Review

A Framework for Implementing Data-Driven Business Operations

Junichi MORI*
Kyohei ISOZAKI
Sachi ISHITANI
Yuta SATOMI
Kuniharu YAMADA
Hiroki ARIHARA
Yasushi IWATA
Yuto HIRATA
Seiji HIGASHI
Masashi YAMAMOTO
Yohei NISHI

Abstract

To deal with the business environment changes surrounding the steel industry, it is important to make objective decisions based on data, rather than take unitary approaches. To effectively implement data-driven business operations, skills and procedures for utilizing data need to be established. In this paper, we describe the framework for implementing the data-driven business operations.

1. Introduction

In the steel industry, there are growing expectations for the improvement and reform of business operation and production processes by utilizing large amounts of accumulated data in order to respond to generational change and to improve labor productivity, in addition to strengthening international competitiveness in terms of cost and quality. Data-driven business operations are those that do not rely solely on subjective judgments based on experience and intuition for improvement and reform, but rather on objective analysis and decision-making based on data in a speedy and sustainable manner.

- Objectivity: The state of making objective judgment based on data
- Rapidity: Data are available for utilization for rapid business judgement.
- Sustainability: Data quality, data utilization model, and the business operations are in a state of sustainable improvement.

The flow of human intellectual work based on data proceeds through the process of "observing" data, "recognizing" the current state, and "predicting" the state to perform appropriate "control" (decision-making). In this paper, the data-driven mechanism that replaces or supports these intellectual tasks is called the data utilization model, and is classified into the following four categories according to the scope of their utilization (**Fig. 1**).

(1) "Visualization Model" which visualizes gathered and accumulated data
(2) "Recognition Model" which judges the abnormality/normality of present Key Performance Indicator (KPI) and/or recognizes the level of KIP
(3) "Prediction Model" which predicts future KPI
(4) "Optimization Model" which determines the optimized action to improve future KPI

In the data driven business operations, it is necessary to productize the data utilization model by using IT technology, and incorporate it into business operations. To this end, in addition to determining the life cycles of the data utilization model and the input data, it is necessary to systemize the products for deployment to actual operations, and to establish the maintenance and operation cycle.

In the data life cycle, in order to establish data-driven business operations, high quality data must be prepared. As the proverb "Garbage in, Garbage out" indicates, for appropriate decision-making, high quality data are required. Therefore, by establishing data management, data can be accumulated, and provided under an appropriate governance. However, even though data are accumulated carefully, paying attention to the governance, since it is difficult to assume various applications, problems frequently appear at the stage when it's going to be used. Therefore, it is necessary to provide a method which improves quality quickly by feedback when problems arise. In the model life cycle, data are firstly prepared and comprehended at the planning stage, then after modelling and the evaluation of the model, the developed model is applied to actual business

* Ph.D., Chief Researcher, Research Section-I, Production Management Research Dept., Intelligent Algorithm Research Center, Process Research Laboratories
20-1 Shintomi, Futtsu City, Chiba Pref. 293-8511

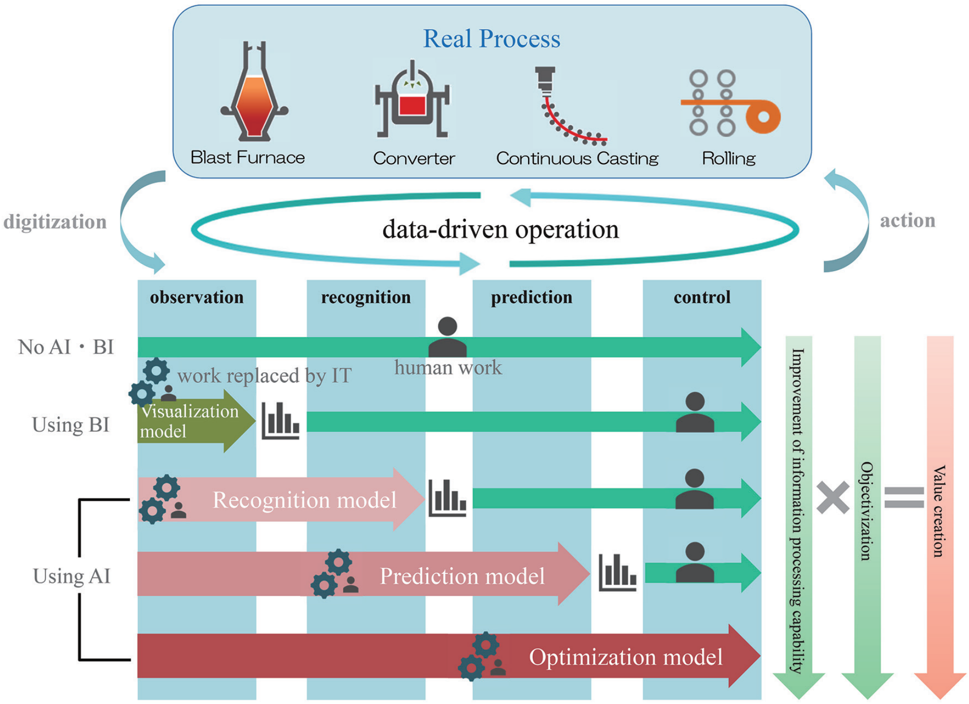


Fig. 1 Data utilization model and the scope of application

operations. When problems occur in the model, it becomes necessary to feedback the problem, and to maintain and improve the quality of the model through relearning of the model and/or the restudy on the model structure.

In the systemization for deployment to actual operations, as well as the usual tasks in system development, the following tasks are necessary: business designs and their operation incorporated in the model, ensuring the functionality or non-functionality of the model, and furthermore, the automation of the release process to sustainably improve the model quality. Furthermore, in the maintenance and operation, in order to enhance the sustainability of the data-driven business operations, it is necessary to monitor the KPI in business and the model data quality, and in the case of problems, to analyze the cause and feedback to the model/data life cycle (**Fig. 2**).

To date, Nippon Steel Corporation has continued to promote the data-driven business operations through the development of several data utilization models, and applications thereof to business operations. As described below, due to various factors in the respective phases, the data-driven business operations were not functioning soundly.

Problems at respective phase and factor

Planning

- Objectives are not appropriately set, and models that do not contribute to the improvement of KIP are developed.
- Unawareness of the possibility of solution by applying similar technologies (in company and/or outside of the company)

Development

- Spend cost (human resources, hour) without reviewing feasibility and/or set objective problem
- Verification items and/or the target values are unclear, and Proof of concept (PoC) is prolonged.

Systemization for deployment to actual operations

- Organization is divided, and the result of the research is not

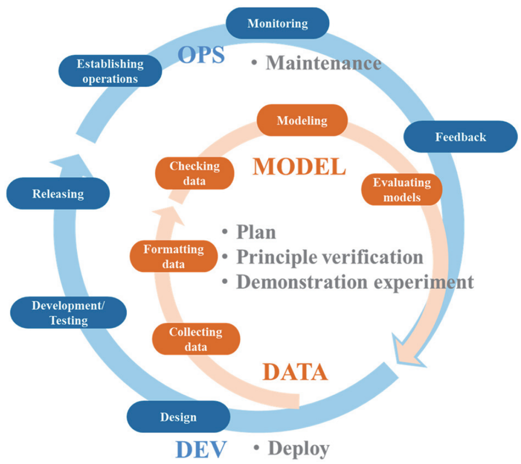


Fig. 2 Process for sustainable improvement of the data-driven business operations

promptly applied to actual operations.

- Insufficiency of the model accuracy recognized at the test operation phase

Maintenance and operation

- The assumption is to continue using it until the end of service life (EOSL), which makes it difficult to keep up with the changes.
  - Lack of response to the change in model (deterioration in performance, change in data)
  - Lack of response to the change in the business environment (sophistication, combination and/or reorganization) and business process
  - Lack of response to the change in IT technology
- No monitoring system, and operation continued without realizing change
- No incentive mechanism to business users, and operations are not

settled

Therefore, Nippon Steel has constructed a framework to realize its target of the improvements of business and production processes from the viewpoint of data-driven operations described at the beginning of this paper by referring to the guidelines summarized by various research organizations and parties.[1–3] The content of such framework is described in this paper. Hereafter, in chapter 2, the skills and the team structure required for the promotion of data-driven business operations are described. In chapter 3, development of the system that supports data-driven business operations is divided into phases, and the content of the tasks to be executed and the requirements for proceeding to the next phase are clarified.

2. Team Structure to Promote Data-driven Business Operation

The skills required for establishing and promoting data-driven business operations consist of the following: "business capability" which enables the extraction and arrangement of subjects based on the comprehension of the subject business, "data science capability" for statistical processing, machine learning and optimization, and "IT capability" for data processing, implementation, and operation (**Fig. 3**).

In each phase for the establishment of data-driven business operations, the balance of the set of the required skills varies.[4] In building the structure, this skills balance must be taken into consideration (**Fig. 4**).

Human resources for data utilization are expected to have all these skills. However, since it is rare for individual human resources to have high level skills in all areas, generally a structure is established by combining personnel that have highly specific skills in their respective fields. Furthermore, even though the structure changes gradually in the respective phases, it is important for the experts having specific skills to mutually deepen their comprehension of skills in other fields in order to carry out smooth communication and not impede the flow

In the conventional development process, the organization is split into the flow of Planning → Principle verification → Demonstration experiment → Systemization for deployment to actual operations → Maintenance and operation, and not only is a long term required from the planning to deployment, but also the organization is structured in such a way that easy rotation of the cycle of the model development is not permitted. In order to rotate the development process speedily without loss of time among organizations, it is necessary to build a structure wherein members responsible for respective roles (data science user, data scientist (citizen/expert), IT engineer, and maintenance engineer) mutually recognize the others' tasks and conduct cooperative work to avoid splitting the process by organization. Further, positioning and the education system for the human resources hierarchically listed corresponding to the knowledge and skills of data science (data science user, data scientist (citizen/expert)) are described in the separate paper No. 21 in this report.

3. Building Products Phase

The building products phase adopts a trial-and-error type approach, which, depending on the situation, allows for retrogression and/or retrial[5, 6] (**Fig. 5**).

The phase is merely a guideline, and when there are few uncertain elements, the planning and the principle verification are mixed.

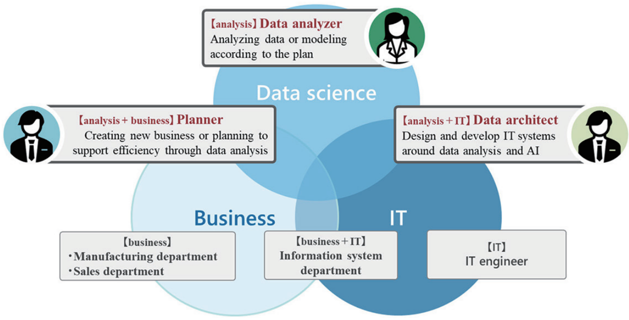


Fig. 3 Skills needed for data-driven business operations

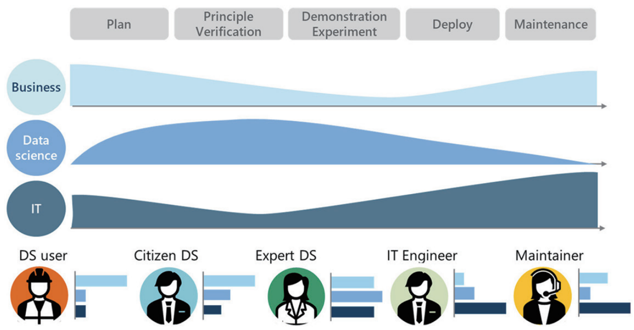


Fig. 4 Skill balance at each phase

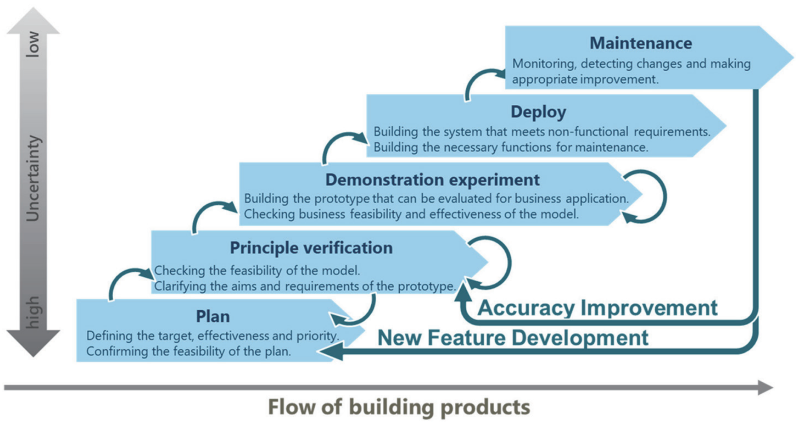


Fig. 5 Step for building products

The phase can then be omitted, and thus various patterns are conceivable. The important point is to rotate each step as speedily as possible, to judge whether "to proceed" or "to retry" at the phase gate set in advance, and to steadily reduce the elements of uncertainty.

The products must be built in such a way as to allow for successive improvement and/or functional expansion even after entering the maintenance and operation phase. To achieve this, in line with the progress along the phases, it is important to establish and optimize the aforementioned life cycles of models and data, and furthermore, conduct successive system improvement (DevOps). Hereafter, the contents of execution in each step, and conditions for advancing to the next phase are clarified.

3.1 Planning

In the planning phase, the objective of the business operation improvement and the approach based on data are clarified, and data are arranged to be shared by the persons concerned. Herein, first, in order to grasp the present situation, business analysis and hearing regarding business are conducted using such tools as a data flow chart and/or business operation flow chart. Then, through analysis of the gap between the present state and the ideal state, the objectives of data utilization are clarified. At the time, it is important to set KPI corresponding to the objective to measure the level of achievement. KPI can be an indicator to quantitatively measure the effect obtained by achieving the objective. The expression of KPI in the form of economic effect will facilitate the formation of consensus among the persons concerned, as well as evaluation. However, it is difficult to define KPI at the early stage of planning. This is because of the lack of concrete decision-making and further actions based on data. Therefore, gradual determination of KPI is required by developing possible actions to be taken through business operation analysis and/or discussions among the persons concerned. To achieve the objective, it is also important to judge whether the AI technology is optimal or not. As the technical subjects to be solved and the technical elements to be verified become clearer through the hearing, the risk of using AI or replacing with AI regarding the subject business is also to be evaluated. For example, since AI can't guarantee 100% accuracy, it should be evaluated what sort of risk will emerge by fully automating the work in the workshop.

Furthermore, the investigation of data at the planning phase contributes to the comprehension of the business, and by clarifying the data quality, whether or not to proceed becomes the judgement standard. For that, in the case of numerical data, an environment is to be established which combines the common business data accumulating foundation and the self Business Intelligence (BI) tool so that the business users can conduct data investigation by themselves. Particularly, if the objective is to construct a machine learning model, it is advisable to scrutinize the influence of absent values and/or abnormal values and the timing at which data can be obtained for data that are candidates for the objective and/or explanatory variables.

In the case of image data, accumulation of data is started at first in some cases. Therefore, as a common foundation, the analysis navigation system which preserves images and data time-sequentially, or the companywide location-free ITV system which gathers, stores, and delivers image data is arranged. In some image processing systems, menu options corresponding to the processing speed and effect/cost are prepared, and thus, by deciding on the utilization of the existing mechanism at the planning phase, labor saving in the subsequent study is considered. Details are explained in paper No. 8 of this report.

In the planning phase, based on the data thus arranged, the feasibility is reviewed, and by comparing the expected effect and the cost, acceptance or rejection is determined. However, since the values of the effect and/or the cost at this stage are just rough estimates, therefore, not being constrained excessively by the largeness or smallness of the value, a few risks may be tolerated.

3.2 Principle verification

In the principle verification phase, validity and feasibility are verified. In particular, it is important to quickly verify the validity of the model of high uncertainty by suppressing time and labor. To this end, in order not to spend time on preparation of the verification environment, the common foundation on which data processing and/or data analysis are carried out is to be arranged in advance so that the repeated processing such as acquisition and/or processing of data is automated to the extent possible. Creation of the verification process like limiting the scope of verification is also effective. Furthermore, the management of experiments also becomes important since the demonstration experiments are conducted on a trial and error basis. Facts about when and under what conditions the demonstration experiment was conducted and what result was yielded are managed, and reserved as the evidence of outcome. Moreover, sometimes in order to secure the reproducibility of verification, the verification environment is virtualized and stored.

In addition, at this time, the arrangement of prerequisites for the demonstration experiment is considered. In the principle verification, since speediness is prioritized, off-line verification is adopted in many cases; therefore, sometimes the comparison and check of the verification with respect to the actual business operations become flawed. Therefore, it is necessary to extract in advance the supplemental verification to compensate for any flawed parts as a prerequisite of the demonstration experiment.

Data to be used in the principle verification should be, as far as possible, the same as those of the actual operations where the system will be deployed. Particularly, the image data are unable to conform to significant verification unless the objectives, environment, and the conditions are in agreement with those of the actual operations where the system is to be deployed (hereafter referred to as "deployed actual operations"). By using the data of the deployed actual operations, problems in the subsequent phases of "accuracy being lower than expected in the actual environment" can be avoided. For example, in the case of studying the image recognition model to read the ID listed on products, if a model constructed based on the image taken indoors is put into practical use outdoors for actual application, considerations on the situation under backlight may become insufficient, and misjudgment may occur. Accordingly, images taken under the actual environment to the extent possible are to be prepared. On the other hand, at this timing, sometimes, preparation of data having the same conditions may fail. At that time, it is important to assume and extract the technical subjects assumed from the difference with the data in the deployed actual operations. Also, depending on the amount of data and/or the algorithms employed, sometimes the processing time is extended, and therefore, an appropriate method meeting the requirements of the business must be studied. Particularly, in the case of deep learning wherein the contribution of the hardware performance is high, the processing time should be estimated in view of the specification of the common foundation on which the model is implemented.

The evaluation of the model is to be reported to business users with not only the general accuracy index values such as of error and/or correct interpretation ratios, but also together with the easy to understand index values showing simply the influence on the business. Predetermined KPI is used therein as an index. Validity is evaluated and confirmed after confirming that the index values fulfil the targeted level, and a result which satisfies business users is yielded. Furthermore, a comprehensive evaluation should be conducted which includes the individual evaluation of whether or not the data processing from the data accumulation/acquisition to the output of the result was smooth without any arbitrariness, whether or not there is any prospect of automation, and the processing time shown as a guideline.

3.3 Demonstration experiment

A mechanism that users can employ in the actual business operations is to be prepared, and by users, the evaluation of the validity of the measures is to be conducted, and the establishment of the business operation system is to be confirmed. The demonstration experiment aims at the repeated execution of the “verification” of the business value of the data utilization model, the “improvement” of the data utilization model after the reception of the feedback from business users, and the “expansion” of the verification scope. For this, it is important to provide the business users with the prototype as soon as possible without costs (term × labor) being incurred, and the demonstration experiment is to be conducted after determining the core function of the prototype from the following viewpoint, and establishing the evaluation method.

- Linkage function with outside data and applications is not built in
- Time is not to be spent on the development of the function of enhancing the usability which does not have a direct influence on the verification of value
- Non-functional requirements not related to the evaluation of the validity of the measures are to be disregarded, and the enhancement of accuracy of the functional requirements is to be targeted.

In the early stage of this step, by limiting the scope of the verification which evaluates the functionality (variety of steel products, division, and so forth) and the subject data, verification of the data utilization value, the feedback from the business users, and the model improvement are targeted. It is also important that the persons concerned are well informed of the approach that starts with the partial optimization, and gradually work toward the entire optimization. Furthermore, not pursuing the real time execution of the prototype synchronized to the data renewal of the existing system, but by tolerating the periodical execution of the task scheduler and/or the Robotic Process Automation (RPA) tool, or human intervention, the business value verification is to be started early. Only when the real time linkage is required for the business value verification, the parallel environment of the actual operations is to be prepared by the mirror environment and/or the database replication.

For the visualization to users, the implementation method using the standard function of the BI tool is considered primarily. In the case when realization with the standard function is difficult, whether manual implementation is tolerated must be studied. When the incorporation of the Graphical User Interface (GUI) for the verification of the business value in addition to the sufficient execution of manual verification is indispensable, the GUI is to be self-manufactured. In the case of the self-manufacturing of GUI, not only are there costs (term and labor) incurred, but there is also a risk of prolonging the development cycle of the prototype.

Then in Nippon Steel, judgement of the applicability to business is speedily reached by arranging the template of the self AI prototype and the guideline, and by making citizen data scientists advance in person from the development of the self AI model to the demonstration experiment. Among the citizen data scientists, there are persons who have skills in developing the self AI model by using Python, but do not have enough experience of developing prototype application and/or demonstration experiment experience. For this, in order to reinforce the skills with the guideline and the development supporting tool based on the current skills base, Nippon Steel provides personnel with the program template (template code) and the development manual as the development support tool for self AI prototype conforming to the process. The guidelines define an AI prototype development process so that citizen data scientists can conduct the demonstration experiments themselves.

On this occasion, for the guideline, the self AI prototype development support tool, and the template code, exposure to many in-company cases is required. By gathering the examples of the in-company self AI prototype development, and by investigating the examples which frequently appear in the demonstration experiment, the functional requirements of the self AI prototype, business requirements and the subjects in development, and by exposing them to the functions and patterns which frequently appear in the examples of the in-company self AI prototype development, the guideline is made to be comprehensively applicable in the company. The self AI prototype development support tool and the template code employ the framework intended for the development of the Web application of Python which is easily recognized by citizen data scientists. However, even if the sample code is taken as the base, for the citizen data scientists who are not specialized in programming, the amount of programming becomes large, and if the difficulty of the program is high, they will feel reluctance, and will not use the sample code. Then, as a device to facilitate the development, by preparing the parts of functions and displays which appear frequently in the in-company self AI prototype examples, and by completing the self AI prototype by simply filling in the blanks, thus the amount of the programming of the citizen data scientists is reduced, avoiding the emergence of any reluctance to use it.

By using the template code as the base, citizen data scientists can develop the self AI prototype by themselves, and keep proceeding to the demonstration experiment. The construction term of a demonstration experiment longer than six months is planned to be shortened to less than half, to 2–3 months.

In order for the operators working on site in steelworks to evaluate the application to business operations, the real time application test on the production site is desirable. Since the image processing differs depending on the objectives of detection and/or control, and the required users’ interface is also different, each interface needs to be self-manufactured. Therefore, the display guidance is designed to be constructed within a short period of time and at low cost by building up the voice alarm function, revolving warning lamp control, and the graphical displaying function on the image processing foundation. The result of the judgement of processing including images is registered with the aforementioned navigation system and analyzed, thus facilitating the performance evaluation.

Through the above attempt, within the predetermined verification scope, the guidance for the decision-making and actions linked to the improvement of KPI is completed, and the effect of KPI improvement even outside the scope of verification is in sight. Possibly, when a sufficient improvement effect is obtained only within the scope of verification, the flow of building products may proceed to the subsequent phase. In the case where the judgement is impossible, uncertainty should be lowered to the extent possible by increasing the case of verification and/or expanding the scope of verification. Furthermore, it is also important to decide whether or not to proceed considering the cost of deploying the system to actual applications and the cost of operation vs. KPI improvement effect obtained.

3.4 Systemization for deployment to actual operation

Since the products which include the data utilization model exert their function through inputting or outputting with users and/or other systems, it is essential to incorporate the input/output function in the products as a part of the business system and business process. Herein, the essence of each of (1) Business design, (2) System design, (3) Data gathering and (4) Operation design is described.

In business design, how the results of the visualization of the output of the data utilization model, image recognition, and/or the numerical estimation are utilized in business are clarified, and the non-functional requirements such as functional requirement, targeted speed and accuracy, and so forth required for the business design are reconfirmed. In the case of introducing AI in particular, based on the mechanism and/or the restriction of the AI, the risk of using the AI is to be evaluated. As AI is unable to guarantee complete accuracy, evaluation should be appropriately implemented by understanding the risk of the variations of accuracy given to business. Furthermore, the business flow for the case of insufficient accuracy actually taking place should also be studied in advance. For instance, for the case wherein the image recognition model fails to read the ID listed on products, the work of humans instead of AI visually checking the ID and inputting it into the system should also be taken into consideration in advance. Realistic business operation flow is also to be studied from the viewpoints of occurrence frequency of problems like this and/or working load. Furthermore, effective result-presentation techniques in the interface with users should be considered, and the image structure should be reviewed based on the feedback from the users obtained in the principle verification and the demonstration experiment phases.

In system design, the system structure is studied based on the content of the business design. The features of the system incorporated with AI is that the life cycle of the entire system differs from that of the AI. The accuracy of the AI model is deteriorated by the change in data characteristics, and appropriate relearning or revision of algorithms becomes necessary. Therefore, from the viewpoints of system maintainability and reusability, the system is to be designed to allow the AI operating part to be loosely coupled. By studying the matching of the error notification rule during the system processing to this system border, the problems of the business system and the AI are separated and clarified. The system implementation environment is selected with the following in mind: the implementation environment which follows the guideline of AIRON-EDGE™ (described in paper No. 18 in this report) is taken as a candidate, the network rapport, speed restriction, and AI model implementation condition. In the system which handles mainly data utilization, as compared with the ordinary business system, in many cases, the data flow amount becomes huge, and therefore, a design which takes into account the communication charge and storage also becomes necessary. In the case that the clients' device such as a mobile phone terminal is utilized, it should be selected after the business requirement (terms of use) and/or the non-functional requirement (speed and/or frequency) have been thoroughly studied, considering the device and/or usage performance.

The purpose of gathering data is to verify whether or not the system exerts, even under the deployed environment, its accuracy as expected with respect to the unknown data which were not included in the demonstration experiment. Verification is implemented repeatedly, and upon necessity, the model is to be improved through relearning of AI. At this time, since the installation condition and the introduction timing of the data-gathering equipment used in the deployed actual operation have already been determined, the data gathering and data accumulating mechanisms corresponding to such equipment are to be studied and constructed.

In the operation design, operation flow accompanying AI is to be embodied. The deterioration in the accuracy of AI due to the change in environment is a concern. For instance, the lenses of the camera installed in a plant may be dusty, and erroneous recognition may occur due to the dust. When such a case is presumed, the condition which regards deterioration of the extrapolation accuracy and/or confirmation frequency are to be determined based on the past verification result. Furthermore, triggering and its execution method upon execution of accuracy correction are to be studied.

3.5 Maintenance and operation

In the maintenance and operation phase, monitoring and improvement are conducted so that utilization of the products is established, and the products continue to exert their intended effect. After the construction of the products, the establishment in business is to be encouraged by not only educating the business users, but also by arranging simultaneously the mechanism linked to the promotion of utilization. In parallel with the maintenance and the operations of the entire system, the model performance during the operation and the data characteristics which affect the model performance are to be monitored at all times. In AI models in particular, since reasoning is established based on the data in the past, large deterioration in accuracy may occur due to the change in environment, variations in operation, secular change of the data-acquiring equipment (sensors, cameras, and so forth), and the variations in the data trend and/or distribution thereby. Therefore, when a change in the trend is detected, the business users should take the initiative in making efforts for the system to exert its effect sustainably by factor-analysis according to the predetermined procedure, study on countermeasures, and the operation of relearning like the model parameter adjustment (**Fig. 6**). In the case that the model accuracy is not improved, the analysis phase is to be repeated in the presence of data scientists. The measurement of the effect is calculated based on the data gathered by the system; however, in the case of other effects exerted by the actions outside the system, a hearing for the business users is also effective.

Regarding the scale to measure the performance of the AI model, there are indexes (RMSE, MAE, and so forth) based on the estimation errors in the case of the regression model which estimates continuous values, and indexes based on the accuracy ratio (accuracy ratio, reproducibility ratio, adaptability ratio, F value, AUC, Log Loss, and so forth) in the case of the classification model which estimates the category value. It is important to set in advance appropriate evaluation indexes corresponding to KPI and/or the model utilization method. For example, in the case where reliable detection of the defects which will become obvious in the subsequent process is desired, since it is important to correctly predict a defect as a defect, the reproducibility ratio should be selected as one of the evaluation

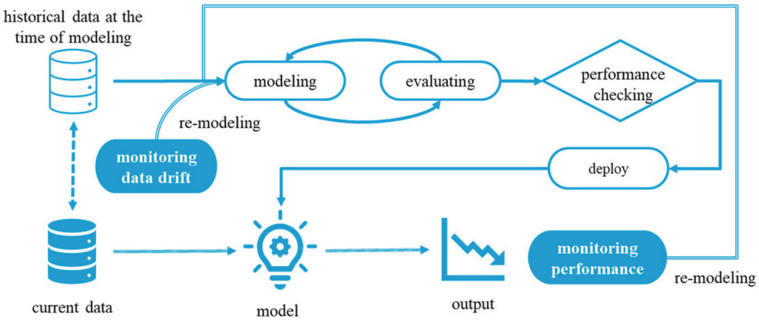


Fig. 6 Life cycle of the data utilization model

indexes.

As a method to detect the change in the characteristics of the input data which influences the model performance, the abnormality detection which detects the input/output data behaves differently from the past data. For example, in the case that the reliability of the measured values is low like in the case of a defective thermometer, relearning after deleting the data is necessary. In the meantime, in the case that abnormal data are detected due to process factors such as facility problems, improvement of the process itself is necessary. Furthermore, even though the individual data are not of abnormal value, sometimes the model estimation accuracy deteriorates due the shift of the statistical distribution of the data from that displayed at the time of model training. It is desirable to detect the shift between the statistical distributions before the deterioration of the estimation accuracy becomes obvious, and to take countermeasures like relearning using the most recent data. As detection methods, methods based on the dimensionality reduction of the amount of characteristic and/or the testing of statistical hypotheses are proposed.

Furthermore, in the system installed with an image-system AI model, sometimes adjustment due to the renewal of the photographic equipment and/or the change in the surrounding environment becomes necessary. In such a case, the image processing foundation copies all recorded image data on the development environment so that the reproducibility test can be conducted. As the off-line readjustment is possible, readjustment is managed without having to station image processing personnel on site. In the meantime, in order to expedite the expansion of the image processing, image processing personnel are allocated on site. Image-processing engineers are being trained, and installation at each steelworks is being positively promoted.

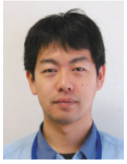
An IT system always requires alignment with version updates of the hardware, OS, middleware, and the library. Combinations of these are restricted. Pluralities of components are often renewed simultaneously, and considerations should be paid to the range affected. Since the behaviors of the AI model before and after version updates do not completely agree with each other due to the change in the internal structure of the library, numerical calculation errors, and the influence of random numbers, it is important to show users the comparison result of the output before and after the version updates, and to confirm that there's no problem in the business operation, and then execute renewal.

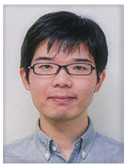
4. Conclusion

In order to correspond to the changes in the business environments which surround the steel industry represented by changes in the steel demand and supply environment, climate change, and decrease in the working population, it is important to conduct data-driven objective analysis and data-driven decision-making rapidly and sustainably. As a framework for implementing data-driven business operations, this paper describes the skills and team structure required to promote data-driven business operations, as well as the steps for building supporting products. This paper explained that skills include business capability, data science capability, and IT capability, and that it is important to form a team structure taking into consideration the balance of each skill. Furthermore, in the products construction step, the objectives, tasks, and methods of each phase (planning, principle verification, demonstration experiment, systemization for deployment to actual operation, and maintenance and operation) are described, as well as the phase gate that determines whether or not to proceed to the next phase. This framework was started with application to actual business in Nippon Steel's system development division, and has been established as the practice guideline of the data-driven business. It is used in the entire company workshop in which 200 staff members participate annually, and thus contributes to the development of DX human resources.

References

1) National Institute of Advanced Industrial Science and Technology: The third Issue of Machine Learning Quality Management Guideline. August, 2022
https://www.cpsec.aist.go.jp/achievements/aiqm/
2) DAMA International: DAMA-DMBOK2: The second Issue of Data Management Knowledge System Guide. December, 2018
https://www.dama.org/cpages/body-of-knowledge
3) The Japanese Society of Artificial Intelligence: AI Map β2.0. June, 2020
https://www.ai-gakkai.or.jp/resource/aimap/
4) Information Technology Promotion Agency, Japan: Skill Checklist & Task list Overview for Data Scientist. July, 2020
https://www.ipa.go.jp/files/000083733.pdf
5) NS Solutions Corporation: (Technical trend) "DataOps" for Data Utilization in the Era of AI. 2019
https://www.nssol.nipponsteel.com/future/stories/assets/nssol-ss-prircas-174-01.pdf
6) NS Solutions Corporation: Enhance Data Utilization Power with DataOps. February, 2021
https://dcross.impress.co.jp/industry/column/column20201007/index.html


Junichi MORI
Ph.D., Chief Researcher
Research Section-I
Production Management Research Dept.
Intelligent Algorithm Research Center
Process Research Laboratories
20-1 Shintomi, Futtsu City, Chiba Pref. 293-8511


Yasushi IWATA
Senior Researcher
Intelligence Research Department
System Research & Development Center
Technology Bureau
NS Solutions Corporation


Kyohei ISOZAKI
Senior Researcher
Intelligence Research Department
System Research & Development Center
Technology Bureau
NS Solutions Corporation


Yuto HIRATA
Chief Manager, Head of Section
AI Solution Section
ICT System Planning Dept.-II
Information & Communication Technology Div.


Sachi ISHITANI
Manager
AI Solution Section
ICT System Planning Dept.-II
Information & Communication Technology Div.


Seiji HIGASHI
Manager
Control Development Section
Systems & Control Technology Dept.
Systems & Control Engineering Div.
Plant Engineering and Facility Management Center


Yuta SATOMI
Researcher
Research Section-I
Production Management Research Dept.
Intelligent Algorithm Research Center
Process Research Laboratories


Masashi YAMAMOTO
Director, Assistant Chief Researcher
R&D Strategy Management Department
System Research & Development Center
Technology Bureau
NS Solutions Corporation


Kuniharu YAMADA
Senior Researcher
Intelligence Research Department
System Research & Development Center
Technology Bureau
NS Solutions Corporation


Yohei NISHI
Senior Researcher
Intelligence Research Department
System Research & Development Center
Technology Bureau
NS Solutions Corporation


Hiroki ARIHARA
Researcher
Intelligence Research Department
System Research & Development Center
Technology Bureau
NS Solutions Corporation