ICS 35.240.01

VDI/VDE-RICHTLINIEN

Oktober 2022
October 2022

VEREIN DEUTSCHER INGENIEURE

VERBAND DER ELEKTROTECHNIK ELEKTRONIK INFORMATIONSTECHNIK

Implementierung und Betrieb von Big-Data-Anwendungen in der produzierenden Industrie

Analyseverfahrensklassen

Implementation and operation of big data applications in the manufacturing industry

Analysis process classes

VDI/VDE 3714

Blatt 4 / Part 4

Ausgabe deutsch/englisch
Issue German/English

Die deutsche Version dieser Richtlinie ist verbindlich.

The German version of this standard shall be taken as authoritative. No guarantee can be given with respect to the English translation.

Inhalt Seite

Contents Page

VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik (GMA)
Fachbereich Digitale Transformation

VDI-Handbuch Informationstechnik, Band 1: Angewandte Informationstechnik
VDI/VDE-Handbuch Automatisierungstechnik

Inhalt Seite

Contents Page

Vorbemerkung

Der Inhalt dieser Richtlinie ist entstanden unter Beachtung der Vorgaben und Empfehlungen der Richtlinie VDI 1000.

Allen, die ehrenamtlich an der Erarbeitung dieser Richtlinie mitgewirkt haben, sei gedankt.

Eine Liste der aktuell verfügbaren und in Bearbeitung befindlichen Blätter dieser Richtlinienreihe sowie gegebenenfalls zusätzliche Informationen sind im Internet abrufbar unter www.vdi.de/3714.

Einleitung

Der Fachausschuss „Big Data“ der VDI/VDE-Gesellschaft für Mess- und Automatisierungstechnik hat sich mit der Erstellung dieser Richtlinie der Aufgabe angenommen, den ökonomischen und ökologischen Nutzen von Big Data aufzuzeigen, den Wissenstransfer über verschiedene Industrien und Branchen hinweg zu verbessern und die Implementierung und den Betrieb von Big-Data-Anwendungen in der produzierenden Industrie voranzutreiben und zu vereinheitlichen.

Die Richtlinienreihe soll eine Orientierung über erforderliche Maßnahmen zur Big-Data-Analyse geben und aufzeigen, welche Methoden für eine zielführende Arbeit geeignet sind und welche Einschränkungen und Hindernisse bestehen. Praktikern und Praktikerinnen sollen Hinweise gegeben werden, welche Methoden und Betrachtungen für den Erfolg eines Big-Data-Projekts hinsichtlich des Einsatzes und des nachhaltigen Betriebs notwendig sind.

Die Richtlinienreihe VDI/VDE 3714 umfasst die Blätter:

Blatt 1 Durchführung von Big-Data-Projekten
Blatt 2 Datenqualität
Blatt 3 Datenbewirtschaftung
Blatt 4 Analyseverfahrensklassen
Blatt 5 Modellierungsverfahren
Blatt 6 Validierung von Modellen
Blatt 7 Online-Anwendung von datengetriebenen Modellen

Die Richtlinienreihe VDI/VDE 3714 ist im Fachausschuss 7.24 „Big Data“ des Fachbereichs 7

Preliminary note

The content of this standard has been developed in strict accordance with the requirements and recommendations of the standard VDI 1000.

We wish to express our gratitude to all honorary contributors to this standard.

A catalogue of all available parts of this series of standards and those in preparation as well as further information, if applicable, can be accessed on the internet at www.vdi.de/3714.

Introduction

The “Big Data” Technical Committee of the VDI/VDE Society Measurement and Automatic Control has taken on the task of creating this standard to demonstrate the economic and ecological benefits of big data, to improve the transfer of knowledge across different industries and sectors, and to promote and standardize the implementation and operation of big data applications in the manufacturing industry.

The series of standards is intended to provide orientation on the measures required for big data analysis and to show which methods are suitable for target-oriented work and which limitations and obstacles exist. The practitioner should be given advice on which methods and considerations are necessary for the success of a big data project in terms of implementation and sustainable operation.

The series of standards VDI/VDE 3714 comprises the parts:

Part 1 Implementation of Big Data projects
Part 2 Data quality
Part 3 Data management
Part 4 Analysis process classes
Part 5 Modelling procedures
Part 6 Validation of models
Part 7 Online application of data-driven models

The series of standards VDI/VDE 3714 is published in the Technical Committee 7.24 “Big Data” of the

„Digitale Transformation“ der VDI/VDE-Gesellschaft für Mess- und Automatisierungstechnik (GMA) entstanden. Damit stellen die Produktion sowie die Mess- und Automatisierungstechnik die Schwerpunkte dar. In den Produktionsprozessen werden beispielsweise für Steuerungs- und Regelungsaufgaben oder für die Qualitätssicherung große Datenmengen erhoben, die mittels Datenanalyse für weitere Prozess- und Geschäftsverbesserungen genutzt werden können. Die Richtlinienreihe gibt eine generelle Orientierung sowie Hinweise auf potenzielle Schwierigkeiten und Hürden bei der Durchführung von Big-Data-Anwendungen von der Entwicklung über die Inbetriebnahme bis zum nachhaltigen Betrieb.

Technical Division 7 “Digital Transformation” of the VDI/VDE Society Measurement and Automatic Control (GMA). Thus, production as well as measurement and automation technology represent the focal points. In production processes, for example, large amounts of data are collected for control and regulation tasks or for quality assurance, which can be used for further process and business improvements by means of data analysis. The series of standards provides a general orientation as well as indications of potential difficulties and hurdles in the implementation of big data applications, from development through commissioning to sustainable operation.

Benachbart zu Big Data finden sich Themen wie das Internet der Dinge (IoT), Vernetzung von Geräten (Smart Devices) oder die zunehmende „Rechnerallgegenwart“ (Ubiquitous Computing) sowie Begriffe wie Business Intelligence, Data Analytics, Advanced Analytics, Data Mining, Smart Data und Data-Warehouse-Systeme, die generell die Nutzung von Daten adressieren.

Adjacent to big data are topics such as the Internet of things (IoT), the networking of devices (smart devices), or the increasing “computer omnipresence” (ubiquitous computing), as well as terms such as business intelligence, data analytics, advanced analytics, data mining, smart data, and data warehouse systems that generally address the use of data.

Die Richtlinienreihe geht von einer generellen Verfügbarkeit aller benötigten Daten aus. Bezüglich der Datenmenge, ihrer Struktur und Integrität wird keine Annahme getroffen. Zur Diskussion und Charakterisierung der Daten helfen die sogenannten „fünf Vs“, die die einzelnen Dimensionen von Big Data bezeichnen. Die Daten werden charakterisiert durch Umfang (*Volume*), Unterschiedlichkeit (*Variety*) und ihre zeitliche Taktung (*Velocity*). Insbesondere bei industriellen Anwendungen sind die Qualität der Daten (*Validity*) und der unternehmerische Mehrwert (*Value*) relevant.

The series of standards assumes a general availability of all required data. No assumption is made regarding the amount of data, its structure and integrity. For the discussion and characterization of data, the so-called “five Vs”, which denote the individual dimensions of big data, are helpful. The data is characterized by *volume*, *variety*, and *velocity*. The quality of the data (*validity*) and the added business value (*value*) are particularly relevant for industrial applications.

Die vorliegende Richtlinie unterteilt Analyseverfahren in zwölf Klassen, erläutert kurz deren Funktionsweise und stellt exemplarische Anwendungsgebiete dar. Sie soll Big-Data-Interessierten einen leicht verständlichen ersten Einblick in die Thematik und einen Überblick in wesentliche Analyseverfahrensklassen geben. Die verschiedenen Aspekte werden durch Beispiele aus der produzierenden Industrie veranschaulicht.

This standard divides analysis methods into twelve classes, briefly explains how they work, and presents exemplary areas of application. It is intended to provide those interested in big data with an easy-to-understand initial insight into the subject and an overview of the main classes of analytics methods. The various aspects are illustrated by examples from the manufacturing industry.

Für den Einblick in die Thematik erläutert die Richtlinie zunächst einige statistische und numerische Verfahren der Datenanalyse. Anschließend wird dargestellt, wie komplexe Sachverhalte durch visuelle Datenexplorationen veranschaulicht und verdeutlicht werden können.

To provide an insight into the topic, the standard first explains some statistical and numerical data analysis methods. It then shows how complex issues can be illustrated and clarified by visual data explorations.

Lösungswege für das häufig auftretende Problem redundanter Daten werden durch die Einführung der Eigenschaftsselektion und -reduktion aufgezeigt. Damit einher geht die daraufhin beschriebene

Solutions for the frequently occurring problem of redundant data are shown by introducing property selection and reduction. This is accompanied by the association analysis described thereupon, which

Assoziationsanalyse, die Zusammenhänge zwischen Variablen ermittelt.

Anschließend wird auf statistische Ansätze eingegangen, die heutzutage oft mit maschinellem Lernen in Verbindung gebracht werden. Die präsentierte Regressionsanalyse bestimmt Beziehungen zwischen einer abhängigen und einer oder mehreren unabhängigen Variablen.

Die daraufhin beschriebene Thematik der Klassifikation kann genutzt werden, um Beobachtungen vorgegebenen Klassen zuzuordnen, wohingegen die darauffolgenden Clusteringverfahren eigenständig neue Gruppen (Cluster) generieren.

Um Datensätze vor der Analyse zu bereinigen, wird ebenfalls auf die Anomaliedetektion eingegangen, durch die Ausreißer erkannt und eliminiert werden können, die andernfalls das Ergebnis negativ beeinflussen würden. Fehlende Datensätze können durch die vorgestellten Imputationsverfahren korrigiert werden.

Data-Mining-Verfahren auf Graphen, strukturierten Objekte sowie multiplen Instanzen werden ebenfalls kurz erläutert.

Die Richtlinie schließt mit einigen Ansätzen der künstlichen Intelligenz (KI) ab. Darunter fallen generische Verfahren, wie neuronale Netze sowie Ensemble Learning, z.B. die sogenannten „Random Forests“.

determines correlations between variables.

Subsequently, statistical approaches that are nowadays often associated with machine learning are discussed. The presented regression analysis determines relationships between a dependent variable and one or more independent variables.

The subsequently described topic of classification can be used to assign observations to predefined classes, whereas the subsequent clustering procedures independently generate new groups (clusters).

To clean up data sets before analysis, anomaly detection is also discussed, through which outliers can be detected and eliminated that would otherwise negatively affect the results. Missing data sets can be corrected by the presented imputation methods.

Data mining methods on graphs, structured objects, as well as multiple instances are also briefly explained.

The standard concludes with some artificial intelligence (AI) approaches. This includes generic methods such as neural networks as well as ensemble learning, e.g. the so-called “random forests”.

1 Anwendungsbereich

Mit dem Begriff „Big Data“ werden – obwohl er bereits seit einigen Jahren verwendet wird – unverändert sehr unterschiedliche Themen und Aspekte assoziiert und entsprechend in der gesellschaftlichen Diskussion differenziert diskutiert. Die immer weiter voranschreitende digitale Kommunikation, der in der Umsetzung befindliche Breitbandausbau und die überall mögliche Verarbeitungsmöglichkeit von Daten beflügeln diese Diskussion sowohl in der Öffentlichkeit als auch in der Fachwelt. Die Themen reichen von Datenschutz über Datensicherheit bis hin zu generellen Strategien für die digitale Wertschöpfung bei kleinen und mittelständischen Unternehmen und auch bei Großunternehmen.

Im Kontext dieser Richtlinie geht es bei Big Data um Technologien zur Datenanalyse. Entsprechende Algorithmen und Werkzeuge können Erkenntnisse über betriebliche Abläufe liefern und zu deren Optimierung beitragen. Hierzu bedarf es der Umsetzung dieser Methoden und Werkzeuge zur Verarbeitung, Analyse und Interpretation von umfangreichen und komplexen Daten in Big-Data-Anwendungen.

1 Scope

Although the term “big data” has been used for several years, it continues to be associated with very different topics and aspects and is accordingly discussed in a differentiated manner in the social debate. The ever-advancing digital communication, the broadband expansion that is currently being implemented, and the processing of data that is possible everywhere are fuelling this discussion both in the public and in the professional world. The topics range from data protection and data security to general strategies for digital value creation for small and medium-sized enterprises as well as for large companies.

In the context of this standard, big data is about data analysis technologies. Corresponding algorithms and tools can provide insights into operational processes and contribute to their optimization. This requires the implementation of these methods and tools for processing, analysing and interpreting extensive and complex data in big data applications.

Die Richtlinienreihe unterstützt Erstellende und Nutzende bei der Vorbereitung, Entwicklung, Inbetriebnahme dieser Anwendungen und ihrem nachhaltigen Einsatz. Letztendlich sollen diese Big-Data-Anwendungen verlässlichere Entscheidungsgrundlagen schaffen, um Produkte und Produktionsprozesse ökonomisch, ökologisch und technisch zu verbessern.

Die Richtlinienreihe soll dazu beitragen, die Vielfalt der in den letzten Jahren durch Forschungs-, Entwicklungs- und Praxisarbeiten entstandenen Erkenntnisse aufzubereiten, die Entwicklung und den Einsatz von Big-Data-Anwendungen in produzierenden Industrien sowie deren Nutzung im regulären Betrieb zu unterstützen.

Zur Zielgruppe gehören alle Stakeholder, von den Praktikern/Praktikerinnen bis zu den Entscheidern/Entscheiderinnen, von der Fertigungs- bis zur Prozessindustrie. Die Richtlinienreihe wendet sich dabei an Nutzenden und Erstellenden von Big-Data-Anwendungen in der produzierenden Industrie, unabhängig und übergreifend für alle Führungs- und Fachaufgaben.

The series of standards supports creators and users in the preparation, development, commissioning of these applications and their sustainable use. Ultimately, these big data applications should provide a more reliable basis for decision-making in order to improve products and production processes economically, ecologically, and technically.

The series of standards is intended to help process the variety of findings that have emerged in recent years through research, development and practical work, and to support the development and use of big data applications in manufacturing industries as well as their use in regular operations.

The target audience includes all stakeholders, from practitioners to decision makers, from operations to process industries. In this context, the guideline series addresses the users and the creators of big data applications in the manufacturing industry, independently and across all management and technical tasks.