International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017),e-ISSN:2455-2585
International Conference on
Recent Explorations in Science, Engineering And Technology (ICRESET'19)
Volume-5, Special Issue-March, 2019.

BIG DATA CHALLENGES FOR E-MOBILITY- INFRA OPERATOR

Roshini.S[1] ,Vrinda Raj.V.P[2],Prof.Jayashree.R[3]
[1,2]UG Students, Department of CSE, KGiSL Institute of Technology, Coimbatore, India
[3]Professor, Department of CSE, KGiSL Institute of Technology, Coimbatore, India

Abstract*— *The E-Charging Operators (equivalent of today's petrol pump operators) need a Big Data approach to ensure their customers (end-users) satisfaction and optimum utilization of the e-Charging stations. The operator may have fast or slow charging infrastructure with their own payment terms and timings. Based on Machine Learning algorithms the operator can offer dynamic pricing at stipulated times and manage the peak demand accordingly.

Keywords*—*E-Charging Operators, Big Data, Machine Learning, dynamic pricing.

INTRODUCTION

Machine learning is an interdisciplinary research area which combines ideas from several branches of science namely, artificial intelligence, statistics, information theory, mathematics, etc.
The prime focus of machine learning research is on the development of fast and efficient learning algorithms which can make predictions on data. When dealing with data analytics, machine learning is an approach used to create models for prediction. Machine learning tasks are mainly grouped into three categories- supervised, unsupervised and reinforcement learning.Supervised machine learning requires training with labeled data. Each labeled training data consists of input value and a desiredtarget output value.
The supervised learning algorithm analyzes the training data and makes an inferred function, which may be used for mapping new values. In unsupervised machine learning technique, hidden insights are drawn from unlabelled data sets, for example, cluster analysis.
The third category, reinforcement learning allows a machine to learn its behavior from the feedback received through the interactions with an external environment. From a data processing point of view, both supervised and unsupervised learning techniques are preferred for data analysis and reinforcement techniques are preferred for decision making problems.

LITERATURE SURVEY

Electric vehicles seem to have finally gained a solid foothold. With continued adoption, there will be an increasing need for access to charging locations. We recognize that many drivers today do most of their charging at home, but many others still require access to a robust nationwide charging station network before even considering the purchase of an electric vehicle. But high costs of equipment and installation are currently impeding the build-out of such a network. Therefore, cost-effective solutions are needed to ensure future investment in charging stations.We recently interviewed over a dozen companies involved with charging station infrastructure—including utilities, automakers, cities, research institutions, and charging station companies—to pull back the veil on current EV charging station infrastructure costs. From there, the next post in this series can then explore charging station business models and strategies to reduce those per-station costs.
We've broken down the cost into several categories: 1) the actual charging station hardware, 2) other hardware and materials, 3) electrician and other labor, 4) mobilization, which we define as time for the electrician and others to prep and get to the worksite (often including an initial on-site consultation), and 5) permitting. This is a look at raw infrastructure costs. We did not include: a) general and administrative overhead, which can easily contribute 20 percent or more to costs but which also reduce rapidly with experience, b) other miscellaneous costs, such as wage requirements for federally funded projects, which can add 15–25 percent to costs, and c) financing (and other costs of capital/debt) for charging station owners.Similarly, we don't include federal, state, and local EV charging station incentives that could reduce per-charging-station infrastructure costs, such as the federal EV charging station infrastructure tax credit that expired at the end of 2013, Connecticut's EV charging station grant program, and the plug-in EV charger rebate program with the city of Anaheim, CA.

PROPOSED SYSTEM

In our proposed system, with the Machine learning algorithms like k means, Linear regression and decision tree we write our own algorithms for each type and also predict the accuracy for each algorithms for the given data set. With the help of the prediction score, we find the suitable algorithm for the given data and provide the correct pricing structure according to the demand.

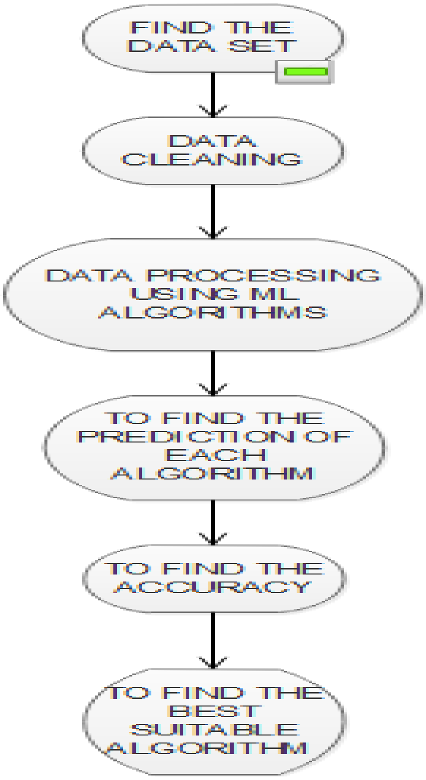
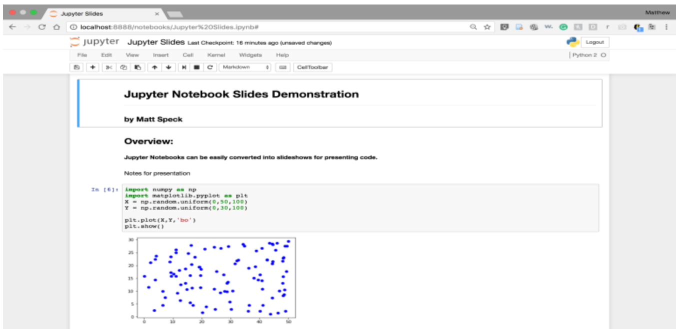


Fig.4. Proposed System Work Flow

OUTPUT AND RESULTS

This section explains the output in the form of screenshots in different levels starting from, registration of users, shortlist the frequent users based on the location, notify the users with the available offers for that particular period.

- **Jupyter Notebook**



Project Jupyter exists to develop open-source software, open-standards, and services for interactive computing across dozens of programming languages.

Data Cleaning:

```
In [47]: df["Energy(kWh)"].fillna("32", inplace = True)

In [48]: print(df[df["GHG_savings(kg)"].isnull()][null_columns])

    No Of Sessions  Accumulated Sessions  No. of Ports  Energy(kWh)  \
91              10                   927            28           35

    GHG_savings(kg)  Charge Fee(in$)
91              NaN               20

In [49]: df["GHG_savings(kg)"].fillna("0", inplace = True)

In [50]: print(df[df["Charge Fee(in$)"].isnull()][null_columns])

Empty DataFrame
Columns: [No Of Sessions, Accumulated Sessions, No. of Ports, Energy(kWh), GHG_savings(kg), Charge Fee(in$)]
Index: []

In [51]: df["Charge Fee(in$)"].fillna("0", inplace = True)

In [52]: print (df.isnull().sum())

Date                         0
No Of Sessions               0
Accumulated Sessions         0
Unique Drivers               0
No. of Ports                 0
Energy(kWh)                  0
Accumulated_Energy(MWh)      0
GHG_savings(kg)              0
Accumulated GHG (kg)         0
Charge time (minutes)        0
Gasoline_Savings(gallons)    0
Charge Fee(in$)              0
dtype: int64
```

Linear Regression:

Linear regression is used to estimate real world values like cost of houses, number of calls, total sales etc. based on continuous variable(s). Here, we establish relationship between dependent and independent variables by fitting a best line. This line of best fit is known as **regression line** and is represented by the linear equation

Y= a *X + b.

Predicting the Test Results

```
In [11]: prediction = linear_regression.predict(X_test)

In [12]: prediction

Out[12]: array([ 35.97218779,  77.31193963,  86.70360178,  85.03249486,
                 61.23372195, 100.87505117, 102.93166294,  67.06454547,
                 73.469528  ,  84.78755339,  68.38189655,  66.37038948,
                 80.39076018, 121.7684763 ,  90.36504756,  84.74201039,
                 90.96758588,  71.59526314,  72.78407473,  95.31489636,
                 77.81737932,  87.45128782,  71.19179737,  82.91267748,
                108.57230878, 103.34322667,  68.86762295,  65.77467476,
                 55.15317102,  90.62089595,  91.88625602, 105.3097544 ,
                 76.41492465,  80.96553886,  72.6505691 ,  65.41746689,
                 85.10503977,  89.08574404, 117.01558388,  82.31899103,
                 46.25052691,  82.42239418,  86.01166079,  86.53530272,
                 86.99572218, 114.07374205,  90.95818465,  64.75672692,
                 76.1651108 ,  51.84867313,  96.43810422,  87.52153879,
                 77.20979401,  78.90987973,  35.01497095,  69.35000014,
                 80.00396888, 102.51764297,  69.37100236,  93.86499806,
                 70.54555262,  93.1229071 ,  88.63334199,  67.6989229 ,
                 71.95569483,  81.36582591,  96.19877888,  85.45186912,
                 70.85546546,  73.89062636,  65.09813048,  86.24810645])
```

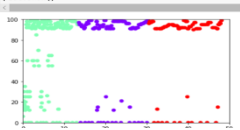
pediction for a same set of data

```
In [13]: linear_regression.predict([[1,1,1,16,7,0.01,3,3,123,0.82]])

Out[13]: array([18.41808282])
```

K_Means:

It is a type of unsupervised algorithm which deals with the clustering problems. Its procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters). Data points inside a cluster are homogeneous and are heterogeneous to peer groups.

```
imputer = Imputer(missing_values='NaN',strategy='mean',axis=0)
imputer = imputer.fit(cluster_X)
cluster_X = imputer.transform(cluster_X)

C:\Users\roshi\Anaconda3\lib\site-packages\sklearn\utils\deprecation.py:58: DeprecationWarning: Class Imputer is deprecate
d; Imputer was deprecated in version 0.20 and will be removed in 0.22. Import impute.SimpleImputer from sklearn instead.
  warnings.warn(msg, category=DeprecationWarning)

]: K_Means = KMeans(3)
   K_Means.fit(cluster_X)

t[24]: KMeans(algorithm='auto', copy_x=True, init='k-means++', max_iter=300,
           n_clusters=3, n_init=10, n_jobs=None, precompute_distances='auto',
           random_state=None, tol=0.0001, verbose=0)

]: cluster_prediction = K_Means.fit_predict(cluster_X)
   prediction_dataset = data.copy()
   prediction_dataset['cluters'] = cluster_prediction

]: plot.scatter(prediction_dataset['Accumulated_Energy(MWh)'],prediction_dataset['Charge Fee(in$)'],c=prediction_dataset['cluter
   plot.xlim(0,50)
   plot.ylim(0,100)
   plot.show()
```



CONCLUSION AND FUTURE ENHANCEMENTS

This study proposed a methodology to analyze EV charging station usage. Data was collected from three ChargePoint level 2 charging stations located in three parking decks at UGA. Information about 3204 charging events from 10 April 2014 to 20 June 2017 was recorded. The data was analyzed using the protocol presented to provide information about EV charging station usage at UGA in terms of charging behavior, charging station occupancy, and geolocation of charging9 station users.

The methods used in this study are reproducible for other cities, and the results will be useful for planning EV charging infrastructure development on other college campuses in the USA. For colleges that plan to build additional charging stations, the methodology presented can be used to examine past charging behavior at existing charging stations. The data collection method is time- and cost-effective because ChargePoint charging stations collect data about every charging event, and this data is quick, easy, and free to obtain. Additionally, the results from this case study can be used as a reference for possible expectations of EV charging station usage on other college campuses. This will be especially useful for colleges that do not have charging stations but want to introduce EV charging infrastructure to their campus.

REFERENCES

1. [1] Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani Introduction to statistical learning (http://www-bcf.usc.edu/~gareth/ISL/)
2. Hastie, Tibshirani, and Friedman's The Elements of Statistical Learning
3. Bishop's Pattern Recognition and Machine Learning
4. David Barber's Bayesian Reasoning and Machine Learning
5. Kevin Murphy's Machine learning: a Probabilistic Perspective
6. Foundations of Machine Learning, Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar
7. Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die by E. Siegel