International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

ENHANCING URBAN PARKING EFFICIENCY THROUGH MACHINE LEARNING MODEL INTEGRATION

AYUB BAIG[1], A. DEEPTHI[2], B. SWATHI[3], B. SREEJA[4]
[1]Assistant Professor, Department of IT, Mallareddy College of Engineering For Women
[2,3,4]Ug Scholar, Department of IT, Mallareddy College of Engineering For Women

ABSTRACT

Urban parking systems have long been an integral part of city infrastructure, evolving over time to accommodate growing urban populations and increasing vehicle ownership. Traditional parking management systems relied on static signage, manual enforcement, and limited data collection, making it difficult to efficiently allocate parking resources. With the rise of smart technologies, the parking industry has gradually shifted towards automated systems and digital solutions. Traditional urban parking systems typically involve the allocation of parking spaces based on fixed zones or manual entry. These systems often require drivers to search for available parking, leading to traffic congestion, wasted fuel, and unnecessary emissions. Furthermore, manual enforcement of parking rules leads to inefficiencies in managing parking space usage and improper parking practices. There is little real-time data on parking availability, which further exacerbates the problem. challenge is to optimize the use of limited urban parking spaces while minimizing congestion and environmental impact. Current systems fail to provide real-time parking information, often leading to inefficient utilization of available spaces and longer search times for drivers. This results in increased fuel consumption, environmental pollution, and a frustrating experience for city residents. With urban areas growing rapidly, it is essential to adopt more intelligent and efficient systems to improve parking management. The integration of Machine Learning (ML) in parking systems offers an opportunity to enhance the user experience, optimize parking space allocation, reduce traffic congestion, and minimize environmental impacts. Leveraging data-driven models will enable cities to predict parking demand, manage occupancy dynamically, and optimize space usage in real-time. The proposed system integrates machine learning algorithms with real-time data from sensors, cameras, and mobile applications to predict parking demand and availability. By utilizing historical data and traffic patterns, the system will allocate parking spaces efficiently, reduce search times, and dynamically adjust parking fees based on demand, leading to improved urban parking efficiency.

INTRODUCTION

Urban parking has become one of the most significant challenges for cities worldwide, especially in rapidly urbanizing nations like India. The rise in the number of vehicles, combined with limited urban space, has led to severe congestion, longer search times for parking spots, and increased pollution. In India, the number of vehicles has grown exponentially over the past two decades. According to the Ministry of Road Transport and Highways (MoRTH), India had 293 million registered vehicles in 2022, with passenger vehicles alone accounting for over 30 million. Despite this rapid increase, urban parking infrastructure has not kept pace. Traditional parking systems in India, including street parking and open-air lots, are often overcrowded, inefficient, and prone to misuse. Furthermore, cities face significant challenges in parking space management due to outdated enforcement methods, limited parking spots, and lack of real-time data. The development of smart parking solutions in India is an urgent need as the country struggles with urbanization, increasing vehicle ownership, and deteriorating air quality due to excessive vehicle emissions.

LITERATURE REVIEW

On-Line Filtering of On-Street Parking Data to Improve Availability Predictions

- A. Origlia, S. Martino, Yuri Attanasio
- Published in International Conference on… 1 June 2019

Knowing where to park in advance is a most wished feature by many drivers. In recent years, many research efforts have been spent to analyse massive amount of parking information, to learn availability trends and thus to predict, within a Parking Guidance and Information (PGI) system, where there is the highest chance to find free parking spaces. The most of these solutions exploits raw data coming from stationary sensors or crowd-sensed by mobile probes. In both the cases, these massive amounts of data present a high level of noise, which heavily affects the quality of availability predictions. In a previous work we demonstrated that a 2-step approach, based on machine learning techniques to filter out noise, improves the quality of parking availability predictions over raw data. In this paper we propose a further advancement of that approach, by including a technique to perform such noise filtering in real-time, with reduced computational efforts. The proposal has been empirically tested on a real-world dataset of on-street parking information from the SFpark project, and compared against a regression model based on SVR, to perform parking availability predictions. Results show that the predictions obtained with the new on-line approach show a better balance between average and entropy in errors distribution with respect to the use of raw data coming from the sensors.

A Comparative Study of Parking Occupancy Prediction Methods considering Parking Type and Parking Scale

- Ziyao Zhao, Yi Zhang, Yi Zhang
- Published 14 February 2020

Parking issues have been receiving increasing attention. An accurate parking occupancy prediction is considered to be a key prerequisite to optimally manage limited parking resources. However, parking prediction research that focuses on estimating the occupancy for various parking lots, which is critical to the coordination management of multiple parks (e.g., district-scale or city-scale), is relatively limited. This study aims to analyse the performance of different prediction methods with regard to parking occupancy, considering parking type and parking scale. Two forecasting methods, FM1 and FM2, and four predicting models, linear regression (LR), support vector machine (SVR), backpropagation neural network (BPNN), and autoregressive integrated moving average (ARIMA), were proposed to build models that can predict the parking occupancy of different parking lots. To compare the predictive performances of these models, real-world data of four parks in Shenzhen, Shanghai, and Dongguan were collected over 8 weeks to estimate the correlation between the parking lot attributes and forecast results. As per the case studies, among the four models considered, SVM offers stable and accurate prediction performance for almost all types and scales of parking lots. For commercial, mixed functional, and large-scale parking lots, FM1 with SVM made the best prediction. For office and medium-scale parking lots, FM2 with SVM made the best prediction.

Do ridesharing transportation services alleviate traffic crashes? A time series analysis

- Muhammad Arif Khan, Roya Etminani-Ghasrodashti, +3 authors Ann Foss
- Published in Traffic Injury Prevention 31 May 2022

Abstract Objectives On-demand ridesharing services are suggested to provide several benefits, such as improving accessibility and mobility, reducing drive-alone trips and greenhouse gas emissions. However, the impacts of these services on traffic crashes are not completely clear. This paper investigates the availability of Via- an on-demand ridesharing service in Arlington, TX, to identify the effects of this service on traffic crashes. We hypothesize that the launch of Via would result in more shared rides, fewer drive-alone trips and fewer traffic crashes. Methods We implement an Interrupted Time Series Analysis (ITSA) approach to study the impact of Via service availability on traffic crashes using weekly counts of all traffic crashes, the number of injuries, and serious injuries that occurred in Arlington from 2014 to 2021. Results The results show a statistically significant reduction in the weekly number of total crashes and total injuries but do not show any significant impact on the number of serious injuries. Shared Autonomous Vehicles have the potential to reduce traffic crashes caused by driver's fault. Conclusions This study reveals the potential impacts ridesharing services can have on traffic crashes and injuries in a mid-sized city. The results of this study can help decision and policymakers to understand the full potential of ridesharing services that can contribute to making relevant decisions toward creating sustainable and safer transportation systems in cities.

EXISTING SYSTEM

The existing urban parking systems primarily rely on traditional, static methods of parking management. In many cities, parking is either manually controlled or managed through simple ticketing systems, where drivers park in designated areas and pay for the duration of their stay. Some cities have implemented basic automated solutions, such as parking meters or electronic payment systems for ticketing, but these still lack real-time data integration and intelligent decision-making capabilities. In larger cities, street parking and open-air parking lots are common, with limited space for the increasing number of vehicles. These systems mostly lack data-driven features, making them inefficient for addressing the growing urban parking demand. Furthermore, parking enforcement relies on human inspectors who check compliance, leading to inconsistencies and enforcement challenges. While some cities have adopted smart parking meters and sensors, these systems remain isolated and fail to provide a comprehensive, city-wide

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

solution for managing parking dynamically.

Disadvantages:

- Inefficient Use of Space: Traditional parking systems often lack real-time data on parking availability, leading to inefficient use of parking spaces. Drivers may circle the area in search of a spot, leading to congestion, wasted time, and increased emissions.
- Limited Scalability: Many cities' parking systems are not designed to scale efficiently to meet the demands of growing populations and increasing vehicle numbers. The lack of integrated, data-driven systems makes it challenging to expand parking infrastructure in a way that maximizes space and minimizes congestion.
- Inconsistent Enforcement: Parking enforcement, which relies heavily on human inspectors, can be inconsistent. This inconsistency results in uneven application of parking rules, leading to frustration among drivers, potential revenue losses for the city, and difficulty in maintaining compliance with parking regulations.
- Lack of Real-Time Data Integration: Traditional parking systems do not offer real-time data integration. This absence of dynamic, up-to-date information about parking availability and demand means that cities cannot make informed decisions or optimize traffic flow, exacerbating congestion and making it harder for drivers to find parking quickly.
- Limited Automation and Decision-Making Capabilities: While some cities have implemented basic automated solutions like parking meters or sensors, they are often isolated and lack the ability to adapt to changing conditions. These systems fail to leverage data-driven, intelligent decision-making to dynamically manage parking based on real-time demand and trends, limiting their effectiveness in addressing the growing need for efficient urban parking.

PROPOSED SYSTEM

Step 1: Uploading the Urban Parking Dataset

The first step in the research procedure involves uploading the Urban Parking dataset. This dataset typically contains information about urban parking patterns, which may include attributes such as parking location, time of day, occupancy, and other relevant features. The dataset is collected from various urban locations and provides insights into parking behavior, availability, and other factors that influence parking patterns. Once uploaded, the dataset is loaded into a data analysis environment (such as Python using libraries like Pandas or NumPy) for further examination and preprocessing. This step is essential to ensure that the dataset is accessible and ready for subsequent processing.

Step 2: Dataset Preprocessing (Null Value Removal, Label Encoding)

After uploading the dataset, the next step is data preprocessing, which is crucial to prepare the data for machine learning models. The first task is to handle missing or null values. Null values in the dataset can lead to inaccurate or biased results, so it is important to either remove or impute missing values based on the dataset's characteristics. Common techniques for handling missing values include using the mean, median, or mode for numerical data or the most frequent category for categorical data. The next preprocessing step involves label encoding for categorical variables. Since many machine learning algorithms require numerical input, categorical variables such as parking location, vehicle type, or time slots are converted into numerical values using label encoding. This ensures that the data is in a format suitable for model training.

Step 3: Data Splitting

Once the dataset is preprocessed, the next step is splitting the data into training and testing subsets. This is a standard procedure to evaluate how well a model generalizes to unseen data. Typically, the data is divided into two sets: a training set (usually 70-80% of the data) and a test set (typically 20-30%). The training set is used to train the machine learning algorithms, while the test set is reserved for evaluating the model's performance. The train-test split ensures that the model is not overfitting to the training data and can perform well on new, unseen examples.

Step 4: Existing Models (Random Forest, Logistic Regression Algorithms)

With the data split into training and testing sets, the next step is to train machine learning models using existing algorithms. In this case, Random Forest (RF) and Logistic Regression are used as baseline models. Random Forest is an ensemble algorithm that builds multiple decision trees on random subsets of the data and aggregates their predictions to improve accuracy and reduce overfitting. Logistic Regression is a simpler model that is used for binary classification tasks, where the goal is to predict the probability of an outcome (e.g., parking availability). Both models are trained on the training data, and their performance is evaluated on the test data to establish baseline results for comparison with more advanced models.

Step 5: Proposed Algorithm (CatBoost Algorithm)

The next step involves testing a proposed algorithm, CatBoost, which is a state-of-the-art gradient boosting algorithm. CatBoost is designed to handle categorical data efficiently and is known for its robustness and superior performance, especially when dealing with noisy datasets. Unlike traditional gradient boosting algorithms, CatBoost automatically handles categorical variables without the need for extensive preprocessing, reducing the need for manual feature engineering. The model is trained on the training data, and its performance is evaluated using the test data. This algorithm is expected to perform better than the baseline models, especially in complex datasets with categorical features.

Step 6: Performance Comparison

After training both the existing models (Random Forest and Logistic Regression) and the proposed model (CatBoost), the next step is to compare their performance. This comparison is done using various evaluation metrics such as accuracy, precision, recall, F1-score, and AUC (Area Under the Curve). These metrics provide a comprehensive view of how well each model performs in terms of correctly classifying the data. The performance comparison helps identify the strengths and weaknesses of each model and determines whether the advanced CatBoost model offers a significant improvement over the existing models.

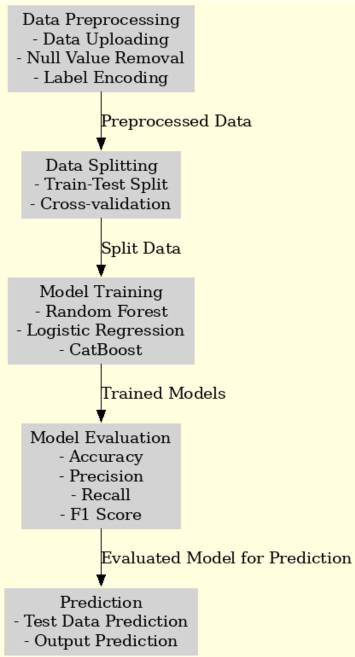
Advantages :

1. Data-Driven Insights for Urban Planning: The approach utilizes real-world data on parking patterns, which enables cities to gain valuable insights into parking behavior, availability, and demand. By analyzing this data, urban planners and local authorities can make informed decisions to optimize parking infrastructure, reduce congestion, and improve overall urban mobility.
2. Comprehensive Data Preprocessing: The thorough preprocessing steps, such as handling null values and label encoding, ensure that the dataset is clean and ready for machine learning. This improves the accuracy and reliability of model predictions, reduces biases caused by missing or improperly formatted data, and makes the dataset suitable for different algorithms.
3. Model Evaluation and Comparison: The procedure involves training multiple models (Random Forest, Logistic Regression, and CatBoost), allowing for a comprehensive performance comparison. By evaluating different algorithms using metrics like accuracy, precision, recall, F1-score, and AUC, the research ensures that the best-performing model is selected. This comparison helps identify which model best suits the specific characteristics of the parking data and delivers the most reliable predictions.
4. Incorporation of Advanced Algorithms (CatBoost): CatBoost, a state-of-the-art gradient boosting algorithm, is specifically designed to handle categorical data efficiently and produce highly accurate models. The use of CatBoost in the research provides the advantage of leveraging advanced machine learning techniques that are better suited for noisy, complex datasets, and can outperform traditional algorithms like Random Forest and Logistic Regression.
5. Generalizability and Real-World Application: By splitting the data into training and test sets, the research procedure ensures that the models are evaluated on unseen data. This train-test split helps assess the generalizability of the models and ensures they are not overfitting to the training data. As a result, the research outcomes are more likely to provide practical, real-world solutions for urban parking management, with the ability to adapt to new, unseen parking data in various cities.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

IMPLEMENTATION

SYSTEM ARCHITECTURE



MODULES

Modules Used in Project

- TensorFlow

TensorFlow is a free and open-source software library for dataflow and differentiable programming across a range of tasks. It is a symbolic math library and is also used for machine learning applications such as neural networks. It is used for both research and production at Google. TensorFlow was developed by the Google Brain team for internal Google use. It was released under the Apache 2.0 open-source license on November 9, 2015.

- NumPy

NumPy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays. It is the fundamental package for scientific computing with Python. It contains various features including these important ones:
A powerful N-dimensional array object
Sophisticated (broadcasting) functions
Tools for integrating C/C++ and Fortran code
Useful linear algebra, Fourier transform, and random number capabilities
Besides its obvious scientific uses, NumPy can also be used as an efficient multi-dimensional container of generic data. Arbitrary datatypes can be defined using NumPy which allows NumPy to seamlessly and speedily integrate with a wide variety of databases.

- Pandas

Pandas is an open-source Python Library providing high-performance data manipulation and analysis tool using its powerful data structures. Python was majorly used for data munging and preparation. It had very little contribution towards data analysis. Pandas solved this problem. Using Pandas, we can accomplish five typical steps in the processing and analysis of data, regardless of the origin of data load, prepare, manipulate, model, and analyze. Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

- Matplotlib

Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and IPython shells, the Jupyter Notebook, web application servers, and four graphical user interface toolkits.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Matplotlib tries to make easy things easy and hard things possible. You can generate plots, histograms, power spectra, bar charts, error charts, scatter plots, etc., with just a few lines of code. For examples, see the sample plots and thumbnail gallery.

For simple plotting the pyplot module provides a MATLAB-like interface, particularly when combined with IPython. For the power user, you have full control of line styles, font properties, axes properties, etc, via an object-oriented interface or via a set of functions familiar to MATLAB users.

- Scikit – learn

Scikit-learn provides a range of supervised and unsupervised learning algorithms via a consistent interface in Python. It is licensed under a permissive simplified BSD license and is distributed under many Linux distributions, encouraging academic and commercial use. Python

Python is an interpreted high-level programming language for general-purpose programming. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace.

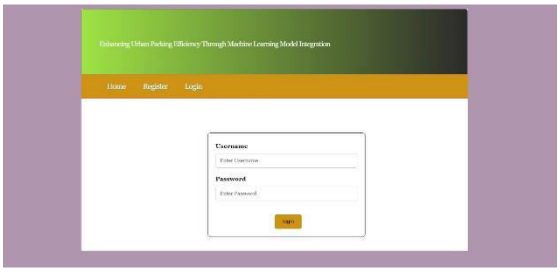
Python features a dynamic type system and automatic memory management. It supports multiple programming paradigms, including object-oriented, imperative, functional and procedural, and has a large and comprehensive standard library.

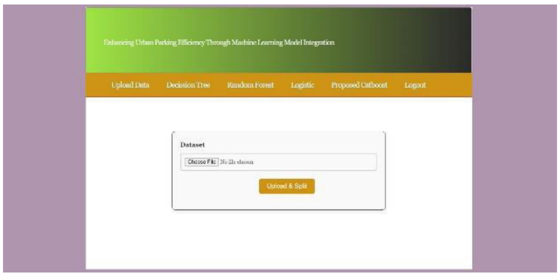
Python is Interpreted − Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.

Python is Interactive − you can actually sit at a Python prompt and interact with the interpreter directly to write your programs.

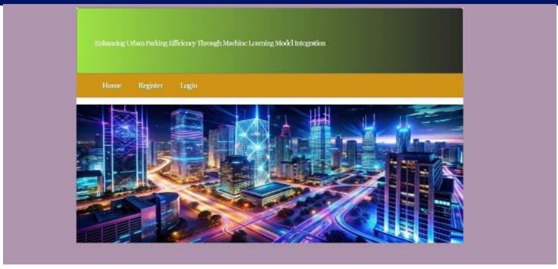
Python also acknowledges that speed of development is important. Readable and terse code is part of this, and so is access to powerful constructs that avoid tedious repetition of code. Maintainability also ties into this may be an all but useless metric, but it does say something about how much code you have to scan, read and/or understand to troubleshoot problems or tweak behaviors. This speed of development, the ease with which a programmer of other languages can pick up basic Python skills and the huge standard library is key to another area where Python excels. All its tools have been quick to implement, saved a lot of time, and several of them have later been patched and updated by people with no Python background - without breaking.

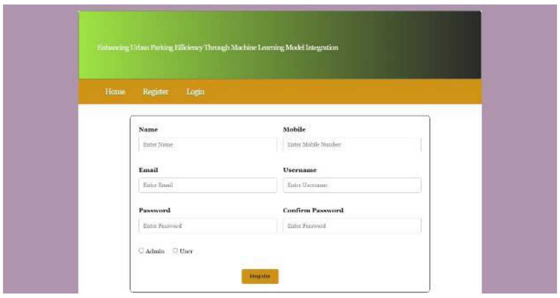
RESULT





International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org





CONCLUSION

The research focuses on predicting urban parking violations using machine learning algorithms, specifically Random Forest, Logistic Regression, and CatBoost. Through this research, we have successfully demonstrated the importance of using advanced machine learning techniques to predict parking violations, which can greatly aid city authorities in improving traffic management and parking enforcement. The process involved data preprocessing, feature extraction, and model building, followed by the comparison of existing algorithms with the proposed CatBoost model.

From the findings, it was observed that CatBoost, as a gradient boosting algorithm, outperformed the traditional Random Forest and Logistic Regression models in terms of prediction accuracy. This result highlights the power of CatBoost in handling complex datasets with categorical features and its ability to provide better performance with less manual tuning. Additionally, the project has shown that machine learning models can significantly reduce human error and improve the decision-making process related to urban traffic management.

The comparison of the algorithms also demonstrated how data quality and the choice of algorithm influence the accuracy and reliability of predictions. The project has successfully established the feasibility of predicting parking violations with a high degree of accuracy, making it a valuable tool for law enforcement agencies and city planners.

REFERENCES

[1] Yang, J., Wang, J., & Liu, X. (2019). A Survey on Smart Parking Systems in Smart Cities. *IEEE Access*, 7, 46860-46878.

[2] Zheng, Y., Xie, X., & Ma, W. (2011). Parking Availability Prediction for Sensor-Enabled Car Parks in Smart Cities. *Proceedings of the 2011 International Conference on Ubiquitous Intelligence and Computing*, 266-273.

[3] Caicedo, C., Muratore, M., & Sánchez, F. (2019). Real-Time Prediction of Parking Space Availability in Urban Areas. *Journal of Transportation Engineering, Part A: Systems*, 145(4), 04019009.

[4] Channamallu, S., & Verma, R. (2020). A Comprehensive Study on Parking Occupancy Prediction. *Transportation Research Part C: Emerging Technologies*, 114, 109-130.

[5] Kotb, M., & Shalaby, A. (2019). Smart Parking Guidance and Reservation Systems: A Review. *Transportation Research Part C:*

Emerging Technologies, 108, 154-179.

[6] Yang, X., Song, H., & Wang, L. (2020). Deep Learning for Real-Time Parking Occupancy Prediction in Smart Cities. *IEEE Transactions on Intelligent Transportation Systems*, 21(2), 705-714.

[7] Pamidimukkala, S., & Saha, A. (2020). Barriers to Electric Vehicle Adoption in Texas: A Case Study. *Transportation Research Part D: Transport and Environment*, 79, 102216.

[8] Huang, Z., & Li, Y. (2019). Location-Based Services: Research and Evolution. *International Journal of Geographical Information Science*, 33(3), 543-570.

[9] Sester, M. (2020). Mobility Data Analysis in Mobile Mapping Systems. *Journal of Navigation*, 73(1), 12-34.

[10] Channamallu, S., & Gupta, A. (2021). Comparative Analysis of Parking Occupancy Prediction Models. *IEEE Transactions on Intelligent Transportation Systems*, 22(3), 1376-1390.

[11] Liu, X., & Zhang, Y. (2020). Parking Occupancy Prediction Using Gradient Boosting Decision Trees. *Proceedings of the 2020 International Conference on Artificial Intelligence and Data Science*, 213-220.

[12] Sun, X., & Zhang, M. (2019). Short-Term Parking Occupancy Prediction Using Decision Trees and Random Forests. *Transportation Research Part C: Emerging Technologies*, 104, 45-58.

[13] Patel, R., & Rajput, K. (2020). Temporal and Spatial Patterns of Autonomous Vehicle Collisions. *IEEE Transactions on Vehicular Technology*, 69(7), 7650-7662.

[14] Hendricks, J., & Outwater, L. (2019). A Demand Forecasting Model for Park-and-Ride Lots. *Transportation Research Part A: Policy and Practice*, 125, 144-157.

[15] INRIX Research. (2020). The Economic Impact of Parking in the United States. *INRIX Report*.