Prediction of Electric Vehicle Charging Stations Distribution Using Machine Learning

Marwa Chendeb EL Rai*, Sabina Abdul Hadi*, Haitham Abu Damis*, Amjad Gawanmeh*
*Department of Electrical Engineering, College of Engineering and IT, University of Dubai, Dubai, UAE.
{mcelrai, sahadi, hadamis, agawanmeh}@ud.ac.ae

2022 5th International Conference on Signal Processing and Information Security (ICSPIS) | 978-1-6654-9265-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICSPIS57063.2022.10002556

***Abstract*—Electric vehicles (EV) are becoming more mainstream as consumer purchasing choices are in favor of this innovative technology. Hence, there is a need to install new electric vehicle charging stations (EVCS). Thus, determining where new stations should be installed is an apparent prospective challenge. In this paper, we tackle the mentioned challenge by creating a dataset enabling us to predict where new EVCS should be be installed. Machine learning models such as K-nearest neighbours, logistic regression, neural networks, and support vector machines can be applied to determine recommended EVCS installation places. City of Dubai located in United Arab Emirates (UAE) is selected to be the model of our case study. Population density, points of interest (POI), and the existence of security cameras are the features considered in this work. The performance of different machine learning models is evaluated using $f1-score$ and accuracy metrics. Experiments show that the validation accuracy is 89% using K-Nearest Neighbors model. The dataset is publicly available at https://github.com/marwarai/Charging-Stations-Dubai.**

***Index Terms*—Charging Stations, Electric Vehicles, K-Nearest Neighbors, Neural Networks, Support Vector Machine, Logistic Regression, Dubai.**

I. Introduction

Prospectively, urban planning decisions will be influenced by the accommodation of hybrid and electric vehicle charging stations. This is particularly true for rapidly developing cities in the United Arab Emirates (UAE). One can see an accelerated adoption of electric vehicles (EV) by the residents of Dubai. Some of the actors involved in making a purchasing decision for a particular EV are vehicle life, battery capacity, electric vehicle charging stations (EVCS) accessibility, and monetary cost. Charging stations could be classified according to their accessibility and power supply rating. Here, we are concerned with direct current (DC) fast charging stations which are the predominant type of public EVCS. Cities are complex topographic and demographic entities. Therefore, multiple variables can be taken into account when planning for the installation of public EVCS in a geographic region. Socioeconomic factors can also come into play. For instance, Roy and Law demonstrate in [1] the inequities of EVCS distribution in Orange County, USA. Their methodology utilizes a machine learning framework based on Multi-Criteria Decision Analysis (MCDA). One of their findings suggests that relatively low-income and low car ownership neighborhoods have the least access to charging stations. In [2], the authors adopted a game theory approach to investigate the interaction among public EVCS availability, traffic route choices, and electric utility pricing. The authors mention that given EVCS usage and consumer behavior, there could exist a substantial impact on power distribution networks and electric utility rates. Hence, it is important to consider this impact when making EVCS installation decisions. It is worth mentioning that relying on only a machine learning algorithm is not enough to guarantee optimal EVCS location choices. Rather the authors in [3] advise that data sets of EVCS usage sessions should be coupled with contextual information about EV owners. This might include places of interest of drivers and travel distances between dispersed EVCS's. The authors have used Multiple Linear Regression (MLR) with their data sets of interest and XGBoost for prediction modeling. In [4], an elaborated approach towards balancing the demands of EV users and the interests of power utility companies is presented. The conflicting interests between the two parties would be more palpable as higher expectations for more EVCS become on the rise. Moreover, [5] demonstrated an optimized combination of different types of EV chargers to efficiently manage EV load distribution while minimizing installation cost, losses, and transformer loading. As the number of electric vehicles is exponentially growing in the UAE, there is an urgent need for more EVCS around the country. In our study, we build a data set of prospective locations and numbers of EVCS to be installed after identifying the main available and relevant features such as population density, Points of Interest (POI), and presence of security cameras. This project can help the city in making optimal EVCS installation and distribution decisions.

Although there are numerous review and research papers that discussed EVCS using machine learning [6] [7], there are still a number of reasons that motivated this research paper:

- The focus on the integration of crucial features such as population density, points of interest and security camera.
- The ability to include much broader geographic scope that is able to take in consideration different population sizes.
- The ability to predict new locations based on the existing EVCS.

The paper is organized as follows: Section 2 presents the proposed methodology. Section 3 describes how the data set is created. The experimental results are reported in Section 4. Finally, Section 5 summarises and concludes the paper.

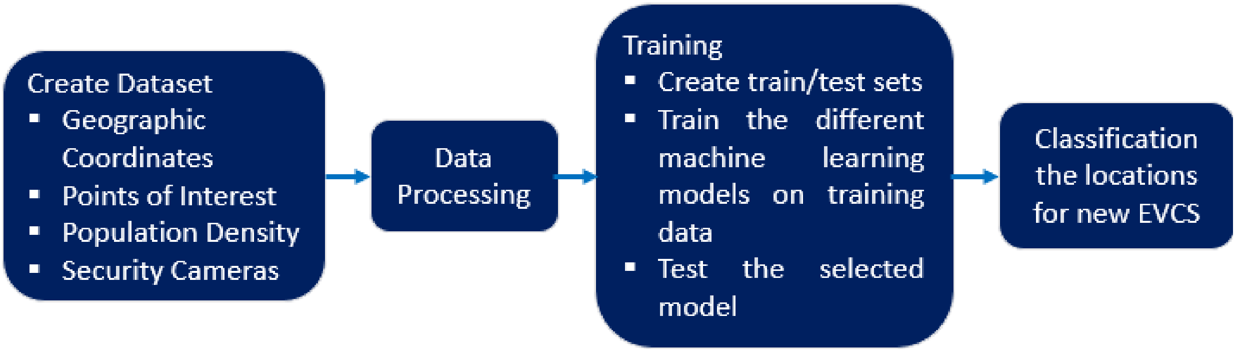

Fig. 1: Framework of the proposed methodology.

OBJECTID	Location	Latitude	Longitude	POI	POI Encoded	Camera	Camera Encoded	Cluster Object ID	Label_positive	Population
1	Dewa EV green charger jumeira beach hotel	25.14181833	55.19190459	hotel	1	yes	1	100	1	7021
2	powertech electrical trading LLC	25.18718951	55.28096242	company	2	yes	1	86	1	19715
3	EV green charging station	25.13367637	55.18563066	charging st	0	not sure	0	104	1	4112
4	EV-box charging station	25.10615745	55.19591335	charging st	0	not sure	0	105	1	35469
5	tesla destination charger (Al Barsha)	25.11906433	55.20037206	charging st	0	not sure	0	105	1	35469
6	DEWA Fast charger	25.09759942	55.16516295	charging st	0	not sure	0	109	1	5592
7	Tesla Destination charger (jumeirah beach road)	25.14224393	55.19123496	charging st	0	not sure	0	100	1	7021
8	tesla destination charger (mina a'slam)	25.1352551	55.18605844	mall	2	yes	1	104	1	4112
9	tesla destination charger (jumeirah Al Naseem)	25.13728513	55.18701605	hotel	1	yes	1	100	1	7021
10	Tesla destination charger (jumeirah Al Qasr)	25.13188308	55.18438736	hotel	1	yes	1	104	1	4112

TABLE I: Some samples of collected data with the different parameters.

II. Proposed Methodology

Our proposed methodology aims to determine ideal locations for EVCS spots based on building up a training/validation/testing data sets of geographic coordinates selected randomly throughout Dubai, and finding the useful parameters for each of these location. The features which we considered are population density, points of interest and the presence of security camera. After the generating of data set, the objective is to find a machine learning model that recommends the optimal locations in Dubai to install the new EVCS. The machine learning classification models considered in this work are logistic regression, neural networks, and support vector machines (SVMs) and K-nearest neighbor. Figure 1 displays the framework of the proposed methodology.

1) Population Density: The population density is collected from Dubai Statistics Center obtained from Dubai Census data which carried out the general census of population [8]. Dubai is divided into clusters using the ArcGIS software. The density population is computed over each cluster as the number of people per square meter for each cluster. This parameter gains high importance because the areas having high population density tend to have higher probability of electric vehicle drivers as part of the local population.

2) Points Of Interest: Points of interest defined how predominantly people are visiting a specific location. As EV charging stations should increase in areas that people drive to often. The points of interest are coded into three levels 0 to 2. Level 2 is assigned to the places where the need for installation is optimal. Level 1 is assigned to the locations with middle interest. Level 0 is the code of the non optimal locations. Points of interest are classified in categories such as hotels, restaurants, malls, hospitals, parks, petrol station, governmental institutions, avenues, and charging stations. For example, charging stations are as 0. Hotels and restaurants are assigned to 1. Malls and hospitals are considered as high important places and they are coded as 2.

3) Security Camera: Some of the charging stations have low power charging output. So, it will take time until the drivers charge their car. The existence of a security camera was used as a index of how safe the area would be for people waiting for their electric vehicles to charge. Binary value 1 is coded to indicate that there is a camera, otherwise the binary value is 0.

A. Machine Learning Approaches

Machine learning (ML) algorithms offer the opportunity to learn and discover patterns from training data. Thus the successful created model serve to predict accurately potential EVCS. ML methods are classified into supervised and unsupervised learning. Regression or classification methods are used for continuous or categorical predictive response, respectively. The algorithms used in this paper for prediction of EVCS are briefly discussed below:

1) Logistic Regression: Logistic regression (LR) is a statistical linear model for binary classification predictive purpose. The mathematical model relates between the different input features $[x_1, x_2, ..., x_n$ and the response variable y as follows:

$$y = b_1x_1 + b_2x_2 + ... + b_nx_n \tag{1}$$

where b_i are the regression coefficients. Initial values are selected randomly then by Using gradient descent [9] method, the regression coefficients are estimated by minimizing the error obtained between selected and estimated values. In this work, Python package $(sklearn.linear.model.LinearRegression)$ [10] is used. The hyperparameter C which is the inverse of regularization strength try to ensure strong regularization for small values.

2) K-Nearest Neighbors: K-Nearest Neighbors (KNN) is a non-parametric method [11] used for classification and regression. For prediction, the euclidian distance is measured to find the k nearest neighbors then it is labeled as the class which includes the most of the neighbors. In this work, the Python package (sklearn. neighbors.KNeighborsRegressor) is applied for the KNN model [10].

3) Neural Networks: One of the most common Neural Neworks methods is multilayer perceptron (MLP). Given a set of input parameters, MLP can exploit non-linear approximation given for regression and classification purposes. MLP incorporates input layer built in the set of features, the hidden layers that acquires the representations and the output layer that predict the final values [12].

4) Support Vector Machine: Support Vector Machine (SVM) is based on the use of best hyperplane that can maximize the edges between the different classes [13]. Using different kernels such as radial basis function, polynomial, Gaussian and linear, the inputs are assigned to high dimensional parameter spaces where they can be separated linearly.

III. Dataset

The dataset is created by collecting the geographic coordinates of 162 existing EVCS in Dubai. ArcGIS software is used in the determination of the population density for each cluster as Dubai is divided into polygons. Figure 2 shows the population in different clusters of Dubai city. The blue indicators represent the center of each cluster reflecting the corresponding geographic coordinates and population density. The red indicators represent some locations of EVCS. The interest of each location is assigned based on the importance of place. The presence of security camera is encoded by assign 0 if there is no camera, and 1 if a camera exists. These data are labeled positive. Table I shows some samples of existing EVCS with the different features. Another 162 locations were selected randomly lacking of EVCS. They represent the negative labeled data. The dataset is split as follows: 80% for training, 10% for validation and 10% for testing. The data is normalized in order to make each column evenly weighted for the model.

TABLE II: Comparison of classification performance with different machine learning approaches.

Method	Precision	Recall	F1-Score	Accuracy
K-Nearest Neighbors	**0.90**	**0.89**	**0.89**	**89%**
Neural Networks	0.79	0.74	0.74	76%
Support Vector Machine	0.85	0.85	0.84	84%
Logistic regression	0.8	0.8	0.79	79%

IV. Results and Analysis

The metrics applied to measure the model performance are $F1-score$ and accuracy defined as follows:

$$F1-Score = \frac{2 \times precision \times recall}{precision + recall} \tag{2}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

where

$$precision = \frac{TP}{TP + FP} \tag{4}$$

$$recall = \frac{TP}{TP + FN} \tag{5}$$

TP, TN, FP and FN stand for True Positive, True Negative, False Positive and False Negative, respectively. Precision quantifies the number of positive class predictions that actually belong to the positive class. Recall quantifies the number of positive class predictions made out of all positive examples in the dataset. The training data is used to train the machine learning algorithms. The validation set is useful to tune the hyperparameters of the obtained models. The parameter C is tuned in the case of logistic regression model. C=400 provided the highest performance. In SVM case, the basis function kernel has been selected after the experimentation of the other kernels such as Linear Kernel, Polynomial Kernel, Sigmoid Kernel, Gaussian Kernel and others. We found that less regularization was helpful in both models and the radial basis function to be the most accurate kernel for SVMs. For KNN classifier, different experiments have been performed with different values of k. The optimal value is $k = 3$ providing the higher accuracy. Finally for the neural network, the best results have been acquired using 4 hidden layers with 150, 100, 50, 25 neurons respectively. Table II offers the precision, recall, $F1-score$ and accuracy gained by applying the different machine learning models on the validation set. K-NN outperforms the other algorithms with $F1-Score$ equal to 0.89 and 89% accuracy. The final K-NN model model (with $k = 3$) had a test set accuracy of 85.8% with an overall accuracy of 87.9%.

A. Prediction

The K-NN model is used to predict the locations of eventual EVCS. 98 locations without having EVCS have been selected randomly. After applying K-NN model, 16 locations were recommended to be the eventual locations to implement new EVCS infrastructures. For example, in a residential cluster (longitude 55.282562 and latitude 25.038423) where the population density is 1889.27, there is a recommendation to implement an EVCS as shown in figure 3.

V. Conclusion and Future Work

In the rapid turn of owning EV, the need for charging stations becomes greater. Being a source of energy that is cheap to charge, an increase in EV drivers will deem strategic planning and building of EVCS necessary. Our project consists

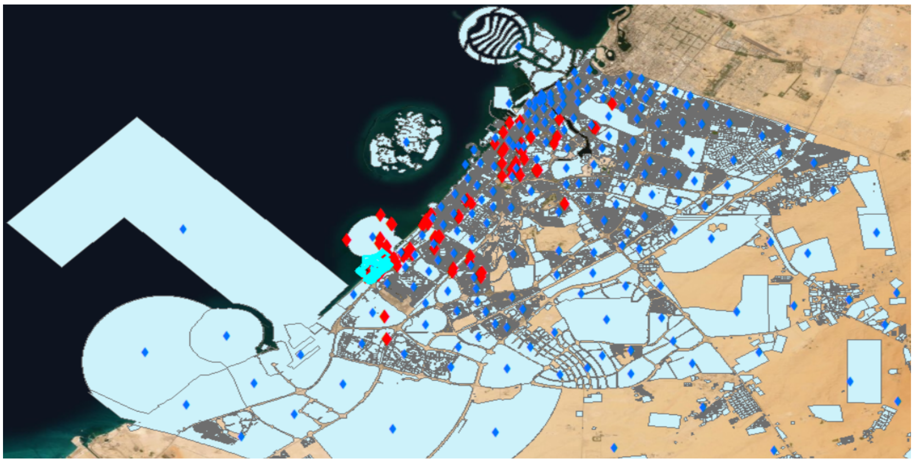

Fig. 2: Map of Dubai city: The population density is calculated for each blue cluster. The red flags correspond to the actual charging stations.

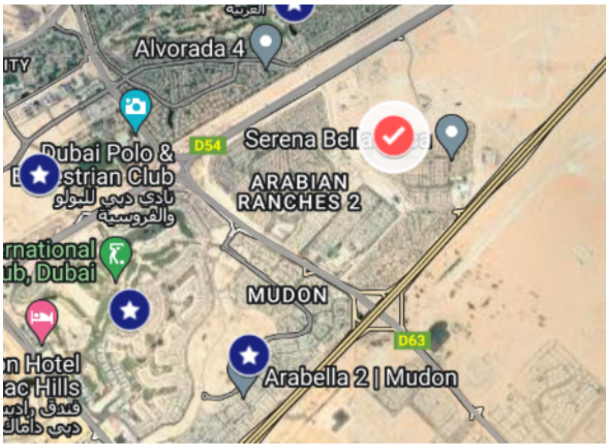

Fig. 3: The red symbol highlights the recommended location to implement an EVCS.

of data collection applied on machine learning models to predict the best EVCS locations. The population density, points of interest, and safety are the relevant factors used to create the dataset. A visualization capability for the predictions by projecting and integrating the recommended points into map is provided. In future work, it is important to extend the work to other cities and over he country. In addition, there is a need to update the data set to obtain better prediction accuracy over time. The realized approach has an enormous potential for urban planning in smart cities. Other environmental and relevant data factors like public transportation, traffic flow and neighboring fuel stations can be added to increase the performance.

VI. Acknowledgement

We deeply thank students Rashed Bin Katfoor, Muhannad Fallatah, Abdulla Salem Mohammed and Saif Karam from University of Dubai for their support in this work.

References

[1] Avipsa Roy and Mankin Law, "Examining spatial disparities in electric vehicle charging station placements using machine learning," *Sustainable Cities and Society*, vol. 83, 2022.

[2] Fang He, Di Wu, Yafeng Yin, and Yongpei Guan, "Optimal deployment of public charging stations for plug-in hybrid electric vehicles," *Transportation Research Part B: Methodological*, vol. 47, pp. 87–101, 2013.

[3] Dario Pevec, Jurica Babic, Martin A Kayser, Arthur Carvalho, Yashar Ghiassi-Farrokhfal, and Vedran Podobnik, "A data-driven statistical approach for extending electric vehicle charging infrastructure," *International journal of energy research*, vol. 42, no. 9, pp. 3102–3120, 2018.

[4] Leonardo Bitencourt, Tiago P. Abud, Bruno H. Dias, Bruno S.M.C. Borba, Renan S. Maciel, and Jairo Quirós-Tortós, "Optimal location of ev charging stations in a neighborhood considering a multi-objective approach," *Electric Power Systems Research*, vol. 199, 2021.

[5] Muhammad Zulqarnain Zeb, Kashif Imran, Abraiz Khattak, Abdul Kashif Janjua, Anamitra Pal, Muhammad Nadeem, Jiangfeng Zhang, and Sohail Khan, "Optimal placement of electric vehicle charging stations in the active distribution network," *IEEE Access*, vol. 8, pp. 68124–68134, 2020.

[6] Christopher Hecht, Jan Figgener, and Dirk Uwe Sauer, "Predicting electric vehicle charging station availability using ensemble machine learning," *Energies*, vol. 14, no. 23, pp. 7834, 2021.

[7] Xingshuai Huang, Di Wu, and Benoit Boulet, "Ensemble learning for charging load forecasting of electric vehicle charging stations," pp. 1–5, 2020.

[8] "https://www.dsc.gov.ae/en-us/pages/default.aspx," 2020.

[9] Sebastian Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, 2016.

[10] Lars Buitinck, Gilles Louppe, Mathieu Blondel, Fabian Pedregosa, Andreas Mueller, Olivier Grisel, Vlad Niculae, Peter Prettenhofer, Alexandre Gramfort, Jaques Grobler, et al., "Api design for machine learning software: experiences from the scikit-learn project," *arXiv preprint arXiv:1309.0238*, 2013.

[11] Naomi S Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992.

[12] Matt W Gardner and SR Dorling, "Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences," *Atmospheric environment*, vol. 32, no. 14-15, pp. 2627–2636, 1998.

[13] William S Noble, "What is a support vector machine?," *Nature biotechnology*, vol. 24, no. 12, pp. 1565–1567, 2006.