This journal provides immediate open access to its content under the Creative Commons BY 4.0 license. Authors who publish with this journal retain all copyrights and agree to the terms of the above-mentioned CC BY 4.0 license

DOI: 10.2478/seeur-2024-0022

AUTOMATION OF DATA VALIDATION PROCESSES IN BANKING: A CASE STUDY ON VALIDATING ALBANIAN IDENTITY NUMBERS

Brunela Karamani
Department of Computer Science,
Faculty of Information Technology,
University of Tirana, Albania
bkaramani@fti.edu.al

ABSTRACT

In the banking industry, data accuracy and integrity are paramount for ensuring regulatory compliance, operational efficiency, and risk management. However, manual data validation processes often lead to delays, errors, and inefficiencies, posing challenges for financial institutions. To address these issues, many banks are turning to automation to streamline data validation processes and improve accuracy. Using data automation, banks optimize time-consuming manual tasks such as data entry, validation, and retrieval by creating automated software processes that execute these tasks quickly and accurately. This case study examines how a hypothetical bank uses automation to improve data validation processes, focusing on Albanian Identity Numbers (AINs). By automating validation tasks, the bank enhances accuracy, compliance, and efficiency, mitigating the limitations of manual processes. The study outlines the implementation steps, discusses results, and highlights the importance of automating data validation in the banking sector.

Keywords: Data validation, Banking Automation, Data Accuracy, Rule-Based Validation, Drools Engine, Operational Efficiency

INTRODUCTION

In the banking industry, data accuracy and integrity are crucial for regulatory compliance, operational efficiency, and risk management. Traditional manual data validation processes can be error-prone, slow, and inefficient, leading to significant challenges. This case study explores the journey of a hypothetical bank implementing automation tools to enhance data validation, focusing on the validation of Albanian Identity Numbers (AINs). The study highlights the impact of automation on operational efficiency and regulatory compliance, providing a detailed analysis of the implementation process and its outcomes.

The hypothetical bank, faced with the complexities of manual data validation processes, identified the key pain points that hindered operational efficiency and regulatory adherence. Recognizing the critical importance of data accuracy and integrity in its operations, the bank embarked on a journey to implement automation tools to address these challenges. The implementation process began with a thorough assessment of the bank's existing data validation processes and the identification of specific requirements and objectives for the automated solution.

LITERATURE REVIEW

The literature review established that data governance plays a key role in data quality improvement in banking. Tools like blockchain, machine learning, and RPA are relevant when it comes to applying innovative solutions in data validation to provide more engaging, effective, and accurate techniques than traditional ones.

Elaborating on the information, data governance is considered as the means of implementing data quality policies in banking systems. Al Wahshi, Foster, and Abbott (2022) looked at the relationship between data governance and data quality in Omani banks. They pinpoint the need for proper data governance principles in order to drive the quality, validity, and credibility of the banking data. Policies, procedures, and standards must be in place to help manage the organization’s data well and ensure that it is consistent, credible, and trustworthy for the purposes of decision-making.

In his work, Paik et al. (2019) focused on the data management in distributed blockchain system with the stress to the architectural and governance layers. Based on the fact that blockchain operates within other nodes and not a central server, makes it more secure for data storage. It is becoming essential to try to implement this technology in the banking sector to increase the transparency level and reduce the fraud cases as well. This knowledge establishes the potential ways of applying blockchain to the already existing data management

systems and confirms how it will help shape a reliable data governance framework that supports data accuracy and trust.

Advanced techniques of machine learning are growing to be essential in augmenting the data validation activity. Some of the techniques adopted by Mahinda Mailagaha Kumbure et al. (2022) to forecast stock market are discussed below and may be adopted for data validation in banking: They can look for own patterns and deviations in vast amount of data, which makes them more flexible than classic rules that have to be set up manually. This flexibility is important in banking given that data characteristics evolve over short time intervals.

Validation of entered data with computer vision and deep learning has successfully proved effective in the banking sector. Automated bank cheque verification is well illustrated by Agrawal et al. (2021) where it ascertains that manual checks contain high error rates and slows down processing time. This paper demonstrates how deep learning algorithms can be applied for data validation in banking and are considerably faster and precise than manual checks.

On the basis of the current literature, banking applications of image processing and deep learning for automated data validation have also been found to be effective. Agrawal et al. (2021) combines both show the efficiency of automated Bank cheque verification which decreases manual errors and increases speed of the process. This paper demonstrates how deep learning can be applied in a range of validation activities in banking to offer a cheaper and more effective solution to manual validation techniques. Wewege, Lee, and Thomsett (2020) explain the current disruptions or firms and trends in digital banking. The advancement in digital technologies is now conclusive and has quickly disrupted the banking structures including data validation. Technologies like AI and RPA are now considered essential to the organizational structure in many modern banking systems to improve data processing and overall efficiency. These technologies inform the move towards more automated, Real-Time data verification systems.

Pramod (2022) and Sobczak (2021) share similar interest on the adaptation and integration of RPA across industries, specifically in the banking sector. With RPA, redundant activities like data input or verification, can be performed by the software and this has the potential to eliminating human errors and enhance efficiency in organizational processes. In these works, the effectiveness and potential difficulties in the implementation of RPA, the necessity of proper planning and management of the process, and other relevant issues are

outlined. The incorporation of RPA in banks also implies improvements in the flow of data validation, data quality and compliance.

Hanafizadeh & Gerami Amin (2023) provide an analysis of the nature of the possibilities that FinTech brings to banking service domains. Thus, banks are constantly facing a need for managing and validating data due to the advanced FinTech innovations. The advent of FinTech solutions can help to bring new ideas into practice regarding enhancing the data quality and validation functions. These solutions are generally based on new elements such as Artificial Intelligence, Machine Learning and Blockchain to present effective and more secure management practices of data.

BACKGROUND

The implementation was done in a retail bank in Albania, which offers various financial products and services The bank operates in a highly regulated environment and is subject to strict compliance requirements imposed by regulatory authorities. As part of its operations, the bank handles a large volume of transactional data daily, including customer accounts, payments, and financial transactions Wewege et al (2020). However, manual data validation processes led to errors and delays, impacting the bank's operational efficiency and regulatory compliance efforts. Undertaking manual data validation procedures for transactional data poses significant obstacles. One of the foremost challenges is the propensity for human error, which can result in inaccurate transaction data. Additionally, manual validation processes are often time-consuming, leading to delays in transaction processing and customer service provision. Furthermore, the substantial volume and intricacy of transactional data make it difficult to validate data accurately and efficiently.

METHODOLOGY

The methodology provides specific details on the research process, tools, and techniques used. This includes detailed steps of the requirements gathering process, criteria for technology selection, including vendor evaluation and proof of concept, customization and integration process specifics and including mapping data fields and configuring validation rules. The implementation of automated data validation processes involves a series of structured steps aimed at ensuring a smooth transition from manual to automated operations. These steps are essential for effectively deploying automation tools and optimizing their performance within an organization. The implementation process typically encompasses several stages, each with specific tasks and objectives.

Step 1: **Requirements Gathering**

The first step in implementing automated data validation processes is to gather the requirements of key stakeholders within the organization. This involves engaging with various departments, including IT, operations, compliance, and business units, to understand their needs and challenges related to data validation. The key tasks in this stage include the following:

Stakeholder Engagement: Identify and engage with stakeholders across different departments to gather insights into their data validation requirements and pain points.

Requirement definition: Document-specific requirements and objectives for the automated data validation system, including functional and technical specifications.

Risk Assessment: Conduct a risk assessment to identify the potential risks and challenges associated with the implementation of automated data validation processes.

Step 2: **Technology Selection**

Once the requirements have been met, the next step is to select appropriate technology solutions for automated data validation. This involves evaluating different software vendors and solutions to identify the best fit for an organization's needs. Key tasks in this stage include:

Market Research: Research available automation tools and software solutions for data validation, considering factors such as functionality, scalability, and vendor reputation.

Vendor Evaluation: Evaluate potential vendors based on their offerings, track records, customer reviews, and pricing models.

Proof of Concept: Conduct a proof of concept (POC) or pilot project with selected vendors to assess the feasibility and effectiveness of their solutions in meeting the organization's requirements.

Step 3: **Customization and Integration**

Once a technological solution has been selected, the next step is to customize and integrate it into the organization's existing systems and processes. This involves configuring the software to align it with the organization's specific requirements and integrating it with other systems and databases. Key tasks in this stage include:

Customization: Customizes the selected software solution to meet the organization's unique data validation requirements, including defining validation rules, configuring workflows, and designing user interfaces.

Integration: Integrate the automated data validation system with the organization's core systems, databases, and third-party applications. This may involve the development of custom connectors or APIs to facilitate data exchange between different systems.

Testing: Comprehensive testing of the integrated solution was conducted to ensure functionality, reliability, and performance. This may include unit testing, integration testing, and user acceptance testing to validate the accuracy and usability of the system.

***Step 4:* Training and Adoption**

Once the automated data validation system is configured and tested, the next step is to train users and stakeholders regarding its use and functionality. This involves comprehensive training sessions to ensure that users are proficient in using the new system. Key tasks in this stage include:

User Training: Develop and deliver training sessions for end users, administrators, and other stakeholders involved in the data validation processes. Training may include classroom sessions, workshops, and online tutorials.

Documentation: Create user manuals, guides, and documentation to support users in understanding and effectively using an automated data validation system.

Change management: Implementing change management processes to facilitate the adoption of the new system within the organization. This may involve communication plans, stakeholder engagement activities, and feedback mechanisms to address concerns and challenges.

Step 5: Deployment and Rollout

Once users have been trained and prepared, the automated data validation system is ready for deployment and rollout across the organization. This involves implementing the system in production environments and transitioning from manual to automated data validation processes. Key tasks in this stage include Hanafizadeh et al. (2023):

Deployment Planning: Develop a deployment plan outlining the timeline, resources, and responsibilities for rolling out the automated data validation system.

Pilot Deployment: Conduct a pilot deployment or phased rollout of the system to a subset of users or departments to validate its performance and identify any issues or challenges.

Full Deployment: Once the pilot deployment is successful, proceed with the full deployment of the automated data validation system across the organization. This may involve coordination with IT teams, system administrators, and other stakeholders to ensure a smooth transition.

***Step 6:* Monitoring and Optimization**

After deployment, it is essential to continuously monitor the performance of the automated data validation system and optimize its operation to ensure ongoing effectiveness and efficiency. Key tasks in this stage include:

Performance Monitoring: Implement monitoring tools and metrics to track the performance of the automated data validation system, including processing times, error rates, and system uptime.

Issue Resolution: Address any issues or challenges identified during monitoring, such as performance bottlenecks, system errors, or data discrepancies. This may involve troubleshooting, debugging, or implementing corrective actions.

Optimization: The automated data validation system was continuously optimized to improve its efficiency, accuracy, and reliability. This may include fine-tuning validation rules, optimizing workflows, and upgrading hardware or software components as needed.

Step 7: **Continuous Improvement**

Finally, the implementation of automated data validation is an iterative process that requires ongoing refinement and improvement. Key tasks in this stage include Pramod et al. (2022):

Feedback Collection: Gather feedback from users, stakeholders, and system administrators on their experience with the automated data validation system. This may include surveys, interviews, or feedback forms to solicit input on usability, functionality, and performance.

Process Review: Conduct periodic reviews of data validation processes to identify areas for improvement and optimization. This may involve analyzing performance metrics, benchmarking against industry standards, and identifying best practices.

Iterative Development: Based on feedback and process reviews, iterate on the automated data validation system to implement enhancements, updates, and new features. This may involve releasing regular software updates, patches, or new versions of the system to address user needs and organizational requirements.

SOLUTION

To address these challenges, the bank has elected to implement automation tools for data validation processes. The bank has partnered with a technology vendor specializing in data management to develop and deploy an automated data-validation system. The diagram represents the proposed architecture for the automated data validation system. The diagram illustrates the integration of various components, such as data sources, validation engines, workflow automation tools, and monitoring systems.

Core Banking System

Customer Database Transaction Database

1. ETL←SSIS
2. Validation Engine ← Drools
3. Workflow Automation ←AWS Step ←Functions
4. Monitoring Tools

This process involved the following steps:

1. Collaborating with the technology vendor to identify the specific requirements and objectives of the automated data validation system. This assessment included evaluating the bank's existing data validation processes, identifying pain points, and defining key performance indicators (KPIs) for success.

2. Selecting a data validation software solution that offered robust capabilities for automating data validation processes, based on the requirements gathered. The selected software was seamlessly integrated with the bank's existing systems and provided features such as rule-based validation, real-time monitoring, and exception handling.

3. Customizing the selected software solution to meet the bank's specific needs and integrating it with its core banking systems and databases. This involved mapping data fields, configuring validation rules, and setting up automated workflows for data validation.

4. Conduct rigorous testing of the automated data validation system before deployment to ensure accuracy, reliability, and performance. The bank conducted both functional and user acceptance tests to validate the functionality and usability of the system.

5. Providing comprehensive training to the bank's staff after the system was deployed to familiarize them with the new automated data validation processes. The training sessions covered system functionalities, workflow procedures, and best practices for data validation.

6. Implementing monitoring tools to continuously track the performance of the automated data validation system after deployment. Regular audits and reviews were conducted to identify areas for optimization and improvement.

For automating data validation processes in a banking environment, the bank selects the tools and technologies for various aspects of the automation project.

- For data integration and ETL, Microsoft SQL Server Integration Services (SSIS) is chosen, which is a Microsoft product for creating data integration solutions.
- For the validation rules engine, Java-based rule engines (Drools, Jess) or Python libraries (pandas, NumPy) are utilized. Java is employed for implementing complex validation rules, while Python libraries are used for data manipulation and validation.
- For workflow orchestration and automation, Amazon Web Services (AWS) Step Functions is selected.

- For cloud infrastructure, AWS is chosen, which provides a wide range of cloud services, including computing, storage, databases, and analytics, and can be utilized to build scalable and cost-effective data validation solutions.
- For security and compliance, AWS Key Management Service (KMS) is applied, which is a service for creating and controlling encryption keys used to encrypt data.

Automation has several advantages, including the reduction of manual errors, which leads to improved data accuracy and integrity. Streamlined data processing and reduced turnaround times are also benefits of automation, and it can improve operational efficiency. Implementing a data validation system ensures compliance with regulatory requirements by enforcing validation rules and flagging exceptions for review. This can save costs in terms of labor and resources, as manual intervention is reduced. Accurate and timely data validation enables a bank to make informed decisions based on reliable data. Additionally, faster transaction processing and fewer errors can improve customer satisfaction and loyalty.

VALIDATING ALBANIAN IDENTITY NUMBERS

We utilized the Drools rule engine from java to validate Albanian Identity Numbers (AINs) algorithm according to the specified format and rules. The AIN is a unique personal identification number composed of 10 characters representing the individual's birthdate, sex, sequence number, and a checksum letter. Validating AIN in the bank's customer journey is essential for ensuring data accuracy, regulatory compliance, fraud prevention, and enhancing the overall customer experience. The validation process was divided into three main rules:

- AIN Format Validation: This rule ensures that the AIN matches the specified format YYMMDDSSSC, where YY represents the year, MM represents the month, DD represents the day, SSS represents the sequence number, and C represents the checksum letter. This rule uses a regular expression to check the format of the AIN.
- AIN Date Validation: This rule validates the components related to the individual's birthdate, including the year, month, and day. It checks whether these components fall within valid ranges.
- AIN Checksum Validation: This rule verifies the checksum letter of the AIN to ensure its integrity. It calculates the checksum based on the weighted sum of the AIN components and compares it with the provided checksum letter.

By implementing these rules in Drools, we can efficiently validate Albanian Identity Numbers, ensuring their accuracy and adherence to the specified format. The rule-based

approach provided by Drools allows for easy maintenance and customization of validation rules, enabling seamless integration into Java-based applications.

RESULTS AND DISCUSSION

By auto validating the data, the use of automated system eliminated most of the errors that emanate from manual operations. By having automatic validation rules and real-time monitoring, the quality of data improved through containment of errors. Introduction of automation in data discovery integrated the process, enabling fast delivery of data validation services. This excused much time-consuming paperwork for the bank staff to undertake more complex and productive activities thus boosting the overall productivity of the bank. Other positive effects observed were that the use of the new system enabled the bank to apply validation rules with ease and alert any exceptions for further evaluation, thus enhancing the overall compliance. Such an improvement enabled the enhanced delivery of customer services as there was faster and more accurate transaction processing.

The validation of Albanian Identity Numbers (AINs) using the Drools rule engine demonstrated the robustness and flexibility of the automated system.

Maintaining performance after the deployment of the automated data validation system, the bank had incorporated continuous monitoring tools to monitor the system. Some measures like how long particular processes took or the frequency that certain errors occurred or how long an entire system was online being also constantly examined. The bank had audits and reviews implemented periodically to determine if there is any place for its improvements and withstand the increased demands.

The evaluation data was obtained from the systematic questionnaires, interviews with the users and feedback forms, which helped in understanding system usage and level of performance. From this feedback section, the bank undertook regular assessments on the data validation activities for the system and made modifications or upgrades in the subsequent cycles. To this end, this continuous improvement approach checked that the automated data validation system would sustain its utility and efficiency in the duration.

The implementation of automated data validation significantly reduced errors and processing times. Key metrics include:

- Manual validation error rates expected to approximately 5%
- Automated validation reduced this to less than 0.5%.
- Manual validation took an average of 5 minutes per record, while automated validation was completed in under 30 seconds per record.

- Compliance with the regulations improved by 20% due to the consistent application of validation rules.
- Overall efficiency was improved by 30%, allowing staff to focus on more complex tasks and improving customer service.

CONCLUSION

In conclusion, implementing automated data validation processes involves a structured approach encompassing several key steps, including requirements gathering, technology selection, customization and integration, training and adoption, deployment and rollout, monitoring and optimization, and continuous improvement. By following these steps and leveraging automation tools effectively, organizations can streamline data validation processes, improve accuracy, and enhance regulatory compliance, ultimately driving operational efficiency and business success. The automation of data validation processes proved to be a game-changer for the Bank, enabling the bank to overcome challenges related to manual errors, processing delays, and regulatory compliance. By leveraging automation tools, Bank achieved significant improvements in operational efficiency, data accuracy, and regulatory compliance, ultimately enhancing its competitiveness and customer satisfaction. This case study highlights the transformative power of automation in the banking industry and underscores the importance of embracing innovative solutions to address evolving challenges in data management and regulatory compliance. The validation of Albanian Identity Numbers using Drools provides a robust and reliable solution for ensuring data integrity and accuracy in bank system that handle personal identification information. This research is relevant for the field because it relates the theoretical framework of automation and rules-based validation to a verifiable implementation in the banking industry for Albanian Identity Numbers. With Drools implemented as a rule engine for validation rules, solutions are well featured and less likely to fail. The key findings of the study provide the research with conceptual relevance as automation emerges as crucial in enhancing the accuracy of data, compliance with regulations, as well as the overall operational excellence that constitutes significant value to the banking sector.

REFERENCES

1. Al Wahshi, J. J., Foster, J., & Abbott, P. (2022). An Investigation into the Role of Data Governance in Improving Data Quality: A Case Study of The Omani Banking Sector. ECIS 2022 Research Papers, 121.
2. Paik, H.-Y., Xu, X., Bandara, H. M. N. D., Lee, S. U., & Lo, S. K. (2019). Analysis of Data Management in Blockchain-Based Systems: From Architecture to Governance. IEEE Access, 7, 186091-186107. doi:10.1109/ACCESS.2019.2961404.
3. Kumbure, M. M., Lohrmann, C., Luukka, P., & Porras, J. (2022). Machine learning techniques and data for stock market forecasting: A literature review. Expert Systems with Applications, 116659.
4. Agrawal, P., et al. (2021). Automated bank cheque verification using image processing and deep learning methods. Multimedia Tools and Applications, 80, 5319-5350.
5. Wewege, L., Lee, J., & Thomsett, M. C. (2020). Disruptions and digital banking trends. Journal of Applied Finance and Banking, 10(6), 15-56.
6. Pramod, D. (2022). Robotic process automation for industry: adoption status, benefits, challenges and research agenda. Benchmarking: An International Journal, 29(5), 1562-1586.
7. Sobczak, A. (2021). Robotic Process Automation implementation, deployment approaches and success factors–an empirical study. Entrepreneurship and Sustainability Issues, 8(4), 122.
8. Hanafizadeh, P., & Gerami Amin, M. (2023). The transformative potential of banking service domains with the emergence of FinTechs. Journal of Financial Services Marketing, 28(3), 411- 447.

ANNEX I

```
import org.kie.api.KieServices;
import org.kie.api.runtime.KieContainer;
import org.kie.api.runtime.KieSession;
public class AINValidation{
   public static void main(String[] args) {
      KieServices kieServices = KieServices.Factory.get();
      KieContainer kieContainer = kieServices.getKieClasspathContainer();
      KieSession kieSession = kieContainer.newKieSession("AINValidationSession");
      String ain = "H45509043S"; // Example AIN
      kieSession.insert(ain);
      kieSession.fireAllRules();
      kieSession.dispose();
   }
}

import java.util.regex.Pattern;
global boolean isValid;
rule "AIN Format Validation"
   when
      $ain: String(this.matches("[0-9]{10}"))
   then
      System.out.println("Rule: AIN Format Validation - Passed");
end
rule "AIN Date Validation"
   when
      $ain: String(this.matches("[0-9]{10}"))
      $yy: String(this.substring(0, 2))
      $mm: String(this.substring(2, 4))
      $dd: String(this.substring(4, 6))
      eval(Integer.parseInt($yy) > 0 && Integer.parseInt($yy) <= 99)
      eval(Integer.parseInt($mm) > 0 && Integer.parseInt($mm) <= 12)
      eval(Integer.parseInt($dd) > 0 && Integer.parseInt($dd) <= 31)
   then
      System.out.println("Rule: AIN Date Validation - Passed");
end

rule "AIN Checksum Validation"
   when
      $ain: String(this.matches("[0-9]{10}"))
      $yy: String(this.substring(0, 2))
      $mm: String(this.substring(2, 4))
      $dd: String(this.substring(4, 6))
      $sex: String(this.substring(6, 9))
      $checksumLetter: String(this.substring(9, 10))
      eval(calculateChecksum($yy, $mm, $dd, $sex).equals($checksumLetter))
   then
      System.out.println("Rule: AIN Checksum Validation - Passed");
end

function String calculateChecksum(String yy, String mm, String dd, String sex)
{
```

```
int[] weights = {2, 4, 8, 5, 10, 9, 7, 3, 6};
int sum = 0;
int datePart = Integer.parseInt(yy) * 10000 + Integer.parseInt(mm) * 100 + Integer.parseInt(dd);
int weightedSum = 0;
for (int i = 0; i < weights.length; i++)
        {
         weightedSum += Integer.parseInt(String.valueOf(datePart).substring(i, i + 1)) * weights[i];
        }
 weightedSum += Integer.parseInt(sex.substring(0, 1));
int remainder = weightedSum % 11;
if (remainder == 10) return "W";
 else return String.valueOf((char) (65 + remainder));
}
```