International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT206552

Using Machine Learning Analytics to Detect Abnormalities and Electricity Theft

Sowrav Saha[a], Utsho Chakraborty[b], Haimanti Biswas[c], Md. Intekhab Rahman Galib[d], Dr. Sheshang Degadwala[e]

[a]UG Scholar, Civil & Infrastructure Engineering, Adani Instiute of Infrastructure Engineering, Ahmedabad, India
[b]UG Scholar, Computer Engineering, Sarvajanik College of Engineering and Technology, Surat, India
[c]UG Scholar, Information Technology, Sigma Institute of Engineering, Vadodara, India
[d]UG Scholar, BBA, Gujarat University, Ahmedabad, India
[e]Associate Professor and Head of Department, Computer Engineering Department, Sigma Institute of Engineering, Vadodara, India

Article Info
Volume 6, Issue 5
Page Number: 271-279
Publication Issue :
September-October-2020

Article History
Accepted : 15 Oct 2020
Published : 23 Oct 2020

ABSTRACT

Abnormalities and Electricity Theft are a major concern for power and economic chaos of one's country. The reason here is, the fraudulent usage of electricity power by customers and broken electric meters or billing errors. Currently, electrical transmission and distribution losses remain a hurdle to the development and sustainability of the sector despite several techniques of energy conservation and electricity distribution analysis that have been employed. While technical failures are regular and predictable, non-technical losses, which are responsible for 80% energy losses, are random and hard to identify and evaluate. Hence it requires more advanced technology. For these reasons, the problem has attracted research interests in many fields, including artificial intelligence, including machine learning and expert knowledge approaches. Here, we have used a linear regression method for anomaly detection. The project therefore showed that the method has improved detection accuracy, sensitivity and reduced magnitude of data required.

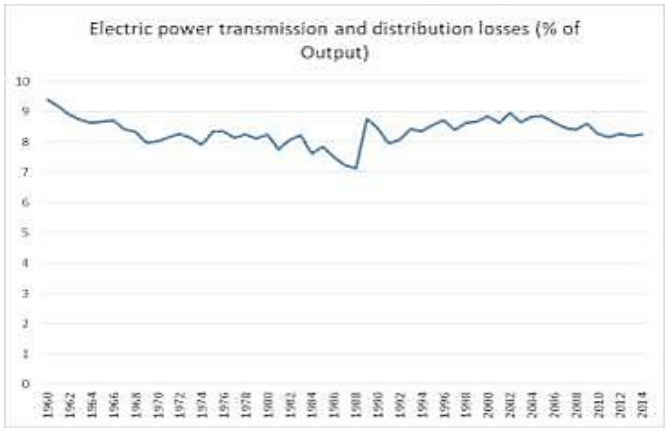
Keywords : Abnormalities, Electricity Theft, Fraudulent, Sustainability, Energy Conservation, Magnitude

I. INTRODUCTION

Over the past fifty years, electricity has become an integral part of our lives. In fact, having electricity is increasingly becoming a basic need for citizens in both developed and developing countries; in any case, electricity connection is as basic as access to water and sanitation. While countries continue to invest heavily in the national grid, not all power generated translates to revenue for the utility companies (Depuru, 2012; Depuru Wang, & Devabhaktuni, 2010). Some of the energy generated is lost through transmission and distribution. Since 1998, the average global electric power transmission and distribution losses as a share of total output have increased from 7.1% in 1998 to 8.251% in 2014 (Nagi et al., 2009; Kumar, Prasad, & Samikannu, 2017; Yakubu, Babu, & Adjei, 2018). However, the peak losses happened in 2002 where a total 8.9% of all electricity produced globally was lost at some point during transmission and distribution as shown in Figure 1. However, these losses are high in the least developed countries where 16% of power

output is lost during transmission and distribution to customers.

Consequently, utility companies lose more than $90 billion every year to electricity losses. Surprisingly, Gaur and Gupta (2016) and Smith (2004) notes that more than 90% of the losses arising from non-technical losses. Non- technical arises mainly from the theft of electricity in form of bypassing meters, tampering with the meters, or connecting directly to the supply lines in low-income settlements where power inspectors may fear enforcement due to security risks.



Source: World Bank

Figure 1: Electric Power Losses as a Percentage of Output

Smart meters and artificial intelligence, however, eliminates the need for constant manual investigation on potential cases of power theft when data lacks to substantiate such claims. They are an effort by utility companies to provide a mean of collecting consumption data more frequently (Gaur & Gupta, 2017; Yurtseven, 2015; Never, 2015). With the help of smart meters, a utility company can collect data that is relayed to servers with the help of mobile networks or fiber-optic network in modern grids. A large amount of data would be meaningless without the application of advanced data mining and analytics techniques (Jamil, 2013; Jamil & Azmad, 2019). Therefore, artificial intelligence such as machine learning has been proposed in the existing literature as crucial in analyzing patterns and periodicity of data collected by smart meters and help stop electricity thieves at their path.

II. LITERATURE REVIEW

Electricity Theft and Detection

Electricity theft has remained an important contemporary issue in electric power transmission and distribution since the early 1990s. Due to the disproportionate share of non-technical losses as a share of overall power system losses, the area of non-technical losses has attracted extensive literature (Dangar & Joshi, 2014; Han & Xiao, 2017; Mashina & Cardenas, 2012; Uparela et al., 2018; Wang et al., 2012). Mashina and Cardenas (2012) focused on evaluating the efficacy of theft detectors in smart grids. Notable in research is the reliance on tamper-proof smart meters by electricity utility companies as the main theft detector in smart grid networks. While such meters are effective in warding off less savvy thieves, they are less effective in preventing the majority of the thefts, which happens when illegal connections bypass the meter (Uparela et al., 2018). Given the weakness of tamper-proof meters in fighting non-technical losses that arise from bypassing of the meter, Uparela et al. (2018) proposes an intelligent system with three building blocks; first module classifies users depending on their power usage and uses; the second module uses ARIMA models to predict future consumptions; the last block uses machine learning to detect fraud. While the model is compelling, the high cost of implementing machine learning solution to electricity theft is highlighted by Shekara & Depuru (2012). Notable hindrance is the lack of specialized skills in machine learning in power systems (Shekara & Depuru, 2012) and high costs of implementing such as solution (Han & Xiao, 2017). While underscoring the value of machine learning in curbing electricity theft, Nabil et al. (2019) decry the lack of privacy in collecting and transmission of electricity consumption data in modern advanced metering infrastructure

(AMI) networks. Electricity theft did not only occur in the instances where the power utilities or their representatives were not aware of. Rather, some of the officers of utility companies take part in the theft of economic gains in forms of bribes from customers (Micheli et al., 2018). Current smart meter methods of detecting fraud fail to account for these occurrences where its employees may collude with customers to aid in electricity theft (Nabil et al., 2019). Along this line, Nabil et al. (2019) propose a Privacy-Preserving Electricity Theft Detection Scheme (PPETD) which leverages convolutional network model to minimize human intervention in smart meters, hence possible fraud by trusted employees.

III.Related Works

```
In [14]:    import pandas as pd
import numpy as np
import matplotlib.pyplot as plt import seaborn as seabornInstance
from sklearn.model_selection import train_test_split fromsklearn.linear_model import
LinearRegression from sklearn import metrics %matplotlib inline
rcParams['figure.figsize'] = 14, 8
RANDOM_SEED = 42
LABELS = ["Normal", "Fraud"] In [2]:

revised-CSV.csv")
data.head(5)
Out[2]:
```

CONS_NO	FLAG	1/1/2014	1/2/2014	1/3/2014	1/4/2014	1/5/2014
0387DD8A07E07FDA6271170F86AD9151	1.0	NaN	NaN	NaN	NaN	NaN
01D6177B5D4FFE0CABA9EF17DAFC2B84	1.0	NaN	NaN	NaN	NaN	NaN
4B75AC4F2D8434CFF62DB64D0BB43103	1.0	NaN	NaN	NaN	NaN	NaN
B32AC8CC6D5D805AC053557AB05F5343	1.0	NaN	NaN	NaN	NaN	NaN
EDFC78B07BA2908B3395C4EB2304665E	1.0	2.9	5.64	6.99	3.32	3.61

5 rows × 1037 columns

```
In [3]: data.info()
<class 'pandas.core.frame.DataFrame'> RangeIndex: 42372 entries, 0 to 42371 Columns: 1037
entries, CONS_NO to Amount dtypes: float64(1036), object(1)
memory usage: 335.2+ MB
```

In [4]:
```
data.describe(include = "all")
```

Out[4]:

	CONS_NO	FLAG	1/1/2014	1/2/2014	
count	42371	42371.000000	25870.000000	25873.000000	2587
unique	42371	NaN	NaN	NaN	
top	7DF7CAEB89C5ACCE3D0FB10C612D44DF	NaN	NaN	NaN	
freq	1	NaN	NaN	NaN	
mean	NaN	0.085318	7.168735	7.057237	
std	NaN	0.279357	34.131237	30.086443	3
min	NaN	0.000000	0.000000	0.000000	
25%	NaN	0.000000	0.000000	0.000000	
50%	NaN	0.000000	3.310000	3.400000	
75%	NaN	0.000000	8.910000	8.570000	
max	NaN	1.000000	3318.000000	2500.000000	267

11 rows × 1037 columns

In [6]:
```
data.replace("?", np.nan, inplace = True)
```

In [9]:
```
missing_data = data.isnull()
missing_data.head()
```

Out[9]:

	CONS_NO	FLAG	1/1/2014	1/2/2014	1/3/2014	1/4/2014	1/5/2014	1/6/2014	1/7/2014	1/8/2014
0	False	False	True	True	True	True	True	True	True	True
1	False	False	True	True	True	True	True	True	True	True
2	False	False	True	True	True	True	True	True	True	True
3	False	False	True	True	True	True	True	True	True	True
4	False	False	False	False	False	False	False	False	False	False

5 rows × 1037 columns

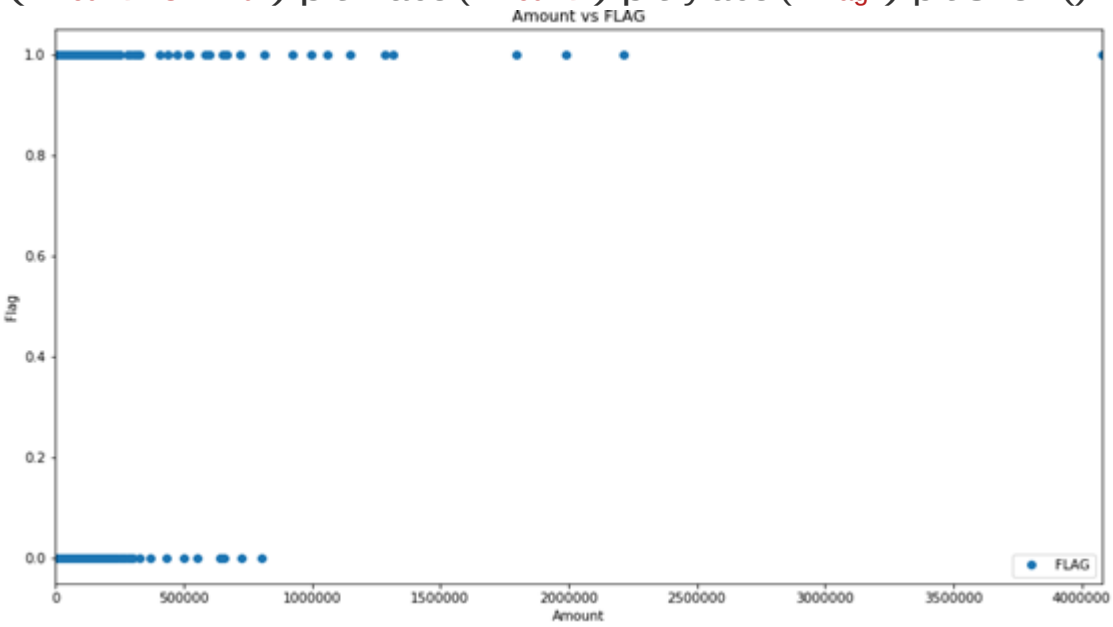
In [10]:
```
for column in missing_data.columns.values.tolist():
print(column)
print (missing_data[column].value_counts()) print("")
```

```
CONS_NO
False  42371
True   1
Name: CONS_NO, dtype: int64
FLAG
False  42371
True   1
```

```
Name: FLAG, dtype: int64

                    1/1/2014
                    False
                    True
                    dtype: int64

Name: 1/1/2014,     1/2/2014
                    False
                    True
                    dtype: int64
Name: 1/2/2014,
                    1/3/2014
                    False
                    True
```

In [ ]:
```
data.fillna(0)
```

In [8]:
```
count_classes = pd.value_counts(data["FLAG"], sort = True)
count_classes.plot(kind = 'bar', rot=0) plt.title("Consumption FLAG Distribution")
plt.xticks(range(2), LABELS) plt.xlabel("FLAG")
plt.ylabel("Frequency")
```
Out[8]:
```
Text(0, 0.5, "Frequency")
```
In [11]:
```
fraud = data[data["FLAG"]==1]

normal = data[data["FLAG"]==0]
```
In [12]:
```
print(fraud.shape,normal.shape)
```
```
(3615, 1037) (38756, 1037)
```
In [16]:
```
data.plot(x="Amount", y='FLAG', style='o')
plt.title("Amount vs FLAG") plt.xlabel("Amount") plt.ylabel("Flag") plt.show()
```



In [21]:
```
X = data["Amount"].values.reshape(-1,1)
```

```
y = data["FLAG"].values.reshape(-1,1)
```

```
In [53]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, rando
```

```
m_state=0)
```

```
In [54]:      regressor = LinearRegression()
regressor.fit(X_train, y_train) #training the algorithm
Out[54]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,
normalize=False)
In [55]:      #To retrieve the intercept:
print(regressor.intercept_)
#For retrieving the slope:
print(regressor.coef_)
```

```
[0.07576006] [[1.37360556e-06]]
In [56]:      #To retrieve the intercept:
print(regressor.intercept_) #For retrieving the slope: print(regressor.coef_)
```

```
[0.07576006] [[1.37360556e-06]]
```

IV.Proposed System

Efficient and comprehensive electricity management techniques are crucial for scientific developments. However, non - technical electrical losses remain a conservation challenge that many scholars have tried to solve for decades but have not been fully implemented. Some of such developments include artificial intelligence approaches such as rule- based expert systems and machine learning and state estimation and network analysis methods, among others. Can electricity distribution credibility be improved further using neural network architecture? This project proposes a linear regression using scikit learn to further improve the accuracy of detecting anomalies of electricity. Given the humongous amount of money utilities lose to electricity theft, the area of electricity theft has attracted extensive literary attention. Conventional electricity theft detection and prevention methods such as frequent inspections of meters and illegal connections have been proposed in multiple studies (Depuru, 2012; Depuru Wang, & Devabhaktuni, 2010; Jamil, 2013; Jamil & Azmad, 2019). Others have proposed the deployment of smart-grids where smart meters transmit electricity consumption data to control centers (Cite). Changing the grid, however, is an expensive undertaking for developing countries struggling with resource limitations and huge budget deficits (Jamil & Azmad, 2019). Coming up with innovative and inexpensive solutions to transforming traditional grids into smart grids, which presents cost-effective ways of fighting electricity theft, improving efficiency, and promoting effective load planning, is crucial for electrical utility firms in the developing countries where electricity theft and inefficiency drive up electricity prices.

Aim and objective signifies the major practice of the frameworks through the project. The project uses a machine learning algorithm called Linear Regression using sickit learn. The algorithm will be trained to predict the electricity usage of customer during three years period. By comparing the predicted value with the actual value consumption patterns of customers can be visualized. When there are values suspected from the training of dataset, the utility company, therefore, can filter out individuals whose power

consumption patterns differ significantly between the two actual and predicted values.

Result Analysis

In [57]:
```
y_pred = regressor.predict(X_test)
```

Out[37]:

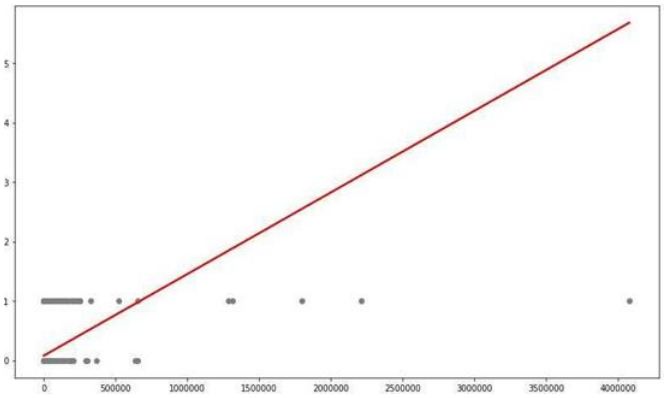
	Actual	Predicted
3.	0.0	0.076553
1	1.0	0.077066
2	0.0	0.079600
3	0.0	0.082553
4	0.0	0.087860
5	0.0	0.084868
6	0.0	0.081590
7	0.0	0.088326
8	1.0	0.083343
9	0.0	0.090207
10	0.0	0.086114
11	0.0	0.080430
12	0.0	0.101879
13	0.0	0.084858
14	0.0	0.077548
15	0.0	0.098423
16	0.0	0.090319
17	0.0	0.091508
18	0.0	0.082809
19	0.0	0.087059
20	0.0	0.077386
21	0.0	0.079228
22	0.0	0.076624
23	0.0	0.078345
24	0.0	0.094601
25	0.0	0.078942
26	0.0	0.088366
27	0.0	0.076547
28	0.0	0.083790
29	0.0	0.084277

	Actual	Predicted
14801	0.0	0.081710
14802	1.0	0.084540
14803	0.0	0.078002
14804	0.0	0.077827
14805	0.0	0.076818
14806	0.0	0.079722
14807	0.0	0.080201
14808	0.0	0.101650
14809	0.0	0.082490
14810	0.0	0.088022
14811	0.0	0.079853
14812	0.0	0.076868
14813	0.0	0.077434
14814	0.0	0.084046
14815	0.0	0.076867
14816	1.0	0.082399
14817	1.0	0.079069
14818	0.0	0.078960
14819	0.0	0.086163
14820	0.0	0.081349
14821	0.0	0.091451
14822	0.0	0.076905
14823	0.0	0.084642
14824	0.0	0.081785
14825	0.0	0.076544
14826	0.0	0.078982
14827	0.0	0.076545
14828	0.0	0.079968

14829	0.0	0.099275
14830	1.0	1.436533

14831 rows × 2 columns

```
In [59]:  plt.scatter(X_test, y_test,
color="gray") plt.plot(X_test, y_pred,
color="red", linewidth=2) plt.show()
```



Progress and Future Scope

Research shows that machine learning.

The algorithm demonstrated optimum accuracy and data processing speed compared to the conventional methods compared. However, other aspects such as high accuracy of predicting could not be maintained as that of other sophisticated algorithms.

Areas such as convolutional neural network and other artificial intelligence algorithms project will be studies in future for better accuracy and computation speed of massive smart meter dataset.

V. CONCLUSION

It is important that private and government power companies are able to recognize fraudulent electricity consumptions so that customers are not required to pay for the amount of power that they did not use.

This is a realistic electricity consumption dataset released by State Grid Corporation of China (http://www.sgcc.com.cn/). This dataset contains the electricity consumption data of 42,372 electricity customers within 1,035 days (from Jan. 1, 2014 to Oct. 31, 2016).

It contains numerical input variables and non-numerical NaN values or missing values. The column „CONS_NO" contains each customer"s smart meter ID. The feature 'Amount' is the total sum of daily electricity consumption of three years (from Jan. 1, 2014 to Oct. 31, 2016, this feature can be used for example-dependant cost-senstive learning. Feature 'FLAG' is the response variable and it takes value 1 in case of "FRAUD" and 0 otherwise.

This project was done for detection of abnormalities and electricity theft which have been a problem for decades. In this project a commonly known linear regression using sickit learn under machine learning algorithm was applied. Since the dataset has some errors, an estimated amount of 70% dataset was trained. Hence, the result indicates that if the predicted value is 1 as the actual value then the customer electricity consumption is "FRAUD" and if the predicted value is 0 then it is "NORMAL" case. Lastly, if the predicted value is more far from the margins and the customer should be suspected and necessary action should be taken. Therefore, the percentage of accuracy for 70% trained dataset is around 76%, this means that our algorithm for these specific-dataset still make reasonably good predictions.

VI. REFERENCES

[1]. Dahringer, N. (2017). Electricity Theft Detection using Machine Learning. arXiv preprint arXiv:1708.05907.

[2]. Dangar, D., & Joshi, S. K. (2014). Electricity Theft Detection Techniques for Distribution System in GUVNL. In International Journal Of Engineering Development And Research| Ijedr (Two Day National Conference (Rteece-2014)-January 2014).

[3]. Depuru, S. S. S. R. (2012). Modeling, detection, and prevention of electricity theft for enhanced performance and security of power grid

Doctoral dissertation, University of Toledo, United States).

[4]. Depuru, S. S. S. R., Wang, L., & Devabhaktuni, V. (2011). Electricity theft: Overview, issues, prevention and a smart meter based approach to control theft. Energy Policy, 39(2), 1007-1015.

[5]. Gaur, V., & Gupta, E. (2016). The determinants of electricity theft: An empirical analysis of Indian states. Energy Policy, 93, 127-136.

[6]. Han, W., & Xiao, Y. (2017). NFD: Non-technical loss fraud detection in smart grid. Computers & Security, 65, 187-201.

[7]. Jamil, F., & Ahmad, E. (2019). Policy considerations for limiting electricity theft in the developing countries. Energy Policy, 129, 452-458.

[8]. Jamil, F. (2013). On the electricity shortage, price and electricity theft nexus. Energy Policy, 54, 267-272.

[9]. Mashima, D., & Cárdenas, A. A. (2012, September). Evaluating electricity theft detectors in smart grid networks. In International Workshop on Recent Advances in Intrusion Detection (pp. 210-229). Springer, Berlin, Heidelberg.

[10]. Uparela, M. A., Gonzalez, R. D., Jimenez, J. R., & Quintero, C. G. (2018). An intelligent system for non-technical losses management in residential users of the electricity sector. Ingeniería e Investigación, 38(2), 52-60.

Cite this article as :

Saha, Utsho Chakraborty, Haimanti Biswas, Md. Intekhab Rahman Galib, Dr. Sheshang Degadwala, "Using Machine Learning Analytics to Detect Abnormalities and Electricity Theft", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 6, Issue 5, pp.271-279, September-October-2020. Available at doi : https://doi.org/10.32628/CSEIT206552
Journal URL : http://ijsrcseit.com/CSEIT206552