SMART DATA

Abstract

Information that truly makes sense is referred to as "Smart Data." It's the distinction between observing a long list of figures corresponding to weekly sales and recognizing the peaks and valleys in sales volume over time. Algorithms transform useless data into practical knowledge. Smart data is data from which clever algorithms have extracted signals and patterns. Large-scale data collection has limited value if there isn't an intelligence layer added to it.

Authors

Anamika Soni Datley
MCA
SRIT,Jabalpur
anamikashailesh0201@gmail.com

Shailesh Soni
MCA
SRIT, Jabalpur
shailesh02101974@gmail.com

I. INTRODUCTION

The amount of information being collected and collated worldwide is still expanding at an incredible rate. One minute is a huge amount of data when it comes to social media alone. During a minute:

- 48 hours of video clips are uploaded;
- 100,000 tweets are posted;
- 600,000 discussions and texts are exchanged.

You may begin to grasp the true nature of data when you realize that it includes IoT sensors, medical records, business communications, and much more. Data is not inherently evil; it is, in fact, the foundation for the resurgence of interest in and advancement of machine learning and artificial intelligence (AI) technologies (ML).

However, there are situations when data quality is questionable. The cost of "poor data" has been calculated to be close to 12% of an organization's income. In this case, the proverb "trash in, rubbish out" is often pretty applicable.

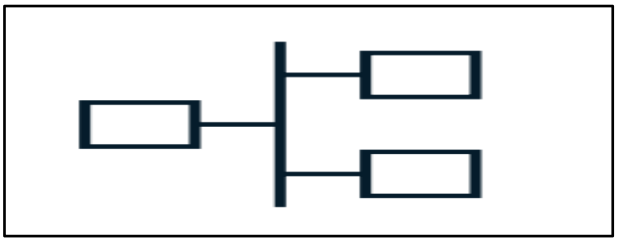
When data is collected, it is already categorized and tagged thanks to smart data. This ensures that data analysis doesn't require any more optimization. Many organizations think that gathering data is a good thing in and of itself. Companies amass vast volumes of information on their clients, transactions, goods, and services in data lakes and data warehouses. They think that gathering data with the intention of using it in the future is the best course of action. In actuality, gathering data that is helpful to their business is a superior choice. It costs money and takes time to collect data. It is more effective to spend time and money gathering useful smart data.

When data is collected, it is already categorized and tagged thanks to smart data. This ensures that data analysis doesn't require any more optimization. Many organizations think that gathering data is a good thing in and of itself. Companies amass vast volumes of

information on their clients, transactions, goods, and services in data lakes and data warehouses. They think that gathering data with the intention of using it in the future is the best course of action. In actuality, gathering data that is helpful to their business is a superior choice. It costs money and takes time to collect data. It is more effective to spend time and money gathering useful smart data.

II. CREATING SMART DATA

Poor integration of operational skills into the data-science process is a typical cause of failure for businesses attempting to use AI. Likewise, we recommend using machine learning only after expert-driven data engineering has been used to evaluate, enhance, and modify process data. We advise the steps shown in Figure 1 in practice:



1. **Define the process**: With the assistance of specialists and plant engineers, draw out the physical changes (such grinding and heating) and chemical changes that will occur throughout each phase of the process (such as oxidation and polymerization). Determine the key sensors and instruments, their limitations, maintenance due dates, measurement units, and whether or not they can be managed. Next, take note of the factors at play as well as the deterministic equations (such as thermodynamic relationships or reaction stoichiometry) that control the process. A literature search should be done in conjunction with the last phase to broaden the scope of thought beyond what the organization already knows. Use of outside specialists may be necessary if process expertise is limited.

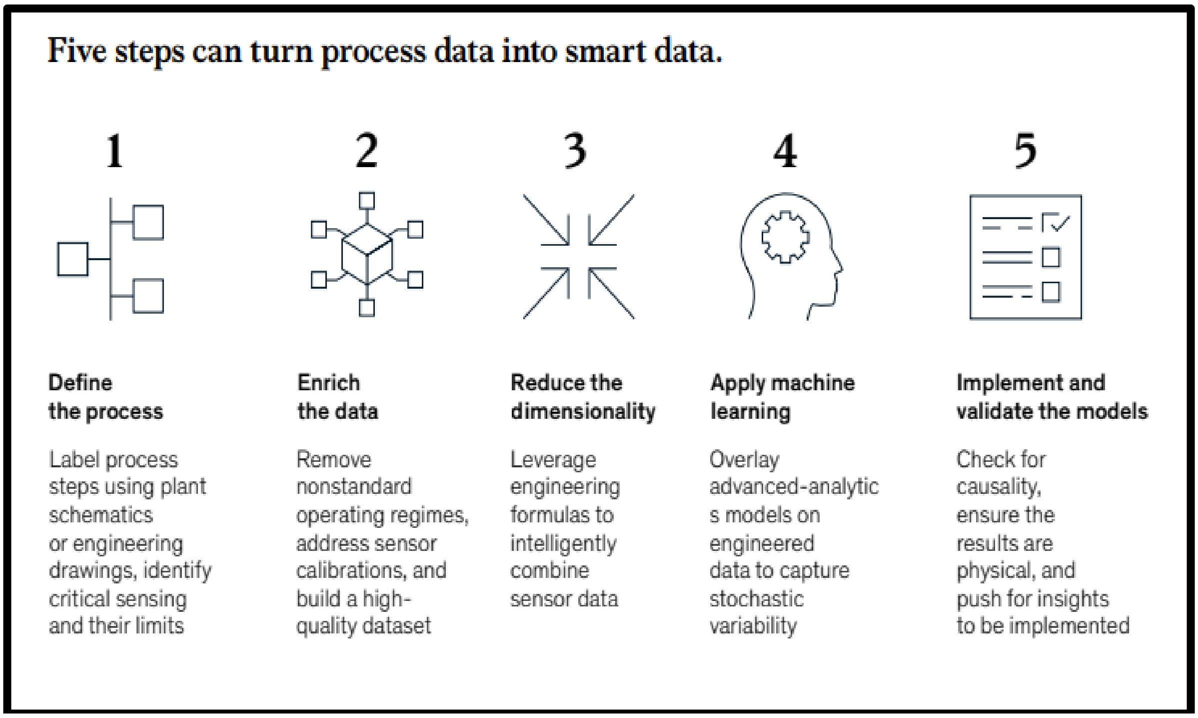
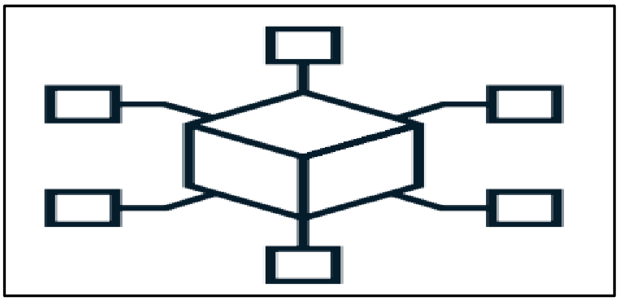


Figure 1: Steps for converting raw data to smart data

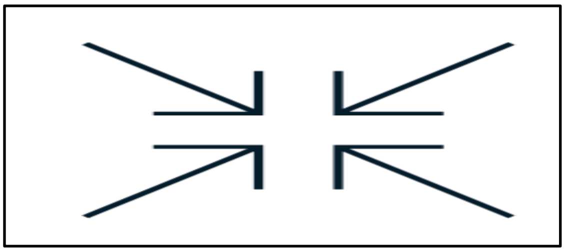
For instance, a North American mining corporation worked to increase the throughput of its grinding operations, which comprised three cyclone "classifiers" and seven grinding mills to sort particles according to size. The process flow, which was broken down into three steps of grinding and separation and each of which was monitored by around a dozen sensors, was presented to the data science team by experts and engineers while they were seated. Together with sensor redundancy and instrument accuracies, data tags (or labels) were recorded. Derivations of the Bond equation7 for grinding energy and Plitt's equation6 for particle separation were among the contributions made by the metallurgical team. The end result was a cohesive team made up of data scientists and plant specialists who knew what to look for in the field that may effect the models that were produced.



2. **Enrich the data:** Raw process data almost usually has errors. So, rather than aiming for the greatest possible number of observables for training, a high-quality dataset should be the primary goal. Teams should be proactive in deleting data from non-related plant configurations or operating regimes as well as non-steady state information, such as the ramping up and down of equipment. Avoid adopting generic techniques to handle missing or abnormal data, such as imputing using averages, "clipping" to a maximum, or fitting to

a hypothetical normal distribution. Instead, teams should carefully fill in data gaps using virtual sensors and physically sound imputations, starting with the important sensors that process experts have identified. A European chemical business, for instance, wanted to use machine learning in its cracking furnace.

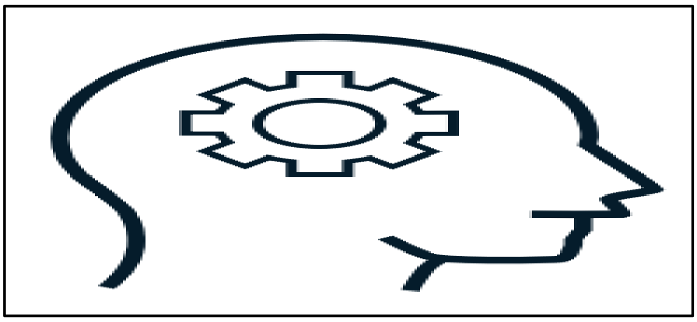
The data-science team discovered the flow meter was defective and that the numbers were occasionally incorrect due to miscalibration, despite experts' claims that it was crucial to the operation. In order to wait for a new flow meter to be placed, the operations team suggested stopping the project. Alternatively, by applying mass-balance formulae and upstream sensor data for temperature and energy consumption, a virtual flow sensor was created to enhance the present values. The analytics team was able to triangulate and rectify the flow numbers using the virtual sensor that was created. The project produced an overall increase in processing throughput of 20%



3. **Reduce the dimensionality:** By matching observable outputs to a collection of characteristics, which are made up of raw sensor data or its derivatives, AI algorithms create a model. To produce a generalized model, the number of observables often has to be significantly more than the number of features. Creating new features by engineering input combinations is a frequent data-science strategy. This requires a tremendous amount of observations, especially given the vast number of sensors present in contemporary plants. Teams should instead narrow the set of features to just include inputs that explain the physical process, then use deterministic equations to construct features that cleverly blend sensor data (such as combining mass and flow to yield density). This is frequently a great approach to incorporate relationships into the data and lower its dimensionality, which reduces the amount of observables needed to properly train a model.

   As an instance, a European chemical business noticed that the feed line to a spray drier occasionally experienced pressure surges, which required breaks or slowdowns in its continuous operation. To forecast the build-up of pressure, a model was created. The outcomes were poor even when all pertinent sensor data was taken into account. In response, the team added some sensor data and specifics of the pipe shape to the Darcy-Weis batch equation. As a consequence, the number of model inputs was decreased, the data quality was improved, and the model's performance went up. Operators might then use the model to their advantage.

   The model was then used by the operators to almost completely eliminate slowdowns, resulting in an 8 percent boost in throughput.



4. **Apply machine learning:** Deterministic and stochastic elements can be found in industrial processes. In actuality, machine learning models should capture the statistical component from auxiliary sensors and data, while first principle-based features should offer the deterministic component. Teams should assess characteristics' relevance and, consequently, their explanatory capacity while evaluating them. Expertly designed elements that, for instance, capture the physics of the process have to be among the most crucial. Generally, rather than fine-tuning a model to obtain the best forecast accuracy, the emphasis should be on developing models that drive plant improvement. Teams should be aware that process data exhibits large correlations by nature. Although model performance may appear to be outstanding in some circumstances, it is more crucial to distinguish the causative factors and manageable variables. Lastly, it is important to assess sensor data inaccuracies in relation to the objective function. Data scientists frequently aim for greater model accuracy only to discover that it is constrained by sensor accuracy.

   For instance, a metal manufacturer in North America sought to develop a model to forecast the temperature required to melt a batch of recycled materials. Based on specific-heat calculations that take into account the mass, heat capacity, and melting point of each alloy, the researchers initially developed one deterministic characteristic for "needed heat." In order to capture stochastic behavior, such as heat loss via the flue or fluctuations in the air temperature, data from 19 sensors were subsequently included as features. The final model performed quite well, with the deterministic feature having a greater than 80% relevance. [9] Operators might use the predictions to order melting when the model output was provided directly to a human-machine interface (HMI).

5. **Implement and validate the models**: Only through implementing models (or their conclusions) can impact be realized. Taking initiative is essential. Teams should regularly evaluate model findings with specialists by validating what is genuinely controllable, assessing key characteristics to make sure they match the physical process, and studying partial dependency plots (PDPs) to understand causality. Further meetings with the operations team should be scheduled to determine what can be introduced and to establish baseline performance. Before spending money on production-grade, automated solutions, teams frequently participate in on-off testing or communicate model findings to operators in a control room in real time. A European bioscience business, for instance, attempted to maximize the production of their fermentation process in the absence of sufficient data. Following early modelling efforts, sensor data and designed characteristics could only account for 40% of the throughput variability. The researchers designed an experiment at the plant using information from the parameter relations in the model, and the findings

were utilized to enhance the model and advise operations on where to install new sensors. Colleagues in data science and operations came to an agreement as a result, which led to a production increase of more than 20%.

Building the Team

Teams made up of operators, data scientists, automation engineers, and process specialists are needed to use AI in heavy industry. We frequently observe that companies have (or are hiring to fill) roles for data science, but they face three main difficulties with regard to process experts: there is a lack of process expertise at a particular facility or throughout the company; there are enough process experts, but they are uneasy using contemporary digital or analytical tools; or process experts are unsure of how to work effectively on digital teams (Exhibit 2). process specialists The retirement of long-tenured employees and the dearth of younger job seekers are two factors contributing to the growing scarcity of process expertise that industrial enterprises are experiencing.

Hence, before implementing AI, businesses frequently need to rebuild their expert pipeline, usually through collaborations with universities and internship programs. Manufacturers and outside consultants can be employed to bolster teams while the pipeline is being rebuilt, but "owning" the capabilities is crucial in the long run because it is a source of differentiating value.

Companies should also upgrade the skills of their present process specialists in analytics software and agile methodologies. Specialists are accustomed to using formulae to explain physical processes; they frequently have engineering or other related backgrounds. Thinking along such lines can help create smart data, but it can also breed mistrust for AI-based methods. Building comfort with the method and outcomes can be accomplished by up skilling process experts through a combination of classroom instruction and on-the-job apprenticeship on cross-functional AI teams. With these abilities, process specialists may support digital teams more effectively. For example, they can collaborate with data scientists to help them comprehend the issue, provide smart data, and pressure-test models to make sure they have picked up the proper first principle-based behaviour.

Also, in our experience, up skilling boosts work satisfaction and employee retention.

Figure 2: Industrial companies have varying level of process expertise

Ways of Working

Because of methodological differences, building high-performing teams with cross-functional positions can be difficult. For instance, although colleagues in data science are more accustomed to iterative workflows like agile, operations staff frequently adhere to unidirectional stage-gated procedures, frequently for safety concerns. Our experience demonstrates that iterative, inclusive, and collocated agile teams frequently have the greatest effect when adopting AI. Coaching is therefore necessary for team members who are not familiar with this strategy.

To establish a method of operation and avoid linear techniques that require thoroughly finishing one stage (such as data extraction) before moving on to the next, it might be beneficial to plan out the model development. In order to swiftly construct a completely functional model with the purpose of maturing individual components in subsequent rounds, elements of each stage should instead be finished concurrently. In reality, this typically entails working with less complex algorithms, a smaller list of characteristics, and a fraction of the available sensor data. The group can then choose what to spend money on for the following step.

To agree on the conclusion and prevent scope creep, a discussion of what "done" means should be included in each iteration. Industrial businesses are relying on AI to improve factory operations—to save downtime, schedule maintenance in advance, raise product quality, and so on. Yet, it is difficult to use AI to have an operational effect.

These businesses will need to design their big data so that it includes operational information in order to succeed (such as misbalance or thermodynamic relationships). Moreover, they will need to put up cross-functional data science teams with personnel who can bridge the gap between machine learning methods and process understanding. A meaningful transformation may be accomplished if these components are coupled with an agile working style that promotes incremental improvement and a predisposition to apply results.

Common Challenges

Notwithstanding this essay, artificial intelligence appears to be present everywhere these days. But honestly, while AI is often thought of as a challenging but creative way to treat illnesses and promote social good, it is already being used for anything from creating stronger fantasy football teams to helping people come up with pick-up lines (how you doing?) Yikes.

Whichever use case AI wants to address, it requires data—smart data. After all, artificial intelligence (AI) can only become clever if it is educated with the appropriate knowledge to aid it in continuously learning and ultimately outsmarting us all. Naturally, none of this should be shocking. The performance of many AI programs is shockingly weak. We all seem to have forgotten how important it is to start with the facts—the data that will enable AI to function—in our haste to implement the trendiest, newest, and most talked-about AI solution. Newsflash: Most AI Projects Fail

Businesses undertaking AI initiatives require a solid base of clean, accurate, and organized data—the kind of training data that computers need to accumulate over time.Common Challenges Sinking Most AI Projects.

Absence of clean, accurate, and high-quality data; Data that is outdated or incomplete; Lack of semantically enhanced data, both structured and unstructured

- Data that is inconsistent across formats and places
- Data sufficiency (enough samples);
- Uncertainty regarding the type of data needed.

Insufficient data quality was the second largest barrier to using AI, only behind a lack of internal expertise, according to a recent study by MIT Technology Review. In addition, 85% of AI programs will "not deliver," per Gartner research. The root of the issue is a lack of smart data. "Without data, you can't feed the algorithms. According to remarks made by Michael Conlin, the Chief Data Officer of the Department of Defense, solid, clean data in high numbers, well-tagged, and well-organized is vital.

You must switch from big data to smart data if you want artificial intelligence to correctly direct your business choices and power your machine learning algorithms.

Big data is everywhere and is produced by everything we do. Indeed, it is abundant. Helpful? Not really. Big data may be both organized and unstructured, but it is not smart data until it has been filtered, cleaned, and analyzed. The following traits can be used to define the most prevalent traits of smart data.

Accuracy

Even though it appears obvious, this quality must be emphasized. Although each of us may define and demand accuracy differently, it is simply reliable knowledge that can be

e-ISBN: **978-93-6252-429-4**
IIP Series

utilized to make effective judgments. According to the required data standards of a certain business model, accurate data should be appropriately defined consistently. Yet, accuracy does not occur automatically; human interaction is required to specify these crucial data properties. Accuracy is a fairly subtle term in many situations, thus it must be understood in the context of the specific quality you're using it for. Your goals may be derailed if your data is even slightly inaccurate.

Completeness

How effectively the data set captures all of the data points that are accessible for a particular instance may be used to quantify this data attribute. There shouldn't be any gaps between the data that was supposed to be gathered and the data that was actually obtained in a full data collection. For instance, the data set will be inaccurate if it only includes the patient's most recent checkups in their medical history. For the data to accurately answer your queries, it must depict a whole picture..

Uniqueness

Data that is self-contained and not included in your database's different formats or locations is referred to by this characteristic. In other words, there shouldn't be any records that are identical to one another. Regrettably, a lot of businesses unknowingly produce the same record repeatedly. Lack of a single source of truth might lead to accuracy issues over time, whether it be a tiny adjustment of the naming practice or incorrect labeling. This is why the word "standardization" is frequently used. Organizations may compare data sets in meaningful ways thanks to standardized data. This is essential for entering data, but it's much more crucial for detecting duplicate..

Timeliness

Data is always changing. So, the ability to gather and update in a timely manner is crucial. Based on time, a thorough knowledge of when the data is no longer usable is required. For instance, the instant a financial agreement's provision is established, it must be taken into consideration. It might have costly repercussions if there is a long delay between the time the data is acquired and the time it is used to guide a business decision. Data that is gathered too soon or too late might skew the results of machine learning.

Quality Always Trumps Quantity

Data quality, which serves as the foundation for wise decision-making, is determined by all of these aspects. As more businesses engage in machine learning and artificial intelligence, data scientists must concentrate on overall quality, particularly that of the metadata. Data is described by metadata, and a dearth of this information is one of the main reasons why data is poor. Without a strong metadata basis, you can never train algorithms to be dependable enough to suit your specific requirements.

In addition to the data itself, analytics and deep learning may be severely hampered by significant restrictions on security, privacy, compliance, IP protection, and physical and virtual obstacles. These limitations must be properly taken into account. The organization

e-ISBN: 978-93-6252-429-4
IIP Series

loses nothing if it gathers and cleans the data only to discover that it is unavailable for a variety of reasons. Frequently, actions like data cleaning are required to ensure that no private material is left behind. Sometimes parties exchanging data need to come to agreements, and other times technical effort needs to be done to get the data to places where it can be evaluated.

Although it may not be the most attractive aspect of your company's AI project, organizing, cleansing, and organizing data is unquestionably the most crucial. You can never get accurate and valid findings from your models without that strong data quality basis. Make sure your data have these "smart" qualities.

Smart AI Means Smart Data Prep

Every company wants to put artificial intelligence (AI) to work. Its potential seems limitless. Big business benefits at the snap of a finger.

But then reality hits: the value that AI can deliver isn't easy. Even IBM, a pioneer in the early age of AI (or AI's rebirth if you want to trace it back to its nascence in the 1940s and '50s), has had its struggles with its Watson AI platform, the most widely known one revolving around the failure of the technology in the healthcare sector on efforts to improve cancer care. There seems to be fairly broad consensus on that point:

- According to a recent IDC poll, the majority of firms experience some failure in their AI initiatives, with a quarter claiming failure rates of up to 50%. The most significant causes of failure were shown to be inadequately skilled personnel and excessive expectations.
- According to Forrester Research, one of the major obstacles facing AI projects is data quality since there is typically a lack of knowledge on the types of data that machine-learning models require and the best ways to prepare such data.
- According to a poll Gartner conducted late last year, AI is now the technology that CIOs most frequently discuss, although VP and analyst Andy Rowsell-Jones warns that this may be due to "irrational enthusiasm." According to Gartner's research on AI and ML Development Strategies.

Arvind Krishna, senior vice president of IBM, stated during the Wall Street Journal's Future of Things Festival that gathering and preparing data accounts for around 80% of the labor involved in an AI project. He claimed that some businesses are just unprepared for the expense and labour involved with it.

"In the general world of IT, around 50% of projects are either delayed, over budget, or are stopped. I'm going to venture a bet that AI is not too different.

Content Expertise for Data Quality

There is a substantial market opportunity, according to Singhal. Amazon Turk, Appen, Figure Eight, and Lionbridge are some of the other companies in the same market as Innodata. For 25 years, Innodata has worked to annotate unstructured information in a range of industries. The company also employs subject matter experts (lawyers, pharmacists, etc.)

to work on projects in the healthcare, pharmaceutical, financial services, and B2B publishing industries.

"You are going through the lifecycle of comprehending and annotating the material while you are developing digital products for your clients," he claims. "To succeed in developing AI applications, you need the competence."

He contends that when businesses use data prep providers who employ the crowd sourcing approach to complete the work, they may not always have that knowledge. To limit risks from incorrectly labeled data, this strategy depends on businesses having their own strict process processes and quality controls in place. "We don't employ the masses."

Teaching the Machine

For precise predictions, a strong ontology and a large amount of training data are needed. According to Singhal, a computer and an algorithm must be trained to comprehend the content and context.

Companies require managed service AI apps that continually monitor machine input in order to develop and deploy genuine AI solutions.

"It is fixing it. Giving the machines this retracted feedback loop enables you to enhance the machine learning model, the author claims. The ability to automate many of these procedures will take years, and it all begins with having excellent, high-quality annotated ground-truth data.

He asserts that there is no universal "workbench" annotation tool. Annotating an SCC (Special Conditions of Contract) legal document requires entirely different tools than annotating a picture. For instance, Innodata had to deal with video footage because one of its customers needed to annotate a lot of license plates. The business had to construct a workstation to accommodate 3000 photos being taken simultaneously. Its developers had to create a workbench to handle high scalability and quick picture loading because of this.

Innodata is chasing the market for complicated document annotation for jobs like pharmacy co-vigilance, which involves watching the effects of prescription pharmaceuticals after they have been given permission to be used. It assists clients that want contract metadata extraction in the financial services industry. It is using machine learning algorithms to look at healthcare data for life insurance.

According to Singhal, "We are also working hard in the reg-tech arena." "We have legal professionals labeling that stuff while looking at other sorts of rules, such as FINRA and FCC 30. They call for advanced knowledge and, quite frankly, better "ground-truth" data that might be used in actual production use cases.

How Smart Data Helps

While there are a few different definitions of "smart data," it is typically understood to be data that has been sorted and prepared at the time of collection so that it is ready and optimized for analytics at the maximum level of quality and speed.

e-ISBN: 978-93-6252-429-4
IIP Series

According to FedTech, Donna Roy, executive director of the Information Sharing and Services Office of the U.S. Department of Homeland Security, claimed at a recent conference that "her teams spend roughly 80% of their time just searching, ingesting, and getting data ready for analysis." Roy thinks that by eliminating the slack in the system, agencies will be able to work more quickly and intelligently.

According to FedTech, Roy described smart data as "data that is independent of software, applications, devices, or networks but still is actionable." Moreover, the data is self-reflective and self-descriptive. It has a specific context and implications. The context of this data is added closer to the original source of the data.

According to Wired in the article Big data, quick data, smart data, "smart data" refers to information that really makes sense. It's the distinction between viewing a long list of figures corresponding to weekly sales and recognizing the peaks and troughs in sales volume over time. Algorithms transform useless data into practical knowledge. Smart data is data from which clever algorithms have extracted signals and patterns..”

Traditional analytics involves gathering, cleaning, and processing data on a regular basis, such as daily or monthly. Because of this approach, the findings are frequently outdated when the data is taken into account. On the other hand, smart data is accessible and converted for analytics at the moment of collection, which helps reduce the time lag associated with data preparation.

What does this entail in terms of business? Smart data primarily assists businesses in extracting pertinent information from the vast amounts of data that are being poured over them. In today's digital business environment, having prior knowledge of what your data is saying is really helpful. A variety of tasks, including patient care and healthcare monitoring, can benefit greatly from the use of smart data.

Consider the issue stated in part one of this blog by Bill Gillis, CIO of the Beth Israel Deaconess Care Group in Boston. His company sought to use claims data to get deeper understanding of patient health, but such information is often not available for study for 90 days following the incident that brought the patient to the healthcare facility in the first place. It certainly isn't enough time for a meaningful response, making the data fairy meaningless. The organization would suddenly have a wealth of fresh knowledge at its disposal that it might use to assist patients if the data could be made available sooner.

Strategic Tactics

While developing a smart data strategy for your business, take into account the following factors:

Think about the data source. Finding the sources that produce the most recent and pertinent data is crucial since not all data sources are created equal. As an illustration, several modern network monitoring solutions make use of unstructured machine data (log files, SNMP, etc.) that is eventually stored and indexed for analysis. There are some drawbacks to the strategy. As a consequence, you have a lot of data to filter through and only gather data that can be documented, which might leave you with blind spots. Second, the procedure results in outdated data. Wire data is a stronger smart data bet for network visibility: It provides a whole picture.

the accuracy of the data. Depending on estimations. On average, poor data loses businesses 12% of their income. Given the crucial use cases smart data is being utilized for, from business analytics to operational responsibilities in data security and application performance monitoring, the adage "garbage in, garbage out" actually takes on an exaggerated meaning. The corporate data governance handbook should have a unified and consistent strategy to developing data quality across the firm.

Examine whether organizational changes are necessary. The majority of data analysis efforts are centralized, but with smart data, value begins to accrue quickly after the data is initially evaluated, providing more possibilities to act—and act more quickly—on data closer to the site of collection. It will have consequences.

Accept automation. The requirement for tools that automate data collection and translation will only increase as you attempt to get value from the ever-increasing data quantities arriving from an ever-increasing variety of sources (read: Internet of Things). There is just no other way to manage that fire hose and hope to sift, prioritize, and understand data in a reasonable manner.

Yet, as said, there are several approaches to the issue. Consider smart data applications for managing the performance of networks and apps, which necessitates incrementing the outermost reaches of the network for complete visibility. The expense and difficulty of extending hardware-based methods to cloud-based systems might be prohibitive.

Converting Big Data to Smart Data

Big Data has emerged as one of the most popular terms in business over the past several years. Capturing and analyzing this knowledge would provide businesses better visibility into their customers and their markets than ever before and perhaps even inspire them to predict what may happen in the future. Each year, record volumes of data of all kinds are collected.

One of the many astounding big data numbers is as follows: 204 million emails are sent every minute, 2.5 million pieces of information are uploaded to Twitter, 277,000 tweets are sent, and 216,000 photographs are published on Instagram.

1. **Hard to find:** Dumb data requires us to be aware of the precise location of a certain piece of information in which we are interested. We could need to learn certain internal identifiers that are used in three distinct systems to identify the same individual, or we might need to know a specific component number that acts as the main key in a database or Hadoop cluster. To address this, we employ prefabricated inquiries or basic keyword searches, which only allow us to retrieve material that is already known and do not assist us in posing new queries or revealing new information.

2. **Hard to combine:** Stupid data has a lot of local relevance. It is identifiable and significant within the confines of the particular silo in which it was produced. But,

worthless information outside of that silo cannot be used. A self-increasing integer key that uniquely identifies a client within a CRM software is exceedingly misleading when used in the same context as data from a dozen other business programs. When released into the world, a short text string like "name" used to identify a particular data property within a key-value store like MongoDB may clash with distinct features from other large data stores, databases, or spreadsheets.

3. **Hard to Understand:** There is a lot of local importance to stupid data. Within the limits of the specific silo in which it was created, it is recognisable and noteworthy. Outside of that silo, useless knowledge, however, cannot be applied. When combined with information from a dozen other business systems, a CRM software's self-increasing integer key that uniquely identifies a client is incredibly deceptive. When a key-value store like MongoDB is published into the public, a short text string like "name" used to identify a specific data property may conflict with distinctive features from other big data stores, databases, or spreadsheets.

The significance is ultimately diminished since the majority of the data is challenging to locate, combine, and evaluate with other data. We only employ dumb data for a select few business challenges, especially those with stable and predictable requirements, due to the large time and financial investment required. For instance, we might use a common BI technique to track and increase product sales by region. But, we are unable to apply the same analytical rigor to staffing client initiatives, identifying competitors' strategies, providing helpful customer service, or any of the hundreds of other routine company operations that might benefit from a data-driven strategy.

Big data is extremely stupid when the data is stupid. Hadoop and other big data platforms may now collect data at will in previously unheard-of volumes and types. Yet it makes it much more difficult for us to find, combine, and comprehend the data we need at any particular time.

Data plus Data (Big Data To Smart Data)

The capacity of huge virtual data to be combined with other data sources is a sometimes underappreciated benefit. Via APIs or by just sending performance, Facebook data may be combined with other consumer data, data from those specific wearable devices, or data from other "connected" devices (i.e., internet of things, or IoT). An even more accurate and nuanced understanding of your client's online and offline activities is produced by the integration of these potent data sets, which provides a more complete image of your customers.

The advantages of big data have been under consideration for a while. Yet taking this route is doubtful. The capacity to exploit this knowledge should only become better as technology develops and big data processing becomes more commonplace. By transforming big data into smart data, businesses may continue to learn important lessons for their industry and make decisions on better, more accurate information.

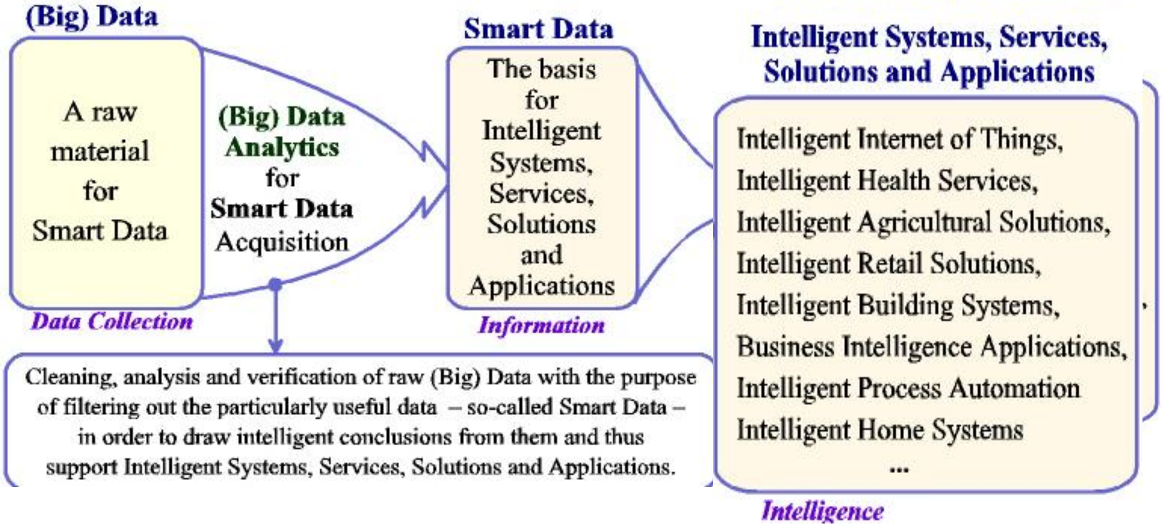


Figure 3: Conceptual model for smart data management.

Conceptual Model for Smart Data Management

This section provides a conceptual framework for smart data management that lays the groundwork for creating a smart service system from a service science perspective. A context-aware service system that can dynamically adapt to a context and help the decision-making process for a particular business circumstance is referred to as a smart service system. Three components make up this perspective: science, management, and engineering. The conceptual paradigm for smart data management, which combines management, scientific, and engineering components, is shown in Figure 1.

Engineering Element

The engineering component, which aims to collect various forms of data, entails the development of new technologies to collect big data and deep data from various data sources and turn them into useable data that can be stored in database management systems. The new data sources and data capture methods have the potential to raise the standard of commercial services and provide brand-new, cutting-edge smart data services. To handle various data sources and support the data collection, supply, and distribution model, this element contains components like data loading, data ingestion, and real-time processing components.

Science Element

The scientific component deals with the structure of service systems and aids the process of service development and the application of competences. It focuses on arranging data into useable information. An interconnected group of concepts from a certain area, relationships between concepts, and connections between concepts and a data source are all considered to be components of a knowledge structure. Our methodology defines ideas in terms of several knowledge elements, such as know-what, know-how, know-why, know-where, know-when, know-who, and know-with. Data sources are linked to pertinent ideas and linkages between concepts are established with the aid of the data analytics and data organization components.

Management Element

The management component, which strives to convert helpful information into actionable insights, addresses strategies and tactics for enhancing smart data-related services through efficient management. Control, discovery, collaboration, learning, and decision support based on actionable insights are the main goals of this component. Based on a specific business scenario, smart data as a service offers a service to a decision-maker. The business decision component is assisted in determining the context of the relevant business circumstance by the context recognizing and context reasoning components. A stakeholder (know-who) conducts actions (know-how) on objects (know-what) at a certain time (know-when) at a place (know-where) because of a contract (know-with) to be consistent with a business rule (know-why), according to the definition of a context.

III. CONCLUSION

The capacity of huge virtual data to be combined with other data sources is a sometimes underappreciated benefit. Via APIs or by just sending performance, Facebook data may be combined with other consumer data, data from those specific wearable devices, or data from other "connected" devices (i.e., internet of things, or IoT). An even more accurate and nuanced understanding of your client's online and offline activities is produced by the integration of these potent data sets, which provides a more complete image of your customers.

The advantages of big data have been under consideration for a while. Yet taking this route is doubtful. With the development of technology and increased processing proficiency, the capacity to manipulate this knowledge should only increase.

This article suggests a conceptual framework for creating a smart service system that can manage and provide smart data as a service from the viewpoint of service science. This research, according to the consensus, is one of the first to emphasize supporting smart data management from a service science standpoint. The suggested model lays a solid platform for change management on smart data as well as for organizational adaptation on company structure and systems to support smart data, which is one of the practical consequences of our study. The necessity for managerial, organizational, and technical adjustments in line with the management, scientific, and engineering components of the model is actually made easier by smart data.

The management shifts put more of an emphasis on creating business plans to provide smart services that are context-aware. The organizational adjustments take into account how important leadership, organizational culture, and organizational structure are to smart data management. The development of new technologies has highlighted the demand for automated solutions for the gathering and transformation of deep and large data for smart data capture. In conclusion, turning businesses that struggle with data into data-driven businesses for smart solutions is the key benefit of smart data. The presented model can serve as a starting point for research on smart data management and use in business, which is one research implication of our study. Based on our prior work in the context of service science, we are now creating the foundation for smart data management.

REFERENCES

[1] Z. Liu, R. Kettimuthu, I. Foster, P.H. Beckman, Toward a smart data transfer node, Future Generation Computer Systems 89: 10-18 (2018)

[2] C. Tan, Y. Sun, G. Li, G. Jiang, D. Chen, H. Liu, Research on gesture recognition of smart data fusion features in the IoT, Neural Computing and Applications, 1-13 (2019)

[3] A. Lenk, L. Bonorden, A. Hellmanns, N. Roedder, S. Jaehnichen, Towards a taxonomy of standards in smart data, in IEEE International Conference on Big Data (Big Data), IEEE, 1749-1754 (2015)

[4] M.H. Alsharif, A.H. Kelechi, K. Yahya, S.A. Chaudhry, Machine learning algorithms for smart data analysis in internet of things environment: taxonomies and research trends, Symmetry 12(1), 88 (2020)

[5] 5 J. Shim, R. Taylor, Purchase-Based Analytics and Big Data for Actionable Insights, IT Professional 21(5): 48-56 (2019)

[6] S. Gupta, A. Leszkiewicz, V. Kumar, T. Bijmolt, D. Potapov, Digital analytics: Modeling for insights and new methods, Journal of Interactive Marketing 51: 26-43 (2020)

[7] Thang Le Dinh1,* and Nguyen AnhKhoa Dam1 1 Business School, Université du Québec à Trois-Rivières, C.P 500, Trois-Rivières, Québec, G9A5H7, Canada

[8] .john dix ,netscout –Aug-13-2018

[9] click workers

[10] *Jay Agarwal, Sean Buckley, Robert Samek, KenSomers, Michel Van Hoey, Joris Van Niel, and Ian Wells, representing views from McKinsey'sGlobal Energy & Materials Practice.2021.*

[11] Tiwari, S.K., Neogi, S.G., Mishra, A. (2023). Design and Implementation of Security Enhancement for Trusted Cloud Computing. In: Kumar, A., Mozar, S., Haase, J. (eds) Advances in Cognitive Science and Communications. ICCCE 2022. Cognitive Science and Technology. Springer, Singapore.

[12] Tiwari,S.K. Sao,N., Kushwaha, A.,Book Title "Fundamental of Machine Learning",lambert Publication, ISBN: 9789391265335, Oct. 2022