

# Energy Meter Data Analysis Using Machine Learning Techniques

<sup>1</sup>K. Anitha Kumari, <sup>2</sup>M. Indusha, <sup>3</sup>D. Abarna Devi, <sup>3</sup>S. Dheva Dharshini

<sup>1</sup>Associate Professor, <sup>2</sup>PG Scholar, <sup>3,4</sup>UG Scholars, Dept. of IT

PSG College of Technology  
Coimbatore, India

**Abstract**— With the advancement of technology, existence of energy meters are not merely to measure energy units. The proliferation of energy meter deployments had led to significant interest in analyzing the energy usage by the machines. Energy meter data is often difficult to analyze owing to the aggregation of many disparate and complex loads. At utility scales, analysis is further complicated by the vast quantity of data and hence industries turn towards applying machine learning techniques for monitoring and measuring loads of the machines. The energy meter data analysis aims at analyzing the behavior of the machine and providing insights on usage of the energy. This will help the industries to identify the faults in the machine and to rectify it. Two use cases with two different motor specifications is considered for evaluation and the efficiency is proved by considering accuracy, precision, F-measure and recall as metrics..

**Keywords**- Energy Meter Data Analysis, Machine Learning Techniques, Energy Usage

## I. INTRODUCTION

In the late 1880's, the commercial use of electrical energy started its phase to full utilization. Today industries continue to remain the major consumer of energy around the world of the total commercial energy production. Significant opportunities emerge to optimize energy use and to minimize electrical energy waste in industries[8]. A manual recording of the electrical energy usage by the industrial machines is a nightmare for gaining insights. Hence the energy meter evolution offers significant opportunities to analyze the behavior of the machines by measuring and analyzing the electricity consumption of the industrial machines.

Energy meter or Watt-Hour Meter is an electrical instrument that measures the amount of electrical energy consumed by the consumers. It is used for measuring the energy utilities by the electric load. Utilities are one of the electrical departments that install these instruments at every place like homes, industries, organizations, commercial buildings to measure the electricity consumption of the industrial machines. Energy meters measure the rapid voltage and currents, calculate their product and provide instantaneous power. This power is integrated over a time interval that

presents the energy utilized over that period of time. Data analysis is the process of examining data sets in order to draw conclusions about the information they contain, increasingly with the aid of specialized systems and software. Commonly used techniques are K-nearest neighbor, decision tree, artificial neural network

and random forest. For further analysis of the data, machine learning approaches are applied. Based on the analysis of the results, faults in the machine are identified and leads to take any remedial actions if necessary.

## II. LITERATURE REVIEW

The literature survey mainly focuses on the applicability of naïve bayes, decision tree, k-nearest neighbor and Support vector machines (SVM) machine learning techniques for analyzing the energy meter data.

### A. Naive Bayes

This classifier functions independently that is each feature in the class is conditionally independent of every other feature of the class. The standard implementation of the naive Bayes' classifier [2] is used in the Weka toolkit [3].

### B. Decision Tree.

Decision tree classifier[4] is trained recursively by partitioning the input space and by defining a local model in each resulting region of the input space according to feature values. It is a good technique when there are missing values in a dataset. During the pre-processing phase one can encounter the challenge of having empty values in the dataset, and the flexibility that decision tree algorithms offer facilitates the data analysis. The approach breaks down a dataset into subsets by decision trees and later an associated decision tree is incrementally developed. Consider the following parameters to mathematically define a decision tree: A vector  $X$  of  $n$  characteristics as input, a

corresponding label  $Y$  as the output and a training set  $S$  that contains  $m$  couples  $(X, Y)$  as given in equation (2.1).

$$\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_m, \dots, \mathbf{x}_n)^T, \mathbf{X} \in \mathbf{R}^n$$

$$\mathbf{Y} \in \mathbf{R}$$

$$S = \{(X_1, Y_1), \dots, (X_n, Y_n)\} \quad (1)$$

A final tree with leaf nodes and decision nodes is the end result after running the algorithm. Also, decision trees can handle numerical data and categorical information, which makes it a very attractive model for various applications. Although finding the optimal data partitioning is NP-complete, greedy approximation algorithms perform well and benefit with very low training overhead.

### C. *k*-Nearest Neighbor (*k*-NN)

*k*-Nearest Neighbor classifies unknown examples based on distance measure like Euclidean to ascertain the distance from that example to every other training example [6]. The distance formula is given as

$$\text{Distance formula: } \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \quad (2)$$

The *k* smallest distances are identified and the most represented class in these *k* classes is considered as the output class label. The value of *k* is generally determined using a validation set or using cross-validation.

### D. Support Vector Machine.

Support vector machine is based on mapping the input feature space to a second linearly separable feature space using a kernel function. An SVM training algorithm builds a model that assigns new examples into one category or the other. New examples are then mapped into the same space and predicted to belong to a category based on which side of the gap they fall. SVM constructs a hyperplane or set of hyperplanes in a high-dimensional space or infinite-dimensional space, that can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier. The most commonly used package is the libSVM implementation [5] of SVMs supported through Weka. The performance of an SVM classifier is greatly dependent on the proper choice of a kernel function than other factors [7].

## III. PROPOSED SYSTEM

The energy meter data analysis aims at analyzing the data recorded by the energy meter connected to the machines. This project contains two main divisions: data collection and data analytics.

### A. Data Collection

The energy meter data is vast in nature that it records 32 parameters for every 20 seconds interval of time. These recorded

data are sent to the Amazon Simple Storage Service (S3) cloud using IoT gateways. The real-time streaming data triggers the machine learning code written in AWS Lambda, a computing service to run machine learning algorithms. The energy meter data collected from the S3 bucket can be used for further data analytics to get the insights.

### B. Data Analysis

Once the data reaches Amazon S3 cloud service, it is analyzed by using Amazon Lambda with the help of boto3 software development kit for reporting the faults in the machine. The machine learning approach is used to classify the data as in-range or out-range. Classification is the process of predicting the class of given data points. Classes are sometimes called as targets/labels or categories. Classification belongs to the category of supervised learning where the targets also provided with the input data. Two machine learning approaches used for detecting the out-range are *k*-nearest neighbor and support vector machine. The out-range data are then used to draw conclusions about the behavior of the machine.

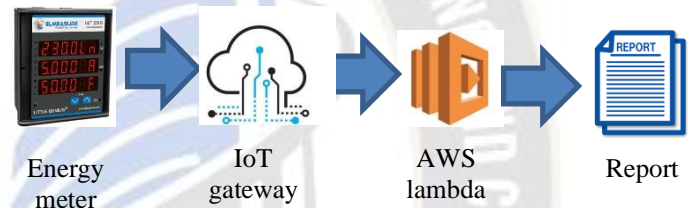


Fig. 1. Overall workflow

Fig. 1 describes the overall workflow of the proposed system. The energy meter records the parameters of the machine and sends it to S3 cloud through IoT gateway. The data collected from the cloud is analyzed using Amazon Lambda for reporting the faults in the machine. The generated report provides an insight of status of the machine and facilitates to identify whether any fault exists or not.

The data set collected for the proposed system is shown in Fig 2.

Date,Time, Voltage, Current, Motor(rmp), vibration, Power factor, Feature
02-01-2019,11:00:00 AM,409.7,560.6,504,0.066721218,0.957,OUT RANGE
02-01-2019,11:00:20 AM,419.6,531.7,863,0.065217462,0.985,OUT RANGE
02-01-2019,11:00:40 AM,411.4,534.7,663,0.069097502,0.96,OUT RANGE
02-01-2019,11:01:00 AM,410.5,526.1,741,0.051837424,0.955,OUT RANGE
02-01-2019,11:01:20 AM,408.9,551.7,719,0.068325587,0.986,OUT RANGE
02-01-2019,11:01:40 AM,417.4,533.6,722,0.076003411,0.975,OUT RANGE
02-01-2019,11:02:00 AM,420.5,61.4,794,0.064422328,0.952,OUT RANGE
02-01-2019,11:02:20 AM,419.9,559,592,0.068936092,0.98,OUT RANGE
02-01-2019,11:02:40 AM,413.2,538.9,695,0.068829919,0.939,OUT RANGE
02-01-2019,11:03:00 AM,414.3,553.9,620,0.060761644,0.952,OUT RANGE
02-01-2019,11:03:20 AM,414.7,530.4,586,0.059540335,0.935,OUT RANGE
02-01-2019,11:03:40 AM,414.6,554.5,530,0.06599821,0.991,OUT RANGE
02-01-2019,11:04:00 AM,414.9,560.1,642,0.058114116,0.956,OUT RANGE
02-01-2019,11:04:20 AM,412.2,530.7,531,0.054149064,0.997,OUT RANGE
02-01-2019,11:04:40 AM,416.7,561.1,498,0.051288372,0.974,OUT RANGE
02-01-2019,11:05:00 AM,409.2,536.6,664,0.050719875,0.959,OUT RANGE
02-01-2019,11:05:20 AM,418.6,549.9,494,0.054794437,0.961,OUT RANGE
02-01-2019,11:05:40 AM,420.4,531.7,506,0.052530393,0.936,OUT RANGE

Fig. 2. Sample dataset

#### IV. USE CASES AND RESULT ANALYSIS

##### A. Use Cases

All electric motors have a predetermined life span. However, this is dependent on proper maintenance without which they are likely to break down much quicker and prone to failures. A deeper understanding of the causes of motor failure and its associated risks prompts to take certain spontaneous actions to enhance the durability and reliability without wear and tear. Use cases considered in the work are overheating, copper loss and vibration as they heavily disturb the lifespan of the machines.

##### a. Overheating

Around 55% of insulating failures in motors occur due to overheating. One must make sure that motor wirings are properly sized and fixed inside the motor. When a motor is too small, it is not possible to dissipate heat at a rapid rate and the motor results in overheat. Excessive heat is caused by excessive current flow within the motor windings by exceeding the acceptable range. Overheating causes major damage to windings and bearings of the motor. It can be prevented by installing effective over-current protection/prevention devices that will detect overcurrent and interrupt supply.

A heat map is a two-dimensional representation of data in which values are represented by colors. A heat map uses a warm-to-cool color spectrum to show the heat dissipation of the 3-phase induction motor. The motor must be kept as cool as possible, ensuring the operating environment is kept cool that may help to prevent breakdowns.

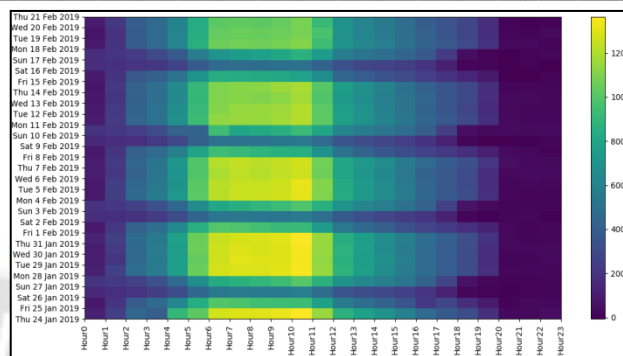


Fig. 3. Heat Map of Induction Motor

In Fig. 3 the vertical scale shows the heat dissipated by an induction motor at different levels.

##### b. Copper Loss

It is the heat produced by electric currents in the conductors of electrical devices. Copper losses increase when the electrical current passing through the conductor's increases. It can be minimized by using conductors of large diameters to reduce the resistance per unit length of the conducting windings of the electrical device. The conductivity of a transmission line can be enhanced by using cables with multiple insulated conductors. Hollow conductors can reduce the cost of transmission.

##### c. Vibration

Vibration causes many issues with the motor, and can eventually cause the motor to fail prematurely. Vibration is often caused by the motor being positioned on an uneven or unstable surface. Vibration can also be a result of reduced lubrication in motors due to less maintenance which results in dust formation. Hence the windings are loosened and cause mechanical damage like misalignment or corrosion. The excessive movements it creates results the lead wires to become brittle. As a solution, motors should be checked regularly for vibration, using a motor analyzing tool such as the SKF EXP4000 Dynamic Motor Analyzer. In order to reduce vibration, one must ensure that the motor is positioned on a flat, stable surface. If vibration still occurs, further verification may be done for signs of wear as well as for loose bearings or misalignment.

##### B. Result Analysis

The performances of the SVM and *k*-nearest neighbor algorithms are measured in terms of four measures: accuracy, F-score, recall and precision. The dataset was split into three ratios as 75:25, 70:30, 60:40 where 75, 70, 60 denotes the amount of data used for training and 25, 30, 40 denotes the amount of data used for testing. The output of SVM and *k*-nn is classified into two different labels as in-range and out-range. When the value is in between 95% and 105% of actual value, it is considered that

the motor is functioning properly and classified as in-range. On the other hand, when the value falls below 95% or above 105%, it is classified as out-range. Energy meter data of two different motor are used for analysis and their specifications are given below in TABLE I

TABLE I. Motor specification

Motor-Id	Current (amp)	Rpm	Current (amp) 95%	Rpm 105%	Current (amp) 105%	Rpm 95%
Motor 1	510	935	485	981	535	888
Motor 2	340	1390	323	1459	357	1320

For three different ratios, the accuracy, recall, precision and F-score values are measured for Support Vector Machine when C=0.01, C=0.1 for the motor-1 as presented in TABLE II and TABLE III respectively. The parameter C decides how much one wants to penalize the misclassified points.

TABLE II. Performance of the Support Vector Machine when C=0.01 of motor-1

Evaluation metrics	Split Size		
	Training size=75% Testing size=25%	Training size=70% Testing size=30%	Training size=60% Testing size=40%
Precision (%)	57.08	46.94	39.06
Recall (%)	75.56	68.51	62.5
F- score (%)	65.03	55.71	48.07
Accuracy (%)	75.56	68.51	62.5

TABLE III. Performance of the Support Vector Machine when C=0.1 of motor-1

Evaluation metrics	Split Size		
	Training size=75% Testing size=25%	Training size=70% Testing size=30%	Training size=60% Testing size=40%
Precision (%)	95.80	94.86	93.75
Recall (%)	95.55	94.44	93.06

F- score (%)	95.39	94.28	92.88
Accuracy (%)	95.55	94.44	93.06

While comparing the performance, the evaluation metrics are found to perform better when C=0.1 for motor-1

Similarly, for three different ratios, the accuracy, recall, precision and F-score values of k-nearest neighbor for k=3, k=7 for the motor-1 are listed in TABLE IV, TABLE V respectively.

TABLE IV. Performance of the k-nearest neighbor when k=3 for motor-1

Evaluation metrics	Split Size		
	Training size=75% Testing size=25%	Training size=70% Testing size=30%	Training size=60% Testing size=40%
Precision (%)	97.84	94.86	96.09
Recall (%)	97.78	94.44	95.83
F- score (%)	97.74	94.28	95.77
Accuracy (%)	97.78	94.44	95.83

TABLE V. Performance of the k-nearest neighbor when k=7 for motor-1

Evaluation metrics	Split Size		
	Training size=75% Testing size=25%	Training size=70% Testing size=30%	Training size=60% Testing size=40%
Precision (%)	97.84	98.19	97.34
Recall (%)	97.78	98.14	97.22
F- score (%)	97.74	98.13	97.19
Accuracy (%)	97.78	98.14	97.22

While comparing the performance measure, the evaluation metrics are found to perform better when k=7 for motor-1.

For the three different ratios, the accuracy, recall, precision and F-score values of Support Vector Machine for C=0.01, C=0.1 for the motor-2 are listed in TABLE VI, TABLE VII respectively.

TABLE VI. Performance of the Support Vector Machine when C=0.01 for motor-2

Evaluation metrics	Split Size		
	Training size=75% Testing size=25%	Training size=70% Testing size=30%	Training size=60% Testing size=40%
Precision (%)	71.30	75.75	60.49
Recall (%)	84.44	87.03	77.77
F- score (%)	77.32	81.00	68.05
Accuracy (%)	84.44	87.03	77.77

TABLE VII. Performance of the Support Vector Machine when C=0.1 for motor-2

Evaluation metrics	Split Size		
	Training size=75% Testing size=25%	Training size=70% Testing size=30%	Training size=60% Testing size=40%
Precision (%)	98.05	97.11	91.99
Recall (%)	97.77	96.29	87.5
F- score (%)	97.83	96.48	88.32
Accuracy (%)	97.77	96.29	87.5

While comparing the performance evaluation, the metrics are found to perform when C=0.1 for motor-2 Accuracy, recall, precision and F-score values of  $k$ -nearest neighbor for  $k=3$ ,  $k=7$  for the motor-2 listed in TABLE VIII, TABLEIX respectively.

TABLE VIII. Performance of the  $k$ -nearest neighbor when  $k=3$  for motor-2

Evaluation metrics	Split Size		
	Training size=75% Testing size=25%	Training size=70% Testing size=30%	Training size=60% Testing size=40%
Precision (%)	94.34	94.59	96.09
Recall (%)	91.11	90.74	95.83
F- score (%)	91.85	91.69	95.77

Accuracy (%)	91.11	90.74	95.83
--------------	-------	-------	-------

TABLE IX. Performance of the  $k$ -nearest neighbor when  $k=7$  for motor-2

Evaluation metrics	Split Size		
	Training size=75% Testing size=25%	Training size=70% Testing size=30%	Training size=60% Testing size=40%
Precision (%)	98.05	97.11	97.35
Recall (%)	97.77	96.29	97.22
F- score (%)	97.83	96.48	97.27
Accuracy (%)	97.77	96.29	97.22

While comparing the performance evaluation, the evaluation metrics are found to perform better when  $k=7$  for motor-2. With the above specification and evaluation of both the motors the  $k$ -nearest neighbor algorithm is performing comparatively better than SVM.

## V. CONCLUSION AND FUTURE ENHANCEMENTS

Manual recordings of the energy meter data are very difficult and time-consuming. With the help of machine learning approaches and services provided by Amazon Web Service (AWS), the analysis, storage of the meter data and gaining timely insights have been made easier. Current work is particular to two induction motors. Future enhancements include induction motors of various specifications and including use cases like power quality and overloading of the induction motor. In addition, other machine learning techniques can be tested for the effectiveness in identifying the faults with industrial machines.

## REFERENCES

- [1]. Barker, Sean K., "Model-Driven Analytics of Energy Meter Data in Smart Homes" (2014). Doctoral Dissertations.
- [2]. John, George H., and Langley, Pat. "Estimating continuous distributions in Bayesian classifiers". *Eleventh Conference on Uncertainty in Artificial Intelligence* (San Mateo, 1995), Morgan Kaufmann, pp. 338–345.
- [3]. Hall, Mark, Frank, Eibe, Holmes, Geoffrey, Pfahringer, Bernhard, Reutemann, Peter, and Witten, Ian H. The weka data mining software: An update. *SIGKDD Explor. Newsl.* (Nov 2009).
- [4]. Murphy, Kevin P. Machine learning: "A probabilistic perspective". The MIT Press, 2012.

- [5]. Chang, Chih-Chung, and Lin, Chih-Jen. Libsvm: "A library for support vector machines". TIST 2, 3 (2011).
- [6]. T. Cover, P. Hart, "Nearest neighbor pattern classification", IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21–27, Jan. 1967
- [7]. M. Hussain, S.K. Wajid, A. Elzaart, M. Berbar, "A Comparison of SVM Kernel Functions for Breast Cancer Detection", *Proceeding Eighth International Conference on Computer Graphics Imaging and Visualization 2011*, pp. 145-150, August 2011.
- [8]. Y. Wang, Q. Chen, T. Hong and C. Kang, "Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges," in *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 3125-3148, May 2019.
- [9]. Zemene, Fikiret and Khedkar, Vijayhsri, "Survey on Machine Learning based Electric Consumption Forecasting using Smart Meter Data" in *International Journal of Computer Applications.*, vol. 180, no. 6, pp.46-52, Decenmber 2017.
- [10]. Alahakoon, Damminda and Xinghuo Yu, "Smart Electricity Meter Data Intelligence for Future Energy Systems", in *IEEE Transactions on Industrial Informatics*, vol.12, no.1, February 2016.
- [11]. Madeline Martinez, Timothy Eveleigh and Bereket Tanju, "Smart meter data analytics for optimal customer selection in demand response programs", in *3rd International Conference on Energy and Environment Research*, vol.7, no. 11, pp. 49-59, September 2016.
- [12]. Mariya, Sodenkamp, Ilya, Kozlovskiy, Konstantin, Hopf, Thorsten Staake, "Smart Meter Data Analytics for Enhanced Energy Efficiency in the Residential Sector", in *13th International Conference on Wirtschaftsinformatik*, pp-1235-1249, February 2017.
- [13]. Akinsola, J E T, "Supervised Machine Learning Algorithms: Classification and Comparison. *International Journal of Computer Trends and Technology*", vol.48, pp. 128 – 138, June 2017.
- [14]. S. Haben, C. Singleton and P. Grindrod, "Analysis and Clustering of Residential Customers Energy Behavioral Demand Using Smart Meter Data," in *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 136-144, January 2016.
- [15]. N.Lu, P. Du, X. Guo and F. L. Greitzer, "Smart meter data analysis," in *PES T&D*, pp. 1-6, May 2012.
- [16]. J. Kwac, J. Flora and R. Rajagopal, "Household Energy Consumption Segmentation Using Hourly Data," in *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 420-430, January 2014.
- [17]. Young TaeChaea, RayaHoresh, Youngdeok Hwang, Young M.Leeb, "Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings", vol.111, pp. 184-194, January 2016.
- [18]. F. L. Quilumba, W. Lee, H. Huang, D. Y. Wang and R. L. Szabados, "Using Smart Meter Data to Improve the Accuracy of Intraday Load Forecasting Considering Customer Behavior Similarities," in *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 911-918, March 2015.
- [19]. A. Monacchi, D. Egarter, W. Elmenreich, S.D.Alessandro and A. M. Tonello, "GREEND: An energy consumption dataset of households in Italy and Austria," in *IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pp. 511-516, November 2014.

## BIOGRAPHY



**Dr K Anitha Kumari** is working as an Associate Professor in the Department of IT in PSG College of Technology, India. She is Highly Passionate and curious about Learning New stuff in Security Protocols. As an Independent Researcher, she had an Opportunity to present her UGC sponsored paper

based on Quantum Cryptography in USA and visited a few Foreign Universities. To her credit, she had filed a PATENT and published around 55 Technical Papers in refereed and Impact Factored International/National Journals/Conferences published by Elsevier, Springer, T&F, etc., Also, she's been an Active Reviewer for Prestigious Journals published by IEEE (IEEE Communications Surveys and Tutorials (IF: 20.230), IEEE Transactions on Industrial Informatics (IF: 5.43)), Springer, Wiley, etc., and Technical Program Committee (TPC) for CECNet 2017, NGCT-2017 WICC-2018 and NCCI-2018 conferences. Her areas of interest include Cloud & IoT Security, Design and Analysis of Security Protocols, Attacks & Defense, Security in Computing, Bioinformatics, Cognitive Security, Quantum Cryptography, Web Service Security, Network Security and Analysis of Algorithms. Out of her research interest, she has contributed a Book chapter in T & F and delivered ample Guest Lectures. Her security project is sanctioned and granted by AICTE for a sum of Rs.11,80,000/-. She's been the mentor for Technovator Projects (2018 & 2014) and 'MEDROIDZ', an ICICI – Trinity 2014 funded project that was selected as one among the 6 projects in India. Academically, she has secured RANK-I and awarded Gold Medal in ME (SE) & in BE (CSE) from Anna University and from Avinashilingam University. She secured Elite+Gold Medal (Top 1%) in NPTEL – Cloud Computing course. She also won prizes in intra and inter institutional cultural events. As a Supervisor, she is currently guiding PhD scholars. You may contact her at [anitha.psgsoft@gmail.com](mailto:anitha.psgsoft@gmail.com) / [kak.it@psgtech.ac.in](mailto:kak.it@psgtech.ac.in).