5-2023

ACCOUNTING AND FINANCIAL STATEMENTS AUTO ANALYSIS SYSTEM

Zhen Jia

ACCOUNTING AND FINANCIAL STATEMENTS

AUTO ANALYSIS SYSTEM

A Project

Presented to the

Faculty of

California State University,

San Bernardino

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

in

Information Systems and Technology

---

by

Zhen Jia

May 2023

ACCOUNTING AND FINANCIAL STATEMENTS

AUTO ANALYSIS SYSTEM

A Project

Presented to the

Faculty of

California State University,

San Bernardino

---

by

Zhen Jia

May 2023

Approved by:

Dr. Lewis A. Njualem, Committee Chair

Dr. Conrad Shayo, Member, Chair of the Information and Decision Sciences Department

ABSTRACT

This project was motivated by the need to revolutionize the generation of financial statements and financial analysis process thus speeding up business decision making. The research questions were: 1) How can machine learning [37] increase the speed of financial statement preparation and automate financial statements analysis? 2) How can businesses balance the benefits of automating financial analysis with potential concerns around privacy, data security, and bias? 3) Can the Java J2EE framework [32] provide a reliable running environment for machine learning?

The findings were: 1) Machine learning can significantly increase the accuracy and speed of financial analysis. Using machine learning algorithms, financial data can be processed and analyzed in real-time, allowing for quicker and more precise financial analysis. Machine learning models can identify patterns and trends in financial data that may not be easily detectable by humans, leading to more accurate financial statements and analysis. Additionally, machine learning can automate repetitive tasks in the financial analysis process, saving time and resources for businesses. 2) Businesses need to carefully balance the benefits of automating financial analysis with potential concerns around privacy, data security, and bias. While machine learning can offer significant advantages in terms of accuracy and speed, it also requires handling sensitive financial data. Therefore, it is crucial for businesses to implement robust data security measures to protect against potential data

breaches and ensure compliance with privacy regulations. Additionally, businesses need to be mindful of potential biases in machine learning algorithms, as biased algorithms can result in biased financial analysis. Regular audits and monitoring of machine learning models should be conducted to address and mitigate any potential biases. 3) The Java J2EE framework can provide a reliable running environment for machine learning. Java J2EE (Java 2 Platform, Enterprise Edition) is a widely used and mature framework for developing enterprise applications, including machine learning applications. It offers scalability, reliability, and security features that are essential for running machine learning algorithms in a production environment. Java J2EE provides robust support for distributed computing, allowing for efficient processing of large financial datasets. Furthermore, it offers a wide range of libraries and tools for implementing machine learning algorithms, making it a viable choice for running machine learning applications in the financial industry.

The conclusions were: 1) Machine learning has the potential to significantly increase the accuracy and speed of financial analysis, thereby revolutionizing the generation of financial statements and the financial analysis process. Various machine learning algorithms, such as decision trees, random forests, and deep learning algorithms, can be utilized to identify patterns, trends, and hidden risks in financial data, leading to more informed and efficient business decision making. 2) Businesses need to carefully balance the benefits of automating financial analysis with potential concerns around privacy, data

security, and bias. While machine learning can offer significant advantages in terms of accuracy and speed, there are ethical considerations that need to be addressed, such as ensuring data privacy, implementing effective data security measures, and mitigating biases in machine learning algorithms used in financial analysis. Businesses should adopt a responsible approach to machine learning implementation, considering the potential risks and benefits. 3) The Java J2EE framework can provide a reliable running environment for machine learning applications, but further research is needed to evaluate the performance and scalability of machine learning models in this framework. Identifying potential optimizations for running machine learning applications at scale in the Java J2EE framework can lead to more efficient and effective implementation of machine learning in financial analysis and decision-making processes. Further research in this area can contribute to the development of robust and scalable machine learning applications for financial analysis in the business domain.

Areas for further study include: 1) Exploring different machine learning algorithms and techniques to further improve the accuracy and speed of financial analysis. 2) Conducting research on the impact of machine learning on financial decision making and business performance. 3) Investigating methods for addressing and mitigating biases in machine learning algorithms used in financial analysis. 4) Evaluating the effectiveness of different data security measures in protecting sensitive financial data in machine learning applications. 5) Studying the performance and scalability of machine learning models in the Java J2EE

framework and identifying potential optimizations for running machine learning applications at scale.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER ONE:

INTRODUCTION

In the business career, there is one long-term topic that has been discussed for decades. The topic is how to increase the speed of business decisions such as running costs, human resources, income, marketing, and company business direction [1]. Business scientists and data analysts have done whatever they can to help businesses. This includes applying efficient working modes, using new technologies to support business processes, and researching new operation methodologies [2]. Currently, around the year 2021, machine learning using artificial intelligence has become one of the famous strategies [3], and many business scientists have significant effort on that. Because of the power of machine learning, large companies, such as Amazon, Google, and eBay, are leading the world in the provision of tailored services to their customers relatively cheaper and faster [4].

With advancements in machine learning technology, these systems can analyze vast amounts of financial data quickly and accurately, allowing businesses to make more informed decisions about their finances. So, this project will focus on how to design and implement a system that integrates with artificial intelligence technologies to increase the speed of generating financial statements and making business decisions. To design and implement the system, it needs a reliable programming language, Java 1.8 plus [5], to design and implement a flexible machine learning algorithm set.

Problem Statement

The theme of this paragraph is the challenges faced by middle-sized and small companies in making business decisions due to limited resources, and the limitations of certain tools and systems commonly used by these companies for financial analysis.

Large companies often have abundant resources, such as data analysis capabilities, research teams, and financial resources, which can be efficiently utilized with the help of new technologies [11]. Additionally, the Balanced Scorecard (BSC) model [12] is a popular framework used by businesses to evaluate performance and financial health. It incorporates financial and non-financial metrics, providing a comprehensive view of a business's performance and aiding in decision-making.

However, middle-sized, and small companies may not have the same level of resources as large companies, which can pose challenges in making business decisions efficiently and accurately. For example, while ERP systems like SAP [36] are excellent for managing accounting and finance components, they may lack analysis capabilities that can generate outcomes to support decision-makers. Similarly, tools like QuickBooks, commonly used by small and middle-sized companies for financial reporting, may not automate importing or fetching data for comparison with market values, limiting their analytical capabilities.

As a result, middle-sized and small companies may resort to using tools like Microsoft Excel to generate financial statements and merge them with online market-searched data. However, these methods can be time-consuming and require significant manpower. Thus, the paragraph highlights the limitations faced by smaller companies in terms of resources and the challenges they may encounter in making business decisions, particularly in financial analysis.

Objective

To obtain accurate and reliable results in business analysis, the data used typically comes from various sources such as data science, data mining, artificial intelligence (particularly machine learning), mathematics, and statistics [13]. This underscores the potential limitations and challenges associated with employing machine learning for financial analysis, including the need to address data accuracy, reliability, bias, privacy, and security concerns. Furthermore, the paragraph provides insights into future developments and advancements in machine learning technology that may impact financial analysis and decision-making, allowing the system to generate recommendations for company managers and utilize big data analytics to support decision-making.

Research Questions

This study attempts to answer the following research questions:

RQ. 1 How can machine learning increase the speed of financial statement preparation and automate financial statements analysis?

RQ. 2 How could businesses balance the benefits of automating financial analysis with potential concerns around privacy, data security, and bias?

RQ. 3 Can Java J2EE framework provide a reliable running environment for machine learning?

Organization of the Study

This project is structured into six chapters. Chapter 1 provides an overview of the culminating experience project. Chapter 2 reviews the literature focused on each of the research questions. Chapter 3 presents the research methods used to answer the research questions. Chapter 4 covers the project's function components, database design, and artificial intelligence calculations. Chapter 5 presents case studies. Finally, Chapter 6 contains the summary and recommendations for further study.

CHAPTER TWO:

LITERATURE REVIEW

RQ. 1 How can machine learning increase the speed of financial statement preparation and automate financial statements analysis?

The research question at hand is how machine learning can enhance the speed of financial statement preparation and automate financial statement analysis. In recent years, there has been a notable increase in interest in the use of big data analytics, workflow, software engineering design, artificial intelligence, microservices, and application programming interfaces in the context of financial statement auto-report generation. Search-Based Software Engineering (SBSE) [38] has also gained momentum as an approach to solving business problems using Search-Based Optimization (SBO) algorithms. By identifying research trends and addressing gaps in the literature, this work provides insights for further research [14].

The article titled "A review on deep learning techniques for network intrusion detection systems" provides an overview of the application of deep learning techniques for network intrusion detection systems (NIDS). The study highlights the importance of NIDS in detecting and preventing cyberattacks, and outlines the challenges associated with traditional NIDS approaches. The article reviews various deep learning techniques, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and autoencoders, that have been applied to NIDS. The study evaluates the performance of these techniques using

metrics such as detection rate, false positive rate, and processing time. The results show that deep learning techniques can provide improved accuracy and efficiency in detecting network intrusions compared to traditional NIDS approaches. The study also identifies challenges associated with the application of deep learning to NIDS, such as the need for large amounts of labeled data and the potential for adversarial attacks [44].

In conclusion, the discussed articles provide valuable insights into the application of machine learning in various contexts, including financial statement preparation, power load forecasting, microservices architecture, and network intrusion detection systems. These studies highlight the potential benefits and challenges associated with the use of machine learning techniques in these domains and suggest avenues for further research. The findings from these studies can inform decision-making and advancements in machine learning technology for improved financial analysis and decision-making in the future.

Moving forward, further research can explore the application of machine learning techniques in other domains, such as risk assessment, fraud detection, and investment prediction. Additionally, studies can investigate the development of new algorithms and methodologies to address the challenges associated with the use of machine learning in financial analysis, such as data quality, interpretability, and fairness. The continued advancement of machine learning technology has the potential to revolutionize the field of finance, and further

research in this area can contribute to the development of innovative solutions for financial analysis and decision-making.

RQ. 2 How could businesses balance the benefits of automating financial analysis with potential concerns around privacy, data security, and bias?

The automation of financial analysis through machine learning techniques offers numerous benefits, such as increased efficiency, accuracy, and decision-making. However, there are also potential concerns around privacy, data security, and bias that need to be addressed. Businesses must carefully balance the benefits of automation with these concerns to ensure responsible and ethical use of machine learning in financial analysis.

The article titled "A Comparative Study of Machine Learning Algorithms for Power Load Forecasting" provides insights into the performance of machine learning algorithms for power load forecasting in the context of an electricity distribution network in Nigeria. The study compares the performance of three algorithms - Artificial Neural Network (ANN), Support Vector Regression (SVR), and Random Forest Regression (RFR) - using historical power load data and performance metrics. The findings indicate that all three algorithms can provide accurate forecasts, with the ANN model performing slightly better. This study highlights the potential of machine learning in power load forecasting, but also emphasizes the need to consider factors [34] that can affect forecast accuracy, such as dataset size and input variables [45].

To ensure effective design and implementation of the system, the use of use case methodology can be employed. Use cases can be created on a high abstraction level or in detail, but detailed use case diagrams can become complex and challenging to comprehend. In this regard, the authors of an article introduce an automated approach to divide a complex use case diagram into smaller use case slices. This approach simplifies the requirements function and non-functional section while ensuring accuracy in the software development process, such as agile development. By breaking down a large and complex use case into smaller areas, this approach simplifies the process and ensures correctness.

In analyzing accounting and financial data, the system may rely on an artificial intelligence (AI) module. AI is defined as the system's ability to interpret external data, learn from it, and use that learning to achieve specific goals and tasks. However, businesses must also consider potential concerns around privacy, data security, and bias when implementing AI in financial analysis. Ensuring responsible and ethical use of AI is essential to mitigate potential risks and ensure compliance with relevant regulations [16].

To analyze the accounting and financial data, the system will rely on a module that includes artificial intelligence (AI). AI is typically defined as the system's ability to correctly interpret external data, learn from it, and use that learning to achieve specific goals and tasks through flexible adaptation [18].

To address concerns around bias and trustworthiness in the system, the framework can incorporate a feedback mechanism from a learning mechanism. This can help make customized selection decisions based on trustworthiness, which is a critical factor [35] in service and provider selection. The framework can be implemented as a prototype system that evaluates and recommends different cloud services, and experiments can demonstrate the effectiveness and efficiency of the proposed approach across different types of services [39] and trustworthiness metrics [19].

In conclusion, businesses need to carefully balance the benefits of automating financial analysis with potential concerns around privacy, data security, and bias. Machine learning techniques offer significant advantages in terms of efficiency and accuracy, but businesses must also consider the ethical implications and potential risks associated with their use. By implementing responsible and ethical practices, businesses can harness the power of machine learning for improved financial analysis while ensuring compliance with relevant regulations and addressing concerns around privacy, data security, and bias.

RQ. 3 Can Java J2EE framework provide a reliable running environment for machine learning?

The project at hand involves several calculation components, and the microservices strategy is well-suited for this purpose. In a Software Engineering Radio interview, Johannes Thönes and James Lewis, principal consultants at Thought Works, discuss the increasing popularity of microservices as an

architectural style. They cover topics such as deployment, size, technical considerations, and decision-making, and compare microservices with other architectures, including service-oriented architecture [15].

As the system is a business analytics platform that involves numerous calculation components for computing business data, an API-based approach will be implemented. However, in a dynamic business environment, API components often need upgrades or may become obsolete over time. To address this issue, API deprecation is necessary, which helps API users to migrate to newer versions. Nevertheless, poorly managed API deprecation may result in unnecessary expenses and negligible benefits for the integrated system.

Therefore, a cost-effective investment strategy will be employed for the API deprecation model. This model utilizes a data-driven mechanism that allows API developers to invest in API deprecation iteratively. The API deprecation strategy for this system will prioritize benefits, enabling API developers to allocate resources appropriately [17].

To ensure the reliable running environment for machine learning in the Java J2EE framework, it is essential to consider factors such as API deprecation, investment strategy, and benefits prioritization. These considerations will contribute to the effective implementation of the project and ensure the system's stability and adaptability in a dynamic business environment.

CHAPTER THREE:

RESEARCH METHODS

RQ. 1 How can machine learning increase the speed of financial statement preparation and automate financial statements analysis?

Machine learning has become a powerful tool in the field of financial analysis, offering opportunities to increase accuracy and speed in various ways. This paragraph will discuss several ways in which machine learning can be leveraged to improve financial statement preparation and automate financial statement analysis.

Firstly, predictive analytics can be used to train machine learning algorithms on historical financial data, identifying patterns and making predictions about future outcomes. This enables financial analysts to make more informed investment decisions and react to market changes in a timely manner. Secondly, automation can be employed to perform repetitive tasks such as data entry and report generation, freeing up analysts' time for more complex and strategic analysis. Thirdly, natural language processing algorithms can analyze large volumes of text data, such as financial news articles, and extract key insights that inform financial analysis, providing valuable information for decision-making. Additionally, machine learning algorithms can detect anomalies in financial transactions, helping to identify potential fraud and ensuring data integrity. Finally, machine learning can be used to analyze large datasets in real-time,

identifying potential risks and opportunities, and enabling proactive risk management.

To support the implementation of machine learning in this project, a robust database system, such as Oracle, will be used to store and manage the data and outcomes. User interfaces will be developed to allow for easy visualization and interaction with the data. Additionally, an application programming interface (API) will be utilized to facilitate data transactions between different systems, ensuring smooth communication and data flow for the machine learning algorithms [8]. To ensure the project operates in a reliable and scalable manner, a microservices-based architecture will be implemented, providing a flexible and scalable environment for the system to run 24/7 [7].

In conclusion, machine learning offers significant potential for increasing the speed and automation of financial statement preparation and analysis. Through predictive analytics, automation, natural language processing, anomaly detection, and real-time risk analysis, machine learning can enable financial analysts to make more informed decisions and streamline their workflow. However, it is crucial to implement the appropriate infrastructure, such as a robust database system, user interfaces, API, and microservices architecture, to support the successful implementation of machine learning in the project.

In addition to the technical components, a workflow module will also be included in the project to guide the application in collecting, processing, and analyzing data. This workflow will involve the participation of an accounting

auditor and a financial advisor to ensure the accuracy and reliability of the outcomes. To implement this workflow module, a BPMN [26] (Business Process Model and Notation) [10] workflow engine will be utilized, providing a structured and efficient approach to managing the workflow [9].

RQ. 2 How could businesses balance the benefits of automating financial analysis with potential concerns around privacy, data security, and bias?

In today's data-driven business landscape, automation of financial analysis has become a prevalent practice. The use of advanced technologies such as artificial intelligence (AI) and machine learning (ML) offers several benefits, including increased efficiency, accuracy, and cost savings. However, along with these advantages, there are also potential concerns around privacy, data security, and bias that need to be carefully addressed to ensure responsible and ethical use of automated financial analysis.

To answer the research question on how businesses can balance the benefits of automating financial analysis with potential concerns, appropriate research methods can be employed. One approach could be conducting a literature review to gather insights from existing academic and industry research on the topic. This could involve reviewing relevant articles, reports, and case studies that discuss the challenges and best practices related to the automation of financial analysis and its impact on privacy, data security, and bias.

Another research method could be conducting surveys or interviews with businesses that have implemented or are considering implementing automated

financial analysis. This could help gather real-world experiences and perspectives on how businesses are managing the benefits and concerns associated with automation. Interviews or discussions with industry experts, financial analysts, and data scientists could also provide valuable insights on the topic.

Additionally, analyzing relevant regulations, guidelines, and legal frameworks related to privacy, data security, and bias in financial analysis can provide a comprehensive understanding of the compliance requirements and potential risks involved in automation. This could involve reviewing laws and regulations such as the General Data Protection Regulation (GDPR), financial industry standards, and ethical guidelines for AI and ML.

By employing appropriate research methods such as literature review, surveys, interviews, and analysis of regulations, businesses can gather insights and information to make informed decisions on how to balance the benefits of automating financial analysis with potential concerns around privacy, data security, and bias. This can help businesses adopt responsible practices, mitigate risks, and ensure ethical use of automated financial analysis in their operations.

RQ. 3 Can Java J2EE framework provide a reliable running environment for machine learning?

Machine learning systems require a reliable running environment to ensure smooth operation and accurate results. To achieve this, we will adopt the

Agile methodology for development and a microservices deployment strategy [20]. The Agile methodology allows for incremental development, reducing upfront planning and design and enabling cross-functional teams to work collaboratively on smaller increments of system components. This approach minimizes risks and facilitates quick adaptation to changes, making the project implementation smoother.

Java, as a programming language, offers several advantages for hosting and integrating machine learning systems across different platforms. Its cross-platform compatibility allows for splitting tasks among different system components, which can then connect with various databases and analysis systems such as SAS, Power BI, and SAP for data import and export. Additionally, Java's core libraries provide support for custom algorithm formulas, such as depth-first search [6], which can be leveraged for the implementation of the project's requirements. With the project's information structure already in place, Java provides a suitable environment for storing data.

While the Agile methodology addresses design and implementation concerns, it may not be sufficient for ensuring a reliable and maintainable information structure in the running environment. Therefore, we will also adopt the Microservices architecture approach, which involves dividing software into small, independent services that communicate over well-defined APIs [21].

These services run individually and are owned by self-contained teams, making it easier to scale and develop applications without affecting each other.

To facilitate communication between these independent nodes, we will use APIs, which are digital communication methods that allow two or more computer programs to interact with each other [22]. APIs offer a consistent interface for the system, even if internal data changes, as they hide internal details and expose only relevant parts. This ensures that the system's running environment remains reliable and maintainable.

In conclusion, Java's J2EE framework, coupled with the Agile methodology and Microservices architecture approach, can provide a reliable running environment for machine learning systems. Java's cross-platform compatibility, support for custom algorithms, and robust libraries make it a suitable choice for hosting and integrating machine learning components. The Agile methodology enables incremental development and collaboration among cross-functional teams, while the Microservices architecture approach ensures independent and scalable services. APIs facilitate communication between system components, making the running environment consistent and maintainable.

Microservices

There are numerous claims regarding the origin of the technology term microservices. One of the earliest known uses of the term was by Fred George, who was working for the company Thoughtworks in 2004. George began researching a prototype architecture based on the "Bayesian Principles." However, it is worth noting that there were similar architectural approaches that existed before the term microservices was coined, such as Service-Oriented Architecture (SOA) and Representational State Transfer (REST) architecture. Nonetheless, microservices have become increasingly popular in recent years due to their scalability, flexibility, and ease of maintenance.

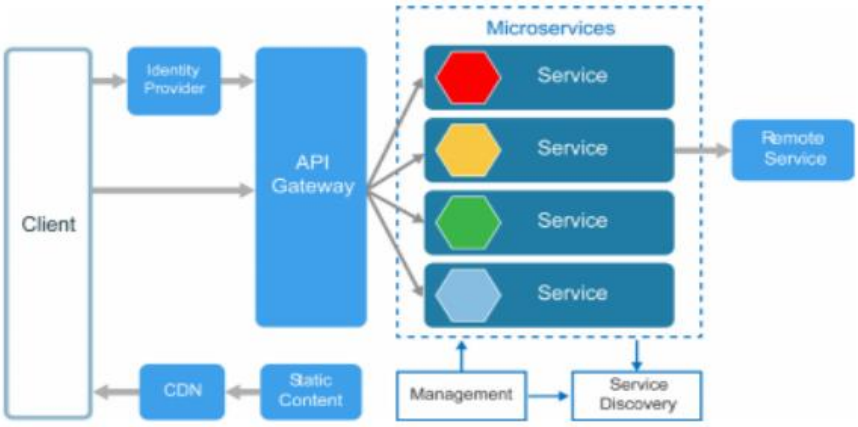

Figure 1 Microservices architectures

In 2005, Peter Rodgers introduced the term "Micro-Web-Services" during a presentation at the Web Services Edge conference, which laid the foundation for dividing a large web application into multiple function components that communicate through REST services. He further explained that microservices

are built using Unix-like pipelines, allowing for a true loose-coupling between services that can call other services, even in multiple language run-times. This enables abstracting complex service assemblies behind simple URI interfaces, allowing for any service, regardless of granularity, to be exposed.

Microservices is a specific implementation approach for service-oriented architectures (SOA) that is designed to create flexible and independently deployable software systems. It represents the evolution of SOA in response to the emergence of DevOps and is increasingly being used to build continuously deployed systems.

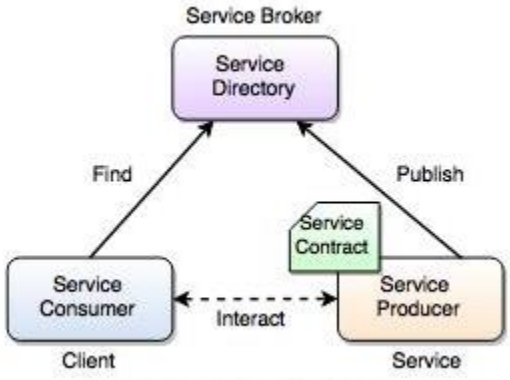

Figure 2 Service-oriented architectures

Java Programming Language

Java is a popular high-level programming language that is class-based and object-oriented. It is designed to minimize dependencies and is used for developing general-purpose software. One of the main advantages of Java is its ability to allow programmers to write code once and run it on any platform that supports Java, without the need for recompilation. This is made possible by compiling the code to bytecode that can be executed on any Java virtual machine, regardless of the underlying computer architecture. Although the syntax of Java is similar to C and C++, it has fewer low-level facilities. In 2019, Java was one of the most widely used programming languages according to GitHub [23].

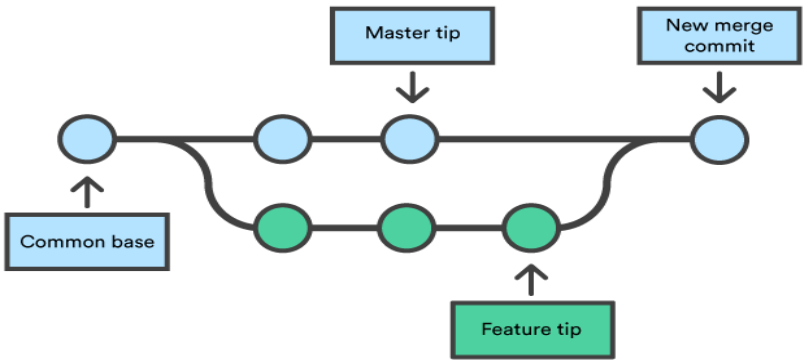

Figure 3 Project code version controller

It's correct that Java was not originally developed by Oracle, but by James Gosling at Sun Microsystems. However, there are a few inaccuracies in the rest of the statement.

Java was released in 1995 as a core component of Sun Microsystems' Java platform, but it was not initially released under a proprietary license. Instead, the first version of Java was released under a proprietary license with source code access provided for a fee.

In 2006, Sun Microsystems released Java under the GNU General Public License (GPL) with the classpath exception, allowing for open-source use and modification of Java while still retaining compatibility with proprietary software. Sun Microsystems continued to maintain Java until it was acquired by Oracle Corporation in 2010. Since then, Java has been developed and maintained by Oracle.

Java Enterprises level System Framework

Maven Plus Spring MVC

Spring MVC is a popular framework within the Java ecosystem that provides a collection of programming and configuration models to simplify the development of robust and testable web applications in Java. With Spring MVC, developers can focus on writing business logic rather than dealing with low-level infrastructure details. Spring MVC is built on top of the Spring Framework, which provides a broad range of features and functionalities, such as dependency injection, aspect-oriented programming, and data access, among others. This makes Spring MVC a powerful and versatile framework for developing web-based applications in Java.

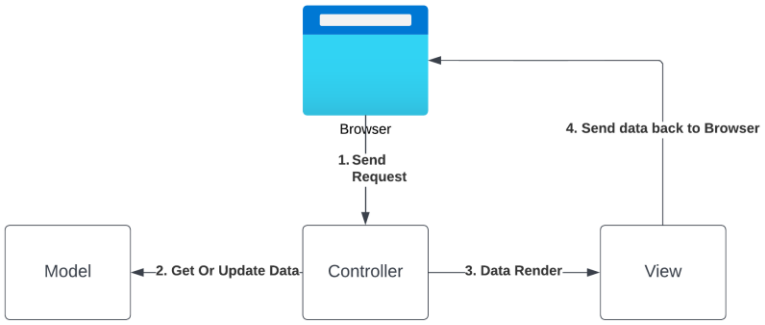

Figure 4 Spring MVC design pattern

Maven is a build automation tool primarily used for Java projects, including Java Spring. Hosted by the Apache Software Foundation, Maven simplifies the management of software builds and their dependencies. In traditional Java project management, developers must manually import Java support libraries

into their projects. This can be a time-consuming and error-prone process. However, Maven automates this process, making it easier for developers to manage their projects' dependencies. It achieves this through conventions for the build procedure, which help ensure consistency and reliability throughout the project. Overall, Maven streamlines Java project management, making it a powerful tool for developers.

Database

In the past, people stored data on paper. However, with the invention of the computer, the digital world began to take shape. To store content in a digital and structured manner, a database was created, which is an organized collection of structured data stored electronically in a computer system. As computers became more prevalent, the need for them to store a larger volume of data for fast retrieval increased. In the 1970s, Dr. Ted Codd, a computer scientist, invented the relational model for database management. This model solved many of the issues that arose from the flat file model. According to the relational model, data is organized into entities and attributes, rather than being combined into a single structure. This made it easier to retrieve specific information and perform complex queries. Overall, the relational model revolutionized database management and laid the foundation for modern databases.

Artificial Intelligence

Since this project involves an artificial intelligence component, it is important to discuss one of the most famous topics in this field: machine learning. To understand machine learning, it's necessary to first understand artificial intelligence. The term "artificial intelligence" typically refers to machines that exhibit cognitive skills associated with the human mind, such as "learning" and "problem-solving". However, this definition has been rejected by many leading AI researchers. Today, AI is more commonly defined in terms of rationality and the ability to act rationally, rather than simply mimicking human cognition. Machine learning is a subset of artificial intelligence that focuses on enabling machines to learn from data without being explicitly programmed. With machine learning, machines can identify patterns and relationships in data and use them to make predictions or decisions. Machine learning is a rapidly growing field that is transforming many industries and has the potential to revolutionize the way we live and work [24].

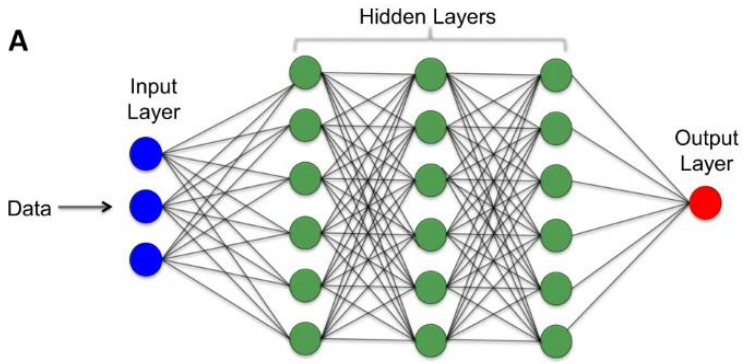

Figure 5 Neural network model

Machine learning is a branch of biotechnology that involves developing machine algorithms that can think and learn in ways like humans to solve problems. It is also an area of research focused on creating methods that "learn," or improve performance on specific tasks by leveraging data. Machine learning is a subset of artificial intelligence, and it uses data to build models that can make predictions or decisions without explicit programming. These models are trained on large sets of data called training data. Machine learning algorithms are used in a wide range of applications, including medicine, email filtering, speech recognition, and computer vision. They are particularly useful when it is difficult or impossible to develop conventional algorithms to perform a given task. Overall, machine learning is a rapidly advancing field with numerous applications in industry, research, and daily life [25].

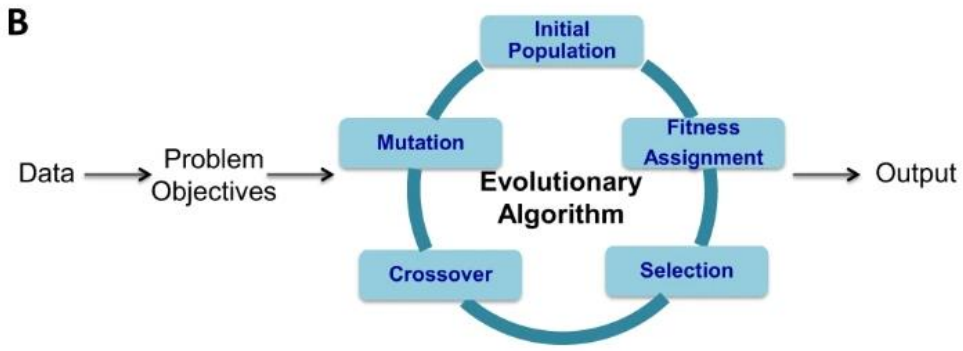

Figure 6 Machine learning model

CHAPTER FOUR:

DESIGN AND IMPLEMENTATION

Machine learning is a subset of artificial intelligence (AI) that involves developing algorithms and models that enable computer systems to learn and make predictions based on data inputs without being explicitly programmed. In the context of financial statement analysis, machine learning algorithms can be trained on historical financial data to identify patterns and make predictions about future financial outcomes. This can help financial analysts make more accurate and informed investment decisions. Machine learning can also automate repetitive tasks such as data entry, reconciliation, and report generation, freeing up financial analysts' time to focus on more complex tasks. Other applications of machine learning in finance include fraud detection, risk management, and natural language processing.

Deep learning is a subset of machine learning that involves developing artificial neural networks with multiple layers of interconnected nodes. Deep learning algorithms can process vast amounts of data inputs and learn to make increasingly complex predictions and decisions. In the context of financial statement analysis, deep learning algorithms can be used to identify patterns and trends in large datasets of financial data, enabling financial analysts to make more accurate predictions about future financial outcomes. Deep learning can also be used for natural language processing, enabling financial analysts to

extract key insights from large volumes of financial news articles and other textual data.

A Java application for financial statement analysis using machine learning and deep learning could be designed and implemented using a microservices [42] architecture approach. This approach involves dividing the software into small, independent services that communicate over well-defined APIs, allowing for easier scaling and faster development. The Java core libraries could be used to support custom algorithm formulas, such as depth-first search, and the project could be hosted and integrated across different platforms. The system components could be split into smaller, specific functions and then merged, with different system components connecting with different databases and analysis systems, such as SAS, Power BI, and SAP, to import and export data automatically. A relational database, such as Oracle, could be used to store the machine learning outcomes and data. The application program interface could be used to facilitate data transactions between different systems, and a workflow module could be implemented using a BPMN workflow engine to guide the application to collect, process, and calculate data. Overall, the Java application could leverage machine learning and deep learning to increase the accuracy and speed of financial analysis, enabling financial analysts to make better-informed decisions and stay ahead of the competition.

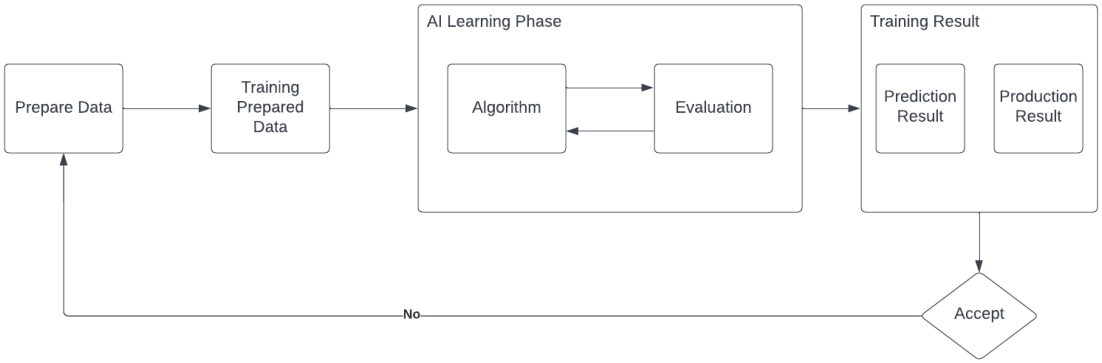

Figure 7 AI training workflow processing flow

BPMN technology provides flexible workflows and three static logic flows, making it an ideal solution for generating and modifying workflows using a BPMN flow tool. In my project, I will utilize this technology for report generation workflows. To control the three static logic flows, namely data collection logic flow, business catalog chooses logic flow, and machine learning logic flow, I will use a flow controller. The flow controller, based on the BPMN workflow, will determine the type of report and the data required when customers make their requirements. Customers will also be able to modify the workflow through the Customer's management platform.

As for the project construction, it is crucial to decompose all data collection and algorithms into code fragments and store them in a database based on the

technology I have chosen.

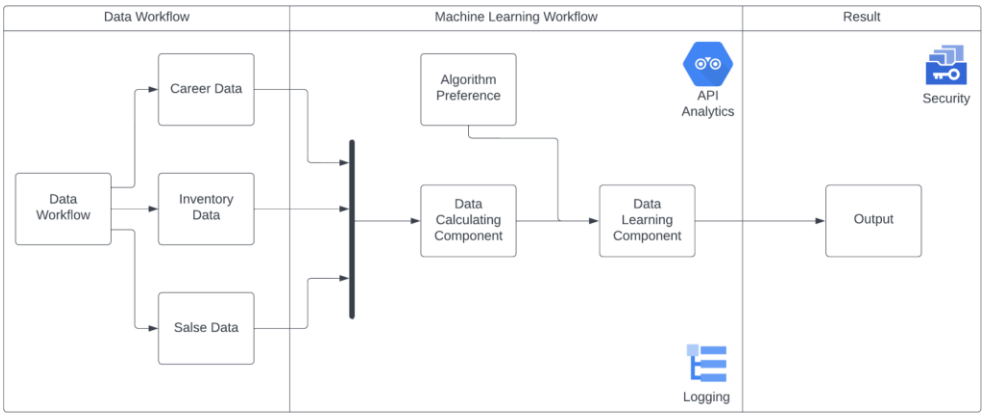

Figure 8 System results generating workflow

Decision Brain's AI cash flow [27] budgeting model uses a "supervised learning" approach to create machine learning models based on historical data and predict future traffic for daily cash flows over the next 3 months [10]. The process includes:

- Identifying groups of cash flows with similar rules and breaking them down for more precise results through more machine learning exercises and time to tune and maintain models if required to produce higher granularity of analysis.
- Identifying the characteristics and rules for cash flow forecasting, such as a time, day of the week or day of the month.
- Selecting the training period, which is the period of historical data used to create the model.

- Analyzing data outliers, assigning new features and either excluding them from the model or creating a separate machine learning model. This requires additional responsiveness adjustments in the face of unexpected events such as COVID-19 and post-COVID-19 periods.
- Finding the right machine learning algorithm for the company's data to provide the lowest error.
- Running tests and comparing model results with manual results by dividing the mean absolute error by the mean of the actual data.

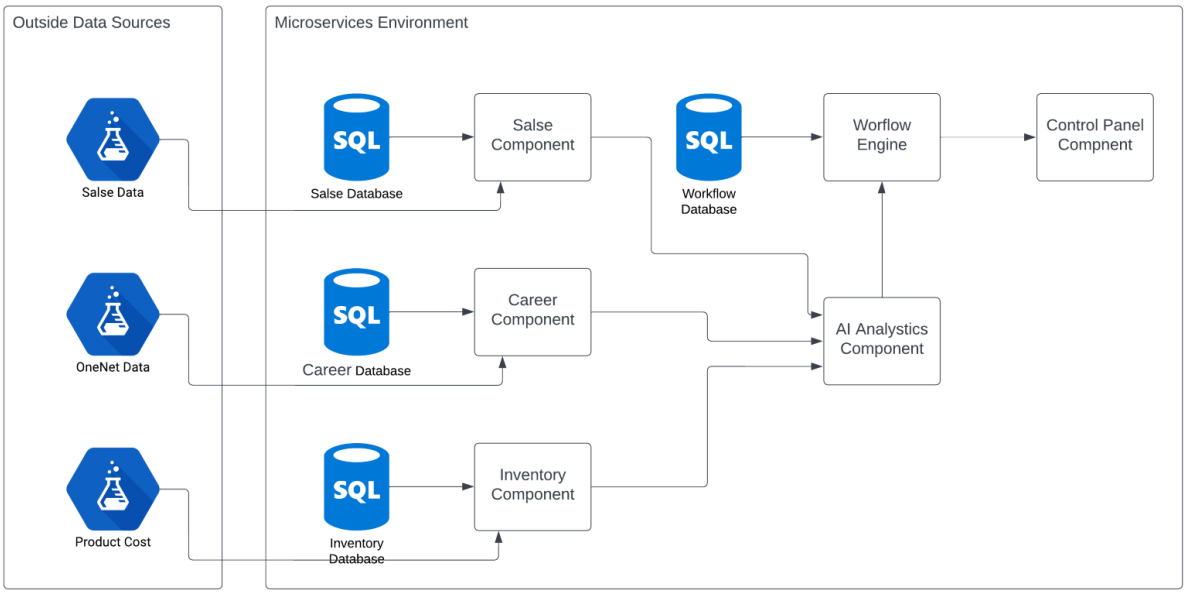

Figure 9 System microservices structure

In addition, BPMN can guide customers in completing their first business catalog report module, report data controller, Email controller, and Database controller. Java 8.0 serves as the basic development language, while HTML [28],

jQuery [29], Tomcat [24], Apache [31], Java SH (Spring and Hibernate) framework [1][24], and Maven are used together. The report generator offers text boxes, area text boxes, submit buttons, reset buttons, customized buttons, search areas, single and slide image boxes, text labels, sub-windows, and section areas on the web application page.

The concept of business income on the balance sheet [33] is widely understood as an expense that boosts income. While this formula can be effective, it can be time-consuming and manual. Additionally, a company's revenue largely determines its budget for years to come. To stay ahead and adapt to market trends, companies can leverage AI by training models using large balance sheets of varying models and scales to predict future years. This approach can help CFOs and stakeholders focus their efforts on increasing revenue.

Financial data analysis involves using mathematical models to forecast conditional expected returns on assets. Typically, asset price movements are modeled as a stochastic process. Machine learning provides a new way to calculate this data and get the expected return on assets within a risk decomposition framework. Researchers have proposed using a nonlinear function to simulate the correlation between financial risk and expected asset returns, thereby recovering the nonlinearity of market data in a model-free manner. For instance, k-means clustering is used to classify analysis data, and a

simple functional form is used to approximate the nonlinear relationship between future and current asset returns.

With the constant expansion of financial data, companies have access to more information than ever before, along with new analytical techniques that can provide faster and better insights into the data. Machine learning, which combines computer science and financial statistics, can be applied to a range of forecasting tasks, including financial fraud detection and future financial forecasting. One of the key considerations when using machine learning is data anomalies, which can be planned and reallocated to understand the impact of active intervention in the system. Pattern recognition is a significant area of machine learning that requires extensive external data. For instance, the level of employee salaries can be modeled using onenet's industry data to predict future patterns. Additionally, recently developed dual machine learning frameworks reduce the impact of imperfect model specification in causal analysis.

Maven

According to Wikipedia, Maven was initially developed to simplify the complex build processes in the Jakarta Turbine project, where multiple projects had their own Ant build files with slight differences and JARs were checked into CVS. Maven aims to provide Java developers with a comprehensive understanding of the development effort in a shorter period. To achieve this, Maven addresses several areas of concern, such as making the build process

easy, providing a uniform build system, offering quality project information, providing guidelines for best practices development, and enabling transparent migration to new features.

Maven serves as a Java-based project constructor and management tool. It allows for an overview of the project construction, including the collection of Java JARs and embedded Tomcat components. In contrast, if another project constructor like Ant is used, library references must be uploaded and imported manually. By using Maven, developers can streamline their workflow and save time on manual tasks, allowing them to focus on developing their applications.

MVC Design Pattern

The Model-View-Controller (MVC) design pattern is a widely used software engineering approach for developing lightweight Java-based projects. This pattern divides a web-based application into three interconnected components, namely the model, view, and controller. The Spring framework plays a significant role in connecting the model and controller components.

The model component is responsible for managing data, logic, and rules of the application, and provides a data manager role. The view component provides the output representation of information, such as a chart or a diagram. It is responsible for presenting information to the user in a way that is visually appealing and easy to understand. The controller component acts as a mediator

between the model and view components. It accepts input from the view layer and then converts it into commands for the model.

One of the key advantages of using the MVC design pattern is that it separates the internal representations of information from the ways that information is presented to or accepted from the customers. This allows for greater flexibility and ease of maintenance, as changes to one component do not affect the others. Additionally, using a framework like Spring simplifies the process of connecting and managing these components, making it easier to develop and maintain Java-based projects.

Spring Framework

The Spring Framework is a powerful application framework and inversion of control container designed for the Java platform. It provides a wide range of core features that can be utilized by any Java application, with extensions available specifically for building web applications on top of the Java EE platform [30]. While the Spring Framework itself does not dictate any specific programming model, it has gained tremendous popularity and has been widely adopted by the Java community.

One of the key advantages of the Spring Framework is its ability to offer a comprehensive programming and configuration model for Java-based web applications. With this, developers can easily embed various components such

as Hibernate [40], BPMN engine, and MVC design pattern into their applications with relative ease. To achieve this, components are first created as objects and then injected into the Spring XML file. This feature makes it easier to manage complex web application projects, as the Spring Framework provides a streamlined approach to building and configuring applications.

Hibernate Framework

Hibernate ORM (shortened to hibernate) is a tool for mapping between object-oriented models and relational databases in the Java language. Its core function is to provide a framework for transforming an object-oriented domain model into a relational database model. By using high-level object handling functions, hibernate solves the problem of object-relational impedance mismatch, which results from direct and persistent database accesses.

Hibernate's primary feature is the ability to map Java classes to database tables and map Java data types to database data types. Additionally, it facilitates data query and retrieval operations by automatically generating SQL calls and handling result set conversion.

Using hibernate to manage a database is much simpler than using direct database connection methods because it allows developers to map Java entity elements to database table elements using annotations in each entity class, which can then be injected into the hibernate configuration file. Without

hibernate, developers would need to manually implement the database connection method and connection pool, execute queue, and map data.

Representational State Transfer

Representational State Transfer (REST) or RESTful Web services is a way to enable interoperability between computer systems on the internet. REST-compliant web services allow requesting systems to access and manipulate textual representations of web resources using a uniform and predefined set of stateless operations. These operations include HTTP verbs such as GET, POST, PUT, and DELETE [13].

REST is often used as a connection between the main system and a BPMN engine. Compared to SOAP, REST requires less work and returns a JSON string to the client, displaying the content on the customer screen. SOAP, on the other hand, requires a client side as a message receiver, adding to the development workload. As a result, using REST can save time and effort in comparison to developing a SOAP connection.

CHAPTER FIVE:

CASE STUDIES

In conclusion, there are several areas for further study in the field of machine learning for financial analysis. Exploring different machine learning algorithms and techniques can lead to improved accuracy and speed of financial analysis. Research on the impact of machine learning on financial decision making and business performance can provide valuable insights for decision-makers. Addressing and mitigating biases in machine learning algorithms is critical to ensure fair and unbiased financial analysis. Evaluating the effectiveness of data security measures in protecting sensitive financial data is essential to safeguard against potential breaches. Additionally, studying the performance and scalability of machine learning models in specific frameworks, such as Java J2EE, can lead to identifying optimizations for running machine learning applications at scale.

Furthermore, case studies play a crucial role in answering specific research questions, and despite the scarcity of case studies in business data analytics in financial statements, they can provide valuable insights. The first case study in this research focuses on decision trees, which are a common method of machine learning with various algorithms introduced to overcome limitations and improve accuracy [43]. Decision trees can be used to identify financial ratios and analyze a company's financial health. The second case study

discusses the use of random forest algorithm, which combines decision trees to improve analysis accuracy and identify patterns and trends in financial data. Deep learning algorithms, based on artificial neural networks [41], are also powerful tools for financial statement analysis as they can analyze large amounts of financial data and predict future performance based on historical data.

In addition to machine learning, the integration of big data analysis in the real estate industry has significant implications for marketing strategies. Real estate companies can leverage big data technology to collect and analyze consumer data, understand customer expectations, and develop targeted marketing methods. Building high-precision marketing channels through network facilities and collaborating with external professional data collection companies can provide valuable information for data analysis and research, leading to improved marketing models and competitiveness in the market.

Traditional methods of consumer data collection through sample surveys have been replaced by more advanced big data technology. Real estate companies must focus on selecting the right technology with strong collection ability and high work efficiency. As younger generations become the primary consumers in the real estate market, companies need to strengthen their integration with big data analysis technology, using data mining and other emerging technologies to better understand their expectations and develop targeted marketing methods.

The article "Artificial Intelligence for the Real World" published in the Harvard Business Review discusses how organizations can successfully implement artificial intelligence (AI) to achieve business benefits. The authors suggest that many companies fail to achieve the full potential of AI because they approach it as a technology problem rather than a business problem.

Define your strategic objectives: Companies should identify specific business problems that can be solved by AI, and prioritize them based on their potential impact; Develop a data strategy: Companies must ensure that they have the necessary data and infrastructure to support AI initiatives, and identify areas where data quality can be improved; Build AI models: Companies must identify the appropriate AI models for their business problems, and develop them using the right data and algorithms; Deploy AI models: Companies should integrate their AI models into their existing business processes and ensure that they are delivering the desired results; Learn and refine: Companies must continually monitor and refine their AI models to ensure that they are delivering value over time.

Overall, further research in machine learning algorithms, addressing biases, data security measures, scalability, and the integration of big data analysis in the real estate industry can contribute to the advancement of financial analysis and marketing strategies, leading to more informed decision making and improved business performance.

CHAPTER SIX:

DISCUSSION, RECOMMENDATIONS AND AREAS FOR FURTHER STUDY

The business analytics market is currently dominated by traditional consulting firms that analyze a company's data to generate marketing reports. However, this process is often costly and time-consuming, making it difficult for small and medium-sized businesses to access these services. This project aims to address this issue by using cutting-edge technologies like artificial intelligence and machine learning to create a general business analytics system that can be easily implemented by businesses of any size.

The system works by allowing customers to choose a business template and then modify or add their own preferences into the AI module. Alternatively, customers can select a minimized configuration setting that only requires them to choose a business catalog. An integrated BPMN engine generates reports based on these preferences, which customers can modify through an administration platform if necessary. The system also allows customers to merge multiple templates to generate future marketing trends.

Based on the findings of this research, there are several areas for further study to enhance the accuracy, speed, and overall effectiveness of the financial analysis system. Firstly, exploring different machine learning algorithms and techniques could further improve the accuracy and speed of financial analysis. This could involve experimenting with various algorithms such as decision trees,

neural networks, or ensemble methods to determine which ones yield the best results for financial data.

Secondly, conducting research on the impact of machine learning on financial decision making and business performance could provide valuable insights. This could involve studying how businesses are incorporating machine learning into their decision-making processes, the benefits, and challenges they encounter, and the impact on their overall performance and financial outcomes.

Thirdly, investigating methods for addressing and mitigating biases in machine learning algorithms used in financial analysis is crucial. Bias in machine learning algorithms can lead to inaccurate results and potentially biased decision-making. Therefore, exploring techniques such as fairness-aware machine learning or developing bias mitigation strategies could help improve the fairness and reliability of the financial analysis system.

Additionally, evaluating the effectiveness of different data security measures in protecting sensitive financial data in machine learning applications is essential. Data security is a critical concern in any machine learning system, especially when dealing with financial data. Therefore, studying and implementing various data security measures such as encryption, access controls, and authentication methods can enhance the security and confidentiality of the financial analysis system.

Finally, studying the performance and scalability of machine learning models in the Java J2EE framework and identifying potential optimizations for

running machine learning applications at scale is crucial. As the financial analysis system is built using Java J2EE framework, understanding the performance and scalability of machine learning models in this environment can help optimize the system for large-scale data processing and improve its overall efficiency.

In conclusion, this research project aims to create a general business analytics system using artificial intelligence and machine learning that can be easily implemented by businesses of any size. The system allows customers to choose business templates, modify preferences, and generate reports based on their requirements. Further areas of study include exploring different machine learning algorithms, researching the impact of machine learning on financial decision making, addressing biases in machine learning algorithms, evaluating data security measures, and studying the performance and scalability of machine learning models in the Java J2EE framework. These areas of research can contribute to the continuous improvement and enhancement of the financial analysis system, making it more accurate, reliable, and efficient in supporting businesses' decision-making processes.

REFERENCES

[1] Glenn Hopper. *Using Artificial Intelligence To Improve Business Decisions*. Forbes Finance Council. Jun 21, 2022. Available: https://www.forbes.com/sites/forbesfinancecouncil/2022/06/21/using-artificial-intelligence-to-improve-business-decisions/?sh=338c391ab494

[2] Kao, Chia Hung; Chang, Cheng-Ying; Jiau, Hewijin Christine. Information and software technology. Vol. 142, p.106746.

[3] Ang, A. (2013) Factor Investing [Online]. Available: https://doi.org/10.2139/ssrn.2277397

[4] L. Sommerville. *Software Engineering, 9th*, Pearson. 2011.

[5] N. S. Williams, *Professional Java for Web Applications,* John Wiley & Sons, Incorporated, February 2014.

[6] (2008, November 11). Practical Artificial Intelligence Programming With Java [Online]. Available: https://resources.saylor.org/wwwresources/archived/site/wp-content/uploads/2011/11/CS405-1.1-WATSON.pdf.

[7] Zhengping Wu, Yu Zhou. *Customized Cloud Service Trustworthiness Evaluation and Comparison Using Fuzzy Neural Networks. IEEE 40th* Annual Computer Software and Application Conference, 2016.

[8] Boca Raton, *Handbook of enterprise operations management*, Fla: Auerbach, 1999.

[9] Cash Flow Forecasting Using Machine Learning for a Retail Group [Online]. Available: https://decisionbrain.com/cash-flow-forecasting.

[10] B. Rücker, *Real-Life BPMN: Using BPMN 2.0 to Analyze, Improve, and Automate Processes in Your Company*, 2 edition, CreateSpace Independent Publishing Platform, Dec. 5, 2014.

[11] Chen, Yangyang; Chen, Cong; Wen, Hao; Jiang, Jian-min; Zeng, Qiong ; Shu, Hongping; Zhong Hong. *From Use Case to Use Case Slices: An Automated Approach. International journal of information system modeling and design.* Vol.10(4), p. 24-50, 2019.

[12] *Balanced scorecard*, Wikipedia [Online]. Available: https://en.wikipedia.org/wiki/Balanced_scorecard.

[13] Parul Gandhi, Surbhi Bhatia, Abhishek Kumar, Mohammad Alojail, Pramod Singh Rathore, *Realization of Business Intelligence using Machine Learning*, Wiley Online Library, 04 January 2021.

[14] Harman, Mark., Mansouri, S.Afshin., Zhang, Yuanyuan. *Search-based software engineering: Trends, techniques and applications*. ACM computing surveys, Vol. 45(1), p.1-61, 2012.

[15] Thones, Johannes. Microservices. *IEEE software, Vol.32(1*), p. 116-116, 2015.

[16] Chen, Yangyang; Chen, Cong; Wen, Hao; Jiang, Jian-min; Zeng, Qiong ; Shu, Hongping; Zhong Hong. *From Use Case to Use Case Slices: An*

Automated Approach. International journal of information system modeling and design. Vol.10(4), p. 24-50, 2019.

[17] G. Zambon and M. Sekler, *Beginning JSP™, JSF™, and Tomcat Web Development: From Novice to Professional*, New York: Apress, 2007.

[18] Haenlein, Michael; Kaplan, Andreas. *A Brief History of Artificial Intelligence: On the Past, Present, and Future and Artificial Intelligence. California management review*. Vol. 61(4), p.5-14, 2019.

[19] John Watson, Roopesh Ramklass, Bob Bryla, *OCA/OCP Oracle Database 12c*, Mc Graw Hill Education, 2015.

[20] John Watson, Roopesh Ramklass, Bob Bryla, *OCA/OCP Oracle Database 12c*, Mc Graw Hill Education, 2015.

[21] (2022, October 20). REST [Online]. Available: https://en.wikipedia.org/wiki/Representational_state_transfer.

[22] (2022, October 20). Maven [Online]. Available: https://en.wikipedia.org/wiki/Maven.

[23] (2022, October 20). Github [Online]. Available: https://en.wikipedia.org/wiki/GitHub.

[24] Mohammed Uddin, Yujiang Wang, Marc Woodbury-Smith, *Artificial intelligence for precision medicine in neurodevelopmental disorders*, NPJ Digital Medicine, November 2019.

[25] Yen-Lin Lee;Pei-Kuei Tsung;Max Wu, *Techology trend of edge AI*, Publisher: IEEE, 2018.

[26] B. Silver, *Bpmn Method and Style, 2nd Edition, with Bpmn Implementer's Guide: A Structured Approach for Business Process Modeling and Implementation Using Bpmn 2*, Cody-Cassidy Press, Oct. 17, 2011.

[27] Cash Flow Forecasting Using Machine Learning for a Retail Group [Online]. Available: https://decisionbrain.com/cash-flow-forecasting.

[28] *HTML & CSS: design and build websites*, Indianapolis: John Wiley & Sons, Inc., 2014.

[29] J. Duckeet, *JavaScript & jQuery interactive front-end web development*, Indianapolis: John Wiley & Sons, Inc, 2014.

[30] N. S. Williams, *Professional Java for Web Applications,* John Wiley & Sons, Incorporated, February 2014.

[31] B. Laurie and P. Laurie, *Apache the definitive Guide 3rd*, O'REILLY, Dec.2002.

[32] A. Leonard, *JBoss Tools 3 Developers Guide,* Packet Publishing Ltd, April 2009.

[33] Beez Labs, Balancing your Balance sheets with Artificial Intelligence and Machine Learning [Online]. Available: https://www.linkedin.com/pulse/balance-sheets-ai-beez-innovation-labs/?trk=articles_directory

[34] Ang, A. (2013) Factor Investing. https://doi.org/10.2139/ssrn.2277397

[35] Homescu, C. (2015) Better Investing through Factors, Regimes and Sensitivity Analysis. https://doi.org/10.2139/ssrn.2557236

[36] (2022, October 20). SAP [Online]. Available: https://en.wikipedia.org/wiki/SAP.

[37] Helmut Wasserbacher, Martin Spindler, Machine learning for financial forecasting, planning and analysis: recent developments and pitfalls, Digital Finance, 4, pages63–88 (2022).

[38] Harman, Mark., Mansouri, S.Afshin., Zhang, Yuanyuan. *Search-based software engineering: Trends, techniques and applications*. ACM computing surveys, Vol. 45(1), p.1-61, 2012.

[39] (2022 April 25). Muminjon [Online]. Available: https://blogs.embarcadero.com/how-to-use-cross-platform-programming-languages-expectations-vs-reality.

[40] (2022, October 20). Hibernate (framework) [Online]. Available: https://en.wikipedia.org/wiki/Hibernate_(framework).

[41] Haenlein, Michael; Kaplan, Andreas. *A Brief History of Artificial Intelligence: On the Past, Present, and Future and Artificial Intelligence. California management review*. Vol. 61(4), p.5-14, 2019.

[42] Thones, Johannes. Microservices. *IEEE software, Vol.32(1*), p. 116-116, 2015.

[43] Ang, A. (2013) Factor Investing. https://doi.org/10.2139/ssrn.2277397

[44] Huijue Kelly Duan, Miklos A. Vasarhelyi, Mauricio Codesso, Zamil Alzamil. *Enhancing the government accounting information systems using social media information: An application of text mining and machine learning*.

International Journal of Accounting Information Systems. Volume 48, March 2023.

[45] Seethiraju L. V. V. D. Sarma, Dorai Venkata Sekhar, Gudipatu Murali. *Stock market analysis with the usage of machine learning and deep learning algorithms*. Bulletin of EEI Stats. Vol 12, No 1.