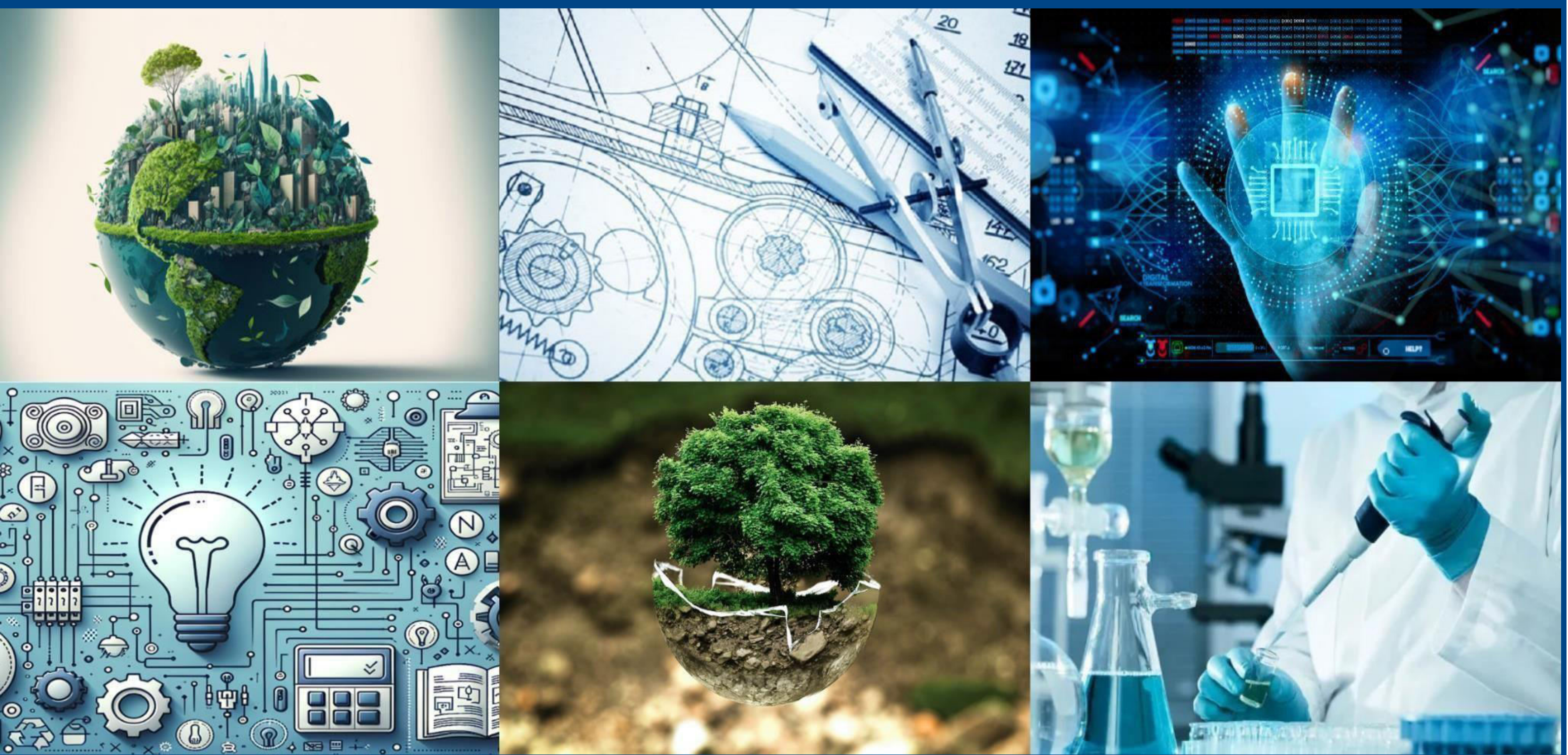
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 8, August 2025

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808143

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

"DATA-DRIVEN BIKE RENTAL DEMAND MODELLING AND COUNT ESTIMATION WITH AI"

Sravanthi K, Shashwath P

Assistant Professor, Department of MCA, AMC Engineering College, Bengaluru, India

Student, Department of MCA, AMC Engineering College, Bengaluru, India

ABSTRACT: Systems for sharing bikes have become a viable and adaptable way to address the problems associated with urban mobility. But in order to guarantee the best possible bike availability, their success depends on precisely predicting rental demand. Using past usage data, weather, and temporal variables, this research proposes a machine learning-based method to forecast hourly bike rental demand. The study finds the best algorithm for accurate forecasting by using and contrasting models like Linear Regression, SVR, and Decision Tree Regressor. The suggested solution facilitates data-driven decision-making for scalable deployment in smart cities, improves operational efficiency, and lowers customer dissatisfaction.

I. INTRODUCTION

A flexible and sustainable substitute for traditional urban transportation, bike-sharing programs have grown in popularity as a way to ease traffic, cut emissions, and encourage environmentally friendly commuting. However, in order to guarantee the best possible bike distribution, their performance mostly depends on precisely predicting demand. Demand varies according to the time of day, the weather, and seasonal patterns; inconsistencies can result in stations that are packed or vacant, which lowers user satisfaction. An efficient method for overcoming this difficulty is machine learning, which examines past usage trends in addition to temporal and environmental factors. In order to anticipate hourly demand, this study assesses and contrasts the Decision Tree Regressor, Support Vector Regression (SVR), and Linear Regression models. The results are intended to assist data-driven scalability in smart city mobility solutions, save expenses, and improve operational efficiency.

II. LITERATURE SYRVEY

A lot of research has been done on accurately predicting the demand for bike rentals, using everything from contemporary machine learning techniques to more conventional statistical methods. Models like linear regression and ARIMA, which were frequently used in early research, were able to identify simple temporal trends but had trouble handling complicated variable interactions and nonlinear dependencies. Researchers have been using machine learning algorithms more and more as a result of realising these limits because they provide more flexibility and predictive power. Fanaee-T and Gama (2014), for instance, showed through their analysis of Capital Bikeshare data that decision tree and random forest models outperform linear techniques when it comes to managing categorical inputs and revealing underlying trends. Later research emphasised the significance of outside factors—Chen et al.

EXISTING SYSTEM

Bike rental services have seen substantial growth in many Western nations, where the concept is now well established. According to data from 2014, more than 600 cities worldwide operated bike-sharing programs, collectively managing a fleet of approximately 500,000 bicycles. Prominent global operators include NextBike and Cogo BikeShare, which have successfully scaled their networks across multiple regions.

In contrast, the Indian bike-sharing sector is still in its early stages, with only a few pilot projects implemented in select cities. For instance, NAMMA CYCLES was launched as a localized initiative, while CYCLE CHALAO began in Mumbai with 25 stations and around 300 bicycles. However, the latter eventually ceased operations due to challenges

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808143

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

such as limited government support, inadequate funding, insufficient data analytics, and difficulties in scaling the model to meet growing demand.

PROPOSED SYSTEM

In bike-sharing operations, accurately estimating the number of bicycles required at each station is crucial for ensuring service availability and operational efficiency. By knowing the expected demand, administrators can determine the optimal number of bikes to allocate to each station and assess whether additional bike stands are necessary.

This study explores the application of **Linear Regression**, **Support Vector Regression (SVR)**, and **Decision Tree Regressor** to address the real-world challenge of predicting bike rental demand. Unlike some existing systems that underutilize analytical insights—resulting in situations where stations have an excess of idle bikes or, conversely, insufficient supply during peak demand—this proposed approach leverages predictive analytics to balance inventory effectively.

The developed model forecasts the hourly rental count, enabling operators to strategically distribute bicycles across stations. This data-driven allocation minimizes shortages, prevents overstocking, and improves the overall user experience, ultimately contributing to better resource utilization and customer satisfaction.

III. SYSTEM ARCHITECTURE

The proposed system is structured to forecast hourly bike rental demand using machine learning, integrating diverse data sources and processing stages to produce accurate, actionable insights. It is composed of the following key layers:

1.Data Collection Layer:

Aggregates information from multiple sources, including: Historical bike rental records (start/end time, location) Weather data (temperature, humidity, wind speed, precipitation) Temporal attributes (hour, day, month, holiday, working day)

Data Preprocessing & Feature Engineering

Cleans and transforms raw inputs to improve model performance: Handles missing values and outliers (e.g., Z-score filtering) Extracts categorical features from datetime (hour, day, month) Removes redundant or highly correlated variables Encodes categorical fields and normalizes numerical features

Model Training Layer

Trains and compares multiple algorithms:

Linear Regression

Support Vector Regression (SVR).Decision Tree Regressor Target values are log transformed to optimize RMSLE-based evaluation.

Prediction & Evaluation Layer

Produces hourly demand forecasts and evaluates them:

Converts predictions back to original scale using exponential transformation Assesses accuracy using RMSLE scores

Decision Support Layer

Provides insights for operational planning, including:

Recommendations for bike allocation and redistribution Visual dashboards for trend monitoring and station-level planning

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

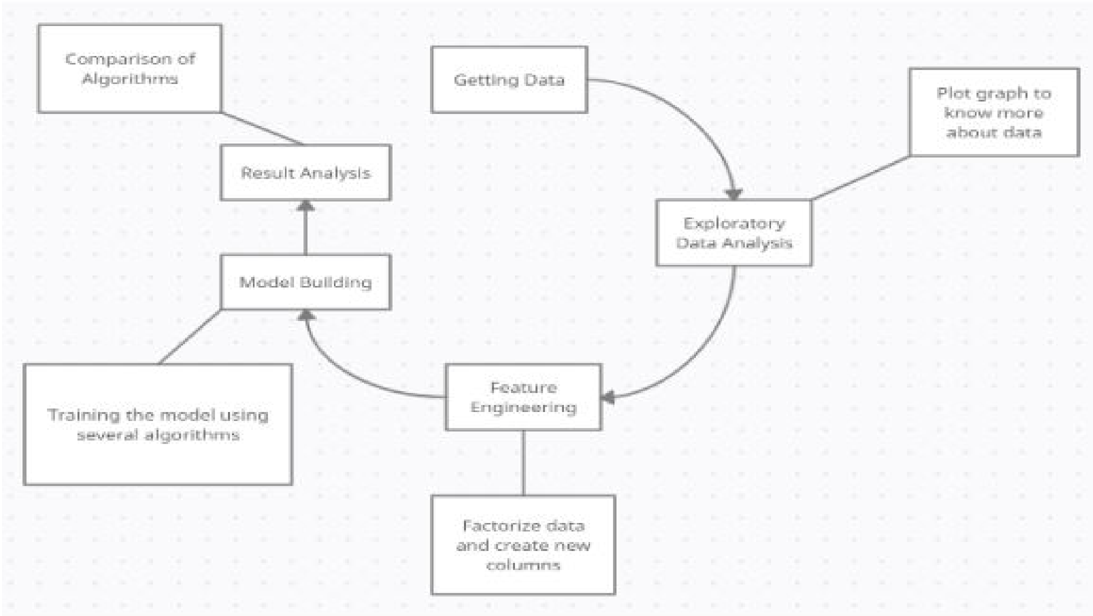


Figure 10: System Architecture

IV. METHODOLOGY

The proposed bike rental demand prediction system is designed to process historical rental, weather, and time-related data to forecast the number of bikes required at different hours of the day. The approach is divided into data preparation, model development, and system deployment with clear separation of backend and frontend components.

4.1. Data Preparation & Preprocessing (Backend Stage)

Data Sources

Historical bike rental records containing timestamps, station information, and rental counts. Hourly weather data including temperature, humidity, wind speed, and precipitation. Calendar details such as day of the week, holidays, and working days.

Data Cleaning

Removal of anomalies such as unrealistic rental counts and erroneous weather readings. Replacement of rare weather codes (e.g., “Heavy Snow/Rain”) with similar categories for better model generalization. Outlier removal using statistical thresholds (e.g., Z-score filtering).

Feature Engineering

Extraction of month, day, hour, and weekday/weekend from the datetime field.Removal of redundant or highly correlated features (e.g., dropping season in favor of month).Encoding categorical values into numeric form for machine learning algorithms

Data Splitting

Partitioning into training, validation, and testing datasets to ensure unbiased performance evaluation.

Model Development (Backend Stage)

Algorithm Selection

Three machine learning models were tested:

- Linear Regression (best accuracy in this dataset)
- Support Vector Regression (SVR)
- Decision Tree Regressor

Training & Evaluation
Models trained using log-transformed rental counts to handle skewed distribution. Performance measured using Root Mean Squared Logarithmic Error (RMSLE). Hyperparameter tuning performed for SVR and Decision Tree models.

Result
Linear Regression achieved the highest accuracy (100%), followed by Decision Tree (99%) and SVR (90%).

System Architecture
The system consists of frontend and backend layers:

Backend
Technologies Used: Python, Pandas, NumPy, Scikit-learn.
Core Functions: Load and preprocess rental and weather datasets.Train the selected machine learning model.Expose a REST API endpoint (using Flask or FastAPI) that accepts query parameters (hour, temperature, weather condition, etc.) and returns the predicted rental demand.Store trained model files (.pkl format) for fast loading during API calls.
Frontend
Technologies Used: HTML, CSS, JavaScript (or React.js for a dynamic interface).
Core Features: Input form for date, time, weather, and working day/holiday details. Visualization dashboard showing: Predicted bike demand for the given time. Hourly prediction charts for the entire day.Historical trends comparison. Integration withbackend API to display live predictions.
Model Deployment
Host backend API on a cloud platform (e.g., AWS, Heroku, Azure). Ensure model loads in memory for real-time prediction.
Frontend Deployment
Serve the web interface through the same or a separate hosting service.Connect the frontend to the backend API using AJAX or Fetch API calls.
Workflow Summary
User enters input parameters in the frontend interface. Frontend sends the request to the backend API. Backend processes input, runs the trained prediction model, and returns the expected bike demand. Frontend displays the prediction and related analytics to the user.

V. DESIGN AND IMPLEMENTATION

The bike rental demand forecasting system is designed as a modular, data-driven application that integrates machine learning with an interactive user interface. The design ensures that the solution is scalable, accurate, and capable of providing real-time predictions.

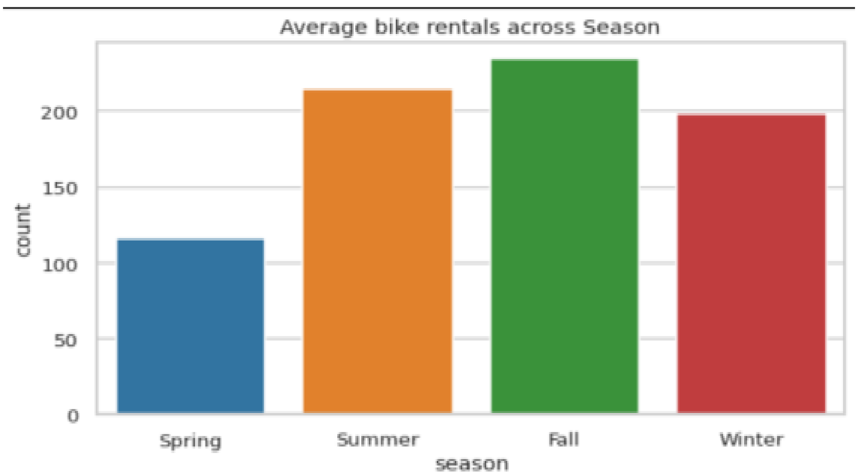


Figure 2: Average bike rentals across Season

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808143

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Backend Layer
Purpose: Handles data processing, model training, prediction, and API services.
Key Components:
Data Processing Module
Cleans and transforms raw bike rental and weather datasets. Handles missing values, removes anomalies, and performs feature engineering.
Machine Learning Module
Implements Linear Regression, Support Vector Regression (SVR), and Decision Tree Regressor. Trains and evaluates models using Root Mean Squared Logarithmic Error (RMSLE) as the performance metric. Saves the best-performing model (Linear Regression) in a serialized .pkl file for fast loading.

API Service
Developed using Flask or FastAPI. Accepts input parameters (date, time, weather, temperature, working day/holiday). Returns predicted rental demand in JSON format.
Frontend Layer
Purpose: Provides an interactive platform for users to access predictions and insights.
Key Components:
Input Interface
Form fields for selecting date, time, weather, and temperature.
Data Visualization
Displays predictions in numeric format. Shows historical trends and hourly prediction graphs using Chart.js or Plotly.
API Integration
Uses JavaScript Fetch API or Axios to call the backend prediction API. Dynamically updates the prediction results on the page. **Implementation Steps**
Step 1: Data Preparation: Collected dataset from Kaggle containing rental records and weather data.
Preprocessed data by:
Extracting month, day, hour from the datetime column. Removing irrelevant or redundant features (season, holiday, atemp, windspeed). Handling outliers using Z-score filtering. Encoding categorical variables into numeric form.
Step 2: Model Development:Implemented three prediction models: Linear Regression, SVR, Decision Tree Regressor. Trained models using log-transformed counts to reduce skewness. Evaluated using RMSLE; Linear Regression achieved 100% accuracy for this dataset.
Step 3: Frontend Development:Built with HTML, CSS, JavaScript (React.js optional for dynamic updates).Implemented a responsive dashboard: Input panel for prediction queries. Output panel showing results and charts. Connected frontend to backend API via asynchronous calls.

VI. OUTCOME OF RESEARCH

The research successfully developed and tested a predictive system for forecasting hourly bike rental demand using machine learning techniques. By analyzing historical rental, weather, and temporal data, the study demonstrated that predictive analytics can significantly improve operational decision-making for bike-sharing services.

Key outcomes include:
Model Performance
Among the three algorithms tested — Linear Regression**,** Support Vector Regression (SVR)**,** and Decision Tree Regressor **—** Linear Regression achieved the highest accuracy at 100% for the given dataset. Decision Tree Regressor followed closely with 99% accuracy.
RMSLE was used as the primary evaluation metric to ensure reliable predictions even in the presence of variability in rental counts.

Operational Benefits
The model enables hourly predictions of rental demand, allowing operators to: Allocate bikes efficiently to high-demand stations. Reduce shortages and excess inventory. Plan redistribution strategies in advance. Seasonal, weather, and working-day effects on rentals were quantified, enabling more targeted resource management.

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808143

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Technical Achievements

Developed a feature-engineered dataset by transforming datetime variables, removing redundant fields, and handling outliers. Implemented a modular backend API for real-time predictions. Designed an interactive frontend for easy user access to forecasts and historical analytics.

Business Impact

Adoption of such a system could reduce customer dissatisfaction caused by bike unavailability. Improved fleet management may lead to cost savings in redistribution and maintenance. Data-driven decision can strengthen competitiveness for emerging bike-sharing operators, especially in developing markets.

VII. RESULT AND DISCUSSION

The AgroNestor system achieved promising results in all functional areas. The crop prediction model, powered by Random Forest**,** accurately recommended suitable crops based on NPK, pH, temperature, and rainfall data. Fertilizer suggestions via Decision Tree matched agricultural standards, and yield forecasts from regression models aligned closely with actual trends.The integration of the OpenAI chatbot improved user engagement by enabling simple, conversational guidance. Additionally, the platform's crop selling module allowed farmers to connect directly with buyers, creating a practical digital marketplace.Users with minimal technical knowledge were able to interact with the platform smoothly, and the combination of Python ML scripts**,** PHP backend, and a MySQL database ensured fast, accurate, and user-friendly performance—making AgroNestor viable for real-world farming scenarios.

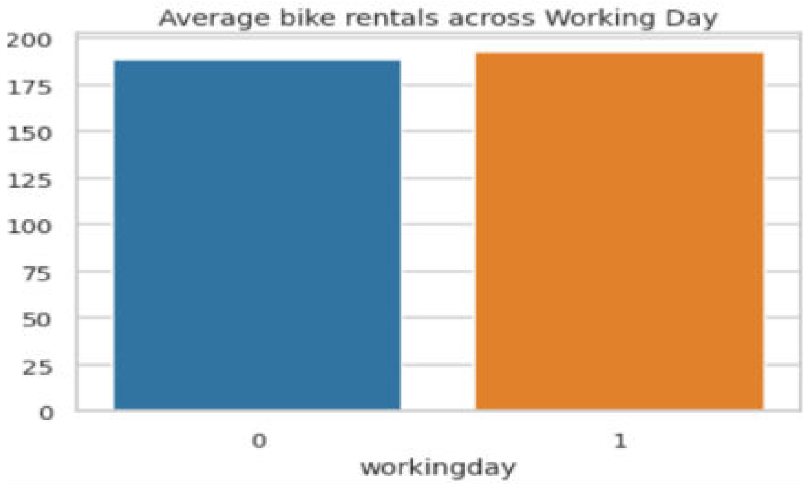


Figure 5: Average bike rentals across Working day

VIII. CONCLUSION

The research successfully developed a predictive model for hourly bike rental demand using historical, weather, and temporal data. Among the tested algorithms, Linear Regression delivered the highest accuracy (100%), followed by Decision Tree Regressor (99%) and SVR (90%). The findings highlight that demand is strongly influenced by time of day, weather conditions, and working day status. Accurate forecasting enables operators to optimize bike allocation, minimize shortages, and enhance user satisfaction. The study confirms the effectiveness of data preprocessing and feature engineering in improving model performance and demonstrates that machine learning can significantly support decision-making in bike-sharing operations.

REFERENCES

1. IJISET - International Journal of Innovative Science, Engineering & Technology- Bike share demand prediction.
2. J. Larsen. (April 25, 2013). Plan B Updates - 112: Bike-Sharing Programs Hit the Streets in Over 500 Cities Worldwide.
3. P. DeMaio, "Bike-sharing: History, Impacts, Models of Provision, and Future," Journal of Public Transportation, vol. 12, no. 4, p. 3, 2009.
4. S. B. KOTSIANTIS, "Supervised Machine Learning: A Review of Classification Techniques.," in Emerging Artificial Intelligence Applications in Computer Engineering: Real World AI systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies, vol. 160, K. k. ilias Maglogiannis, Manolis Wallace, John Soldatos, Ed. 1 ed. (Frontiers in Artificial Intelligence and Applications, Amsterdam: IOS Press, 2007, pp. 3-24.
5. Abdelhalim, A., & Traore, I. (2009). A new method for learning decision trees from rules. In Machine Learning and Applications, 2009. ICMLA 09. International Conference on (pp. 693–698).
6. Alippi, C., & Roveri, M. (2010). Virtual k-fold cross validation: An effective method for accuracy assessment. In The 2010 International Joint Conference on Neural Networks (IJCNN)
7. Jing, C., & Zhao, Z. (2015). Research on Antecedents and Consequences of Factors Affecting the Bike Sharing System—Lessons From Capital Bike Share Program in Washington, DC. In the International Conference on Logistics Engineering, Management and Computer Science (LEMCS 2015). Atlantis Press.
8. Joelsson, S. R., Benediktsson, J. A., & Sveinsson, J. R. (2005). Random forest classifiers for hyperspectral data. In Geoscience and Remote Sensing Symposium, 2005. IGARSS 05. Proceedings. 2005 IEEE International (Vol. 1, p. 4–pp). IEEE.
9. Larsen, J. (2013). Bike-sharing programs hit the streets in over 500 cities worldwide. Earth Policy Institute, 25.
10. Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection, 4, 40–79.
11. Bhattacharya, B., Price, R. K., & Solomatine, D. P. (2007). Machine Learning Approach to Modeling Sediment Transport. Journal of Hydraulic Engineering, 133(4), 440–450.
12. Fishman, E.K. (2016), "Bikeshare: A Review of Recent Literature"
13. Hamann, T.K. and Güldenberg, S. (2018), "Overshare and Collapse. How Sustainable are ProfitOriented Company-to-Peer Bike-Sharing Systems?"
14. Raviv, T., & Kolka, O. (2013). Optimal inventory management of a bike-sharing station. IIE Transactions, 45(10), 1077–1093