Journal of Autonomous Intelligence (2024) Volume 7 Issue 5
doi: 10.32629/jai.v7i5.1087

ORIGINAL RESEARCH ARTICLE

Anomaly detection in Smart Traffic Light system using blockchain: Securing through proof of stake and machine learning

Shamneesh Sharma, Nidhi Mishra*

Department of Computer Science & Engineering, Kalinga University, Naya Raipur 492101, India

*** Corresponding author:** Nidhi Mishra, mishra.nidhi1980@gmail.com

ABSTRACT

The Smart Traffic Light system plays a crucial role in smart cities, contributing significantly to enhancing the overall urban living standards, supporting sustainable practices, assuring public safety, and optimizing operational efficiency. Smart lighting systems leverage advanced technology such as the Internet of Things (IoT), data analytics, and networking to construct a complex and flexible lighting infrastructure. Anomalies within the context of a Smart Light system's data pertain to patterns, events, or behaviors that are unexpected or uncommon, displaying a considerable deviation from the system's typical operational state. The distributed and tamper-resistant ledger of blockchain technology renders it well-suited for maintaining an open and immutable record of data and events pertaining to smart lighting systems. The integration of blockchain technology and anomaly detection techniques enables the establishment of a resilient and reliable smart lighting system. The primary objective of this study is to investigate the application of the Isolation Forest algorithm for anomaly detection and its enhancement through the implementation of Proof of Stake for enhanced security measures. The Isolation Forest algorithm is utilized for anomaly prediction and evaluation of accuracy and precision. This is achieved by employing synthetic ground truth labels. Subsequently, the non-anomalous data is incorporated into the system as blocks using blockchain technology. The experimentation involves the use of synthetic numerical data derived from a dataset accessible on Kaggle. This process is conducted in two distinct phases. The first phase focuses on the analysis of synthetic numerical data, specifically targeting the identification of anomalies, prior to the implementation of blockchain technology. In the second phase, the same technique is applied to the synthetic numerical data following the integration of blockchain technology. Various machine learning (ML) models were utilized to analyze the data, resulting in improved accuracy from 48% to 94%, as evidenced by the validation of the obtained results. The empirical results indicate that the utilization of blockchain technology in data processing leads to an improvement in accuracy. Furthermore, the random forest algorithm exhibits robust performance when integrated with blockchain technology.

Keywords: smart cities, Smart Traffic Lights, information security, blockchain, machine learning

ARTICLE INFO

Received: 2 August 2023
Accepted: 24 November 2023
Available online: 20 March 2024

1. Introduction

The rapid advancement of technology has resulted in the metamorphosis of cities into intelligent and interconnected urban landscapes. Incorporating intelligent lighting systems represents a fundamental component of the digital revolution[1]. The implementation of intelligent lighting solutions offers numerous benefits, encompassing enhanced energy efficiency, reduced operational expenditures, heightened safety measures, and improved quality of life for community residents. Smart lighting systems that utilize Internet of Things (IoT) technology enable cities to effectively manage their energy consumption and provide adaptable lighting based on real-time data[2]. This integration of IoT technology ensures seamless connectivity and efficient management of lighting systems.

Significant challenges arise because of the sophistication and dependency of Smart Light Systems, in particular with regard to issues of safety[3]. Traditional security approaches that rely on a centralized structure are susceptible to cyberattacks due to the flaws that emerge from having a single point of failure in each method[4]. A breach in the security of a smart light system can have catastrophic ramifications, including the cessation of illumination services and large increases in the risk of accidents involving automobiles and pedestrians[5]. To effectively mitigate these significant safety concerns, it is imperative to seek innovative solutions that prioritize safeguarding the integrity of data transmitted through smart lights, while also guaranteeing its confidentiality. The identification of irregularities in data from the Internet of Things (IoT) is of utmost importance in guaranteeing the security, effectiveness, and dependability of the management infrastructure for intelligent traffic systems. Internet of Things (IoT) devices collect up-to-date information regarding the state of roads, the flow of vehicles, and various environmental factors[6]. The identification of anomalies has the potential to mitigate hazards, alleviate congestion, enhance traffic flow optimization, and ultimately enhance road safety and efficiency. Various techniques and approaches encompass the utilization of cameras, sensors, and other related devices. The common anomalies which can be there in IoT networks for Smart Light Systems are addressed and described briefly in **Table 1**.

Table 1. Anomalies in IoT fetched data for smart light systems in smart cities.

Sr. No.	Anomaly	Description	Research reference
1	Unusual Traffic Patterns	Anomalies may be found in the dataset used to estimate traffic during the data analysis and visualization stage. Anomalies or unexpected occurrences in the smart lighting system may be indicated by unusual spikes or dips in traffic that do not follow regular patterns. To comprehend their underlying causes and the consequences for the security paradigm, these anomalies may merit additional examination.	[7]
2	Outliers in Security Data	Anomalies may appear while examining outliers in the dataset's numerical properties that pertain to security. These abnormalities may point to security data discrepancies, such as unauthorized access attempts or unusual lighting consumption habits. Maintaining the integrity and dependability of the security of the smart lighting system depend on identifying and fixing these abnormalities.	[8] [9]
3	Blockchain Consensus Failures	Potential anomalies in the form of consensus failures may appear during the testing phase of the blockchain-based security paradigm. The integrity of the blockchain may be compromised if the consensus process is unable to bring nodes to a consensus on the authenticity of transactions. Such oddities can be indicators of consensus mechanism weaknesses that need more research and optimization.	[10] [11]
4	Blockchain Tampering efforts	To gauge the resilience of the model during testing, intentional blockchain tampering efforts may be simulated. Any unauthorized changes or additions to the blockchain would be seen as an anomaly, perhaps indicating assaults or security holes in the smart lighting system.	[12]
5	Inconsistent Correlations	Anomalies in the correlations may be seen when visualizing the correlation matrix between various parameters. Unexpected correlations or the absence of correlations between specific variables may point to anomalies or hidden linkages in the data that might affect the performance and accuracy of the security model.	[13] [4]
6	Performance bottlenecks in Scalability Testing	Performance abnormalities may show up as the blockchain network is being scaled. Performance issues, such as longer transaction times or resource constraints, may be signs that implementing the security model in real-world smart city systems with increasing data volumes may be difficult.	[14]

The present research introduces an innovative security model based on blockchain technology, specifically designed for smart lighting systems, to address the aforementioned challenges. The introduction of blockchain technology, initially associated with cryptocurrencies, has garnered attention across various sectors due to its inherent characteristics of decentralization, transparency, and immutability[15].

The blockchain operates as a decentralized and immutable ledger, offering secure and auditable data storage and transmission, thereby eliminating the requirement for a central authority. By utilizing blockchain

technology, it is possible to improve the reliability and resilience of smart lighting systems inside smart cities, which makes it easier to design a secure and reliable security architecture[16]. This can be accomplished through the implementation of blockchain technology. The technique that was used in this study takes a methodical approach to the design, implementation, and analysis of the security model that is based on blockchain technology. The first step in reproducing the operation of an intelligent lighting system is to import and then preprocess a relevant dataset on traffic prediction. Both the data analysis capabilities of the Pandas library and the data visualization skills of Seaborn are utilized in the process of analyzing the dataset. The purpose of this study is to unearth the underlying patterns and correlations included within the information and ultimately to locate important parameters that are essential for the security model. The subsequent stage encompasses the development of the architectural framework of the blockchain network, encompassing a comprehensive elucidation of its fundamental constituents. The fundamental constituents in question are commonly referred to as blocks, transactions, and cryptographic hashing algorithms. This subsequent phase aims to facilitate the advancement of a resilient and tamper-resistant blockchain technology capable of efficiently safeguarding vital security data furnished by the smart lighting system. To achieve the objective of developing a blockchain, two distinct consensus mechanisms, namely Proof-of-Work (PoW) and Proof-of-Stake (PoS), have been investigated and incorporated as integral components of this study. The optimal strategy for fulfilling the requirements of the intelligent lighting security model was determined through a comprehensive examination of the advantages and disadvantages associated with each consensus approach. To validate the efficacy of the blockchain-based security strategy, exhaustive testing and simulations must be conducted. To evaluate the network's resistance to potential cyberattacks and manipulation, synthetic data simulating various traffic and illumination scenarios are generated[17]. This procedure is conducted to determine the extent of network resilience. A comprehensive evaluation is conducted to evaluate the model's efficacy and scalability in diverse urban settings. The purpose of this evaluation is to determine whether the model can effectively manage the growing volume of data and transactions in a smart city environment.

This study offers a thorough investigation of the convergence of anomaly detection, blockchain technology, and machine learning in a specific domain of a Smart Traffic Light System. The main focus of this research is the incorporation of a resilient blockchain framework, with a particular emphasis on utilizing the Proof of Stake consensus mechanism to strengthen security measures. Additionally, researchers have employed machine learning techniques to reinforce the anomaly detection capabilities of traffic light systems. The integration of these technologies seeks to address the changing complexities associated with safeguarding vital assets and guaranteeing the dependability of smart city applications. The next section undertakes an in-depth examination of the proposed framework. Initially, it presents a comprehensive exposition of the selected Proof of Stake consensus mechanism and its utilization in safeguarding our blockchain-oriented system. Subsequently, this paper explains the integration of machine learning algorithms, highlighting their crucial function in effectively detecting and addressing abnormalities inside an intelligent traffic signal system. The objective is to assess the efficacy of our proposed solution in relation to conventional anomaly detection approaches, utilizing empirical evaluations and case studies. As the investigation progresses, it becomes evident that our methodology not only improves the security stance of the Smart Traffic Light system, but also boosts its performance through the implementation of real-time anomaly detection. The combination of blockchain with machine learning in a comprehensive manner represents a notable advancement in enhancing the dependability and robustness of smart city infrastructure, with a specific focus on the crucial area of traffic control.

2. Literature review

Anomaly detection plays a vital role in the context of smart cities, as it contributes to the optimization of efficiency and security through the analysis of various data streams, including sensor data, traffic patterns, and

energy use. One of the key benefits of this technology is its ability to enhance security and surveillance measures by the detection and identification of atypical behaviors, such as unauthorized access, loitering, or the presence of suspicious things. The development of smart lighting systems is an ongoing process aimed at enhancing both energy efficiency and operational capabilities inside urban environments sometimes referred to as smart cities[18]. One of the prevailing developments in the field pertains to linked lighting, wherein the lighting systems may be remotely controlled. This advancement is made possible using Light Emitting Diode (Light Emitting Diode) technology, which is known for its energy-efficient characteristics. Additionally, adaptive lighting systems have also gained prominence as part of the ongoing trends[19]. The integration of sensors, such as motion detectors and occupancy sensors, enables lighting systems to exhibit intelligent behavior in response to the surrounding environment. Predictive maintenance and optimization are facilitated through the application of advanced analytics to the data gathered from smart lighting systems. The development of human-centric lighting aims to replicate the patterns of natural light to improve human well-being and productivity. The current status of smart cities and smart lighting is characterized by a dynamic nature, since continuous research and development efforts are actively contributing to its evolutionary progress[20]. Traffic management leverages deviations in traffic patterns to enhance the flow of vehicles and increase operational effectiveness. Anomaly detection has been employed in environmental monitoring to identify atypical air quality, noise levels, and other environmental characteristics. Anomaly detection is employed in smart grids to discern deviations in energy-consumption patterns, thus signifying possible flaws or instances of tampering. Anomaly detection is employed in infrastructure monitoring to assess the well-being and efficiency of the essential infrastructure, thereby identifying deviations at an early stage to mitigate future breakdowns.

This section serves as an introduction to the background of work undertaken in the field of Security in Smart Systems using Blockchain. One recent study claim that these is a need of integration of blockchain technology in smart cities[21]. The researchers conducted a thorough examination of existing literature to discern the relevance of incorporating blockchain within the smart city framework, with a primary focus on enhancing security measures. The rationale behind this investigation likely stems from the growing complexity and interconnectivity of smart city infrastructures, where various components such as IoT devices, sensors, and data networks interact seamlessly to provide efficient urban services. Some of the researchers have also advocated for its integration in supply chain management to improve the overall performance of smart cities[22]. One more recent study of 2023 claims that there is a need of blockchain based information security framework for smart cities[23]. This investigation employed the Latent Dirichlet Allocation (LDA) technique as a methodological framework to identify and forecast the dominant research patterns in the security field of smart cities. The Latent Dirichlet Allocation (LDA) technique is frequently utilized in the fields of natural language processing and machine learning for the purpose of topic modeling. It facilitates the discovery of hidden topics present in a collection of documents.

Solutions in literature have pursued a variety of methods for anomaly detection in IoT-based architectures[24]. A novel approach for anomaly detection and a lightweight blockchain-based framework is proposed to build an IoT detection model in a distributed setting by Yisroel Mirsky, Tomer Golomb and Yuval Elovici[25]. This research presents a novel approach utilizing blockchain technology to address the issue of autonomous collaborative anomaly detection within a vast network of Internet of Things (IoT) devices. The researchers have tested this approach on a distributed IoT simulation environment to show how it can enhance network and device safety. Conceptually identical work, utilizing a similar method, was proposed in 2018 by Tsuyoshi Ide[26]. This paper introduced a collaborative framework for anomaly detection on Blockchain, which involves the extension of Smart Contract functionality to accommodate noisy sensor data. The task was formalized as a multi-task probabilistic dictionary learning approach, which aimed to tackle challenges related to validation, consensus building, and data privacy concerns.

To enhance the human experience, the implementation of secure and autonomous technologies is imperative for the development of smart cities. There is a need for improvement in the areas of centralization, reliability, and data integrity. The utilization of blockchain-based architectural solutions, coupled with the implementation of outlier detection mechanisms, serves to ensure the preservation of data integrity, immutability, and availability[27]. This technology effectively safeguards data and mitigates instances of vehicle-related criminal activities. Machine learning is employed in various applications such as traffic tracking, analysis of criminal behavior, and accident detection to effectively identify anomalies within datasets. A concurrent work[28] has also considered a similar approach based on traffic monitoring, criminal activity profiling, accident detection and reporting. A comparable study was conducted in the domain of smart homes for the purpose of anomaly detection[29]. In this study, the researchers put forth a trustworthy system based on machine learning for the environment of IoT-based smart homes. The interconnectedness of numerous internet-enabled devices gives rise to potential privacy concerns, necessitating the implementation of stringent measures to establish trust within the network. Another study was carried out in the year 2019 which employs big data and machine learning techniques to detect abnormal behaviors within smart home settings[30]. The utilization of a Hidden Markov Model (HMM) involves the training of said model on network sensor data, resulting in a notable accuracy rate of 97% in the detection of attacks.

Linear regression, decision tree regressor, and random forest regressor are often employed machine learning models for the purpose of regression analysis[31]. The utilization of these techniques can be employed in the context of anomaly detection within a Smart Traffic Light System. This involves the prediction of anticipated traffic circumstances and subsequently discovering anomalies by evaluating deviations from these predicted patterns[32]. Linear regression models are utilized to represent the association between input variables and anticipated traffic flow. On the other hand, Decision Tree Regressor is employed to capture non-linear associations between input features and traffic conditions[33]. Anomalies are identified when the observed traffic conditions exhibit substantial deviations from the anticipated values. The Random Forest Regressor algorithm expands upon the decision tree methodology by amalgamating numerous decision trees, hence improving both accuracy and generalization[34]. The process of anomaly detection encompasses the establishment of suitable thresholds, regular model updates, the aggregation of predictions from numerous models, and the integration of the system with the intelligent traffic light management system[35]. The selection of a particular model is contingent upon the intricacy of the relationships present within the data, as well as the desired level of interpretability and robustness of the model[36].

The existing body of scholarly research pertaining to anomaly detection in smart city infrastructures, with a specific focus on traffic management systems, has discovered several inherent constraints. Some of the factors that need to be considered are the susceptibility to cyber threats, difficulties in scaling, and the ability to respond in real-time. To tackle these concerns, the study suggests a novel incorporation of blockchain technology, notably the Proof of Stake consensus mechanism, and machine learning algorithms within the Smart Traffic Light System. The aforementioned methodology improves security by employing a decentralized and tamper-resistant structure. It also boosts scalability by dividing the workload throughout a decentralized network. Additionally, it facilitates real-time anomaly detection by utilizing machine learning methods. The objective of this technique is to address the latency issues that have been identified in prior research and enhance the resilience of the Smart Traffic Light System against anomalies.

3. Methodology

The inclusion of blockchain technology into anomaly detection systems can yield both advantageous and disadvantageous effects on performance. Although blockchain technology offers enhanced security and an unalterable ledger, it might potentially create challenges related to latency and scalability. These challenges

may impede the performance of time-sensitive applications such as smart lighting systems. Figure 1 describes the methodology used for inculcating blockchain technology in Smart Light Systems for anomaly detection.

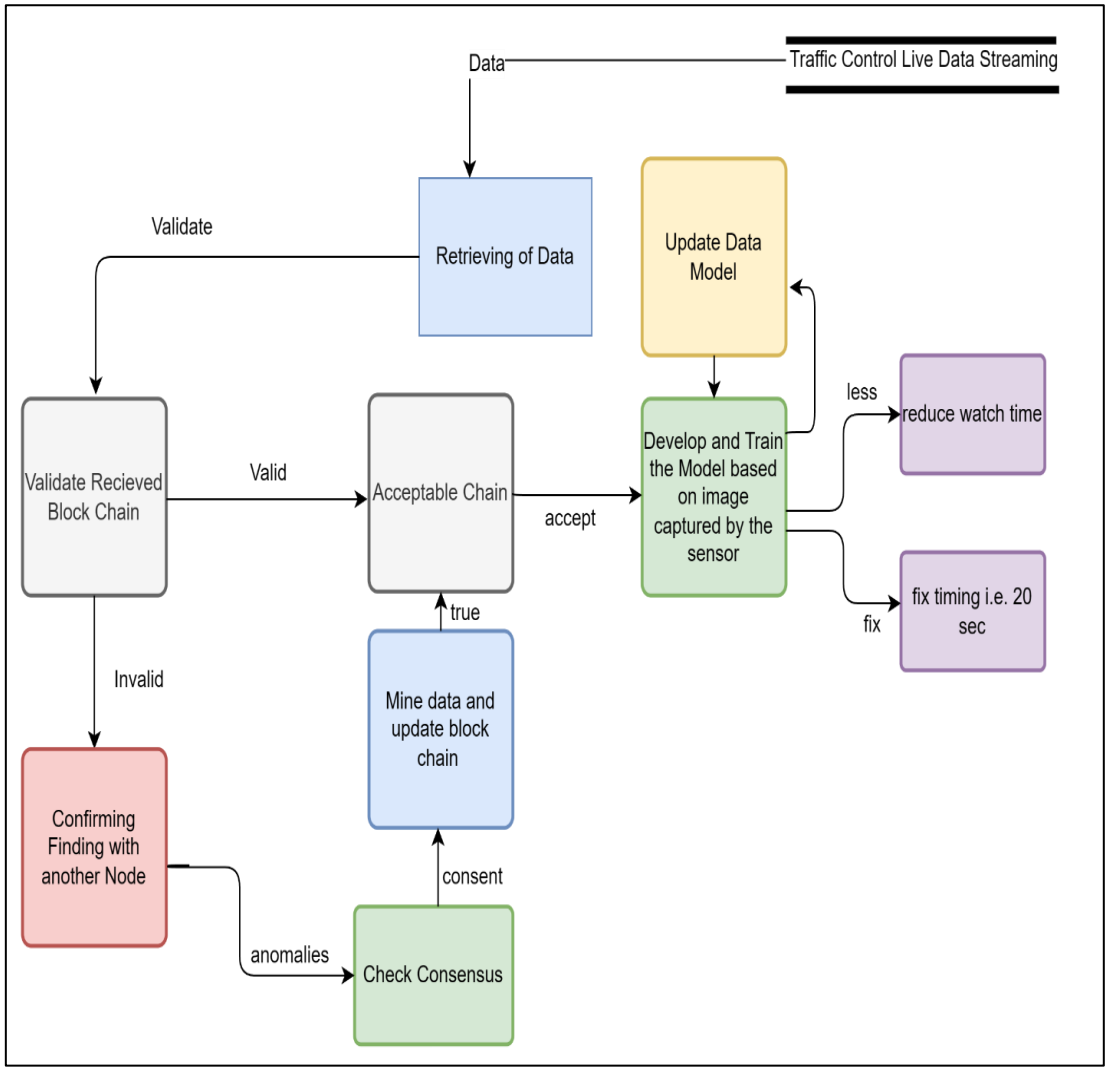

Figure 1. Process of including blockchain in Smart Traffic Lights for anomaly detection.

Traditional blockchains sometimes impose a substantial burden on IoT devices that possess constrained processing capabilities and limited energy resources. Moreover, the incorporation of blockchain technology brings a level of intricacy that has the potential to affect the effectiveness of anomaly detection in terms of efficiency. Mitigation measures encompass many techniques like off-chain processing for jobs that require timely execution, hybrid methodologies, optimized consensus processes, localized blockchains, and dynamic adjustment mechanisms. Achieving a harmonious equilibrium between the advantages offered by blockchain technology and the imperative for swift data processing and efficient data transfer is of paramount importance. To enhance the overall performance of the system, a multitude of approaches can be implemented.

The implementation of the proposed security paradigm, which is founded on blockchain technology, should be extended to encompass multiple crucial stages. The first step in simulating the functionality of an intelligent lighting system entails the importation and preprocessing of data related to traffic prediction. To enhance comprehension of the underlying patterns and interrelationships, the data analysis is conducted using Panda's library, while data visualization is facilitated through the utilization of Seaborn. The further development of the blockchain network will involve the installation of nodes that are responsible for processing security data and facilitating communications, as well as the interlinking of these nodes with one another. The purpose of this study is to investigate various consensus procedures with the intention of promoting agreement

among nodes over the legitimacy of transactions and preserving the integrity of the system. The comprehensive design and development of the blockchain network constitutes an integral component of the procedural framework. This study aims to explore the intricate intricacies of the architecture of the blockchain, elucidate its fundamental components, and establish robust communication protocols among the interconnected nodes. The aforementioned stage holds significant importance in the establishment of a robust and impervious security infrastructure for the present intelligent traffic light system. The initial step involves a meticulous analysis of the dataset's characteristics, followed by the selection of a limited number of essential parameters for the security model. These may encompass data pertaining to traffic patterns, trends in lighting usage, and other relevant information necessary for the effective monitoring and control of the smart lighting system.

The commencement of constructing the architectural framework of the blockchain network is initiated once the crucial variables have been identified. The primary aim of this study is to create a distributed and decentralized ledger system that enables the secure and immutable storage of data related to security issues. Each individual network node operated as a participant with the responsibility of verifying and validating transactions. The researchers have conducted an analysis of different consensus mechanisms, including Proof-of-Work (PoW) and Proof-of-Stake (PoS), to establish consensus among the nodes within the network. A comprehensive assessment of the benefits and drawbacks of each mechanism is conducted, considering the specifications of the intelligent traffic light system. Moreover, it is important to contemplate the integration of a consensus mechanism that effectively caters to the unique requirements of the network, while concurrently ensuring a balanced state of security, scalability, and energy. This research additionally examined the potential vulnerabilities and attack vectors that could pose a threat to the security framework of blockchain technology.

The selection of a consensus process for an intelligent traffic light system is contingent upon various criteria, including but not limited to energy efficiency, scalability, speed, trustworthiness, and decentralization. Proof of Stake (PoS) is considered a viable option owing to its notable advantages in terms of energy efficiency and reduced resource demands in comparison to Proof of Work (PoW). The system enables nodes to generate blocks and verify transactions by considering the quantity of cryptocurrency they possess, hence reducing ecological consequences and upholding network security. Delegated Proof of Stake (DPoS) is a consensus mechanism that improves scalability and transaction processing speed by enabling shorter confirmation times for blocks and increasing throughput. The Practical Byzantine Fault Tolerance (PBFT) protocol prioritizes the achievement of prompt consensus within a low-latency setting, hence ensuring both fault tolerance and definitive agreement. Proof of Authority (PoA) is particularly well-suited for situations in which the identification and reliability of participants hold significant importance. The proposed approach circumvents the resource-intensive characteristics of Proof of Work (PoW) and offers expedited block creation durations. Hybrid consensus models, which include components from many techniques, may be investigated to customize the consensus mechanism according to the individual needs of an intelligent traffic light system. Nonetheless, the development and execution of efficient hybrid models necessitate a comprehensive comprehension of the merits and limitations inherent in distinct consensus methods.

4. Experimental analysis

The present study employed the Isolation Forest method to detect anomalies within a dataset comprising traffic estimates. This section focuses on providing the systems architecture information, dataset information, existing Artificial Intelligence (AI) Models used for the prediction validation of the anomalies in the data algorithm of the model develop and Blockchain integration algorithms.

System information: The experimentation was done on a computing machine and software given in **Table 2**. This table contains local machine and the software used for simulation.

Table 2. Hardware and software information used for simulation.

Local hardware information	*Processor*	Intel(R) Core(TM) i5-6300U CPU @ 2.40GHz 2.50 GHz
	Installed RAM	8.00 GB (7.88 GB usable)
	Device ID	87143E1C-741E-4688-AE27-153E7C682FBA
	Product ID	00330-51622-10013-AAOEM
	System Type	64-bit operating system, x64-based processor
	SSD	256 GB
Operating system	Microsoft Windows 10 (Pro) 22H2	
Software used	Online IDE-Google Colab hosted Jupyter Notebook service	

Dataset information: The phenomenon of traffic congestion in metropolitan areas across the globe is on the rise, mostly attributed to factors such as the expansion of urban populations, the deterioration of infrastructure, suboptimal signal timing, and the absence of up-to-date information. According to INRIX's estimation, the economic impact on United States commuters in 2017 amounted to around $305 billion, resulting from factors such as fuel inefficiency, time missed, and escalated transportation expenses. To enhance traffic conditions, urban areas must implement novel techniques and technologies. The dataset has a total of 48.1 k (48,120) observations, documenting the hourly count of automobiles across four distinct junctions. The dataset comprises information pertaining to the date and time of road crossings, junction details, and vehicle identifiers. The data collected by the sensors at each junction was obtained at various time intervals, resulting in the inclusion of traffic data from distinct temporal segments. Certain intersections have shown a dearth of data, necessitated careful consideration while formulated future estimates.

The target variable in the dataset was denoted as the 'Cross 1' column. This dataset was derived from a Smart traffic dataset and consisted of multiple features. The main objective of this research is to develop a blockchain system aimed at ensuring the integrity of data by identifying and eliminating any anomalies detected in the dataset using the Isolation Forest model. The dataset contains 6 columns, and 16,128 rows contain the data about crossing of 6 lanes. The dataset was downloaded from Kaggle repository available on the following link:

https://www.kaggle.com/datasets/fedesoriano/traffic-prediction-dataset

Blockchain integration to the Algorithm of the developed model: Once the dataset was imported, the relevant variables were subsequently extracted from the dataset. Subsequently, the contamination parameter for the Isolation Forest model was configured to a value of 0.05 to initiate its operation. The model underwent training using the feature data to detect any potential outliers or anomalies present within the dataset. Furthermore, the trained model was employed to generate predictions for any anomalies present in the dataset. During the investigation, anomalies were identified by conducting a search for instances in which the model predicted a value of −1. A record was created documenting the indices of the identified aberrations. To ensure the dependability of the data, it is advisable to establish a foundational blockchain infrastructure. Blocks served as the fundamental components of the blockchain structure, encompassing essential elements such as a timestamp, a hash value representing the preceding block, and a hash value pertaining to the current block under consideration. Initially, a blockchain instance was instantiated, followed by the addition of a genesis block to the instance. To facilitate the construction of the blockchain, it was imperative to iterate through the rows of the dataset and ascertain the presence or absence of the index within the anomaly indices list. Upon transforming the raw data into a dictionary, a thorough examination was conducted to ensure that each index did not contain any unforeseen or extraneous information. A novel block was appended to the blockchain after receiving data through the "add_block(data)" function. The blockchain was intentionally designed to identify and exclude any anomalies detected, ensuring that only valid data is stored. To ascertain the authenticity of the blockchain, a comprehensive analysis was conducted on each individual block to draw a conclusive

determination. The process commenced with the second block and proceeded sequentially through the series of blocks within the blockchain. During this progression, the hash value of each block was recalculated, followed by a comparison with the previously computed hash value. Furthermore, an examination is conducted to ascertain whether the hash value of the preceding block corresponded to the hash value of the present block, which was accomplished by inspecting the "previous_hash" attribute of the current block. If every block successfully satisfies each of these criteria, then the blockchain is deemed to be genuine. The experimental findings elucidated the number of anomalies detected and assessed the extent to which the blockchain technology preserved the integrity of the dataset. Given the simplicity of the blockchain implementation and Isolation Forest method employed in this study, it is imperative to acknowledge that their applicability to more intricate real-world scenarios may be limited. Additional refinement and advancement of the model and algorithm would be necessary to effectively accommodate authentic applications across diverse domains. Overall, this experiment yielded valuable insights into the data integrity and anomaly detection capabilities of machine learning algorithms and blockchain technology. The integration of these two technological advancements holds the promise of enhancing data security and reliability in contexts where the detection of data tampering and fraudulent activities is of utmost importance. Before implementing the strategies in real-world scenarios, it is imperative to conduct a thorough evaluation and testing of the techniques' actual performance and their suitability across various use cases. The stepwise description of the pseudocode used for the development of the model is as follows:

- Step 1: Import the data from a traffic dataset for smart lights.
- Step 2: Separate the target variable and the features from the dataset, then initialize an isolation forest model with a given contamination parameter and using the features, train the Isolation Forest model, then use the learned model to forecast anomalies in the dataset.
- Step 3: Create a Blockchain class with two methods: create_genesis_block() and add_block(data) and identify the dataset's anomaly detection indexes and a genesis block is added to the network using the create_genesis_block() function.
- Step 4: The add_block(data) function creates a Block class with the following attributes: timestamp, data, previous_hash, and hash, as well as a method calculate_hash(). It also adds a new block to the blockchain with the provided data and the hash of the previous block.
- Step 5: The block's hash is determined by the calculate_hash() function utilising its characteristics then Make a "blockchain" instance of the Blockchain class and add blocks to it while omitting the anomalies found in the dataset:
- Step 6: Convert the row data for each row in the dataset to a dictionary if the row index is not one of the anomaly_indices list.
- Step 7: Use the add_block(data) function to add the block containing the data to the blockchain and check the validity of each block to confirm the blockchain's integrity:
- Step 8: Start with the second block (index 1) and iterate through the blockchain chain and calculate each block's hash, then compare it to the previously saved hash and check to see whether the current block's previous_hash matches the previous block's hash.
- Step 9: If any of these tests are unsuccessful, the blockchain is invalid. Print the indices of the found anomalies and the blockchain's validity status.

Algorithm used for Anomaly Detection: This research uses the Isolation Forest Algorithm for anomaly detection. The utilization of intelligent traffic management technologies has become prevalent as a response to growing concerns regarding traffic congestion and safety within urban environments[37,38]. Traffic light anomalies have a detrimental effect on both the flow of traffic and the safety of drivers. The incorporation of this model serves the objective of enhancing the dependability and efficiency of the Smart Traffic Light system[39]. The Isolation Forest model, which was proposed by Liu, Ting, and Zhou in 2008, is a commonly

employed machine learning algorithm utilized for the purpose of anomaly detection in datasets with a high number of dimensions. The primary objective of this study is to discern anomalies, outliers, or infrequent occurrences, while simultaneously addressing obstacles such as distance-based and density-based methodologies. The model functions based on the fundamental principle of "isolating" anomalies by means of random partitioning, employing a tree-like structure to segregate data points. Anomalies exhibit shorter average path lengths, whereas normal instances necessitate a greater number of splits, leading to longer average path lengths. The approach based on isolation assigns a score of anomalies to each data point, where shorter average path lengths suggest potential anomalies and longer path lengths suggest normal instances. The Isolation Forest model employs a mathematical formula to compute the anomaly score for every data point within the dataset. The anomaly score serves as a metric for quantifying the level of isolation exhibited by a given data point within the feature space, thereby facilitating the detection and identification of anomalies. The mathematical expression representing the anomaly score in the Isolation Forest model is given by the formula given in Equation 1.

Anomaly Score (A) for Data Point n:

$$A(n) = 2^{-\frac{G(h(n))}{y(m)}} \tag{1}$$

where:

- $G(h(n))$ represents the average path length of data point n in the Isolation Forest, which is calculated during the model training process.
- *y(m)* is the average path length of unsuccessful search in a completely random binary tree with m data points.
- The term *y(m)* is a constant representing the average path length of unsuccessful search in a completely random binary tree. It depends on the number of data points in the dataset *(m)*.
- The anomaly score *A(n)* is a value between 0 and 1. Data points with a lower anomaly score are considered more likely to be anomalies, while data points with higher scores are deemed to be normal instances.

The Isolation Forest model creates a forest consisting of random trees, and the final anomaly score for each data point is obtained by agglomerating the anomaly scores obtained from all of the trees in the forest. The Isolation Forest model is an unsupervised machine learning technique, which means that it does not require labelled data during training. Its main application is in anomaly detection jobs; thus, this fact should be brought to your attention as soon as possible.

Machine Learning Algorithms used for the validation of detected anomalies: Various machine learning (ML) algorithms have been implemented in the field of smart transportation to enhance traffic flow, transportation efficiency, and overall user experience. This study employed three ML algorithms, namely Linear Regression, Decision Tree Regression, and Random Forest Regression to forecast the accuracy of blockchained data in relation to its binding in the consensus.

- Linear Regression Algorithm: To evaluate data, make predictions, and get a deeper knowledge of the relationships between variables, linear regression is used extensively in many different domains, including economics, finance, the social sciences, and engineering. Linear regression model is used in this research for training the machine for performing the feature scaling using standard scaler and obtain the prediction based on the testing dataset.
- Decision Tree Algorithm: The Decision Tree algorithm is commonly employed in supervised machine learning for tasks involving classification and regression. The utilisation of anomaly prediction is commonly observed in smart lighting systems to forecast atypical readings. The Decision Tree algorithm can be employed in the domain of anomaly value prediction to construct a model that detects atypical patterns within the smart lighting system, utilising a diverse range of features or sensor data. The generation of a model can be achieved through the utilisation of a specific methodology.

- Random Forest Algorithm: The Random Forest algorithm is an effective ensemble learning technique that is commonly employed in intelligent lighting systems for the purpose of forecasting anomalous values and atypical behavior by leveraging sensor data and environmental variables. The process encompasses various stages, including data collection, preprocessing, labelling, training, anomaly prediction, threshold selection, evaluation, and real-time prediction. The algorithm employs a combination of decision trees to construct a varied ensemble, thereby mitigating the risk of overfitting and enhancing its ability to generalize. The ultimate prediction is determined through a consensus voting process among all decision trees. The performance of the model is assessed through the utilization of established metrics and cross-validation methodologies. Real-time prediction facilitates the ongoing monitoring of sensor data and enables the timely detection of anomalies as they manifest.

The dataset is partitioned into two subsets, with 80% of the data allocated for training the model and the remaining 20% reserved for testing purposes. The data is scaled through the application of feature scaling using standard scalar. Subsequently, various AI models are fitted to the scaled data to generate predictions based on mean square error and R-squared (R2) scores. In summary, the implemented model incorporates all three AI algorithms to train and validate the target variable. The training data is utilized to train each regression model prior to generating predictions using the test set. The performance of each model is evaluated by utilizing two metrics, namely the mean squared error (MSE) and the R-squared (R2) score.

5. Result and discussion

The efficacy of the present security approach, which is based on blockchain technology, is evaluated through the implementation of various simulated real-world scenarios. The simulations generate artificial data to replicate diverse traffic and lighting patterns, along with attempts to breach security systems. The objective of these simulations is to validate the resilience and reliability of the blockchain network, thereby demonstrating its ability to detect and prevent unauthorized access and manipulation.

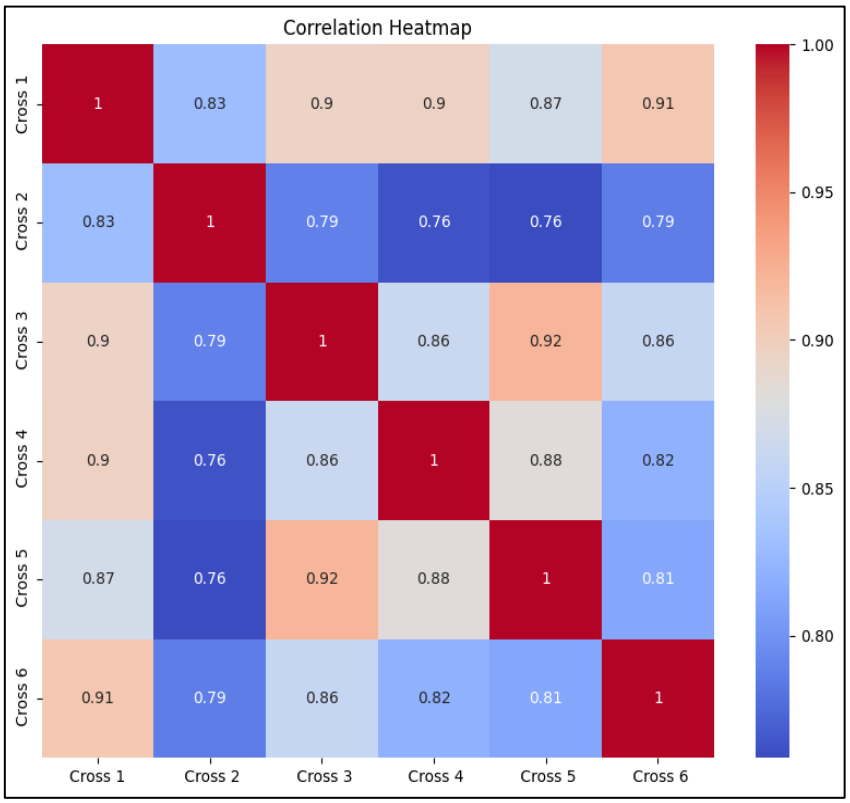

Figure 2. Visualization of the correlation matrix.

The graphical representation in **Figure 3** illustrates the distribution of each numerical characteristic within the dataset. The histogram plot within each subplot illustrates the frequency or count of values within distinct ranges or bins, representing a singular column within the dataset. This visualization facilitates the comprehension of the dataset's numerical elements by providing insights into their range, central tendency, and spread. **Figure 2** is a graphical representation of the correlation matrix that pertains to the numerical characteristics of the dataset. The intensity of the colors is used by the heatmap to indicate the degree to which two features are related to one another as well as the direction in which that relationship points. Hues with a warm temperature denote a positive association, whilst hues with a cool temperature denote a negative correlation. The exact correlation coefficients are represented by the values that are indicated on the heatmap. The employment of this representation makes it easier to recognize linkages and interactions among various aspects contained inside databases.

Figure 3 employs scatter plots to visually represent the pair-wise correlations that are present among the numerical attributes of the dataset. A scatter plot is a visual representation that compares two characteristics and displays the data in the form of a graph, where each data point is represented by a dot. This visualization enables the identification of patterns, trends, and potential correlations among diverse pairs of features. This facilitates the comprehension of the connections and interrelationships that are present among the numerical variables within the dataset.

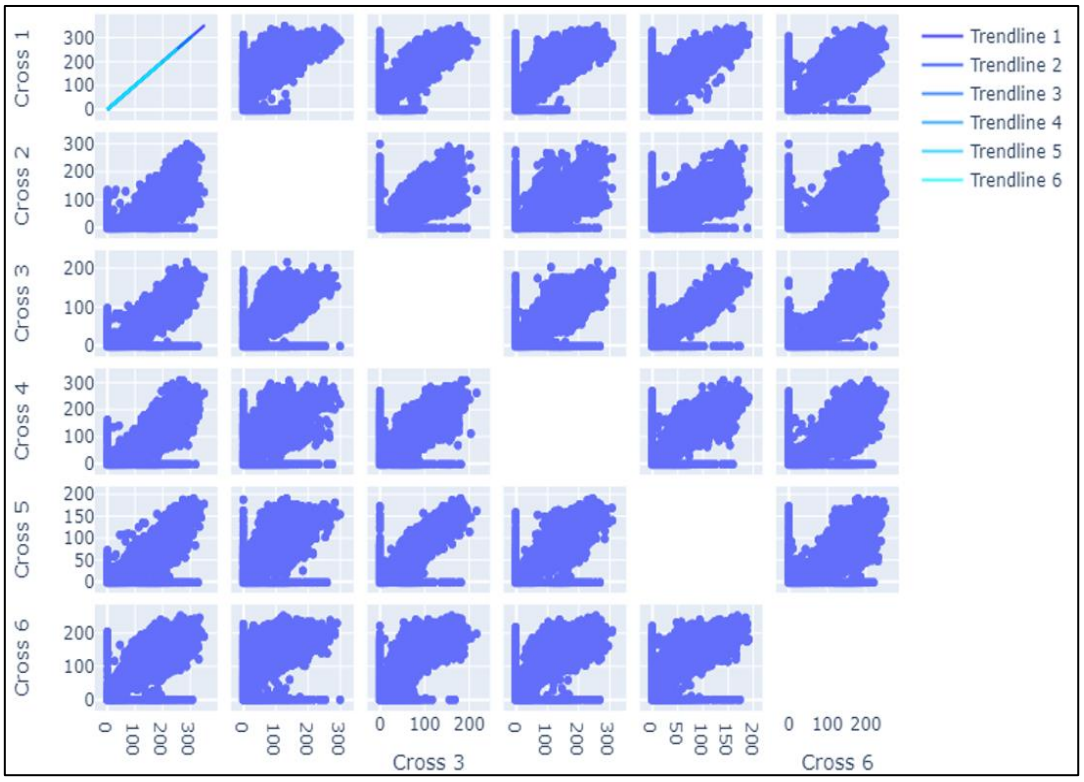

Figure 3. Pair-wise scatter plots with trendlines.

Figure 4 shows a specific pairwise scatter plot between 'Cross 1' and 'Cross 3'. **Figure 4** shows the relationship between 'Cross 1' and 'Cross 3', emphasizing a detailed analysis with a dedicated trendline depicted in a contrasting color for clarity. The graphical representation of the relationship between specific attributes and the label of the dataset is depicted in **Figure 5**. The heatmap's presentation of correlation coefficients utilizes warmer colors to indicate a positive correlation, while colder colors are used to represent a negative correlation. The heatmap's annotated values display the exact correlation coefficients. This visualization facilitates comprehension of the relationship between specific traits and the label by emphasizing the potentially crucial or noteworthy aspects for prediction.

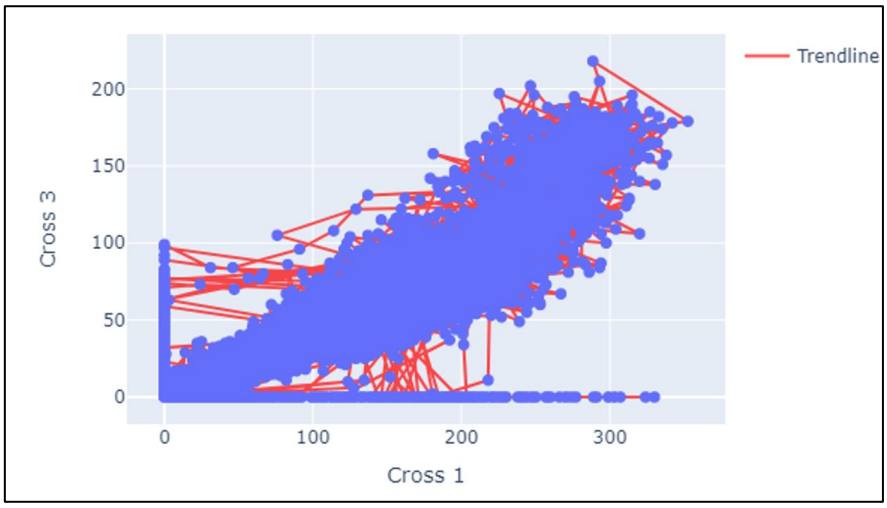

Figure 4. Scatter Plot between Cross 1 and Cross 3 with a trendline.

When conducting a comparative analysis of several models, we discovered significant variances in key indicators when considering the integration of blockchain technology. In the absence of blockchain, the Linear Regression model yielded a Mean Squared Error (MSE) of 0.487, an R2 Score of 0.250, and an Accuracy of 48.69%.

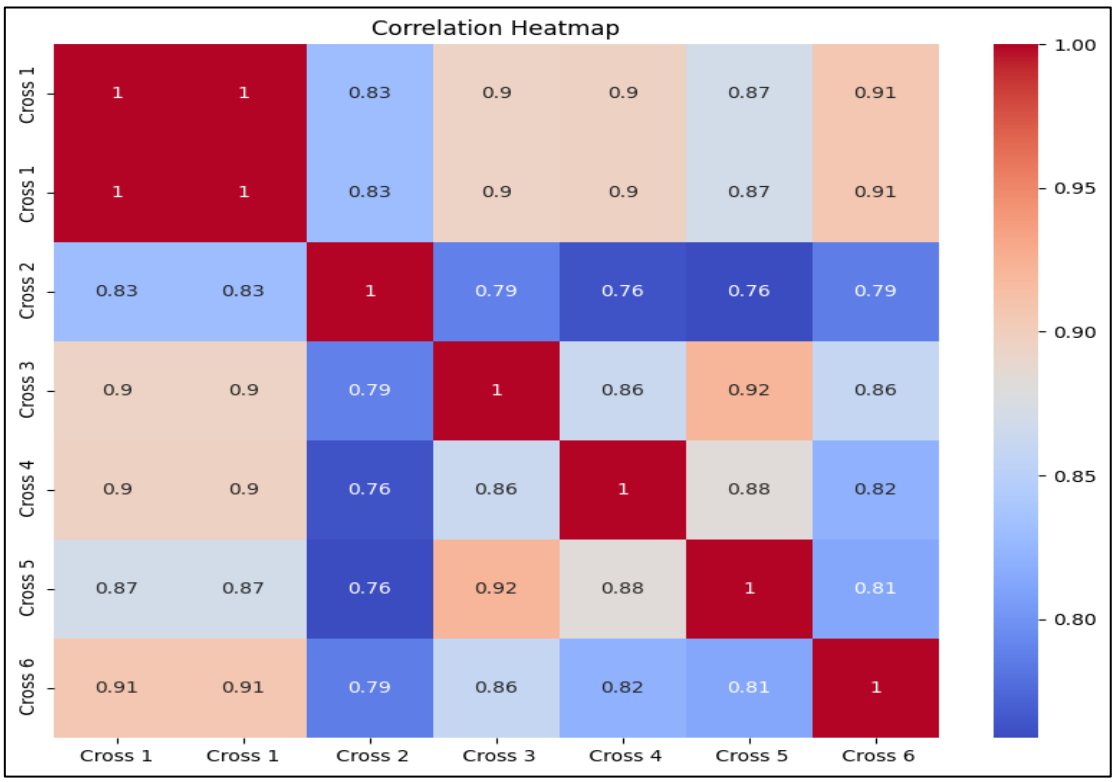

Figure 5. Visualization of the correlation between features and the label.

Nevertheless, the incorporation of blockchain technology resulted in a notable enhancement in performance, as evidenced by a decrease in the mean squared error (MSE) to 665.074, an increase in the R2 score to 0.911, and an improvement in accuracy to 91.1%. In a similar vein, the inclusion of blockchain technology resulted in notable enhancements in the performance of the Decision Tree Regressor. In the absence of blockchain technology, the model exhibited a Mean Squared Error (MSE) of 0.467, an R2 Score of 0.868, and an Accuracy rate of 46.69%. The implementation of blockchain technology resulted in significant enhancements in the metrics, yielding an MSE (Mean Squared Error) of 830.398, an R2 Score of 0.888, and

an Accuracy of 88.7%. The introduction of blockchain technology resulted in improved performance of the Random Forest Regressor. The model lacking blockchain technology exhibited a mean squared error (MSE) of 0.268, an R-squared (R2) score of 0.073, and an accuracy rate of 48.88%. The implementation of blockchain technology resulted in notable enhancements in the metrics, including a notable decrease in the mean squared error (MSE) to 433.808, a substantial gain in the R2 score to 0.942, and a notable improvement in accuracy, reaching 94.2%. The results of this study indicate that the incorporation of blockchain technology had a beneficial impact on the predictive abilities of the models. This resulted in enhanced accuracy and decreased mistakes in the regression tasks conducted by the Linear Regression, Decision Tree Regressor, and Random Forest Regressor algorithms. Considering the primary objective being prediction accuracy, it is recommended to employ the Random Forest Regression model for this dataset. However, the optimal model can also be influenced by additional factors, such as interpretability and computational efficiency. When choosing the most suitable regression model for implementation, it is imperative to carefully evaluate the specific requirements of the given scenario. The efficacy of the Random Forest algorithm within a Smart Traffic Light System is contingent upon various aspects. In the absence of blockchain technology, the integrity of data, the process of selecting relevant features, and the availability of appropriate training data assume paramount importance. The utilization of blockchain technology has the potential to augment both the integrity and security of data. However, it is important to acknowledge that other factors such as decentralization, smart contracts, real-time updates, and privacy considerations all contribute significantly to this process. These characteristics may have an impact on both the training and deployment of the model. The incorporation of blockchain technology should be strategically aligned with the objectives of the Smart Traffic Light System, thereby enhancing its overall precision and effectiveness. It is imperative to consider the selection of features, training data, and privacy considerations. **Table 3** depicts all the results achieved with and without applying blockchain technology.

Table 3. Anomalies detection Metrix result analysis.

Model	**Without blockchain**			**With blockchain**		
	MSE	**R2 Score**	**Accuracy**	**MSE**	**R2 score**	**Accuracy**
Linear regression	0.487	0.250	48.69%	665.074	0.911	91.1%
Decision tree regressor	0.467	0.868	46.69%	830.398	0.888	88.7%
Random forest regressor	0.268	0.073	48.88%	433.808	0.942	94.2%

6. Conclusion

The current study used linear regression, decision tree regression, and random forest regression models to assess the validity of the model's accuracy. The research demonstrated that the incorporation of blockchain technology yielded notable enhancements in the efficacy of Linear Regression, Decision Tree Regressor, and Random Forest Regressor models. In the absence of blockchain technology, the Linear Regression model exhibited a mean squared error of 0.487, an R2 score of 0.250, and an accuracy rate of 48.69%. Nevertheless, the introduction of blockchain technology resulted in a reduction of the mean squared error to 665.074, an enhancement of the R2 score to 0.911, and an increase in accuracy to 91.1%. This implies that the utilization of blockchain technology has the potential to augment prediction capabilities and mitigate errors in regression activities. The current body of research pertaining to the application of blockchain and machine learning in smart traffic signal systems shows considerable potential for anomaly identification. Future research directions encompass the optimization of anomaly detection algorithms, exploration of blockchain consensus methods, integration of blockchain and machine learning to bolster security, mitigation of privacy concerns, and the creation of energy-efficient blockchain systems. The issue of scalability is also being actively tackled, with the development of solutions that can effectively manage an increasing volume of transactions and accommodate the proliferation of smart devices, all while maintaining optimal performance levels. The evaluation is focused

on assessing the system's resilience to adversarial assaults, including those that are designed to exploit vulnerabilities in machine learning models. Empirical investigations and practical examinations are underway to evaluate the feasibility and efficacy of the proposed anomaly detection system through real-world deployments and case studies. Efforts are being made to tackle the problems related to interoperability and standardization by establishing standardized interfaces that facilitate the integration of security solutions based on blockchain technology and machine learning. The integration of user-centric design concepts is being implemented to guarantee that the system is designed in a manner that prioritizes the needs and preferences of the users, while also being in accordance with the values and norms of society.

Author contributions

Conceptualization, SS and NM; methodology, SS; software, SS; validation, SS and NM; formal analysis, SS and NM; investigation, SS; resources, SS; data curation, SS; writing—original draft preparation, SS; writing—review and editing, SS and NM; supervision, NM; project administration, NM. All authors have read and agreed to the published version of the manuscript.

Conflict of interest

The authors declare no conflict of interest.

References

1. Kyvelou SS, Bobolos N, Tsaligopoulos A. Exploring the Effects of "Smart City" in the Inner-City Fabric of the Mediterranean Metropolis: Towards a Bio-Cultural Sonic Diversity? Heritage. 2021, 4(2): 690–709. doi: 10.3390/heritage4020039
2. Wang C, Gu J, Sanjuán Martínez O, et al. Economic and environmental impacts of energy efficiency over smart cities and regulatory measures using a smart technological solution. Sustainable Energy Technologies and Assessments. 2021, 47: 101422. doi: 10.1016/j.seta.2021.101422
3. Paiva S, Ahad M, Tripathi G, et al. Enabling Technologies for Urban Smart Mobility: Recent Trends, Opportunities and Challenges. Sensors. 2021, 21(6): 2143. doi: 10.3390/s21062143
4. Kishore K, Sharma S. Information Security & Privacy in Real life-Threats & Mitigations: A Review. International Journal of Computer Science and Technology. 2013, 4(3): 38–41.
5. Sarker IH. Smart City Data Science: Towards data-driven smart cities with open research issues. Internet of Things. 2022, 19: 100528. doi: 10.1016/j.iot.2022.100528
6. Shamneesh S, Manoj M, Keshav K. Node-Level Self-Adaptive Network Path Restructuring Technique for Internet of Things (IoT). Intelligent Communication, Control and Devices. Published online August 28, 2019: 453–461. doi: 10.1007/978-981-13-8618-3_48
7. Wan Y, Xu K, Wang F, et al. Characterizing and Mining Traffic Patterns of IoT Devices in Edge Networks. IEEE Transactions on Network Science and Engineering. 2021, 8(1): 89–101. doi: 10.1109/tnse.2020.3026961
8. Wang Q, Chen L, Wang Q, et al. Anomaly-Aware Network Traffic Estimation via Outlier-Robust Tensor Completion. IEEE Transactions on Network and Service Management. 2020, 17(4): 2677–2689. doi: 10.1109/tnsm.2020.3024932
9. Sharma S, Manuja M, Kishore K. Vulnerabilities, Attacks and their Mitigation: An Implementation on Internet of Things (IoT). International Journal of Innovative Technology and Exploring Engineering. 2019, 8(10): 146–150. doi: 10.35940/ijitee.f3761.0881019
10. Alaslani M, Nawab F, Shihada B. Blockchain in IoT Systems: End-to-End Delay Evaluation. IEEE Internet of Things Journal. 2019, 6(5): 8332–8344. doi: 10.1109/jiot.2019.2917226
11. Sharma C, Sharma S, Sakshi. Latent DIRICHLET allocation (LDA) based information modelling on blockchain technology: A review of trends and research patterns used in integration. Multimedia Tools and Applications. 2022, 81(25): 36805–36831. doi: 10.1007/s11042-022-13500-z
12. Gupta M, Sharma S, Sakshi, et al. Security and Privacy Issues in Blockchained IoT. Blockchain Technology. Published online February 24, 2022: 27–56. doi: 10.1201/9781003138082-3
13. Angiulli F, Fassetti F, Serrao C. Anomaly detection with correlation laws. Data & Knowledge Engineering. 2023, 145: 102181. doi: 10.1016/j.datak.2023.102181
14. Schäffer M, di Angelo M, Salzer G. Performance and Scalability of Private Ethereum Blockchains. Lecture Notes in Business Information Processing. Published online 2019: 103–118. doi: 10.1007/978-3-030-30429-4_8
15. Bhutta MNM, Khwaja AA, Nadeem A, et al. A Survey on Blockchain Technology: Evolution, Architecture and Security. IEEE Access. 2021, 9: 61048–61073. doi: 10.1109/access.2021.3072849

16. Meisami S, Meisami S, Yousefi M, et al. Combining Blockchain and IoT for Decentralized Healthcare Data Management. International Journal on Cryptography and Information Security. 2023, 13(1): 35–50. doi: 10.5121/ijcis.2023.13102
17. Chen HY, Sharma K, Sharma C, et al. Integrating explainable artificial intelligence and blockchain to smart agriculture: Research prospects for decision making and improved security. Smart Agricultural Technology. 2023, 6: 100350. doi: 10.1016/j.atech.2023.100350
18. Kishore K, Sharma S. Evolution of wireless sensor networks as the framework of Internet of Things—A review. International Journal of Emerging Research in Management &Technology. 2016, 5.
19. Sharma K, Sharma C, Sharma S, et al. Broadening the Research Pathways in Smart Agriculture: Predictive Analysis Using Semiautomatic Information Modeling. Journal of Sensors. 2022, 2022: 1–19. doi: 10.1155/2022/5442865
20. Bali S, Sharma S. Anticipating Legal Issues Associated with the Cyber Security and Privacy of Automated Driving Systems in India. Autonomous Driving and Advanced Driver-Assistance Systems (ADAS). Published online November 12, 2021: 389–400. doi: 10.1201/9781003048381-20
21. Sharma N, Mishra S. Integration of Blockchain Technology into Smart Cities for Information Security: Research Constituents and Need of Research. 7th International Joint Conference on Computing Sciences (ICCS-2023). 2023.
22. Singh S, Malik A, Batra I, et al. Need for Integration of Blockchain Technology in Supply Chain Management of Health Supplements. 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE). Published online May 12, 2023. doi: 10.1109/icacite57410.2023.10183099
23. Sharma S, Mishra N. Horizoning recent trends in the security of smart cities: Exploratory analysis using latent semantic analysis. Journal of Intelligent & Fuzzy Systems. Published online November 7, 2023: 1–18. doi: 10.3233/jifs-235210
24. Zhao H, Zhang W, Wu X, et al. Outlier Detection and Trust based Distributed Cooperative Spectrum Sensing in Internet of Vehicles. 2022 International Conference on Computing, Communication, Perception and Quantum Technology (CCPQT). Published online August 2022. doi: 10.1109/ccpqt56151.2022.00027
25. Mirsky Y, Golomb T, Elovici Y. Lightweight collaborative anomaly detection for the IoT using blockchain. Journal of Parallel and Distributed Computing. 2020, 145: 75–97. doi: 10.1016/j.jpdc.2020.06.008
26. Mathew SS, Hayawi K, Dawit NA, et al. Integration of blockchain and collaborative intrusion detection for secure data transactions in industrial IoT: A survey. Cluster Computing. 2022, 25(6): 4129–4149. doi: 10.1007/s10586-022-03645-9
27. Kumar P, Kumar R, Srivastava G, et al. PPSF: A Privacy-Preserving and Secure Framework Using Blockchain-Based Machine-Learning for IoT-Driven Smart Cities. IEEE Transactions on Network Science and Engineering. 2021, 8(3): 2326–2341. doi: 10.1109/tnse.2021.3089435
28. Maskey SR, Badsha S, Sengupta S, et al. BITS: Blockchain based Intelligent Transportation System with Outlier Detection for Smart City. 2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). Published online March 2020. doi: 10.1109/percomworkshops48775.2020.9156237
29. Cook AA, Misirli G, Fan Z. Anomaly Detection for IoT Time-Series Data: A Survey. IEEE Internet of Things Journal. 2020, 7(7): 6481–6494. doi: 10.1109/jiot.2019.2958185
30. Ramapatruni S, Narayanan SN, Mittal S, et al. Anomaly Detection Models for Smart Home Security. In: Proceedings of the 2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS); 27–29 May 2019; Washington, DC, USA. pp. 19-24.
31. Balogun AL, Tella A. Modelling and investigating the impacts of climatic variables on ozone concentration in Malaysia using correlation analysis with random forest, decision tree regression, linear regression, and support vector regression. Chemosphere. 2022, 299: 134250. doi: 10.1016/j.chemosphere.2022.134250
32. Erhan L, Ndubuaku M, Di Mauro M, et al. Smart anomaly detection in sensor systems: A multi-perspective review. Information Fusion. 2021, 67: 64–79. doi: 10.1016/j.inffus.2020.10.001
33. Thakur A, Sharma S, Sharma T. Design of Semantic Segmentation Algorithm to Classify Forged Pixels. 2023 IEEE 12th International Conference on Communication Systems and Network Technologies (CSNT). Published online April 8, 2023. doi: 10.1109/csnt57126.2023.10134649
34. Sharma S, Kishore K. Internet of Things (IoT): A Review of Integration of Precedent, Existing & Inevitable Technologies. AGU International Journal of Engineering and Technology, 2017, 4, pp. 442–2455.
35. Ray PP. An Introduction to Dew Computing: Definition, Concept and Implications. IEEE Access. 2018, 6: 723–737. doi: 10.1109/access.2017.2775042
36. Sharma S, Manuja M, Puri D. Performance Analysis of Commodity Server with Freeware Remote Terminal Application in Homogeneous and Heterogeneous Mutli-computing Environments. Recent Innovations in Computing. Published online 2021: 3–12. doi: 10.1007/978-981-15-8297-4_1
37. Farman H, Khan Z, Jan B, et al. Smart Transportation in Developing Countries: An Internet-of-Things-Based Conceptual Framework for Traffic Control. Islam SH, ed. Wireless Communications and Mobile Computing. 2022, 2022: 1–11. doi: 10.1155/2022/8219377

38. Sharma C, Batra I, Sharma S, et al. Predicting Trends and Research Patterns of Smart Cities: A Semi-Automatic Review Using Latent Dirichlet Allocation (LDA). IEEE Access. 2022, 10: 121080-121095. doi: 10.1109/access.2022.3214310
39. Oladimeji D, Gupta K, Kose NA, et al. Smart Transportation: An Overview of Technologies and Applications. Sensors. 2023, 23(8): 3880. doi: 10.3390/s23083880