IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, December 2023

Designing Smart Cities Models Using Machine Learning Methods in India

Dr. Ashad Ullah Qureshi[1], Praveen Kumar[2], Arshee Naz[3]
Technical Officer, Indian Institute of Information Technology, Sonepat, Haryana[1]
Website Administrator, Shri Vishwakarma Skill University, Dudhola[2]
Junior Research Fellow, National Institute of Technology, Kurukshetra, Haryana[3]
aqureshi@iiitsonepat.ac.in[1], praveen.kumar@svsu.ac.in[2], arshee_jrf@nitkkr.ac.in[3]

Abstract: *Discovering important patterns in data can help cities to plan, monitor, and assign resources more efficiently, converting them in smart cities with more organized communities. Machine learning models can take advantage of this large amount of data to improve and scale these cities' duties. In this work, we explore machine learning approaches to solve different problems in the smart cities domain related to water consumption, energy consumption and emergency events. More specifically, our work sheds light on the design of ensemble learning, sequential models and the combination of probabilistic graphical and deep learning models to this type of problems. Moreover, we carefully compare, adapt and implement methods to address the particular characteristics of the data and the problems of smart cities.*
We are going to focus on four specific problems:
1. Classifying the water pump operation status, quality and quantity,
2. Predicting the future water consumption based on historical consumption,
3. Time resolution prediction for emergency events and
4. Dis-aggregating energy signals into their component appliances.

Keywords: Smart Cities, Machine Learning, Artificial Intelligence, Models of smart city, ML Applications, etc

I. INTRODUCTION

Metropolises across the globe are generating substantial volumes of data on a daily basis. Machine learning algorithms are capable of transforming this data into valuable outputs that can enhance decision-making for public services. Furthermore, they have the potential to enhance comprehension of population behavior in order to more effectively impact public policies. Public servers can enhance their cities and countries by analyzing consumption and mobility trends to develop more effective solutions and optimize resources. The machine learning problems addressed in this thesis encompass various areas, including water and energy consumption and management, as well as emergency event predictions. The user's text is "v". This section outlines the issues and primary contributions of this thesis. We begin by tackling the issue of water supply and management, particularly in the least developed nations. The availability of water in these places is influenced by various factors, such as topographical, political, management, and environmental considerations.

Hence, we construct an ensemble-learning driven predictive-analytic framework for intelligent water management, aimed at forecasting the operational state (e.g., functional or non-functional), as well as the quality and quantity of water pumped. Initially, we engage in feature engineering to carefully choose pertinent features. Subsequently, we employ the Boost and Random Forest ensemble learning models to produce our predictions. Finally, we conduct thorough feature analysis to determine the most influential features for each of the aforementioned prediction problems. We assess the effectiveness of our framework using two publically accessible datasets on smart water management in Tanzania and Nigeria. Our results demonstrate that our suggested models outperform a " *Supper Vector Machine strategy across various measures, including precision, recall, and F1 score*". Furthermore, we showcase that our models exhibit exceptional predictive capabilities in determining the operational health of water pumps across various water extraction techniques. We do a comprehensive feature analysis to examine the significance of different feature

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, December 2023

groups (such as geography and management) in predicting the performance of models for water pump operation status, water quality, and water quantity.

Subsequently, we conduct a meticulous examination of features to determine the specific influence of individual features, rather than merely feature groups, on performance. It is observed that among the individual attributes, the position at coordinates 1 (x, y, z) has the greatest influence on performance. Our analysis provides valuable insights into the specific data that should be gathered in the future to enhance the accuracy of water problem predictions.

II. LITERATURE REVIEW

[1] Diana Gaifulina, Andrey Fedorchenko, and Igor Kotenko" Network Protocols Determination Based on Raw Data Analysis for Security Assessment under Uncertainty" IEEE Xplore2019

This study focuses on the analysis of network traffic in situations when the specifications of the network protocols are unknown. Our proposal involves using text-inspired algorithms to analyze raw data and discover the common structures of network protocols. We also aim to find the lexical specifications of these structures. The research topic is highly relevant due to the presence of significant heterogeneity, partly lexical uncertainty, and the utilization of new, proprietary, or customized data transfer protocols in computer networks.

Pertinence to the subject matter: We provide the method of network traffic analysis and provide the experimental findings that validate the practicality of the proposed strategy.

[2] Nils Maurer, German Aerospace Center, Oberpfaffenhofen, Germany Corinna "Towards Successful Realization of the LDACS CyberSecurity Architecture: An Updated Data Link Security Threat and Risk Analysis" IEEE Xplore2018

Currently, there are significant modifications taking place in the field of Communication Navigation and Surveillance (CNS) in civil aviation, particularly in the European SESAR and the U.S. NextGEN research efforts. The digitalization progress in communication, particularly in essential infrastructure, leads to frequency saturation and automated data processing, which are responsible for solving data storage issues. Consequently, a comprehensive assessment of potential dangers and vulnerabilities for LDACS was conducted, leading to the creation of an initial draft of the cybersecurity architectural specification.

Conclusion: This study takes an additional stride by introducing a fitting collection of techniques and protocols to enhance security for LDACS. The set is assessed in terms of performance and security to align with the cybersecurity architectural definition defined in previous research.

[3] Souparnika Jayaprakash, Kamalanathan Kandasamy" Database Intrusion Detection System Using Octraplet and Machine Learning" IEEE Xplore2018

In recent years, there has been a significant rise in digitization, resulting in the constant automation and online availability of every service. Online services have become widely popular and trusted for securely storing all personal and private user information in databases. Consequently, the attackers shifted their attention onto the databases that house vital information. Despite the presence of security procedures for host-based systems and networks, security breaches persist on a daily basis, resulting in data theft. Therefore, prioritizing database security becomes imperative. This study presents a comprehensive automated database intrusion detection system that effectively deals with both internal and external threats, which have the potential to bypass traditional network or host-based intrusion detection systems without being noticed. The proposed solution is adaptable and optimized to accommodate the growing intricacy and ever-changing nature of databases. Our Architecture is a detection approach that utilizes anomaly-based methods and incorporates Role-based Access control (RBAC). They design the framework to implement the role-based access control and introduce a novel data structure called triplet, which records the SQL queries.

Relevance to the subject: This system employs a Naive Bayes Classifier, which is a supervised Machine Learning technique used to identify abnormal queries. The proposed methodology has the potential to enhance both the detection rates and overall performance of the system.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, December 2023

[4] Alexandria Farar and Hayrettin Bahşi, Bernhards Blumbergs" A Case Study: The Use and Evaluation of Cyber Deceptive Methods Against Highly Targeted Cyber Attacks" IEEE Xplore2018

Conventional security measures, such as intrusion detection systems, firewalls, and antivirus software, are insufficient in thwarting security breaches induced by specifically tailored cyber attacks. This paper presents the findings of a case study that utilizes a technique to identify, map, develop, test, and monitor deceptive measures for the purpose of early detection, considering that many of these attacks often go unnoticed. Metrics are created to verify the efficacy of the deception implementation. Deception mechanisms are assigned to the initial three stages of the intrusion kill chain, namely reconnaissance, weaponization, and delivery. Subsequently, Red Teams were enlisted to evaluate the deceptions in two specific case scenarios.

Conclusion: By utilizing metrics, it is evident that the deceptions employed in the case studies efficiently identify cyber threats prior to the exploitation of the target asset, and effectively generate confusion and ambiguity among attackers regarding the organization's network architecture, services, and resources.

[5] T.T. Teoh 1a, Y.Y. Nguwi, Yuval Elovici1b, N.M. Cheung, W.L.Ng ``Analyst Intuition Based Hidden Markov Model On High Speed, Temporal Cyber Security Big Data" IEEE Xplore2017

Hidden Markov Models (HMM) are stochastic models used for predicting time series data. It has achieved success in the fields of finance, bioinformatics, healthcare, agriculture, and artificial intelligence. However, the application of Hidden Markov Models (HMM) in the field of cyber security has been limited. The predicted probabilistic nature of HMM and its capacity to represent several naturally occurring states make it an ideal framework for modeling cyber security data. Therefore, the purpose of this study is to present the preliminary findings of our efforts to forecast security breaches using Hidden Markov Models (HMM). This work utilizes a comprehensive network dataset that represents cyber security attacks to construct an expert system. Statistical data can be generated by extracting the features of attackers' IP addresses from our linked datasets. The cyber security specialist assigns a numerical value to each attribute and creates a grading system based on the analysis of the log history. We utilized Hidden Markov Models (HMM) to differentiate between a cyber security attack, uncertain events, and the absence of an assault. This was achieved by partitioning the data into three clusters using Fuzzy K-means (FKM) algorithm, followed by human labeling of a small dataset based on analyst intuition. Finally, we employed a state-based method utilizing HMM.

Conclusion: Our findings are highly promising in contrast to detecting irregularities in a cyber security log, which typically leads to a significant number of false positives.

2.1 OBJECTIVES

- Improved traffic management: Machine learning can be used to optimize traffic flow, reduce congestion and improve overall transportation efficiency.
- Enhanced public safety: Machine learning can be used to predict and prevent crime, as well as to improve emergency response times.
- Increased energy efficiency: Machine learning can be used to optimize energy usage and reduce waste, resulting in cost savings for residents and businesses.
- Improved public services: Machine learning can be used to predict and prevent service outages, as well as to improve overall service quality.
- Increased citizen engagement: Machine learning can be used to analyze data from social media and other sources to gain insights into the needs and preferences of citizens, allowing for more effective and responsive government services

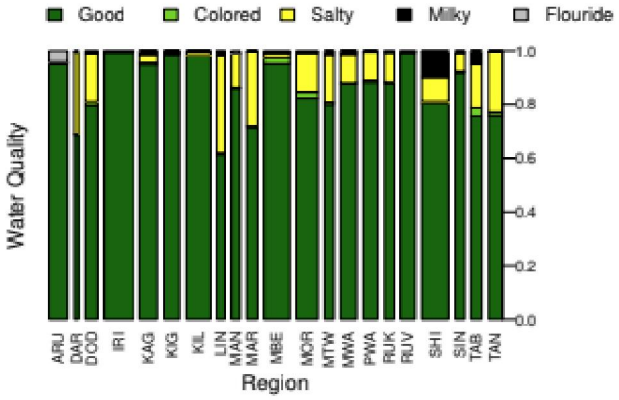
III. RESEARCH METHODOLOGY

DATASET

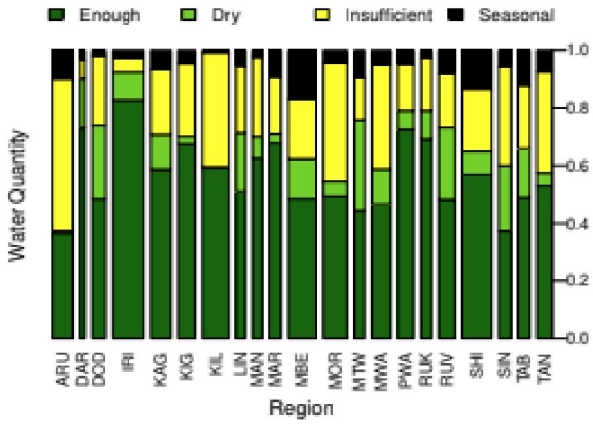
We utilize two datasets sourced from Africa, specifically Tanzania and Nigeria. Both datasets have been released to the public by Taarifa and the respective Ministries of Water in Tanzania and Nigeria, as referenced in sources [22] and [23]. The Tanzania dataset was acquired through the utilization of portable sensors, written reports, and input from individuals utilizing mobile phones. The dataset consists of 59,401 cases and includes information on the operational

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, December 2023

condition of the pump, water quality, water quantity, pump location, source type, extraction technique, and population demographics in the region where the pump is located. The Nigeria dataset contains 132,542 cases and shares certain properties with the Tanzania dataset, however it has less features in comparison.



(b) Water quality



(c) Water quantity

Figure 3.1: Tanzania: Comparison of pump operation status, water quality and

Water Quantity for Different Regions

Smart system for Bike-sharing management

An intelligent bike sharing system can be defined as a machine that integrates a server (computer) connected to the Internet that can collect and analyse data and communicate with other systems.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, December 2023

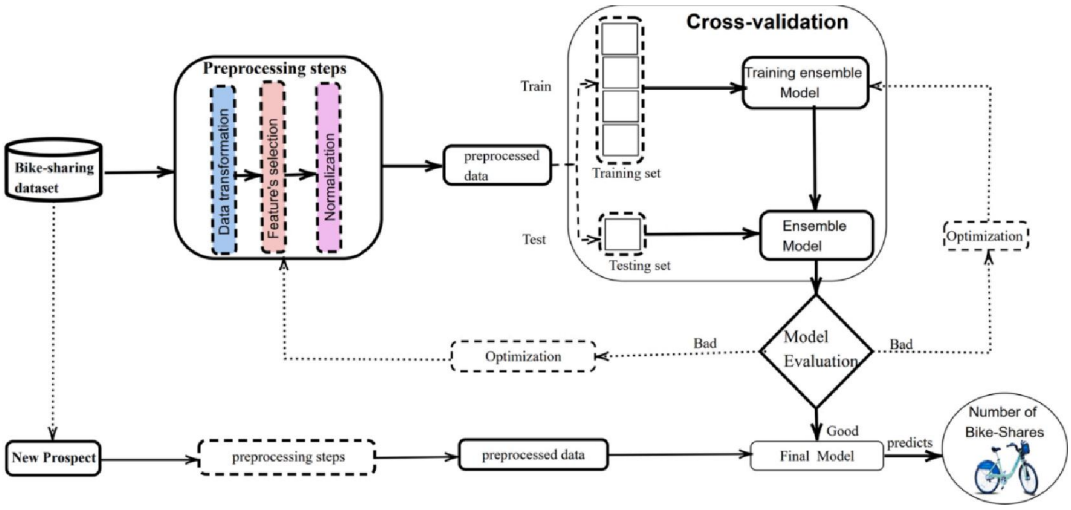


Fig. 3. Global ensemble-based system for real-time average number of bike-share prediction.

IV. QUALITATIVE RESULTS

In this paragraph, we assess the qualitative prediction performance of the GCRF and LSTM models in comparison to the baselines. This analysis aims to demonstrate the superior performance of our models. Display the 1-hour and 12-hour forecasts for GCRF and linear regression, as well as the 1-hour and 12-hour forecasts for LSTM and ARIMA, specifically for the residence hall RA. When making a 1-hour forecast, we see that linear regression, which tends to closely track the actual values in the previous time step, does not perform well because the recent past may not accurately reflect the future. By contrast, GCRF produces smoothed predictions because to its training on complete input sequences, resulting in greater performance. Furthermore, it is evident from Figure b that the 12-hour forecast for linear regression is much inferior to its 1-hour forecast. When comparing, it is observed that GCRF, which considers whole sequences and captures the underlying fluctuations in the data, does not experience a substantial decline in its 12-hour prediction ability. Like GCRF, LSTM is a sequence-to-sequence model that effectively captures the dependencies in the data. Importantly, its prediction performance remains unaffected even with bigger time steps.

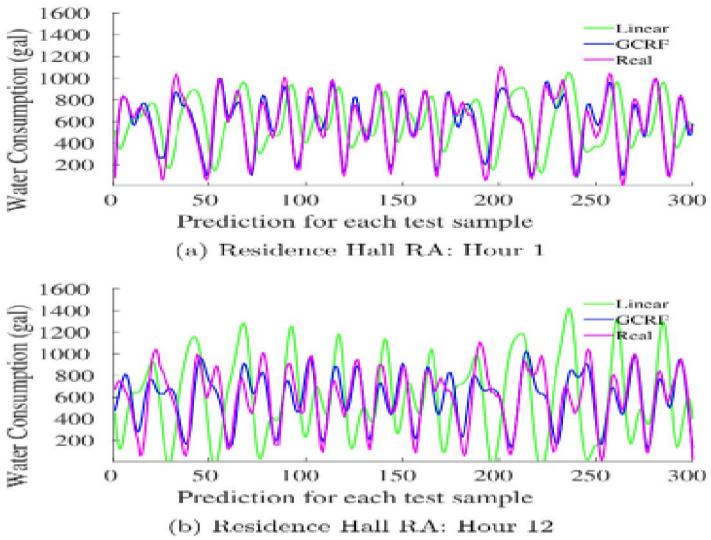


Figure: Qualitative Results: GCRF vs Linear Regression

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, December 2023

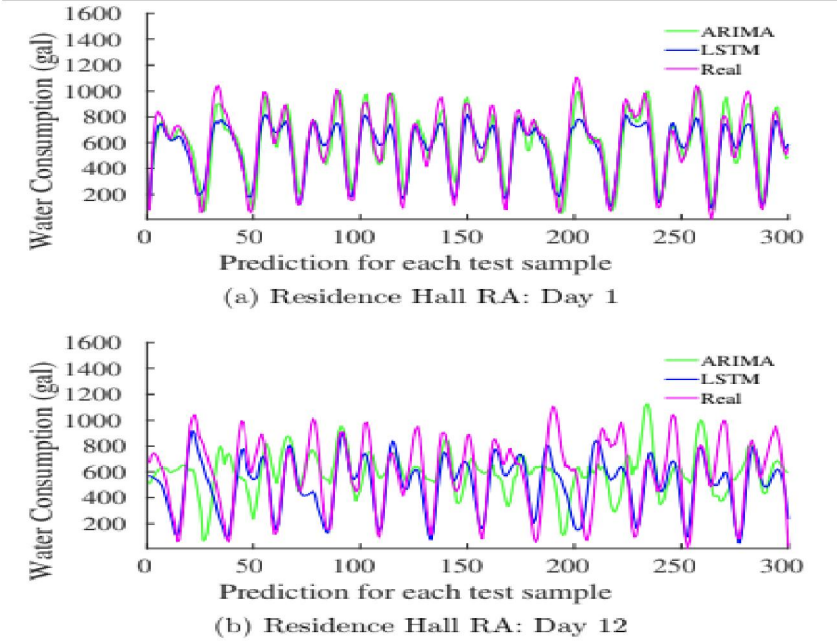


Figure: Qualitative Results: LSTM vs ARIMA

V. CONCLUSION

This chapter focused on the topic of predicting hourly water use using data gathered from various buildings within a university campus. SWaP, a Smart Water Prediction system, was developed to precisely forecast future hourly water consumption using historical data. In order to optimize SWaP (Size, Weight, and Power) and improve prediction accuracy, we developed discriminative probabilistic graphical models and deep learning models. Specifically, we utilized sparse Gaussian Conditional Random Fields (GCRF) and Long Short-Term Memory (LSTM) based deep models to effectively capture the interdependencies within the water consumption data. Our experimental evaluation demonstrates that SWaP outperforms linear regression and ARIMA baselines in terms of prediction accuracy for all buildings, as measured by RMSE and MAE. Furthermore, we have noted that a model based on the GCRF algorithm exhibits superior performance compared to a deep learning model based on LSTM. Hence, we suggest using the computationally efficient and interpretable GCRF-based SWaP, which enhances the practical appeal of our approach.

We introduced an innovative deep generative framework that incorporates a recently created generative model, VRNNs, for the purpose of energy disaggregation. We have shown that our model can accurately forecast individual appliance consumption signals by disaggregating the aggregated energy consumption using a sequence-to-many-sequence approach. We have additionally showcased the ability of our models to achieve exceptional performance in two widely recognized real-world energy disaggregation datasets, DataPort and REDD. Our models have achieved a remarkable 29% and 41% enhancement in Mean Absolute Error (MAE) compared to the current state-of-the-art methods. We have also showcased the proficiency of our framework in precisely forecasting the energy usage of low-power appliances that lack any noticeable repetitive pattern. This achievement sets the stage for a detailed and knowledgeable energy disaggregation process.

REFERENCES

[1] E. O'Dwyer, I. Pan, S. Acha, N. Shah, Smart energy systems for sustainable smart cities: current developments, trends and future directions, Appl. Energy 237 (2019) 581–597.

[2] Y. Liu, C. Yang, L. Jiang, S. Xie, Y. Zhang, Intelligent edge computing for IoT-based energy management in smart cities, IEEE Netw. 33 (2) (2019) 111–117.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, December 2023

[3] R. Petrolo, V. Loscri, N. Mitton, Towards a smart city based on cloud of things,a survey on the smart city vision and paradigms, Trans. Emerg. Telecommun. Technol. 28 (1) (2017) e2931.
[4] U. Aguilera, O. Peña, O. Belmonte, D. López-de Ipiña, Citizen-centric data services for smarter cities, Future Gener. Comput. Syst. 76 (2017) 234–247.
[5] P. Neirotti, A. De Marco, A.C. Cagliano, G. Mangano, F. Scorrano, Current trends in smart city initiatives: some stylised facts, Cities 38 (2014) 25–36.
[6] F. Al-Turjman, I. Baali, Machine learning for wearable iot-based applications: a survey, Trans. Emerg. Telecommun. Technol. (2019) e3635.
[7] F.M. Al-Turjman, Information-centric sensor networks for cognitive IoT: an overview, Ann. Telecommun. 72 (1–2) (2017) 3–18.
[8] F. Al-Turjman, Information-centric framework for the Internet of Things (IoT):
Traffic modeling & optimization, Future Gener. Comput. Syst. 80 (2018) 63–75. [9] Z. Allam, Z.A. Dhunny, On big data, artificial intelligence and smart cities, Cities 89 (2019) 80–91.
[10] H. Li, T. Wei, A. Ren, Q. Zhu, Y. Wang, Deep reinforcement learning: framework, applications, and embedded implementations, in: 2017 IEEE/ACM International Conference on Computer-Aided Design, ICCAD, IEEE, 2017, pp. 847–854.
[11] S. Ramchurn, P. Vytelingum, A. Rogers, N.R. Jennings, Putting the ‘‘smarts’’ into the smart grid: a grand challenge for artificial intelligence, Commun. ACM 55 (4) (2012) 86–97.
[12] Z. Allam, P. Newman, Redefining the smart city: culture, metabolism and governance, Smart Cities 1 (1) (2018) 4–25.
[13] H. Habibzadeh, T. Soyata, B. Kantarci, A. Boukerche, C. Kaptan, Sensing, communication and security planes: a new challenge for a smart city system design, Comput. Netw. 144 (2018) 163–200.
[14] A. Ferdowsi, U. Challita, W. Saad, Deep learning for reliable mobile edge analytics in intelligent transportation systems: an overview, IEEE Veh. Technol. Mag. 14 (1) (2019) 62–70.