DOCUMENT : MEASURE ENERGY CONSUMPTION

PROBLEM DEFINITION:

The problem at hand is to create an automated system that measures energy consumption, analyzes the data, and provides visualizations for informed decision-making. This solution aims to enhance efficiency, accuracy, and ease of understanding in managing energy consumption across various sectors. The ratio of the energy consumption (E) of a society to its economic output (gross domestic product, GDP), measured in dollars of constant purchasing power (the E/GDP ratio).

DESIGN THINKING

❖ Data source :

A data source may be the initial location where data is born or where physical information is first digitized, however even the most refined data may serve as a source, as long as another process accesses and utilizes it. Concretely, a data source may be a database, a flat file, live measurements from physical devices, scraped web data, or any of the myriad static and streaming data services which abound across the internet

❖ Data preprocessing :

Data preprocessing is an important step in the data mining process. It refers to the cleaning, transforming, and integrating of data in order to make it ready for analysis

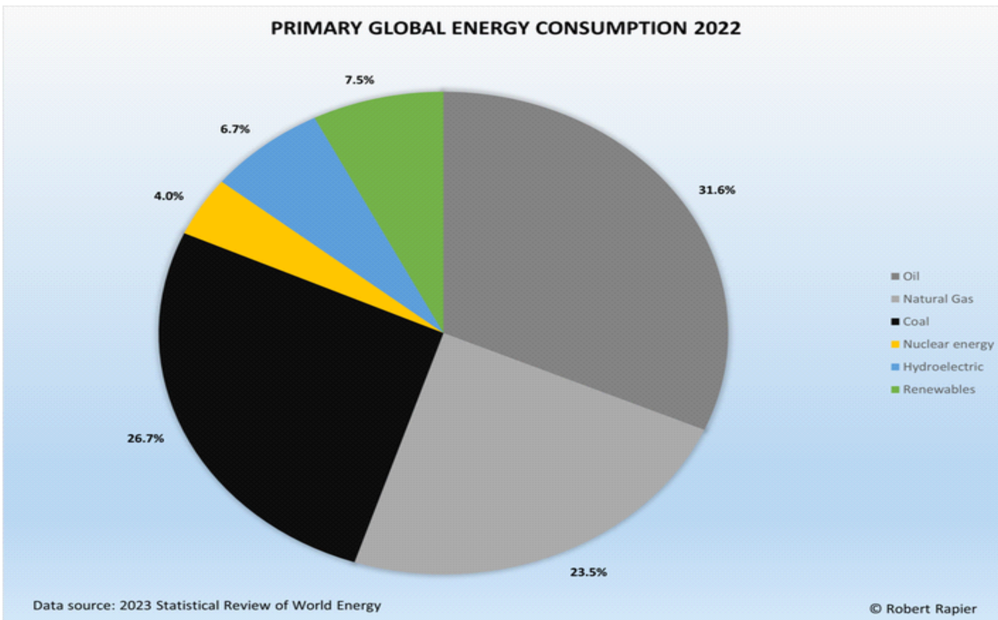
❖ Feature extraction :

Feature extraction is a process of dimensionality reduction by which an initial set of raw data is reduced to more manageable groups for processing.

❖ Model development :

Model development is an iterative process, in which many models are derived, tested and built upon until a model fitting .

❖ visualization



-

❖ Automation :

Automation is the use of technology to perform tasks with where human input is minimized

CONCLUSION :

Sentiment analysis is a marketing tool that helps you examine the way people interact with a brand online. This method is more comprehensive than traditional online marketing tracking, which measures the number of online interactions that customers have with a brand

INNOVATE TECHNIQUES SUCH AS ENSEMBLE METHOD:

The most popular ensemble methods are boosting, bagging, and stacking. Ensemble methods are ideal for regression and classification, where they reduce bias and variance to boost the accuracy of models.

DEEP LEARNIN ARCHITECTURE:

Deep-learning architectures such as deep neural networks, deep belief networks, deep reinforcement learning, recurrent neural networks, convolutional neural networks and transformers have been applied to fields including computer vision, speech recognition, natural language processing, machine translation.

TO IMPROVE THE PREDRCTION SYSTEM ACCURACY:

Linear Regression: Linear regression is perhaps one of the most well-known and well-understood algorithms in statistics and machine learning. Predictive modeling is primarily concerned with minimizing the error of a model or making the most accurate predictions possible, at the expense of explainability.

TO IMPROVE THE ROBUSTNESS:

A common ove model robustness is adversarial training which follows two steps-collecting adversarial examples by attacking a target model, and fine-tuning the model on the augmented dataset with these adversarial examples

TIME SERISE ANALYSIS:

Time series analysis is a specific way of analyzing a sequence of data points collected over an interval of time. In time series analysis, analysts record data points at consistent intervals over a set period of time rather than just recording the data points intermittently or randomly.

MACHINE LEARNING MODELS:

Factoring performance, accuracy, reliability and explainability, data scientists consider supervised, unsupervised, semi-supervised and reinforcement models to reach best outcomes.

MACHINE LEARNING MODELS TO PREDICT FUTURE ENERGY CONSUMPTION PATTERNS:

The five machine learning methods used are multi-layer perceptron (MLP), support vector machine (SVM), radial basis function (RBF) regressor, REPTree, and Gaussian process.

DATA PREPROCESSING STEP:

By the developments on the measurement and communication infrastructures in power systems, it has become possible to collect data from more points and with higher resolutions compare to the past. Increasing data volume, on the one hand, increases the quality of the information possessed, on the other hand, it has made the processing of data more complicated [16]. With the increase in data volume, the size and variety of data quality problems has also increased. The success of data analysis is closely related to data quality. In order to obtain consistent results, missing or outlier data must be determined and removed from the data sets, and the data should be formatted in accordance with the study.

Data integration:

In general, electrical energy consumption clustering studies are based on consumption data only. However, in some studies, various data affecting electricity consumption can also be included in the analysis. In such multivariate studies,

different data sets should be combined and analyzes should be performed on a single data set.

Data Cleaning:

In data analyses, it is not feasible to use raw data directly. Quality problems in the raw data may cause problems in the implementation of the analyses or in obtaining consistent results after the analysis. Some of the reasons leading to quality problems in electricity consumption.

Outlier Data:In its most general definition, it is the values that are far from the general data distribution and are statistically inconsistent with other data.

Noisy Data:It is low-quality data that is not possible to be used with the help of any software or device to make sense of the information it contains

Missing Data:Missing data are empty or meaningless sections in the data set as the result of problems in the phase of measurement, transfer, or storage processe.

Data Reduction:

Datasets may have more features or instances than required. Working with an unnecessarily crowded data set increases the computational effort and time in the analysis.

ANALYSIS THE ENERGY CONSUMPTION DATA:

Analysing how much energy your facility

consumes lets you quantify the energy resources associated

with your service, and identify and correct consumption

inefficiencies.

THE ROLE OF STANDARD FOR ENERGY USING DATA:

These standards specify how energy consumption data from meters, especially smart meters, is to be stored and shared. For example, there are standards for how data should be stored in utility databases and in electricity meters. Other standards are data models that specify how different data elements are interrelated – such as the connection between the meter, the consumer, the transformer, and the distribution circuit.

DATA AND METHOD :

County-level energy consumption data were obtained from the CEC. Electric power and natural gas consumption data were available for two temporal scales, monthly and yearly. Monthly data consisted of energy consumption estimates from 2005 to 2007. Yearly data consisted of residential and nonresidential energy actual consumption from 1990 to 2009.

CREATING VISUALIZATION:

- **Data visualizations should have a clear purpose and**

audience.

• Choose the right type of viz or chart for your data.

• Use text and labels to clarify, not clutter.

• Use color to highlight important information or to differentiate or compare.

• Avoid misleading visualizations.

PROCESS DEPENDENCY:

By this stage, it will come as no surprise that plastics processes are not equally energy-intensive. Therefore each process again requires a specific external benchmarking reference but sufficient data are not available for all production processes. Fortunately the data that are available cover the majority of the materials processed and are relevant to the bulk of the industry machines.

Average machine SEC data across the range of machines surveyed (which includes any machine base load) has been calculated and this is shown on the upper right. These average values are not weighted in any manner for the number of machines surveyed or the production volume. They are a simple average of the SEC values for the variety of machines that were physically surveyed. As with site SEC

values the use of a simple average SEC for any process is not particularly useful due to the effect of production rate on the apparent SEC of the process.

Comparing machine SEC values without taking production rate into account is meaningless.

Average machine SEC for various processes1

Average machine SEC values from machine data and typical process loads from PCL information (see Section 2.4) show some comparison but the machine SEC must be corrected for the production rate to correct for base loads and to allow machine comparisons. Average SEC values have very limited usefulness for assessment or comparison purposes.

Unlabelled Image

Sign in to download full-size image

The machine base load is amortised over higher process loads at higher production rates

As with site performance, increasing machine production rates amortises the machine base load over greater process loads and the apparent SEC of the machine decreases. Despite this, the values reinforce the strong process dependency of energy use in plastics processing, i.e.,

extrusion is significantly less energy-intensive at the machine level than injection moulding.

THE ROLE OF STANDARDS FOR ENERGY USAGE DATA:

These standards specify how energy consumption data from meters, especially smart meters, is to be stored and shared. For example, there are standards for how data should be stored in utility databases and in electricity meters. Other standards are data models that specify how different data elements are interrelated – such as the connection between the meter, the consumer, the transformer, and the distribution circuit. These data models are used in software applications that manipulate or analyze energy usage data, including presentment of data to consumers via online websites.

The most important use of standards is in exchanging data between computer systems operated by different entities.

DESIGN AND OPTIMIZATION METHODOLOGY:

In order to generate the energy consumption data for the designs of conventional and thermally coupled quaternary sequences, a multiobjective genetic algorithm

with constraints handling is used. We decided to employ this method since it has been proved to be a robust tool for the optimization of chemical processes. The genetic algorithm is coupled to the process simulator Aspen Plus; thus the complete rigorous model for distillation columns is used. A main problem with the use of the process simulator is the long time required to evaluate the objectives and constraints functions. Thus, the speed of the strategy is improved through the use of neuronal networks, which are used as surrogate models for the evaluation of the objectives and constraints functions. The code is implemented in Matlab, which is linked to Aspen Plus using ActiveX Technology. For more details the reader is referred to the original contribution (Gutiérrez-Antonio and Briones-Ramírez, 2015). For each system, a tuning process is performed to determine the number of generations and the number of individuals required. For the configurations analysed in this work, 1000 individuals are required per generation. The numbers of generations were 150 and 250 for conventional and thermally coupled distillation sequences, respectively. The main decision variables for the conventional

configurations are the number of stages, the reflux ratio and location of feed stages. In the case of the thermally coupled sequences, the flow rates and locations of interlinking streams are also important design variables. The objective function involves the simultaneous minimization of the number of stages and heat duty for each sequence, while the constraints are the purities and recoveries established for all cases.

CONCLUSIONS:

The primary objective of an energy management system is to collect energy consumption data, process it, and eventually, either present the results to the user to influence behaviors and recommend specific actions, or perform such actions directly in a closed-loop control arrangement. These functions are ensured by specifically designed mathematical methods that – when implemented properly - allow to achieve energy savings.

Ontological engineering does not contribute to these energy savings but it brings several useful benefits to the overall software design process. Firstly, it helps to develop a deeper insight into the modeled domain, intended system

functionality, considered types of data and mechanisms of its exchange between individual components. This is beneficial especially in case of larger development teams. Secondly, it allows this knowledge be applied in a coherent way in the implementation of the information system and make it more modular, flexible, and open. Potential three party providers of new services have not only some data from the database, they have also an explicitly stated semantics about them. Furthermore, such a system should be more adjustable in next installations.

All these features, realized by ontological modeling, make the implementation of services easier and thus help to reduce the time needed for system development. Then, in an open platform like ENERsip, the services are subject of free competition.

DATASET:

In order to accurately evaluate the load demand of the Cellini clinic, the electrical energy consumption data, measured every 15 min, was gathered from the local distribution system operator. The collected data in the year 2012 amounts to 2662.325 MWh or 131.06 kWh/m2.

From the available dataset it is possible to note that the energy consumption in summer is 30% higher than in winter due to air-conditioning. The winter load peaks are between 450 and 480 kW, while the summer peaks are between 600 and 650 kW.

The dataset also contains the ambient temperature that correlates with electrical consumption as presented in Table 9.1 (Matlab Manual, n.d.). In this case, the correlated vectors are the daily electrical energy consumption and the minimum/maximum temperature of the day. At the beginning of this study we considered the humidity, but removed it in the final dataset composition because of its low correlation with the load (see Table 9.1, where the correlation has been evaluated as in the temperature case).

Table 9.1. Correlation values between electrical consumption and minimum and maximum temperature as well as humidity.

Empty Cell Minimum temperature Maximum temperature Minimum humidity Maximum humidity

Correlation year 2012 0.72 0.70 0.10 0.12

Correlation Summer 2012 0.45 0.41 0.03 0.04

Correlation Winter 2012 0.43 0.32 0.02 0.03

The dataset was partitioned into a training set and a testing set. The training set was used to train an ANN and consists of an input vector and a score vector. The testing set was used to assess the strength and effectiveness of the predictive relationship and to calculate the forecasting error in terms of Mean Absolute Percentage Error (MAPE), Daily Peak MAPE, Coefficient of Variance of the Root Mean Squared Error (CVRMSE), Maximum Percentage Error (MPE), and Percentage of Test set with a MAPE under 5% (PE5%). These measures of accuracy are defined as follows:

where N is the number of data points, Yt is the actual load, Ft is the predicted load, and is the mean of the testing set loads.

The Daily Peak MAPE analyzes the results with respect to daily load peaks, while the CVRMSE indicates the uncertainty in the forecast procedure. This last performance index has been exploited by the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) to define a target for general model accuracy. ASHRAE recommends a CVRMSE value under 30% on