Review Article

Towards Autonomous Analytics: The Evolution of Self-Service BI Platforms with Machine Learning Integration

Shakir Syed [1,*]

[1] Self-Service Data Science Program Leader, Cummins Inc, USA

*Correspondence: Shakir Syed (shakir.syed.microsoft@gmail.com)

Abstract: Self-service business intelligence (BI) platforms have become essential applications for exploring, analyzing, and visualizing business data in various domains. Here, we envisage that the business intelligence platform will perform automatic and autonomous data analytics with minimal to no user interaction. We aim to offer a data-driven, intelligent, and scalable infrastructure that amplifies the advantages of BI systems and discovers hidden and complex insights from very large business datasets, which a business analyst can miss during manual exploratory data analysis. Towards our future vision of autonomous analytics, we propose a collective machine learning model repository with an integration layer for user-defined analytical goals within the BI platform. The proposed architecture can effectively reduce the cognitive load on users for repetitive tasks, democratizing data science expertise across data workers and facilitating a less experienced user community to develop and use advanced machine learning and statistical algorithms.

Keywords: Self-Service Business Intelligence (BI), Data Analytics, Autonomous Analytics, Data Visualization, Business Data Exploration, Intelligent Infrastructure, Scalable BI Systems, Hidden Insights, Large Datasets, Exploratory Data Analysis

1. Introduction

Accuracy and actionability are pivotal in generating value from big data volumes. From reporting to greater autonomy with analytics, using ML embedded directly in the self-service BI suite enables the target audience to make data-driven decisions more frequently and confidently. While mature for enterprises embracing the cloud, challenges remain including ease of use by those with varying skill levels, security, cost control, and custom analytics development timescales. The different success criteria and trade-offs among these organizations therefore also influence what vendors have produced in their products as well as where independents fit into this space.

Taking these lessons into account and using recent project experiences, a master blueprint of the preferred architecture for this next evolutionary step in self-service BI is drawn. It has greater portability of analytics than present with business content packages; scales transparently to simplify security, performance, and governance for vast volumes of concurrent predictions using different data; and will be routinely and fully used by broader audiences since it holds great promise in making machine learning techniques as smart and easy to use, understand, and trust as, graphical user interfaces, drag-and-drop, visual programming, and the concealment of software, branching, modeling diagrams, data profiling, and executive dashboards combined [1].

1.1. Background and Significance

Unprecedented volumes of data are continuously generated by individuals, businesses, devices, and applications within a broad spectrum of industries, making it

essential for organizations to be data-driven. Many modern enterprises rely on data and business intelligence (BI) platforms to provide insights that drive key business decisions and performance. With the evolving complexity and diversity associated with big data integration, application, and analysis, advances in AI, machine learning, and data science are increasingly essential. Advanced analytics provide unique advantages and intelligence where traditional BI and descriptive analytics fall short. While years ago, expensive data scientists and IT departments were essential for running predictive models against vast volumes of data, today this is no longer the case. The growing trend of implementing data democratization programs within companies has resulted in businesses opening up their vast volumes of data to a much broader range of business users, enabling these everyday workers to obtain valuable insights without a data scientist's assistance. Initially a relatively simple BI tool for querying databases, OLAP systems, and building reports and dashboards, today's BI market is increasingly composed of more complex analytics platforms that offer not only descriptive but also diagnostic, predictive, and prescriptive capabilities. As advanced analytics have become increasingly integrated into modern analytical tool sets, this has led to the rise of "self-service BI"—tools enabling business professionals across an organization to derive insights from an increasing range of structured and unstructured data. Business analysts are more frequently conducting complex analyses than before, performing tasks like clustering customers into groups with common buying behaviors, predicting inventory turnover ranges, and classifying website visitors. As a reverse innovation, self-service BI platforms encourage more frequent ad hoc use of BI reporting from departments, including marketing, finance, human resources, and non-data-driven executives. Thanks to self-service BI, insights can be obtained and acted upon more quickly than through traditional independent reporting that passes through layers of IT for implementation [2].

Figure 1. Self-Service Analytics Evolves to 'Smart' Analytics

1.2. Research Objective

Thus, the overall objective of this paper is to identify the current state and direction of research in the technology behind these systems to enable a move away from "tool-driven" programming towards a more user-friendly analytics capability. The specific technology is centered on two areas: data management and machine learning algorithms. Automated data management is used to find, select, and transform data to enable the execution of ML algorithms most efficiently and effectively. Machine learning algorithms can then be used to build models that can be used to answer business questions in a predictive manner [4].

Thus, the specific research question is whether it is possible, given any arbitrary set of analytic requirements and constraints, to automatically find the right data and then use the appropriate ML algorithm to generate the analytic result using data that is directly managed and owned by the business user. Once the analytical model has been created,

the question then expands to include searching for additional data and other sources that could be included to further refine the analytic results. The motivation for the research is that although ML strategies have now become widely available to business end-users, they still face significant challenges in terms of knowledge and skills in these areas. The outcome would be to enable a BI system that is both self-serve for the end-user and automatically adaptive depending on the requirements of the BI analysis being carried out [6].

1.3. Scope and Organization of the Paper

The scope of this research is to explore the evolution of self-service business intelligence (BI) platforms with machine learning functionalities, a class of platforms unique in offering both intuitive data discovery and prediction capabilities to the end user. There are theoretical and computational challenges in the development of these platforms: from technological aspects like the definition of visual predictive analytics primitives to those related to good practice-oriented tool use, such as the design of data provisioning and predictive model performance quality assessment methods. These are the perspectives from which this work is organized, the latter focuses on opening up new frontiers for research in this domain [28].

In more detail, the remainder of this work is organized as follows. To calibrate the expectations of the reader, we clarify the scope of the paper and outline the structure of this document. In the next section, we provide an overview of the contextual matters surrounding the main topic of this paper. First of all, we discuss the state of the art in the cardinal branch of research that gave a major contribution to the evolution of self-service BI: data visualization. In section 3, we show how the democratization of predictive modeling is transitioning from a trend to a consolidated paradigm in the industry. These two discussions set the background for the fourth section, as the theoretical evolution of the relationship between data scientists and business analysts envisions that some aspects of the former profession are inherited by the latter. With this in mind, in this part of the study, we analyze the main theoretical and technological aspects of the new class of platforms we are leading researchers in [3].

Equation 1: Data Preprocessing

Normalization:

$$x' = \frac{x - \mu}{\sigma}$$

Where: x' = normalized value
x = original value
μ = mean of the dataset
σ = standard deviation of the dataset

2. Foundations of Self-Service BI

"The phrase "self-service Business Intelligence" (or simply "self-service BI") typically refers to software that enables business users to perform queries and generate reports. Once the sole province of IT professionals with database skills and a solid understanding of data warehousing, BI applications are increasingly deployed by organizations to embrace self-service BI as an enterprise goal, empowering their workers to ask and answer business questions themselves rather than rely on BI professionals to ensure that key insights are available when needed.

However, despite the outward appearance of intuitiveness, the process of asking (and answering) questions about the company remains inherently cumbersome and time-

consuming for the typical businessperson who cannot afford the time required to learn how to use the traditional BI systems properly. These systems have long used design principles that "pre-select" the types of questions likely to be asked of the data while streamlining necessary data manipulations. Although this rigidity was once necessary, such tools appear woefully out of date in a world where there is a consensus that the best insights typically come from exploratory analysis, a class of tasks that are explicitly not pre-defined [5]."

2.1. Definition and Components of Self-Service BI

The term self-service BI is commonly used to designate the process where end users are empowered to develop their reports and analyses without the need for IT intervention. However, self-service BI means much more than this. It is, in reality, the process itself that makes access to information, to enable decision-making, more agile and simpler for users. The success of self-service BI depends on technology so that users can develop their analyses in a simple, intuitive, and reliable way. Business intelligence workers need fast, easy access to reliable data, and those who have analytical abilities should be able to perform dynamic, intuitive, and predictive analyses without necessarily having advanced BI skills. Self-service is how users can create their layout for data visualization, and it also involves the discovery of data. This process involves exploring all available data through reports, dashboards, and ad hoc analysis, affording the discovery with the ability to create variable relationships that were not previously perceived, allowing exploration of data behavior and the analysis of potential causes and effects without the need to define questions in advance, thus becoming a discovery inherently driven by the primary data rather than by predetermined hypotheses [8].

2.2. Benefits and Challenges

Self-service BI platforms are used to generate analytics insights and business outcomes based on data collected from numerous data sources such as spreadsheets, risk marts, data warehouses, operational data stores, and big data systems. Self-service BI comes with a unique set of business and technical benefits and challenges. Business benefits of self-service BI include greater business user independence and autonomy. This means reduced time to intelligence, a boost to business user capabilities, and the organization's reactivity to change, as well as overall operational flexibility. This increased time to intelligence includes data self-discovery through understanding real data and better usage via improved, faster data organization and flexible self-service access. The technical benefits of self-service BI implementation include a reduction in demand on the IT service desk and time savings in organizations, particularly for IT service provisioning. On the other hand, self-service BI end users, being isolated, inadvertently replicate work and data. Since these users are business task owners with dominant discipline expertise and few dedicated IT resources, it can be unclear whether they are using the best data analysis tools or where opportunities lie to exploit these tools for the best results [7].
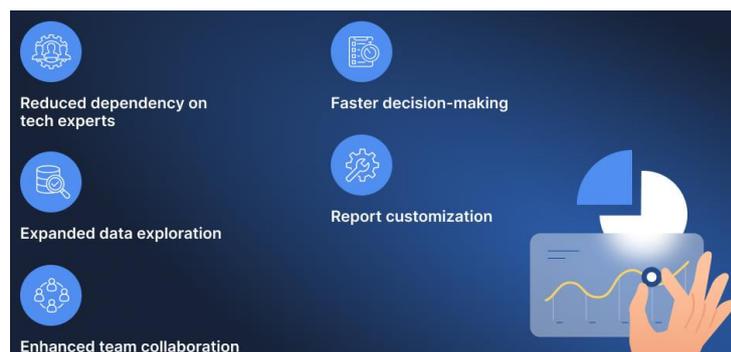
Figure 2. Benefits of Self-Service BI

3. Integration of Machine Learning in Self-Service BI

Recent developments in data management platforms have made it possible to integrate sophisticated machine learning algorithms into decision support systems. While these platforms have demonstrated their capabilities in terms of algorithms and technologies, they still lack a level of user interaction and automation that is required for real business environments. Self-service BI platforms, on the other hand, provide a high level of user interaction and automation, but they are limited in terms of advanced machine learning capabilities. In this work, we demonstrate the integration of machine learning into a self-service BI platform and propose a new class of platform called Autonomous Analytics that combines the capabilities of both platforms for improved decision support in businesses. Our architecture uses the metadata and interaction logs of self-service BI platforms to provide feedback to machine learning algorithms and to optimize models or strategies for ongoing queries for every level of user expertise.

In traditional BI environments, there is usually an IT and a BI department, which are responsible for designing the BI applications and the user side. The development of a new query can be a lengthy process with a complex feedback cycle between the users and the IT department. Our approach reduces the feedback cycles because only the correct modeling of the data used in the interaction and the algorithm can create the value expected by the user. The users in recurring cases only have to send the IT department the data and a model update call; this is necessary to keep the actual value of the refinements. Such a system cannot create value without user feedback as well; in this case, the users have to refine their context usually by doing more research to get more education, the latter might be supported by model-validated results. Furthermore, Autonomous Analytics tries to validate the results in a self-service manner by augmenting the tools with automated methods for generating refinement proposals and visualizing the effect of the refinements [9].

3.1. Overview of Machine Learning

Machine learning and other types of predictive modeling apply statistical techniques to discover or infer relationships in data that can be used to predict future behavior or other outcomes of interest. We define ten common categories of supervised machine learning techniques, as well as their most widely used methods. A model is trained using a labeled dataset, which contains the outcome (or class label) of interest, as well as all other key variables or features, usually mapped as columns of a table, spreadsheet, or file. The outcome can be a numerical value, a binary state, one of a set of discrete classes, or an ordering of greater or lesser value. Which type of outcome dictates the machine learning category into which the problem fits? Each row shows a specific instance of known outcomes and associated values of features. The model eventually discovers some type of mapping relationship between the input attributes to the outcome values.

Thereafter, the model can apply this mapping to unseen cases, the set of which is referred to as a test dataset. To add a predictive indicator for later regression use, we can append a column reflecting the predicted value to the original dataset, our source of training data. If we use a large testing dataset, we can validate the model and assess its goodness of fit by comparing the predicted values to the actuals, thereby evaluating the model's performance in terms of precision, accuracy, and other standard statistics. Models can be tested for their ability to predict outcomes that have not been previously seen using hold-out or cross-validation methods [11].

3.2. Applications in Business Intelligence

There has been considerable research in the field of machine learning that focuses on the problem of business analytics. It is easy to intuit the connection between the two

spheres; machine learning expertise is, in itself, a form of knowledge and can be used to intelligently inform decisions. In this section, various applications at the intersection of these two fields are discussed, illustrating both the challenges and the considerable potential that the existing body of machine learning knowledge can bring to bear on an important class of real-world problems. Additionally, thoughts about promising areas worthy of deeper study and consideration are shared.

Sales Forecasting and Revenue Management Revenue management and sales forecasting are classic problems in the business intelligence toolbox. The goal of revenue management is to maximize the firm's revenue by engineering the demand for products at different times and price points while accommodating the constraints of supply and market interests. Traditional applications have been focused on industries such as airlines, with the problem of putting pricing rules into place that automatically adjust the price of tickets over time as demand for those tickets rises. This approach increases the fares as the date of the available ticket becomes closer, encouraging passengers to commit to booking, and also increases the last-minute fares purchased by business travelers [10].
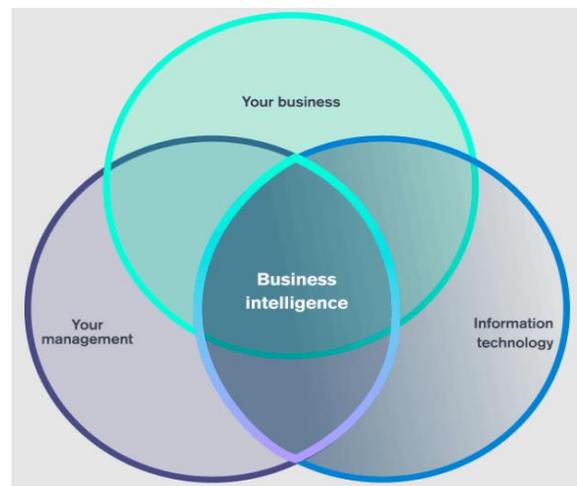
Figure 3. The Business Intelligence Platform Revolution

3.3. Benefits and Challenges of Integration

With the evolution of AI and ML capabilities, it becomes critical to enable users to quickly and easily tap into the power of machine learning to see, understand, and act on all the data as it fluctuates. Here, we dissect the BI inside self-service business intelligence into its components and describe how machine learning and analytics techniques and algorithms, with a particular focus on time series analysis and forecasting, can be integrated to work not only in an interoperable manner but as components of the same system. We describe our design choices and outline an integrated platform we build that embodies those principles. As data volumes and complexity grow, and as data analysis needs become ever tighter, we believe that the conventional segregation between capacities used to establish system-level situational awareness by big data experts and analysts should evolve into a single joint effort. This drives the natural desire to create a single policy guiding system that encapsulates analytical capabilities at all scales in a single architecture serving as policy advisors. Such a system needs to not only allow big data experts and analysts to move from non-trivial to more complex analytical and forecasting or future state envisioning tasks but also to support the needs of emerging generations of users who might not be experts in the subject domain or have deep quantitative knowledge. As automated future state modeling and forecasting continue to hold increased importance, we need to ensure analytics and ML capacity become easily accessible to a much broader class of users than was possible previously. Significant

factors include seamless integration of substantial-sized data with embedded intelligent forecasting or classification, and support for widely varying users' skill sets and overlapping roles [13].

4. Autonomous Analytics: Concepts and Technologies

People perform tasks, make decisions, use resources, and execute everyday activities. These everyday activities are full of data that are recorded regularly by a plethora of different devices and tools. Data recording is no longer a scarce resource but an integral part of carrying out an activity. Data carries information that is valuable for understanding how to enhance these data-informed activities, how to optimize them, and how to make them more efficient and effective. Through data analysis and the use of business intelligence techniques, professionals are enabled to make sense of this valuable information and are guided to optimal solutions. The eventual result of the professional activities, however, is still carried out manually. We envision a much more automated approach to problem-solving through data analysis, leading us to the concept of autonomous analytics [27].

Autonomous analytics is not only about enabling people to undertake data-informed tasks better but also about improving the tasks and processes for which data is informative. The idea is that job activities can become much more adaptive through machine learning, leading activities to be an optimizing and preserving driver for results and, going beyond that, making the job activities much more automated. Reaching a higher level of automation in the decision-making process will greatly benefit from business intelligence platforms, which nowadays mainly focus on producing reports and visualizations that are interactive and work exactly as the user dictates. We propose an evolutionary path that transforms these well-structured platforms into self-service business intelligence and analytics platforms that are naturally collaborative companions, allowing for the identification of critical events and situations in data and arriving at joint optima for these critical situations [14].

4.1. Definition and Components of Autonomous Analytics

First, since the term "autonomous" may be a little broad in scope, let's make sure we understand it in the same way. Autonomous analytics is the achievement of a paid and intended objective outcome from an analytics and BI platform with minimal manual human effort. The components of autonomous analytics are the tools and techniques that help minimize the human effort to achieve a paid and intended outcome. When humans do need to "get involved" in some aspect of operations, there are mechanisms that facilitate this integration and manage the interaction between humans and machines. In other words, the goal of autonomous analytics is to minimize operational intervention in the analysis process over time so that practitioners can concentrate on delivering more valuable content with our analytics skill set [12].

To become autonomous, a company's analytics team utilizes one or more machine learning models in a BI platform to automate more of the analytics process over time, with minimal expert intervention. This could include setting intelligent alerting to notify of conditions or situations that "merit" human intervention. The BI platform typically uses a combination of role-based, multi-tenanted systems designed to enable mostly self-service exploration and to visualize data in the form in which it can make the most meaningful impact. This platform currently requires some group effort to deliver initial self-service and analytic capabilities, maintaining the roles and enabling security models over time while ensuring both data quality and platform long-term performance and security [26].

Equation 2 : Regression Analysis

Linear Regression Model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \in$$

Where: y = dependent variable
β_0 = intercept
β_i = coefficients for independent variables x_i
ϵ = error term

4.2. Key Technologies Enabling Autonomous Analytics

The availability of new technologies is increasing the degree to which Business Intelligence (BI) platforms can be considered autonomous. Technologies increasing the autonomy of BI platforms encompass software applications, interfacing technologies, as well as linkages to software environments. Machine learning is a key technology to automate or algorithmically improve analytics. At its core, it enables the extraction of deep patterns and trends in vast data sets that no data miner, no analyst, and certainly no business decision-maker could ever ferret out. Where the rampant creation of so-called Big Data, consisting of complex structured, semi-structured, and unstructured data types, features the business-events era, machine learning, and self-service BI tools are emerging as key to integrating structured and unstructured data types [29].

Current BI tools are capable of interfacing with such deep-insight technologies; moreover, BI capabilities, like dashboarding, enterprise reporting, and ad hoc analysis, are designed to be embedded in larger applications or environments, such as customer relationship management applications or finance and risk management systems. Also, non-BI tools supporting BI workflows no longer work as stand-alone software. They, too, are mostly web-based, encapsulated in mobile BI applications. In this section, we move one level up to assess the major technologies enabling autonomous and educative BI capabilities in the self-service era. Of course, this focus, in turn, directs our attention to the role of machine learning and deep insights in self-service BI software [15].

5. Case Studies and Use Cases

We describe in this section several cases of machine learning embedded in commercial self-service BI systems and then review industrial use cases to show how machine learning in self-service BI systems could meet real industrial needs. Keep in mind that embracing the biggest audiences: citizen data scientists and business users with ML using self-service BI systems does not mean these embedded ML methods are optimized. Sophisticated feature engineering and regularization or comparing various model capacities, etc. say the trade-off between bias and variance could be done behind the scenes of the model building. However, automatic embedded models could fail to match the performance of model building traditionally done by data scientists for individual purposes. Anomalous insights with built-in anomaly detection. Economies are moving to data-driven decision-making for citizens using AI. Non-AI specialists can generate 27 KPIs of a business to receive an alert digest similar to an email, which means citizens may also benefit from it. Anomalies in data are identified, explanations are tailored by citizen data scientists, and impact analyses are driven by using cloud services [16].

5.1. Real-World Examples of Self-Service BI Platforms with ML Integration

In this section, we will illustrate our definition of what represents a self-service BI platform with machine learning integration with commercially successful companies. Our goal is to showcase the rapidly growing ecosystem of companies that satisfy the criteria we defined. We would like to point out that these are only examples of existing platforms. The fast growth of this area can be assumed, and new platforms may have been created since we conducted our research. Instead, our focus is to future-proof the concept of the central issue of this paper [25].

5.1. Real-World Examples of Self-Service BI Platforms with ML Integration. Under the first category (guided analytics), we will present the BI platforms that fully incorporate or expand their products with advanced analytical capabilities like forecasting, which is essential for business practitioners. Our examples from the first category include 1) Microsoft's Power BI, a leading on-premises and cloud business analytics tool. Companies from multiple domains utilize it extensively. Microsoft achieved these sorts of analytics by hooking its Azure-based capabilities along with other parts of the Microsoft stack. 2) Tableau, which emphasizes visualization and discovery in an ongoing modeling and visualization loop. 3) Qlik, which supports user interaction in exploratory ways from any device and has a so-called associative data engine. 4) Their value is improved with the use of explainable AI technologies, and we show the XAI capabilities of Power BI, Tableau, and Qlik using an "explainable" augmented analytics cloud-based software. 5) TIBCO, which has gone a long way to try to integrate data preparation and modeling. Furthermore, an open-source BI and data analytics platform that has led to successful predictors is called Apache Zeppelin. RunLoop-Rostov presents the integration of MLflow's model packaging with Azure Databricks and Power BI. These BI tools allow analysts to store their Databricks training logs and, once the model is ready, register the model and consume it in Power BI using the provided API endpoint [17].
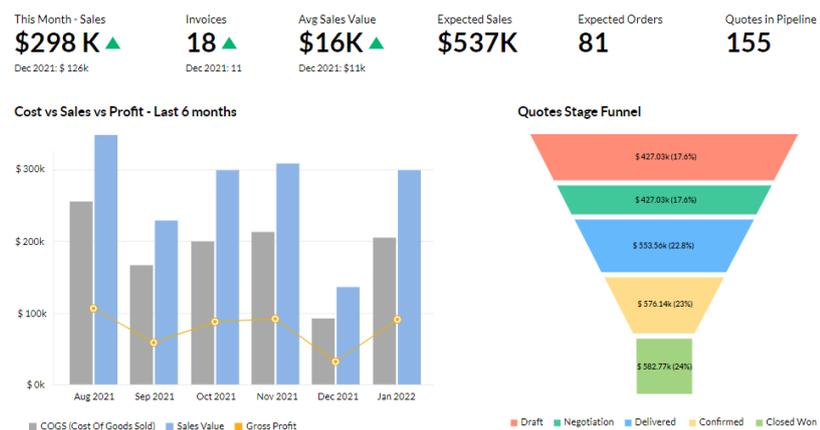
Figure 4. An example of an interactive dashboard for a sales department. Source

6. Future Directions and Implications

In this paper, we have examined two distinct lines of research—the evolution of self-service Business Intelligence (BI) platforms, often known as self-service BI, and the line of research involving machine learning integration with BI platforms. The evolution of self-service BI platforms has come a long way over the past decade. Going beyond the simple definition of serving business users in spearheading data discovery and exploration, a new era of self-service BI is on the horizon, one where the ultimate end goal is the enablement of all business users with access to more meaningful and thorough insights. Machine learning integration with BI platforms is the initial step in enabling a broader range of end users to unravel insights that were exclusively the province of data scientists. Machines are typically very good with data and hence, this interplay between self-service BI and data science should enable a seamless transition towards greater autonomous analytics. However, the integration of advanced machine learning capabilities with traditional self-service BI platforms creates a complex and unfamiliar mix of capabilities for business users [30].

Although a few pioneering platforms have ventured down such an integrated path, most organizations are still trying to figure out how they can bridge the advanced analytics divide. In this research, we have identified several limitations and challenges in the development of robust, user-friendly interfaces (a) catering to two different personas:

data scientists and business analysts; (b) managing machine learning models through their entire life cycle on these platforms; and (c) creating easy access to advanced machine learning without inherently involving a data scientist. Several future directions are needed to address these and other related challenges and advance this strand of research [19].

6.1. Emerging Trends and Innovations

The emergence of SSBI platforms over the last decade has already started to change the landscape of data and analytics within the business by bringing analytics into the hands of business users who capture, maintain, and understand data. The concept has evolved further as SSBI and ADS are starting to come together to provide necessary automation around prediction. Along with SSBI, there is a shift towards what is being called Smart Business Intelligence, where insights from both structured and unstructured data are molded into decisions and actions. New emerging trends include the adoption of in-memory platforms for data preparation and SSBI, efficient SQL-on-Hadoop engines, the emergence of discovery and exploration platforms built on access to datasets, and a mixed-source economy that brings technology cost advantages and builds collaborative ecosystems among value players and third-party vendors while promoting the deployment of high-performance big data stacks. The convergence of data lineage platforms that incorporate both structured and unstructured data lineage with business rules validation for SSBI supports audit and regulations. SSBI increases the importance of data governance and data management to improve and share data tales effectively. Integration of data visualization and predictive analytics for self-consumption leads to new convergences and applications including diagnostics, forecasting, and decision-making. Capabilities model reports describe these capabilities in detail, which are useful to assess SSBI, and interlocks with big data technologies. Incumbents are making significant advancements in data technologies as they start to interlock and integrate SSBI into their stack. Nuances in big data and SSBI competitors include which keep improving and announcing new features. The success or failure of vendors will be due to various factors including integrating BI and analytics, pushing their traditional platforms, scaling, dealing with broad acceptance, and managing public and private clouds. M&A activities to create the platform will also shape the scene, contenders, and hierarchy [18].

6.2. Implications for Business and Industry

We have argued that one way to overcome key challenges for the successful application of machine learning and auto-analytics in the industry is for the platforms to evolve as clouds where algorithms become data and data become analytics. We highlight how autonomous analytics systems on the cloud will become powerful intermediaries by enabling very large numbers of micro-contributors and consumers to manipulate and use algorithms and data through simple self-service mechanisms. This promises a new era of citizen analytics. Implications for business and industry: Cloud and big data have shifted the value emphasis of digital data services from information about individual transactions to the automatic capture, analysis, aggregation, and use of the data in those transactions. Autonomous analytics represents a quantum step further by making the resulting knowledge-based asset of that data generalized and reusable across different contexts and by aggregating the expertise that makes the analytics queries successful. To remain relevant in knowledge-based commerce, businesses must embrace and evolve their analytics platforms so that real-time data and learning are shared outcomes of the digital services offered [21].
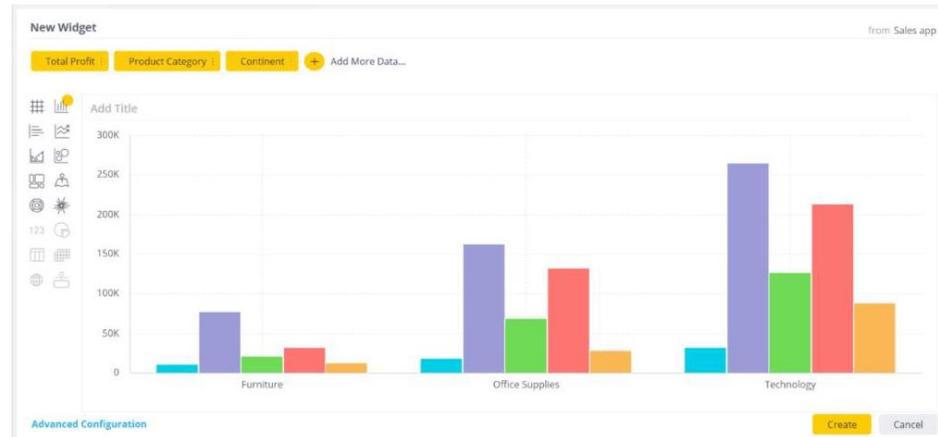
Figure 5. A Comparison of Self-Service Analytics Tools

7. Conclusion

In this paper, we proposed an architecture that takes the vision of self-service BI tools one step further by making them more intelligent using machine learning. Our architecture introduces several machine learning techniques, such as feature importance to rank datasets and columns, outlier detection on columns to enhance data profiling, column clustering, and component analysis to facilitate data understanding and column type detection, allowing self-service BI tools to automatically generate visualizations for the user without requiring any input, user-defined pattern recognition to learn the user's typical interaction with the system, adaptive behavior, and reinforcement learning to give the tool experience and let it learn from experience across users and over time. The results are promising and show that embedding recent machine learning techniques in self-service BI platforms can assist the user in quickly creating useful visualizations and support a wider audience of less experienced users. While we have implemented the proposed architecture on a self-service BI tool, many other self-service BI tools could benefit from the machine learning pipeline we introduced. In the future, we intend to expand the scope of applications to include a more evolved understanding of the user, the ability to improve tasks, the ability to optimize visualization for the target device and user, more advanced chart type detection from user intent, and to provide different grades between user input and system automatic output to better automate less trivial results. The main objective of this work is to move self-service BI from self-service business intelligence to autonomous self-service business intelligence [20].

Equation 3: Clustering

K-Means Clustering:

$$J = \sum_{i=1}^{k} \sum_{j=1}^{n} \left\| x_j^{(i)} - \mu_i \right\|^2$$

Where:
J = objective function (total within-cluster variance)
k = number of clusters
$x_j^{(i)}$ = data point in cluster i
μ_i = centroid of cluster i

7.1. Summary of Key Findings

Machine learning models are increasingly integrated into business intelligence platforms that also support self-service analytics. ML can unleash the potential of BI for

effective analytics toward autonomous data exploration and enable new data discovery processes for non-experts. Such advancements transform users from mere information consumers to information producers. To explore how state-of-the-art business intelligence platforms are currently augmenting self-service analytics with machine learning and the challenges these advancements present in current deployment scenarios, we perform an empirical study of the three leading commercial platforms [31].

This paper reports on potential biases and limitations of the examined platforms and proposes a data trust cycle method for engaging with users in the iterative process toward autonomous business intelligence. The potential for bias is identified in data curation, model training, and deployment phases, while the limitations of the examined platforms are rooted in platform integrations and predominantly technical user roles. The paper discusses the implications of these findings for future work and the software industry. We conclude with implications for future research and practice [22].

7.2. Limitations and Future Research Directions

The paper's contribution is a step towards self-service BI platforms that focus on making the execution of predictive analysis more user-friendly for users with limited machine learning knowledge. However, it also has certain limitations. Our system, Anomaly Discovery, is currently a prototype, and there are opportunities to extend and refine our work. The prediction results of the system are currently presented as a list. While presenting the prediction results as a list is user-friendly and can be insightful, this may need to change as the number of predictors grows. Additionally, the use of many visuals may speed up the process and allow the user to explore the data more deeply [32].

One potential future direction for extending Anomaly Discovery is to make it easier for users to compare different results. The anomaly model is currently trained based on only historical data, but it can be intriguing to allow the user to compare models with different features and even tweak the model by using a subset of the entire data sets. This will become more important when the predictor repository grows, and users need to figure out which predictors seem to have the most impact on anomalies. While we have tried to provide a curated list of predictors, we would like to be able to adjust these lists to reflect the characteristics of the data, thus giving the predictive model a higher fidelity to the data. To achieve this, the indicator part needs to be more accurate to allow the user to go back and adjust successive predictions to zero in on the anomaly and associated context [24].

References

[1]   Aravind, R., Shah, C. V., & Surabhi, M. D. (2022). Machine Learning Applications in Predictive Maintenancefor Vehicles: Case Studies. International Journal of Engineering and Computer Science, 11(11), 25628–25640.https://doi.org/10.18535/ijecs/v11i11.4707

[2]   Kommisetty, P. D. N. K. (2022). Leading the Future: Big Data Solutions, Cloud Migration, and AI-Driven Decision-Making in Modern Enterprises. Educational Administration: Theory and Practice, 28(03), 352-364.

[3]   Mandala, V., & Mandala, M. S. (2022). ANATOMY OF BIG DATA LAKE HOUSES. NeuroQuantology, 20(9), 6413.

[4]   Perumal, A. P., Deshmukh, H., Chintale, P., Desaboyina, G., & Najana, M. Implementing zero trust architecture in financial services cloud environments in Microsoft azure security framework.

[5]   Avacharmal, R. (2022). ADVANCES IN UNSUPERVISED LEARNING TECHNIQUES FOR ANOMALY DETECTION AND FRAUD IDENTIFICATION IN FINANCIAL TRANSACTIONS. NeuroQuantology, 20(5), 5570.

[6]   Pamulaparthyvenkata, S. (2022). Unlocking the Adherence Imperative: A Unified Data Engineering Framework Leveraging Patient-Centric Ontologies for Personalized Healthcare Delivery and Enhanced Provider-Patient Loyalty. Distributed Learning and Broad Applications in Scientific Research, 8, 46-73.

[7]   Korada, L., & Somepalli, S. (2022). Leveraging 5G Technology and Drones for Proactive Maintenance in the Power Transmission Industry: Enhancing Safety, Continuity, and Cost Savings. In Journal of Engineering and Applied Sciences Technology (pp. 1–5). Scientific Research and Community Ltd. https://doi.org/10.47363/jeast/2022(4)260

[8]   Vaka, D. K. "Artificial intelligence enabled Demand Sensing: Enhancing Supply Chain Responsiveness.

[9]   Bansal, A. Advanced Approaches to Estimating and Utilizing Customer Lifetime Value in Business Strategy.

[10] Mandala, V., Premkumar, C. D., Nivitha, K., & Kumar, R. S. (2022). Machine Learning Techniques and Big Data Tools in Design and Manufacturing. In Big Data Analytics in Smart Manufacturing (pp. 149-169). Chapman and Hall/CRC.

[11] Perumal, A. P., & Chintale, P. Improving operational efficiency and productivity through the fusion of DevOps and SRE practices in multi-cloud operations.

[12] Avacharmal, R., & Pamulaparthyvenkata, S. (2022). Enhancing Algorithmic Efficacy: A Comprehensive Exploration of Machine Learning Model Lifecycle Management from Inception to Operationalization. Distributed Learning and Broad Applications in Scientific Research, 8, 29-45.

[13] Korada, L. (2022). Low Code/No Code Application Development - Opportunity and Challenges for Enterprises. In International Journal on Recent and Innovation Trends in Computing and Communication (Vol. 10, Issue 11, pp. 209–218). Auricle Technologies, Pvt., Ltd. https://doi.org/10.17762/ijritcc.v10i11.11038

[14] Pamulaparthyvenkata, S., & Avacharmal, R. (2021). Leveraging Machine Learning for Proactive Financial Risk Mitigation and Revenue Stream Optimization in the Transition Towards Value-Based Care Delivery Models. African Journal of Artificial Intelligence and Sustainable Development, 1(2), 86-126.

[15] Vaka, D. K. " Integrated Excellence: PM-EWM Integration Solution for S/4HANA 2020/2021.

[16] Bansal, A. (2022). Establishing a Framework for a Successful Center of Excellence in Advanced Analytics. ESP Journal of Engineering & Technology Advancements (ESP-JETA), 2(3), 76-84.

[17] Chintale, P., Korada, L., WA, L., Mahida, A., Ranjan, P., & Desaboyina, G. RISK MANAGEMENT STRATEGIES FOR CLOUD-NATIVE FINTECH APPLICATIONS DURING THE PANDEMIC.

[18] Avacharmal, R. (2021). Leveraging Supervised Machine Learning Algorithms for Enhanced Anomaly Detection in Anti-Money Laundering (AML) Transaction Monitoring Systems: A Comparative Analysis of Performance and Explainability. African Journal of Artificial Intelligence and Sustainable Development, 1(2), 68-85.

[19] MULUKUNTLA, S., & VENKATA, S. P. (2020). AI-Driven Personalized Medicine: Assessing the Impact of Federal Policies on Advancing Patient-Centric Care. EPH-International Journal of Medical and Health Science, 6(2), 20-26.

[20] Laxminarayana Korada, Vijay Kartik Sikha, & Satyaveda Somepalli. (2022). Importance of Cloud Governance Framework for Robust Digital Transformation and IT Management at Scale. Journal of Scientific and Engineering Research. https://doi.org/10.5281/ZENODO.13348757

[21] Vaka, D. K. (2020). Navigating Uncertainty: The Power of 'Just in Time SAP for Supply Chain Dynamics. Journal of Technological Innovations, 1(2).

[22] Bansal, A. (2022). REVOLUTIONIZING REVENUE: THE POWER OF AUTOMATED PROMO ENGINES. INTERNATIONAL JOURNAL OF ELECTRONICS AND COMMUNICATION ENGINEERING AND TECHNOLOGY (IJECET), 13(3), 30-37.

[23] Chintale, P. SCALABLE AND COST-EFFECTIVE SELF-ONBOARDING SOLUTIONS FOR HOME INTERNET USERS UTILIZING GOOGLE CLOUD'S SAAS FRAMEWORK.

[24] Mulukuntla, S., & VENKATA, S. P. (2020). Digital Transformation in Healthcare: Assessing the Impact on Patient Care and Safety. EPH-International Journal of Medical and Health Science, 6(3), 27-33.

[25] Laxminarayana Korada. (2022). Optimizing Multicloud Data Integration for AI-Powered Healthcare Research. Journal of Scientific and Engineering Research. https://doi.org/10.5281/ZENODO.13474840

[26] Dilip Kumar Vaka. (2019). Cloud-Driven Excellence: A Comprehensive Evaluation of SAP S/4HANA ERP. Journal of Scientific and Engineering Research. https://doi.org/10.5281/ZENODO.11219959

[27] Bansal, A. (2021). OPTIMIZING WITHDRAWAL RISK ASSESSMENT FOR GUARANTEED MINIMUM WITHDRAWAL BENEFITS IN INSURANCE USING ARTIFICIAL INTELLIGENCE TECHNIQUES. INTERNATIONAL JOURNAL OF INFORMATION TECHNOLOGY AND MANAGEMENT INFORMATION SYSTEMS (IJITMIS), 12(1), 97-107.

[28] Korada, L. (2021). Unlocking Urban Futures: The Role Of Big Data Analytics And AI In Urban Planning–A Systematic Literature Review And Bibliometric Insight. Migration Letters, 18(6), 775-795.

[29] Bansal, A. (2021). INTRODUCTION AND APPLICATION OF CHANGE POINT ANALYSIS IN ANALYTICS SPACE. INTERNATIONAL JOURNAL OF DATA SCIENCE RESEARCH AND DEVELOPMENT (IJDSRD), 1(2), 9-16.

[30] Mandala, V. (2022). Revolutionizing Asynchronous Shipments: Integrating AI Predictive Analytics in Automotive Supply Chains. Journal ID, 9339, 1263.

[31] Bansal, A. (2020). An effective system for Sentiment Analysis and classification of Twitter Data based on Artificial Intelligence (AI) Techniques. International Journal of Computer Science and Information Technology Research, 1(1), 32-47.

[32] Chintale, P., Korada, L., Ranjan, P., & Malviya, R. K. ADOPTING INFRASTRUCTURE AS CODE (IAC) FOR EFFICIENT FINANCIAL CLOUD MANAGEMENT.