Optimizing Waste Management Using IoT and Blockchain Based Machine Learning Forecasting Techniques

MSc Research Project
Data Analytics

Raksha Muddegowdana Koppalu Prakasha
Student ID: x19193181

School of Computing
National College of Ireland

Supervisor: Dr. Rashmi Gupta

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Raksha Muddegowdana Koppalu Prakasha
Student ID:	x19193181
Programme:	Data Analytics
Year:	2020
Module:	MSc Research Project
Supervisor:	Dr. Rashmi Gupta
Submission Due Date:	17/12/2020
Project Title:	Optimizing Waste Management Using IoT and Blockchain Based Machine Learning Forecasting Techniques
Word Count:	9830
Page Count:	28

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	1st February 2021

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Optimizing Waste Management Using IoT and Blockchain Based Machine Learning Forecasting Techniques

Raksha Muddegowdana Koppalu Prakasha
x19193181

Abstract

With the growing urbanization, modeling an efficient waste management system has become a huge challenge. The current process for the same is hindered by the requirement of vast physical and financial resource. In cities with vast populations, we often find garbage bins overflowing with wastes, which leads to spread of harmful disease and environmental pollution. Additionally, the poor management of these accumulated wastes also reflects badly on the economy of a state/country/-city. Introduction of smart city using IoT technology has led to raise of innovative solutions to improve the process of waste management. But it is still not enough to envision an optimized system because of the own drawbacks of the technology used(IoT).In this research project, a novel approach to optimize the waste management system is explored using an ecosystem of IoT-Blockchain and Machine Learning. The IoT is used to provide smart monitoring of waste levels at garbage bins in a city, and decentralized Blockchain network is used to create an ecosystem of trust, reliability, and security to the IoT collected data. And finally Machine Learning forecasting algorithms are used to carry out analysis and forecast future waste levels of a garbage bin and thereby facilitate the build of an efficient waste collection and disposal system with minimized physical and financial resource use. Additionally, in this research carried out, various machine learning forecasting techniques, ARIMA, LSTM, CNN, MLP are implemented to forecast future waste fill percentage of a garbage bin and their performances are evaluated.

Keywords: ***Machine Learning, IoT, Blockchain, LSTM, ARIMA, CNN, MLP***

1 Introduction

With the growing urban population, there is a huge spike in the quantity of waste accumulated. In the current growing world, the waste accumulated is a prime indicator of socioeconomic development of a countryFerrera de Lima (2012), a city, or a town. And as a result, we can say that the progressing economy corresponds to the enormous volumes of waste produced. But the stumbling block in this context is the ineffectiveness and bottlenecks in the process of waste management. If on one hand, waste produced acts as an indicator of socioeconomic development, the effective management of these produced wastes also plays a vital role in the overall development of a country, city, or town.

In the current scenario, the effective management of the waste is posing to be an ever-increasing problem in all major urban cities and towns. The waste management process involves activities for monitoring the industrial and domestic waste produced in a city or town, from the origination of waste through the collection of these waste, transportation, and final dumping/recycling of the collected waste. And establishing a proper waste management system has been a significant challenge for cities across the globe. The current system of waste management includes a labor-intensive process involving a set and pre-decided program for waste collection, and this process has failed to meet or uphold the requirements to address the efficiency of resource allocation to preserve a safer and clear environment in the cities. The ineffectiveness of this process can be observed by the accumulation of enormous volumes of waste at bins across the cities, wherein the cases of waste lying on the ground, on the roads, and bins overflowing with the collected wastes are usual. And the absence of an effective waste management process is leading to a prime cause of environmental pollution and the growth of deadly diseases posing a colossal risk to human healthNoche et al. (2010). The effectiveness of the system depends on the understanding of the accessible resources and economic requirements because even the punctual system of waste disposal involves the bottlenecks of the need for vast physical labor resources, the high cost of scheduling, and improper scheduling.

The growing concept of IoT (Internet of Things) and smart devices is a step towards refining the process of efficient waste management by using enhanced analytical solutions and smart monitoring. IoT-powered waste management solutions are based on improving the efficiency of waste collection and its treatment(recycling/dumping) with significant cost savings, effective resource use, and increased productivity, which also involves combining other low-cost technologies. The biggest drawback of the IoT based solution is the lack of data privacy, security, and reliability, and these drawbacks are a huge hindrance in implementing a system of automation and smart monitoring in the scope of waste management.

Blockchain is yet another emerging trend in the technological world that is focused around the concepts of decentralization, data security, privacy, and reliability and in this research paper, this revolutionary concept of Blockchain is used in combination with the IoT to overcome the drawbacks of IoT solutions mentioned above. And this combination of IoT-Blockchain ecosystem is extended to carry out Realtime Analysis and forecasting of future waste production trends at the IoT fitted smart bins using the concepts of Machine Learning. The aim of the solution proposed is to identify the trends of garbage collected at IoT fitted smart bins across a city or a town using various Machine Learning algorithms and thereby designing an efficient waste collection system with route optimization.

1.1 Motivation

Efficient waste management is a prime requirement in the daily life of an urban city or a town, and the current scenario of waste management is restricted in terms of inefficient resource management and high cost of operation. We often observe waste bins overflowing with collected wastes and at times take days for its removal because of a fixed pre-set schedule of the collection as part of the current system. And in the retrospect, there are also cases wherein the bins that are scheduled to be emptied might not have any waste collected for disposal. This leads to a waste of fuel and time of garbage trucks scheduled to pick these wastes. With the increasing population, an increase in the production of industrial and domestic waste has seen a huge spike, and an efficient waste management process is necessary to regulate it. The introduction of IoT in the form of smart bins paved a path for optimizing the process of waste collection through route optimization and timely notification of filled garbage bins, but this still isn't enough to build a fully automated system of waste management which is less labor-intensive and cost-efficient. The current IoT based waste management solution is as seen in Figure 1. The other serious drawback of IoT-equipped smart waste management is its inability to establish security, trust, privacy, and reliability in its ecosystem. As of 2020, it was reported that more than 98% of the IoT sensor collected data are unencrypted and susceptible to cyber attacks, making the whole system unreliable. The revolutionary technology of Blockchain which works on the grounds of data security, privacy, and trust can be used to overcome the above-mentioned drawbacks of the IoT system. In order to automate the process of waste management and make the human intervention as negligible as possible, intelligent analytical techniques such as AI and Machine Learning is necessary. Machine-learning techniques such as ARIMA, Convolutional Neural Network, LSTM (Long Term Short Memory), Multi-Layer Perceptron, etc, are predominantly used for forecasting. Keeping all the above-mentioned points in consideration, this research aims at answering the question in section 1.2,

Figure 1: Current IoT Based Waste Management System

1.2 Reasearch Question

RQ: "How efficiently can the machine learning forecasting techniques help forecast waste fill levels of smart bins and thereby help optimize the Waste Management Systems in urban cities?"

1.3 Research Objective

The objective of this research are:

- To analyse the patterns and trends of the waste fill levels of smart bins.
- To forecast waste fill levels of smart bins.
- To carry out performance evaluation of different forecasting models used to forecast waste fill levels.

The research report consists of sections, detailing the review of previous work done in the field of the research topic, the methodology carried out in the research, implementation carried out, evaluation of the implemented models, conclusion, and future work.

2 Related Work

With the growing population, waste management becomes a necessary action of any developing/developed city and its economy. The current ecosystem of waste management is limited by its bottlenecks of being hugely labor intensive, unsystematic, and costly. And because of the above-mentioned bottlenecks, we often observe the cases of garbage bins overflowing with collected waste, and waste accumulated at the sides of the roads and empty yards, and these scenarios represent the negative impact on the socio-economic development of an urban city or a town. The current process of waste management follows a predetermined schedule for waste collection, which leads to the cases wherein the garbage bins overflowing with waste go uncollected as opposed to one which is half full or even empty. This also has led to inefficient use of city resources.

IoT has managed to induce smart intelligence to resources involved in the process of waste management, but it has failed to achieve optimizing efficiency in the entire process because of its own drawbacks. Blockchain which is yet another revolutionary technology will help overcome the drawbacks of the IoT ecosystem and take a step further towards optimizing the system of waste management. Blockchain induces the properties of reliability, trust, and decentralization in the IoT ecosystem facilitating the smart devices to work autonomously with no dependencies on a centralized entity Atlam et al. (2018). Plugging in the concepts of Machine Learning to an ecosystem of IoT-Blockchain will help automate the whole process of waste management wherein it can be used to forecast the future trends and patterns of waste collection across the city and thereby help design an efficient, optimized system of waste management with the route, cost, and resource optimization. In this section, we will carry out the review of the past academic works and research in the scope of our proposal for efficient waste management, and accordingly, this section is divided into the following subsections:

- Literature Based on Working of Current Waste Management Process and their Drawbacks
- Literature Based on Application of Internet of Things, Blockchain and Machine Learning in Scope of Waste Management
- Literature Based on Forecasting Techniques of Machine Learning

2.1 Literature Based on Current Waste Management Process and their Drawbacks

The growth in Urbanization has seen a huge spike in the amount of waste collected, especially the solid waste collected. In most developing cities across the globe, the process of waste management still remains an outdated process hindering the growth of the city and country in its socio-economic terms. Efficient waste management is necessary for a cleaner and healthier environment. The current process of waste management has many bottlenecks such as improper resource utilization, high cost of operation, local environmental regulations, etc. Hence Waste Management is a challenging task for any developing country.

A case study carried out by Khatib et al. (2007), on solid waste management in Palestinian territories, compares the different elements of the waste management process such as collection, cost allocation, local authorities and regulation, disposal, etc, and reviews its shortcomings. The research revealed the lack of proper waste disposal, collection, and treatment for more than 90% of the population. The high cost imposed on the residents for the process of waste collection led to the illegal burning of wastes in open grounds leading to environmental pollution and decrease in air quality.

Research carried out by Contreras et al. (2010) on the city of Boston and Yokohama listed the four driving categories in the management of solid waste. Legal drivers which include laws and regulations, technology development and institutional drivers including available technologies, regional and international drivers including solid waste flow as recyclable resources and socio-economic drivers including population trends and public awareness are the four drivers listed in the research. Improper disposal facilities and increasing population were the prime challenges in the system of waste management.

The research carried out by Mohee et al. (2015) reviews the status of the waste management process in Small Island Developing States(SIDS). Comparison of waste generated between Caribbean SIDS, Pacific SIDS, and Mediterranean SIDS is carried out, and it was seen that on average 1.29 kgs of waste were produced per capita per day in these regions. These countries reported facing shortcomings in the process of waste collection, transfer, and treatment with outdated collection systems, inefficient resource management, outdated equipment, and inaccessible waste dump sites. The major problem of waste management in these countries was the practice of landfill disposal, illegal dumping, and waste burning at the expense of sustainable waste management practices such as recycling, composting, etc.

India, according to the research carried out by Kumar et al. (2017), has seen a huge spike in development in terms of urbanization, economy, and industrialization, and as a result, the amount of solid waste generated has also witnessed a huge leap in the measurement of per-person quantity. The country that is densely populated and home to many ethnicities and traditions face a massive challenge in dealing with the issue of solid waste management. The public population of the country lacks awareness of waste management and depends completely on the local municipal bodies to take care of the waste generated. And this has lead to a complete lack of proper segregation and management of solid waste. The research carried out by Narayana (2009) shows that local municipal council authorities of cities across India are responsible for waste collection and its management. Municipal authorities have appointed designated waste collectors to collect solid waste (industrial and domestic) with the help of associated resources such as dump trucks, pull carts, etc. Littering is yet another serious problem in India, and the authorities have invested a large sum to introduce schemes to regularly sweep the roads and control littering, but it has shown no impact on the problem of deducing huge financial loss. In conclusion, we can say that for a developing country like India, waste management is a massive challenge and requires a strict waste policy and proper management.

Joshi and Ahmed (2016) emphasizes the urgent need to promote active participation and educate the public with proper and safe disposal techniques. They mention the importance of implementing an integrated waste management system, taking into account the merits of the existing sectors of waste management for their cost-effectiveness and sustainability. The research concludes the need for awareness for safe waste disposal, the use of fitting technology, and public-government partnership to achieve an efficient waste management system.

2.2 Literature Based on Application of Internet of Things, Blockchain and Machine Learning in Scope of Waste Management

2.2.1 Application of Internet of Things in Scope of Waste Management

Internet of Things has helped grow the system of waste management towards efficiency. It optimizes the process of waste management by installing the smart intelligence to the resources(garbage bins) through sensors and in turn, transforms the whole system to a data-driven process, and thereby help to drive the process towards being cost and resource efficient. The smart sensor fit garbage/waste bins keep track of their fill levels and notify for collection on being full. Additionally, the data-driven process established by IoT can be used to carry out pattern analysis, route optimization, etc.

Route Optimization using intelligent transportation system us Intelligent Transportation System is another critical implementation of IoT in the scope of a smart city. Research by Medvedev et al. (2015) presented with a novel solution using the combination of smart waste management and intelligent transport system. The intelligent transportation system includes smart surveillance system that comprises of IoT components like actuators, sensors, and cameras for establishing QoS in the ecosystem of waste collection and route optimization. The novel approach also includes a built in decision system to handle ineffectiveness in waste collection with a goal to facilitate a high quality of service to each individual involved.

Gutierrez et al. (2015) carried out research to investigate the efficiency of IoT system integrated in combination to Geographical Information Systems(GIS), and data-access networks in the scope of waste management. The research presented with an unconventional approach for waste collection using the concept of IoT sensors fit smart trashcans, that are induced with the capabilities to read, collect, and transmit data about their fill levels to a spatio-temporal network, which in turn uses the concept of graph theory to effectively carry out the waste collection. The research also gives an insight into the analogy of traditional waste management systems.

Internet of Things has been playing a prominent role in maintaining a clean environment and introducing the feature of smart city in many cities across the globe. The affordability of IoT devices are pushing the boundaries of IoT solutions and making it more feasible, and IoT used in the scope of waste management will help reduce the operational costs and investments involving both citizens and governing authority of a city or a town. The research carried out by Al-Jabi and Diab (2017) sheds light on the need to grow and expand the horizons of IoT technology, additional to the need and significance for a IoT driven design and implementation of a smart waste management solution. They propose the design of a smart waste bin equipped with ultrasonic sensors, RFID tags, which will facilitate the measure of waste production and human (citizen) interaction with the process of waste management.

Saha and Auddy (2017) explores the major research conducted in the scope of waste management and subsequently have proposed eight major strategies of waste management to increase its efficiency. These strategies include, recycling, composting, landfills, etc. These strategies are exploited in combination to IoT ecosystem at different stages of waste management to reduce cost, reduce waste generation and efficient resource management.

2.2.2 Application of Blockchain to Internet of Things

Wörner and von Bomhard (2014) proposes the concept of Blockchain of things that can enable the exchange and trading of crypto-coins (bitcoins) in a regulated IoT environment. The IoT sensors are established as nodes on the distributed ledger, made detectable through a sensor repository and data from these sensors are processed using Smart Contracts.

The use of IoT technology has seen a huge growth and impact on various fields ranging from military, smart city to medical science, etc. Many researches have been carried out to review the concerns of security in IoT solutions, by carrying out investigations on publicly available IoT datasets. In the same context, the research carried out by Banerjee et al. (2018) implied the need for a standard sharing platform for sensitive IoT data amongst the research community and proposed the use of Blockchain technology to ensure the secure and reliable exchange of IoT data. The proposal includes two conceptual blockchain solutions for data security and integrity.

The key challenge in the success story of a IoT system lies in the extent of its data security, reliability, and privacy. Blockchain is a revolutionary technology disrupting the world of technology with it unique features of decentralization and distributed system that facilitate data privacy, security, and reliability. Research carried out by Restuccia et al. (2018) explores the possibility of combining the blockchain features with IoT ecosystem wherein the computation-intensive blockchain is used to process bulk volumes of IoT data. The research also implies that the "Smart contracts" features of blockchain will help attain anonymous and autonomous consensus amongst different stakeholders of IoT ecosystem.

The research carried out by Ferrag et al. (2019) explores the use of blockchain technology in various fields and proposes a solution including the threat model developed on blockchain protocols and fundamentals. The threat model built classifies different categories of threat ranging from manipulation attacks, reputation attacks, service attacks, crypto-analytic attacks to identity attacks. Additionally, the paper also proposed a novel taxonomy for privacy-preserving IoT ecosystem keeping in consideration the blockchain threat model, performance of IoT system, communication, and computational complexity of the ecosystem.

2.2.3 Application of Blockchain to Machine Learning

Blockchain technology has the ability to deal with millions of data in a decentralized ecosystem with a promise of accountability for every data. Machine Learning, which is used to carry out analysis of the data, requires reliable data for improved accuracy of the results. And when Blockchain technology which is characterized to induce data reliability is used in combination with Machine Learning, produces highly precise outcomes. The research carried out by Vyas et al. (2019) accentuate the data reliability, data security, and data privacy features of Blockchain technology and in turn, proposes the solution by applying these properties of blockchain technology to Machine Learning to carry out accurate decision making in the field of healthcare.

The success of a Machine Learning model depends on the reliable, dependable, and trustworthy source of data. Blockchain is a platform that provides accountability to the data stored in it. It is a platform where data stored in a cryptographically backed validated ecosystem. Blockchain stored data are characterized by high integrity, transparency, and resiliency. The research carried out by Salah et al. (2019) emphasizes that the combination of Machine Learning and Blockchain helps establish a secure, decentralized data store for Machine Learning driven system. The smart contracts features of blockchain can be used with the algorithms of machine learning to carry out smart analytics and decision making that are trustworthy and undisputed.

In a different scenario, Machine Learning can be used to overcome issues of blockchain such as double spending and majority attacks. The research carried out by Tanwar et al. (2020) gives an insight into a detailed review of the combination of these two technologies applied to build resilient applications. The research explores various Machine Learning techniques such as Support Vector Machines(SVM), Convolutional Neural Network(CNN), and Long Short Term Memory(LSTM) to analyze various attacks on blockchain-based applications.

2.2.4 Application of Blockchain to Internet of Things, and Machine Learning

Rathore et al. (2019) presented the concept of BlockDeepNet which is a Blockchain-based Deep Learning solution to support secure alliance with IoT ecosystem. BlockDeepnet solution helps overcome the issues of privacy leaks and security in the IoT ecosystem and further enables the feed of secure data to Deep Learning methods to carry out analytics and decision making. Blockchain is used to establish confidentiality and integrity in the solution. The solution is concluded to showcase high accuracy and low latency.

The research carried out by Lee et al. (2020) presented with an abstract of research efforts that have been carried out to explore security and privacy issues in the application of the combination of Machine learning algorithms and Blockchain technology in the IoT ecosystem. The research sheds the light on blockchain technology being a key factor in accomplishing a privacy-preserving data analytics and decision-making solution.

As discussed in the review of researches in the above subsections, blockchain applications help realize promising solutions to overcome the problems in the field of Machine Learning and IoT. As summarized in the research by Singh et al. (2020), the convergence of the fields of Machine Learning, IoT, and Blockchain technologies is disrupting the smart city solutions and is helping design a network architecture for sustainable and intelligent smart city and smart society.

2.2.5 Application of Internet of Things, Blockchain and Machine Learning in Scope of Waste Management

Shyam et al. (2017) presents a waste management solution that supplies intelligence to waste bins using IoT smart sensors. The data obtained from these sensors are used by machine learning algorithms to carry out the analytics and trend extraction. The analysis carried out is used to drive decision making. The solution by Fonseca et al. (2019) uses the machine-learning method of Convolutional Neural Network to classify fill levels in waste bins and to carry out real-time waste monitoring by processing images of the waste bins.

Efficient Recycling is also a vital part of the waste management system, and an approach presented by Gupta et al. (2019) uses the concept of Machine Learning and IoT to refine the process of waste recycling by eradicating human intervention in the whole process. Intelligent automated systems, such as electromagnetic sensors for metal sorting, sound wave detection to sort plastic and glass, etc are used to segregate the collected wastes.

The research carried out by Gopalakrishnan and Radhakrishnan (2019) focuses on the need for incentivizing people's participation in waste management processes such as segregation of waste, recycling, and disposal, additional to the need for a supportive technology that could automate the monitoring of important events in waste management. The blockchain technology that provides immutability to data through cryptographic methods and incentivization in real-time through smart contracts and cryptocurrency are used in combination with Artificial Intelligence and Machine Learning to provide solutions for effective waste management process.

In recent times, various new approaches have been proposed to optimize the system of waste management. The new approaches include generic algorithms, colony optimization, deep neural networks, etc. In the paper by Khoa et al. (2020), a novel approach to optimize the waste collection process is proposed. The solution uses low-cost IoT devices in combination with Machine Learning techniques and graph theory to optimize the system of waste collection. The solution implemented showcased reduced time and route optimization.

2.3 Literature Based on Forecasting Techniques of Machine Learning

Forecasting includes analyzing the trends and patterns in the historical data and using the observed trends and patterns to forecast/predict future values. The field of Machine Learning comprises of various algorithms and models to carry out time series forecasting to predict/forecast the future values. R. and T. (2004) presents a novel approach for defining weights for the combination of forecasting models. The features of the time series are used by the machine learning techniques to define the weights. To evaluate the proposed approach, Multilayer Perceptron(MLP), a class of Feed Forward Artificial Neural Network is used to combine the two widespread methods of forecasting, Random Walk and Autoregressive model. And the evaluation of the prototype revealed remarkably accurate results.

Ahmed et al. (2010) carried out a comparative study of time-series forecasting models keeping in consideration the new advances in the same scope. The study includes the comparative study of eight different statistical time series forecasting methods and eight machine learning forecasting models. Multilayer Perceptron(MLP), Bayesian Neural Network(BNN), k-Nearest Neighbor, Generalized Regression Neural Network(GRNN) are some of the few different machine learning forecasting models that were subjected to pre-processing and evaluated for their performances. The results of the evaluation showed MLP to be a better-performing model in terms of its accuracy compared to all other statistical and machine-learning forecasting models. The study also concluded the higher performance of machine learning models than the statistical models for forecasting.

In another experiment carried out by Kandananond (2012), comparison study of machine learning forecasting models, Multilayer Perceptron(MLP) and Support Vector regression(SVR) with traditional model Auto Regressive Integrated Moving Average(ARIMA) is carried out. The experiment was carried out in the scope of forecasting consumer demands of various products of an organization. The evaluation of the models were carried out on the values of evaluation metric Mean Absolute Percentage Error(MAPE) value. The results revealed Multi-Layer Perceptron(MLP) model to be better performing in case of both auto-correlated and non-auto-correlated data.

The experiment carried out by Safi (2013), the forecasting models ARIMA model, and

Multi-Layer Perceptron(MLP) are compared on their performance to forecast monthly electricity consumption in a smart city setup. The outcome of the experiment revealed that the performance of the different forecasting models differed with the quantity of hidden layers (in the case of deep learning) and learning rates. The result of the comparison also concludes that the forecasting model MLP outperforms other traditional method by providing better consistency.

An experiment has been carried out by Kochak and Sharma (2014) used-Multi Layer Perceptron(MLP) in combination with various back-propagation algorithms to carry out the forecast of future fuel sales. The evaluation carried out with evaluation metrics, Mean Average Error(MAE) and Root Mean Squared Error(RMSE), showed backpropagation models to perform better than any other traditional models. In yet another experiment by Massaro et al. (2018), comparative study of Deep Learning models and traditional supervised learning models such as k-Nearest Neighbour(kNN), Decision trees, Support Vector Machine(SVM), etc is carried out to forecast future sales of Walmart products. The deep learning models outperform the traditional models with an accuracy of 70% in forecasting future sales.

With the latest developments in the field of computation, advanced forecasting techniques are been developed. Long Short Term Memory(LSTM), a deep Learning model for forecasting is one such advanced technique. Lots of experiments and studies have been carried out to forecast future results using LSTM model. Siami Namini et al. (2018) carried out an empirical study to get insights into the performance of LSTM model as opposed to traditional models such as ARIMA. The results of the study showed LSTM outperforms the other model and revealed commendable reduction in the error rate (by 80-85%). The study also revealed the absence of any effect between the training times and performance of LSTM. In another study carried out by Alhirmizy and Qader (2019), LSTM is used to carry out the forecasting of Multivariate time series using TensorFlow library and Keras. LSTM model with an accuracy of 85% outperformed ARIMA, AR, MA, and other models in forecasting future values.

Other time series forecasting method which has been observed to have better performance are Convolutional Neural Network(CNN) and Recurrent Neural Network(RNN). Keeping this in consideration Zhang and Dong (2020) carried out an experiment to forecast future temperature values using Convolutional Neural Network(CNN), and Recurrent Neural Network(RNN). The experiment also consists of a novel hybrid model for forecasting built using the combination of CNN and RNN known as CRNN model. CRNN model is a multilevel neural network that processes both spatial correlation and time correlation in the time-series data. The evaluation of the models concluded CRNN to have a higher performance with a MAE value of 0.907 and RMSE value of 1.697.

2.4 Summary and Conclusion

In Table 1 summaries of few of the forecasting techniques reviewed are listed. And based on the list, keeping the findings in consideration, the following models are chosen to carry out forecasting in this research project.

- ARIMA
- Multi Layer Perceptron
- Long Short-Term Memory Networks(LSTMs)

- Convolutional Neural Network(CNNs)

Table 1: Summary Table for Literature Based on Forecasting Techniques of Machine Learning

Model	Reference	Title	Findings
MLP	R. and T. (2004)	Using Machine Learning Techniques to Combine Forecasting Methods	Multilayer Perceptron(MLP), used to combine the two widespread methods of forecasting, RandomWalk and Autoregressive model produces a remarkably accurate results.
MLP	Ahmed et al. (2010)	An Empirical Comparison of Machine Learning Models for Time Series Forecasting	Comparitive study of eight different time series forecating techniques revealed MLP to be a better performing model i terms of accuracy.
ARIMA, MLP	Safi (2013)	Artificial neural networks approach to time series forecasting for electricity consumption in Gaza Strip	ARIMA model and MLP model are evaluated using evaluation metrics MAE and RMSE on their performance to forecast future fuel sales. And the results showed that the backpropagation models perform better than any other traditional models.
ARIMA, LSTM	Siami Namini et al. (2018)	A Comparison of ARIMA and LSTM in Forecasting Time Series	An empirical study to get insights into the performance of LSTM model with traditional model was carried out and the results showed that LSTM outperforms the other model with commendable reduction in error rate
CNN	Zhang and Dong (2020)	Temperature Forecasting via Convolutional Recurrent Neural Networks Based on Time-Series Data	CNN and RNN are tested for their performance to forecast time series data and the results showed a lower rate of RMSE and MAE for CNN model indicating its better performance
LSTM	Alhirmizy and Qader (2019)	Multivariate Time Series Forecasting with LSTM	LSTM is used to forecast Multivariate time series using TensorFlow library and Keras and with an accuracy of 85%, it outperformed ARIMA, AR, MA, and other models.

3 Methodology

The methodology CRISP-DM is used in this research project. CRISP-DM, which stands for Cross Industry Process for Data Mining, provides a robust methodology and structured approach to carry out any Machine Learning Project. Independence of the methodology from the technology used and industry sector and its feature of flexibility was the main motivation behind using CRISP-DM for this research project(Figure 2). As part of the CRISP-DM methodology, the research project is divided into a sequence of events that can be repeated and backtracked when needed. These events are as follows:

- Research Understanding
- Data Extraction & Understanding
- Data Preparation and Processing
- Data Modeling
- Model Evaluation and Results
- Research Findings

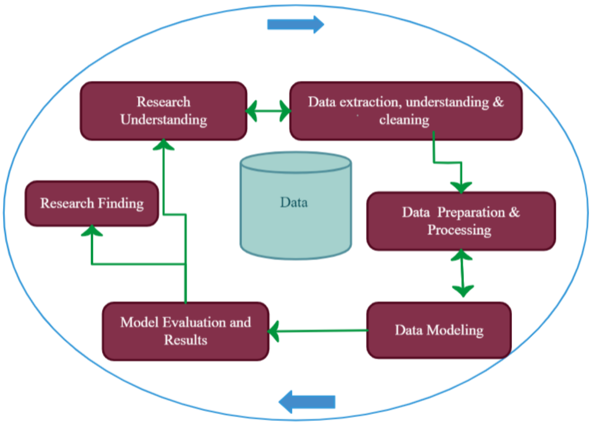


Figure 2: CRISM-DM Methodology for the Research Project

3.1 Research Understanding

The first event, part of the research methodology is to understand the business and implementational aspect of the research project.

Waste management in a developing or developed urban city is a very complex challenge, which is constrained by its system of high labour-intensive and resources. An average $300 per capita is spent every month for a functional waste management system. This expense is a huge financial burden, even though it is equally shared between the citizens of an urban city and its governing entity. And there are only few cases of innovations and technical solutions as a step towards optimizing and improving the ecosystem of waste management. Route optimization is one such innovation to improve the efficiency of routes traversed for waste collection by a collection truck. Even with the introduction of innovations and solutions in the scope of waste management, the system still remains inefficient in its working.

With the introduction of smart waste management, an initiation of smart city garbage bins across the smart cities are equipped with IoT sensors, which bestow information on the garbage fill levels, the quantity of waste collected, the location of garbage bins, threshold for waste collection, etc. In this research project, a novel solution is proposed using this IoT emitted data from the smart bin in combination with Blockchain and

Machine Learning to optimize the waste management process. This solution aims at making the process of waste management less labor intensive and also aims at reducing operational costs of the whole process. Machine Learning forecasting techniques are used to forecast the waste fill levels of the smart bins, which helps plan an efficient collection system of the waste wherein efficient pre-planned schedules for collection with route optimization for the collection trucks can be designed using the forecasted values. This helps in a considerable reduction in the use of resources and its operational costs (e.g. fuel costs of trucks). This also creates a huge impact on the environment, wherein the cases of overflowing garbage bins can be avoided and thereby keep the environment pollution and disease free. And Blockchain in the same context helps establish a decentralized ecosystem of trust for the IoT data collected and additionally induces the features of data reliability, security, and integrity to the IoT data that are used in Machine Learning models for forecasting, and hence making the results of forecasting reliable and trustworthy. Hence a solution for an ecosystem built with the combination of IoT, Blockchain, and Machine Learning to optimize waste management is the focus of this research paper.

3.2 Data Extraction & Understanding

Data for the implementing the proposal is collected from the reliable source by, Nikolaos (2019), which is a real time monitoring data of garbage bins set up as part of Composition EU Project. The data contains the fill levels of the garbage bin and is open-sourced with consent of KLEEMANN Hellas SA and ELDIA SA. The dataset has various attributes such as Fillpercentage, Battery Level, eventDate. The dataset is composed of hourly observations of garbage fill levels between the dates 20/06/2019 to 08/08/2019.

3.3 Data Preprocessing/Preparation

Data preparation and processing is an important part of any machine learning research project wherein data to be modeled are checked for impurities such as missing values, the degree of correlation, etc. In this research project, the data which is a IoT sensor collected data on waste fill levels is extracted as a JSON file from its source. Keeping the proposed solution, the scope of the project and the academic requirement of the research project module in consideration, the data to be modeled is considered to be Blockchain validated and equipped with properties of trust, security, and integrity.

The JSON data extracted is converted into a CSV file using Python programming in a Google Colab environment. The CSV generated is converted into a data frame using Python Pandas and is pre-processed and checked for any missing/null values and outliers using the same. The data is converted to a time series using pandas library and Augmented Dickey-Fuller(ADF) test is carried out to check for stationarity of the time series data.

3.4 Data Modeling

Multiple machine learning forecasting techniques are applied on the cleaned pre-processed data to forecast waste fill levels of the smart bin. The models used in this research project are:

- **ARIMA:**

ARIMA which stands for "AutoRegressive Integrated Moving Average" is a traditional model for time-series forecasting formed by the combination of three models namely Autoregressive(AR), Integrated(I) and Moving Average (MA), represented using terms, p, d, and q. ARIMA(p, d, q), where, the term "p" represents the order of Autoregresssive part, "q" represents the order of Moving Average term and "d" represents the stationarity of the time series data, ARIMA is widely used for its properties of capturing complex relationships of time taking into consideration error terms and lagged observations, and forecasting values based on variable regression of historic values.

- **LSTM:**

  LSTM also known as Long Short Term Memory is a class of Recurrent Neural Network(RNN) in deep learning. It is characterized by its ability to learn the longterm dependencies and thus makes it a better choice than RNN to remember information for an extended amount. LSTM model are trained using backpropagation and has an ability to overcome vanishing gradient problem. This makes the model better resistant to errors and a better choice to train larger volumes.

  LSTM is a network that is comprised of multiple layers that include an input and output layer and corresponding gates to control the flow of data in and out of these layers. Additional to the input layers it contains multiple hidden layers, pointing to the higher efficiency of learning. LSTM has the capacity to learn dependency order between the variables of a time-series data and it learns the context for predictions and forecasts. These features of LSTM makes it a choice of the model to carry out the forecast in this research project.

- **Multi Layer Perceptron(MLP)**

  Multi-Layer Perceptron commonly known as MLP is a class of feed forward artificial neural network. It is comprised of minimum three layers, an input layer, an output layer, and the hidden layer/'s. Except the input layer, all the other layers of MLP are subjected to non-linear activation function. MLP layers are trained using the supervised learning method of back propagation. The learning method, the activation function used(non-linear), and multi-layer architecture make MLP different from linear perceptron. Its property to forecast values of the univariate time series data with temporal ordered observations makes it an efficient forecasting algorithm to be implemented in this research project.

- **Convolutional Neural Network(CNN)**

  Convolutional Neural Networks commonly called as CNN's are a class of deep neural networks. It is a more regularized version of Multi-Layer Perceptron. CNNs consists of an input layer, an output layer, and multiple hidden layers. The hidden layers of CNNs are typically convolutional layers that are a product of matrix multiplication of their inputs. ReLu is the most commonly used activation function for the CNNs. The dilated convolutional layers makes CNNs effectively learn time-series data. And its property, to learn the recurring patterns and trends in the time series makes it another prime choice to be modeled in this research project.

3.5 Model Evaluation

The models implemented are measured for their performance using various evaluation metrics,

- **Mean Error**

  Mean Error commonly regarded as ME (Equation 1) is the average of all individual errors, which is nothing but the difference between the "Actual" and "Forecasted" values. Mean Error measures the error type, forecast bias. And the ME value will be positive for consistently low forecasts and negative for consistently high forecast.

$$\mathrm{ME} = \frac{\sum_{i=1}^{n} y_i - x_i}{n}. \tag{1}$$

- **Mean Absolute Error**

  Mean absolute error also known as MAE(Equation 2) is the absolute measure of the average deviation of the forecasted values. It is very beneficial in measuring forecast errors and gives insights into how huge the forecast error is. It is robust to outliers and is useful for the measuring the performance of time-series data corrupted with outliers. It considered as a measure of accuracy.

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} = \frac{\sum_{i=1}^{n} |e_i|}{n}. \tag{2}$$

- **Root Mean Squared Error**

  Root Mean Squared Error abbreviated as RMSE (Equation 3) is the measure of the average deviation of the forecasted values from the original/actual observations. It gives insights into how far the forecasts have deviated from the actual observations. RMSE can be considered as a measure of accuracy, penalizing the bad forecasts points.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{t=1}^{T} (\hat{y}_t - y_t)^2}{T}}. \tag{3}$$

- **Mean Percentage Error**

  Mean Percentage Error or MPE is the measure of average of percentage errors, forecasted values differ from the actual values. Similar to Mean Error, it is also used to measure the bias in the forecasts. It is measured using the formula,

$$\mathrm{MPE} = \frac{100\%}{n} \sum_{t=1}^{n} \frac{a_t - f_t}{a_t} \tag{4}$$

- **Mean Absolute Percentage Error**

  Mean Absolute Percentage Error(MAPE) also regarded as Mean Absolute Percentage Deviation(MAPD) is a of measure the average of the absolute percent error of the forecast. It gives insights into the relative errors and is considered as a measure of forecast accuracy. Formula to calculate MAPE is,

$$\mathrm{M} = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|, \tag{5}$$

4 Implementation

In this research project, a novel approach to optimize waste management system using the combination of IoT, Blockchain, and Machine Learning forecasting techniques is proposed. The research design (proposal) is divided into two parts, the first being the overall solution for smart waste management proposed and keeping the academic requirements and scope of the subject module in mind, the second part of the research design focuses on the implementation of machine learning part of the solution proposal, which includes implementing machine learning techniques to forecast waste fill levels of a garbage bin in a smart city ecosystem.

4.1 Smart Waste Management Solution Proposal

Like mentioned before, in this research paper, a novel solution involving the combination of smart IoT, Blockchain, and Machine Learning, to optimize waste management process is proposed (Figure 3). The solution considers the eco-system of a smart city, wherein garbage bins across the city are equipped with IoT sensors that gives real-time information on their waste fill levels. The real-time data collected from these sensors are fed into a decentralized Blockchain network, to check for their integrity and induce the property of privacy, security, reliability, and trust to the IoT data collected from different stakeholders. Machine Learning forecasting techniques are applied onto the Blockchain processed data, and the use of reliable data makes the model performs better, and results more promising. Different time series forecasting techniques are applied to forecast the waste fill levels, and future fill levels of smart garbage bins across the city and the models implemented are evaluated using various evaluation metrics such as RMSE, MAPE, MAE, etc to determine the better-performing model in forecasting waste fill levels. These forecasted values can be used to design an efficient waste collection scheduling system, keeping in consideration, route optimization and resource allocation.

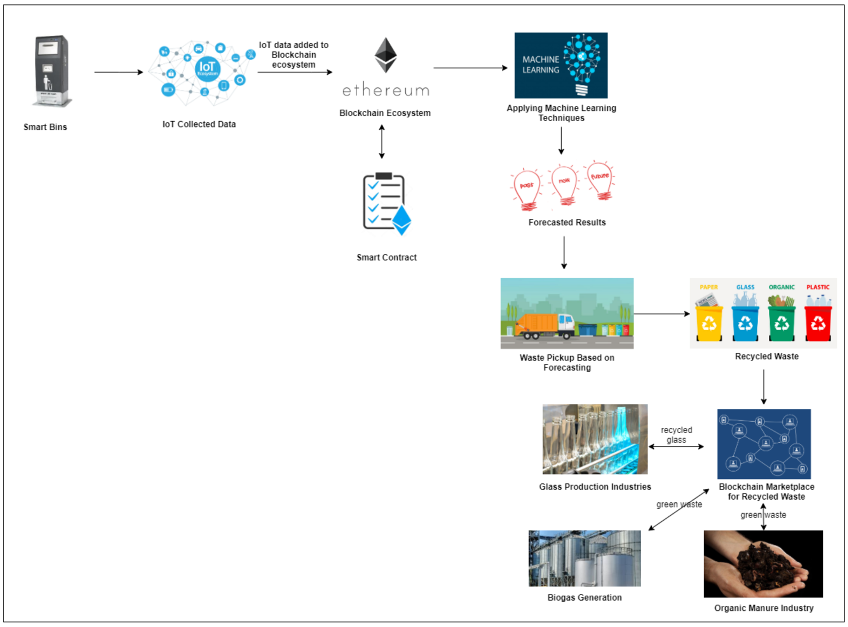


Figure 3: Smart Waste Management Solution Proposal

4.2 Implementation Design for Waste Fill Levels Forecasting

In this research project we will implement the machine learning part of the smart waste management solution proposed in the previous subsection (4.1). Different forecasting techniques are implemented to forecast the waste fill levels of waste at smart garbage bin in a city. The implementation flow for the following is as seen in Figure 4.

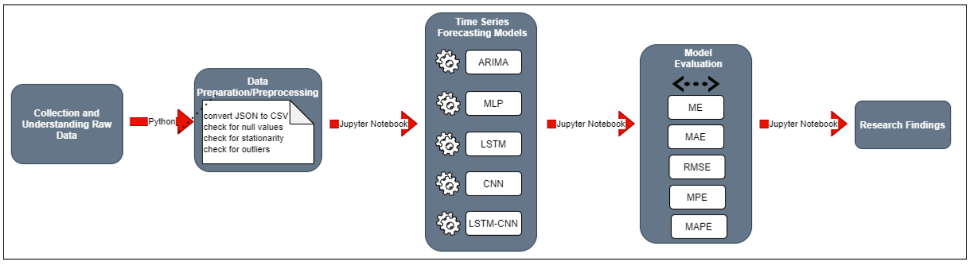


Figure 4: Implementation Process Flow

The design process follows the CRISP-DM Methodology. In the first step of implementation, the IoT sensor data on waste fill levels of the garbage bin is extracted from the reliable source. The extracted data is in JSON format, and it is converted to CSV format in python programming environment using pandas and python libraries.

In the next step, the data is explored to get deeper insights into the trends and patterns of the variables in the dataset. Exploratory Data Analysis is carried out to get more insights into the trends and patterns in the dataset. A heatmap (Figure 5) is plotted to get insights into the degree of correlation between the attributes of the dataset. The dataset was also explored to analyze the trend of waste fill levels (Figure 6), and Battery levels of IoT sensors by date and plots were plotted for the same respectively. Additionally as part of data exploring, plots for average waste fill levels (Figure 7) and battery percentage by time were also plotted.

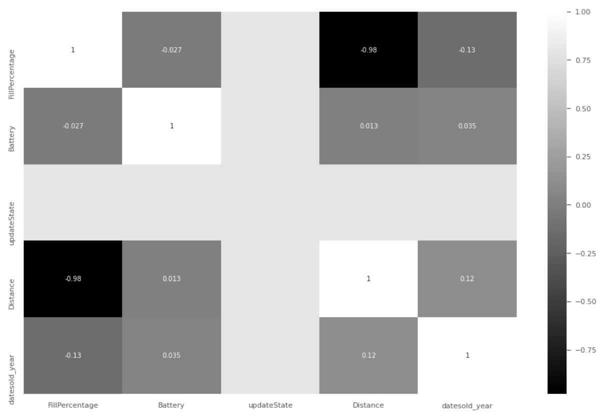


Figure 5: Heatmap of data attributes

In the third step, the data are pre-processed and prepared to be modeled. As part of the data pre-processing, for the traditional model ARIMA, the time-series data are checked for stationarity using ADF-Fuller test, estimated for the trend and ACF and PACF functions are plotted. The figure 8 shows the plot of PACF function for the time-series data that are checked for stationarity and seasonal trend. And for the other neural network models, LSTM, CNN, MLP, CNN-LSTM, as part of the data pre-processing and

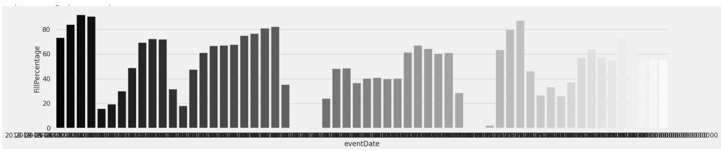


Figure 6: Trend of Waste Fill Levels by Date

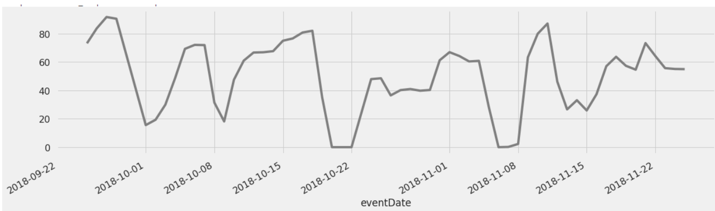


Figure 7: Average Waste Fill Levels by Date

preparation step, the data is split into train and test set. The data is split into 80% train set and 20% test set. Additionally, the data which is still inform of a sequence are converted to supervised data to be trained in a neural network. Neural Network needs the data to be of the form ”features”, ”target”, and hence conversion of the data into this format is necessary and for this we use the concept of **Lookback**. In our implementation we use Keras module **TimeseriesGenerator** to carry out this step as shown in Listing 1. In the experiments, the lookback value is set at 15.

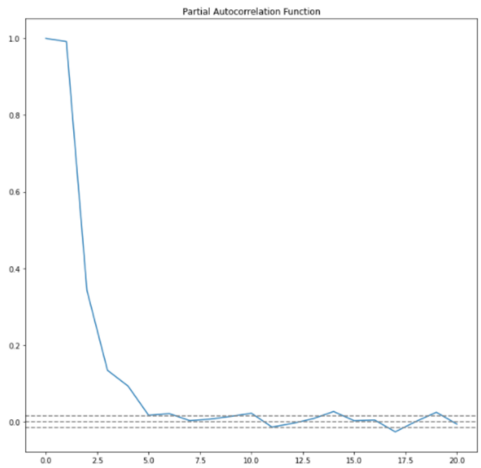


Figure 8: ARIMA PACF plot for the data

Listing 1: Preprocessing for Neural Network

```
look_back = 15

train_generator = TimeseriesGenerator(fill_train, fill_train,
    length=look_back, batch_size=20)
test_generator = TimeseriesGenerator(fill_test, fill_test, length=look_back,
    batch_size=1)
```

In the fourth step, different machine learning forecasting models are implemented, to forecast waste fill levels at IoT sensor fit garbage bin. Python programming, and their packages is used to carry out the implementation. Jupyter Notebook in Google Colab is used as the programming environment.

- **ARIMA Model**

  The data in the form of a CSV file is loaded into Jupyter Notebook, and is accessed as a data frame, which is then indexed by the column eventDate. The python packages numpy, pandas and matplotlib are used for numerical computations, pre-processing and visualizations respectively. Rolling Statistics for the variable "FillPercenatge", which is the main focus of our experiment, is computed and plotted (visualized). The stationarity of the time series is checked, carrying out the Augmented Dickey-Fuller(ADF) test. The ADF test resulted in a value of 0.003, less than 0.05, hence the null hypothesis can be rejected and the time series can be considered stationary. Once the time series are tested and confirmed for their stationarity, ACF(Autocorrelation) and PACF(Partial Autocorrelation) function plots are plotted to get the values of "p" from the PACF plot and "q" from the ACF plot. ARIMA model for order p, d, q, ARIMA(p, d, q) are trained/fitted using **.fit()** function from "statsmodel" package. Fitted values are plotted using **.fittedvalues** function and checked for model fit. **.forecast** function is applied on the fitted values, to forecast values for the specified time steps and these forecasted values are converted to actual numbers using log function. Additionally **pmdarima** package (module) is used to auto plot and train ARIMA model, computing the optimum value of p, d, q. And lastly predicted/forecasted values are plotted using **plot_predict** function.

- **LSTM**

  The preprocessed data that was split into train and test set are trained using Keras **LSTM** package. A simple LSTM architecture with single LSTM layer is used to train the data, with one Dense Layer. The model is trained using Adam Optimizer and Mean Squared Error Loss Function for 500 epochs, and verbose value of 1. The LSTM layer uses **reLu** activation function. **model.fit_generator** function is used to fit the model because, data generator(TimeseriesGenerator) is used to preprocess the data in the format required in the neural network. Predictions from the model are obtained using **model.predict_generator** function using the test data and goodness of fit of our model is validated. Furthermore, future forecasts are carried out by feeding into the model the past n(look_back) days. Plots for forecasts and future forecasts are plotted using Python's **plotly** package.

- **CNN**

  The model is trained using keras.layers package, **Conv1D**. Convolutional Neural Layer consists of the convolutional layer, with a kernel size of 2, and **reLu** activation function. The model is trained using Adam Optimizer and Mean Squared Error Loss Function for 500 epochs, and verbose value of 1. **model.fit_generator** function is used to fit the model and the predictions from the model are obtained using **model.predict_generator** function. Future forecasts are carried out by feeding into the model the past n(look_back) days and plots for forecasts and future forecasts are plotted using Python's **plotly** package.

- **MLP**

A simple MLP architecture with single layer and **reLu** activation function is used to train the data. The model is trained using Adam Optimizer and Mean Squared Error Loss Function for 500 epochs, and verbose value of 1. Similar to LSTM and CNN model, **model.fit_generator** function is used to fit the model and the predictions from the model are obtained using **model.predict_generator** function. And similarly future forecasts are carried and forecasts are plotted.

- **CNN-LSTM**

  A hybrid model combining the layers of CNN and LSTM is implemented. The model has a CNN layer, followed by a LSTM layer, both activated using the **reLu** activation function. And similar to the other neural networks mentioned above, the model is trained using Adam Optimizer and Mean Squared Error Loss Function for 500 epochs, and verbose value of 1, and **model.fit_generator** function is used to fit the model and the predictions from the model are obtained using **model.predict_generator** function.

After the models are trained, the models are evaluated for their performance using the evaluation metrics, ME, MAE, RMSE, MPE, and MAPE, discussed in detail in the section 5

5 Evaluation

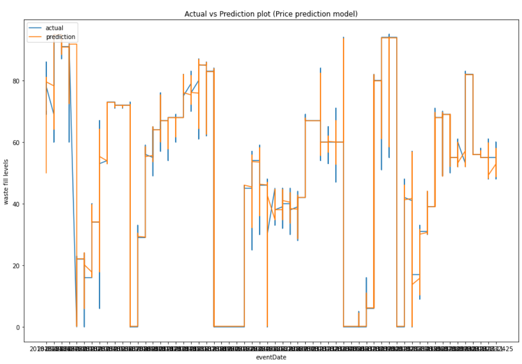
The models implemented are evaluated using evaluation metrics, Mean Error(ME), Mean Absolute Error(MAE), Root Mean Squared Error(RMSE), Mean Percentage Error, and Mean Absolute Percentage Error(MAPE).

5.1 ARIMA Model

ARIMA model is evaluated using the above-mentioned evaluation metrics. The evaluation is carried out by comparing the **ARIMA_predicted** values with the original data. The evaluation resulted in ME = -0.00304, MAE = 1.165 , RMSE = 3.351 , MPE = -0.1235 , MAPE = 0.20344 as shown in Figure 9a and the plot for forecasted and actual values are plotted as in Figure 9b.

```
{'mae': 1.1658169857453924,
 'mape': 0.2034472718930649,
 'me': -0.0030441185166198603,
 'mpe': -0.12354420515697011,
 'rmse': 3.351773455717107}
```

(a) ARIMA Evaluation metrics



(b) ARIMA: Plot for Forecasted vs Actual Values

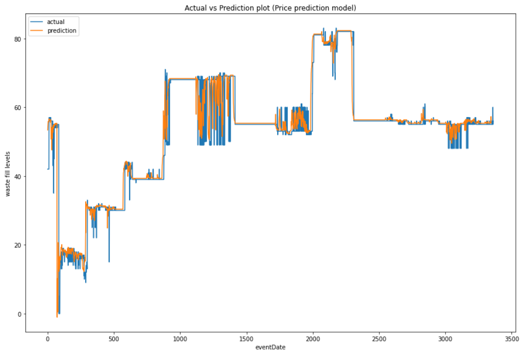
Figure 9: ARIMA Model

5.2 LSTM Model

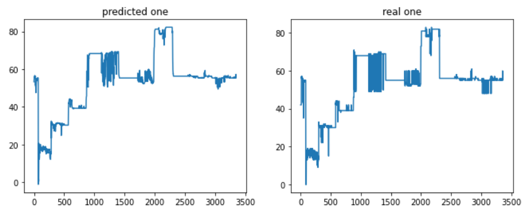
LSTM model is evaluated using the above-mentioned evaluation metrics. The evaluation is carried out by using the forecasted values against test data values for fill levels. The evaluation resulted in ME = -0.3364 , MAE = 1.2925 , RMSE = 3.056 , MPE = -0.0014 , MAPE = 0.0374 as shown in Figure 10a and the plot for forecasted and actual values are plotted as in Figure 10b and Figure 10c.


```
{'corr': 0.9817121733561912,
 'mae': 1.2925003715621022,
 'mape': 0.03742320960830484,
 'me': -0.336418977885708,
 'mpe': -0.0014298838946333907,
 'rmse': 3.056053678508206}
```


(a) LSTM Evaluation metrics



(b) LSTM: Plot for Forecasted vs Actual Values



(c) LSTM: Plot for Forecasted vs Actual Values

Figure 10: LSTM Model

5.3 CNN Model

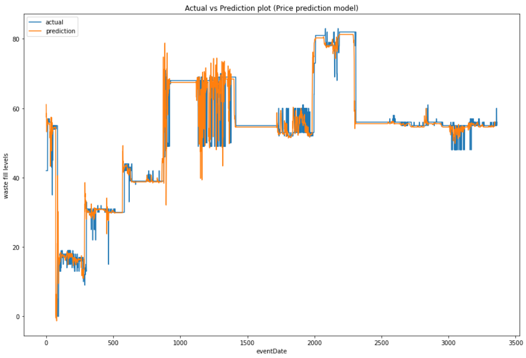
CNN model is evaluated using the above-mentioned evaluation metrics. The evaluation is carried out by using the forecasted values against test data values for fill levels. The evaluation resulted in ME = 0.1610 , MAE = 1.5012 , RMSE = 3.497 , MPE = -0.0021 , MAPE = 0.0414 as shown in Figure 11a and the plot for forecasted and actual values are plotted as in Figure 11b and Figure 11c.

5.4 MLP Model

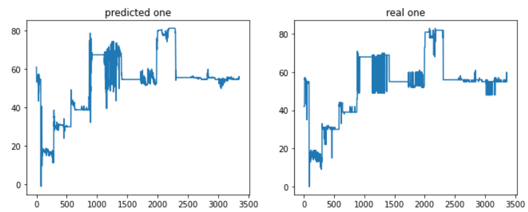
MLP model is evaluated using the above-mentioned evaluation metrics. The evaluation is carried out by using the forecasted values against test data values for fill levels. The evaluation resulted in ME = 0.1137 , MAE = 1.7178 , RMSE = 4.1068 , MPE = 0.0015 , MAPE = 0.0369 as shown in Figure 12a and the plot for forecasted and actual values are plotted as in Figure 12b and Figure 12c.

```
{'corr': 0.9757586396735556,
 'mae': 1.5012710167844354,
 'mape': 0.04141112426758809,
 'me': 0.16107183399132036,
 'mpe': -0.0021401447652247122,
 'rmse': 3.4973778196924044}
```

(a) CNN Evaluation metrics



(b) CNN: Plot for Forecasted vs Actual Values

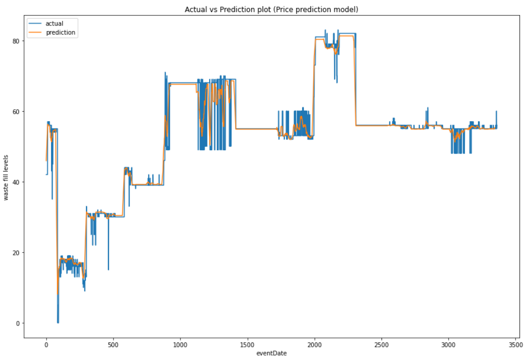


(c) CNN: Plot for Forecasted vs Actual Values

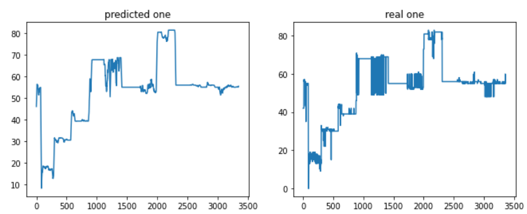
Figure 11: CNN Model

```
{'mae': 1.7178297048547064,
 'mape': 0.03691107446936132,
 'me': 0.1137253156480715,
 'mpe': 0.0015841864346290472,
 'rmse': 4.106868712002064}
```

(a) MLP Evaluation metrics



(b) MLP: Plot for Forecasted vs Actual Values



(c) MLP: Plot for Forecasted vs Actual Values

Figure 12: MLP Model

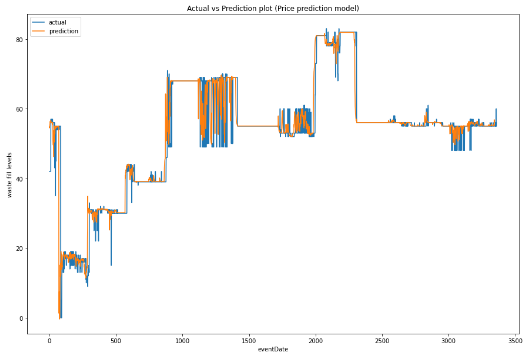
5.5 CNN-LSTM Model

CNN-LSTM model is evaluated using the above-mentioned evaluation metrics. The evaluation is carried out by using the forecasted values against test data values for fill levels.

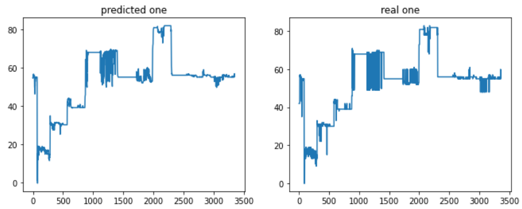
The evaluation resulted in ME = -0.1280 , MAE = 1.1259 , RMSE = 3.0633 , MPE = 0.0364 , MAPE = 0.0663 as shown in Figure 12a and the plot for forecasted and actual values are plotted as in Figure 12b and Figure 12c.

```
{'corr': 0.9814380481986593,
 'mae': 1.1259478810513297,
 'mape': 0.0663914517847701,
 'me': -0.12808988408563526,
 'mpe': 0.036460133644956,
 'rmse': 3.063365611079236}
```

(a) CNN-LSTM Evaluation metrics



(b) CNN-LSTM: Plot for Forecasted vs Actual Values



(c) CNN-LSTM: Plot for Forecasted vs Actual Values

Figure 13: CNN-LSTM Model

5.6 Discussion

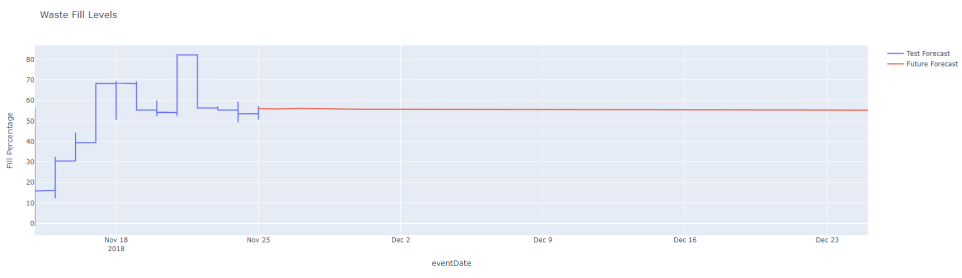
The Table 2 summarizes the evaluation metrics computed from all our models implemented. A positive Mean Error (ME), represents under forecast of the model and a negative value indicates over forecast, and a model is said to be good if the ME is near to zero. Mean Absolute Error(MAE) is the average of forecast errors and for a better-performing model MAE value is zero or near to zero. A Root Mean Squared Error(RMSE) value of zero indicates a 100% accurate forecasting model, and low values of RMSE(almost zero), indicates a better model. Similarly Mean Percentage Error(MPE) and Mean Absolute Percentage Error(MAPE) values closer to zero indicate better-performing model.

Table 2: Evaluation Metrics Table

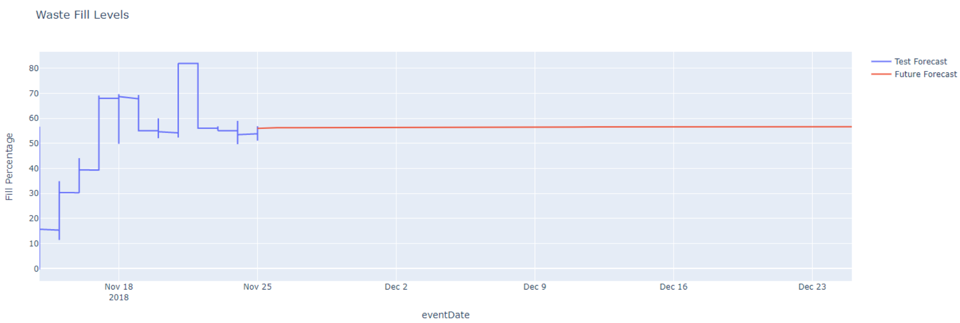
Evaluation Summary					
Metrics	ARIMA	LSTM	CNN	MLP	CNN-LSTM
ME	-0.00304	**-0.3364**	0.1610	0.1137	-0.1280
MAE	1.165	**1.2925**	1.5012	1.7178	1.1259
RMSE	3.351	**3.056**	3.497	4.1068	**3.0633**
MPE	-0.1235	**-0.0014**	-0.0021	0.0015	0.0364
MAPE	0.20344	**0.0374**	0.0414	0.0369	**0.0663**

Keeping the above points in consideration and referring the Table 2 , we can say that all the models, more or less have the evaluation metrics in the similar range, indicating good performances from the model. But the **LSTM** model with ME = 0.1610 , MAE = 1.5012 , RMSE = 3.497 , MPE = -0.0021 and MAPE = 0.0414 to be the model with the best performance. ARIMA model, which is the only traditional model implemented, has not fallen behind in its performance compared to the neural networks. Even though its MAPE value is 0.20344, highest of all the models, its RMSE(3.351), ME(-0.00304), and MAE(1.165) values are less than other neural network. The other model, which showed exceptional results was the hybrid model of **CNN-LSTM** with ME = -0.1280 , MAE = 1.1259 , RMSE = 3.0633 , MPE = 0.0364 and MAPE = 0.0663. Even though its MAPE value is greater than that of CNN and MLP, its RMSE value and MAE value are relatively very less. CNN model with RMSE value of 3.497 and MAPE value of 0.0414, can also be considered a good model, but in comparison with other models it falls short in its performance. MLP value was the least performing model in the experiment carried out with the highest values of RMSE(4.1068) and MAE(1.7178). Hence we can conclude than LSTM, followed by CNN-LSTM model were the best-performing models from the experiment carried out.

The models trained are used to forecast fill levels of waste in garbage bins for the future dates. The forecasts of the better-performing models identified are considered for this purpose and these forecasts can be used to build an efficient waste collection system, wherein we have knowledge of what the fill levels of a garbage bin will be for the future dates and can schedule its collection accordingly. The forecasts of the better-performing models, which in our case are, **LSTM** and **CNN-LSTM** models can be seen in Figure 14.



(a) LSTM Forecasts



(b) CNN-LSTM Forecasts

Figure 14: Model Forecasts

From the future forecasts observed, we can see that there is not much variations in the level of waste forecasted for the next 20 days. This can be due to the insufficient training data fed to the model. IoT data for smart waste management is sensitive and limited for the public use. In case of accessibility to the real-time data from these smart bins, collected over a long period of time, the models can be trained on the huge bulk of data and the forecasts made, will be more reliable and precise.

The forecasted values from the smart bins can be used to analyse the future trends of waste collection across a city or a town and thereby these predicted values can be used to design an efficient waste collection system with route optimization.

6 Conclusion and Future Work

With growing urbanization, the need for an efficient waste management system is growing every day. With the introduction of technologies such as IoT, Blockchain, and Machine Learning, there have been new solutions built to refine the process of waste management. In this research paper, a novel solution for optimizing and efficiently carrying out waste management, combining the technologies of IoT-Blockchain and Machine Learning, is proposed. The solution involves the IoT extracted data from IoT sensor fit smart bins, measuring their fill percentage being processed in a decentralized eco-system of Blockchain to induce the properties of trust, reliability, and security, which the IoT data lacks. And finally machine learning techniques are applied to the Blockchain processed data to forecast future fill levels of waste at these smart bins to build an efficient waste collection system with route optimization, which is both cost and resource effective. Different machine learning forecasting techniques including both traditional (ARIMA) and Deep Learning (LSTM, CNN, MLP) models are implemented to forecast future fill percentage of a smart bin. The models implemented shows promising results. In the experiment carried out, deep learning model LSTM, and the hybrid CNN-LSTM model was identified the better performing model with RMSE value of 3.056 and 3.0633 respectively, and MAPE value of 0.0414 and 0.0663 respectively. The traditional model ARIMA, with RMSE = 3.351 and MAPE = 0.20344, was also found to be almost equal in its performance against deep learning models.

As part of the future work, a decentralized marketplace can be built to recycle the waste collected. The waste collected can be segregated and advertised for recycling on this marketplace, wherein stakeholders of various fields and sectors who are in need of these recyclable wastes can bid on them and purchase for use. For example, a manure-producing company can purchase biodegradable wastes to convert it into manure, or a glass production company can purchase the recycled glass to mold a new one. This solution helps efficient management of waste and additionally induce the business perspective to the whole process.

Acknowledgement

I would like to thank my research supervisor, Dr. Rashmi Gupta, for her thorough support and guidance through every stage of development of my research project. Her invaluable suggestions and feedback helped a great deal in shaping my research project. I would also like to thank my family and friends for their constant encouragement and support.

References

Ahmed, N., Atiya, A. F., Gayar, N. E. and El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting, *Econometric Reviews* pp. 594 – 621.

Al-Jabi, M. and Diab, M. (2017). Iot-enabled citizen attractive waste management system, *2017 2nd International Conference on the Applications of Information Technology in Developing Renewable Energy Processes Systems (IT-DREPS)*, pp. 1–5.

Alhirmizy, S. and Qader, B. (2019). Multivariate time series forecasting with lstm for madrid, spain pollution, *2019 International Conference on Computing and Information Science and Technology and Their Applications (ICCISTA)*, pp. 1–5.

Atlam, H., Alenezi, A., Alassafi, M. and Wills, G. (2018). Blockchain with internet of things: Benefits, challenges and future directions, *International Journal of Intelligent Systems and Applications* .

Banerjee, M., Lee, J. and Choo, K.-K. R. (2018). A blockchain future for internet of things security: a position paper, *Digital Communications and Networks* (3): 149 – 160.
URL: *http://www.sciencedirect.com/science/article/pii/S2352864817302900*

Contreras, F., Ishii, S., Aramaki, T., Hanaki, K. and Connors, S. (2010). Drivers in current and future municipal solid waste management systems: cases in yokohama and boston, *Waste Management & Research* (1): 76–93. PMID: 19942650.

Ferrag, M. A., Derdour, M., Mukherjee, M., Derhab, A., Maglaras, L. and Janicke, H. (2019). Blockchain technologies for the internet of things: Research issues and challenges, *IEEE Internet of Things Journal* (2): 2188–2204.

Ferrera de Lima, J. (2012). New model of municipal solid waste management, *International Journal of Environment and Sustainable Development* p. 238.

Fonseca, B. J., AliFelerminoD.M., A. and Saide, S. M. (2019). A deep convolutional neural network for classifying waste containers as full or not full, *2019 IEEE International Smart Cities Conference (ISC2)* pp. 54–59.

Gopalakrishnan, P. and Radhakrishnan, R. (2019). Blockchain based waste management.

Gupta, P. K., Shree, V., Hiremath, L., Gupta, P. and Rajendran, S. (2019). *The Use of Modern Technology in Smart Waste Management and Recycling: Artificial Intelligence and Machine Learning.*

Gutierrez, J. M., Jensen, M., Henius, M. and Riaz, T. (2015). Smart waste collection system based on location intelligence, *Procedia Computer Science* pp. 120 – 127. Complex Adaptive Systems San Jose, CA November 2-4, 2015.
URL: *http://www.sciencedirect.com/science/article/pii/S1877050915030008*

Joshi, R. and Ahmed, S. (2016). Status and challenges of municipal solid waste management in india: A review, *Cogent Environmental Science* (1): 1139434.

Kandananond, K. (2012). A comparison of various forecasting methods for autocorrelated time series, *International Journal of Engineering Business Management* .

Khatib, A., Arafat, H., Basheer, T., Shawahneh, H., Salahat, A., Eid, J. and Ali, W. (2007). Trends and problems of solid waste management in developing countries: A case study in seven palestinian districts, *Waste management (New York, N.Y.)* pp. 1910–9.

Khoa, T., Phuc, C., Lam, P., Nhu, L., Trong, N., Phuong, N., Dung, N., Tan-Y, N., Nguyen, H. and Duc, D. (2020). Waste management system using iot-based machine learning in university, *Wireless Communications and Mobile Computing* pp. 1–13.

Kochak, A. and Sharma, S. (2014). Demand forecasting using neural network for supply chain management.

Kumar, S., Smith, S. R., Fowler, G., Velis, C., Kumar, S. J., Arya, S., Rena, n., Kumar, R. and Cheeseman, C. (2017). Challenges and opportunities associated with waste management in india, *Royal Society Open Science* (3): 160764.
URL: *https://royalsocietypublishing.org/doi/abs/10.1098/rsos.160764*

Lee, Y., Rathore, S., Park, J. H. et al. (2020). A blockchain-based smart home gateway architecture for preventing data forgery, *Hum. Cent. Comput. Inf. Sci. 10* (9).

Massaro, A., Maritati, V. and Galiano, A. (2018). Data mining model performance of sales predictive algorithms based on rapidminer workflows, *International Journal of Computer Science and Information Technology* pp. 39–56.

Medvedev, A., Fedchenkov, P., Zaslavsky, A., Anagnostopoulos, T. and Khoruzhnikov, S. (2015). Waste management as an iot-enabled service in smart cities, *Lecture Notes in Computer Science Internet of Things, Smart Spaces, and Next Generation Networks and Systems* p. 104–115.

Mohee, R., Mauthoor, S., Bundhoo, Z. M., Somaroo, G., Soobhany, N. and Gunasee, S. (2015). Current status of solid waste management in small island developing states: A review, *Waste Management* pp. 539 – 549.
URL: *http://www.sciencedirect.com/science/article/pii/S0956053X15004262*

Narayana, T. (2009). Municipal solid waste management in india: From waste disposal to recovery of resources?, *Waste Management* (3): 1163 – 1166.
URL: *http://www.sciencedirect.com/science/article/pii/S0956053X08002626*

Nikolaos, A. (2019). Composition fill level sensors datasets, *Zenodo Dataset* .
URL: *https://doi.org/10.5281/zenodo.3375560*

Noche, B., Rhoma, F., Chinakupt, T. and Jawale, M. (2010). Optimization model for solid waste management system network design case study, *2010 The 2nd International Conference on Computer and Automation Engineering, ICCAE 2010* .

R., P. and T., L. (2004). Using machine learning techniques to combine forecasting methods, *In: Webb G. I., Yu X. (eds) AI* .

Rathore, S., Pan, Y. and Park, J. H. (2019). BlockDeepNet: A Blockchain-Based Secure Deep Learning for IoT Network, *Sustainability* (14): 1–15.

Restuccia, F., d'oro, S., Kanhere, S., Melodia, T. and Das, S. (2018). Blockchain for the internet of things: Present and future.

Safi, S. (2013). Artificial neural networks approach to time series forecasting for electricity consumption in gaza strip.

Saha, H. and Auddy, S. (2017). Waste management using internet of things (iot).

Salah, K., Rehman, M. H. U., Nizamuddin, N. and Al-Fuqaha, A. (2019). Blockchain for ai: Review and open research challenges, *IEEE Access* pp. 10127–10149.

Shyam, G. K., Manvi, S. S. and Bharti, P. (2017). Smart waste management using internet-of-things (iot), *2017 2nd International Conference on Computing and Communications Technologies (ICCCT)*, pp. 199–203.

Siami Namini, S., Tavakoli, N. and Siami Namin, A. (2018). A comparison of arima and lstm in forecasting time series, pp. 1394–1401.

Singh, S., Sharma, P. K., Yoon, B., Shojafar, M., Cho, G. H. and Ra, I.-H. (2020). Convergence of blockchain and artificial intelligence in iot network for the sustainable smart city, *Sustainable Cities and Society* p. 102364.
URL: *http://www.sciencedirect.com/science/article/pii/S2210670720305850*

Tanwar, S., Bhatia, Q., Patel, P., Kumari, A., Singh, P. K. and Hong, W. (2020). Machine learning adoption in blockchain-based smart applications: The challenges, and a way forward, *IEEE Access* pp. 474–488.

Vyas, S., Gupta, M. and Yadav, R. (2019). Converging blockchain and machine learning for healthcare, *2019 Amity International Conference on Artificial Intelligence (AICAI)*, pp. 709–711.

Wörner, D. and von Bomhard, T. (2014). When your sensor earns money: Exchanging data for cash with bitcoin, pp. 295–298.

Zhang, Z. and Dong, Y. (2020). Temperature forecasting via convolutional recurrent neural networks based on time-series data, *Complexity* pp. 1–8.