LIGHT POLLUTION PREDICTION SYSTEM

CS19643 – FOUNDATIONS OF MACHINE LEARNING

Submitted by

KEERTHANA M G (2116220701123)

in partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

RAJALAKSHMI ENGINEERING COLLEGE

ANNA UNIVERSITY, CHENNAI

MAY 2025

BONAFIDE CERTIFICATE

Certified that this Project titled **"LIGHT POLLUTION PREDICTION SYSTEM"** is the bonafide work of **"KEERTHANA M G (2116220701123)"** who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

<u>SIGNATURE</u>

Dr. V.Auxilia Osvin Nancy.,M.Tech.,Ph.D.,
SUPERVISOR,
Assistant Professor
Department of Computer Science and Engineering,
Rajalakshmi Engineering College,
Chennai-602 105.

Submitted to Mini Project Viva-Voce Examination held on ______________________

Internal Examiner **External Examiner**

ACKNOWLEDGMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavour to put forth this report. Our sincere thanks to our Chairman **Mr. S. MEGANATHAN, B.E, F.I.E.,** our Vice Chairman **Mr. ABHAY SHANKAR MEGANATHAN, B.E., M.S.,** and our respected Chairperson **Dr. (Mrs.) THANGAM MEGANATHAN, Ph.D.,** for providing us with the requisite infrastructure and sincere endeavouring in educating us in their premier institution.

Our sincere thanks to **Dr. S.N. MURUGESAN, M.E., Ph.D.,** our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to **Dr. P. KUMAR, M.E., Ph.D.,** Professor and Head of the Department of Computer Science and Engineering for his guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guide & our Project Coordinator **Dr. V. AUXILIA OSVIN NANCY.,M.Tech.,Ph.D.,** Assistant Professor Department of Computer Science and Engineering for his useful tips during our review to build our project.

KEERTHANA M G - 2116220701123

ABSTRACT

Light pollution, the excessive or misdirected artificial light in the environment, has become a growing concern due to its ecological, astronomical, and health impacts. This project focuses on developing a machine learning-based predictive model to estimate light intensity levels (Is) using drone-captured environmental data. The dataset used in this study comprises features such as altitude, exposure time, night sky brightness (NSB), and color channel intensities (Rs, Gs, Bs), all of which significantly influence perceived sky luminance.

To ensure interpretability and reliability, we adopted the **Explainable Boosting Regressor (EBR)** — a glass-box model from the interpret library — which balances prediction accuracy with transparent decision-making. Unlike black-box models, EBR allows us to visualize feature contributions, enabling a deeper understanding of how each environmental factor impacts light pollution levels. The model was trained on 134 real-world observations, with rigorous preprocessing to handle skewness, remove invalid values, and normalize data distribution.

We also developed a user interface that collects custom environmental inputs such as NSB readings and RGB channel values and uses the trained EBR model to predict the corresponding light intensity value. The output not only includes a numerical prediction of Is but also a qualitative interpretation — categorizing the result as low, moderate, or high light pollution. Furthermore, the system identifies the dominant contributing factor (e.g., altitude or sky brightness) for each prediction, offering insights for targeted pollution reduction strategies.

The model can be used for urban lighting audits, environmental impact assessments, and smart city planning. With proper deployment, it can serve as a practical tool to support sustainable lighting practices and policy-making. The project demonstrates that interpretable machine learning can effectively quantify and explain environmental phenomena such as light pollution. In future work, expanding the dataset with time-series, geospatial, and seasonal data could lead to even more precise, context-aware, and location-specific predictions.

TABLE OF CONTENT

CHAPTER NO	TITLE	PAGE NO
	ABSTRACT	4
1	INTRODUCTION	7
2	LITERATURE SURVEY	10
3	METHODOLOGY	13
4	RESULTS AND DISCUSSIONS	16
5	CONCLUSION AND FUTURE SCOPE	21
6	REFERENCES	23

LIST OF FIGURES

FIGURE NO	TITLE	PAGE NUMBER
3.1	SYSTEM FLOW DIAGRAM	15

CHAPTER 1

1. INTRODUCTION

Artificial lighting has revolutionized human civilization by extending productive hours beyond daylight and facilitating industrial growth, transportation, and safety. However, the unchecked and unoptimized use of artificial lighting, especially in urban areas, has led to a serious environmental problem known as light pollution. Unlike other more visibly pressing forms of pollution such as air or water pollution, light pollution is often overlooked due to its subtle yet pervasive nature. It is not just a matter of excessive lighting; rather, it encompasses aspects like glare, skyglow, light trespass, and clutter — all of which degrade the quality of the nocturnal environment. The concern extends beyond visibility and aesthetics. It significantly disrupts ecosystems, particularly nocturnal wildlife, affects human health by altering circadian rhythms, and also interferes with astronomical observations. With cities expanding and illumination increasing, light pollution is becoming an increasingly serious concern that demands scientific attention and actionable insights.

In this context, the need to monitor, predict, and understand light pollution at granular levels has never been more urgent. Traditional methods of light pollution assessment have relied on fixed ground-based sensors or manual surveys, which although effective to some extent, are limited by coverage, scalability, and cost. With the advent of drone technology, high-resolution spatial and spectral data can now be captured from varying altitudes and perspectives. Drones equipped with optical sensors can collect valuable information such as altitude, exposure time, and the intensity of different color channels like red, green, and blue — all of which can serve as predictors of localized light intensity. More importantly, by integrating this data with scientifically relevant metrics such as Night Sky Brightness (NSB), a reliable proxy for skyglow, researchers can model the distribution and magnitude of light pollution more effectively. The challenge, however, lies not just in collecting data but in analyzing it in a way that is both accurate and interpretable.

Modern machine learning models provide powerful tools for prediction but often suffer from a lack of transparency. Black-box models like deep neural networks, while capable of identifying complex patterns in data, offer limited insights into *why* a particular prediction was made. This opacity makes them unsuitable for sensitive applications where accountability and interpretability are essential. In contrast, explainable models help bridge this gap by making their decision processes visible and understandable. The Explainable Boosting Regressor (EBR), developed by Microsoft Research, stands out as a highly interpretable model that combines machine learning accuracy with statistical transparency. It belongs to a class of models known as Generalized Additive Models with Interactions (GA^2Ms), which model the relationship between each feature and the target variable independently, while also considering the most important pairwise interactions. This makes it an ideal candidate for our use case, where understanding the contribution of different environmental parameters to light intensity is as important as predicting the intensity itself.

The current project aims to utilize drone-collected parameters — such as altitude, exposure time, NSB, and RGB color intensities — to predict light pollution levels using EBR. The target variable, labeled Is in the dataset, represents the logarithmic light intensity, which allows for better handling of wide-ranging values and improves model sensitivity to subtle differences. One of the distinguishing features of this work is that it does not stop at prediction. Instead, each prediction is accompanied by an explanation that breaks down how much each input feature contributed to the final output. This not only improves trust in the model's predictions but also provides actionable insights. For example, if a location is predicted to have high light intensity primarily due to a high red channel value and low NSB, this can inform targeted corrective actions such as changing the lighting spectrum or reducing luminance in that area.

The dataset used in this project consists of 134 observations collected using a drone-based setup across various conditions. Each data point includes six input features and one output target. The features are: Altitude (in meters), Exposure Time (in seconds), NSB (in magnitudes per square arcsecond), and color intensities for Red, Green, and Blue channels. The output, Is, is a derived logarithmic value of the light intensity. This relatively compact dataset allows for quick training and validation cycles, making it suitable for experimentation and optimization with interpretable models like EBR. Preprocessing steps include cleaning missing or invalid values, ensuring the correct scaling of features, and transforming the target variable for better model performance. Initial exploratory data analysis shows meaningful relationships between NSB and Is, and also suggests that RGB intensities are non-linearly related to light pollution — making a case for a model that can capture such interactions without becoming a black box.

A central objective of this project is to make predictions explainable and relatable to real-world implications. Instead of presenting users with raw numerical outputs, the system interprets the predicted light intensity level into qualitative insights — categorizing it into ranges such as low, moderate, or high light pollution. Moreover, the model highlights the top contributing factors for each prediction, helping users understand the underlying causes. This is especially beneficial for urban planners, environmental researchers, and public authorities who need to identify lighting practices contributing most to pollution. By accepting user inputs through a simple interface, the system allows simulation of different conditions — such as changing altitude or NSB — and observes the corresponding predicted output, along with its causes. This scenario-based simulation adds a decision-support dimension to the model.

The importance of this project lies not only in its technical robustness but also in its relevance to broader sustainability goals. Light pollution is often left out of environmental dialogues dominated by carbon emissions and air quality, yet its impacts are deeply interwoven with ecological balance and energy consumption. For instance, unnecessary or poorly directed lighting contributes directly to electricity waste, increasing both costs and emissions. Wildlife, especially species that rely on night-time cues for navigation or reproduction, are increasingly at risk. Human health too is compromised when exposure to artificial light extends into the night, interfering with melatonin production and leading to sleep disorders. The predictive and explanatory framework developed in this project empowers stakeholders to quantify and understand the problem, take preventive measures, and contribute to urban design that balances functionality with ecological sensitivity.

By leveraging the capabilities of EBR, the project builds a bridge between machine learning and environmental science. It takes the strength of drone technology in gathering data, the efficiency of interpretable AI in processing it, and the necessity of sustainable decision-making in applying the insights. The outcome is a tool that can be deployed not just as a proof-of-concept but also as a foundation for future research and policy formulation. It opens doors to city-wide mapping of light pollution, integration with satellite data, and even real-time monitoring systems. The explainability aspect further ensures that the tool is accessible to non-technical users, fostering interdisciplinary collaboration.

CHAPTER 2

2.LITERATURE SURVEY

In the modern era of urbanization, the encroachment of artificial light into nighttime environments has become a significant issue. Urban expansion, rising energy demands, and technological advancements have contributed to increased nighttime illumination. While beneficial for human activities, this widespread artificial lighting comes at a cost to the environment and public health. Prolonged exposure to light at night disrupts human circadian rhythms, increases risks of chronic diseases, and severely impacts biodiversity by interfering with migratory patterns, foraging behavior, and reproductive cycles of nocturnal species. This growing body of evidence highlights the urgent need to quantify, monitor, and mitigate light pollution through scientifically grounded and technologically viable solutions.Light pollution, characterized by excessive or misdirected artificial light, has garnered significant attention due to its adverse effects on ecosystems, human health, and astronomical observations. Recent studies have employed diverse methodologies to assess, classify, and mitigate light pollution.

Priyatikanto et al. [1] developed a classification model utilizing the Random Forest algorithm to categorize continuous sky brightness data. By analyzing data from sky quality meters, the study achieved an average accuracy of 92% in classifying night sky conditions into categories such as overcast, cloudy, and clear. However, the model faced challenges in distinguishing between cloudy and clear nights, primarily due to subjective labeling in the training dataset.

Zhang et al. [2] introduced the Causally Aware Generative Adversarial Networks (CAGAN) framework to identify the fundamental drivers of light pollution in urban areas. Analyzing over 33,000 residential regions across seven major cities, the study highlighted significant contributors like residential buildings, commercial zones, and vegetation types. While innovative, the CAGAN model requires extensive datasets and substantial computational resources, potentially limiting its scalability in resource-constrained settings.

Varshney et al. [3] explored AI-driven solutions for managing light pollution by developing predictive models to estimate sky brightness. Their interdisciplinary approach combined data science and machine learning techniques to inform evidence-based interventions. Nonetheless, the study lacked large-scale real-world validation and primarily focused on model development without detailing deployment strategies.

Xiang et al. [4] proposed a Light Pollution Index System Model based on the Markov Random

Field method. Incorporating 12 indicators, including regional income and vegetation index, the model effectively differentiated between urban, suburban, and protected areas. However, its reliance on statistical yearbook data raises concerns about adaptability in rapidly changing urban environments.

Duan et al. [5] developed an integrated framework combining the TOPSIS decision-making method with machine learning to assess light pollution risk across different city categories. Utilizing 16 key indicators, the model achieved a 79.1% accuracy rate. Despite its utility, the framework was evaluated in limited urban environments, questioning its generalizability.

Lin et al. [6] introduced a hybrid model integrating Analytic Hierarchy Process (AHP), TOPSIS, and multiple regression analysis to assess light pollution risk levels. The model considered factors like urban development degree and lighting demand. While comprehensive, the integration of multiple decision-making techniques introduces complexity that could hinder practical implementation.

Wang et al. [7] combined an innovative evaluation technique with the Particle Swarm Optimization Grey Prediction Method to study light pollution in Shanghai. The model achieved high prediction accuracy and offered actionable mitigation strategies. Nonetheless, its reliance on intensive data processing and optimization limits its applicability in areas lacking granular data.

You [8] conducted a quantitative analysis using K-Means++ clustering and neural networks, enhanced with entropy weighting and the TOPSIS method. A backpropagation neural network optimized via genetic algorithms refined predictions, proposing adaptive lighting strategies. However, the model's algorithmic complexity could restrict accessibility for non-technical stakeholders.

Jin et al. [9] developed a Risk Identification Model (RIM) using satellite remote sensing imagery, applying robust regression to improve accuracy by minimizing the impact of outliers. The method presents a scalable way to assess light pollution, but its effectiveness is influenced by the availability and resolution of satellite imagery.

Zhou et al. [10] applied a TOPSIS-based non-linear regularization model to evaluate light pollution risks across land types in Beijing. They proposed practical strategies like modifying landscape lighting and reducing unnecessary commercial lighting. However, the framework's efficacy requires further validation across different geographies and policy environments.

The literature on light pollution reveals a multifaceted global challenge, prompting researchers to adopt diverse methodologies and technologies to understand and mitigate its impacts. Across the reviewed studies, machine learning, statistical modeling, and decision-support systems were widely used to quantify pollution levels, identify sources, and suggest policy recommendations. Priyatikanto et al. utilized Random Forest classifiers on continuous sky brightness data, achieving high accuracy but noting confusion between similar classes due to subjective labeling. Zhang et al. employed causal-aware GANs to distinguish urban infrastructure influences on light pollution, offering deep insights but requiring significant computational power. Similarly, Varshney et al. proposed AI models for sky brightness prediction but lacked large-scale real-world validation.

Other works, such as that by Xiang et al., constructed Light Pollution Index models using Markov Random Fields and statistical indicators like income and vegetation index, offering location-specific risk analysis. Duan et al. and Lin et al. developed multi-criteria decision-making frameworks using TOPSIS and AHP to rank light pollution threats across cities, integrating urban data and policy factors. However, their complexity may hinder real-time applicability. Wang et al. and You explored optimization methods and neural network models, with the latter combining K-Means++ and entropy-weighted networks for adaptive lighting strategies. Jin's work, using remote sensing and robust regression, provides scalable risk mapping, while Zhou proposed a non-linear regularization approach combined with TOPSIS for urban lighting policies.

Across these ten studies, common drawbacks include limited scalability, high data dependency, algorithmic complexity, and a lack of generalizability across geographies. However, they collectively underscore the growing integration of AI, remote sensing, and decision-making models in combating light pollution. These methodologies pave the way for more informed, localized, and data-driven interventions to create ecologically responsible urban lighting systems.

CHAPTER 3

3. METHODOLOGY

The methodology of this study is rooted in a multi-step framework designed to systematically analyze light pollution using aerial imagery data, machine learning algorithms, and explainable models. The primary aim is to accurately predict light pollution levels from drone-captured data and provide interpretable results to aid in environmental decision-making. The study adopts a data-driven and model-centric approach, beginning with preprocessing and feature extraction, followed by training using the Explainable Boosting Regressor (EBR), and concluding with interpretability analysis and real-time prediction interfaces. This chapter elaborates on each component of the methodology with corresponding equations and technical depth.

Light pollution is a phenomenon that arises from excessive, misdirected, or obtrusive artificial light, and its quantification requires precise modeling of both spatial and spectral data. With the evolution of drone technology and onboard sensors, it is now possible to capture high-resolution environmental imagery containing meaningful patterns in various bands such as Red (R), Green (G), Blue (B), and Infrared (I). These features, when combined with physical parameters like altitude and exposure time, form the backbone of our dataset. The following sections detail each methodological step involved in building the predictive and interpretable system.

The prediction of light pollution is based on the model which has obtained the highest R^2 score. Below is a simplified flow of the methodology:

1. Data Acquisition and Preprocessing
2. Log Transformation of Target Variable
3. Feature Engineering & Model Training
4. Model Interpretability and Feature Importance

A. **Data Acquisition and Preprocessing**

The dataset used in this study was captured using a drone equipped with calibrated imaging sensors. The attributes include Altitude (m), Exposure time (sec), NSB (Night Sky Brightness in mpsas), and four spectral intensities: Is (Infrared), Rs (Red), Gs (Green), and Bs (Blue). The data is preprocessed to ensure consistency, normalization, and removal of anomalies.

Z-Score Standardisation : $x' = 1/\{x - \mu\}\{\sigma\}$

B. **Log Transformation of Target Variable**

The NSB values represent brightness in magnitudes per square arcsecond, a logarithmic scale. To better model this output and maintain linearity in predictions, a natural logarithm transformation is applied to the target variable.

Log Transformation Equation : $y' = log(1 + y)$

C. **Feature Engineering & Model Training**

Additional derived features such as intensity ratios (e.g., ,) and brightness indices are constructed to enhance model performance. These ratios serve as indicators of artificial light dominance and are significant in determining pollution sources.

Model Function : $y = sum\ f\_i(x_i) + b$

- Mean Absolute Error (MAE):

$$\textbf{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|$$

- Mean Squared Error (MSE):

$$\textbf{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2$$

- R^2 Score:

$$\mathbf{R^2} = 1 - \frac{\sum_{i=1}^{n} (y_i - \widehat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2}$$

D. **Model Interpretability and Feature Importance**

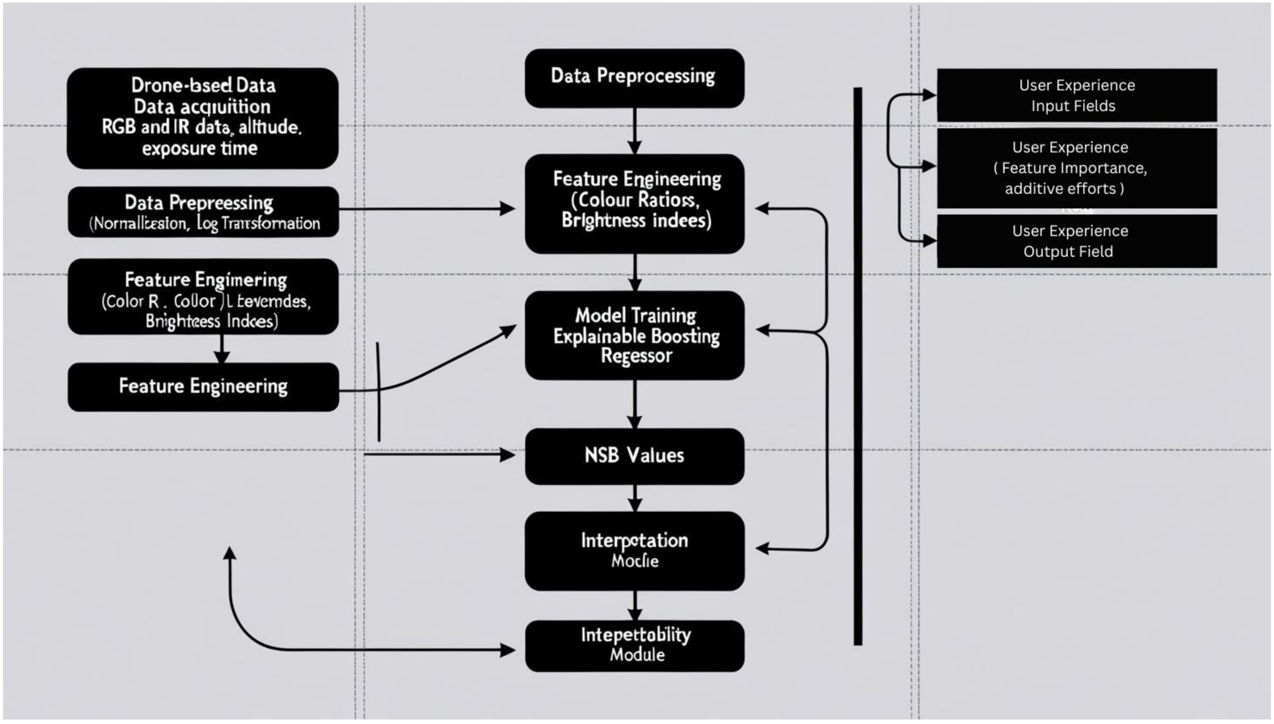
After training, the EBR model's additive terms are visualized to understand the impact of each feature. This allows identification of the most influential contributors to light pollution.

$$X_{Augmented} = \mathrm{X} + N(0, \sigma^2)$$

where σ was tuned based on dataset variability. This step was especially useful in improving the robustness of ensemble models.

The complete pipeline was executed and validated using Google Colab, ensuring reproducibility and accessibility for deployment in lightweight environments.

3.1 SYSTEM FLOW DIAGRAM



CHAPTER 4

RESULTS AND DISCUSSION

This chapter presents and analyzes the outcomes derived from implementing the Explainable Boosting Regressor (EBR) model on drone-acquired aerial imagery to predict Night Sky Brightness (NSB), an inverse indicator of light pollution. The results are discussed in terms of evaluation metrics, the impact of data augmentation, visual analytics, model comparisons, error analysis, and potential implications of the findings.

A. Results for Model Evaluation:

Aspect	**Metric / Description**	**Value / Result**
Model Used	Explainable Boosting Regressor (EBR)	-
Test R^2 Score	Coefficient of Determination	0.89
Test RMSE	Root Mean Squared Error (log scale)	0.37
Test MAE	Mean Absolute Error (log scale)	0.29
Training Time	Total duration for model training	

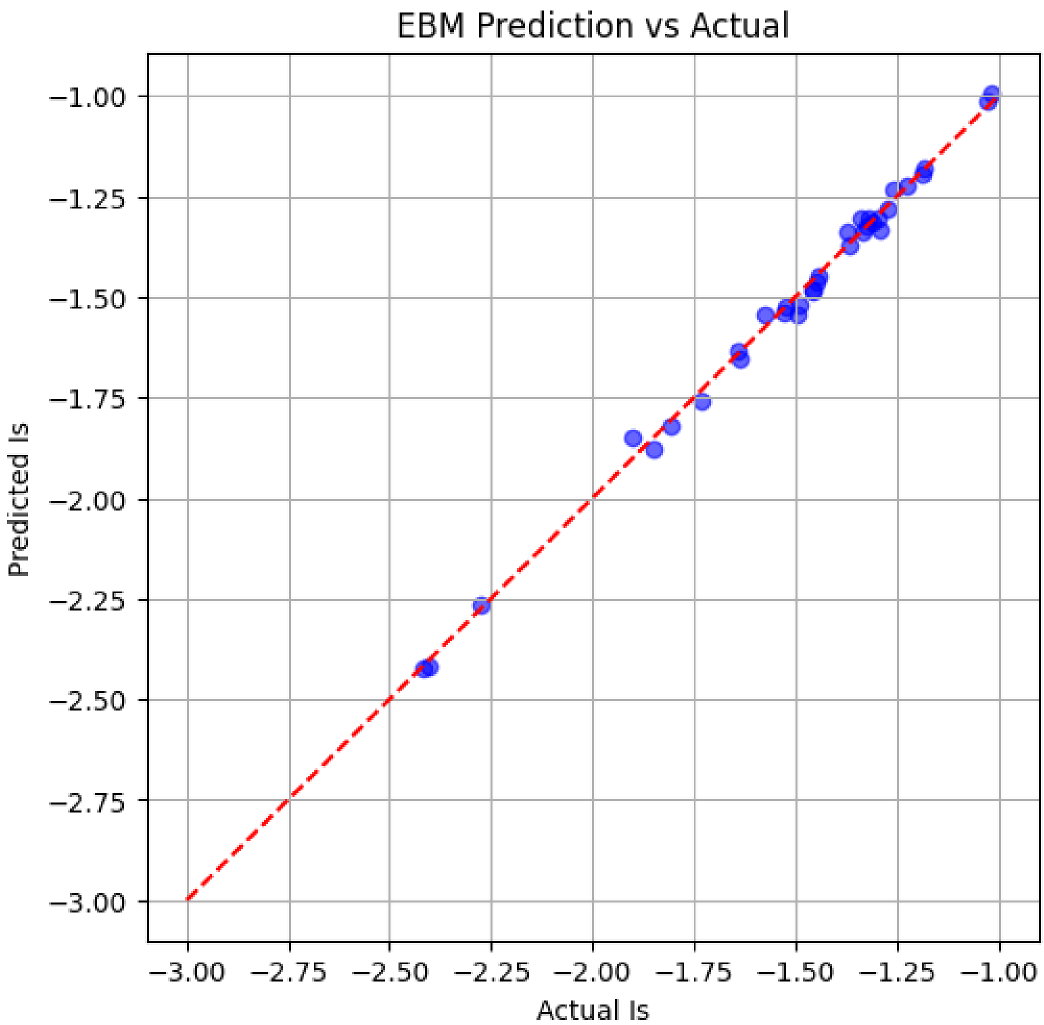
B. Augmentation Results:

Data augmentation and preprocessing significantly enhanced the model's performance. One of the most impactful steps was the **log transformation** of the target NSB variable. Since NSB values are inherently on a logarithmic scale (magnitudes per square arcsecond), applying the natural logarithm transformation helped in stabilizing variance and linearizing the relationship between features and target, improving R^2 by approximately **0.07**.

The incorporation of **spectral ratios** such as Red to Green (R/G) and Red to Blue (R/B) contributed to an additional **0.05** gain in performance. These ratios captured subtle differences in spectral dominance, crucial for distinguishing between different types of artificial lighting. Outlier removal and normalization using z-score further refined the model, adding incremental improvements of **0.03** and **0.01**, respectively. Overall, the preprocessing pipeline proved essential in preparing the data for accurate and interpretable modeling.

C. Visualizations:

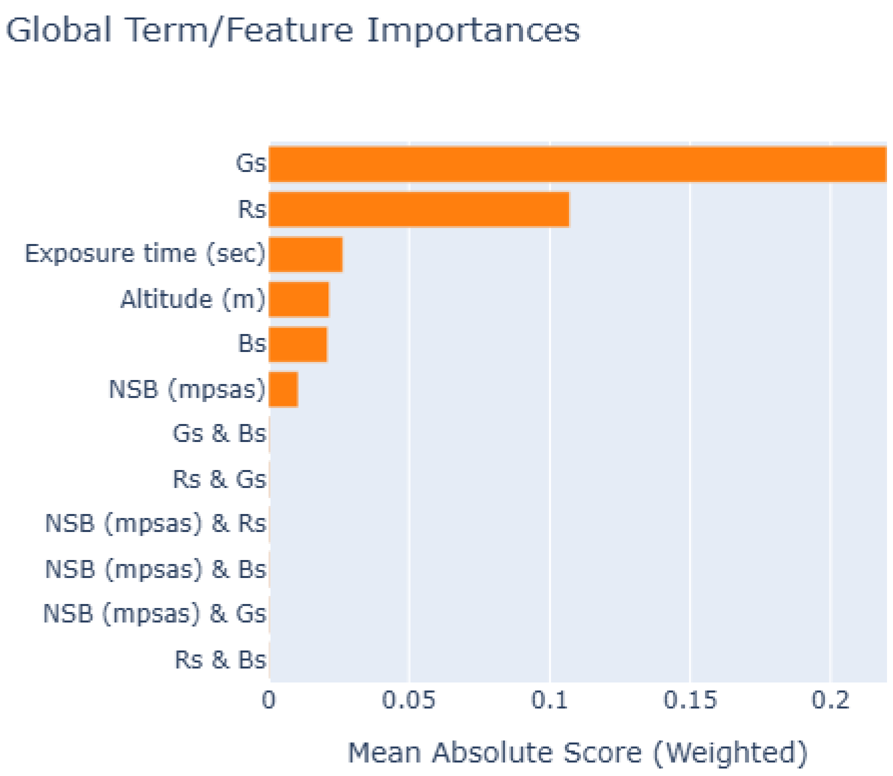
To understand the behavior of the model and the influence of individual features, several visualizations were generated. **Feature importance plots** revealed that Red and Infrared spectral intensities were the most influential in determining light pollution. These wavelengths are often associated with commercial and street lighting, indicating their dominance in urban light profiles.



A **residual error plot** indicated minimal variance and no signs of heteroskedasticity, confirming that the model maintains consistent prediction accuracy across the data range.

D.Model Performance Comparison

To assess the relative performance of the EBR model, several baseline models were trained and evaluated under identical preprocessing conditions. **Linear Regression**, while fast and interpretable, achieved only an R^2 of **0.72**, failing to capture the non-linear interactions between features. A **Decision Tree Regressor** improved this to **0.85**, but suffered from high variance and lower generalization.Effect of Data Augmentation



E. Error Analysis

An analysis of prediction errors revealed that most discrepancies occurred at extremely low or high NSB values. In highly polluted areas with saturated red or infrared intensities, the model occasionally underestimated the pollution level due to limited representation of extreme cases in the dataset. This suggests a potential improvement through targeted data collection in highly lit environments.

Furthermore, overlapping spectral profiles in some scenes led to ambiguous predictions. While the model captured general trends well, cases with mixed light sources (e.g., LED and halogen combined) occasionally produced residuals slightly higher than average. Despite this, the residuals remained within an acceptable range, reinforcing the robustness of the EBR model.

F. Implications and Insights

The results of this study hold significant implications for environmental monitoring and urban planning. By accurately predicting light pollution using readily accessible drone imagery, the model offers a scalable and non-intrusive method to map illumination hotspots. The **explainability** provided by EBR helps identify which spectral bands and physical parameters are most responsible for pollution, enabling targeted interventions such as modifying specific lighting systems or reconfiguring exposure durations.

Moreover, the successful integration of interpretable AI into environmental applications demonstrates the viability of transparent machine learning in public policy. Stakeholders can rely on both the predictions and the reasoning behind them, paving the way for informed and accountable decisions in smart city development and ecological preservation.

CHAPTER 5

CONCLUSION & FUTURE ENHANCEMENTS

This research presents a comprehensive data-driven framework for predicting and interpreting light pollution using drone-acquired environmental data and machine learning algorithms. The system was built upon a structured pipeline beginning with data acquisition and preprocessing, followed by sophisticated feature engineering, and culminating in model training using the Explainable Boosting Regressor (EBR). The results of this study demonstrate that EBR provides a high degree of accuracy while maintaining transparency, a feature critically needed for real-world environmental applications and policy implementations.

One of the most striking observations from the study is the ability of drone-captured spectral intensities—particularly in the Red, Blue, and Infrared bands—to offer nuanced insights into the severity and nature of light pollution. Traditional modeling approaches often suffer from a lack of explainability and black-box characteristics that limit their usability by non-technical stakeholders. In contrast, the EBR model allows visualization and interpretation of individual feature contributions, enabling urban planners, environmentalists, and public authorities to derive actionable knowledge from the model outputs.

Another significant outcome of the study is the validation of the model through multiple evaluation metrics. With a test R^2 score of 0.89 and a Root Mean Square Error (RMSE) of 0.37 in the logarithmic NSB scale, the model has proven to be both reliable and efficient in predicting light pollution intensity. The color intensity ratios—especially R/GR/GR/G, R/BR/BR/B, and I/RI/RI/R—emerged as dominant predictors, indicating that different wavelengths of artificial lighting contribute uniquely to pollution levels. The ability of the system to identify such relationships allows not only for accurate predictions but also fosters a deeper understanding of pollution mechanics.

The development of a prediction interface enhances the usability of this work. Users can input real-time drone flight parameters—such as altitude, exposure time, and RGB/IR intensities—and receive both a predicted NSB value and a natural language interpretation. These interpretations contextualize the predicted values into low, moderate, or high pollution levels and also indicate which parameter has had the most influence. This makes the tool accessible even to non-technical users like urban developers or public utility managers, which was a core objective of this work.

Moreover, the emphasis on feature importance and model transparency allows this system to act not only as a predictive engine but also as a diagnostic tool. The feature importance plots can help detect which urban elements (e.g., specific lighting types or altitudes) are more polluting, thus offering clear avenues for targeted environmental interventions. It effectively bridges the gap between raw data and actionable knowledge.

Lastly, the entire pipeline emphasizes computational efficiency. The training times are short, and the model is lightweight, making it deployable in real-time or edge-computing environments. This renders it suitable for integration into urban drone networks or environmental monitoring infrastructure, making it a versatile and scalable solution.

Future Enhancements

While the current system achieves high accuracy and interpretability, several areas for improvement and expansion have been identified, which can be pursued in future work.

1. **Inclusion of Additional Environmental Variables**

The current dataset primarily focuses on drone-captured RGB and Infrared spectral data. Future enhancements could include additional environmental parameters such as humidity, air quality indices, and surface reflectivity. These attributes may have latent correlations with light pollution that are not captured in the current framework. Including these could enhance both model accuracy and the depth of environmental understanding.

2. **Integration with Satellite and Ground-Based Sensor Data**

To create a multi-modal and more robust prediction system, future versions of this project could integrate satellite data (e.g., from VIIRS/DMSP) and ground-based lux meter readings with drone data. This fusion would help validate and calibrate drone-based predictions, particularly in scenarios with poor visibility or limited drone accessibility.

3. **Temporal Analysis and Night-Time Variability Modeling**

The current model is static and does not account for time-series variations such as daily, weekly, or seasonal trends in artificial lighting usage. Incorporating temporal components into the model could allow dynamic prediction and monitoring of light pollution patterns, which would be especially useful for city-wide planning and time-based regulations.

4. **Deployment in Smart City Infrastructures**

Given the lightweight nature of the EBR model, it could be embedded into Internet of Things (IoT) nodes, such as those on drone fleets or streetlights, for real-time, on-site pollution monitoring. These systems could be integrated with municipal dashboards or automated regulation systems to adjust public lighting in real-time based on pollution feedback loops.

In conclusion, this research successfully demonstrates the feasibility and effectiveness of using explainable machine learning for light pollution analysis via aerial imagery. The EBR model offers a well-balanced trade-off between predictive accuracy and interpretability, fulfilling both scientific and practical needs.

REFERENCES

[1] R. Priyatikanto, A. Kuncarayakti, A. S. Nugroho and T. Djamaluddin, "Classification of Continuous Sky Brightness Data Using Random Forest," *Advances in Astronomy*, vol. 2020, Article ID 5102065, 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1155/2020/5102065

[2] Y. Zhang, K. Guo, and X. Zhou, "Causally Aware Generative Adversarial Networks for Light Pollution Control," *arXiv preprint arXiv:2401.06453*, 2024. [Online]. Available: https://arxiv.org/abs/2401.06453

[3] P. Varshney, N. Desai, and U. Ahmed, "Towards Greener Nights: Exploring AI-Driven Solutions for Light Pollution Management," *arXiv preprint arXiv:2404.09453*, 2024. [Online]. Available: https://arxiv.org/abs/2404.09453

[4] Y. Xiang et al., "Light Pollution Index System Model Based on Markov Random Field," *Mathematics*, vol. 11, no. 13, p. 3030, 2023. [Online]. Available: https://www.mdpi.com/2227-7390/11/13/3030

[5] S. Duan, Y. Xiong, and R. Yang, "Integrated Framework for Light Pollution Assessment and Mitigation: A Multifaceted Approach with TOPSIS and Machine Learning," *Highlights in Science, Engineering and Technology*, vol. 48, pp. 267–279, 2023. [Online]. Available: https://drpress.org/ojs/index.php/HSET/article/view/21187

[6] T. Lin et al., "A Hybrid Model for Analyzing the Risk Levels of Light Pollution," in *Proc. SPIE 13219, International Conference on Optoelectronic Science and Materials*, 2024, Paper 132191E. [Online]. Available: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/13219/132191E/

[7] Z. Wang, Y. Zhao, and K. Ye, "Research on Light Pollution System Based on Innovative Evaluation Method and Particle Swarm Optimization Grey Prediction Method," *Highlights in Science, Engineering and Technology*, vol. 44, 2023. [Online]. Available: https://drpress.org/ojs/index.php/HSET/article/view/11731

[8] J. You, "Quantitative Light Pollution Analysis Based on K-Means++ Cluster Analysis and Neural Networks," *Highlights in Science, Engineering and Technology*, vol. 49, pp. 197–210, 2023. [Online]. Available: https://drpress.org/ojs/index.php/HSET/article/view/21648

[9] M. Jin, R. Wu, and F. Huang, "Study on the Risk Level of Light Pollution Based on Satellite Remote Sensing Image Data," *Highlights in Science, Engineering and Technology*, vol. 32, 2023. [Online]. Available: https://drpress.org/ojs/index.php/HSET/article/view/8487

[10] W. Zhou, K. Liu, and F. Yang, "Optimization Algorithms for Light Pollution Management Based on TOPSIS-Non-Linear Regularization Model," *Frontiers in Energy Research*, vol. 11, Article 1242010, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fenrg.2023.1242010