**MORE EFFICIENT USE OF PUBLIC LIGHTING WITH**

**DATA ANALYTICS BASED ON MACHINE LEARNING**

**Abstract**

Street lightings are less important in the general perspective of citizens. However, our research results found and explained in this paper state that it is not the case. Even if street lightning may be underestimated, they are found effective in analyzing the energy consumption of potentially intelligent cities. Their contribution to the economy might be enormous when these anomalies are resolved. Pattern detection is significant in understanding and assisting in avoiding anomalies by obtaining consumption data properly. Researchers are studying pattern detection and anomaly detection in specific lightings such as buildings and apartments, whereas those studying street lightings are fewer than the city as a whole. In addition to multiple application areas, there are various methodologies, such as the internet of things or the calculation of time series for this purpose. This paper discusses the results of implementing a comprehensive utility software platform to analyze the energy consumption of street lighting in Turkey, where electricity consumption patterns and anomaly detection research in Turkey are minimal. Finally, analysis for public lighting in Turkey, related to electricity consumptions and anomalies in these consumptions, is explained in detail compared to other implementations in Europe.

**Keywords**: Machine Learning, Energy Efficiency, Information Systems, Smart Cities, Time Series Analysis, Anomaly Detection

**Introduction**

Streetlights are not lambs that enlighten their ambiance; they are more than a lamb on the road. There can be various types (older or modern; automatic or manual) and dimensions (shorter, taller) for streetlights. They serve to enhance citizens' safety, comfort, and security as one portion of several municipal properties. They can be located on sidewalks, paths, parks, pedestrian areas. An adequately designed street lighting system is crucial (Leccese et al., 2020). The number of streetlights in a city generally depends on the city’s surface or demography.

This paper discusses how streetlight energy savings can be optimally planned and implemented through integrated information systems and based on data analytics. Using multiple techniques to analyze data from different perspectives and implementation of a proper information system are found effective to calibrate energy utilization at the governmental level. The comparison with other similar approaches of various countries is also underlined and Turkish use case was explained in this paper as a complementing explanatory study for researchers working on this field of study.

In contrast to this perspective, streetlights on the road can also affect the citizens’ habits and demography. For example, people prefer the enlightened road to walk alone at night even if that road is not the shortest way to their destination. In all cases, their number consists of great expense for the government. Some countries have regulations for establishing a reliable ratio between road traffic volume and minimum illuminance level (Leccese et al., 2020). Moreover, a flawed designed lighting system can contribute to light pollution. Therefore, governments should take precautions related to useless energy utilization and detect abnormal energy usage. Before taking precautions and detecting abnormal utilization, the pattern of regular electricity consumption should be obtained.

**Literature Review**

In the literature, there is not directly related to detecting the pattern of streetlights utilization, which is our statement of the problem. Therefore, in this part, we tried to get information from many related articles of previous research where a similar problem has been discovered in different geographies under different conditions. These variations within each case also bring in numerous parameters to analyze the similar problem from multiple angles and differences of conditional effects for each specific use case in itself. From this perspective, Leccese et al. (2019) focused on analyzing correlations between spatial properties and level of lighting to get beneficial information in the early stage of urban design. Leccese et al. (2019) get the correlation through the I-index, which describes the mean depth according to all other lines in the axial map. They selected this spatial indicator because it shows the best goodness-of-fitness as regards traffic data that is observed. It should be said that this spatial indicator is correlated with the leading lighting parameters such as the average-maintained luminance in carriageways, the average-maintained illuminance, and the minimum maintained illuminance.

During the study, three different scenarios are examined by Leccese et al. (2019). These are the present, the design, and the ideal scenario in which European technical standards are included as minimum lighting requirements values. As a result, they concluded that in a scenario, there was no link between the traffic conditions in terms of the I-index and the lighting levels in roads analyzed in Scenario-1. Scenario-2, because of applying a proposed methodology from another article, they concluded that if a road is more trafficked or on central and requires more restrictive lighting, then it is expected to higher illuminance levels with respect to other peripheral or lesser trafficked roads. Lastly, in Scenario-3, the values of minimum lighting requirements are set to enlighten the roads, which is not realizable in real life all the time. Nevertheless, it can be said that this scenario has the best energy performance. Finally, to set the lighting requirements formerly, with no need of special methods of road classification according to traffic volumes, would be possible.

Leccese et al. (2020) studied the applicability of the space syntax methodology to determine street lighting classes. They mentioned that lighting classes are selected according to several concepts. The most complicated concept to evaluate is traffic volume, measured by particular companies observing the traffic. Since these companies are not affordable or available for many municipals, Leccese et al. (2020) suggested an alternative way to assess lighting classes utilizing a space syntax approach. Their case study for a town in Italy, Pontedera, showed a good correlation between traffic volume measured by a particular company and traffic volume estimated by their approach. With the help of their study, they can estimate the traffic volume precisely and fast. Then, they can design a proper lighting system to save unutilized energy consumption.

Serrano-Guerrero et al. (2020) suggested a new time-series treatment method: seasonality analysis of electricity consumption (SAEC) since several other methods do not produce the expected results. Moreover, the SAEC method obtains a better electricity consumption pattern with less time and computational effort. Then, they used electrical consumption patterns for anomalies detection improvement in electrical consumption profiles. Also, they analyzed the electricity consumption amount of two different university facilities such as Universitat Politècnica de València in Spain and Universidad Politécnica Salesiana in Cuenca, Ecuador. The results are compared with widespread methods such as without seasonality analysis (WSA), detrending method (DM), and seasonality filter (SF). Serrano-Guerrero et al. (2020) contributed to the literature in several ways. The first contribution was about the conditions that electricity demand data must be provided to make time-series treatment beneficial. The second contribution was an improved anomaly detection through the suggested method, with high precision and with less FPR and FNR. The final contribution was that contextual anomalies are detected by the SAEC method.

Chen and Wu (2018) considered that although data mining techniques are developed and datasets are emerging, no reliable regional-scale empirical models use real data. Therefore, they attempted to identify the concealed distribution patterns in regional energy consumption. Before the experimental data mining process, the crucial steps are explorations with hypothesis tests and statistical analysis. After preprocessing for both missing and abnormal data, information mining, results validation, and application, these patterns are obtained. After the first step of preprocessing, to give a general estimation about energy distribution inequality, the Lorenz curve was proposed by Chen and Wu (2018). In the final step, validations of distinguished distribution patterns and further discussion on application potential were done. Then, these patterns were used to model the region with limited information from 212 and 66 samples belonging to Beijing and Hangzhou.

**Statement of Problem**

Finding the regional streetlight energy consumption level can be done by more advanced technological tools of the internet of things and information technology in the developed countries, whereas, for developing countries, it is not a similar scenario. In developing countries, consumption and data analysis levels cannot be identified without being grounded on the bulk of samples, including multi-dimensional parameters, as Chen and Wu (2018) have explained. Therefore, systems that are more simplified in the implementation but still highly analytical in controlling data would be a critical requirement for a proper analysis of streetlight energy consumption in developing countries.

As Kaygic et al. (2013) also mentioned, the Belgrade Domestic Energy Model (BEDEM) for predicting the energy consumption and carbon dioxide (CO2) emissions, a model for each use case would be required to understand in-depth analysis of each country or city plan of electricity use schema to analyze it better and optimize the overall system for more efficiency. For BADEM, the distribution of energy use has been found as space heating (71%), light and appliances (15%), and water heating (9%). Thus, the streetlights of a city still play a considerable role in energy consumption. Their ﬁndings showed that the uncertainty in this model’s predictions still requires more work to make it more specific and valuable in optimization.

From a similar point of view, Rossi et al. (2016) found that public lighting accounted for the total electricity consumption. There should be solutions to reduce its footprint with, unfortunately, costly lamp replacement projects, which were causing a burden on municipalities. Their conclusion was to solve this situation with a control system using the internet of things that would “exploit the ubiquity of cellular networks.” Rossi et al. (2016) projected that the impact at the country level could scale up beneﬁts at the city level of major Italian cities.

Additionally, Singh and Yassine (2018) presented an intelligent data mining model to analyze and forecast energy time series to uncover various temporal energy consumption patterns. One of which models of this kind explained the potentiality of patterns to define associations with time (hour, day, weekday, week, month, and season of the year) and appliances are vital factors to infer the impact of energy consumption behavior analyze the energy forecasting trend. However, as they underlined, this is a challenge since it is hard to determine multiple relationships and usage from concurrent streams of data to be collected.

According to previous research, we define that the problem is constructed from a set of issues: a collection of energy consumption data is critically important, and the internet of things shall be used for proper construction of a model and setup of complex analyses before optimization. Nevertheless, after the data is cleaned and prepared for analysis, the models and analytical processes are still complex to interpret similarities. Therefore, as the collection of data and multi-variate models are stated, there is a need for a machine learning platform for making both data and models work in alignment and relatively rapid to reach anomalies faster more accurately and propose an optimization based on that result.

**Findings and Analysis**

The case from Turkey on energy consumption analysis of region-based street lighting has been studied to determine possible anomalies on monthly-based net consumption values ​​in meters. This problem is handled as anomaly detection in time series. The FbProphet library developed by the Facebook team was used to solve the problem. This library contains procedures that enable making periodic, weekly, monthly and daily forecasts on non-linear time series data. It has the feature of taking change points, seasonality, and special day effects into consideration while forecasting.

In the study, net consumption values ​​were forecasted for each month using FbProphet. Confidence intervals were determined using the lower and upper limits of the forecasted value obtained. Values ​​outside this confidence interval were defined as anomalies. Apache Spark was used to analyze the counters simultaneously in the analysis process. While enabling parallel processing, Spark not only reduces the analysis time but also enables the operations defined as "job" to be viewed and a possible error to be examined in detail, thanks to the interface (Spark-Shell) it offers.

Within the scope of this study, 38 thousand meters of 21 different private Electricity distribution companies in 71 different provinces were analyzed. The monthly data of these meters for the last two years (2019/01 - 2021/01) were collected. Data assignment processes were made for the missing observations in the counters. On an annual basis, we see that consumption decreases in summer months due to the variation of sunshine duration according to months, while consumption increases in the winter months. This shows us that there is seasonality in the data. The following graph is obtained when the data is visualized to prove this. This graph gives us the average change in net energy consumptions within the relevant months. While a decrease is observed in consumption in June, July, and August, it is seen that there is an increase in December, January, and February. This graph reveals seasonality in the data set. In order to give this seasonal effect to the model, this information is given with the relevant parameters in FbProphet.

Chart

Description automatically generated with medium confidence

Figure . Average Change of Net Energy Consumption in Relevant Month/Year

However, with the data structure at hand, 38 thousand different models have been structured for each electricity meter based on the net consumption data of 24 months. This study of models also delivered a set of forecasts the monthly consumptions of each electricity meter within a confidence interval. Consumptions outside this confidence interval are defined and accepted as anomalies. The graph below gives us the distribution of found anomalies in the months and years analyzed.

Chart

Description automatically generated

Figure . Distribution of Anomaly Numbers by Month/Year

As shown in Figure 2, while net consumption measurements with fewer anomalies were found in summer months, net consumption measurements with more anomalies were observed in winter months. In parallel to the trend of net consumption, there is also seasonality in the observations with anomalies. Prediction models were based the collection of data from a system for energy management that was connected to central data-pipe platform from national level energy distribution companies.

When we examine the net consumption values with anomalies in Turkey, the first place is in the province of Istanbul. Antalya follows it. The graph in Figure 3 shows the number of anomalies as the color intensity increases. In other words, the darkest provinces contain the most anomalies. Provinces not included in the analysis are depicted in gray.

Map

Description automatically generated

Figure . Distribution of the Number of Net Consumption Measurements Containing an Anomaly Based on Provinces in Turkey

In order to measure the success of the established models, MAPE (Mean Absolute Percent Error), which is one of the error metrics frequently used in the literature, was used. This metric calculates the average of the absolute percentage difference between the predicted value and the true value.

As a result of the application, the error value in models was calculated as 8% on average. In addition, the importance degrees of anomalies were calculated. While calculating the degrees of importance, the absolute distances of the true values of the detected anomalies from the relevant upper or lower limits of the forecast were measured. Afterwards, the min-max normalization was performed for the distances in model and the importance levels between 0-1 were obtained. This calculation allowed us to take quick action and scale accurately.

**Conclusion**

General lightings such as streetlights, road, and building lightings significantly affect regions’ economies. Therefore, one can find varied papers using varied methods to analyze and from varied perspectives from different countries. Thus, the statement of the problem in each paper was also distinctive. In this paper, our statement problem included obtaining a well-cleaned dataset and utilizing this dataset to detect anomalies with time series analysis.

Data of 38 thousand electricity meters from 71 cities, which belonged to 21 various electricity companies, were used in our study to understand the changes in the usage patterns and the potential for effective measurement and energy savings. Multiple data visualizations and histograms were prepared according to these data from cities with intense and unintense anomalies, and detections were made from these differences. The results concluded an accord with the net electricity consumption distribution, and the distribution of the sum of anomaly consumptions was observed.

Based on our analyses from the installed internet of things infrastructure, complementing data analytics software platform and machine learning tools and techniques with its applicable predictive models integrated is the solution for smart cities' energy efficiency problems for the time being. This comprehensive set of systems is one of the ways to provide support to decision-makers and optimize the grid with advanced analytics.

**References**

1. Gutierrez-Escolar A., Castillo-Martinez A., M. Gomez-Pulido J., Gutierrez-Martinez J,, Stapic Z. , Medina-Merodio J. (2015) A Study to Improve the Quality of Street Lighting in Spain. Energies, 8, 976-994; doi:10.3390/en8020976
2. Serrano-Guerrero X., Escrivá-Escrivá G., Luna-Romero S., Clairand J.M. (2020) A Time-Series Treatment Method to Obtain Electrical Consumption Patterns for Anomalies Detection Improvement in Electrical Consumption Profiles. Energies, 13, 1046; doi:10.3390/en13051046
3. Chena Y., Wu J. (2018) Distribution patterns of energy consumed in classified public buildings through the data mining process. Applied Energy, 226, p 240–251
4. Leccese F., Lista D., Salvadori G., Beccali M., Bonomolo M. (2020) On the Applicability of the Space Syntax Methodology for the Determination of Street Lighting Classes. Energies, 13, 1476
5. Kostica M., Djokic L. (2009) Recommendations for energy efficient and visually acceptable street lighting. Energy, 34, p 1565–1572.
6. Kavgic M., Mumovic D., Summerfield A., Stevanovic Z., Ecim-Djuric O. (2013). Uncertainty and modeling energy consumption: Sensitivity analysis fora city-scale domestic energy model. Energy and Buildings, 60, p 1–11.
7. Millera C., Nagyb Z., Schlueter A. (2017) A review of unsupervised statistical learning and visual analytics techniquesapplied to performance analysis of non-residential buildings. Renewable and Sustainable Energy Reviews. <http://dx.doi.org/10.1016/j.rser.2017.05.124>
8. Rossi C., Gaetani M., Defina A. (2016) AURORA: an Energy Efficient Public Lighting IoT System for Smart Cities. Performance Evaluation Review, Vol. 44, No. 2.
9. Denardin G.W. , Barriquello C.H., Campos A., N. do Prado R. (2009) An Intelligent System for Street Lighting Monitoring and Control. IEEE
10. Bullough J.D., S. Rea M, Zhou Y. Analysis of Visual Performance Benefits from Roadway Lighting Final Report. Project No. 5-19. June 2009
11. Bennett M., Newborough M. (2001) Auditing energy use in cities. Energy Policy, 29, p. 125-134.
12. Singh S., Yassine A. (2018) Big Data Mining of Energy Time Series for Behavioral Analytics and Energy Consumption Forecasting. Energies. 11, 452. doi:10.3390/en11020452
13. Hayes M. A., Capretz M. (2015) Contextual anomaly detection framework for big sensor data. Journal of Big Data. 2:2 DOI 10.1186/s40537-014-0011-y
14. Zhanga S, Jiaoa Y., Chena W. (2017) Demand-side management (DSM) in the context of China's on-going powersector reform. Energy Policy. 100, p 1–8.
15. Werley S. dos Angelos E., R. Saavedra O., Carmona Cortés O. A., Nunes de Souza A. (2011) Detection and Identification of Abnormalities in Customer Consumptions in Power Distribution Systems. IEEE Transactions on Power Delivery. vol. 26, no. 4.
16. Fenza G., Gallo M., Loia V. (2019) Drift-Aware Methodology for AnomalyDetection in Smart Grid. Special Section On Urban Computing & Well-Being In Smart Cities:Services, Applications, Policymaking Considerations. vol 7.
17. Leccese F., Leonowicz Z. (2012) Intelligent wireless street lighting system. IEEE.
18. Uddin S., Shareef H., Mohamed A., Hannan M A., Mohamed K. (2011) LEDs as Energy Efficient Lighting Systems: A Detail Review. IEEE Student Conference on Research and Development.
19. Cheng V., Steemers K. (2011) Modelling domestic energy consumption at district scale: A tool to support national and local energy policies. Environmental Modelling & Software 26 p. 1186-1198.
20. Al-Jarrah O., Al-Hammadi Y., Yoo P.D., Muhaidat S. (2017) Multi-Layered Clustering for Power Consumption Profiling in Smart Grids. IEEE Access. 10.1109/ACCESS.2017.2712258
21. Huang Y., Sun Y., Yi S. (2018) Static and Dynamic Networking of Smart Meters Based on the Characteristics of the Electricity Usage Information. Energies. 11, 1532; doi:10.3390/en11061532
22. Brownsworda R.A., Flemingb P.D., Powell J.C., Pearsall N. (2005) Sustainable cities – modelling urban energy supply and demand. Applied Energy. 82, p.167–180
23. Patterson J.L. (2018) Traffic modelling in cities –Validation of space syntaxat an urban scale. Indoor and Built Environment. 0(0) p. 1–16.