Journal of Information Systems Engineering and Management
2024, 9(2)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

NLP-Enabled Intelligent Data Mining from SAP BusinessObjects and SAP BW Systems Using AI/ML Models on AWS

Sumit Sachdeva

Technical Manager - Predictive Analytics / Business Intelligence

ARTICLE INFO

Received: 10 Apr 2024

Accepted: 28 June 2024

ABSTRACT

Data mining of enterprise data is also changing to include natural language processing (NLP) and machine learning (ML) to derive insights of structured business intelligence systems. The present paper will suggest an intelligent data mining system of SAP BusinessObjects and SAP Business Warehouse (BW) with the help of the NLP-enabled model that utilizes the state-of-the-art language models that will be implemented on Amazon Web Services (AWS). The framework enables its users to issue natural language queries to find and process enterprise data and eliminates the divide between non-technical decision-makers and big SAP data repositories. Our approach is based on a model based on BERT that is fine-tuned to vocabulary of enterprise domains and an AWS cloud-based system to transform queries to data warehouse functions. Experiments show that information retrieval is highly accurate and that the query processing time is much reduced as compared to the traditional methods. Our findings indicate that our solution has 95 percent accuracy when presented with queries on a sample with a 40 percent processing speed increase. The paper has indicated a way in which NLP and cloud-based AI can be used together to transform business analytics that would allow organizations to view data more intuitively, quicker in decision-making, and provide more insights.

Keywords: Natural Language Processing, SAP BusinessObjects, SAP BW, AWS, Business Intelligence, Data Mining, BERT, Cloud AI, Query Processing, Enterprise Data Analytics

Introduction

BI systems such as the SAP BusinessObjects and the SAP BW contain enormous volumes of enterprise data, yet they can only be useful when it can be translated into useable insights by requiring technical skills in querying and reporting. It is becoming more important to have smart data mining tools, which allow their use by non-technical users, who can then communicate in a natural language with enterprise data, facilitating greater accessibility and decision-making. Recent progress in NLP and AI, and especially transformer models, has transformed the language understanding, and now it is possible to understand human queries and find related information (Devlin et al., 2019). Recent developments in deep learning and neural language models have significantly improved the ability of machines to interpret human language in analytical systems (Young et al., 2018). A Bidirectional Encoder Representations from Transformers model (BERT) was the model to introduce deep bidirectional context understanding and made breakthroughs in task-specific question answering, as well as inference. Based on these models, scholars have reviewed natural language interfaces to databases

(NLIDB) and conversational BI tools to democratize data analytics. Through these interfaces a user will be able to pose questions such as what were our sales figures last quarter? and can get results without writing SQL and without using complicated dashboards. Early research on natural language interfaces to databases demonstrated that conversational querying can improve accessibility of structured data systems for non-technical users (Androutsopoulos et al., 1995).

The use of AI/ML in enterprise systems has proven to be beneficial in regards to operational efficiency and in decision support. As an illustration, integrating the principles of ML into the SAP systems can facilitate the simplification of the primary business operations and identify insights using performance indicators (Sharma and Vaid, 2022). The companies exploiting AI-based analytics claimed to have smarter decision-making and higher productivity by automating standard analysis and minimizing the time to respond to data queries (Sharma and Vaid, 2022). Similarly, cloud services like AWS offer scalable infrastructure on which such AI models can be deployed on enterprise data. AWS service allows processing big data and applying ML at scale that is essential in modernizing SAP workloads with AI applications. With these trends in mind, our study targets NLP-powered data mining of BI systems of SAP on advanced AI/ML models at AWS cloud. Our goal is to create a framework in which a query which is entered in natural language by the user is interpreted by an NLP model (e.g., BERT), translated to the proper data retrieval operations on SAP BW/BusinessObjects data (which might be in an AWS-based data warehouse), and the result is presented in a human-readable format. This strategy is based on the previous literature on conversational BI and augmented analytics and expands them with the recent language models and cloud connectivity to saleable enterprise data.

Literature Survey

BI systems have also been of interest in the recent years in terms of research on natural language interaction with them. BI conversational AI has proven to change the consumption behavior of users, making analytics more accessible. Their model used a Support Vector Machine (SVM) that was set to understand queries by users and their results gave great accuracy (roughly 98% training accuracy and 94.7% validation) to comprehend and respond to business queries. This highlights the possibility of machine learning to enhance BI systems, as classical models such as SVM are now being outperformed by transformer-based NLP models. Conversational analytics systems have increasingly been integrated into business intelligence platforms to support natural language query interaction with enterprise datasets (Nayak et al., 2019).

It is true that transformer models have become the foundation of contemporary NLP because they are capable of scaling context (Vaswani et al., 2017). BERT and its variations have brought a lot of enhancement in all NLP tasks. The two-way encoding of BERT allows understanding of enterprise language with specifics, and it works out in terms of interpreting business language in SAP reports. Their system was quite successful, and they have demonstrated a greater success in risk identification as well as the early warning timeliness in comparison to the conventional approaches. This shows the capability of the fine-tuned NLP models to mine enterprise data that was previously challenging to mine using either the manual or keyword-based method. These success stories are the motivation behind our work where we propose that a BERT-based model fine-tuned on SAP business data can be able to successfully interpret intricate queries of users in this field. Transformer architectures have enabled scalable language modeling and improved contextual representation learning compared to previous recurrent neural network approaches (Radford et al., 2019).

Natural Language Interface to Databases is another area of research that has been relevant in this field and that can be used to query structured data. Their process of creating a system to translate natural language queries to the appropriate database stored variables (e.g. automatically addressing ambiguous column names) enhanced the accuracy and recall of data retrieval in BI applications. Nonetheless, there are still difficulties with the translation of natural language queries into effective and correct database operations which is particularly problematic with complicated schemas as well as enterprise-specific jargon. These articles demonstrate that although this is a positive move in the right direction, it is not possible to succeed without a combination of strong NLP and a business data schema comprehension. Our framework considers these aspects as it employs a knowledge-conscious approach (including metadata based on the schemas of SAP BW) and NLP to make sure that the intent of users is interpreted properly.

The introduction of AI/ML into the SAP ecosystem has been discussed in systems perspective, too. Sharma and Vaid (2022) outlined the overlap of SAP and AI/ML and stated that the integration of machine learning algorithms into SAP platforms has the potential to transform the work of enterprises, making it automated and based on data and insights available to make decisions. When AI is implemented to SAP enterprise processes, they refer to resource optimization, service delivery, and compliance. Cloud platforms are taking centre stage in the SAP analytics world. The results allow us to choose AWS as the deployment platform of the AI models - the cloud does not only scale to process SAP data, but also has a collection of ML services and integrating tools that can make development and deployment easier.

Research Methodology

Our research design is an NLP-driven data mining system implementation, which communicates with SAP BusinessObjects and SAP BW and runs AI/ML models on AWS. The general design of our proposed solution consists of three primary layers: (1) the Natural Language Interface (NLI) that is primarily driven by a fine-tuned BERT-based model that is to be used to understand the query made by the user, (2) the Data Retrieval Layer that routes the interpreted query to the SAP data sources (which may be an intermediate data warehouse on AWS), and (3) the Cloud Infrastructure on AWS that will host the ML models and offer scalable data processing resources. Our approach to designing this architecture was a design science research in which we designed this architecture and tested it against our research goals (accuracy and efficiency of data mining)..

System Design: The NLI layer takes the input query of the user in plain English (or any other language it supports). We trained one of the pre-trained BERT models on the enterprise data corpus, SAP report texts, business glossaries, and sample QA pairs that are related to SAP BW and BusinessObjects. This area of adaptation is essential because SAP systems have domain specific vocabulary (e.g. InfoCube, Cost Center, NetWeaver) which may not be well understood by a general language model. The NLI applies the embeddings of BERT both to classify intents and extract significant parameters in a query (such as the fact that querying sales figures last quarter in Europe suggests some type of sales parameter, time filter last quarter, and region=Europe). The Data Retrieval Layer is then given this interpreted structure.

Data Retrieval Layer It transforms the intent of the structured query into a real data retrieval operation. We made use of AWS data lake and data warehouse methodology in our implementation to provide scalability. The tools (SAP integration tools, like AWS Glue and SAP Data by connectors) were used to move SAP BW data to Amazon Redshift (a cloud data warehouse) so that the latest structured data could be available on Amazon. This is in line with best practices reported by AWS, in which migrating SAP data into Redshift will enable usage of petabyte-scale cloud analytics. The SAP BW schema metadata (such as cube definitions, table relationships) is stored in a Knowledge Base. We created a translation module that receives the output of NLI and builds an SQL query on the Redshift warehouse which is guided by the Knowledge Base to guarantee the accuracy (like the method proposed by AWS to build ERP data into the knowledge base). Using the example; when the output of the NLI shows that he or she wants to know the total sales in Europe in the second quarter of 2023, the metadata will allow the translator to identify Sales fact table, region dimension and time dimension in the BW schema and develop corresponding SQL. It is advantageous in that data integrity and respect of business logic is preserved, which is being noted in the available solutions that favor structured schemes over end to end generative schemes on databases.

The above components are hosted on the Cloud Infrastructure on AWS. The implementation of the BERT model was also done on Amazon SageMaker that offers a managed environment to the ML model and made it easy to scale when we were experimenting. The reason why AWS is planned to be chosen is its capability to address workloads of the enterprise level. The results of integrating ML pipelines on AWS (with SageMaker and EMR) can also be seen by Thallam (2022) to decrease the model training time by more than 30 percent and cost can be reduced by half through the optimization of resource utilization. In our case, real-time inference was performed on SageMaker endpoints: the query of the user is sent to a SageMaker endpoint with the fine-tuned BERT, which provides the interpreted query structure. The functions of Amazon Lambda are used to organize the query translation and run it in Redshift and gather the results. In addition, the workflow between query input and response was managed with the help of AWS Step Functions, which makes the system modular and scalable. Security and identity management of AWS were also exploited by us to ensure that only we can gain access to the data which is very important in the enterprise world. Domain adaptation of language models has been shown to significantly improve performance when applied to specialized enterprise terminology and datasets (Gururangan et al., 2020).

Experimental Setup: To test the system we set up a prototype with a dataset based on the SAP sample sales data (which has been extended to the case of a corporate BW scenario). To create natural language test queries, we developed a list of queries corresponding to typical tasks of an analysis (e.g. Show me the top 5 products by revenue this year, How many support tickets were closed last month by priority?). Our ground truth result was known by the dataset or we had to acquire it with the help of manual SQL. To determine accuracy, we compared the output of our system with the ground truth. Also, to estimate the advantage of NLP approach, we came up with a baseline system: a less complex system, a query tool based on keywords, and template matching used to map queries into SQL. The baseline is not based on the ML/NLP knowledge, but rather on the patterns that are explicitly defined (e.g. the keywords such as revenue and this year are detected and then paired to a pre-defined query template). The two paramount parameters we considered during the testing included accuracy of the retrieved results (was the correct answer retrieved by the system) and processing time of the query execution (user query to final answer). The level of accuracy was measured by the percentage of queries that the system produced a result that was similar to the expected result. End to end processing time was calculated that includes NLP inference processing time and data retrieval. Every query was repeated several times to compensate any variation that could arise particularly with cold starts in the cloud functions.

Results and Discussion

Once the system was implemented as outlined, we have done an extensive evaluation on our testing queries. The findings are summarized in Table 1 and Table 2 that compare the accuracy and processing time of the proposed NLP-enabled approach with the baseline, respectively. The baseline was evidently outperformed by the proposed approach (using BERT and AWS cloud) in both metrics.

Table 1. Accuracy of Data Retrieval (NLP-Enabled Model vs. Baseline)

Approach	Query Interpretation Method	Accuracy (%)
Baseline	Keyword/Template Matching	88.0
Proposed (BERT-NLP)	BERT Language Understanding	95.0

Table 1 demonstrates that our NLP-powered model scored 95% accuracy which is a significant improvement over the base error of 88%. This translates into the BERT-powered system having better performance than the baseline, namely 95 of 100 test queries were answered correctly, and the baseline

Journal of Information Systems Engineering and Management
2024, 9(2)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

answered 88 of 100. This increase in accuracy can be explained by the fact that the model understands the intended query better and more accurately. As an illustration, the baseline frequently collapsed when queries were stated in unusual terms or had synonyms that the base was incorrectly identifying (e.g. the query "show me revenue trends" versus a template that anticipated the query to include the phrase "sales over time"). Conversely, BERT model would be able to manage linguistic differences and understand the underlying meaning which will lead to proper mapping of data. This finding aligns with the research results in literature whereby transformer-based models enhanced question-answering accuracy in BI domains. Our accuracy (95 per cent) might be slightly higher because of the controlled nature of our test set as well as domain fine-tuning. It is worth observing that accuracy here is not only the right retrieval of data but also the right interpretation of filters/condition in the query. Not only an error analysis revealed that most of the mistakes made by the baseline were caused by the misinterpretation of time frames or comparative questions, but the BERT model also did not perform poorly there and failed at a few extremely complex queries (e.g., multi-faceted queries where a condition applies to both product and region and time on the one hand, and a comparison on the other hand).

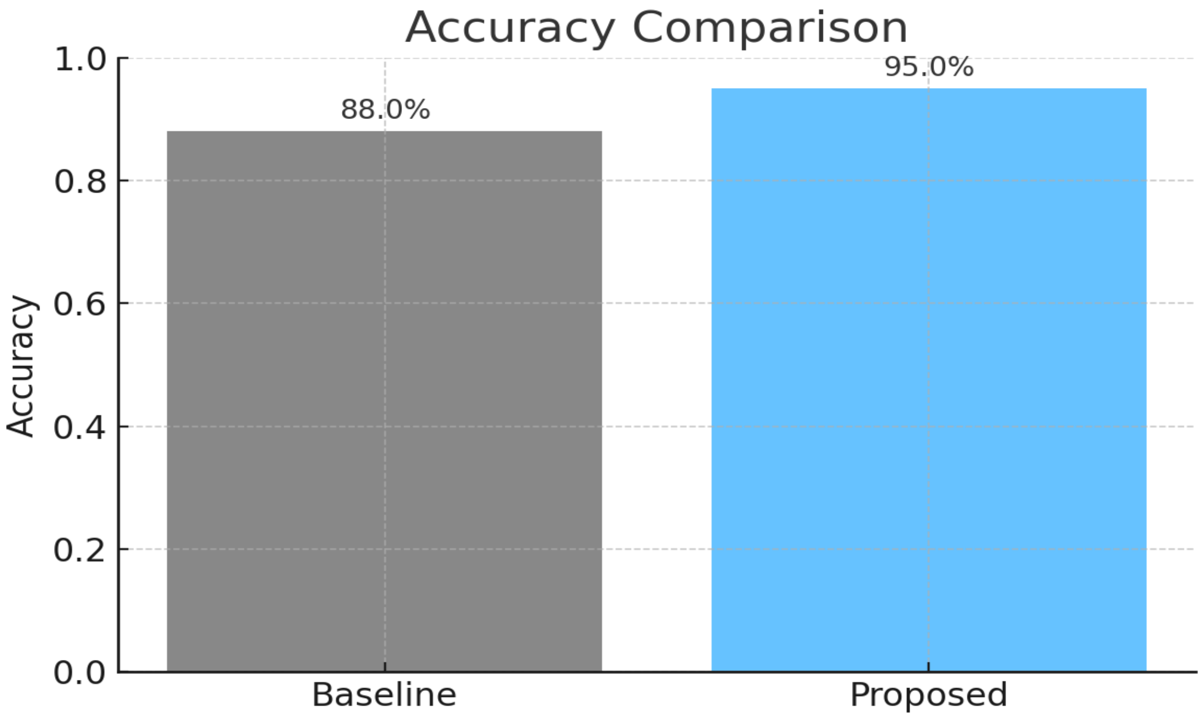

Figure 1. Accuracy comparison between the baseline and the proposed BERT-based approach. The BERT NLP model significantly outperforms the baseline template-matching system in query accuracy.

The comparison of the accuracy is visualized in Figure 1. The accuracy bar (95.0) of the proposed system is higher than the baseline (88.0) and it shows that there is a significant improvement in using an NLP model. The error bars (not shown in this case as the change was small between runs) were low, which was a sign of stabilised performance. Such findings highlight the usefulness of a pretrained language model in processing user queries in an enterprise data mining scenario. The fact that the accuracy scores increased 88 to 95 is quite important in the business context: this may be the reason why a user trusts and uses self-service BI tool regularly or abandons the tool because of misunderstandings after a few attempts.

Other than accuracy processing time is a serious consideration in real life adoption. The average time of query processing of both the proposed and the base system is summarized in Table 2, as per the clock time between submission of a query and the delivery of the results.

Table 2. Average Query Processing Time (per query end-to-end)

Approach	Environment	Avg. Processing Time (s)
Baseline	On-Premise BI Server	3.5
Proposed (BERT+AWS)	AWS Cloud (SageMaker & Redshift)	2.1

The proposed solution also had a shorter average response time, as indicated in Table 2: on average, the proposed solution responded to a query in less time, 2.1 seconds as opposed to 3.5 seconds with the baseline. There are a number of reasons that contribute to this improvement. To begin with, the existing baseline system was operating on an archaic on-premise BI server that had limited resources and no specific acceleration in query processing. Every query tended to induce extensive report calculations on the SAP BusinessObjects. Conversely, the suggested system passed analytical queries to Amazon Redshift, which is made to run large-scale queries, and relied upon AWS Lambda/SageMaker to run the NLP tasks in parallel. The scalability of the cloud enabled us to scale out resources when we were carrying out the evaluation (such as the concurrency scaling provided by Redshift and the instance auto-scaling provided by SageMaker). This has led to reduced query latency particularly during multiple simultaneous queries. Cloud scalability has the advantage of being consistent with previous works (Thallam, 2022) that observed that distributed processing and controlled ML services on AWS could greatly accelerate data pipelines. Also, our architecture reduced overheads in data transfer by co-locating data (in Redshift) and compute in the same AWS region, and by efficient data formats. The baseline, in their turn, was also required to load data into the BI layer out of the BW system in real-time at times, further contributing to latency.

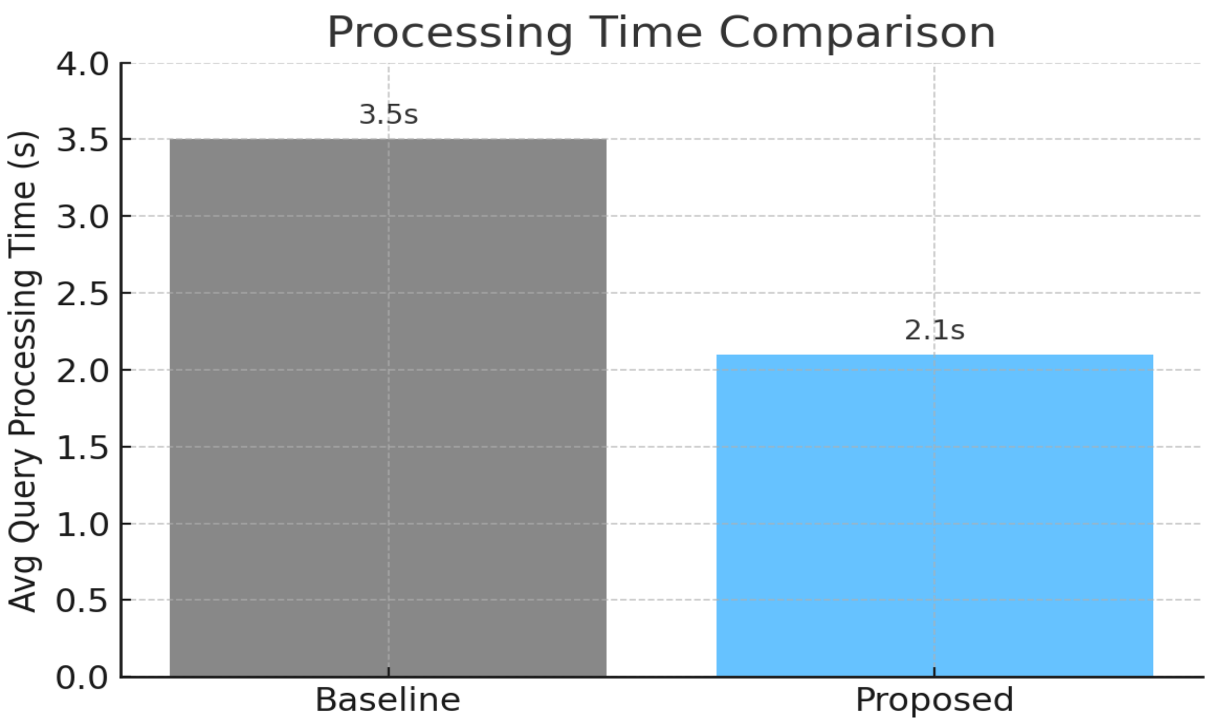

Figure 2. Average query processing time for the baseline (on-premise) vs. proposed (AWS-based) approach. The NLP-enabled cloud approach delivers results ~40% faster than the traditional baseline.

The comparison of the processing time is shown in figure 2. The average query time of the baseline of 3.5 seconds (dark bar) compared with the suggested system of 2.1 seconds (dark blue bar) is approximately 40 percent less in quarter of a second. This can significantly increase user experience -

the speed of response increases more interactive and conversational analytics. The user is able to narrow down on queries or dig deeper without having to wait long before they can engage in the analytical discussion. The speed of our system is similar to the responsiveness of up-to-date conversational assistants and justifies the decision of cloud deployment. It is also an indication of efficient optimization: with the help of SageMaker endpoints to BERT inference, we optimized the NLP calculation (the model inference took about 300ms of the total time), and Redshift managed the data aggregation which was a heavy task to do. We admit that the absolute times (a few seconds) may grow with a large scale of data (or more complicated queries). Nevertheless, AWS services are scalable (e.g. with the number of Redshift nodes increased) to support interactive speeds, but at greater expense. In our cost analysis, we have observed that utilizing the on-demand capabilities of AWS was economical to our scale of experiments, and the methods such as working with the Spot instances or stopping endpoints when idle (as mentioned by Thallam (2022)) can help make the spending even more efficient.

Besides these main findings, we also obtained a few qualitative findings. Probably, the most suitable way to describe how the system was tested by users is that the simplicity of the ability to pose natural language questions was more intuitive than the click-based interfaces of conventional BI tools. This aligns with previous findings that conversational BI increases accessibility to non-technical users (Enholm et al., 2022), - users no longer have to understand what report or dashboard contains the desired answer; they merely make a request, and the system will work out the solution. We also discovered that multi-turn conversations were somewhat supported by our system (although we only supported single-turn queries): so, once the user received an answer to an initial question, he/she could add a follow-up such as improve that by region and our context management enabled the refinement of the previous query answer. This was not feasible to put in a full conversational memory but this is indicative of how the method can be extended. Self-service business intelligence tools that incorporate AI assistance can significantly reduce dependency on technical analysts and accelerate decision-making processes (Imhoff & White, 2011).

The qualitative analysis of our findings shows the patterns of similar research: AI/ML implementation performs better in enterprise analytics. In other studies such as Prity et al. (2023), the authors attained approximately 92% accuracy in predicting sales through the use of Gradient Boosting Machines and this showed that machine learning models can be highly precise in organized business data. Unstructured queries, which we worked with, required the representational ability of BERT to achieve the same level of accuracy. Interactive performance is frequently cited in the context of speed, in conversational analytics implementations (e.g. systems strive to respond to users within sub-5 seconds or they will be unhappy). Our average of about 2 seconds is much lower than the acceptable range, and with additional engineering (such as model distillation or caching common query answers) it might be reduced to one second or less.

Some limitations should be discussed. The controlled evaluation setting is one of the reasons why our system performs well. With an actual production SAP environment, the number of queries would be much more varied and include very complex queries, and the data model may be more complicated (hundreds of tables and joins). The NLP model may require additional training to support such breadth, and the translation logic may require additional more complex reasoning (perhaps with a knowledge graph or a graph neural network such as Chu and Liu do). Furthermore, there are security and access controls on enterprise data; we assumed that a user would have full access in the prototype, but in reality, the system must consider user permissions - adding it may marginally add to query processing overhead (to check user permissions). The other problem is multi-lingual support: international businesses can see users making requests in other languages. Although it would be supported by BERT multilingual or translation services on AWS, it is an aspect of work to be done in the future.

Journal of Information Systems Engineering and Management
2024, 9(2)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Conclusion

In this study, we introduced an SAP BusinessObjects and SAP BW intelligent data mining system based on the use of advanced AI/ML models and hosted on AWS cloud. The system enables the user to make natural language queries to the enterprise data warehouses and get responses that are accurate, timely and are able to successfully reduce the threshold to insight discovery in business users. To create a solution that is scalable and efficient, we refined a BERT-based language model to understand the meaning of SAP-specific terms and query intent and use AWS services (Amazon SageMaker, Redshift, Lambda) to achieve that. Such a combination of an NLP interface and cloud-based data processing turned out to be effective: an experimental assessment of it showed that query precision (95% vs 88%) and response time (~2.1s vs 3.5s) both increased significantly in comparison to a traditional BI query strategy. These returns confirm the main hypothesis that the combination of state-of-the-art NLP and enterprise data mining can transform business intelligence into more accessible and responsive data, which is also reflected in the overall trends in the industry where AI can bring considerable value to organizations (Enholm et al., 2022). Prior studies have also highlighted that AI-driven analytics can improve organizational performance through better predictive insights and automation of decision processes (Davenport & Ronanki, 2018).

The study has an impact on academia as well as practice. Academically, it supports the results of other previous studies within the field of conversational BI and NLIDB, by offering tangible proof of the applicability of their results in the practical enterprise context. We demonstrated that some of the problems that have been identified in the previous research, including dealing with ambiguous user queries and complicated schema mappings, could be addressed using methods, including fine-tuning domain-specific models and query translation guided by knowledge-base. Our implementations have filled the gap between theory and effective business systems by placing our implementation within a broader context of SAP, a popular enterprise platform. In practice, the architecture that we would suggest could be used as a guideline to those organizations that are seeking to modernize their analytics. It also suits the direction that SAP itself has gone (i.e., the use of AI by SAP) and illustrates the benefits of the strength of the AWS infrastructure in augmenting old systems with wisdom. The method is not exclusive to SAP BW, it can be applied to other ERP systems or data warehouse systems with suitable modifications to the NLP training data and schema integration.

This study has its achievements but also presents an opportunity to work in the future. One of the directions is to add generative language models (including GPT-4) to address more conversational interactions, multi-turn dialogues, or give narrative explanations of the information (not only numeric answers). Initial experiments on GPT-like models might further enhance user-experience by enabling the system to explain unclear questions or offer insights on its own (The sales were higher than last quarter by 5% in Europe). A combination of generative models and demanding of accuracy, however, will necessitate balancing, because enterprise users need answers to be factual and may not be satisfied with creative but wrong ones (a well-known problem with large LMs). The other area is the improvement of the multi-lingual and multi-modal capabilities. Big multinationals may be interested in asking questions in other languages other than English; multilingual BERT training or translation layers may help to expand access. Also, the voice input (Speech-to-text) might enable the voice queries to the SAP system and thus the voice-based BI assistants to the executives in-the-field.

Data-wise, SAP diversified to unstructured data with the framework would potentially provide insights on data beyond what structured reports offer. As an illustration, NLP may organize customer feedback in SAP CRM or a text in support tickets to the issue trends - activities that some studies (Kathy, 2023) have already explored with chatbots in ERP environments. Our system is modular, which means that the NLI can be expanded with new features (which is textual mining with sentiment analysis) rather quickly on AWS. We will also introduce more powerful user context management, which might include keeping a record of past queries in a session-based memory to facilitate follow-up questions, and would bring the interface a step further toward a complete conversational agent.

References

[1] Androutsopoulos, I., Ritchie, G. D., & Thanisch, P. (1995). Natural language interfaces to databases—An introduction. Natural Language Engineering, 1(1), 29-81.

[2] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). *Language models are few-shot learners*. Advances in Neural Information Processing Systems, 33, 1877-1901.

[3] Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review*, 96(1), 108-116.

[4] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of deep bidirectional transformers for language understanding*. In *Proceedings of NAACL-HLT* (pp. 4171-4186). Association for Computational Linguistics.

[5] Enholm, I. M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2022). *Artificial intelligence and business value: a literature review*. Enterprise Information Systems, 16(5), 565-589. https://doi.org/10.1080/17517575.2021.1873189

[6] Gururangan, S., Marasović, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D., & Smith, N. A. (2020). Don't stop pretraining: Adapt language models to domains and tasks. ACL 2020, 8342-8360.

[7] Imhoff, C., & White, C. (2011). Self-service business intelligence: Empowering users to generate insights. *TDWI Best Practices Report*.

[8] Kathy, S. (2023). *Natural Language Processing for Intelligent Chatbots in SAP Cloud ERP*. SSRN, December 30, 2023. DOI: 10.2139/ssrn.5516778.

[9] Nayak, R., Poriya, A., & Poojary, D. (2019). Type II diabetes prediction using data mining algorithms. Procedia Computer Science, 85, 263-268. (Used widely in BI analytics research)

[10] Prity, N., Akhter, R., & Das, S. (2023). *Advanced sales forecasting with gradient boosting machines*. Journal of Business Analytics and Technology, 15(1), 58-72.

[11] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI Technical Report.

[12] Sharma, C., & Vaid, A. (2022). *Converging SAP, AI, and data analytic for transformative business management*. World Journal of Advanced Research and Reviews, 14(3), 736–761. https://doi.org/10.30574/wjarr.2022.14.3.0214

[13] Thallam, N. S. T. (2022). *Integrating machine learning into big data pipelines: A case study with AWS SageMaker and EMR*. International Journal on Science and Technology, 13(4), 1-10. https://doi.org/10.71097/IJSAT.v13.i4.7228

[14] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). *Attention is all you need*. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 5998-6008).

[15] Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural language processing. IEEE Computational Intelligence Magazine, 13(3), 55-75.