DBP Insights

2024 Data, Analytics, and AI Predictions and Prescriptions

www.dbp-institute.com

Foreword

Thank you very much for your interest in our research: **DBP Insights: 2024 Data, Analytics, and AI Predictions and Prescriptions.**

The world's economy will dramatically transform in the coming years and data will play a big role in this change. A report from MIT says digitally mature firms are 26% more profitable than their peers. McKinsey Consulting indicates that data-driven organizations are 23 times more likely to acquire customers. Industry analyst firm Forrester found that organizations that use data to derive insights for decision-making are almost three times more likely to achieve double-digit growth. Overall, organizations are looking at Analytics and AI for improved business performance.

In this backdrop, this research work **DBP Insights: 2024 Data and Analytics Predictions and Prescriptions**, not only talks about the **ten predictions**, but it also gives a **prescriptive recommendation and the required capabilities** for organizations to prepare for these future changes in 2024 and beyond. It is prepared by 4 very senior industry experts whip bring over 100 years of consulting and research experience. Each of the ten DBP Insights, has three main sections.

1. Background i.e., the **context** for the prediction
2. The rationale for the prediction
3. Finally, the **prescriptive recommendation** on how to respond and capitalize on this prediction. It is not just enough to predict or know the future state, prescription is also important so that organization can plan and respond for its future state.

Last year too, we produced such a similar report with **ten predictions for 2023** and majority of the predictions were visible in 2023.

This research report is a teamwork from some of the best data and analytics experts in the industry. Firstly, I thank the entire Analyst team i.e. **Arun Marar, PhD**, **Sanjiv Chib, Tobias Zwingmann, and Prashanth Southekal, PhD, MBA,** who collaborated with me with their inputs. I sincerely thank DBP's advisors **Steve Rosvold** and **Gary Cokins** for giving their valuable feedback on this report. Also, special thanks **CFO.University, Arcalea, and Grihasoft Technologies** for supporting this research.

I hope, you will find the **DBP Insights** report useful for your organization. For any questions or clarifications, do not hesitate to contact me. Have a great 2024!

Regards and Thank you

Prashanth Southekal, PhD., MBA, MS, ICD.D

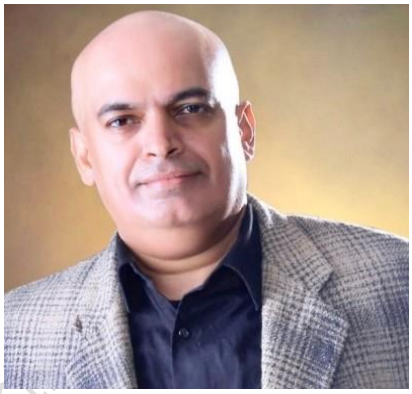
Founder and Managing Principal
Data for Business Performamce (DBP) Institute

M: +1 587 434 2581
E: psouthekal@dbp-institute.com
W: **www.dbp-institute.com**

Our core values: Simplicity - Passion - Integrity - Commitment - Excellence

The Contributors: The Analyst Team

The **DBP Insights report** is researched, analyzed, and prepared by the following Analysts.



Arun Marar, Ph.D. is an expert at implementing AI/ML models utilizing a wide range of technologies such as supervised and unsupervised learning, text analytics and natural language processing and deep learning. Arun has consulted with clients all around the world across domains such as Banking and Financial Services, Retail, Healthcare and Transportation Logistics. Arun is also an expert in optimization, and this has helped him aid companies in moving up the value chain of analytics from predictive analytics to prescriptive analytics. Dr. Marar can be reached at **arunmarar@gmail.com**

Prashanth Southekal, PhD, MBA, ICD.D is a Consultant, Author, and Professor. He has worked and consulted for over 80 organizations including P&G, GE, Shell, Apple, FedEx, and SAP. He is the author of three books — "Data for Business Performance", "Analytics Best Practices" and "Data Quality" — and writes regularly on data, analytics, and ML in Forbes and CFO University. His second book, ANALYTICS BEST PRACTICES was ranked #1 analytics books of all time in May 2022 by BookAuthority. Apart from his consulting and advisory pursuits, he has trained over 4,000 professionals worldwide in Data and Analytics. Dr. Southekal is also an Adjunct Professor of Data and Analytics at IE Business School (Madrid, Spain). Dr. Southekal can be reached at **psouthekal@dbp-institute.com**

Sanjeev Chib is a visionary Product and Data Executive with an obsession for constructing innovative solutions, leveraging technology and data, that transform complex problems into sustainable ventures. He is currently the VP of Digital Security Products at Symcor with a mandate to build data-driven solutions that tackle financial crimes and fraud. Previously, Sanjeev founded the Data Solutions subsidiary at Canada's largest payment processor (Moneris) leading cross-functional teams to incubate data products using over 3.5 billion annual transactions. In this role, he pioneered Canada's first marketing attribution platform (Offlinx®) connecting digital marketing spend to in-store sales, winning international patents for the underlying mechanism. Sanjeev is a sought-after advisor for start-ups and speaker on data, innovation, and risk. His experiences are backed by CPA, CA, and CISA designations, professional certifications in Strategy and Design Thinking, and a Bachelor of Commerce degree from the University of Toronto (CA). Connect with Sanjeev at linkedin.com/in/sanjeevchib or email him at **sanjeevchib@yahoo.com**

Tobias Zwingmann is Co-Founder of the German AI advisory firm RAPYD.AI and on a mission to help companies implement machine learning and AI faster while delivering meaningful business value. He is the author of the book AI-Powered Business Intelligence and has more than 15 years of professional experience working in a corporate setting where his responsibilities included building data science use cases and developing an enterprise-wide data strategy.

Mr. Zwingmann can be reached at **tobias@rapyd.ai**

DBP Insight 1: Business performance improvement will take the centerstage (again).

Background

Analytics, at its core, is leveraging data for derving insights to measure and improve business performance. In this regard, measurements reflect the performance of the company's goals in meeting business outcomes. Basically Measurements provide visibility, and visibility drives performance and, ultimately, business outcomes. Management guru Peter Drucker once said, "What gets measured gets managed." Aberdeen Group examined the use of KPIs in more than 350 enterprises and found that the measurements derive performance improvements, including a 10% increase in the time-to-decision making, a 9% increase in both profitability and revenue growth and customer performance improvements of 9% in both net-new customers gained and customer satisfaction. Basically, effectively measuring business outcomes can help enterprises grow their revenue, reduce costs (direct and indirect) and mitigate risks.

Rationale for this Prediction

Basically, a well-designed measurement framework holds the key to providing the right feedback and track business performance. Howeber in the ;asyt few yesrs, many data and analytics projects have failed. According to Gartner, over 85% of the data and analytics do not offer business outcomes. Hemce compnaies will be going back to the drawing board and looking at some fndamental questions including strategies for measuring and improving business performance from data and analytics.

Prescriptive Recommendation

Measuring business outcomes using D&A is expensive, time-consuming, and resource-intensive, measurement methodology can vary widely. Below are the top three measures enterprises can take to effectively measure and improve business performance.

1. Only a few entities are worth measuring. Albert Einstein once said, "Not everything that can be counted counts, and not everything that counts can be counted." So, what should businesses measure? Effective performance measurement with D&A using KPIs should directly or indirectly measure three things: revenue, costs (direct and indirect) and risks.
2. Measurement should be based on transactional data for three key reasons:
    a. Transactional data is not only dynamic in nature but also large in volume.

b. Transactional data records financial value as it deals with the consumption of business resources such as money, time, materials, plants and more.
c. Transactional data fosters relationships between the counterparties with two-fold effect on accounting.

3. Measure performance using quality data. So, what is quality data from a measurement perspective? Quality data can help the firm grow revenue, reduce costs, and mitigate risks by emabling imprpved oprrations, complaice, and analytics. In this regard, the following five questions will help ensure the validity of the overall performance measurement engagement:
    a. Why do you want to know?
    b. How much do you want to know?
    c. What is the value of knowing and not knowing?
    d. Who owns the entity that is measured to realize the change?
    e. Do you have the quality data to derive the insights?

DBP Insight 2: Enterprise Data Governance (EDG) will drive business enablement.

Background

Data Governance is the specification of decision rights and an accountability framework to manage data in the entire data lifecycle (DLC) so that right people are managing the right data in the right manner. Effective data governance improves the utilization of data in operations, compliance, and analytics in business. In addition, in the last few months, enterprises are increasingly considering internal LLMs (Large Language Models) to derive business benefits by addressing their concerns on customization, control, data privacy, security, and more. But many business enterprises which avoid risks and play safe, have transformed Data Governance into a regulatory compliance function. This has resulted in the poor utilization of data in other areas especially in analytics, ML (Machine Learning) and AI (Artificiaal Intelligence) thereby adversely affecting innovation and growth.

Rationale for this Prediction

In the last few years, implementation of regulations such as GDPR, CCPA, HIPAA and other compliance requirements, have transferred data governance into a purely compliance function. Too much of data governance has resulted in low levels of innovation and business growth. Organizations that started to realize the data governance is much more than compliance; it is enabling the business to transform data into a valuable business asset.

Prescriptive Recommendation

Below are the top three strategies enterprises can take to effectively implement data goverence and truly transform data into a business asset.

1. The Data Governance should be performed in the SoR (system of record) pertaining to all stages of the DLC (data lifecycle). A SoR is simply a data system that can be relied upon to have authoritative data source i.e., the Golden Record - the most accurate and correct values.
2. Demoncratzie the data by ensuyring that authorized users having access to use data to improve operations, compliance and analytics. The key in data demoncratization is authorized users. But if unauthorized and untrained users are given access to critical and sensitive data in the name of data democratization, the competitive advantage of the business might be jeopardized. Overall data

governance must balance data democratization with data protection with the right use-cases and RBAC (Role-based access control) controls.

3. Data Governance is not a one time project; it is a continuous improvement process that relies on data literacy, which is also an ongoing process. This means empower every team member with the right data and analytics skills to access and use data for improved business performamce.

DBP Insight 3: Data footprint reduction will move closer to ESG Initiatives

Background

A digital or data footprint is the body of data that is creates when an individial or a IT system performs some actions in an digital system. These actions could be directly or or indirtectly related to the three main purposes– operations, compliance, and analytics. The data footprint would be a profound issue in the coming years because businesses will capture a lot of data in different formats and timeframes resulting in increased data or digital footprint.

Rationale for this Prediction

Every minute, users add 300 new hours of video on YouTube, totalling more than 1 billion gigabytes of data. Facebook users share more than 100 terabytes of data daily. Video conferencing such as Zoom and Microsoft Teams also generate lot of data. According to an IDC report, "The volume of data stored in the Global Storage Sphere is doubling approximately every four years." AI also has a tremdous carbon footprint. Accoridng to MIT, generating one image in AI takes as much energy as fully charging your smartphone [MIT, 2023]. In 2019, University of Massachusetts Amherst researchers found that training a single AI model can emit over 626,000 pounds of CO2, equivalent to the emissions of five cars over their lifetimes [Cho, 2023].

But unfortunately, most of the data that is captured and stored is not being used. Forrester says that "73% of the data in enterprises is never used for any strategic purposes. All this creating more complexity not only in effectively trying to govern, manage, and secure it, but also in the management of costs, ESG metrics, and risk. Today, the data centers are contributing less than 2% of the world's carbon footprint. But in the next few years, as more data is captured, the data centers will double their carbon footprint. Hence there is a pressing need look for ways to reduce their company's data footprint.

Prescriptive Recommendation

Increase in data footpront affects the costs, ESG (Environment, Social, and Governanace) metrics and risk. So, what can enterprises do to reduce carbon footprint?

1. Identify your data and analytics strategy based on the level of digitization required for your business. In other words, understand the role and the importance of

data in the industry and inside the enterprise. Associate data capture, storage, and processing to the three main business objectives i.e., operations, compliance, and analytics.

2. Profile and catalog the existing enterprise data in the entire data lifecycle (DLC). Assign the data owner or custodians for critical data elements – especilaly the referece datra and master data elements. Instead of storing the data in native and unstructured format, label, structure and process the data to further enhance the value of data.
3. Define the retention and disposition schedule for the data. Build rules for efficient discovery of unused data assets i.e., dark data. For data objects that fall outside the retention and disposition schedule archiving, purging, and back-up routines, or even purge the data if needed. For example, in Florida, physicians must maintain medical records for five years after the last patient contact. But the IoT notofcation for a turbine vibrations are relevent for a few seconds. So there is no point in mainationiong the data beyond the relevent period.

DBP Insight 4: Synthetic data will enhance AI model development.

Background

AI models and more specifically Generative AI (GenAI) are advancing at a rapid pace. Hence it is imperative to understand that the models and the data they are trained upon. Very often the real-world data required to train the models is limited, and one must rely on synthetic data to make the training data more comprehensive including offering support in performance testing and exploring scalability. Basically, synthetic data which is algorithmically creating new data by mimicking real data has four key characteristics:

- Mirrors the statistical properties of the original data.
- Reflects the original data structure and format.
- Does not have mappings or traceability/linkage.
- Has no adherence to privacy or security (RBAC) or regulations.

Rationale for this Prediction

There are three scenarios of real-world data that can provide impetus in adopting synthetic data – scarcity, sensitivity, and difficult to obtain. In critical applications such as self-driving cars it is imperative that synthetic data is used to ensure events or outliers that are not adequately represented in the real-world data are represented adequately using synthetic data. Regulations prevent one from using personally identifiable information in training data and synthetic data has a role to play in anonymization of real-world data so they can provide valuable inputs to the AI models. Synthetic data improves data diversity making training data sets more comprehensive and representative. Synthetic data can also be employed to introduce controlled biases that would mitigate the problem of bias and fairness in the training data.

Prescriptive Recommendation

So, when do enterprises take advantage of synthetic data to build and deploy data, analytics, and AI models?

1. Identify the three real-world scenarios in the enterprise that prompt the need for synthetic data – scarcity, sensitivity and difficult to obtain. Any of these three situations demand a need for synthetic data. To manage security and privacy controls use, Data clean rooms (DCR) and share aggregated data rather than granular data.

2. Be sure to frequently benchmark and validate synthetic data with real-world data so that the AI models are always realistic and do not suffer from model degradation.
3. Identify opportunities for innovation such as new products or ideas. Utilize synthetic data in training the analytics and AI models in such situations.

DBP Insight 5: Human potential will be augmented with AI.

Background

In the evolving landscape of artificial intelligence, augmentation, not replacement, is the future of human-AI interaction. The challenge lies in leveraging AI to enhance human capabilities without crossing into the territory of replacement, a balance essential for the sustainable integration of AI into the workforce. Generative AI, currently at the forefront of technological innovation, is making significant strides. However, while GenAI often creates awesome results in 99% of all cases, in 1% it fails miserably. How do we combat this situation? The answer is straightforward: We either need (A) strong testing systems or safety nets, or (B) keep a human in the loop to oversee the results. And since no one has figured out A yet, the only viable option in the short term is B, ensuring that the technology complements rather than overrules human judgment.

Rationale

Users have become the default 'human in the loop' in AI applications, indicating a shift towards more interactive and collaborative AI systems. This trend towards augmentation is fueled by the desire to improve decision-making and efficiency, not to replace human workers. Real-world instances, such as AI-enhanced customer service and product design, exemplify this synergy, showcasing how AI can suggest improvements while leaving room for human expertise and discretion. The rise of copilots is another strong indicator. By 2024 every major enterprise software company has announced or is somehow working on a “Copilot” feature for their software. Popular examples include Microsoft, SAP, and Salesforce.

Prescriptive Recommendations

Augmentation, while transformative, demands careful implementation to truly benefit enterprises and their workforce. The secret lies in the 'sweet spot' of AI-human collaboration – enough to significantly boost efficiency without overwhelming or sidelining human input. To prepare for the augmentation trend, businesses should adapt:

1. Balancing AI and Human Input. The blend of AI and human skills is vital in various sectors. For instance, in customer service, chatbots manage routine inquiries, freeing up human agents for complex issues. This synergy is crucial for maximizing efficiency and maintaining a high standard of service.

2. Designing AI to Support Humans. AI should be designed to complement human tasks, not take them over. This approach is evident in fields like product design, where AI tools analyze customer feedback and trends, offering valuable insights that guide human designers in their creative process. The right 'dose' of AI can significantly enhance decision-making, but this varies widely across different tasks and user profiles. The user must not be overwhelmed by the technology, but it needs to blend in seamlessly.
3. Contextual Application of Augmentation. Understanding the user's level of analytical maturity and data literacy is key to effective augmentation. AI tools need to be customized to the user's specific requirements, ensuring that they provide the right level of support without overwhelming or underutilizing human capabilities.
4. Finding the Augmentation Sweet Spot. The key lies in finding the right balance between AI capabilities and human intuition. If the AI suggests are only weak (consider a spell check in a text writing software), users won't see enough value. If it's too much (consider generating a whole article from a single keyword), people might rely on the technology too much, neglecting their own unique skill set. This 'sweet spot' varies across different tasks and industries. It requires careful consideration of the specific problem at hand and the unique capabilities of both the AI system and the human users involved.

DBP Insight 6: AI model drift will make AI adoption risky.

Background

Organizations have long since struggled to implement AI/ML models for a variety of business use cases in their enterprise. Gartner has claimed that 85% of the AI/ML model fail eventually. While there are several reasons for this such as misalignment of ROI, issues regarding quality and quantity of data etc., one of the main issues is AI model "drift". This happens when model degradation is observed in time. One of the main reasons for this happening is the dislocation of the real-world data from the data the model was trained on before it was implemented in production. This results in model failing to perform as per the expected metrics of model performance.

Rationale for this Prediction

In the burgeoning age of Generative AI (GenAI), ChatGPT revolutionized the industry. However, an interesting observation was made by a group of researchers from Stanford University and University of California, Berkeley. They observed that ChatGPT was getting "dumber" based on comparing performance in March 2023 outperforming the performance in June 2023. This can be attributed to AI model drift even as systems like ChatGPT were expected to perform better in time because of the feedback system inherent in their functioning.

Prescriptive Recommendation

So, how can enterprises address the risks of AI model drift?

1. Data Quality: The main reason for observing AI model drift or model degradation is because of the change in data distribution in time. Thus, it should be ensured that data quality is ensured at every stage the model is retrained and is representative of the real-world in real time.
2. Adaptive Learning: AI models must learn to adapt to the changing environment such as change in business processes in operations that necessitates reinventing the AI model paradigm from a different perspective altogether.
3. Feedback Loops with Human Oversight and Continuous Monitoring: One of the mitigation techniques of AI model drift is to have human in the loop at every stage of the process of putting the model into production – model design and inception to model implementation. Also, wherever amenable, implement feedback loops that can help in realizing that the AI model is drifting, and the situation needs to be addressed at the earliest. In addition, AI models should be

continuously monitored to ensure that the model is performing to expectations. The expectations are governed by a list of performance metrics such as accuracy, precision, and recall.

DBP Insight 7: Data privacy will dictate data connectivity solutions.

Background

The power of data to deliver effective insights and solutions grows exponentially when you connect internal and external data sources to enrich datasets. The benefits of connecting data becomes even more evident when feeding AI models, enabling them to create relationships across different datasets to produce better results. However, connecting these different data sources is often challenging and not always accurate, resulting in less reliable insights or even adding additional bias to results.

Rationale for this Prediction

The importance of connecting data is increasing. The significant benefits of connecting and enriching data to feed AI algorithms and analytics models will drive the need to find better mechanisms to connect this data effectively. Determining the best linking identifiers or other methods, such as creating robust ID Graphs, to stitch customer data or multiple data sources in a privacy-compliant manner, may further be supported using AI powered methods. Further challenging the connectivity is the rise of privacy regulations requiring explicit consents and restrictions with 3rd party cookies on browsers that have degraded some of the connectivity mechanisms and identifiers. Without an effective data connection mechanism, various data-driven and analytics solutions such as targeted advertising, marketing attribution, customer insights and business performance measurement are impacted.

Prescriptive Recommendation

Enriching data sets by connecting multiple internal and external sources will drive significant value to organizations by delivering better performance insights and enabling them to develop better products. Organizations that seek to develop these capabilities should consider the following:

1. Effectiveness: As your organization determines their data strategy and the various data sources required, the mechanisms to connect these data sources should also be a consideration in the data strategy. More importantly, the effectiveness of the connection mechanism and the acceptable match rate should be part of the calculation.
2. Achieving a 100% match rate is often not possible or expensive, and in some cases, much smaller match rates to create a representative sample that can be used to infer population-level results by the algorithms and models may be

sufficient. However ensure that the sample selected is good in count, is representative of the sample, and is randomly selected.

3. Privacy sensitivity: Organizations should be aware of the various privacy regulations that need to be followed and ensure that the data connectivity methods used do not violate compliance with these regulations. Non-compliance may prove costly to the organization, both financially and reputationally. Therefore, if 3rd party connectivity tools or solutions (e.g., interest-based audience segments) are used, ensure the solutions are compliant with the changing regulations.

DBP Insight 8: AI Flexibility will be bolstered by Open-Source ecosystem.

Background

Organizations must prioritize adaptability over commitment to rigid solutions in the fast-moving AI landscape. The challenge is to balance a field that is still maturing, with a growing number of commodity and off-the-shelf solutions. Open-source tools provide a good vehicle for this. The AI industry is a constant flux. OpenAI almost stumbled over a weekend and established leaders like Google and Apple are still looking for their place. This volatility underscores the importance of agility for organizations working with AI technologies. Prematurely committing to a single AI solution can lead to missed opportunities as the field rapidly evolves and is ultimately hard to predict.

Rationale

Statistics show a skyrocketing open-source AI ecosystem: over 400,000 open-source models available on Hugging Face (with the vast majority added in the last 12 months) and over 60,000 Github stars for LangChain (a popular and Sequoia-backed LLM application framework less than two years old) are just two examples of this thriving ecosystem. In addition, customizing LLM applications and fine-tuning LLM models has never been easier, thanks to advances in both technology (see Huggingface AutoTrain as an example) and foundational research (e.g., see the ChipNeMo paper on domain-adapted continued pre-training). These building blocks represent an unprecedented opportunity for custom AI solutions. However, maintaining a balance between customization and pre-built solutions is critical, especially when AI developer talent is scarce.

Prescriptive Recommendations

Here are some ways in which businesses can embrace this trend and prepare now:

1. Avoid Vendor Lock-In. Long-term contracts with AI vendors can limit flexibility. Organizations should seek agreements that allow for adaptation and change, ensuring that they can pivot as new AI technologies and approaches emerge – especially for critical, strategic business use cases.
2. Embrace An Open-Source Tool Stack. Adopting an open-source stack provides the flexibility that is critical in the volatile AI market. This approach allows organizations to stay at the forefront of innovation and adapt quickly to new

developments, without being tied to a single vendor or solution - especially for critical use cases that are moved from prototype to production.

3. Experiment with Adaptation. Fine-tuning pre-trained AI models to meet specific business needs is no longer a distant possibility, but a present reality. Experimentation not only leads to more tailored solutions, but also fosters a deeper understanding of AI capabilities and is especially relevant for larger organizations that wouldn't consider themselves a "tech" company.
4. Develop Experimentation & Performance Tracking. Without a system in place to measure the performance of AI models, organizations cannot gauge the effectiveness of their custom solutions. Developing robust performance metrics is essential to understanding the value of customization and guiding future AI strategies.

In conclusion, the path to leveraging AI in business is not through a rigid commitment to single solutions, but through a flexible, customizable approach that can adapt to the rapid advances in the AI space.

DBP Insight 9: Risk intelligence will evolve to tackle AI-powered cyber threats.

Background

As the world continues to become more digitized, with larger amounts of data and payments flowing through digital streams, there has been a marked increase in sophisticated fraud, financial crimes, and cyber threats. As a result, solutions will need to evolve, utilizing complex data analytics methods and AI-powered defences, to detect, deter, and prevent attacks across a broader spectrum of financial crimes and cyber threats. For instance, machine learning and AI may be used increasingly to detect fake payment requests, fraudulent patterns in payments flowing across institutions, or to detect suspect ("mule") accounts used in money-laundering. These next-generation threat prevention solutions will increasingly be adopted by institutions and companies as they face more sophisticated threats.

Rationale for this Prediction

Driving this need for more sophisticated, data and AI-enabled "risk intelligence" solutions is the increase in AI capabilities, geo-political uncertainties and payments digitization enabling a significant predicted rise in financial crimes and cyber threats.

1. Rise of AI Capabilities to Replicate Identities: As AI capabilities advance in replicating people's faces, voices, or other aspects of their identity in content, this escalation will heighten the susceptibility to financial crimes. Fraudsters will increasingly rely on these capabilities to, for example, make requests for payments or other transactions appear legitimate.
2. Global geo-political uncertainties: Continuing tensions and wars (as well as influencing election results to shift power centres) across the globe need increasing funding and will force the different factions to seek new, stealth ways of raising and moving money used to fund their efforts and causes.
3. Payments Digitization: The modernization of payment platforms is emerging. Real-time Rails (RTR) and open-banking regulations and frameworks are gaining increased interests and, along with the continued growth of cryptocurrencies, these will enable consumers to transact and move money faster and globally. Vulnerabilities within these digital payments' methods will be exploited by fraudsters to further fuel their financial crimes.

Prescriptive Recommendation

Organizations facing these increased cyber threats and exposures to financial crimes (e.g., fraudulent requests for payments) are recommended to consider the following:

1. Fortify existing processes and data: Zero-trust security frameworks have been gaining prominence and should be considered to strengthen data security. In addition, additional financial controls should be implemented to ensure outgoing payments are valid and secured before release.
2. Apply data governance controls as they relate to data classification and protection should be reinforced, and data loss protection solutions implemented. Tokenization of critically sensitive data (e.g., personal information, payment card data, etc.) should also be considered to protect data.
3. Implement enhanced Risk Intelligence and Security Solutions: As better data and AI-driven solutions emerge, consider implementing these new zero-trust security platforms that will mitigate cyber threats by effectively detecting, deterring, and preventing loss from ransomware and other cyber attacks to your apps, data, and networks. These new solutions may also be effective in mitigating financial crimes such as detecting and alerting fake requests for payments.

DBP Insight 10: "AI-as-a-Service" will accelerate embedding of AI into existing solutions.

Background

Over the past decade, there's been a shift toward smarter products integrating data and predictive analytics. Examples like thermostats adjusting temperature based on user-predicted behaviour, and HR tools predicting candidates' suitability are now commonplace. Yet, some solutions have shown mixed success, like flawed applicant tracking software rejecting strong candidates. The emergence of new AI capabilities, like LLMs to create new content (e.g., ChatGPT, Dall-E), offers possibilities for more effective problem-solving and smarter products. For instance, Canva links its design platform to multiple AI image generators like Dall-E by OpenAI, Magic Media's Text to Image, and Imagen by Google Cloud. Additionally, Khan Academy's AI tutor Khanmigo and UpToDate, an AI-enhanced diagnosis tool for doctors, signify this integration.

Rationale for this Prediction

In an environment of high interest and inflation causing declines in consumer spending and growth, companies are looking for new approaches to attracting consumers and maintaining revenue by creating new value and experiences with existing products. As AI capabilities continue to evolve and become easier to connect and embed, the trend is towards the democratization of AI services. Basically, as AI improves, companies will hasten embedding these capabilities into existing products. But developing these AI capabilities is complex and expensive. Hence, their development will be led by specialized tech companies and startups who will offer AI services ("AI-as-a-Service") to early adopter companies via API plug-ins, granting a temporary competitive edge until widespread adoption.

Prescriptive Recommendation

Companies looking to take advantage of AI services to enhance customer value and experiences are advised to consider the following:

1. Problem-Focused: Start by being clear on the problem your product is solving, and ask the "What if." questions where AI can be utilized to make the product more effective, or to enhance the customer experience. Determine if appropriate AI

services exist yet that can be easily integrated and will be able to effectively solve the problem.

2. Viability: Adding AI features and services is not cheap. Determine the growth strategy for your product, such as increasing the Life-Time Value (LTV) of existing customers or attracting new customers. Understand how the new AI features in your product will support these growth drivers (e.g., by making customers more 'sticky' to increase LTV). As with any new product feature, measuring these metrics and taking a 'test and learn' approach will help mitigate risks and improve the chance of success.
3. Protect the Downside: AI features may require feeding additional data, and with this comes added risks to ensure privacy, ethical uses and prevent data loses. Some questions you need to consider include whether your new AI feature requires passing a lot of user generated data to the AI service, does this data include privacy-sensitive or confidential data, how will you and the 3rd party AI service provider protect this data, and how will biases in the data impact the results presented through your product platform? In addition, as your product's reliance on these AI services grows, you need to understand the impact of outages and downtimes of AI services on your customers and your business reputation. You also need to understand whether the AI services will be able to scale effectively as your user base grows.

Required Data, Analytics, and AI Capabilities

Fundamentally, the purpose of data and analytics is to offer insights to **know or to act** for the organization. If the purpose of analytics is to just **know**, ask - why do you want to know? how much do you want to know - the time horizon? what is the value of knowing and not knowing? But if the purpose of insights is to **act**, ask what are the events that are adversely affecting the organization's ability to plan, respond, and control from the predictive analytics insight? Also, spare a thought on who will own or consume the derived insights? How can you realize the change?

In this backdrop, here are the 12 important capabilities required in effectively adopting the ten DBP insights in the organization. The below capabilities are specific to **Data, Analytics, and AI** and assumes that organizations have good capabilities on problem solving, collaboration/teamwork, customer service, talent management, and more.

1. Office of the CDO
2. Enterprise Performance Management (EPM)
3. Digital Solutions
4. Master Data Management (MDM)
5. Data Engineering
6. Canonical Analytics System (DWH/Data Lake/Data Lakehouse)
7. Semantic Layer and Data Catalog
8. BI/AI/ML Platforms
9. Data Literacy and Change Management
10. Data Governance, Security, Privacy, Ethics, and Bias
11. Data Products Mindset
12. Decision making using data under Uncertainty and Risk

Criteria to Identify Data, Analytics and AI Projects

So, how do you implement these Data, Analytics and AI predictions? Selecting the Data and Analytics Projects for tangible business impact is based on **five key criteria**.

1. The Data, Analytics, and AI Projects selected should measure and improve/optimize business objectives and performance. Specifically, the problem definition should be linked to 3 key business objectives: enhancing revenue, reducing costs and/or minimizing risks.
2. The Data and Analytics Projects selected should be practical and relevant for the business. The selection should be based on questions and factors that are under the influence and control of the business. The business question (dependent variable) should be contextual, and the data sourced (independent variables) should be relevant for deriving the answer to the question.
3. Data and Analytics Projects should apply appropriate concepts and techniques such as EDA, association, segmentation, prediction, prescriptive Analytics, and so on to measure and improve/optimize business performance. The pre-requisite for this is availability of data: and its ease of integration. Fundamentally Data and Analytics help in leveraging the past performance to drive future results and performance. In addition, the project should apply quality data supported by realistic assumptions.
4. Data and Analytics Projects drive change. Hence the Data and Analytics Projects should address the implementation and change management aspects. Specific questions include:
    a. Whose problem (person or persona) are you solving?
    b. Who owns the Data and Analytics solution?
    c. Availability of Resources and Capabilities/Skills
    d. What are the risks of implementing this solution? Is the Data and Analytics solution ethical?
    e. What solution and governance components are in play in the Data and Analytics Projects
5. Finally, the Data and Analytics Projects should deliver quantifiable business impact measured using KPIs such as ROIC (Return on Invested Capital), FCF (Free Cash Flow), Growth, CX (Customer Experience), etc.

Conclusion

In today's volatile, uncertain, complex, and ambiguous (VUCA) global market, change is the only constant factor. However, the digital business offers unlimited possibilities for enterprises to create business value through data and cope up with the change. However, just capturing and storing data will not make an enterprise data driven. Data can be an asset only if its managed well; if not data can become a limitation, or even a liability. There must be a planned and conscious effort to realize value from data so that it becomes a valuable business asset. Hence business enterprises must seek to acquire quality data, adopt AI/ML models, build the right talent pool, re-engineer business models, improve the data maturity, governance, data literacy, and so on to keep pace with the change.

In general, holistic insights from data and analytics are based on four key factors – data, models, assumptions, and ethics. Often, most discussions on Analytics are centered on data and models. But the real challenges in deriving insights are on the latter two: assumptions and ethics. At the core, getting results from Data and Analytics centers on change management. The former CEO of GE, Jack Welsh, said – "An organization's ability to translate insights into action rapidly is the ultimate competitive advantage.".

References

- IDC, "IDC FutureScape: Worldwide Digital Transformation 2021 Predictions", https://www.idc.com/research/viewtoc.jsp?containerId=US46880818, 2021
- MIT, https://www.technologyreview.com/2023/12/05/1084417/ais-carbon-footprint-is-bigger-than-you-think/, 2023
- Cho, Renee, https://news.climate.columbia.edu/2023/06/09/ais-growing-carbon-footprint/, 2023
- Southekal, Prashanth, "Data Quality: Empowering Businesses with Analytics and AI", Wiley, 2023
- Statista, "Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025", https://www.statista.com/statistics/871513/worldwide-data-created/, 2022

Data, Analytics, and AI Acronyms

- AI - Artificial Intelligence
- CDO - Chief Data Officer
- CDW - Cloud Data Warehouses
- COTS - Commercial-off-the-shelf
- DDI - Data-driven innovation
- DLC – Data (and Analytics) Lifecycle.
- EDW – Enterprise Datawarehouse
- EDG – Enterprise Data Governance
- EPM – Enterprise Performance Management
- IT - Information Technology
- KPI - Key Performance Indicator
- LLM – Large Language Models
- LMA – Last Mile Analytics
- MDM - Master Data Management
- ML - Machine Learning
- ROI - Return on Investment
- RBAC – Role Based Access Control
- SSA – Self Serve Analytics
- TAVI - Text, Audio, Video and Images.
- TCO - Total Cost of Ownership
- VUCA - Volatile, Uncertain, Complex and Ambiguous

Data, Analytics, and AI Glossary

- Algorithm: A mathematical formula or statistical process used to perform analysis of data.
- API (Application Program Interface): A set of programming standards and instructions for accessing or building web-based software applications.
- Artificial Intelligence: The ability of a machine to apply information gained from previous experience accurately to new situations in a way that a human would.
- Big Data: Big data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze. Big data sets are characterized by 3Vs i.e., volume, velocity, and variety.
- Business Intelligence: The general term used for the identification, extraction, and analysis of multi-dimensional data.
- Change Management: Change management is the discipline that guides how we prepare, equip and support individuals to successfully adopt change to drive organizational success and outcomes.
- Cloud Computing: A distributed computing system hosted and running on remote servers and accessible from anywhere on the internet.
- Dashboard: Dashboard is a visual snapshot of business performance using KPIs to help users make smarter, data-driven decisions.
- Data Analytics: The process of answering business questions using data. Businesses typically use the three types of analytics: Descriptive, Predictive and Prescriptive Analytics.
- Data Analysis Workspace: It is a flexible browser-based interface that allows you to quickly build analyses and share insights. Using the drag-and-drop interface, you can craft your analysis, add visualizations to bring data to life, curate a dataset, share and schedule projects with anyone in your organization.
- Data Architecture: It is the mechanism in which data is collected, and how it is stored, arranged, integrated, and used in data systems and in organizations.
- Data Governance: A set of processes or rules that ensure data integrity and that data management best practices are met.
- Data Integration: The process of combining data from different sources and presenting it in a single view.
- Data Lake: A large repository of enterprise-wide data in raw format – structured and unstructured data.

- Data Mart: The access layer of a data warehouse used to provide data to users.
- Data Mining: It is finding meaningful patterns and deriving insights in large sets of data using sophisticated pattern recognition techniques. To derive meaningful patterns, data miners use statistics, machine learning algorithms, and artificial intelligence techniques.
- DataOps: DataOps as defined by Gartner is a collaborative data management practice focused on improving the communication, integration, and automation of data flows between data managers and data consumers across an organization.
- Data Product: A data product is the application of data for improving business performance; it is usually an output of the data science activity.
- Data Science: A discipline that incorporates statistics, data visualization, computer programming, data mining, machine learning and database engineering to solve complex problems.
- Data Warehouse: A repository for enterprise-wide data but in a structured format after cleaning and integrating with other sources. Data warehouses are typically used for conventional data (but not exclusively).
- Database: A digital collection of data and the structure around which the data is organized. The data is typically entered into and accessed via a database management system.
- Deep learning: Deep learning is a function of AI that imitates the human brain by learning from how it structures and processes information to make decisions. Instead of relying on an algorithm that can only perform one specific task, this subset of machine learning can learn from unstructured data without supervision.
- Descriptive Analytics: Condensing big numbers into smaller pieces of information. This is like summarizing the data story. Rather than listing every single number and detail, there is a general thrust and narrative.
- ETL (Extract, Transform and Load): The process of extracting raw data, transforming by cleaning/enriching the data to make it fit operational needs and loading into the appropriate repository for the system's use.
- Hallucination: Hallucination refers to an incorrect response from an AI system, or false information in an output that is presented as information.
- Hypothesis. A hypothesis is an assumption, an idea, or a gut feeling that is proposed for the validation so that it can be tested to see if it might be true.

- Insight. It is the understanding of a specific cause and effect within a specific context. In this book, the terms insight and information are used interchangeably.
- KPI. A Key Performance Indicator (KPI) is a measurable value that demonstrates how effectively the entity is achieving key objectives or targets.
- Machine Learning: A method of designing systems that can learn, adjust and improve based on the data fed to them. Using statistical algorithms that are fed to these machines, they learn and continually zero in on "correct" behavior and insights and they keep improving as more data flows through the system.
- Metadata. Any data used to describe other data — for example, a data file's size or date of creation.
- ModelOps: ModelOps as defined by Gartner is on the governance and life cycle management of a wide range of operationalized artificial intelligence (AI) and decision models, including machine learning, knowledge graphs, rules, optimization, linguistic and agent-based models.
- Online analytical processing (OLAP). The process of analyzing multidimensional data using three operations: consolidation (the aggregation of available), drill-down (the ability for users to see the underlying details), and slice and dice (the ability for users to select subsets and view them from different perspectives). OLAP systems are used in BI reports.
- Online transactional processing (OLTP). The process of providing users with access to large amounts of transactional data in a way that they can derive meaning from it. OLTP systems are used in Transactional reports.
- Predictive Analytics: Using statistical functions on one or more data sets to predict trends or future events.
- Prescriptive Analytics: Prescriptive analytics builds on predictive analytics by including actions and make data-driven decisions by looking at the impacts of various actions.
- Prompt: A prompt is an input that a user feeds to an AI system to get a desired result or output.
- RBAC (Role Based Access Control): RBAC is an access-control mechanism defined around factors such as authority, responsibility, and competency.
- Semantic Layer: The semantic layer is the representation of data that helps different business end-users discover and access the right data efficiently, effectively, and effortlessly using common business terms.

- SQL (Structured Query Language): A programming language for retrieving data from a relational database.
- Stable Diffusion: It is a latent text-to-image diffusion model capable of generating photo-realistic images given any text input, cultivates autonomous freedom to produce incredible imagery, empowers billions of people to create stunning art within seconds.
- Systems of Insight (SoI). It is the system used to perform data analysis from the data that is combined from the SoR or transactional systems.
- System of Record (SoR). The authoritative data source for a data element. To ensure data integrity in the enterprise, there must be one — and only one — system of record for a data element.
- Structured Data: Data that is organized according to a predetermined structure.
- Unstructured Data: Data that has no identifiable structure, such as text, audio, video and images.

DBP Insights: 2024 Data and Analytics Predictions and Prescriptions

Research Supported By

ARCALÉA

Grihasoft
Benefit from us

CFO.UNIVERSITY

DBP (Data for Business Performance) Institute is a Consulting, Research, and Education firm**,** that helps organizations leverage digital technologies, data, and analytics for improved business results. For more information, visit https://www.dbp-institute.com.