Report

***“Optimizing Power Consumption on Home Appliances using Machine Learning”***

**By,**

Sudhanva Narayana,

Intern, Machine Learning Engineer,

**Phone:** +91 9632350260

**Email:** [nsudhanva@gmail.com](mailto:nsudhanva@gmail.com)

**Faststream Technologies,**

6, 2nd Main Rd, Arekere Off Bannerghatta Road, Bengaluru, KA, 560076

**Phone:** +91 80–4161–3204/3205

**Email:** [info@faststreamtech.com](mailto:info@faststreamtech.com)

|  |  |
| --- | --- |
| **Table of Contents** | **Page** |
| **Executive Summary** | **No** |
| 1. **Introduction**   *1.1. Background*  *1.2. Home Automation*  *1.3. Machine Learning*  *1.4. Project Objectives*  *1.5. Dependencies and Tools*  *1.6. Limitations of the project* | **1**  2  3  4  5  6  6 |
| 1. **Project Methodologies**   *2.1.*  *2.2.*  *2.3. Applying Machine Learning model to predict intensities* | **7**  10  10  11 |
| 1. **Preparing Dataset**   *3.1.*  *3.2.*  *3.3.* | **12**  13  14  16 |
| 1. **Prediction**   *4.1.*  *4.2.* | **19**  20  23 |
| 1. **Custom Algorithm**   5.1.  *5.2.*  *5.3.* | **25**  26  26 |
| 1. **Results**   *6.1. Performance of algorithms*  *6.2. Conclusion* | **51**  52  58 |
| **References** | **59** |

**1. INTRODUCTION**

Power (electricity) optimization will be a crucial task in the coming years as there will be a limited supply. While we run out of electricity, it will affect both domestic and industrial applications. This report demonstrates techniques and strategies to optimize power consumption when there is a limited supply, using Machine Learning to predict the amount of electricity consumed at a given hour for a given device specifically for home appliances. It also notifies the user of excess power consumption and also suggests measures to save power. It also considers factors like weather, time of the day and type of the device and the device priority to make the final decision. Machine Learning algorithms such as Multiple Linear Regression, Decision Tree Regression, Random Forest Regression are used to predict the values. We present the result of each algorithm by showing in how much each algorithm is better (or worse) compared to the rest of the algorithms.

***1.1. Background***

Faststream Technologies is a vanguard of technology solutions, specializing in Product & System Engineering, IoT, Big Data, Security, and Application Development with a global footprint across North America, EMEA, and APAC. With over 200+ clients, Faststream Technologies enables Digital Transformation for enterprises by delivering a flawless customer experience, business competence, and deep insights through an integrated set of disruptive technologies and expertise. We are passionate about delivering well-organized, inventive and world-class hardware and software solutions, with a focus on Healthcare, Aerospace, Semiconductors, Automotive, Consumer Electronics, Home Automation, Telecommunications, Security, Retail, and E-Commerce.

Faststream Technologies works at the juncture of business and technology, assisting clients with advancing their product and business performance through sustainable information technology solutions. Faststream Technologies drives innovation to help clients advance their product design, business processes, and application development. Our engineering team’s deep expertise in transforming design specs into marketable hardware products — through ASIC design services that include RTL design, design verification and physical design for digital and analogue/mixed-signal semiconductors — is a key differentiator to our suite of application development capabilities.

For today’s challenges like embedded processor SoC specifications, Faststream Technologies delivers all of the required firmware/embedded software, positioning us as the turnkey ‘concept-to-product’ design company. The team is led by a group of focused senior executives and Technologists who complement each other with significant industry experience in building turnkey solutions. Many of our technologists have multiple patents to their credit in the areas of Analog/Mixed-Signal Design, IoT and embedded systems.

***1.2. Home Automation***

Home automation or demotics is building automation for a home, called a smart home or smart house. A home automation system will control lighting, climate, entertainment systems, and appliances. It may also include home security such as access control and alarm systems. When connected with the Internet, home devices are an important constituent of the Internet of Things. A home automation system typically connects controlled devices to a central hub or "gateway". The user interface for control of the system uses either wall-mounted terminals, tablet or desktop computers, a mobile phone application, or a Web interface, that may also be accessible off-site through the Internet. While there are many competing vendors, there are very few worldwide accepted industry standards and the smart home space is heavily fragmented. Manufacturers often prevent independent implementations by withholding documentation and by litigation. The home automation market was worth US$5.77 billion in 2013, predicted to reach a market value of US$12.81 billion by the year 2020.

***1.3. Machine Learning***

Machine Learning is the science (and art) of programming computers so they can learn from data. For example, your spam filter is a Machine Learning program that can learn to flag spam given examples of spam emails (e.g., flagged by users) and examples of regular (non-spam) emails. The examples that the system uses to learn are called the training set. Each training example is called a training instance (or sample). In this case, the task T is to flag spam for new emails, the experience E is the training data, and the performance measure P needs to be defined; for example, you can use the ratio of correctly classified emails. This particular performance measure is called accuracy and it is often used in classification tasks.

***Supervised Learning***

Machine Learning systems can be classified according to the amount and type of supervision they get during training. There are four major categories: supervised learning, unsupervised learning, semi-supervised learning, and Reinforcement Learning. In supervised learning, the training data you feed to the algorithm includes the desired solutions, called labels. A typical supervised learning task is classification. The spam filter is a good example of this: it is trained with many examples emails along with their class (spam or ham), and it must learn how to classify new emails.

***Unsupervised learning***

In unsupervised learning, as you might guess, the training data is ranged. The system tries to learn without a teacher.

* **Clustering** 
  + k-Means
  + Hierarchical Cluster Analysis (HCA)
  + Expectation Maximization
* **Visualization and dimensionality reduction** 
  + Principal Component Analysis (PCA)
  + Locally-Linear Embedding (LLE)
  + t-distributed Stochastic Neighbour Embedding (t-SNE)

***1.4. Project Objectives***

We use multiple regression techniques to predict power consumption based on previously given data. After the values are predicted, a custom algorithm is used to limit power consumption by identifying devices that are consuming more power, by type of weather, by rooms and notify the user of its excess power consumption. It also considers factors like a number of people in the room and time they have stayed in that room. In this way, the total power consumption of devices is reduced per month. However, the power consumption per day is not considered limited. Machine Learning is used only to predict the power consumption at that point in time and not to predict or understand the behaviour of people. As there is no real data, we used a random number generation and time series generator to generate a dataset of certain attributes on which we used to train our models. We later used another generated dataset to predict values from the trained model using regression.

Regression is a set of statistical processes for estimation of the relationship among data points and variables. It predicts the conditional expectation of the dependent variable when given the independent variables, that is the average of both dependent and independent variables. A function of the independent variables is estimated and the values are predicted using probability distribution.

**List of regression algorithms used:**

* Multiple Linear Regression
* Random Forest Regression
* Decision Tree Regression

*The result is interpreted and plotted measuring the performance of the mentioned algorithms above.*

***1.5. Dependencies and Tools***

1. Python – a general-purpose interpreted, interactive, object-oriented, and high-level programming language
2. Anaconda – a free and open source distribution of the Python and R programming languages for data science and machine learning related applications, that aims to simplify package management and deployment
3. Numpy – the fundamental package for scientific computing with Python.
4. Scipy - a Python-based ecosystem of open-source software for mathematics, science, and engineering
5. Pandas - pandas is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language.
6. Matplotlib – a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms
7. Sci-kit learn – a free software machine learning library for the Python programming language.

***1.6. Limitations of the project***

The images are generated using numpy’s random number generator and mesh-grid technique which uses sine and cosine function to generate contour like matrices. However, these generated images are not real and should be used for experimental purposes only. The accuracy, methodology of the machine learning algorithms is true whereas the images are not. The project is also dependent on certain Python environments and related tools. It is independent of the development environment.

**2. PROJECT METHODOLOGIES**

***Steps involved***

* Something

Data is checked for validity, accuracy, completeness and consistency. Data auditing is made to detect anomalies and contradictions. The detection and removal of anomalies is performed by a sequence of operations on the data known as the workflow. After executing the cleansing workflow, the results are inspected to verify correctness. Data that could not be corrected during execution of the workflow is manually corrected.

**2.1. Something**

**2.2. Something**

**2.3. Applying Machine Learning model to data**

1. Training set is determined. Here, 75% of the generated dataset is used as training data and the rest 25% is test data.
2. The training set represents the real-world use if the function. A set of input objects is gathered and corresponding outputs are also gathered.
3. The input representation of the learn function is determined and the accuracy of the learned function depends strongly on how the input object is represented. The input object is transformed into a feature vector, which contains a number of features that are descriptive of the object.
4. The structure of the learned function is determined (any Machine Learning algorithm can be used)
5. The design (model) is completed and I run on the gathered training set. Some of the supervised algorithms require the user to determine certain control parameters. These parameters are adjusted by optimising performance on the test set of the validation set and cross validation is also applied
6. Accuracy of the learned function is determined. After the parameter is adjusted, the performance of the resulting function is measured on the test set that is separate from the training set.

**3. PREPARING DATASET**

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Maecenas porttitor congue massa. Fusce posuere, magna sed pulvinar ultricies, purus lectus malesuada libero, sit amet commodo magna eros quis urna.

Nunc viverra imperdiet enim. Fusce est. Vivamus a tellus.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin pharetra nonummy pede. Mauris et orci.

Aenean nec lorem. In porttitor. Donec laoreet nonummy augue.

Suspendisse dui purus, scelerisque at, vulputate vitae, pretium mattis, nunc. Mauris eget neque at sem venenatis eleifend. Ut nonummy.

**3.1. Something**

**3.2. Something**

**3.3. Something**

**4. PREDICTION**

Cross-validation, sometimes called rotation estimation, or out-of-sample testing is any of various similar model validation techniques for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in the goal prediction, we estimate how accurately a predictive model will perform in practice. In a prediction problem, a model is usually given a dataset of known data on which training is run (training dataset), and a dataset of unknown data (or first seen data) against which the model is tested (called the validation dataset or testing set). The goal of cross-validation is to test the model’s ability to predict new data that were not used in estimating it, in order to flag problems like overfitting and to give an insight on how the model will generalize to an independent dataset (i.e., an unknown dataset, for instance from a real problem).

**5. RESULTS**

**5.1. Something**

**5.2. Conclusion**

The performance data is collected based on the confusion matrix produced by the algorithms. In the field of machine learning and specifically the problem of statistical classification, a confusion matrix, also known as an error matrix, is a specific table layout that allows visualization of the performance of an algorithm.

Kernel SVM and K Nearest Neighbours tabulated similar results. Logistic Regression and SVM showcased a significant improvement in the results. However, Decision Tree and Random Forest classifiers performed better than linear classifiers. Random Forest and Decision Tree both topped at 99.5% making them the best algorithm to use for this kind of dataset. Overall the results are similar in nature. We come to the conclusion that for a dataset of 1,000 rows, there are major differences in accuracy on machine learning models.

**References**

Something