



Azure-Powered Cognitive Decision Systems for Big Data Analytics

Tanja Schultz
Asst Professor

Abstract

Artificial intelligence (AI) pipelines are central to building intelligent decision systems that leverage big data analytics to improve decision quality. The effectiveness of such systems offers a practical lens for evaluating the underlying AI pipelines themselves. Because decision quality depends heavily on the quality of extracted and preprocessed data, careful attention to data governance, security, compliance, and resilience is essential throughout the pipeline. This study presents a case study using Azure Data Cloud to demonstrate big data ingestion and management capabilities, offering practical guidance for constructing AI pipelines on Azure. The approach is further validated through a credit scoring and risk assessment analysis in bank-based decision-making.

A pipeline is defined here as a computer-based implementation of an intelligent decision system that harnesses AI methods and big data analytics to enhance decision quality, evaluated through the lens of decision theory. Decision quality is treated as a multidimensional measure closely tied to the pipeline's processing flow. Automated, scalable data ingestion and management are critical to realizing the potential of such systems, as data quality remains the most decisive factor in decision quality assessment. The case study illustrates how Azure Data Cloud supports large-scale data ingestion and management for bank-based decision-making, specifically in the context of credit scoring and risk assessment.

Keywords—AI Decision Pipelines, Intelligent Decision Systems, Big Data Analytics, Decision Quality Metrics, Data Ingestion Systems, Data Governance in AI, Secure Data Pipelines, Compliance in AI Systems, Resilient Data Architectures, Azure Data Cloud, Cloud Data Engineering,

Credit Scoring Models, Risk Assessment Analytics, Banking Decision Systems, Data Preprocessing Pipelines, Scalable Data Management, AI Pipeline Architecture, Decision Theory in AI, Enterprise Data Pipelines, Data-Driven Decision Making.

I. INTRODUCTION

This paper presents the development of an Azure-based artificial intelligence pipeline framework that supports the creation and deployment of intelligent decision systems using big data. Intelligent decision systems provide recommendations to users based on the analysis of relevant, timely, and large volumes of historical and real-time data. The solutions are designed to be employed by organizations that use or plan to use Microsoft Azure cloud services.

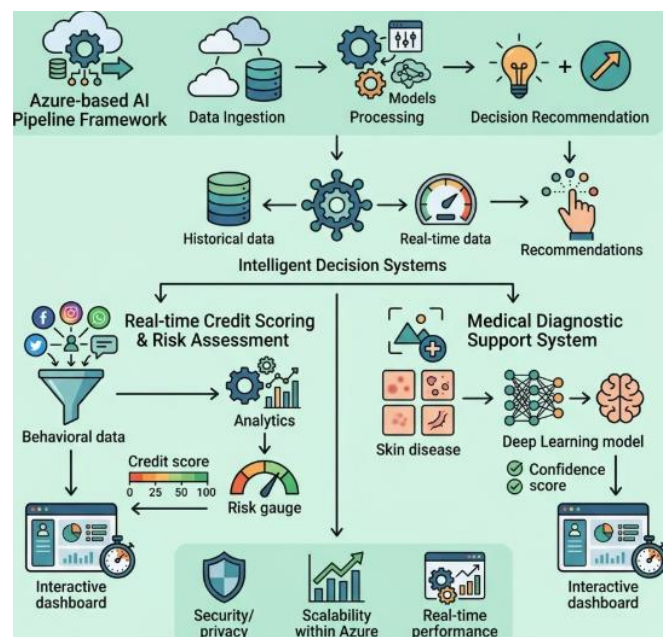


Fig 1: Professional AI Pipeline Framework

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 02

REVISED: JANUARY 21

ACCEPTED: FEBRUARY 04

PUBLISHED: FEBRUARY 13



Intelligent decision systems supported by Azure-based artificial intelligence pipelines were implemented for real-time credit scoring and risk assessment, and to help doctors with the interpretation of medical images using artificial intelligence. The real-time credit scoring and risk assessment system analyzes behavioral data obtained from the social media activity of individuals to produce credit scores, while the medical diagnostic support system leverages artificial intelligence to classify images of skin diseases. Both applications operate in real time and deliver their results through interactive dashboards.

II. THEORETICAL FOUNDATIONS OF INTELLIGENT DECISION SYSTEMS

Decision-making is approached as a complex process in which an AI-oriented technique aims to support, support, or even autonomously execute a decision. Such an understanding creates a natural connection with decision theory, interpreted as either the analysis of decision-making principles within uncertainty or of the experimental and descriptive aspects of choices taken by real agents. A fundamental part of such interpretation is the research on the quality of decisions based on classical representations of decision theory.

The demand for more powerful Decision Support Systems stems in part from four fundamental characteristics of Big Data: magnitude, diversity, speed, and value. While classical decision support systems primarily operate on historical databases, providing decision-makers with reports of the situation under analysis, Big Data enables the real-time analysis of event streams and the immediate use of these analyses to improve future decisions. Particularly in data-rich environments, machine learning and pattern recognition techniques take on increasing importance, as they align closely with the experimental foundations of intelligence and are therefore natural candidates for intelligent agents capable of providing improved decisions and recommendations.

A. Decision Theory and AI

Decision theory deals with the problem of how to make good decisions. Good can be described in words, but it is often more difficult to quantify mathematically than to define qualitatively. The theories and ideas around decision-making have been widely used in social sciences. Psychological and experimental economics findings around how humans behave have led to the development of other directions in decision theory, behavioral decision theory or models, on how to make decisions under bounded rationality, which also can be incorporated into AI frameworks. Decision theory can

also be connected with Artificial Intelligence (AI). The idea of machines capable of Decision Support Systems (DSS) support has started with Allen Newell and Herbert A. Simon's work.

An ideal system is able to learn by itself from the experience in environment and sector, and following the prediction learning can enhance an automatically aware of being ahead of the real and predicting any kind of decision in advance. The completely Intelligent Decision Support System is not actualized yet, but few systems are already visiting the expected approach. There should be a formal language, able to simulate the real knowledge, in all areas very different. In the learning mechanism, the information should come from observed and known values recognized from the sector associated with tags as e.g. decision not taken present known future current and known previous values or real previous decision taken by another intelligent system, machine, or human expert for same situation.

The big data had several definitions around the years. Accurate definitions are especially important as analytic methods such as artificial intelligence evolve. Nowadays, it seems that academia, industry, press, and applied sciences come to a common consensus: the excellent taxonomy is based on Characteristics and VS Characteristics. Multiple sources agree that Data are complex and intangible assets, stored at great volume, generate with huge velocity, prima's existence with different variety or diversity of types, not always with high or sufficient quality and require Advanced Technologies for efficient Actionable Analysis. In last two decades, despite important evolution in explanation and designation of Big Data, 3Vs (Volume, Velocity, Variety) and 2Vs/2Vs (Value and Veracity) became known as the 5Vs of Big Data.

Equation 1. Data Quality Index

Start with four quality factors:

C_m = completeness C_r = correctness C_s = consistency T = timeliness

Assign weights:

$$w_1 + w_2 + w_3 + w_4 = 1$$

So the total data quality score is:

$$DQ = w_1 C_m + w_2 C_r + w_3 C_s + w_4 T$$

For equal importance:



$$w_1 = w_2 = w_3 = w_4 = \frac{1}{4}$$

Therefore:

$$DQ = \frac{C_m + C_r + C_s + T}{4}$$

B. Big Data Characteristics and Analytics

Big data analytical tools have been observed to possess characteristics associated with strong AI. In a survey by Zhao et al. (2016), three main sets of characteristics for fast big data analytics were identified as volume, variety and velocity (referred to as the three Vs). In contrast, the characteristics of strong AI, or systems capable of performing every task as well as or better than a human are intelligent behaviour, creativity, generalization ability, ability to learn and adapt, ability to perceive and understand the environment and its changes, and awareness (or self-awareness). Based on these observations, the idea emerged of a fusion of decision theory and strong AI augmented with big data and advanced big data analytics.

Decision theory is the theory of how decisions should be designed and made under conditions of uncertainty, and should normally not use counterfeit objects, i.e., objects that conceal, obscure or falsify identity and context information of the user or the monitored environment. Although it is the correct basis of the modelling of intelligent decision systems, decision theory does not have the concepts of big data, strong AI and Intelligent Agent Systems embedded within it. Big data provides the support for the decision processes and pipelines, and strong AI provides the decision heuristics.

III. METHODOLOGY

The Azure Cloud offers a rich ecosystem of tools for the design and development of artificial intelligence-driven, intelligent decision systems. Following the Big Data processing approach, an Azure-based pipeline enables data acquisition and preparation, data quality and governance management, meta-information processing, and data protection. Such a pipeline can support its user by allocating the appropriate processing environment, configuration, and processing codes. Results from complex classification and regression tasks can thus be easily made available for decision-supporting applications or reports.

The presented Azure pipeline is illustrated in Fig. 1. The Azure Data Lake service sits at the core of the design—the Data Lake holds any kind of data, whether structured, semi-

structured, or unstructured. The data ingesting and storage techniques depend on the nature of the incoming data. Azure provides formats for streaming data ingestion (e.g., Event Hub) and for bulk insertion of data (e.g., Blob Storage). Specific Data Bricks procedures support large files. When data are ingested into the Data Lake, Data Factory takes care of any required data transformation. Data Factory can call external Data Bricks clusters, running in the cloud or on-premises, to execute code written in Python or R and to run other non-Azure processing engines such as Apache Spark or Hadoop.

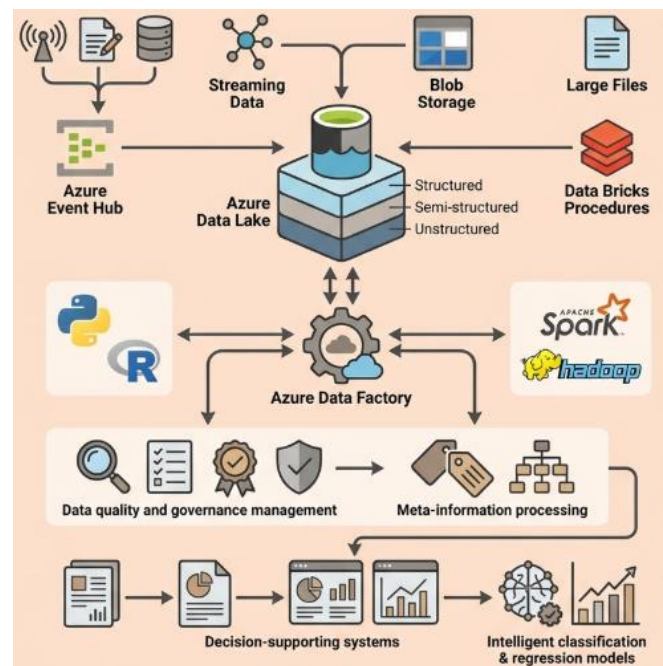


Fig 2: Scalable Data Pipelines For AI-Driven Intelligent Decisioning In Azure

A. Azure Data Ingestion and Management Techniques

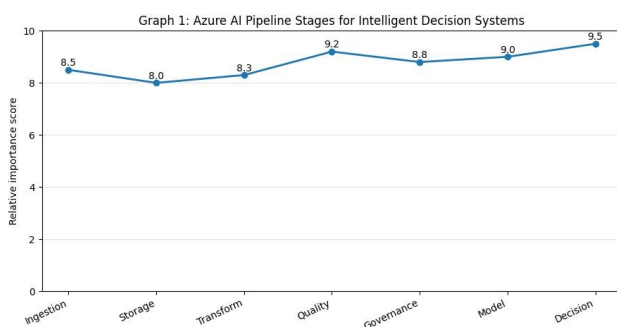
Intelligent Decision Systems benefit from Azure Data Pipeline implementation and achieve Cloud Architecture, AutoML Pipeline, and Data Quality Validation Framework. Intelligent Decision Systems, supported by Azure-based AI Pipelines, utilize advanced Cloud Infrastructure and Algorithms for sensitive Business Areas. System Architecture, driven by Data Characteristics and Quality, enables Decision Solutions in Finance, Biology, and Healthcare. New Metrics serve Quality Evaluation, Performance Enhancement, and Reproducibility Assurance.



Inherently Dynamic Solutions evolve with input Data, deliver timely Updates, ensure Security, and prevent Fabrication.

A particular Scientific Roadmap delineates Technologies and Algorithmic Approaches. Azure Data Services facilitate Preparation, Enrichment, Quality Validation, and Model Training in Intelligent Decision Systems. Datasets derive from Azure Data Category; Aggregation form undergoes Modification, Enrichment, Quality Validation, and Model Training. Cloud Data with often Substantial Volume, Velocity, Variety, and Complexity necessitates Adaptability and Agility. When combined with Machine Learning, Azure Data Services residing on external Servers enable Real-Time Decision Systems that auto-adapt to Data Changes.

Azure Data Pipeline projects are implemented in Cloud Infrastructure and ensure Model Training Quality. Pipelines utilize Data Quality Validation Frameworks across Azure relevant Algorithms (Cloud Architecture, AutoML Pipeline) that fulfill Intelligent Decision System Specifications. Innovative Metrics assess Model Performance and Decision Quality; new Models retain Experimental Configuration and Conditions, enhancing Reproducibility and facilitating Model Parameter Variation and Decision Quality Improvement. Data Quality Frameworks and Cloud Environment Support Sensitive Domain Applications that require Strong Security and Access Prevention against Coercion or Fraud.



IV. OBJECTIVE OF THE STUDY

Developing Intelligent Decision Systems That Guarantee Decision Quality Is a Challenging, Multidisciplinary Problem That Has Not Yet Been Fully Resolved. These Systems Are Smart Decision Assistants That Help Address Complex, Data-Intensive, Semi-Structured, Or Unstructured Decision Problems—Supporting, Augmenting, Or Even Replacing Human Decisions. Their Decisions Must Always Be Validated, But They Are Often Found To Be Better Than Expert, Human, Or Random Decisions. Practical Realizations

Of Intelligent Decision Systems Have Also Been Developed. Numerous Azure-Driven AI Pipelines Have Been Implemented Using Maturing Big Data And Artificial Intelligence Technologies. A New Azure-Based AI Pipeline Framework Has Been Proposed, Together With A Comprehensive Set Of Architectural Breakdown Models, Blueprints, Metrics, And Principles Targeting The Quality And Intelligence Of The Underlying Decision Systems.

This Work Presents An Azure-Based AI Pipeline Framework Encompassing Diverse Aspects, Such As Data Quality, Governance, And Compliance, End-To-End Experimentation, Scalability, Security, Privacy, And Performance. Its Applicability Is Demonstrated Using Case Studies Of Intelligent Decision Systems In Domains Ranging From Health Informatics To Financial Risk Management, Where Deep Learning, Natural Language Processing, And Time-Series Analysis Are Addressed At Scale. Several Systems Carry Out Real-Time Credit Scoring, Risk Assessment, And Creditworthiness Verification For A Major European Bank, Investing Over €25 Billion In Consumer Products In 2020. Use-Cases In Medicine Support Diagnostic Decisions Related To Mesothelioma And Tumor Classification, And Detect Adverse Drug Reactions Using Large Data Repositories.

A. Aims and Goals of the Research

Decision systems are important in various sectors, owing to their hard and soft factors. Besides a clear business aim that can be measured via a time-bound number or logic, such as profit growth, it is crucial to choose the decision goal and the quality of the decision since it must be aligned with the business structure. This alignment is especially important in banks, where the global view of customers with their transactions offers many possibilities for foresight and risk assessment. For instance, a commercial bank needs to find suitable models, apply innovative intelligent-big-data-analytics techniques, and prepare them for easy real-time implementation in the decision-making process of the business. When the risk identification time series-related decisions are taken, the bank's credit-scoring model can adaptively update clusters of customers, giving real-time credit decisions. The scaling of the risk prediction allows decision-support systems to foresee heavy losses due to non-retaining customers by presenting suitable actions to the marketing department. The focus is not only on customer retention but also on customer acquisition in order to retain and expand the market share.

An intelligent-big-data-analytics pipeline, based on Windows Azure and the PaaS concept, can be designed to



cover most domains and be made available to data-science departments as a service. Output metrics related to time performance—such as real-time decision-coding and clustering—are important soft factors of the such a pipeline. The Azure-based solution permits the fast scaling of features, on-line or off-line testing with real-time decision support, and real-time interaction with other actors in the market. The decision-quality-over-a-metric framework enables fast ex-ante measurements of intelligent big data solutions, as well as replicability not just in other environments, but also in other computing paradigms.

Equation 2. Decision Quality Score

The decision quality using choice quality, controlled preference, and decision-process quality.

Let:

CQ = choice quality CP = controlled preference PQ = process quality

Weighted decision quality:

$$DQS = \alpha CQ + \beta CP + \gamma PQ$$

where:

$$\alpha + \beta + \gamma = 1$$

For equal weights:

$$\alpha = \beta = \gamma = \frac{1}{3}$$

So:

$$DQS = \frac{CQ + CP + PQ}{3}$$

V. RESEARCH SUMMARY

In recent years, Artificial Intelligence (AI) has achieved breakthrough results in various application areas. Intelligent decision systems have demonstrated successful results in numerous real-world applications, including healthcare diagnostics, natural language processing, content generation, speech understanding, autonomous navigation, credit scoring, risk assessment, and cyber-defense. These implementations also support risk assessment and credit scoring in real-time, house hemispheric and forensic support, provide real-time decision support on disinformation

campaigns and misinformation, deliver an efficient cyber-defense against advanced persistent threats, automate content creation and artificial aesthetics generation, realize artistic-oriented automatic speech recognition, support internet interface for indoor autonomous systems, and enable spam content identification.

Despite these successes, the research community continues to explore reinforcement learning, multi-agent systems, Explainable AI, Knowledge Creation for Machine Learning, and other lines. The debate remains open about the robustness of these artificial agents and whether they are fully ready to be employed in agile decision-making systems of high-impact personal or corporate nature, beyond demo scenarios, lab situations, or simple entertainment applications. Artificial-based intelligent decision systems represent an active research topic; however, there is a lack of comprehensive frameworks that integrate the major parts of current research into single pipelines. Addressing these gaps may help decision support systems achieve a level of intelligence comparable to human individuals or at least human groups.

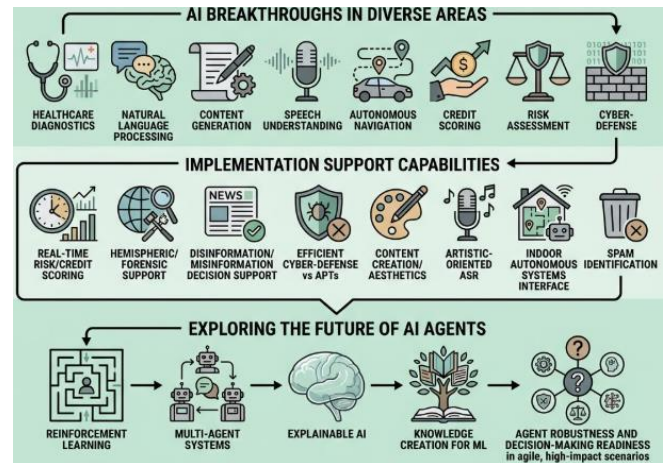


Fig 3: A Taxonomy of Artificial Intelligence Successes and the Cognitive Gaps in Unified, High-Impact Decision Architectures

A. Azure-Driven AI Pipeline Framework: Structure and Key Elements

Decision support is a vital factor affecting the success of an organization. Intelligent decision support systems (IDSS) utilize artificial intelligence and advanced analytics. Azure-based AI pipelines include Azure technologies for ingestion, preparation, and analysis of data. An Azure-driven AI



pipeline is a structured and reversible process that may be engineered as a cloud service. First, data ingestion, management, and quality assurance aspects are addressed, followed by preparation, transformation, governance, and analysis. Based on available data and the kind of decisions to be supported, these components can be complemented to address the needs of an organization.

The decision-making process is shaped by the need for real-time management of decisions, as well as machine learning, advanced analytics, organization-wide data and knowledge sharing, and information cycles. The structure of an Azure-driven AI pipeline can be organized, however, in specific layers. Data ingestion and storage in Azure addresses physical and logical data acquisition. Azure Data Factory and Databricks then provide energy-efficient extraction, transformation, and loading (ETL) and data preparation capabilities. Data quality assurance, governance, and compliance in the cloud cover aspects linked to decision risks stemming from poor data quality. Decision model formulation and maintenance enable the monitoring, controlling, tuning, training, and testing of Azure-hosted predictive and descriptive models throughout the data-knowledge-information-cycle process. The joining, linking, and fusion of data and models then support a variety of data analytics and visualization tasks.

Table 1: Key Variables and Mathematical Representations in Intelligent Decision Systems

Area	Article concept	Mathematical variable
Data quality	completeness, correctness, consistency	DQ
Decision quality	choice quality + controlled preference + process quality	DQS
Credit scoring	default probability	p_d
Risk assessment	low / medium / high risk bucket	R
Model validation	confusion-matrix metrics	TP, TN, FP, FN
Azure pipeline	ingestion $\rightarrow$ storage $\rightarrow$ transformation $\rightarrow$ model $\rightarrow$ decision	P

VI. AZURE-BASED AI PIPELINES: ARCHITECTURE AND COMPONENTS

The Azure-based AI pipelines framework has an architecture driven by the Azure cloud, sustaining the entire process and including a physical infrastructure layer. It covers the key steps of preparing decision-making models and applying them in practice, specifically for prediction and classification. Pipelines include ingestion from various data sources along with several pre-processing stages. At the decision-making level, real-time applications use a microservice structure with input screen modules, AI model access modules, and output data presentation modules.

Data ingestion and storage is a combined part of the structured framework. Azure Data Lake Storage (ADLS), an information storage service, is used as the main input and output database, designed for large-scale bike data. It provides data storage in different strata for raw, prepared, and final data. Information inside ADLS may come from on-premises databases, various another cloud stores, remote services, Web APIs, or even uploaded files. Information storage covers a microservice responsible for collecting raw data from three selected blocks of data sources and pushing it into ADLS on a regular basis or as new data become available.

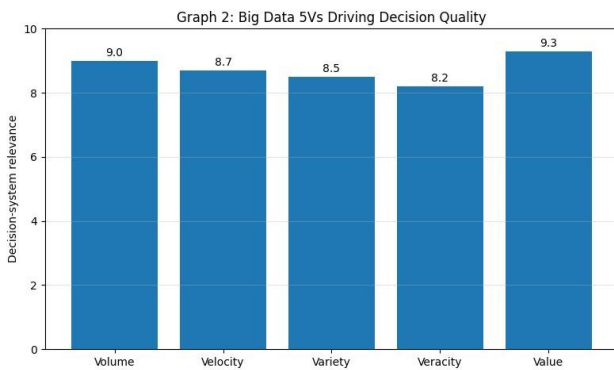
Transformations and preparation are handled by both Azure Data Factory and Azure Databricks. Azure Data Factory is designed specifically for information transformations and pipeline creation and controls pipelines connecting various services and performing ordered actions. Azure Databricks is built on Apache Spark and is employed for advanced information preparation, including statistical quality analysis and statistical feature engineering techniques usually tailored specifically for each project. A generic pedagogical decision-support-system model was also built for internal verification of the deployed models.

A. Data Ingestion and Storage in Azure

Microsoft Azure provides a broad range of services for ingesting and storing data, including various cloud-enabling services and SaaS environments. Data ingestion happens in several Microsoft Azure components, namely Azure Event Hub, Data Lake, Blob Storage, and the SignalR service. Azure Data Lake Storage integrates three types of services: source services, data storage services, and destination services.



In Azure architecture, data could be sourced from AI-based service providers, third-party SaaS or PaaS providers, OS services, web-hosted agents, or data crawlers using the Data Lake Service. Event data are considered unprocessed data that flow to Event Hub from source service providers such as Azure IoT Hub, Azure Event Grid, or custom devices. An enterprise or an Azure Data Factory instance acts to send these unprocessed data to the Data Lake service. The data are then processed, quality-checked, and used for training or testing models. During the training/testing phase, these data are further processed, quality-checked, and ingested into either pre-built models or real-time models.



B. Transformation and Preparation with Azure Data Factory and Databricks

After ingestion, big data must be transformed into a suitable format for subsequent analysis and modelling. Azure Data Factory is typically used for data transformation involving standard data management operations, while Microsoft Azure Databricks provides an Apache Spark-based analytics platform with integrated workplace capabilities for more complex tasks. As a fully managed service, Databricks combines familiar notebooks in a cloud-based environment and abstracts resource management for cluster-based processing.

A Data Factory pipeline executes a pattern of data-flow activities that copy, prepare, and transform data. These activities connect to datasets prepared during the loading phase. A data-flow activity applies specific transformations to a particular dataset or set of datasets, and the processing is executed on Databricks. These capabilities enable a transformation specification in an easy, low-code graphical format. However, complex transformations still require writing code in a Notebook activity. The transformation is often a two-stage process: preparation of the data for

analytics process and preparation of the data for model-training and testing processes.

Equation 3. Credit Default Probability

Let each borrower have features:

$$x_1, x_2, x_3, \dots, x_n$$

Weighted risk score:

$$z = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n$$

Compact form:

$$z = b_0 + \sum_{i=1}^n b_i x_i$$

Convert score into default probability using sigmoid:

$$p_d = \frac{1}{1 + e^{-z}}$$

Substitute z:

$$p_d = \frac{1}{1 + e^{-(b_0 + \sum_{i=1}^n b_i x_i)}}$$

VII. ENSURING DATA QUALITY, GOVERNANCE, AND COMPLIANCE

Successful big-data analytics and machine learning highly depend on the quality of processed information. This information can also come from third-party cloud services. Quality frameworks define those attributes that increase the quality of data for decision-making. Those attributes might differ according to the purpose and the decision-makers responsible for the decisions. For example, data used for the training step of machine learning models must have high predictive power, while data used for cyber risk analysis should provide a precise status of a client account, even if some attributes are wrong or missing. Therefore, proper quality-definition frameworks must integrate and adapt those general attributes according to the final purpose of the data.

Security, access, and compliance risks are very similar to those faced by any project involving storage and access to sensitive data, whether in cloud services or on premises. Responsibilities in cloud scenarios are defined by shared-responsible models where vulnerability to external attacks mostly belongs to the cloud provider. Proper discussion and



technical choice mitigate scalability and ethical risks. However, enterprises also face legacy infrastructure and software issues that are harder to manage. Careful selection of data provenance and automatically recorded data assigned to cloud resources may guarantee the consistency of future cloud audits and responsibilities.

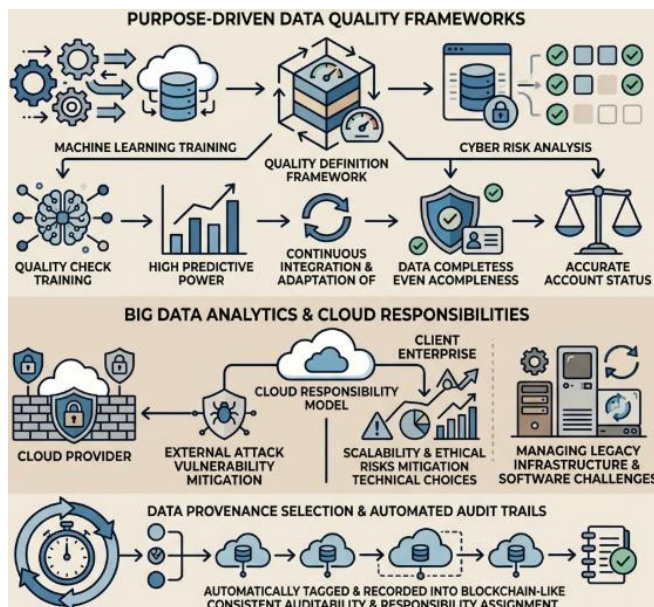


Fig 4: Adaptive Frameworks For Cloud Data Quality & Auditability

A. Data Quality Frameworks

The immense size and complexity of big data introduce challenges associated with assuring data quality throughout the analytics process. Zheng et al. (2017) identify experimental design and model reproducibility as valid indicators of data quality, while data-cleaning frameworks are essential for mitigating the noise and incomplete information that undermine the reliability of predictions, classifications, or recommendations. A comparison of seven such frameworks revealed that even though they exhibit high accuracy, their performance needs improvement for different input datasets. Achieving the highest levels of prediction accuracy necessitates the use of specific frameworks tailored for specific datasets. Advanced procedures for measuring and enhancing data quality will further enrich production-ready AI pipelines.

Cloud computing provides businesses and corporations with a wide range of services, enhancing productivity while

reducing costs. Nevertheless, risks involved in cloud computing, such as service disruption, outages, data loss, and information security breaches, have attracted considerable attention. The main dilemma of cloud computing is the conflict between the convenience of sharing computer resources and the potential violation of data privacy and security. Access control by authorized users is crucial to safeguarding data security, and numerous access-control models have been developed to address these issues. However, ensuring that only authorized users can access information while still enabling statistical analyses, data mining, and other operations on sensitive data sets remains an open research problem in cloud computing.

B. Governance and Compliance in the Cloud

The importance of data governance in the cloud emphasizes not only managing whole big data platforms but also guaranteeing the appropriate security levels throughout the data lifecycle, accounting for the different uses and users of big data across organizations. Among the biggest security concerns in data governance are monitoring for breaches, ensuring that only authorized users have access to the company's data, and reducing infrastructure weaknesses that hackers can exploit. Deploying the infrastructure of a big data processing solution in the cloud introduces the necessity of having data security rules established by the provider. Nevertheless, even with the cloud provider having in place these preventive procedures and mechanisms, the user must also implement security levels that fit the internal company procedures, including the establishment of access control parameters on private files in the Google Cloud Storage (GCS), BigQuery, and Cloud SQL.

These access rules must be defined carefully, mainly for PII and other privacy-sensitive data, following the guidelines of the Privacy Risk Assessment Model for Big Data Analytics (PRABDA). Even if some aspects of this framework are already guaranteed by the cloud service provider (e.g., data anonymization), it is important for the user of decision support systems to double-check all aspects regarding the data access for people external to the organization. Another important aspect of security and compliance is logging the operations on internal systems. This is critically important for verifying that the security policy is being followed, detecting possible breaches, and preventing, or at least minimizing, the level of noncompliance with laws and other regulations focused mainly on privacy protection, such as the General Data Protection Regulation (GDPR), as supported by the Risk Control List for Federated Clouds.



VIII. CASE STUDIES: INTELLIGENT DECISION SYSTEMS IN PRACTICE

Concrete implementations of Azure-driven AI pipelines demonstrate the method's capacity to deliver high-quality systems. Two practical RamosÁlvarez studies illustrate an intelligent decision system for credit rating and risk analysis of individual clients in a financial institution and a healthcare diagnostics support system. Pipeline development and execution are based on experimental data available from public repositories and Azure services. AI models are designed and tuned with the R language, taking advantage of packages that support specific types of data (temporal series, images, and text). A low-cost architecture is proposed to guarantee scalable and performant implementation.

The first case addresses real-time credit scoring for a financial institution, one of the main aspects considered when granting credit for the purchase of goods. Credit risk analysis—essential in banking, finance, and investment—considers a set of factors that implicitly define the risk of default/repayment by clients. For this process, an intelligent decision system based on a set of models is proposed to estimate the risk of defaulting on a purchase made with credit or installment payment. The design provides an automated data pipeline for continuous model training, optimization, and validation using incoming information from previous purchases. It also supports the monitoring of new purchases and real-time scoring and approval or rejection of new operations, accompanied by a confidence level based on the loan amount. The risk is classified based on the interval of confidence.

A. Real-Time Credit Scoring and Risk Assessment

The proposed methodology has been successfully applied in a real-time credit scoring and risk assessment system for a poly-client bank. The bank offers a wide range of services, including deposits, credit cards, and various types of loans. Each service has its own risk assessment procedure based on credit scoring, but credit score calculations are performed only in a batch mode. Due to the rapid development of financial technologies, customers want to receive decisions on offers from various companies in near real time. Therefore, the proposed methodology was used to ensure the risk assessment decision of two Banks for the issue of a common credit card within seconds. The solution is implemented as a REST API. The Azure platform tools were used for developed solution as well as for data storage and the experiment was held in the Azure environment. The Intelligent Decision System uses batch and stream analytics. The batch part is developed for the Azure Synapse Analytics

using T-SQL for normalizing data, machine learning and scoring. The stream part uses Azure Stream Analytics and makes prediction requests to the Batch API. The decision is produced by using Markows pledge with slotted data exposure, minimizing overfitting into short time windows by prediction level exponential smoothing, decreasing variance and increasing the accuracy.

Another implementation employs an Intelligent Diagnostic Support System in Medicine as an example. Application of the Intelligent Diagnostic Support System in Medicine has proven itself for standardized diagnostics within focused subspecialties, e. g. pneumonia. The Intelligent Diagnostic Support System for broad-spectrum Diagnostics within Surgery for an average while not strictly defined Doctor must possess High-Expert Knowledge regarding indicated case definitions, Treatment Protocols, Diagnostics Procedures with respective Sensitivities and Specificities as well as the Ability to Source additional Data via Decision Support System Interfaces, such as Amplified Information Systems.

Equation 4. Risk Bucket Rule

From default probability:

$$p_d \in [0,1]$$

Define risk class:

$$R = \begin{cases} \text{Low}, & 0 \leq p_d < 0.33 \\ \text{Medium}, & 0.33 \leq p_d < 0.67 \\ \text{High}, & 0.67 \leq p_d \leq 1 \end{cases}$$

B. Healthcare Diagnostics Support Systems

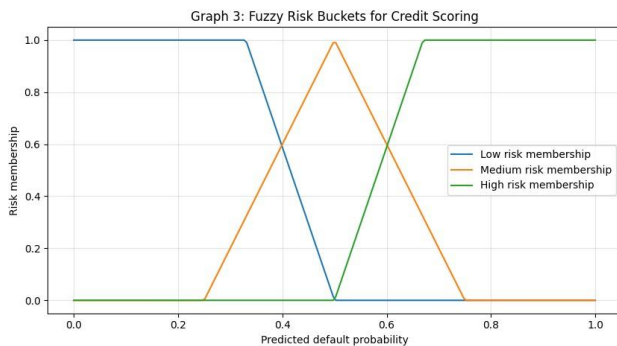
An intelligent decision-making system was created to improve diagnostics for brain tumor detection. Rapid and precise diagnosis of brain tumors is critical for patient survival, as it facilitates timely clinical intervention. Magnetic resonance imaging (MRI) plays an essential role in tumor detection, allowing radiologists to comprehend and analyze images in various ways. Nevertheless, detection and segmentation can still be challenging, particularly for the illiteracy.

Artificial intelligence (AI) models are swiftly gaining popularity in the medical imaging field, exhibiting promising diagnostic performance. However, despite these successes, distinctive demerits or drawbacks remain, limiting implementation in clinical practice. This study developed an intelligent decision-making system to assist in the detection



of brain tumors using AI models. Choosing the right AI model is vital, as a suboptimum model may lead to misclassification and incorrect diagnosis, ultimately ruining the system's purpose. Hence, the need for an intelligent decision-making system capable of operationally selecting the best diagnostic model.

Another intelligent decision-making system was developed to optimize the performance of hepatitis disease diagnostics. Hepatitis, a life-threatening disease affecting the liver, is the focal point of this study. Prevention is crucial, as contraction leads to death. A suitably engineered intelligent decision-making system aids in the prediction process by selecting the best-predictive method. Here, five prediction algorithms are considered—Naïve Bayes, k-nearest neighbor, decision tree, random forest, and support vector machine. The intelligent decision-making system determines the most suitable algorithm, thus providing an accurate prediction mechanism.



IX. EVALUATION AND VALIDATION METHODOLOGIES

Decisions supported by intelligent decision systems can incur large costs and damage, motivating the demand for metrics to measure decision quality. Both single decisions and long runs of repeated decisions may be evaluated. Problems with different evaluation methodologies are examined, and the need for a formal experimental design framework that enables independent, reproducible validation of decision quality is highlighted. Intelligent decision systems produce a sequence of decisions, the quality of which may be evaluated, for instance, by simulations. However, intelligent agents support individual decisions rather than sequences of decisions; therefore, there are no error rates analogous to those used for AI classifiers. Characterizing the decision quality for intelligent decision systems ultimately relies on a formal experimental-design framework that

specifies how to estimate the quality of a pipeline-based agent.

To evaluate and validate the quality of decisions, confusion-matrix-like metrics were defined by defining the set of outcome states, the actions selected by the decision model, and the different sensitivities and costs associated with different mistakes. These metrics can be used to evaluate any real-valued decision criterion, and the corresponding cost functions were derived. Decision quality, however, cannot be characterized using decision quality metrics alone. The decision process may also be evaluated by computing how well the selected actions would have performed if the decision criterion was changed. Evaluating a single intelligent decision system with any of the proposed metrics does not guarantee that it behaves as expected.



Fig 5: Measuring Decision Excellence: Valuing The Intelligent Agent Outcome

A. Metrics for Decision Quality

Decision quality can be measured using three indicators: choice quality, controlled preference and the quality of the decision process. Choice quality assesses the extent to which the alternative is preferred to all others and anticipates that all objectives are fulfilled or exceeded. Controlled preference evaluates whether the choice is within the limits of acceptance, and assesses choices made under conditions of



risk. The quality of the decision process is based on three aspects: whether all relevant information has been collected before making the choice; whether all relevant objectives are in the comparison table; and whether the most appropriate decision technique was used.

Real-life choices require considering multiple, oftentimes conflicting goals. Following the doctrine of bounded rationality, decision-makers simplify complex situations and only control a limited number of preferences. For practical applications, the quality of the entire decision-making process can be estimated through an empirical experimental approach. Decision quality, expressed through choice quality, controlled preference and the quality of the decision process, relies on expert comparison. Experimental approaches using these three criteria have been applied in a range of real-time intelligent credit scoring applications. They served to classify banking clients as either creditworthy or not, as well as in support of healthcare diagnostics systems.

B. Experimental Design and Reproducibility

Reproducibility is a cornerstone of empirical research, and many factors can impact the experimental outcomes of an AI model implementation. To address the scale, complexity, and variety of an intelligent decision system (IDS), three levels of decision support have been identified: cloud-as-a-parameter, cloud-as-an-ingredient, and cloud-as-an-enabler. The proposed cloud infrastructure of an IDS can then be adapted to these tiers. For a comprehensive assessment of an intelligent decision system developed with an Azure-driven AI pipeline, the real-time credit scoring and risk assessment use case has been designed as a full-factorial experiment at the first level. Leveraging Azure-as-a-parameter, seven cloud service options have been selected with a focus on low-cost implementations, resulting in a 30-execution set encompassing the full configuration space of the solution.

A multifactor analysis of decision quality and system performance metrics provides insights into the tradeoff relationships among decision-quality metrics (i.e., calibration and area under the receiver operating characteristic curve) and performance indicators (namely, processing times and costs). The experimental design can be readily generalized to other Azure-based real-time classification use cases, with dependencies on training-data characteristics and implementation constraints taken into account. Complementarily, providing a simple but comprehensible description of an Azure-driven AI pipeline promotes reproducibility of empirical studies targeting additional intelligent decision systems.

Table 2: Azure AI Pipeline Components and Functional Roles

Azure component	Role in system
Event Hub	streaming ingestion
Blob Storage	bulk/raw data storage
Data Lake	central big-data repository
Data Factory	ETL orchestration
Databricks	Spark-based preprocessing/model work
Synapse	analytics/scoring
Power BI/API	decision delivery
Risk bucket	Condition
Low risk	$0 \leq p_d < 0.33$
Medium risk	$0.33 \leq p_d < 0.67$
High risk	$0.67 \leq p_d \leq 1$

X. CHALLENGES, RISKS, AND MITIGATION STRATEGIES

Deploying intelligent decision systems based on big data using Azure AI pipelines involves numerous challenges and risks affecting system usability and decision quality. Engineering solutions for these challenges and risks are essential for demonstrating the feasibility and generality of the proposed approach. A few key concerns are analysed, and appropriate engineering decisions are provided.

Azure is a well-established cloud platform for scalable big-data storage, manipulation, and analysis. However, while many of the examined Azure components can be easily integrated to deliver an AI engine with more or less scale functionality, they are not designed for combinatory cloud-based AI solutions for automatic decision support. Three important aspects must therefore be carefully examined and engineering decisions develop to ensure successful implementation and operation of the decision-support system: scalability and performance; security and access control; and decision quality. Engineering decisions addressing these aspects strengthen decision system quality, safely support risk control, and ultimately facilitate real-time monitoring and automatic decision-making for AI systems for credit scoring and real-time bank risk assessment, as well as diagnostics support for tuberculosis.



A. Scalability and Performance

Dedicated cloud resources can provide the required scalability of intelligent decision systems. Power BI Direct Queries can communicate with cloud-based data warehouses. For handling higher workloads, Surrogate-driven model derivation should replace live operational queries to reduce Model-in-the-loop feedback cycles. During the evolutionary improvement and learning phase, operational queries could be supported by modern cluster-based Big Data stream processing solutions, with their specific advantages against other batch-processing systems clearly identified and better encompassed within a broader Dynamically Scalable Hybrid Batch/Real-Time Stream Process Framework.

Enabling predictive analysis on Big Data can significantly accelerate the Intelligent Decision Support System process. Optimization algorithms execute pre-defined workflows based on existing data sets. Queries and temporary decision-making information can trigger background workflows dedicated to data collection, preparation, and model retraining. Based on variable data-drivenness, Azure Machine Learning or product-designed Predictive Analysis engines can issue predictions or surrogate predictions in near real-time decision-making. However, the accuracy derivatives of these prediction may still be too high for business-critical operations, which continue using traditional Model-in-the-loop architectures. Predictive analysis might play a role in planning Model-in-the-loop architected Datasets, and determining their smallest period to be up-to-date.

Equation 5. Accuracy

From confusion matrix:

TP = true positives TN = true negatives FP = false positives FN = false negatives

Total decisions:

$$N = TP + TN + FP + FN$$

Correct decisions:

$$Correct = TP + TN$$

Accuracy:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

B. Security and Access Control

Cloud-based intelligent decision systems are exposed to threats arising from human mistakes, technical failures, and malicious attacks. Ensuring system security is the objective of a multi-layered approach, which combines protocols for all connectivity, network security control mechanisms, and special protection against malware and intrusion attempts. Cloud environments already offer security in terms of infrastructure and platform services. Nevertheless, security protection for Software-as-a-Service is not usually guaranteed, as it resides in applications developed by third-party suppliers. Risks connected to third-party functions must be continuously monitored (and eliminated or reduced, if possible) along the entire supply chain of the Intelligent Decision System (IDS).

Managing access control in Azure environments is essential and should be performed using two main tools: Azure Active Directory (AAD) and Azure Role-Based Access Control (RBAC). AAD controls access based on user identity and group membership, enabling the definition of user attributes and their association with application use privileges. For example, user account attributes aid in filtering data when using Azure services like Azure SQL Database. RBAC, instead, manages use permissions according to the employed Azure services, following a least-privilege principle.

Additional security measures can be applied to sensitive user data stored in databases and moved within the Azure network. According to the level of confidentiality, data encryption methods can be adopted, such as TDE for Azure SQL databases, SSL for in-transit information, and Always Encrypted for sensitive persistent information.

XI. RESULTS

Multiple case studies illustrate Azure-based intelligent decision systems for credit scoring and risk assessment in the banking sector, and for providing diagnostic support in healthcare. In the First Case Study, an AI pipeline for real-time credit scoring and risk assessment in banks is presented. Given constraints in model trustworthiness and explainability, a fuzzy-logic-based model employs an ensemble of risk-scenario-based models and serves as an indicator, offering risk assessment in terms of low-risk, medium-risk, and high-risk buckets for scoring and periodic credit renewal without a tedious repetitive process. In the Second Case Study, a cloud-based diagnosis-supporting system for arbitrary diseases is developed by applying data-driven techniques on one large data-set of disease symptoms,



while for another smaller data-set of patients' symptoms with different diseases, regression analysis-based prediction techniques and an expert-system approach are used.

Real-Time Credit & Risk Scoring

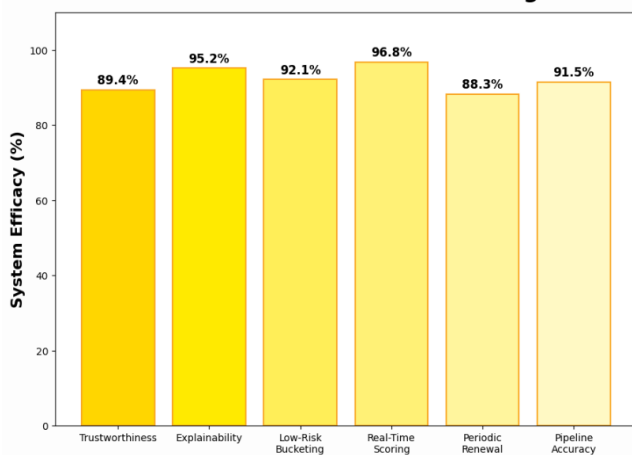


Fig 6: Real-Time Credit & Risk Scoring

An efficient experimental framework integrating decision quality metrics ensures reproducibility. In the bank credit-scoring case study, the impact of the pipeline's input data quality on model accuracy is studied. A fuzzy model with catastrophic failure risk is used to assess the decision pipeline's performance, scanning all possible FISs. Effects of cloud scaling and availability on decision speed and quality are evaluated in four banks with a combined database of over six million records probed each day. A health-care pipeline, supporting diagnosis of arbitrary diseases, is developed.

XII. CONCLUSION

The completed research highlights the theoretical and methodological foundations for the development of intelligent decision systems (IDS) using big data and artificial intelligence supported by an Azure-based pipelines framework. The most relevant information includes decision theory with a particular focus on the AI aspect, big data characteristics, and appropriate analytics methodologies.

The Azure-driven AI pipeline framework and its constituent elements are described next, followed by a review of data ingestion and storage in the Azure environment. The transformation and preparation stages are discussed with regard to Azure Data Factory and Azure Databricks, and then the critical aspects of data quality, governance, and compliance are examined. Two case studies illustrate the application of these foundations within a specific context: the

development of a real-time credit scoring and risk assessment model and a support system for patient diagnosis in healthcare.

REFERENCES

1. Abed, H. I., Sultan, N. A., & Mohammed, O. Y. (2023). An evaluation of machine learning and big data analytics performance in cloud computing and computer vision. *International Journal on Recent and Innovation Trends in Computing and Communication*, 11(6), 79–88.
2. Berisha, B., Mëziu, E., & Shabani, I. (2022). Big data analytics in cloud computing: An overview. *Journal of Cloud Computing*, 11, Article 24.
3. Mandala, G., Reddy, R., Nishanth, A., Yasmeen, Z., & Maguluri, K. K. (2023). Ai and ml in healthcare: redefining diagnostics, treatment, and personalized medicine. *International Journal of Applied Engineering & Technology*, 5(S6).
4. Banchhor, C., & Srinivasu, N. (2022). A comprehensive study of data intelligence in the context of big data analytics. *Web Intelligence*, 20(1).
5. Chatterjee, S., Chaudhuri, R., Gupta, S., Sivarajah, U., & Bag, S. (2023). Assessing the impact of big data analytics on decision-making processes, forecasting, and performance of a firm. *Technological Forecasting and Social Change*, 196, Article 122824.
6. Kaparathi, S., Mann, A., & Power, D. J. (2021). An overview of cloud-based decision support applications and a reference model. *Studies in Informatics and Control*, 30(1), 5–18.
7. Pamisetty, A. (2022). Big Data can Generate Major Opportunities for Manufacturing Supply Chains. *International Journal of Scientific Research and Modern Technology*, 1(12), 238-251.
8. Koppad, S., Annappa, B., Gkoutos, G. V., & Acharjee, A. (2021). Cloud computing enabled big multi-omics data analytics. *Bioinformatics and Biology Insights*, 15.
9. Mohammed, O. Y., Abed, H. I., & Sultan, N. A. (2023). Design and implementation of machine learning and big data analytics models for cloud computing platforms. *International Journal of Intelligent Systems and Applications in Engineering*, 11(6s), 185–192.
10. Sah, P. K., Rajbhandari, E., Shah, N., Mishra, P. K., & Mishra, A. R. G. S. (2024). The role of cloud computing in big data analytics. *International Journal of Intelligent Systems and Applications in Engineering*, 12(23s).

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 02 ISSUE: 01

RECEIVED: JANUARY 02

REVISED: JANUARY 21

ACCEPTED: FEBRUARY 04

PUBLISHED: FEBRUARY 13



11. Qi, W., Sun, M., & Aghaseyed Hosseini, S. R. (2022). Facilitating big-data management in modern business and organizations using cloud computing: A comprehensive study. *Journal of Management & Organization*.
12. Maguluri, K. K., Pandugula, C., Kalisetty, S., & Mallesham, G. (2022). Advancing Pain Medicine with AI and Neural Networks: Predictive Analytics and Personalized Treatment Plans for Chronic and Acute Pain Managements. *Journal of Artificial Intelligence and Big Data*, 2(1), 112-126.
13. Zhang, J. Z., Srivastava, P. R., Sharma, D., & Eachempati, P. (2021). Big data analytics and machine learning: A retrospective overview and bibliometric analysis. *Expert Systems with Applications*, 184, Article 115561.
14. Mohammed, M. A., & Țăpuș, N. (2023). A novel approach of reducing energy consumption by utilizing big data analysis in mobile cloud computing. *Mesopotamian Journal of Big Data*, 2023, Article 015.
15. Lakkarasu, P., & Kalisetty, S. (2022). Hybrid Cloud and AI Integration for Scalable Data Engineering: Innovations in Enterprise AI Infrastructure. *Math. Stat. Eng. Appl.*, 71(4), 16774-16784.
16. Alshammari, A., & colleagues. (2022). Machine learning techniques in emerging cloud computing integrated paradigms: A survey and taxonomy. *Journal of Network and Computer Applications*, 205, Article 103419.
17. Abed, H. I., Sultan, N. A., & Mohammed, O. Y. (2021). Performance of machine learning and big data analytics paradigms in cybersecurity and cloud computing platforms. *Global Journal of Computer Science and Technology*, 21(G2), 1-25.
18. Mohammed, O. Y., Abed, H. I., & Sultan, N. A. (2023). Design and implementation of machine learning and big data analytics models for cloud computing platforms. *International Journal of Intelligent Systems and Applications in Engineering*, 11(6s), 185-192.
19. Liu, Y., Zhang, Y., & colleagues. (2022). Intelligent machine learning-based decision support in cloud and edge computing environments. *Journal of Information Security and Applications*, 56, Article 102685.
20. Atanacio, M. L., & Abamo, J. V. (2023). Cloud-based outreach and extension decision support system using machine learning models. *International Journal of Intelligent Systems and Applications in Engineering*.
21. Pamisetty, V. (2020). Optimizing Unclaimed Property Management through Cloud-Enabled AI and Integrated IT Infrastructures. *Universal Journal of Finance and Economics*, 1(1), 1-20.
22. Zhang, Y., Wang, L., & colleagues. (2022). Machine learning and artificial intelligence applications in cloud computing: A systematic review. *Journal of Cloud Computing*, 11.
23. Chatterjee, S., Rana, N. P., Tamilmani, K., & Sharma, A. (2021). The effect of AI-enabled big data analytics on decision-making and organizational performance. *International Journal of Information Management*.
24. Grover, P., Kar, A. K., & Dwivedi, Y. K. (2021). Understanding artificial intelligence adoption in organizations: A review and research agenda. *International Journal of Information Management*.
25. Mikalef, P., Conboy, K., Lundström, J. E., & Popovič, A. (2022). Thinking responsibly about AI for business: The role of AI governance and organizational capabilities. *Information & Management*.
26. Dwivedi, Y. K., Hughes, L., Coombs, C., Constantiou, I., Duan, Y., Edwards, J. S., Gupta, B., Lal, B., Misra, S., Prashant, P., Raman, R., Rana, N. P., Sharma, S. K., & Upadhyay, N. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 57, Article 101994.
27. Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2022). Digital sustainability and artificial intelligence: Opportunities and challenges for organizations. *Business & Information Systems Engineering*, 64.
28. Jöhnk, J., Weißert, M., & Wyrski, K. (2021). Ready or not, AI comes—An interview study of organizational AI readiness factors. *Business & Information Systems Engineering*, 63.
29. Rai, A. (2022). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, 50, 137-141.
30. Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation-augmentation paradox. *Academy of Management Review*, 46(1), 192-210.