Comparative Analysis of Light Pollution Impacts on Moths Using Machine Learning and XAI

No Author Given

No Institute Given

Abstract. Light pollution is an increasingly prevalent issue that has a harmful impact on the health of human beings, animals and plants. Light pollution causes a disruption that alters the habitats of nocturnally active species. Because nocturnal insects are lured to artificial light sources, the consequences can be fatal. Moths are nocturnal insects that have an abnormal sensitivity to artificial light. Artificial night exposure has been identified as one of the main causes of the fast decrease in moth populations. Artificial night lighting causes moths to invest shorter periods of feeding than moths in the dark. Therefore, artificial lighting shortens feeding periods, heightens sublethal effects, and increases moth mortality. In this research, to analyze the effects of light pollution and for the observation of foraging by moths, different machine learning models have been used in the moth foraging data. To perform analyses, Decision Tree, Random Forest, KNN, XGBoost, and SVM have been used. SMOTE was applied to balance the unbalanced dataset, which considerably improved accuracy. All of the models performed excellently in classifying the foraging phase, with SVM above 95% accuracy and others over 99%. Explainable AI models - LIME and SHAP had also been applied to explain and analyze the models' predictions.

Keywords: Light pollution · Moths · Machine Learning.

1 Introduction

Artificial light is a type of man-made light source and is used for a variety of reasons. It can be turned on and off at will, and its brightness and color temperature can be changed to accommodate a variety of requirements and tastes [1]. ALAN (artificial light at night) is a significant source of anthropogenic pollution. However, artificial light is a growing environmental issue due to the proliferation of streetlights, business buildings, and residential properties that are illuminated at all hours of the day and night [2]. Light pollution has negative effects on the ecosystem, altering the cycles of life and posing a serious danger to biodiversity. The presence of light pollution has been found to disturb the existence of various animal species.

The appearance of artificial nightlight can have a disruptive effect on the behavior of many species, as they are adapted to survive and flourish in the dark. It is having harmful consequences for the animal and human kingdoms, including nocturnal and migratory species, flying creatures, and plants [3,4]. It has a deleterious influence on natural ecosystems, interfering with moth communication, feeding, migration, and reproduction [7].

Artificial light may disrupt adult moth foraging behavior, which might have serious effects on population dynamics. The feeding behavior of moths treated with artificial light is the topic of this study. Because moths, particularly bigger species, are particularly drawn to light that is relatively abundant in short-wavelength radiation, the use of longer-wavelength light is frequently advocated as a technique for moth conservation in lighted environments [6].

Previously, researchers used the hawkmoth's visual system as a model to explore the impact of different light sources on the visual ecology of nocturnal hawkmoths, as well as harmonic radar technology on hawk and lappet moths [8,9]. Longer and shorter-wavelength light (amber and reddish light) is used for moth conservation in lit situations. Researchers also conducted an experiment to investigate the impact of this conservation technique on the eating behavior of moths using green, white, red, and no artificial night lighting [5].

Despite several studies in this area, few have employed machine learning to characterize foraging behavior. This study will look at the feeding habits of adult moths exposed to artificial light. In this paper, we used different types of machine learning methods (Decision Tree, Random Forest, SVM, and KNN) to find out the foraging behavior of four distinct moth species. We applied statistical validation approaches to assess a model's performance. To empirically compare these five models and select the best one, we used the Confusion Matrix on the models. XAI is a popular analysis AI that uses AI-powered decision-making to characterize model accuracy, transparency, fairness, and outcomes. Interpretability methods like LIME can help select the most suitable models, and this report uses the LIME and SHAP frameworks to explain model performance.

The paper will be discussed as follows: In Section II, related works on the subject will be discussed. In Section III, the proposed system's methodology, including data set, preprocessing, and data splitting, will be discussed. The results will be evaluated in Section IV, and the discussion and conclusion will be discussed in Sections V and VI, respectively.

2 Related Works

This study [10] shows the lack of feeding behavior of male moths and the calling behavior of female moths due to the effects of artificial light at night. For this research, they chose Yponomeuta cagnagellus moths. The moths were chosen from both illuminated and dark locations and experimented with different levels of light intensity in the laboratory. It was found that male moths did not show any activity when feeding under the light. Again, the calling behavior of female moths was more visible in dark conditions than in dim or bright treatments. Moths from both locations were found to be affected by the ALAN.

In this research, the effects of various light types on these ecologically significant insects' visual ecology are tested, and they have used the hawkmoth's visual system as a model. They have evaluated how various artificial light emission spectra may impact nocturnal hawkmoths' capacity to engage in visually-driven behaviors such as flower selection for pollination, intraspecific communication, and predator-repelling actions. The most striking result is the prediction that broadband amber light sources will interfere with the hawkmoth's flower color perception. The moths may choose poor backdrop resting locations as a result of narrow-band sources' inability to perceive color contrasts in moth wings and in flowers and plants, making them more susceptible to nocturnal predators [8].

The findings of this research show for the first time through experimental research that streetlights fragment landscapes, and they also show that light pollution affects moth movement patterns beyond what was previously thought to be the case. This could potentially have an adverse effect on moth reproductive success and impair a critical ecosystem service. To capture individual flight trajectories of moths at previously unheard-of spatial and temporal resolution within a 1 km range, the authors of this study used harmonic radar technology on a number of nocturnal pollinators to understand how street lights affect movement behavior and orientation performance. They have used hawk moths and lappet moths for their experiment [9].

The researchers conducted a study in the Berlin-Brandenburg region 164 of Germany to examine how artificial light affects 165 the moth species Agrotis exclamationis. They focused on traits such as body length, eye size, and 167 forewing length and hypothesized that increasing light pollution would lead to a decrease in 169 these traits. The researchers used a "light pollution 170 map" to categorize artificial light at night (ALAN) 171 levels at different sites and years. They analyzed 172 moths collected over the past 137 years and found 173 that while there were no differences in traits across 174 different locations, there were changes over time that were dependent on sex and trait. They observed a trend of decreasing eye size in female 176 moths in areas with higher levels of light pollution, 177 indicating that the morphological traits of these 178 moths are being impacted by artificial light [11].

Table 1: Dataset-'Data Moth Foragin' Attributes

Attribute Name	**Attribute Explanation**
Day	Date of the experiment
Block	Refers to the block design
Treatment	Lamp type (Green, White or Red) or no lamp (Dark)
Species	Four moth species
Sex	Male versus female;
unique.id	Unique id of every individual, measured over approximately 8 hours
Obs	Observation of foraging by moths every six minutes, where 1 = foraging and 0 = not foraging
Size	Size classes of moths, where 1 = large (forewing length $<$ 15mm) and 0 = small (forewing length $\geq$ 15mm)

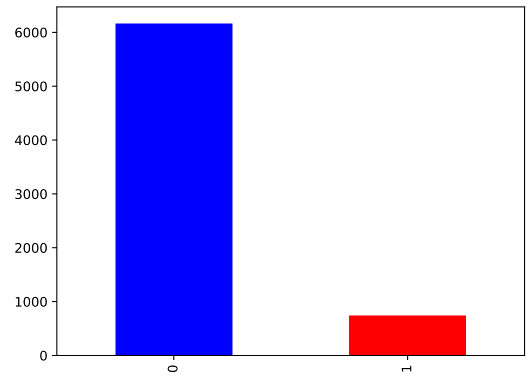

Fig. 1: ‘Obs’(Observation) classes before applying SMOTE

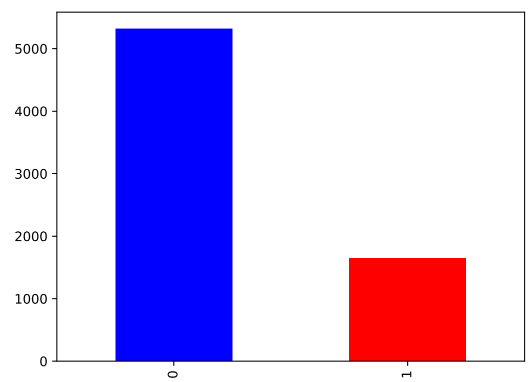

Fig. 2: ‘Obs’(Observation) classes after applying SMOTE

3 Methodology

3.1 Dataset

The dataset is referred to as ‘Data Moth Foraging’ [14]. It contains 6900 annotated data points and 8 characteristics of moth behavior when exposed to artificial nightlight (Table 1). In addition to moth ‘species’ and ‘sex’, the files also contain general information like ‘day’, ‘block’, and ‘treatment’ type. Data like foraging ‘Obs’(observations), moth 1size’, and ‘unique.id’ are included in the dataset. The portions were designated using the identifier ‘Day’.

We examine the dataset to check whether it includes any null values. There is no null value in the dataset. We utilized label encoding on the dataset to convert the columns from categorical (Boolean and Objective type) to numerical for the classifier and divided the Dataset into Training (70%) and Test (30%) data. Finally, the dataset was ready to be used to train a model.

The imbalance within the dataset is depicted in Figure 1. The 'Obs' 1 was difficult to categorize, and it did not work well. We used a variation of the SMOTE technique called SMOTE-ENN (Edited Nearest Neighbour), which combines both oversampling and undersampling, to balance the dataset. While the ENN performed the undersampling, the SMOTE performed the oversampling [15].

3.2 Decision Tree

A decision tree classifier predicts the value of a target class given an unknown test case. It is simple and easy to implement. It can be expanded into two types: classification trees with symbolic class labels and regression trees with numerically valued class labels. In our model, we used DT to predict the value of observation of foraging by moths [12].

3.3 Random Forest

Randomness is essential to the operation of a Random Forest (RF) algorithm. By randomly picking a portion of characteristics and data points from the training set, the method generates several decision trees. This randomization aids in reducing overfitting and improving model generalization. Furthermore, each decision tree in the forest is trained on a separate sample of the data, increasing the model's variety and resilience. The random forest aggregates the output of all decision trees to generate a final forecast, which is typically more accurate and less prone to mistakes than a single decision tree [12].

3.4 SVM

Support vector machines (SVM) is supervised learning models. The usual class of objects is described using training data, while the other items are regarded as anomalies. The classifier created by SVM divides the input space into a limited area that contains normal items, while the remaining space is presumed to include anomalies [13].

3.5 KNN

K-Nearest Neighbor (KNN) is a nonparametric approach to classifying samples by finding the k number of samples closest to the test sample and assigning the most frequent class label. It is an instance-based learner [12].

3.6 XGBoost

Gradient-boosted decision trees (XGBoost) have been presented as a new ensemble tree-based approach for improving CART decision tree performance. It optimizes the tree split against a given loss function using a gradient-boosting method. It has received a lot of interest from the Kaggle community as well as applications in real life. Federated learning (FL) has been proposed to allow numerous parties to collaborate on ML model training without requiring data to be sent to a central place. This method has been effective in cases where customer data privacy is critical and the law restricts data mobility [13].

3.7 Explainable AI

Explainable AI (XAI) is a popular analysis AI that is known for being honest and transparent. It uses AI-powered decision-making to characterize model accuracy, transparency, fairness, and outcomes. Interpretability methods like Local Interpretable Model Agnostic Explanation (LIME) can help select the most suitable models. This report uses the LIME and SHAP frameworks of XAI to explain the model's performance by calculating the contribution of each feature to the prediction [16].

4 Results

Table 2: Evaluation Metrics Results Before Applying SMOTE

	Precision		**Recall**		**F1-score**		**Accuracy**
Models	Obs = 0	Obs = 1	Obs = 0	Obs = 1	Obs = 0	Obs = 1	
Decision Tree	0.94	0.75	0.98	0.52	0.96	0.61	0.926
Random Forest	0.94	0.75	0.98	0.52	0.96	0.61	0.9266
SVM	0.89	0.00	1.00	0.00	0.94	0.00	0.888
KNN	0.94	0.61	0.96	0.52	0.95	0.56	0.9097
XGBoost	0.94	0.75	0.98	0.52	0.96	0.61	0.927

The performance of a model is evaluated and understood using evaluation metrics. This also shows how effectively the models are working. Additionally, evaluation metrics are crucial for comparing various models. Some of the metrics we have used for evaluating our models include Confusion Matrix, Precision, Recall, F1-Score and Accuracy. Figure 3 and Figure 4 are the Confusion Metrics before and after applying SMOTE respectively.Table 2 and Table 3 illustrate how the models are overfitted before applying SMOTE by analyzing the Precision, Recall and F1-score of the models.

The performance of the models after SMOTE application is summarized in Table 3. It demonstrates that every model provides very good accuracy. For

Table 3: Evaluation Metrics Results After Applying SMOTE

	Precision		**Recall**		**F1-score**		**Accuracy**
Models	Obs = 0	Obs = 1	Obs = 0	Obs = 1	Obs = 0	Obs = 1	
Decision Tree	1.00	1.00	1.00	1.00	1.00	1.00	0.999
Random Forest	1.00	1.00	1.00	1.00	1.00	1.00	0.999
SVM	0.98	0.90	0.97	0.92	0.97	0.91	0.956
KNN	1.00	1.00	1.00	0.99	1.00	0.99	0.997
XGBoost	1.00	1.00	1.00	0.99	1.00	1.00	0.998

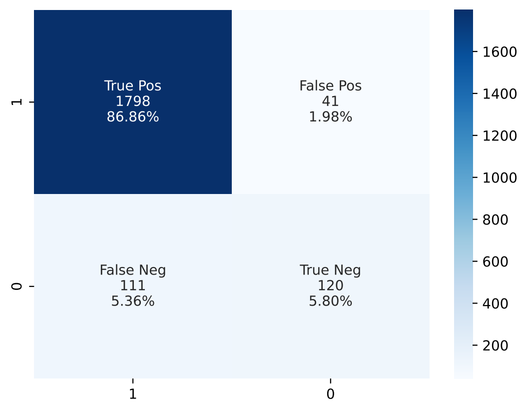

Fig. 3: Confusion Matrix of Decision Tree before applying SMOTE

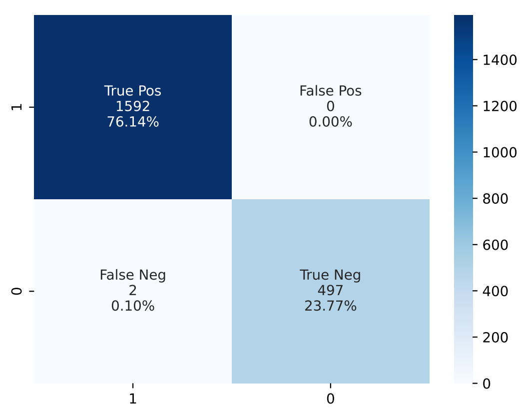

Fig. 4: Confusion Matrix of Decision Tree after applying SMOTE

precision, recall, and f1-score of 1.00, Decision Tree and Random Forest provide an accuracy of 99.9%. The accuracy rates provided by KNN and XGBoost are 99.7% and 99.8%, respectively. Only SVM provides an accuracy that is 95.6%,

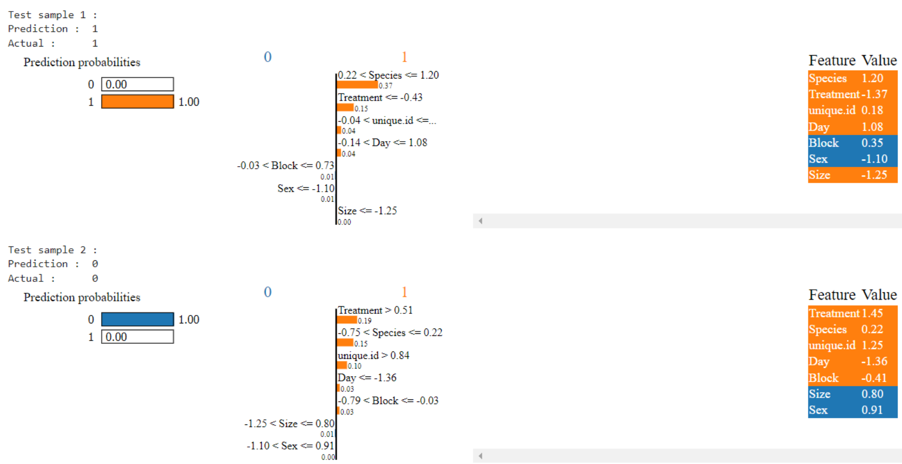

Fig. 5: LIME explanation of Decision Tree model

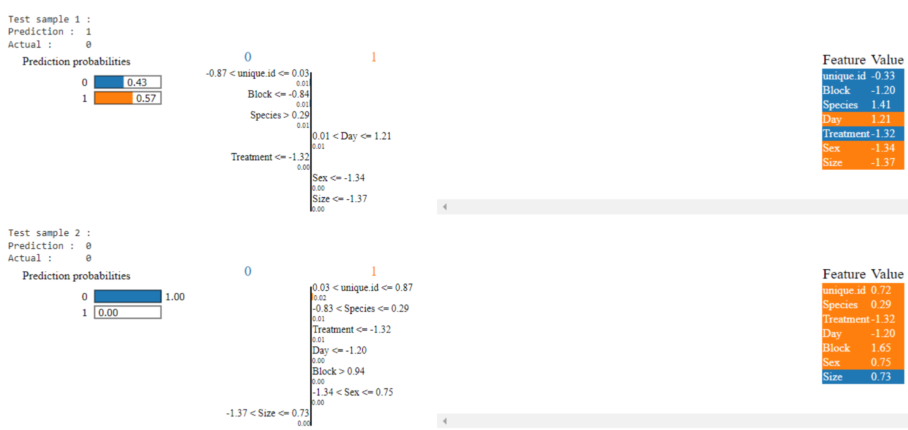

Fig. 6: LIME explanation of XGBoost model

which is comparably lower than that of other models. Although the SVM model's accuracy, recall, and f1-score are 0.98, 0.97, and 0.97 for 'Obs' = 0, they are comparatively lower for 'Obs' = 1.

5 Discussion

The Decision Tree model of the two test samples is explained by LIME in Figure 5. Essentially, the LIME explanation combines a prediction probability with a feature value table where each attribute is assigned a value and a weight for each prediction. The model is absolutely confident that it can predict the 'Obs' class 1 for test sample 1. We can see from the feature value table that the

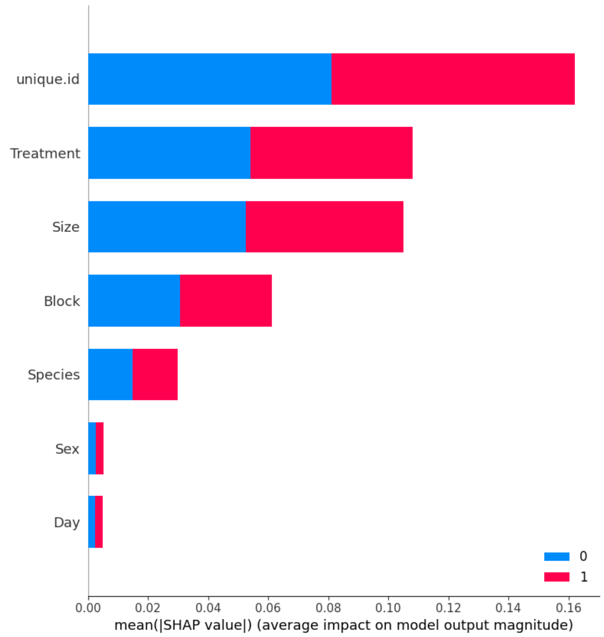

Fig. 7: SHAP Summary Bar plot for Decision Tree model

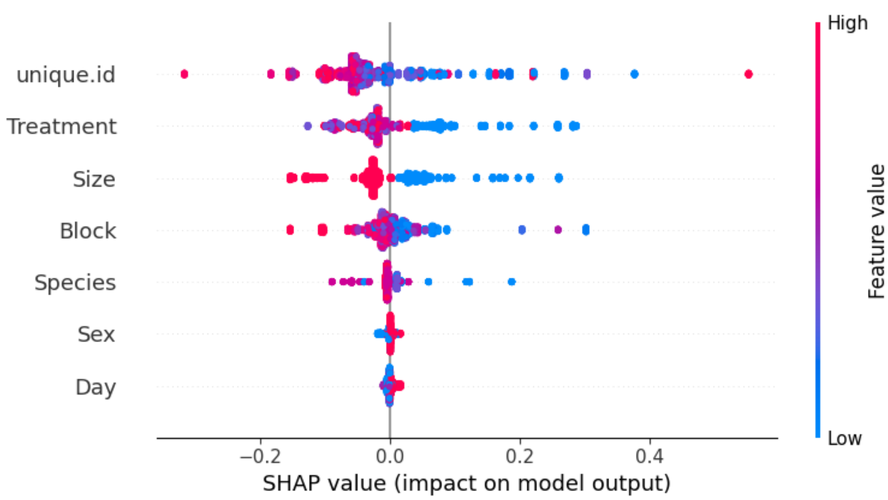

Fig. 8: SHAP summary plot for Decision Tree

features 'Treatment', 'Species', 'unique.id', 'Day', and 'Size' support class 1 in the prediction because their feature value is below or within the threshold. Here, the feature value for the 'Species' feature is 1.20, which is between 0.22 and

1.20. Thus, class 1 is supported by the 'Species' feature. 'Size' and 'Sex' features support class 0 for test sample 2.

Figure 6 illustrates that test sample 1 provides an incorrect LIME prediction for the XGBoost Model. Although the prediction has a confidence level of 57% and supports class 1, the actual result is class 0. 'Day', 'Sex', and 'Size' features are supported for Class 1. Only the Size attribute is supported for the Obs class 0 in Test Sample 2.

The SHAP bar summary plot for the Decision Tree Algorithm is shown in Figure 7. An overview of the important features of the global model is provided in this plot. Here, Blue denotes the 'Obs' class 0, and Red denotes the 'Obs' class 1. The X-axis displays the mean absolute values, which illustrate the average impact on the model output magnitude, and the Y axis displays the contribution of the features to decision-making. From this plot, it is clear that 'unique.id', 'Treatment', 'Size', 'Block', and 'Species' are the most significant features. 'Sex' and 'Day', however, are less significant features for the model prediction.

Figure 8 shows the summary plot for the decision tree model. SHAP summary graphs provide an overview of feature importance and the factors influencing it. Here all the dots are the data points. The color indicates how significant that attribute was for that dataset row. A prediction's greater or lower value is indicated by the value's horizontal placement. As can be seen from the figure, 'unique.id' is the property that matters the most, while 'Day' is the least important. The features 'Sex' and 'Day' contribute the least to the prediction. When the feature value decreases for the 'unique.id', 'Treatment', 'Size', and 'Block' features, the positive impact of these characteristics on the prediction increases. The 'unique.id' feature, however, is an extreme example where a high feature value has a positive impact.

6 Conclusion

Light pollution is a problem that is getting worse and is hazardous to the health of people, animals, and plants. The disruption brought on by light pollution changes the habitats of creatures that are active at night. Moths are nocturnal insects with an abnormal sensitivity to artificial light. Artificial night exposure has been recognized as one of the primary drivers of the rapid decline in moth populations. Multiple machine learning models are applied to moth foraging data to examine the effects of light pollution and to observe moth foraging. Analyses are carried out using Decision Tree, Random Forest, KNN, SVM and XGBoost algorithms. The imbalanced dataset is balanced using SMOTE, which significantly increased accuracy. All of the models have performed well in classifying the foraging phase, with SVM above 95% accuracy and others over 99% accuracy. Explainable AI models such as LIME and SHAP are also used to explain the models' performance. LIME gives a local explanation of a particular data instance with the prediction probability, feature value table and weight of the features on the prediction. SHAP Summary and summary bar plot has been used to explain for global model prediction.

References

1. Vedantu. (n.d.). Light Sources. VEDANTU. https://www.vedantu.com/physics/light-sources
2. International Dark-Sky Association. (2017, February 14). Light Pollution - International Dark-Sky Association. https://www.darksky.org/light-pollution/ (video)
3. Altermatt, F., Ebert, D. (2016). Reduced flight-to-light behaviour of moth populations exposed to long-term urban light pollution. Biology Letters, 12(4), 20160111. https://doi.org/10.1098/rsbl.2016.0111 (Journal)
4. Rajkhowa, R. (2014). Light pollution and impact of light pollution. International Journal of Science and Research (IJSR), 3(10), 861-867. (journal)
5. Frank, Kenneth. (1991). Impact of Outdoor Lighting on Moths. International Astronomical Union Colloquium. 112. 51. 10.1017/S0252921100003687. (Journal)
6. Van Langevelde, F., Van Grunsven, R. H. A., Veenendaal, E., Fijen, T. P. (2017). Artificial night lighting inhibits feeding in moths. Biology Letters, 13(3), 20160874. https://doi.org/10.1098/rsbl.2016.0874 (Research Article)
7. Longcore, T., amp; Rich, C. (2004). Ecological light pollution. Frontiers in Ecology and the Environment, 2(4), 191–198. https://doi.org/10.1890/1540-9295(2004)002[0191:elp]2.0.co;2
8. Briolat, E.S., Gaston, K.J., Bennie, J. et al. Artificial nighttime lighting impacts visual ecology links between flowers, pollinators and predators. Nat Commun 12, 4163 (2021). https://doi.org/10.1038/s41467-021-24394-0 (article)
9. Degen, Jacqueline Storms, Mona Lee, Chengfa Jechow, Andreas Stöckl, Anna Hölker, Franz Jakhar, Aryan Walter, Thomas Walter, Stefan Mitesser, Oliver Hovestadt, Thomas Degen, Tobias. (2022). Streetlights affect moth orientation beyond flight-to-light behaviour. 10.1101/2022.10.06.511092. (research article)
10. Cieraad, E., Van Grunsven, R. H. A., Sman, F., Zwart, N., Musters, C., Strange, E. F., Van Langevelde, F., Trimbos, K. B. (2022). Lack of local adaptation of feeding and calling behaviours by Yponomeuta cagnagellus moths in response to artificial light at night. Insect Conservation and Diversity, 15(4), 445–452. https://doi.org/10.1111/icad.12568 (article)
11. Keinath, S., Hölker, F., Müller, J., amp; Rödel, M.-O. (2021). Impact of light pollution on moth morphology–a 137-year study in Germany. Basic and Applied Ecology, 56, 1–10. https://doi.org/10.1016/j.baae.2021.05.004
12. Chih-Fong Tsai, Yu-Feng Hsu, Chia-Ying Lin, Wei-Yang Lin, Intrusion detection by machine learning: A review, Expert Systems with Applications, Volume 36, Issue 10, 2009, Pages 11994-12000, ISSN 0957-4174,https://doi.org/10.1016/j.eswa.2009.05.029. (https://www.sciencedirect.com/science/article/pii/S0957417409004801) (journal)
13. Ong, Y. J., Zhou, Y., Baracaldo, N., Ludwig, H. (2020). Adaptive histogram-based gradient boosted trees for federated learning. arXiv preprint arXiv:2012.06670.
14. Van Langevelde, F., van Grunsven, R. H. A., Veenendaal, E. M., Fijen, T. P. M., Fijen, T. P. M., & van Grunsven, R. H. A. (2017, February 8). Data from: Artificial night lighting inhibits feeding in moths. Zenodo. Retrieved April 29, 2023, from https://zenodo.org/record/5015437.ZEyWWXZBxPa
15. 7 smote variations for oversampling. KDnuggets. (n.d.). Retrieved April 29, 2023, from https://www.kdnuggets.com/2023/01/7-smote-variations-oversampling.html
16. Ribeiro, M. T. (2016, February 16). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. arXiv.org. https://arxiv.org/abs/1602.04938