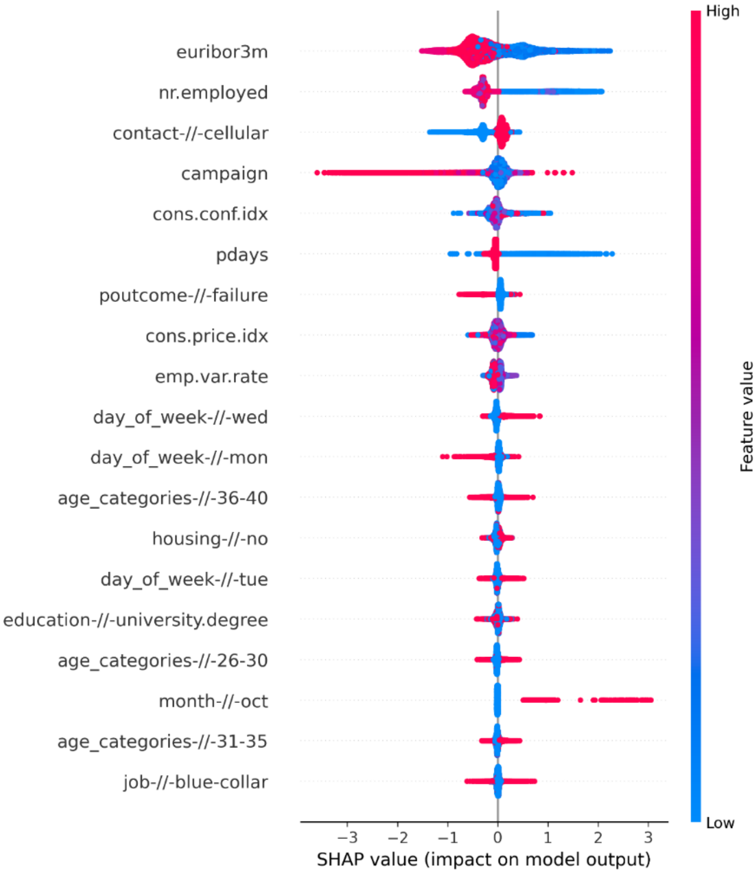

Figure 5. Local SHAP explanation summary plot.

Figure 6 shows the clustering results, with the algorithm identifying 19 segments. Figure 7 explains the key features used for this segmentation. For instance, the "cons.price.idx" (consumer price index) is significant for segments 6, 17, and 12. Additionally, the feature "month-aug" distinguishes segment 16 from others.

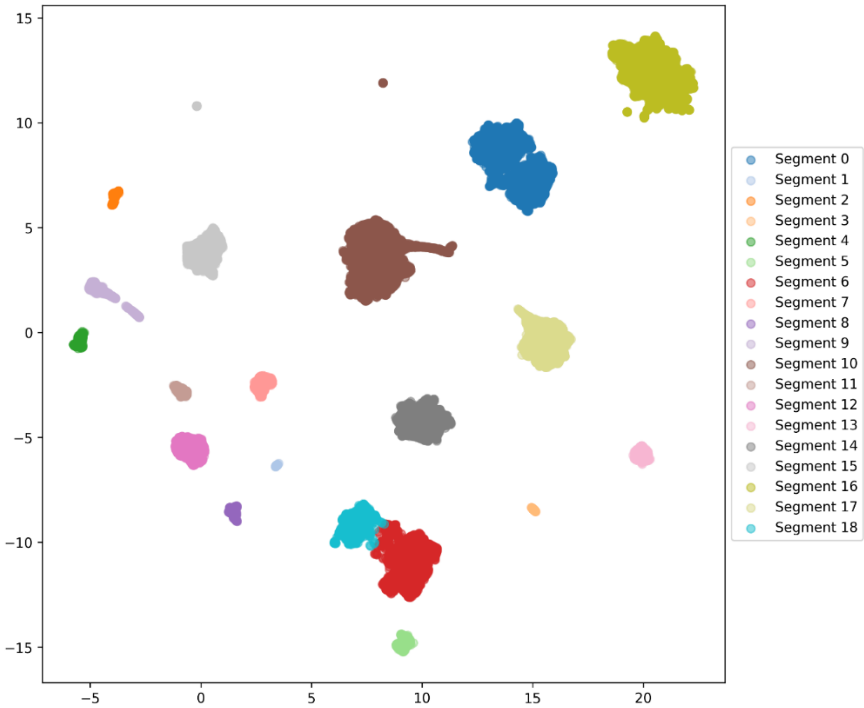

Figure 6. Clustering results.

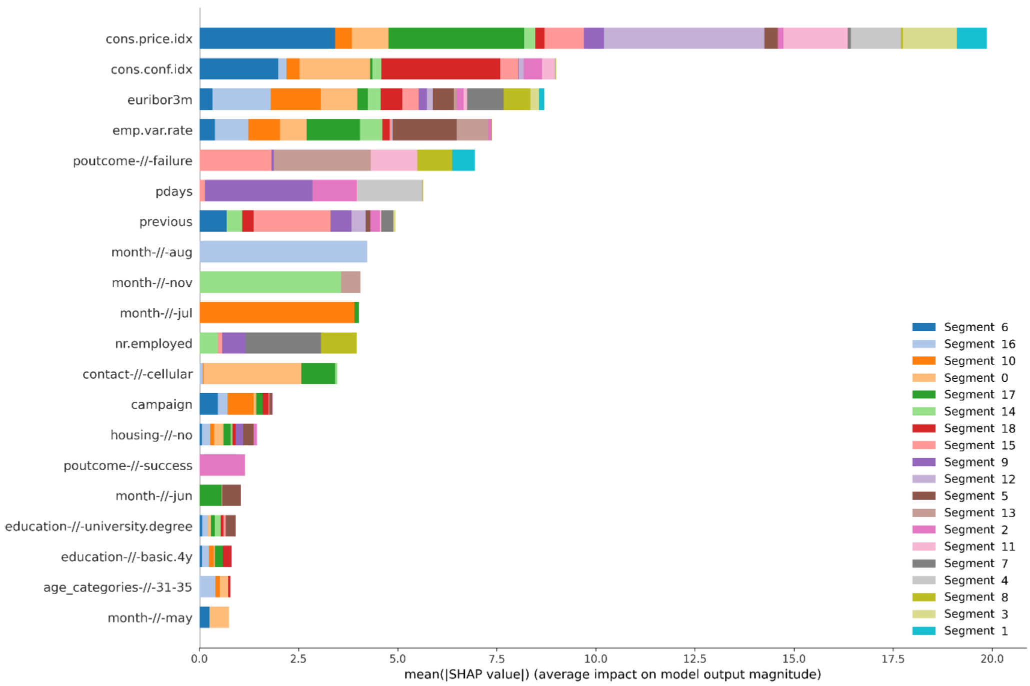

Figure 7. Global SHAP feature importance plot for the segments.

Combining the classification and clustering experiments provides new insights on the results. The key features for classification are “euribor3m” and “nr.employed”. Segment 7 stands out due to "nr.employed", but its value is average; thus, it is not linked to success. “contact-//-cellular”, the third most important feature, differentiates segment 0, which consists of people not contacted via cell phone. This segment shows low promise for targeted marketing with only a 4% success ratio. Segment 9 has the highest success ratio at 301%, with 464 positive and 154 negative targets. It is distinguished by the feature “pdays”, representing the number of days since the last contact from a previous campaign, averaging 6.7 compared to 999 for other segments.

5.2.1 Sales call outcome prediction

LeadDesk provided an anonymized dataset comprising scripts and call outcomes to support the experimentations. The objective of VTT activities in this use case was to analyze which scripts were most effective and develop a model to forecast script success.

The dataset included scripts for various campaigns utilized by the calling agents. Each script consisted of one or more pages. For each call, a campaign identifier, call result, and the specific script pages used were recorded, along with other pertinent details such as call duration. Exploratory data analysis was employed to examine the dataset.

For certain campaigns, the result was a binary value (deal/no deal), while for others, the result was encoded to be a more detailed target. For instance, in one case, the target was categorized as "Positive," "High potential," "Negative," or "Other." Descriptive statistics were computed, and figures displaying relevant information were generated for illustrative purposes.

Data from two of LeadDesk's customers were selected for analysis; ultimately, the dataset with larger amount of data available was selected for the experiment. The objective was to predict the outcome of sales calls based on features accessible prior to making the calls. The analysis process followed the steps outlined in Section 5.2. The sales call results were categorical but were binarized into "Deal" or "No Deal" categories to support the experiment.

Relevant features were extracted from the dataset, including attributes of the manuscript used, temporal features such as the hour of the day, day of the week, and month, as well as attributes of the call itself. These features were one-hot encoded, and an XGB classifier was utilized for the analysis. The train-test split ratio was 85-15%. SHAP was employed to identify the most influential features and their impact on the model's output.

The XGB model demonstrated a balanced accuracy of 64.78%, compared to the benchmark (coin flip) accuracy of 55.98%. Therefore, it can be argued that the model could potentially enhance sales outcomes in operational use.

Additionally, experiments were conducted using the call transcript data. The data provided by LeadDesk contained anonymized conversations between the salesperson and the customer in text format (data was processed through speech-to-text tools by LeadDesk). Research question in the experiment was to study, if the outcome of the call can be predicted while the call is still ongoing. This could provide guidance to the salesperson when deciding on the next steps during the sales process. In the experiments, limited number of characters were used from the transcript (counting from the beginning of the transcript). This was repeated with different length limitations and experimented with different machine learning approaches. As a conclusion, the prediction accuracy starts to be better that a random guess around 1000 characters from the beginning of the transcript and reaches a plateau of ~0.9 around 2000 characters (the accuracy does not increase when using more chacters than this). However, this can be explained by the fact that the calls that do not lead into sales are shorter and around 1000-2000 characters it is often evident how the call ended. The end reason prediction for transcripts were experimented with using word frequencies as an input for a machine learning model, augmenting the word frequencies with sentiment information, using LLM model embeddings as an input for a ML model and using RAG (retrieval augmented generation) approach. Out of these approaches, the first three approaches provided approximately equal results while the RAG approach was significantly less accurate.

5.2.2 Offer outcome prediction

The objective of this task was to predict the outcome of sales offers (i.e., whether an offer was won or lost). Konecranes provided an anonymized dataset containing customer information, sales cases, and product configuration data in CSV format. VTT developed the machine learning model for this purpose.

The dataset included anonymized customer information such as country and city, along with sales case data. Each sales case record contained an offer result (won or lost), which served as the target variable for the model. Additional business-related information, including area, organization, and sales case value, was also provided. Supplementary features were calculated, such as the number of updates made to each offer. Moreover, configuration data of the offered product was also included in the dataset. Based on this comprehensive data, a model was trained to predict the outcome of future offers.

The most relevant features for predicting offer outcomes were identified. In our analysis, some features within the dataset were determined to be of low importance. Techniques such as Shapley Additive Explanations (SHAP) and Random Forest feature importance were utilized to identify these unimportant features. We achieved a balanced accuracy of 83.9% in the dataset, which is a very good result.

5.3 Intelligent configuration recommendation

The objective of configuration recommendation, as discussed in Section 3.3, is to assist sales experts during the offer creation process. Complex industrial B2B products, such as cranes or waste containers, involve numerous parameters that must be configured when generating an offer. This procedure demands significant manual effort and the application of tacit knowledge accumulated by the sales expert over time. To expedite the configuration process and leverage historical offers effectively, data analytics and machine learning methods were employed to analyze past data and generate initial configurations for offers. These configurations serve as the foundation for new offers and can be further refined by the salesperson as necessary.

Molok served as a use case provider for a configuration recommendation project in the domain of waste management. VTT developed a method to generate initial configurations utilizing data exported from the CRM system. The approach involved the following high-level steps:

- Data was provided in an Excel spreadsheet format. Given the objective of generating recommendations for a specific geographical area, the data was initially converted into a machine-readable format and filtered based on the country and postal code specified by the user. Geographical information could be optionally excluded with no preliminary filtering applied.
- Each offer in the dataset included a variable number of rows, depending on the products encompassed within the offer. Since this format is unsuitable for data analytics tasks, it was transformed into a hierarchical representation where each offer comprises subcomponents detailing the offered product and configuration specifics.

- Frequency analysis was subsequently performed. This step aimed to assign a numerical representation to each offer, presenting them as vectors of uniform length, which facilitates further processing.
- The offers, depicted as vectors, were then clustered employing the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) method.
- The most common configurations within each cluster were identified and provided as recommended configurations.

An implementation was also carried out to visualize the clustering results for development purposes.

A method for evaluating the performance of the approach was developed. The evaluation involves dividing the dataset into training and testing data based on time. The training data is used to generate recommendations, and the testing data is then used to determine how many times the recommended configuration was applied. This simulates the scenario where a salesperson requests a recommended configuration based on historical data collected up to the present day, and the success of the offer in the future is assessed. The results provided by the algorithm were compared to “dummy recommender” (using random past offer as a recommendation). The recommended configurations provided by the algorithm occurred more often in the testing data.

In addition to the previously described evaluation of the recommendation algorithm, the algorithm was integrated into a demonstration version of a CPQ system used by Molok sales experts for expert evaluation and collection of feedback. The 10 participants worked in sales and marketing (8 persons) or customer service (2 persons). Most of them were not familiar in the usage of AI tools, but some of them had at least tried an AI tool or were using them regularly. The experts were relatively critical regarding the challenge the developed application is trying to solve. As they are experienced salespersons, they did not consider having such a tool necessary. From their perspective the demonstration was for example lacking essential features or the recommendations were outdated. However, the less experienced sales experts found the demonstrated application slightly more beneficial and the persons with no background on AI usage were more positive towards the demonstrated application. Overall, the experts were also relatively positive on the future usage of such AI-driven tools and provided many useful suggestions for the future developments. The expert feedback is highly beneficial in the future development of the intelligent configuration recommendation system.

6. Key findings from the InnoSale project

6.1 Importance of data quality

High-quality data is the foundation for any successful AI-driven solution in B2B sales. Throughout the InnoSale project, it became clear that the effectiveness of AI models—whether for semantic search, sales outcome prediction, or configuration recommendation—depends directly on the accuracy, completeness, and consistency of the underlying data. Poor data quality can lead to unreliable predictions, misinformed recommendations, and ultimately, a lack of trust in the system. The project’s experiments, especially those involving real business data from partners, highlighted the need for careful data preprocessing, cleaning, and integration from multiple sources (ERP, CRM, support tickets, etc.) to ensure that AI tools deliver meaningful value.

6.2 Importance of trust

Trust emerged as a critical factor in the adoption and effectiveness of AI tools. For users, especially sales experts with deep domain knowledge, to rely on AI-driven recommendations or predictions, they must trust both the data and the algorithms. Transparency in how AI models reach their conclusions (e.g., using SHAP values to explain feature importance) and clear communication about the limitations of the models are essential. The project found that trust is built not only through technical accuracy but also through openness about what the AI can and cannot do, and by involving users in the evaluation and feedback process.

6.3 User acceptability

User acceptability is closely linked to both trust and perceived usefulness. The InnoSale project's feedback sessions revealed that experienced sales professionals may be skeptical of AI tools, especially if they feel these tools do not address their real-world challenges or lack essential features. Conversely, less experienced users or those unfamiliar with AI were more open and positive about the demonstrated applications. In addition, the evaluation results revealed management's more positive attitude towards AI solutions, while the sales experts were more skeptical or even afraid of losing their own positions. This suggests that successful adoption requires not only robust technology but also user-centered design, training, and change management to address varying levels of expertise and openness to new tools.

6.4 Added-Value of AI

A recurring insight from the project was that sometimes even basic automation ("usein ATK:kin riittää", often IT is enough) can deliver significant value. While advanced AI models can provide powerful insights, many sales process improvements can be achieved through simpler digitalization and automation of routine tasks. The project demonstrated that AI adds the most value when it augments human expertise, automates repetitive work, and provides actionable insights—freeing up experts to focus on higher-value activities. The feedback from partners like Konecranes, Molok, LeadDesk, and Wapice reinforced that the real-world impact of AI is maximized when it is integrated thoughtfully into existing workflows and addresses concrete user needs.

7. Discussion and future work

The InnoSale project provided a starting point for the research and development work on AI-driven tools for B2B sales. However, there were some limitations:

- **Customer segmentation without specific customer data:**
  One of the main limitations was that customer segmentation experiments were performed without access to actual, detailed customer information. Instead, segmentation relied on indirect or anonymized data, which limits the precision and applicability of the findings to real-world B2B sales scenarios.
- **Use of B2C data for B2B contexts:**
  Due to the unavailability of suitable B2B datasets, some experiments (such as sales outcome prediction) used open datasets from the B2C domain in order to generate public research results. This means that these results may not fully reflect the complexities and nuances of B2B sales processes.

- **Generalizability of results:**
  Since some models were trained and validated on data that does not perfectly match the intended industrial context or on data from a specific company, the generalizability of the results to other companies or industries may be limited.
- **Expert feedback and user acceptance:**
  Feedback from experienced sales professionals indicated scepticism about the usefulness of AI-driven tools, especially if the tools did not address their real-world challenges or lacked essential features. This highlights the importance of user-centered design and the need for further development to increase user acceptance.
- **Data quality and integration:**
  The effectiveness of AI models depends heavily on the quality, completeness, and integration of data from multiple sources (ERP, CRM, support tickets, etc.). Any shortcomings in data quality can reduce the reliability of predictions and recommendations.
- **Evaluation in real business settings:**
  Solutions were only evaluated in demonstration environments or with limited real-world feedback, which may not capture all practical challenges or user needs.

The InnoSale project has laid a strong foundation for the development and deployment of AI-driven tools in B2B sales, but several important avenues for future work have emerged. One of the primary recommendations is to improve access to richer and more detailed customer data. Future research should prioritize partnerships and data-sharing agreements that enable the use of authentic, granular B2B sales data, or investigate possible usage of synthetic data. This will allow for more precise segmentation, better model training, and ultimately, more actionable insights for sales teams.

Another key area for future work is the adaptation and validation of AI models in genuine business environments. While many of the project's solutions were tested in demonstration settings or with anonymized data, their effectiveness and usability in real-world sales processes remain to be fully proven. Piloting these tools in live sales environments, gathering feedback from end-users, and iteratively refining the solutions will be essential steps to ensure practical impact and user acceptance.

User acceptance itself is a critical theme for future development. The future work should emphasize user-centered design, involving sales experts throughout the development process to ensure that solutions are both relevant and intuitive. Training, change management, and transparent communication about the capabilities and limitations of AI will also be vital to foster trust and encourage adoption.

Data quality and integration remain ongoing challenges. The effectiveness of AI models is directly tied to the accuracy, completeness, representativeness, and consistency of the underlying data. Future initiatives should invest in robust data governance, improved data integration across systems (such as ERP, CRM, and support ticket platforms), and advanced preprocessing techniques to ensure that AI tools are built on a solid foundation.

Finally, the project highlighted the value of even basic automation and digitalization in improving sales processes. While advanced AI offers significant potential, it also uses substantially resources while many organizations can achieve substantial gains through simpler IT solutions that automate routine tasks and streamline workflows. Future work should balance the pursuit of cutting-edge AI with pragmatic steps toward digital transformation, ensuring that technology augments human expertise and delivers tangible value in everyday operations.

8. Summary

The InnoSale project, from 2021 to 2025, set out to transform the sales processes for complex industrial products in B2B environments by harnessing the power of artificial intelligence. The project brought together a diverse international consortium of industry leaders and research organizations, all united by the goal of making sales processes more efficient, accurate, and customer-centric. At the heart of InnoSale was the recognition that traditional B2B sales, especially for highly configurable products, relied heavily on manual expertise, fragmented data, and intuition, which often led to inefficiencies and inconsistent outcomes.

To address these challenges, the project focused on developing and piloting AI-driven tools that could automate routine sales tasks, provide data-driven insights, and support decision-making throughout the sales funnel. VTT and its partners co-created use cases that reflected real-world needs, such as semantic search tools to help sales and support teams quickly find relevant information from historical support tickets, machine learning models to predict the outcomes of sales calls and offers, and intelligent recommendation systems to streamline the configuration of complex products. These solutions were built on a foundation of integrating data from multiple enterprise systems (ERP, CRM, support tickets, and pricing tools) highlighting the critical importance of data quality, completeness, and integration.

Experiments conducted during the project demonstrated that AI models could achieve promising accuracy in tasks like offer outcome prediction and customer segmentation, even when working with anonymized or B2C datasets due to the scarcity of real B2B data. The project also underscored the importance of trust and user acceptance: while management tended to be optimistic about AI's potential, experienced sales professionals were often skeptical, emphasizing the need for transparent, explainable models and user-centered design. Feedback from these users will be invaluable in refining the tools and identifying areas for further development.

Ultimately, InnoSale showed that while advanced AI can deliver significant value, even basic automation and digitalization can lead to substantial improvements in sales processes. The project concluded that the most successful AI solutions are those that augment human expertise, automate repetitive work, and provide actionable insights freeing up sales experts to focus on higher-value activities. Looking ahead, the report recommends further work to secure richer B2B datasets, pilot AI tools in real business settings, and invest in robust data governance and user training. InnoSale thus lays a strong foundation for the continued evolution of AI-driven sales in complex industrial domains, balancing technological innovation with practical, user-focused implementation.

References

[1] Syam, Niladri; Sharma, Arun. Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice. Industrial marketing management, 2018, 69: 135-146.

[2] Cuevas, Javier Marcos. The transformation of professional selling: Implications for leading the modern sales organization. Industrial Marketing Management, 2018, 69: 198-208.

[3] Hall, Kelly R., et al. Understanding salesperson intention to use AI feedback and its influence on business-to-business sales outcomes. Journal of Business & Industrial Marketing, 2021

[4] Mattila, Malla; Yrjölä, Mika; Hautamäki, Pia. Digital transformation of business-to-business sales: what needs to be unlearned? Journal of Personal selling & sales Management, 2021, 41.2: 113-129.

[5] Prieto, Fábio; BRAGA, H. Contribution of Artificial Intelligence in B2B Sales: A Danfoss Case Study. International Journal of Advanced Engineering Research and Science, 2021, 8.4: 225-234.

[6] Saura, Jose Ramon; Ribeiro-Soriano, Domingo; Palacios-Marqués, Daniel. Setting B2B digital marketing in artificial intelligence-based CRMs: A review and directions for future research. Industrial Marketing Management, 2021, 98: 161-178.

[7] Paschen, Jeannette; Wilson, Matthew; Ferreira, João J. Collaborative intelligence: How human and artificial intelligence create value along the B2B sales funnel. Business Horizons, 2020, 63.3: 403-414.

[8] Jelinek, Ronald. All pain, no gain? Why adopting sales force automation tools is insufficient for performance improvement. Business Horizons, 2013, 56.5: 635-642.

[9] Kristjansdottir, Katrin, et al. The main challenges for manufacturing companies in implementing and utilizing configurators. Computers in Industry, 2018, 100: 196-211.

[10] Singh, Jagdip, et al. Sales profession and professionals in the age of digitization and artificial intelligence technologies: concepts, priorities, and questions. Journal of Personal Selling & Sales Management, 2019, 39.1: 2-22.

[11] Mahlamäki, Tommi, et al. Adoption of digital sales force automation tools in supply chain: Customers' acceptance of sales configurators. Industrial Marketing Management, 2020, 91: 162-173.

[12] Blazek, Paul; Pilsl, Klaus. The importance of analyzing configurator usage data to leverage product innovation and sales performance. 9th International Conference On Mass Customization and Personalisation, Sep. 23 -25. 2020.

[13] Rohaan, D.; Topan, E.; Groothuis-Oudshoorn, C. G. M. Using supervised machine learning for B2B sales forecasting: A case study of spare parts sales forecasting at an after-sales service provider. Expert Systems With Applications, 2022, 188: 115925.

[14] Hannah Bast, Björn Buchhold and Elmar Haussmann (2016), "Semantic Search on Text and Knowledge Bases", Foundations and Trends® in Information Retrieval: Vol. 10: No. 2-3, pp 119-271. http://dx.doi.org/10.1561/1500000032