



Computers, Materials &  
Continua

[Make Submissions](#)
[Propose a Special Issue](#)

**Vol.70, No.3, 2022**

Views 15  
PDF Downloads 9

**Download & Citation**

Full-Text PDF  
Full-Text HTML  
Full-Text XML  
Full-Text Epub

**Citation Tools**

BibTex  
EndNote  
RIS

Vol.70, No.3, 2022, pp.4429-4446, doi:10.32604/cmc.2022.019755

OPEN ACCESS

ARTICLE

## Blockchain and Machine Learning for Intelligent Multiple Factor-Based Ride-Hailing

Zeinab Shahbazi, Yung-Cheol Byun\*

IIST, Department of Computer Engineering, Jeju National University, Jeju-si, Jeju Special Self-Governing Province, Korea

\* Corresponding Author: Yung-Cheol Byun. Email: ycb@jejunu.ac.kr

(This article belongs to this Special Issue: [Advancements in Lightweight AI for Constrained Internet of Things Device](#))

Received 24 April 2021; Accepted 09 July 2021; Issue published 11 October 2021



Download PDF



View HTML

### Abstract

One of the common transportation systems in Korea is calling taxis through online applications, which is more convenient than traditional taxis. However, the driver's passenger taxi request can be rejected based on the driver's location and distance. The acceptance and rejection of the received request. The security of this system is another main core to save the traffic and drivers. In this study, the origin and destination of the Jeju island South Korea were captured from T-map and Google Maps. The tree and XGBoost techniques. The blockchain framework is implemented in the Hyperledger Fabric platform. The socio-economic. The cross-validation was accomplished. Distance is another factor for the taxi trip, which in total presents the successful matching of ride-hailing taxi services with the specialty of distance, the trip request, and safe

### Keywords

Taxi trip; blockchain; machine learning; ride-hailing; prediction; trip distance

### Cite This Article

BibTex EndNote RIS

Shahbazi, Z., Byun, Y. (2022). Blockchain and Machine Learning for Intelligent Multiple Factor-Based Ride-Hailing Services. *Computers, Materials & Continua*, 70(3), 4429–4446.



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits reproduction in any medium, provided the original work is properly cited.

### We recommend

Data Based Violated Behavior Analysis of Taxi Driver in Metropolis in China

Jiao Yao et al., CMC-Computers, Materials & Continua

Service-Aware Access Control Procedure for Blockchain Assisted Real-Time Applications

Research on the big data of traffic flow: A systematic review

Tao Lyu et al., Journal of Traffic & Transportation Engineering (English Edition), 2021

Uber agrees to share passenger data with researchers

# Blockchain and Machine Learning for Intelligent Multiple Factor-Based Ride-Hailing Services

Zeinab Shahbazi and Yung-Cheol Byun\*

IIST, Department of Computer Engineering, Jeju National University, Jeju-si, Jeju Special Self-Governing Province, 63243, Korea

\*Corresponding Author: Yung-Cheol Byun. Email: ycb@jejunu.ac.kr

Received: 24 April 2021; Accepted: 09 July 2021

**Abstract:** One of the common transportation systems in Korea is calling taxis through online applications, which is more convenient for passengers and drivers in the modern area. However, the driver's passenger taxi request can be rejected based on the driver's location and distance. Therefore, there is a need to specify driver's acceptance and rejection of the received request. The security of this system is another main core to save the transaction information and safety of passengers and drivers. In this study, the origin and destination of the Jeju island South Korea were captured from T-map and processed based on machine learning decision tree and XGBoost techniques. The blockchain framework is implemented in the Hyperledger Fabric platform. The experimental results represent the features of socio-economic. The cross-validation was accomplished. Distance is another factor for the taxi trip, which in total trip in midnight is quite shorter. This process presents the successful matching of ride-hailing taxi services with the specialty of distance, the trip request, and safety based on the total city measurement.

**Keywords:** Taxi trip; blockchain; machine learning; ride-hailing; prediction; trip distance

## 1 Introduction

The increasing usage of online channels in the COVID19 pandemic fastens the business models and gives passengers the potential to have comfortable and safe choices. Increasing the Online-to-Offline (O2O) service increases the transportation sector for ride-hailing service, e.g., based on personal mobility in china, the transportation system expands into different passengers [1]. Ride-hailing and ride-sharing are two different topics. The ride-hailing is the requested taxi, which pick-up the passenger from a known place and drop-off at the requested destination, but ride-sharing is sharing the ride between passengers who don't have the same destination. This means ride-sharing is different from personal service, and it's not a private ride. Kakao taxi and T-map in Korea are popular apps for ride-hailing requests.

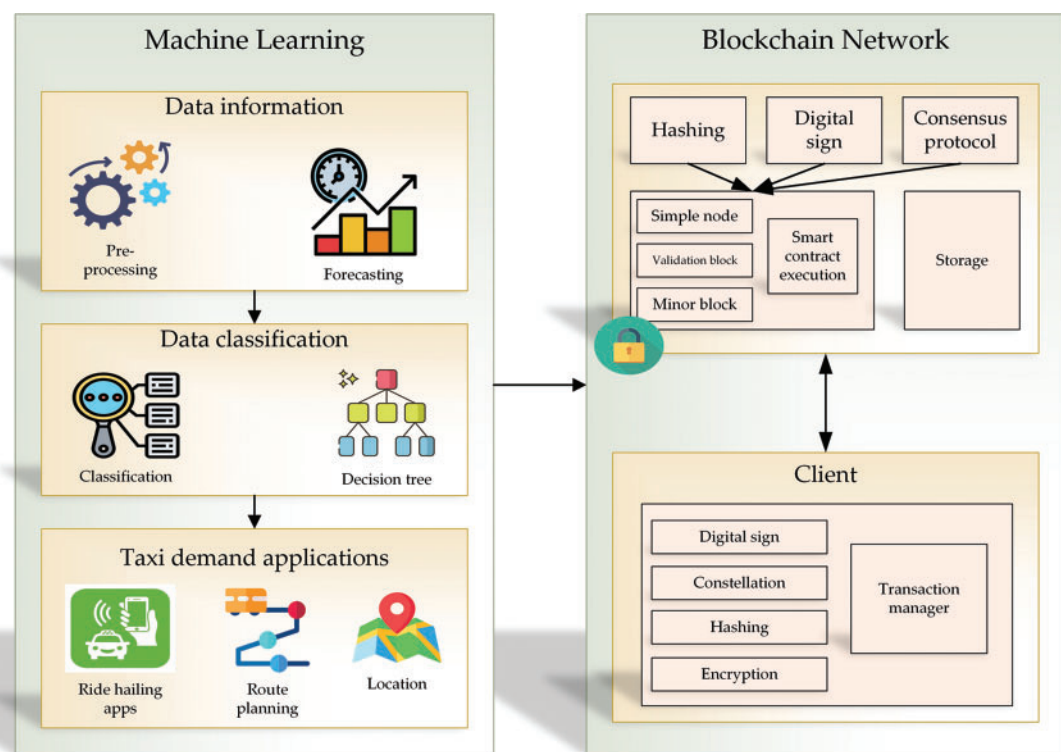
Regarding the passenger's convenience by using the taxi apps, some complaints are related to the drivers if they accept the passenger's requested location or deny it. Thus, the possibility of



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

not matching the balance between driver and passenger demand causes the ride-sharing service to be well-fit to passenger balance. This aspect makes the drivers protest against the taxi-sharing service in Korea.

Blockchain technology comes with huge changes in the taxi demand service with securing the database and transactional records [2]. Ride-hailing platforms based on blockchain technology can contact the passenger with the driver who accepts the requested location [3]. A Ride-hailing system based on blockchain reduces the mentioned problem by adding a manager between the passenger and driver. Passengers share the transactional data in nodes instead of the need to trust any authority. This process causes removing any intermediaries who carry out the role of gatekeeping. The transactional information keeps the distributed ledger possible to access through the defined nodes to make the blockchain network more transparent. In this system, the main power of blockchain is to record the data based on the time-series information to aggregate and share the transactions processed into the blockchain network. Every block in this process is not accessible, and it's encrypted, which is connected with a hash code. The transactions are the constant information which shared, unalterable and secure [4]. Fig. 1. shows the simple architecture of the proposed system overview. The machine learning process is to clarify the passenger and driver information based on the requested location and their pick-up and drop-off distance. The total data analysis and classification is evaluating the quality of ride-hailing and the satisfaction of passengers. The blockchain input analyzes and extracts data from the machine learning section to avoid fake transaction information and secures the driver and passenger profile information to avoid fake information.



**Figure 1:** Simple overview of the proposed integrated taxi demand service based on blockchain and machine learning approaches

The main contribution of this paper defined as:

- Identifying the match factors in ride-hailing service using machine learning techniques.
- Minimizing the overfitting and bias selection based on the cross-validation process.
- Connecting the gap between passenger and driver.
- Improving the ride-hailing service in big cities.
- Using blockchain technology to improve the security and transparency of the ride-hailing service.
- Deploying the smart contract prototype to explore the expenses between passenger and driver.
- Linking the problems in transportation to machine learning state-of-art.
- Exploring the taxi demand prediction in terms of network architecture.
- Capturing the demand service prediction based on temporal and spatial information.

The remaining of this paper, divided as Section 2, presents a brief explanation of the related study in taxi demand ride-hailing service. Section 3 presents the proposed taxi demand service based on the blockchain and machine learning integrated system methodology. Section 4 presents the implementation of the proposed system. Finally, Section 5 presents the generated results, and we conclude this paper in section conclusion.

## **2 Related Work**

This section presents a brief explanation of taxi demand service background and related applications and methods. There are three parts: service operation platform, taxi demand service based on machine learning, and taxi demand service based on blockchain.

### **2.1 Service Operation Platform**

One of the useful sharing economy operations is platform operation [5–10]. Employing platform-based technology is known as a more convenient way for companies [11–16]. Mantin et al. [17] process a platform for peer-to-peer marketplace based on uncertain values with consumer possessing. In [18], time establishing and optimal versioning proposed. In [19], the operation platform investment was reported with considering the network externals. Car sharing is presented in [20] and becomes a key role for the product design line. There are many research areas which focus on pricing decision in service of platforms. E.g., exploring ride-sharing optimal pricing platforms [21], supporting the taxi-hailing service based on optimal policies [22–27], and controlling the self-scheduling based on the capacity of service [28].

### **2.2 Taxi Demand Service Based on Machine Learning**

In recent years, it is enhancing the trip duration dedicated with many kinds of research for forecasting. The four-step process is one of the taxi demand prediction applications for spatiotemporal factors [29]. Moreira-Matias et al. [30] proposed the demand service based on spatial distribution considering real-time forecasting data for the passenger spatiotemporal demand distribution. Zhang et al. [31] proposed the forecasting of hotspot locations based on the adaptive method. Davis et al. [32] presented the taxi demand prediction based on the time-series data using the mobile application of taxi for regulation mining. In [33], New York City taxi transportation data used contains 1 billion uses of taxis and the complaints related to this system and Google Trends' information. The analysis is based on the competition between ride-sharing and ride-hailing. Seow et al. [34] present the taxi dispatching system in Singapore based on simulation

techniques to reduce passenger waiting time. Comparing this technique with other state-of-arts shows it successfully reduced 33.1% of waiting time and reduced 26.3% of idling taxis.

### 2.3 Taxi Demand Service Based on Blockchain

Different studies in the field of blockchain are appointed for taxi demand to secure the data-sharing service. The blockchain characteristic gives trust, security and establishes the transportation system with the intelligent ecosystem. The data-sharing issue gives better utilization for transportation resources [35]. In [36], feature exploration for reward-based communication of vehicles was explored. This mechanism is standing based on the Proof of Work (PoW) consensus. Liu et al. [37] applied the Byzantine Fault Tolerance algorithm to design their consensus phase based on blockchain framework to secure the vehicle announcement network using aggregation protocol. In [38], data sharing security processed based on DPoS consensus and applied blockchain for the Internet of Vehicles (IoV) to enhance the proposed system. Lu et al. [39] proposed the mechanism of trust points in vehicles based on a blockchain framework for securing the communication between vehicles. Applying public blockchain in this system contains the high cost and limitation for vehicle resources; therefore, this system is not suitable for Peer to Peer (P2P) data-sharing networks.

## 3 System Model

This section contains the proposed system detail model. This research aims to improve the taxi demand service performance and security by applying machine learning and blockchain integration. This section has six parts: machine learning framework for taxi demand service, decision tree and XGBoost performance in the proposed system, the optimization algorithm and cross-validation process, the blockchain framework, smart contract deployment for driver and ride requesting phase in this system.

### 3.1 Decision-Performance in Taxi Demand Service

The decision tree is used to create and specify the matching factors and build a suitable model that fits this process. The decision tree is a type of algorithm which contains various decision-making rules to classify and predict the information into smaller groups. These small groups contain sub-models that avoid the overuse of data splitting, including automatic intersection detection, regression, classification, etc. This model is easy to understand, and the presented system is non-parametric, which has no requirement for normalization and assumption. Thus, it's not absorbent for outliers, and the comparison result of this system with other state-of-art shows the superior of this algorithm in terms of prediction [40]. In this process, the Classification and Regression Tree (CART) executes the binary split, which differs from other splitting areas. The classification tree is a separate dependent variable, and the regression tree is the continuous dependent variable. Eq. (1) shows the probability of the two elements related to various groups.

$$A = \sum_{i=1}^n R(i)(1 - R(i)) = 1 - \sum_{i=0}^n R(i^2) = 1 - \sum_{i=0}^n (m_i/m^2) \quad (1)$$

where A represents the Gini index which is the measurement of each node diversity. R(i) is each node that belonged to objects probability in the  $i$ th category. The number of parent node observations present as m and  $m_i$  is the target variable of  $i$ th category.

### 3.2 Optimization and Cross-Validation Process

Machine learning algorithms contain better output than conventional ways in terms of prediction and represent the overfitting problem. The cross-validation process distributes the defined model without overfitting and similarly optimizes the classification and complexity balance. This technique presents how well the defined model performing and tests the accuracy. The decision tree errors have also taken using this algorithm. The tree's growth contains more terminal nodes and fewer errors, which shows the system has no good output with new data. To overcome this problem, cross-validation uses the cost function as shown in Eq. (2).

$$C_{\beta}(I) = C(I) + \beta(I^{\sim}) \rightarrow \text{Min } I \quad (2)$$

$C(I)$  defines the errors of misclassification in the tree. The measure complexity defines with the  $I$  and  $\beta(I^{\sim})$  that is based on  $I^{\sim}$  Which are the tree total node records and  $\beta$  define as a parameter. Machine learning data processing is in two train and test sets in this system. The data process in tree form and the  $\beta$  parameter defines regularization, which shows the training set input observations based on the defined model. The rest of the data is random for the test set. The depth of each node sample size was selected based on an exhaustive brute-force search to optimize the model. This method evaluates the minimum size between 1 to 10 nodes and the depth of the tree. The main reason for applying optimization and cross-validation in this system is overcoming the overfitting and balancing the decision-tree errors. Feature importance in this process is shown in Fig. 2. The feature importance is based on the population and distance of the trip according to the travel purposes and means.

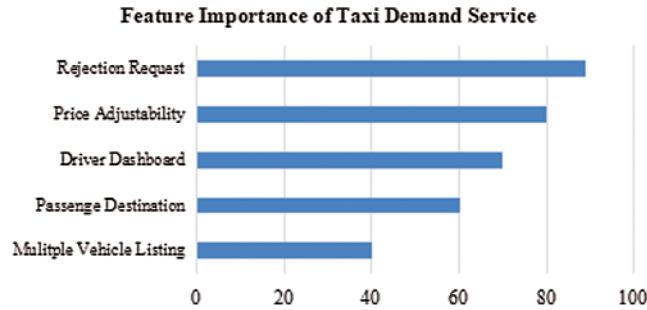


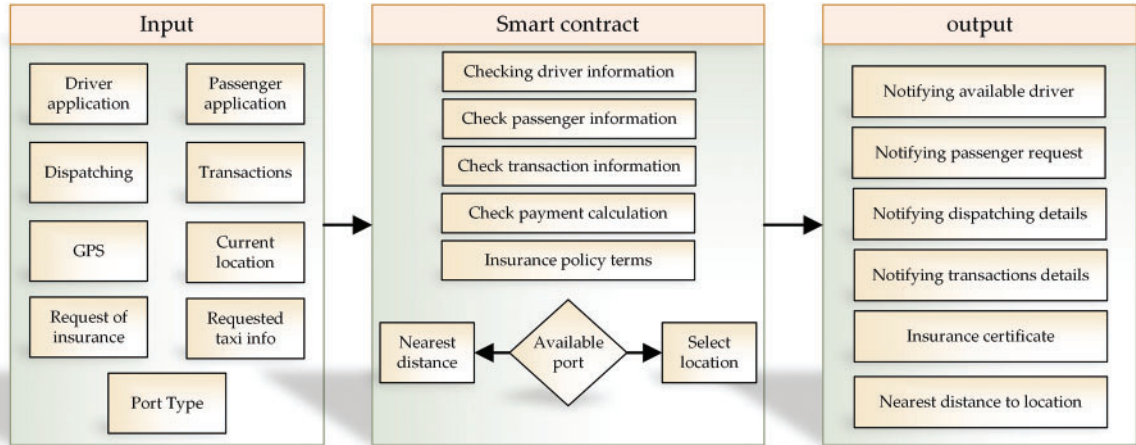
Figure 2: Feature importance of proposed taxi demand service

### 3.3 Blockchain-Based Framework for Taxi Demand Service

The blockchain provides tamper-resistant transaction records. Blockchain service reading is based on timestamp records which is difficult to tamper. In [41–44], the consensus protocols of blockchain are presented which shows that the blockchain server is strong and reliable in terms of the consensus system. Hyperledger Fabric uses the consensus of practical Byzantine fault method appropriate for the private blockchain. One of the important modules of blockchain is the smart contract. In this system, we have used smart contracts as a decision-making inference engine. Smart contracts are used for verification of run-time and actuating tasks. Task management based on IoT devices is also presented. The recommendation from sensors sends to the sensing function submitted by calling the smart contract. In this case, the smart contract verifies the sensor's license if it is available on the blockchain server. If it's not registered, then there is no access allowance. The generated data tracking analyzes with the smart contractor. The transaction proposal shows



the detailed analysis of sensing information based on set rules. Fig. 3. shows the inference engine data flow within the smart contract. Once all the conditions are matched, the smart contract gives the information to the client to perform the task. In the opposite case, the client gets a notification that the rules are not met.



**Figure 3:** Inference engine data flow

### 3.4 Driver Smart Contract Deployment

Driver identification is based on the driver ID in the blockchain network. The driver registration information expands the smart contract and saves it as primary information in smart contract storage. Fig. 4. defines the smart contract detail process in taxi demand service. As explained above, the transaction information between the passenger and driver saves into the blockchain network. The first step is the data service registration which contains the driver information. Next is requesting the driver information, selecting the driver based on provided data, and sending the request to the driver. Next, the driver responds to the passenger if the request for a defined location is acceptable or not and payment records between passenger and driver for a specified location. Finally, upload the records of transactions and investigate the block into the blockchain network based on the defined contract in this system. The smart contract's main role is to set the rules for taxi drivers and passengers based on conditions for identifying the transactions and details of payments for the defined location.

### 3.5 Phase of Ride Requesting

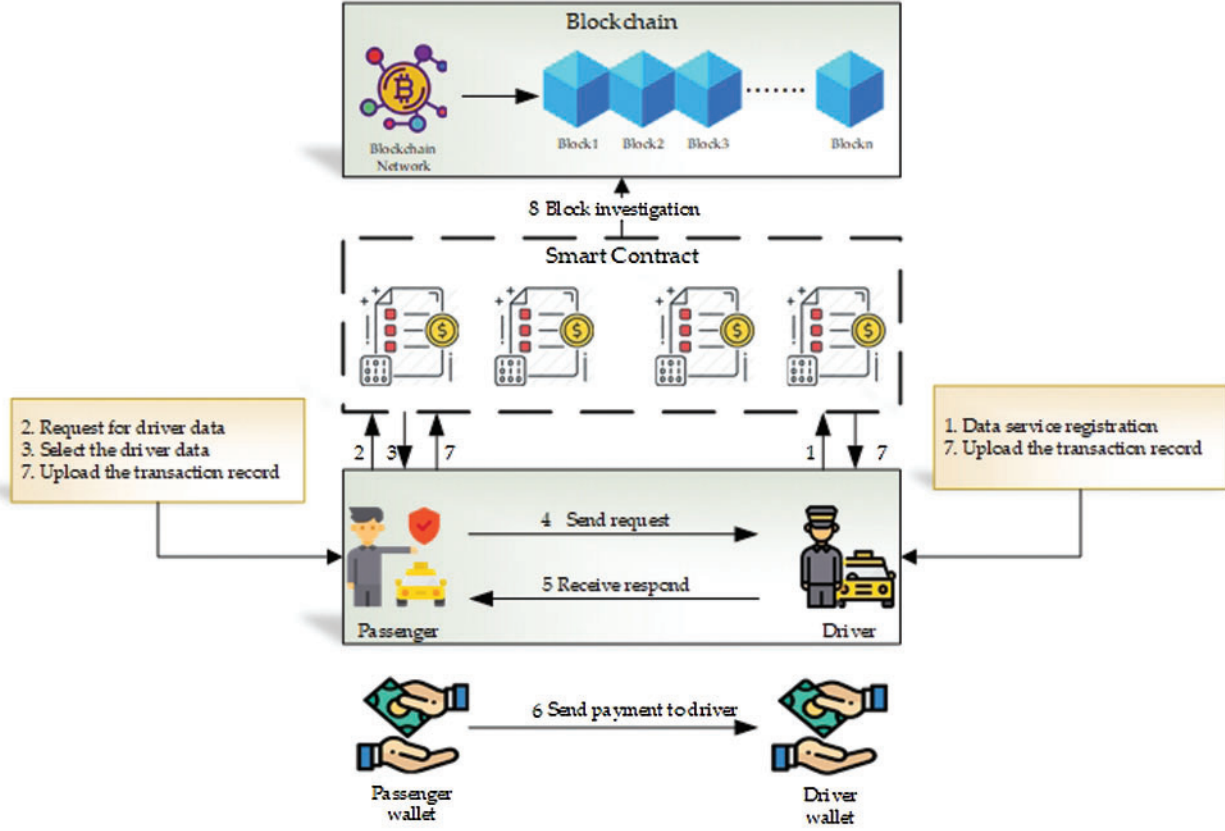
Ride requesting is the process that the passenger is doing based on the nearest options. The passenger who is registered and looking for a ride has the identity number or passenger ID, reputation value, ride request, current location, requested location, and travel distance. Eq. (3) defines the mentioned process.

$$R_p = \{ID_p, Location_p, destination_p, travel_p\} \quad (3)$$

When the request is received, the ID holder  $ID_p$  tries to execute the matching option based on the passenger location and requested destination for the possible driver. Then, the interested driver checks the  $Request_p$  of the passenger and sends the detail of the accepted response, such

as price. Finally, the response of the driver is defined in Eq. (4).

$$Request_d = \{ID_d, Reputation\ value_d, Price_d\} \quad (4)$$



**Figure 4:** Smart contract detail process in taxi demand service

## 4 Implementation

This section provides the detailed implementation process of the proposed system. The environmental setting applied dataset information, the performance analysis of machine learning, and blockchain techniques explained.

### 4.1 Environmental Setting

In this section, a detailed explanation of the environmental setting is discussed. We have presented the system implementation environment detail specification in Tab. 1. The main components of the system are Integrated Development Environment (IDE) that is the composer-playground. The used memory for this process is 32GB. CPU is Intel(R) Core (TM) i7-8700 @3.20 GHz. The programming language is python version 3.6.2. The suitable operating system is Ubuntu Linux 18.04.0 LTS. We have used the Docker engine and composer with version 18.06.1-ce and version 1.13.0. The used environment of Hyperledger Fabric is with the version of 1.2. The applied CLI tool is composer REST server, and the node version is 8.11.4. Tab. 2. shows the development environment of the machine learning algorithms in taxi demand service.



**Table 1:** Development environment of the proposed system

Components	Description
IDE	Composer-Playground
Memory	32 GB
CPU	Intel(R) Core (TM) i7-8700@3.20 GHz
Python	3.6.2
Operating system	Ubuntu Linux 18.04.0 LTS, Windows 10 64bit
Docker engine	Version 18.06.1-ce
Docker composer	Version 1.13.0
Hyperledger fabric	V1.2
CLI tool	Composer REST Server
Node	V8.11.4
Programming language	WinPython-3.6.2, IDE Jupyter Notebook
Optimization algorithm	Decision tree, XGBoost
Browser	Google Chrome

**Table 2:** Dataset overview

Type of matching	N	Rate (%)
Accept	8.010	34.4
Fail (Driver)	7.775	33.7
Fail (Passenger)	6.069	28.2
Others	247	3.7
Total	22.101	100.0

#### 4.2 Data Description

Collected data for the proposed system is from T-map, which focuses on identifying the conditions that taxi drivers reject the passenger drive request. There are in total 22.101 cases in Jeju Island. 10.000 cases were used in this study for processing. [Tab. 2.](#) shows the overview of the applied dataset.

Type of matching shows the acceptance or rejection of request based on 1 and 0. If the request is accepted is one, and if it's rejected is 0. There are many aspects of the destination's place of origin and location, recorded as independent variables in Open Data (ODs) named time, day, weekdays, weekends, land users, socio-economics, etc. The variables related to socio-economics contain the density of population, density of business, and employee density related to destination and origin. This information is captured from the OD portal in Jeju.

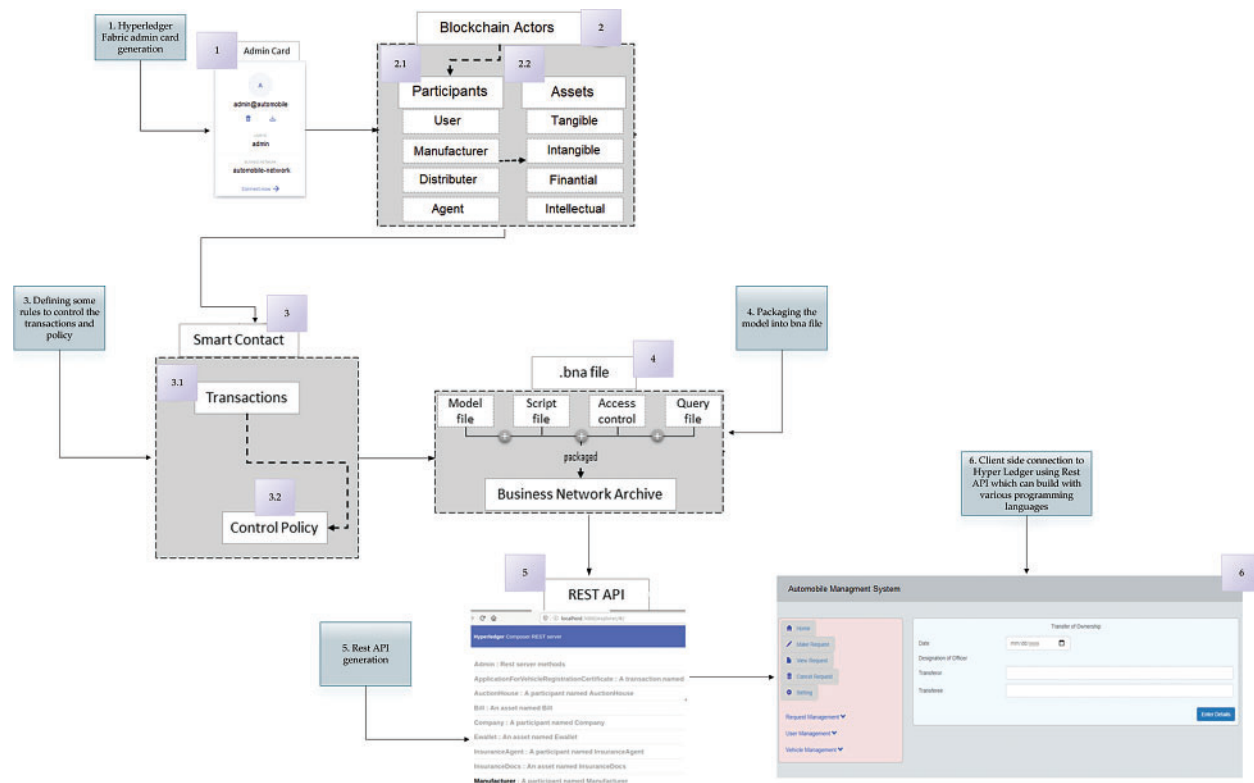
#### 4.3 Performance Analysis of Machine Learning Framework

In this study, the prediction process is based on supervised learning to predict the dependable targets of input data. The prediction process is based on the train set and test set. 80% of data used in the train set, and 20% used in the test set. The analysis of machine learning models summarized in [Tab. 3.](#) The results are applying cross-validation and without cross-validation and recorded as 0.82% without cross-validation and 0.83% with cross-validation in Jeju Island.

Based on the results, the distance was the main reason in Jeju. Cross-validation prominent the importance of variables and non-important ones. There are in total five significant variables in Jeju, named as: A(6) Distance, A(8) Mid-night, A(9) pick-time, A(5) the density of D\_employee, and A(2) the density of O\_employee.

**Table 3:** Machine learning-based analysis records

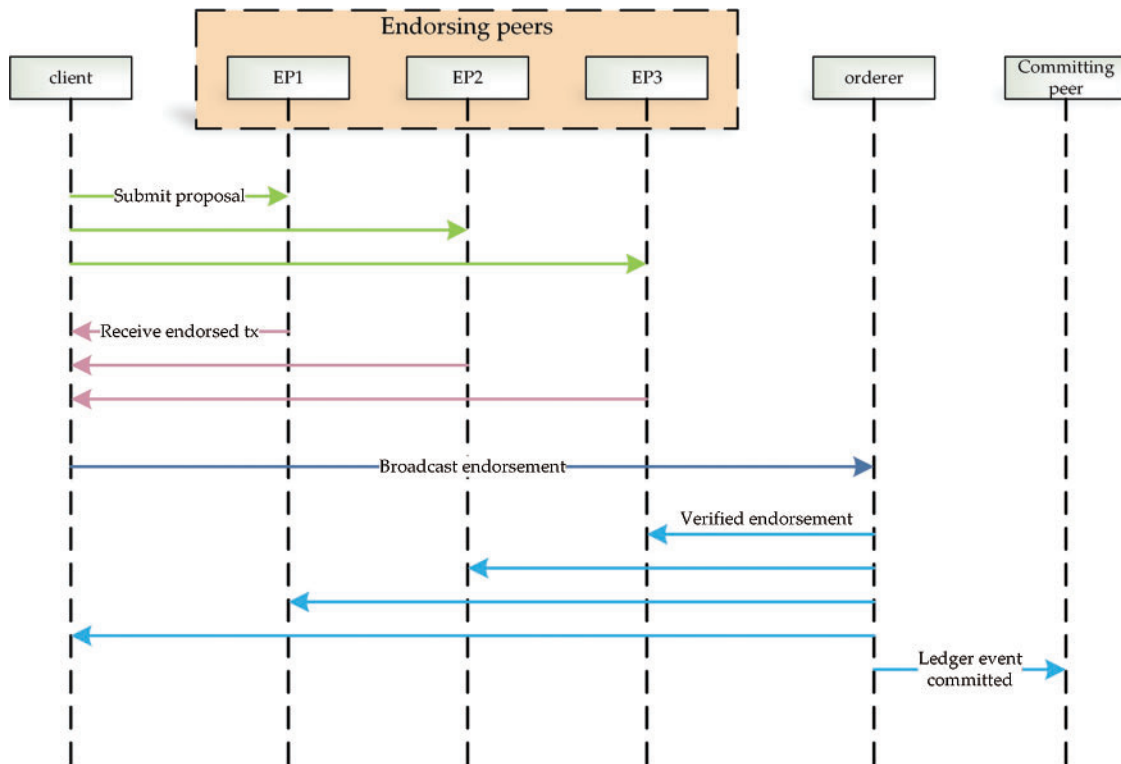
Variable		With cross-validation		Without cross-validation	
		Importance	Rank	Importance	Rank
A (0)	O_density of population	—	—	0.063	6
A (1)	O_density of business	—	—	0.052	7
A (2)	O_density of employee	0.011	4	0.073	2
A (3)	D_density of population	—	—	0.082	2
A (4)	D_density of business	—	—	0.066	5
A (5)	D_density of employee	0.016	3	0.070	3
A (6)	Distance	0.424	1	0.220	1
A (7)	Cloudy	—	—	0.020	9
A (8)	Mid-night	0.172	1	0.070	4
A (9)	Peak time	0.122	2	0.026	8
A (10)	Weekend	—	—	0.014	10



**Figure 5:** Hyperledger fabric framework process for taxi-hailing

#### 4.4 Performance Analysis of Blockchain Framework

In some scenarios, the Blockchain framework needs private data processing that just authorized users can access, endorse, submit, and query. The Hyperledger Fabric can apply special APIs on data series, e.g., put private data, get private data, etc., accessible for authorized peers. Fig. 5. shows the design of the Hyperledger Fabric process. The composer permits generating the admin card for every network. Based on the generated admin card, it becomes possible to define the participant and assets. Rules are defined in the smart contract and packaging process in the business network archive (BNA). The front-end interconnection is processing based on bna files in REST API, and various programming languages can build the client-side. The main reason for using the Hyperledger Fabric in this system is the advantages, which fit well to the presented goal. The need to provide permissioned membership, high level of scalability and trust, need to define the data basis to provide privacy, sensitive data protection and digital keys for protection of the system, etc. Hyperledger Fabric provided IDE integrated development environment (IDE) with a Hyper Ledger Composer name that allows developers to develop customized applications on the blockchain.



**Figure 6:** Transaction flow sequence diagram

#### 4.5 Transaction Process in Blockchain System

Transactional process of the blockchain environment presented in Fig. 6. sequence diagram. The client information contains the ClientID, ChaincodeID, Transaction (tx), a payload of tx, time stamp, and client sign. EP presents the endorser peer to check the signature and perform the transaction. The formation and redundancy of transaction processes by EPs. The signature validity

also checks by the membership service provider. After the checking process, the EP adds the sign into tx and sends the response. After assembling the client, the application-level inspects the responses of the proposal, the transaction reply to the orderer peer. At last, the validation of the transaction is decided by committing peer and update the ledger. Hyperledger Fabric has the highest transaction per second.

## 5 Results

This section predicts the taxi-hailing service based on machine learning and controlling the transactional records and system security based on blockchain.

### 5.1 Taxi-Hailing Demand Forecasting

Taxi-hailing forecasting and various indicators processed by XGBoost algorithm. The main factors of this process defined as time, environment, and taxi demand based on XGBoost. Tab. 4. shows the various indicators prediction in trip demand service. There are three columns named as a type of model, input, and output.  $T$  defines the time, and  $E$  defines the environment. The system output is the next week's demand of the presented system.

**Table 4:** Trip demand predictive indicators

Model	Input	Output
XGB + T	Data of time	Demand of next week
XGB + T + E	Data of time + environment	Demand of next week
XGB + T + TX	Data of time + environment + taxi	Demand of next week

The gride search algorithm was applied to regulate the XGBoost-based hyperparameters. Tab. 5. shows the records of defined hyperparameters. There totally five defined hyperparameters as gamma, learning-rate, max depth, min\_child\_weight, and n-estimators.

### 5.2 Performance Analysis of Taxi Demand Service Based on Machine Learning Algorithms

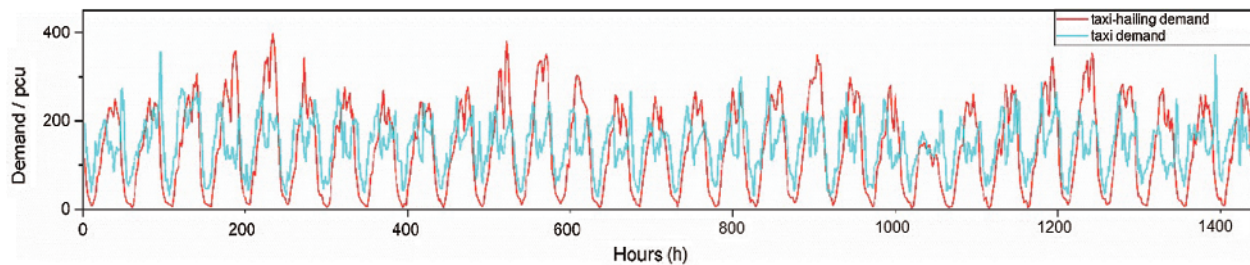
In this process, Jeju island selected for our research object because it contains various tourist spots and has lots of visits and trip demands. The taxi-hailing information is divided based on the time, as shown in Fig. 7. Taxi demand and ride-hailing service are defined as regular and based on increasing the ride-hailing service, taxi demand service decreases. Detail records are presented in Tab. 6.

Fig. 8. presents the taxi demand prediction results based on the XGBoost algorithm in the train and test set. The prediction results of both train and test sets are recorded.

The experimental results present the XGBoost performance in defined hyperparameters and optimal value and clearly show that XGBoost model inputs and outputs in terms of demand in the coming week. Similarly, the MAPE and RMSE of the presented system were evaluated based on various situations in terms of time, environment, and taxi. Evaluating the performance of XGBoost prediction in various hours defined based on MAPE and RMSE. Fig. 9. Presents the MAPE of XGBoost in three different time stamps as eleven in the morning and five in the afternoon, which in this time the model performance is better than other times.

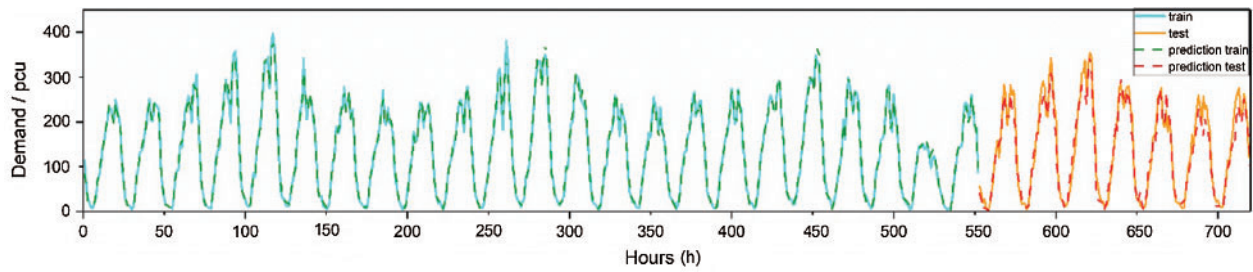
**Table 5:** Model hyperparameter settings

Model	Hyperparameters	Optimal value
XGB + T	Gamma	0.5
	Learning-Rate	0.5
	Max-Depth	14
	Min-Child-Weight	13
	n-estimator	49
XGB + T + E	Gamma	0.5
	Learning-Rate	0.4
	Max-Depth	12
	Min-Child-Weight	13
	n-estimator	39
XGB + T + TX	Gamma	0.5
	Learning-Rate	0.4
	Max-Depth	12
	Min-Child-Weight	12
	n-estimator	46

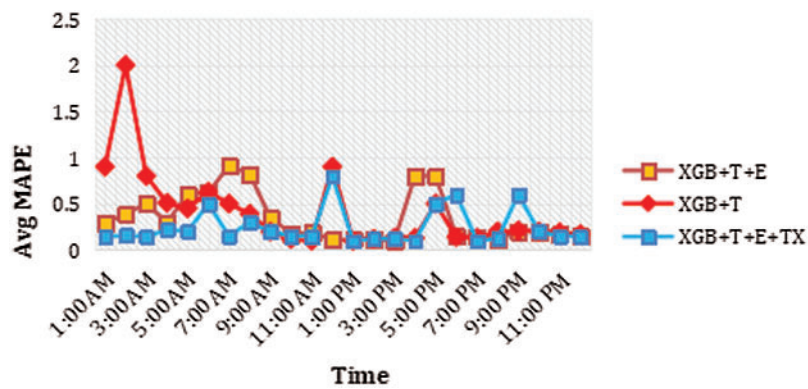
**Figure 7:** Online-taxi hailing and taxi demand service in Jeju**Table 6:** Proposed system resource utilization analysis

Type	Name	CPU (Max %)	CPU (Min %)	Memory (Max)	Memory (Min)	Traffic In	Traffic Out
Process	Local-client	12.86	6.36	85.3 MB	84.0 MB	583 KB	253.4 KB
Docker	Peer1.Trainer	12.93	6.69	80.6 MB	87.8 MB	576.5 KB	240.9 KB
Docker	Peer0.Trainee	11.76	7.62	513.5 MB	512.5 MB	2.8 MB	934.5 KB
Docker	Peer0.Trainer	10.99	7.19	512.3 MB	511.2 MB	2.8 MB	929.7 B
Docker	Peer1.Trainee	0.00	0.00	9.7 MB	9.7 MB	0 B	0 B
Docker	Orderer	5.63	2.40	24.6 MB	21.9 MB	2.3 MB	3.4 MB
Docker	ca_nodeTaxi	0.00	0.00	11.0 MB	11.1 MB	657 B	0 B

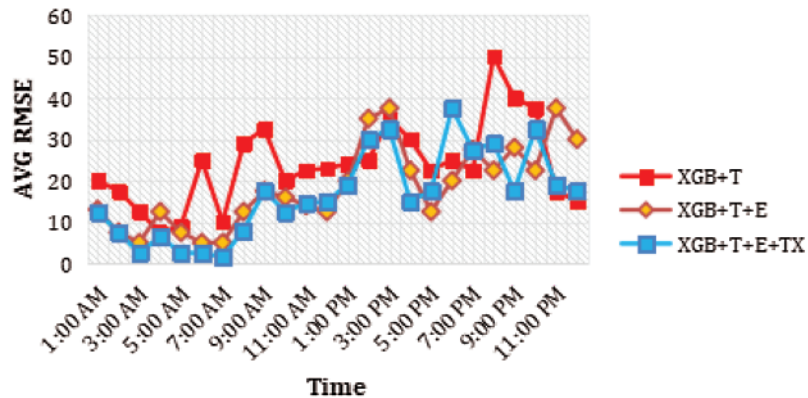
Fig. 10. The RMSE records show that the model performance in XGB + T + E + TX is good except eleven and twelve in the morning, four and five in the afternoon.



**Figure 8:** Taxi demand service based on XGBoost



**Figure 9:** Average MAPE of XGBoost prediction model in various hours



**Figure 10:** Average RMSE of XGBoost prediction model in various hours

### 5.3 Blockchain Performance Analysis for Secure Taxi Demand Service

The security analysis was processed against the probability of attacks for the taxi demand environment. Four factors were covered for this process: key attack, false data injection attack, replay attack, and man in the middle attack. The key attack secures the taxi demand service based on encryption to make it difficult for the attacker, which requires high computational power that is tough for the intruder. The false data injection performs before validating records, and every



node necessary verification and authority for the successful consensus mechanism. The replay attack requires the private key for agreement between nodes and man in middle attack authorize nodes based on the temporary private key for all sessions to avoid a middle attack. The presented framework carries out the secure framework for protecting data, locking access points of various devices, enabling fault tolerance, data encryption, and decentralization. The functionality of smart contracts in this system is to scale down the costs.

#### 5.4 Simulation Results

Fig. 11. presents the transaction latency based on three different user groups following the Hyperledger Fabric. Each group contains 700, 1200, and 1700 users. The statistical measurement, e.g., Min, Max, Avg, also evaluates the performance of the blockchain platform. For example, the 700 users become stable with the Min of 37 TPS and Max of 40 TPS.

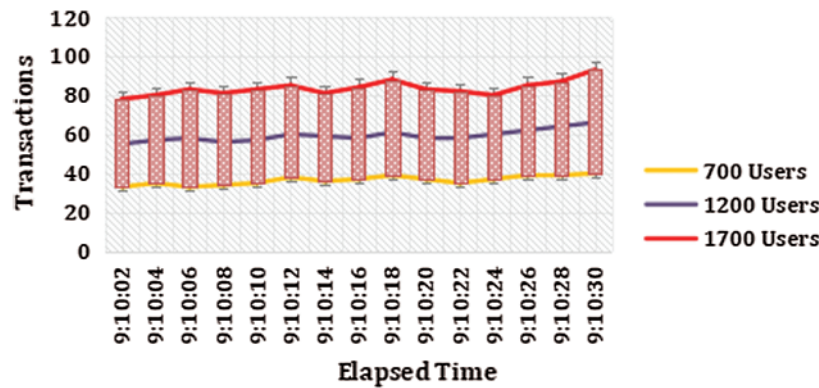


Figure 11: Hyperledger fabric transaction records per second

Similarly, the 1200 user groups contain the Min of 57 and Max of 63 TPS. Finally, the 1700 user group contains the Min of 79 and Max of 95 TPS. Based on the presented records, it is visible that the throughput increase based on increasing the number of users. However, the user numbers do not affect the performance of the system in terms of transactions.

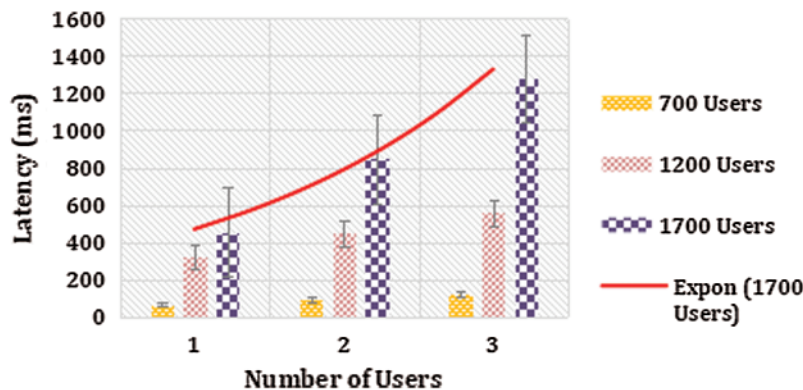


Figure 12: Query transaction latency

Fig. 12. shows the query transaction function of the proposed taxi demand case study in terms of latency. Three groups performance designed in the blockchain platform. The latency is based on getting a response from the blockchain. Based on increasing the number of users, the query transaction latency also increases in the blockchain platform.

The performance of the presented system was analyzed based on the resource utilization records, such as CPU usage in terms of Max and Min. The usage of memory (Max and Min). The traffic in and out. Tab. 6. Shows the presented blockchain-based taxi demand service improves the performance of the network with efficient system utilization resources. Tab. 7. shows the result comparison of the machine learning algorithm with other state-of-arts. It is shown that XGBoost has quite higher accuracy than others, with an accuracy of 92%.

**Table 7:** Proposed system resource utilization analysis

Algorithms	Accuracy (%)
SVM	79
RF	85
LSTM	87
Proposed system	92

## 6 Conclusion

In this study, the factors show the reliability of the taxi apps in Jeju island. Weather conditions, variables, distances, time of the day, etc., are the applied factors to measure the success of the proposed system. The key approach of the proposed system is an integration of machine learning and blockchain. Accomplishing XGBoost and Decision tree algorithms for analysis before cross-validation and after cross-validation. The performance analysis of the blockchain framework is evaluated in this system. Hyperledger Fabric framework shows the maximum efficiency for securing the taxi demand service using smart contracts and other factors. The overall result presents the good performance and power of the system in the training and testing set. Consequently, using the machine learning algorithms found the importance of distance for the acceptance rate of drivers in terms of passenger ride requests. The main focus of this manuscript is to improve the security and intelligence of the ride-hailing system based on the integration of machine learning and blockchain framework. We have analyzed this process based on the taxi usage records in Jeju Island, South Korea, with the passengers rating and satisfaction from the kakao taxi and t-map. Ride-hailing requests from passengers and probability of the acceptance from the driver for the defined location were analyzed. The designed framework improves the passenger's transaction records' security and reduces the waiting time of passengers for the requested location.

**Funding Statement:** This research was financially supported by the Ministry of Small and Medium-sized Enterprises (SMEs) and Startups (MSS), Korea, under the “Regional Specialized Industry

Development Program (R&D, S3091627)” supervised by Korea Institute for Advancement of Technology (KIAT).

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] L. Ma, T. Li, J. Wu and D. Yan, “The impact of E-hailing competition on the urban taxi ecosystem and governance strategy from a rent-seeking perspective,” *The China E-Hailing Platform. Journal of Open Innovation: Technology, Market, and Complexity*, vol. 4, no. 3, pp. 35, 2018.
- [2] M. Dabbagh, K. K. Choo, A. Beheshti, M. Tahir, N. S. Safa *et al.*, “A survey of empirical performance evaluation of permissioned blockchain platforms: Challenges and opportunities,” *Computers & Security*, vol. 100, pp. 102078, 2021.
- [3] K. Kato, Y. Yan and H. Toyoizumi, “Blockchain application for rideshare service,” in *8th Int. Conf. on Logistics, Informatics and Service Sciences (LCISS)*, Toronto, Ontario, Canada, pp. 1–5, 2018.
- [4] Z. Zheng, S. Xie, H. Dai, X. Chen, H. Wang *et al.*, “An overview of blockchain technology: Architecture, consensus, and future trends,” *IEEE Int. Congress on big Data (BigData Congress)*, Orlando, United States, pp. 557–564, 2017.
- [5] T. M. Choi, S. W. Wallace and Y. Wang, “Big data analytics in operations management,” *Production and Operations Management*, vol. 27, no. 10, pp. 1868–1883, 2018.
- [6] F. Jamil and D. H. Kim, “Enhanced kalman filter algorithm using fuzzy inference for improving position estimation in indoor navigation,” *Journal of Intelligent & Fuzzy Systems*, vol. 40, pp. 1–15, 2021.
- [7] S. Ahmad, S. Malik, I. Ullah, D. H. Park, K. Kim *et al.*, “Towards the design of a formal verification and evaluation tool of real-time tasks scheduling of IoT applications,” *Sustainability*, vol. 11, no. 1, pp. 204–226, 2019.
- [8] S. Ahmad, L. Hang and D. Kim. “Design and implementation of cloud-centric configuration repository for DIY IoT applications,” *Sensors*, vol. 18, no. 2, pp. 474–494, 2018.
- [9] S. Ahmad and D. Kim, “A multi-device multi-tasks management and orchestration architecture for the design of enterprise IoT applications,” *Future Generation Computer Systems*, vol. 106, pp. 482–500, 2020.
- [10] F. Jamil, H. K. Kahng, S. Kim and D. H. Kim. “Towards secure fitness framework based on IoT enabled blockchain network integrated with machine learning algorithms,” *Sensors*, vol. 21, no. 5, pp. 1640, 2021.
- [11] B. Jiang and L. Tian, “Collaborative consumption: Strategic and economic implications of product sharing,” *Management Science*, vol. 64, no. 3, pp. 1171–1188, 2018.
- [12] S. Ahmad, F. Jamil, N. Iqbal and D. H. Kim, “Optimal route recommendation for waste carrier vehicles for efficient waste collection: A step forward towards sustainable cities,” *IEEE Access*, vol. 8, pp. 77875–77887, 2020.
- [13] S. Ahmad, F. Jamil, A. Khudoyberdiev and D. H. Kim, “Accident risk prediction and avoidance in intelligent semi-autonomous vehicles based on road safety data and driver biological behaviours,” *J. Intell. Fuzzy Syst.*, vol. 38, no. 4, pp. 4591–4601, 2020.
- [14] N. Iqbal, F. Jamil, S. Ahmad and D. H. Kim, “A novel blockchain-based integrity and reliable veterinary clinic information management system using predictive analytics for provisioning of quality health services,” *IEEE Access*, vol. 9, pp. 8069–8098, 2021.
- [15] F. Jamil, N. Iqbal, S. Ahmad and D. H. Kim, “Peer-to-peer energy trading mechanism based on blockchain and machine learning for sustainable electrical power supply in smart grid,” *IEEE Access*, vol. 9, pp. 39193–39217, 2021.
- [16] S. Ahmad, I. Ullah, F. Jamil and D. H. Kim, “Toward the optimal operation of hybrid renewable energy resources in microgrids,” *Energies*, vol. 13, no. 20, pp. 5482, 2020.
- [17] L. Jiang, S. Dimitrov and B. Mantin, “P2P marketplaces and retailing in the presence of consumers’ valuation uncertainty,” *Production and Operations Management*, vol. 26, no. 3, pp. 509–524, 2017.

- [18] H. K. Bhargava, B. C. Kim and D. Sun, "Commercialization of platform technologies: Launch timing and versioning strategy," *Production and Operations Management*, vol. 22, no. 6, pp. 1374–1388, 2013.
- [19] J. Anderson, G. G. Parker and B. Tan, "Platform performance investment in the presence of network externalities," *Information Systems Research*, vol. 25, no. 1, pp. 152–172, 2014.
- [20] S. Fay and J. Xie, "The economics of buyer uncertainty: Advance selling vs. probabilistic selling," *Marketing Science*, vol. 29, no. 6, pp. 1040–1057, 2010.
- [21] S. Banerjee, C. Riquelme and R. Johari, "Pricing in ride-share platforms: A queueing-theoretic approach," Available: SSRN 2568258, 2015.
- [22] X. Wang, F. He, H. Yang and H. O. Gao, "Pricing strategies for a taxi-hailing platform," *Transportation Research Part E: Logistics and Transportation Review*, vol. 93, pp. 212–231, 2016.
- [23] F. Jamil, S. Ahmad, N. Iqbal and D. H. Kim, "Towards a remote monitoring of patient vital signs based on IoT-based blockchain integrity management platforms in smart hospitals," *Sensors*, vol. 8, no. 20, pp. 2195, 2020.
- [24] F. Jamil, M. A. Iqbal, R. Amin and D. H. Kim, "Adaptive thermal-aware routing protocol for wireless body area network," *Electronics*, vol. 8, no. 1, pp. 47, 2019.
- [25] F. Jamil, N. Iqbal, S. Ahmad and D. H. Kim, "Toward accurate position estimation using learning to prediction algorithm in indoor navigation," *Sensors*, vol. 20, no. 16, pp. 4410, 2020.
- [26] F. Jamil and D. H. Kim, "Improving accuracy of the alpha–beta filter algorithm using an ANN-based learning mechanism in indoor navigation system," *Sensors*, vol. 19, no. 18, pp. 3946, 2019.
- [27] N. Iqbal, F. Jamil, S. Ahmad and D. H. Kim, "Toward effective planning and management using predictive analytics based on rental book data of academic libraries," *IEEE Access*, vol. 8, pp. 81978–81996, 2020.
- [28] G. P. Cachon, K. M. Daniels and R. Lobel. "The role of surge pricing on a service platform with self-scheduling capacity," *Sharing Economy*, vol. 19, no. 3, pp. 101–113, 2017.
- [29] B. Kazaz, S. Webster and P. Yadav, "Interventions for an artemisinin-based malaria medicine supply chain," *Production and Operations Management*, vol. 25, no. 9, pp. 1576–1600, 2016.
- [30] L. Moreira-Matias, J. Gama, M. Ferreira and L. Damas, "A predictive model for the passenger demand on a taxi network," in *15th Int. IEEE Conf. on Intelligent Transportation Systems*, pp. 1014–1019, 2012.
- [31] K. Zhang, Z. Feng, S. Chen, K. Huang, G. Wang *et al.*, "A framework for passengers demand prediction and recommendation," in *IEEE Int. Conf. on Services Computing (SCC)*, San Francisco, CA, USA, pp. 340–347, 2016.
- [32] N. Davis, G. Raina and K. Jagannathan, "A multi-level clustering approach for forecasting taxi travel demand," in *2016 IEEE 19th Int. Conf. on Intelligent Transportation Systems (ITSC)*, Rio de Janeiro, Brazil, pp. 223–228, 2016.
- [33] Wallsten, "The competitive effects of the sharing economy: How is uber changing taxis," Technology Policy Institute, Washington, D.C., USA, vol. 22, pp. 1–21, 2015.
- [34] K. T. Seow, N. H. Dang and D. H. Lee, "Towards an automated multiagent taxi-dispatch system," in *IEEE Int. Conf. on Automation Science and Engineering*, Scottsdale, Arizona, pp. 1045–1050, 2007.
- [35] L. Yue, H. Junqin, Q. Shengzhi and W. Ruijin, "Big data model of security sharing based on blockchain," in *3rd Int. Conf. on Big Data Computing and Communications (BIGCOM)*, Chengdu, China, pp. 117–121, 2017.
- [36] M. Singh and S. Kim, "Blockchain based intelligent vehicle data sharing framework," arXiv preprint arXiv: 1708.09721, 2017.
- [37] L. Li, J. Liu, L. Cheng, S. Qiu, W. Wang *et al.*, "Creditcoin: A privacy-preserving blockchain-based incentive announcement network for communications of smart vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 7, pp. 2204–2220, 2018.
- [38] J. Kang, Z. Xiong, D. Niyato, D. Ye, D. I. Kim *et al.*, "Toward secure blockchain-enabled internet of vehicles: Optimizing consensus management using reputation and contract theory," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 2906–2920, 2019.

- [39] Z. Lu, W. Liu, Q. Wang, G. Qu and Z. Liu, "A privacy-preserving trust model based on blockchain for VANETs," *IEEE Access*, vol. 6, pp. 45655–45664, 2018.
- [40] A. M. Prasad, L. R. Iverson and A. Liaw, "Newer classification and regression tree techniques: Bagging and random forests for ecological prediction," *Ecosystems*, vol. 9, no. 2, pp. 181–199, 2006.
- [41] M. Salimitari and M. Chatterjee, "A survey on consensus protocols in blockchain for iot networks," *ArXiv Preprint ArXiv:1809.05613*, 2018.
- [42] F. Jamil, L. Hang, K. H. Kim and D. H. Kim, "A novel medical blockchain model for drug supply chain integrity management in a smart hospital," *Electronics*, vol. 8, no. 5, pp. 505, 2019.
- [43] Z. Shahbazi and Y. C. Byun, "A framework of vehicular security and demand service prediction based on data analysis integration with blockchain approach," *Sensors*, vol. 21, no. 10, pp. 3314, 2021.
- [44] Z. Shahbazi and Y. C. Byun, "Smart manufacturing real-time analysis based on blockchain and machine learning approaches," *Applied Sciences*, vol. 11, no. 8, pp. 3535, 2021.