PREDICTION OF ELECTRICITY BILL USING SML TECHNIQUE

S. Muthukumaran, Headof the Department,
Abitha. C, Preethi. R, Final Year,
Department of Information Technology
St. Joseph College of Engineering, Sriperumbudur, Chennai.

Abstract:

Generally, predicting how the Electricity price will perform is one of the most difficult things to do. It can be described as one of the most critical processes to predict that. This is a very complex task and has uncertainties. To prevent this problem in One of the most interesting (or perhaps most profitable) time series data using machine learning techniques. Hence, Electricity price prediction has become an important research area. The aim is to predict machine learning based techniques for electricity price prediction results in best accuracy. The analysis of dataset by supervised machine learning technique (SMLT) to capture several information's like, variable identification, uni-variate analysis, bi-variate and multi-variate analysis, missing value treatments and analyze the data validation, data cleaning/preparing and data visualization will be done on the entire given dataset. To propose a machine learning-based method to accurately predict the Electricity price Index value by prediction results in the form of electricity price increase or stable state best accuracy from comparing supervise classification machine learning algorithms. Additionally, to compare and discuss the performance of various machine learning algorithms. dataset with evaluation classification report, identify the confusion matrix and to categorizing data from priority and the result shows that the effectiveness of the proposed machine learning algorithm technique can be compared with best accuracy MAE, MSE, R.

KEYWORDS – predict electricity bill, pre-processing, regression algorithm, feature extraction, Machine Learning.

Introduction:

Machine learning is the process of predicting the future based on historical data. Machine learning (ML) is an artificial intelligence (AI) technique that allows computers to learn without having to be explicitly programmed. Machine learning is concerned with the creation of computer programmes that can adapt to new data, as well as the fundamentals of machine learning, such as the construction of a simple machine learning algorithm in Python.

Specialized algorithms are used in the training and prediction process. It feeds the training data to an algorithm, which then applies the training data to new test data to make predictions. Machine learning can be divided into three distinct areas. There are three types of learning: supervised, unsupervised, and reinforced.

A supervised learning algorithm is given both the input data and the accompanying labelling to learn data, which must first be tagged by a person. There are no labels in unsupervised learning. It was made available to the learning algorithm. This method must determine how the input data is clustered. Finally, reinforcement learning interacts with its environment in a dynamic manner and receives positive or negative feedback in order to enhance its performance.

To uncover patterns in Python that lead to meaningful insights, data scientists utilise a variety of machine learning methods. These algorithms can be divided into two classes based on how they "learn" about data in order to generate predictions: supervised and unsupervised learning. The method of predicting the class of supplied data points is known as regression. Targets, labels, and categories are all terms used to describe classes. The task of approximating a mapping function from discrete input variables (X) to discrete output variables is known as regression predictive modelling (y).

Objectives:

The goal is to develop a machine learning model for Electricity Bill Prediction, to potentially replace the updatable supervised machine learning classification models by predicting results in the form of best accuracy by comparing supervised algorithm.

Literature Survey:

Review For Electricity Price Forecasting Techniques In Electricity Markets **Author:** Ankur Jain, Ankit Tuli, Misha Kakkar.. Electricity Price forecasting has been an important and crucial issue in every nation nowadays. This paper summarizes the various proposed models and techniques for the price forecasting. Accuracy of the price forecasting significantly influence the profits for transmission companies, distributors, suppliers etc. It presents an extensive review of different approaches for price forecasting. Different approaches and methods like ARIMA (Auto Regressive Integrated Moving Average), LSSVM (Least Square Support Vector Machine), LLWNN (Local Linear Wavelet Neural Network), ANN (Artificial Neural Network) etc. are compared and weakness and strengths of these methods are analyzed.

Price Forecasting for the Balancing Energy Market Using Machine-Learning Regression **Author:** Alexandre Lucas, Konstantinos Pegios, Evangelos Kotsakis and Dan Clarke. The importance of price forecasting has gained attention over the last few years, with the growth of aggregators and the general opening of the European electricity markets. Market participants manage a tradeoff between, bidding in a lower price market (day-ahead), but with typically higher volume, or aiming for a lower volume market but with potentially higher returns (balance energy market). Companies try to forecast the extremes of revenues or prices, in order to manage risk and opportunity, assigning their assets in an optimal way.

Energy Markets Forecasting. From Inferential Statistics to Machine Learning' **Author:**.Emma Viviani, Luca Di Persio and Matthias Ehrhardt. In this work, we investigate a probabilistic method for electricity price forecasting, which overcomes traditional ones. We start considering statistical methods for point forecast, comparing their performance in terms of efficiency, accuracy, and reliability, and we then exploit Neural Networks approaches to derive a hybrid model for probabilistic type forecasting. We show that our solution reaches the highest standard both in terms of efficiency and precision by testing its output on German electricity prices data.

System Design:

Data Wrangling:

In this section of the report will load in the data, check for cleanliness, and then trim and clean given dataset for analysis. Make sure that the document steps carefully and justify for cleaning decisions.

Data collection

The data set collected for predicting given data is split into Training set and Test set. Generally, 7:3 ratios are applied to split the Training set and Test set. The Data Model which was created using Random Forest, logistic, Decision tree algorithmsand Support vector classifier (SVC) are applied on the Training set and based on the test result accuracy, Test set prediction is done.

Preprocessing

The data which was collected might contain missing values that may lead to inconsistency. To gain better results data need to be preprocessed so as to improve the efficiency of the algorithm. The outliers have to be removed and also variable conversion need to be done.

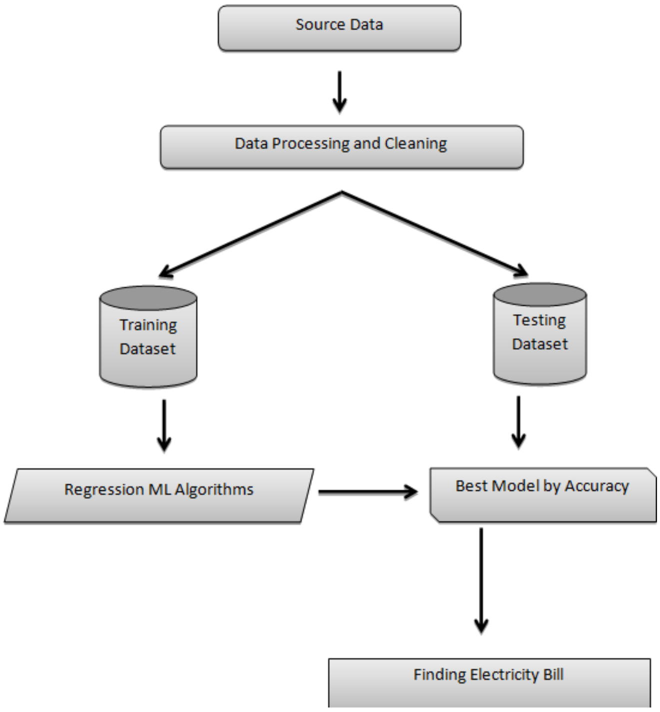
Building the Regression model

The prediction of Electricity Bill, A Random Forest Algorithm prediction model is effective because of the following reasons: It provides better results in regression problem.It is strong in preprocessing outliers, irrelevant variables, and a mix of continuous,categorical and discrete variables. It produces out of bag estimate error which has proven to be unbiased in many testand it is relatively easy to tune with

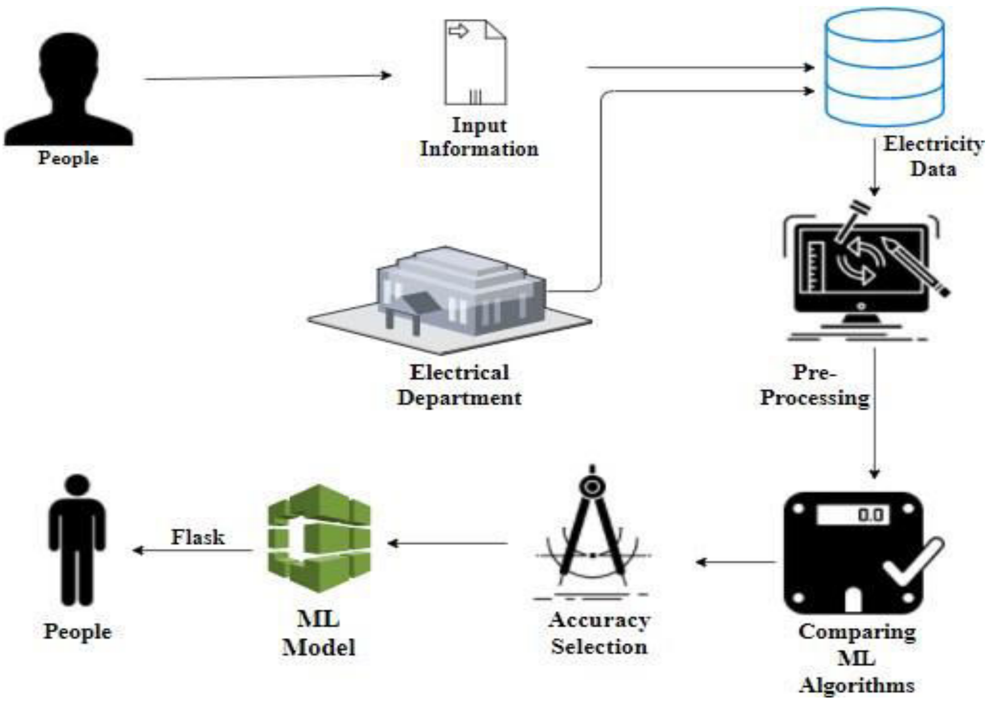
Construction of a Predictive Model

Machine learning needs data gathering have lot of past data's. Data gathering have sufficient historical data and raw data. Before data pre-processing, raw data can't be used directly. It's used to preprocess then, what kind of algorithm with model. Training and testing this model working and predicting correctly with minimum errors. Tuned model involved by tuned time to time with improving the accuracy.

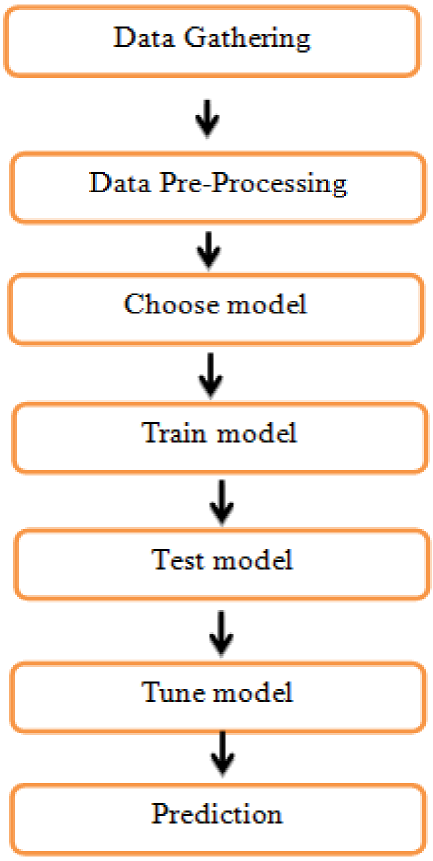
Work flow diagram



System Architecture



Dataflow Diagram



Implementation:

Algorithm Explanation:

```
#import library packages
importpandasaspd
importmatplotlib.pyplotasplt
importseabornassns
importnumpyasnp
importwarning

warnings.filterwarnings('ignore')
data=pd.read_csv("Household energy bill data.csv")

df=data.dropna()

df.columns
pd.crosstab(df.num_rooms,df.amount_paid)
pd.crosstab(df.num_rooms,df.is_ac)
#Histogram Plot of Amount distribution
df["amount_paid"].hist(figsize=(10,8),color="red")
plt.title("Amount Distribution")
plt.xlabel("Amount")
plt.ylabel("No of data")
plt.show()
#Histogram Plot of Income distribution
```

```
df["ave_monthly_income"].hist(figsize=(10,8),color="blue")
plt.title("Income Distribution")
plt.xlabel("Income")
plt.ylabel("No of data")
plt.show()
#barplot for Children and Cost
fig,ax=plt.subplots(figsize=(15,8))
sns.barplot(x="num_children",y="amount_paid",ax=ax,data=df)
plt.title("Number of Children vs Cost")
#Propagation by variable
defPropByVar(df,variable):
dataframe_pie=df[variable].value_counts()
ax=dataframe_pie.plot.pie(figsize=(10,10),autopct='%1.2f%%',fontsize=12)
ax.set_title(variable+' \n',fontsize=15)
returnnp.round(dataframe_pie/df.shape[0]*100,2)
PropByVar(df,'num_children')
fig,ax=plt.subplots(figsize=(15,6))
sns.boxplot(df.num_people,ax=ax)
plt.title("Number of People")
plt.show()

sns.pairplot(df)
plt.show()
# Heatmap plot diagram
fig,ax=plt.subplots(figsize=(15,10))
sns.heatmap(df.corr(),ax=ax,annot=True)
```

Spliting Train/Test:

```
#preprocessing, split test and dataset, split response variable
X=df.drop(labels='amount_paid',axis=1)
#Response variable
y=df.loc[:,'amount_paid']
#We'll use a test size of 30%. We also stratify the split on the response
variable, which is very important to do because there are so few fraudulent
transactions.
fromsklearn.model_selectionimporttrain_test_split
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.3,random_sta
te=1)
print("Number of training dataset: ",len(X_train))
print("Number of test dataset: ",len(X_test))
print("Total number of dataset: ",len(X_train)+len(X_test))

defqul_No_qul_bar_plot(df,bygroup):
dataframe_by_Group=pd.crosstab(df[bygroup],columns=df["is_urban"],normalize
='index')
dataframe_by_Group=np.round((dataframe_by_Group*100),decimals=2)
ax=dataframe_by_Group.plot.bar(figsize=(15,7));
vals=ax.get_yticks()
ax.set_yticklabels(['{:3.0f}%'.format(x)forxinvals]);
ax.set_xticklabels(dataframe_by_Group.index,rotation=0,fontsize=15);
ax.set_title('Number of Rooms or not by given attributes (%) (by
'+dataframe_by_Group.index.name+')\n',fontsize=15)
ax.set_xlabel(dataframe_by_Group.index.name,fontsize=12)
ax.set_ylabel('(%)',fontsize=12)
ax.legend(loc='upper left',bbox_to_anchor=(1.0,1.0),fontsize=12)
rects=ax.patches
```

The output for the prediction of electricity bill must be a clear cut one for the assumption of the electricity price which is getting from the place.

Conclusion:

The analytical process started from data cleaning and processing, missing value, exploratory analysis and finally model building and evaluation. The best accuracy on public test set is higher accuracy score is will be find out. This application can help to find the Electricity Bill.

Future Enhancement:

- Electricity Bill prediction to connect with AI model.
- To automate this process by show the prediction result in web application or desktop application at cloud.
- To optimize the work to implement in Artificial Intelligence environment.

References:

[1] ISO New England, "Market rule 1," 2019. [Online]. Available: www.iso-ne.com/participate/rules-procedures/tariff/market-rule-1

[2] PJM Interconnection, "Operating agreement of pjm interconnection, l.l.c." 2011. [Online]. Available: pjm.com/directory/merged-tariffs/oa.pdf

[3] F. C. Schweppe, *Spot pricing of electricity / by Fred C. Schweppe ... [et al.]*. Kluwer Academic Boston, 1988.

[4] V. Kekatos, G. B. Giannakis, and R. Baldick, "Online energy price ma- trix factorization for power grid topology tracking," *IEEE Transactions on Smart Grid*, vol. 7, no. 3, pp. 1239–1248, 2016.

[5] Q. Zhou, L. Tesfatsion, and C. Liu, "Short-term congestion forecasting in wholesale power markets," *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2185–2196, 2011.

[6] G. Hamoud and I. Bradley, "Assessment of transmission congestion cost and locational marginal pricing in a competitive electricity market," *IEEE Transactions on Power Systems*,vol. 19, no. 2, pp. 769–775, May 2004.

[7] W. Deng, Y. Ji, and L. Tong, "Probabilistic forecasting and simulation of electricity markets via online dictionary learning," 2016.

[8] J. F. Toubeau, T. Morstyn, J. Bottieau, K. Zheng, D. Apostolopoulou,Z. De Gre`ve, Y. Wang, and F. Valle´e, "Capturing spatio-temporal dependencies in the probabilistic forecasting of distribution locational marginal prices," *IEEE Transactions onSmart Grid*, pp. 1–1, 2020.

[9] X. Geng and L. Xie, "A data-driven approach to identifying system pattern regions in market operations," in *2015 IEEE Power Energy Society General Meeting*, July 2015, pp. 1–5.

[10] Y. Ji, R. J. Thomas, and L. Tong, "Probabilistic forecasting of real-time lmp and network congestion," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 831–841, March 2017.