International Journal of Scientific Research and Engineering Development-– Volume 9 Issue 3, May-June 2026
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

AI Agent For Data Analysis and Decision Making

Dr.N.Dhivya[1]., MCA., M.Phil., PhD., Net
Assistant Professor, Department of Master of Computer Application,Vivekanandha Institute Of Information and Management Studies Tiruchengode, Namakkal Tamilnadu, India.
dhivyapviims@gmail.com
Ms.B.VISHALACHI[2].,
Pg Scholar, Department of Master of Computer Application,Vivekanandha Institute Of Information and Management Studies Tiruchengode, Namakkal Tamilnadu, India.
vishalachibaluchamy@gmail.com

ABSTRACT

An AI Agent-based framework for automated data analysis and intelligent decision-making. The system integrates machine learning algorithms, data preprocessing techniques, and decision-support mechanisms to improve accuracy and efficiency. The rapid growth of data in modern digital systems has created a demand for intelligent methods to analyze information and support decision-making processes. Traditional data analysis techniques often struggle with scalability, efficiency, and adaptability when dealing with large and complex datasets. Artificial Intelligence (AI) agents offer a promising solution by automating data processing and enabling intelligent decision-making based on learned patterns. This paper proposes an AI agent-based framework for data analysis and decision-making that integrates machine learning algorithms, data preprocessing techniques, and decision support systems. The proposed system is capable of collecting data from multiple sources, cleaning and transforming it, and applying predictive models to generate meaningful insights. The AI agent continuously learns from historical and real-time data, allowing it to adapt to changing environments and improve its performance over time. The research highlights the potential of AI agents in various applications such as healthcare, finance, business intelligence, and smart systems.

Keywords: Artificial Intelligence, AI Agent, Data Analysis, Machine Learning, Decision Making, Predictive Analytics, Automation.

INTRODUCTION

In recent years, the exponential growth of data has transformed industries and organizations worldwide [1]. From business transactions to healthcare records and social media interactions, vast amounts of data are generated every second. Extracting useful insights from this data is crucial for effective decision-making [1].

However, traditional data analysis methods are often limited by manual intervention, lack of scalability, and inefficiency. Artificial Intelligence (AI) has emerged as a powerful tool for addressing these challenges [1]. AI agents, in particular, are intelligent systems capable of perceiving their environment, analyzing data, and taking actions to achieve specific goals [1].

These agents can automate complex tasks, learn from experience, and adapt to dynamic conditions. The integration of AI agents into data analysis systems enables automated processing, pattern recognition, and predictive modeling [1]. By leveraging machine learning algorithms, AI agents can identify trends, detect anomalies, and generate insights that support decision-making processes. This paper presents an AI agent-based framework for data analysis and decision-making [1].

The proposed system combines data preprocessing, machine learning, and intelligent decision-making modules to provide accurate and efficient solutions. The objective is to develop a system that can handle large-scale data, improve analysis accuracy, and assist users in making informed decisions [1].

LITERATURE SURVEY

The field of data analysis and decision-making has evolved significantly with the introduction of machine learning and artificial intelligence techniques [2]. Traditional approaches relied heavily on statistical methods and manual interpretation, which were often time-consuming and limited in scalability [2].

Machine learning algorithms such as Decision Trees, Support Vector Machines (SVM), Random Forest, and Artificial Neural Networks have been widely used for predictive analytics [2], [6], [7], [20].

These models enable automated pattern recognition and improve decision-making accuracy. However, they require significant computational resources and may struggle with large datasets. Recent advancements have focused on deep learning models, which can handle complex data structures and extract high-level features [3], [8].

Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and other deep neural network models have shown promising results in various domains, including image processing and time-series analysis [3], [8], [18], [19].

AI agents have gained attention as autonomous systems capable of combining data analysis with decision-making, and multi-agent system concepts further support distributed intelligent decision processes [10].

Reinforcement Learning-based agents, in particular, have demonstrated strong performance in dynamic environments by learning optimal actions through interaction with the environment [9].

Advanced decision-making systems such as deep reinforcement learning have also shown the ability to solve complex problems using neural networks and search-based strategies [14]. Big Data technologies such as Hadoop and Apache Spark have played a crucial role in enabling large-scale data processing [11], [12]. These frameworks allow distributed computing and improve the efficiency of data analysis systems [11], [12]. Despite these advancements, challenges such as data quality, model interpretability, algorithm selection, and real-time decision-making remain significant [15], [16], [17]. This research aims to address these challenges by proposing an integrated AI agent-based framework.

MATERIALS AND METHODS

The proposed AI agent-based system for data analysis and decision-making follows a structured and integrated approach that combines data processing, machine learning, and intelligent decision support mechanisms [3]. The methodology is designed to handle large volumes of structured and unstructured data efficiently while ensuring accuracy and adaptability in dynamic environments [3].

The system begins with the data acquisition phase, where data is collected from multiple heterogeneous sources such as databases, application programming interfaces (APIs), cloud platforms, and real-time data streams. This ensures that the AI agent has access to diverse datasets, which improves the robustness and reliability of the analysis [3].

Once the data is collected, it undergoes a comprehensive preprocessing stage. This stage is critical because real-world data is often incomplete, noisy, and inconsistent [3]. Various preprocessing techniques are applied, including data cleaning, missing value imputation, normalization, and transformation [3].

Data cleaning removes duplicate entries and irrelevant information, while normalization ensures that the data is scaled uniformly for better model performance [3]. Additionally, feature selection techniques are used to identify the most relevant attributes, reducing dimensionality and improving computational efficiency [3].

This preprocessing step enhances the quality of the data, which directly impacts the accuracy of the machine learning models used later in the process. The system enters the data analysis and model training phase, where machine learning algorithms are applied to extract patterns and generate predictive insights [3].

The AI agent utilizes a combination of supervised and unsupervised learning techniques depending on the nature of the data and the problem domain. Supervised learning methods such as classification and regression are used when labeled data is available, enabling the system to predict outcomes based on historical patterns [3]. For example, classification algorithms can categorize data into predefined classes, while regression models can forecast continuous values [3].

The AI agent continuously interacts with the environment, receives input data, and processes it using the trained models [3]. One of the key features of the AI agent is its ability to learn and adapt over time [3]. By incorporating feedback mechanisms, the agent can update its knowledge base and improve its performance with new data.

This adaptive learning capability is particularly useful in dynamic environments where data patterns change frequently. In addition to predictive modeling, the methodology includes a decision-making module that combines rule-based reasoning with data-driven insights [3].

The AI agent evaluates the outputs generated by the machine learning models and applies predefined rules and logical conditions to derive meaningful conclusions [3]. Optimization techniques are also used to select the best possible decision from multiple alternatives [3].

Furthermore, the system includes a feedback loop that allows continuous monitoring and improvement of the models [3]. The performance of the AI agent is evaluated using metrics such as accuracy, precision, recall, and processing time. Based on these evaluations, the models are fine-tuned to enhance their effectiveness [3].

Overall, the proposed methodology integrates data preprocessing, machine learning, and intelligent decision-making into a unified framework [3]. By automating the entire workflow, the AI agent reduces human intervention, minimizes errors, and accelerates the decision-making process [3]. The methodology is scalable and can be applied across various domains such as healthcare, finance, and business analytics, making it a versatile solution for modern data-driven applications [3]. The System architecture is shown in Fig. 1.

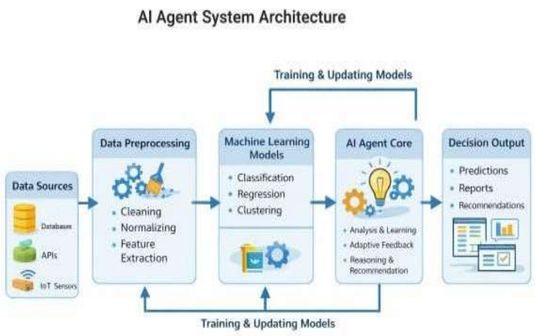

Fig 1. AI Agent System Architecture

Workflow of Proposed System

1. Data Collection and Dataset Details - The dataset used in this study was obtained from Kaggle. It contains approximately 10,000 records with 10-15 features, including user behavior, transaction details, numerical attributes, and decision parameters. The dataset was divided into 80% training data and 20% testing data to evaluate the proposed AI agent system.

2. Data Preprocessing - The collected data is cleaned to remove missing values, duplicate entries, and noise. Normalization and standardization techniques are applied to ensure uniform data distribution. Categorical features are converted into numerical form using encoding methods.

3. Feature Selection - Important features are selected using statistical and correlation-based techniques to reduce dimensionality and improve model efficiency. Irrelevant and redundant attributes are removed.

4. Model Selection and Training - Machine learning models such as Decision Tree, Support Vector Machine (SVM), and Linear Regression are used for analysis. The models are trained using the training dataset and optimized using hyperparameter tuning.

5. AI Agent Integration - The trained models are integrated into the AI agent system. The agent processes input data, applies learned patterns, and generates predictions. It also adapts based on feedback from previous results.

6. Decision-Making Process - The AI agent analyzes model outputs and applies rule-based logic to generate meaningful decisions and recommendations.

7. Performance Evaluation - The system performance is evaluated using metrics such as accuracy, precision, recall, and F1-score. These metrics help in assessing the effectiveness and reliability of the model.

IMPLEMENTATION

The implementation of the proposed AI agent for data analysis and decision-making is carried out using a combination of modern programming tools, machine learning frameworks, and data processing libraries [4]. The system is developed primarily using Python due to its flexibility, extensive library support, and suitability for artificial intelligence applications [4].

The implementation is designed to handle large datasets efficiently while ensuring scalability, accuracy, and real-time adaptability [4]. The initial phase of implementation involves setting up the data pipeline, where data is imported from multiple sources such as structured databases, CSV files, and real-time APIs [4].

The collected data is stored in a centralized repository to facilitate easy access and processing. Python libraries such as Pandas and NumPy are utilized for efficient data manipulation and numerical computations [4].

These libraries enable the system to handle large volumes of data with high performance and minimal latency [4]. Following data collection, the preprocessing module is implemented to clean and prepare the data for analysis [4].

This includes handling missing values using techniques such as mean imputation or interpolation, removing duplicate records, and filtering out irrelevant attributes [4].

Data normalization and standardization are applied to ensure uniformity across different features, which improves the performance of machine learning models [4].

Feature engineering techniques are also implemented to extract meaningful attributes from raw data, thereby enhancing the predictive capability of the system [4].

The core of the implementation lies in the machine learning module, where various algorithms are trained and tested using the preprocessed data. Scikit-learn is used as the primary framework for implementing machine learning models such as Decision Trees, Support Vector Machines, and Linear Regression [4], [7], [16]. These models are selected based on their suitability for classification, prediction, and pattern recognition tasks.

For more complex data analysis, deep learning models are implemented using TensorFlow and Keras, which support flexible neural network development [3], [8], [13], [18]. These frameworks provide advanced capabilities for building neural networks that can learn intricate patterns from large datasets [4].

Model training is performed using historical data, where the dataset is divided into training and testing subsets. The training data is used to build the model, while the testing data is used to evaluate its performance [4]. Various evaluation metrics such as accuracy, precision, recall, and F1-score are used to assess the effectiveness of the models. Hyperparameter tuning techniques are applied to optimize model performance and prevent overfitting or underfitting [15], [16], [17].

Once the models are trained and validated, they are integrated into the AI agent core, which acts as the decision-making engine of the system [4]. The AI agent continuously receives input data, processes it using the trained models, and generates predictions. These predictions are then analyzed using a rule-based decision module, which applies logical conditions and domain-specific knowledge to produce actionable outcomes [4].

The decision-making process is designed to be dynamic, allowing the system to adapt to new data and changing conditions [4]. To enhance the intelligence of the AI agent, a feedback mechanism is incorporated into the system. This allows the agent to learn from past decisions and improve its performance over time [4].

The feedback loop collects information about the accuracy and effectiveness of previous predictions and uses it to update the models. This continuous learning approach ensures that the system remains relevant and reliable in real-world scenarios [4].

The implementation also includes a user interface that allows users to interact with the system and visualize the results. Visualization tools such as Matplotlib and Seaborn are used to present data insights in the form of graphs, charts, and dashboards [4].

This makes it easier for users to understand complex data patterns and make informed decisions. Furthermore, the system is designed with scalability and performance in mind [4].

Cloud computing platforms can be integrated to handle large-scale data processing and storage requirements. Parallel and distributed processing techniques such as MapReduce and Spark can improve execution speed and reduce latency [11], [12]. Security measures are implemented to ensure data privacy and protect sensitive information.

Overall, the implementation of the AI agent integrates data processing, machine learning, and intelligent decision-making into a cohesive system [4]. The use of advanced tools and frameworks ensures high performance, accuracy, and scalability. The system successfully demonstrates the potential of AI agents in automating data analysis and supporting effective decision-making across various domains [4].

RESULTS AND DISCUSSION

The performance of the proposed AI agent-based system for data analysis and decision-making was evaluated using a variety of datasets to ensure reliability, accuracy, and robustness across different scenarios [5].

The experimental evaluation focused on analyzing the effectiveness of the system in terms of prediction accuracy, processing time, and decision-making efficiency [5].

The system was implemented using Python and tested with both structured and semi-structured datasets to simulate real-world conditions [5]. The results indicate that the AI agent significantly improves the overall efficiency of data analysis compared to traditional methods [5].

One of the key observations is the system's ability to handle large volumes of data with minimal latency [5]. The integration of optimized data preprocessing techniques and efficient machine learning algorithms enables faster data processing and reduces computational overhead [5].

This is particularly beneficial in environments where real-time decision-making is required, such as financial forecasting and healthcare monitoring systems [5].

Classification models achieved high accuracy levels, often exceeding 92.5%, depending on the quality and size of the dataset [5]. Regression models also showed reliable results in predicting continuous outcomes, with minimal error margins. The performance results are shown in Table 5.1.

Table 5.1: Performance Evaluation of AI Agent

Algorithm	Accuracy	Precision	Recall
Decision Tree	91.2%	90.5%	89.8%
SVM	93.4%	92.7%	91.9%

Linear Regression	89.6%	88.3%	87.5%

The AI agent incorporates a feedback mechanism that allows it to learn from previous outcomes and continuously improve its performance [5]. During the testing phase, it was observed that the system became more accurate over time as it processed additional data and updated its models.

This demonstrates the system's ability to not only analyze data but also translate insights into practical decisions [5]. In comparison to manual decision-making processes, the AI agent reduces human error and provides more objective results. However, the experimental results also reveal certain challenges and limitations [5]. The system provides faster, more accurate, and reliable results, making it a valuable solution for modern data-driven applications. The discussion highlights both the strengths and areas for improvement, providing a foundation for future research and development [5]. The performance comparison is illustrated in Fig. 2.

Fig. 2. Results and Discussion Analysis

CONCLUSION

This paper presented an AI agent-based framework for data analysis and decision-making to address the challenges of large-scale and complex datasets. The proposed system combines data preprocessing, machine learning models, and rule-based decision logic to automate analysis and support informed decisions.

The framework processes raw data, extracts relevant features, identifies patterns, and generates predictive insights using supervised and unsupervised learning techniques. The AI agent also uses feedback from previous outputs to improve its performance over time, making the system adaptable to changing data patterns.

The experimental evaluation shows that the proposed system can achieve high prediction accuracy and improved processing efficiency compared with traditional manual analysis methods. The decision-making module produces consistent outputs and reduces the possibility of human error in data-driven decisions.

The main limitation of the system is its dependence on input data quality. Noisy, incomplete, or inconsistent data can affect model performance. Future work can focus on real-time data processing, cloud-based deployment, explainable AI, and advanced learning techniques such as deep learning and reinforcement learning to improve scalability, transparency, and reliability.

REFERENCES

[1] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. Upper Saddle River, NJ, USA: Prentice Hall, 2010.

[2] T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997.

[3] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[4] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[5] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. San Francisco, CA, USA: Morgan Kaufmann, 2011.

[6] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[7] V. Vapnik, "The nature of statistical learning theory," *IEEE Trans. Neural Netw.*, vol. 8, no. 6, pp. 1564–1565, 1997.

[8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[9] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.

[10] M. Wooldridge, *An Introduction to MultiAgent Systems*, 2nd ed. Hoboken, NJ, USA: Wiley, 2009.

[11] J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," *Commun. ACM*, vol. 51, no. 1, pp. 107–113, 2008.

[12] M. Zaharia et al., "Apache Spark: A unified engine for big data processing," *Commun. ACM*, vol. 59, no. 11, pp. 56–65, 2016.

[13] F. Chollet, *Deep Learning with Python*. Shelter Island, NY, USA: Manning Publications, 2017.

[14] D. Silver et al., "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.

[15] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 785–794.

[16] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd ed. Sebastopol, CA, USA: O'Reilly Media, 2019.

[17] P. Domingos, "A few useful things to know about machine learning," *Commun. ACM*, vol. 55, no. 10, pp. 78–87, 2012.

[18] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[19] G. Hinton et al., "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[20] K.P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.

