

Problem Description

The goal of the problem is to develop a system that will allow the energy provider AENERGY to determine whether a client would have trouble covering rising electricity rates.

Approach

Data science, artificial intelligence, and machine learning have improved the corporate sector in the modern period. The use of conventional programming techniques is not the only way to solve this issue. These days, using machine learning to solve these kinds of issues is popular. Both methods have some advantages and disadvantages. A comparison between conventional programming methods and machine learning techniques is shown in the following table.

Machine Learning	Traditional Programming
A model is built from the data, and that model is the logic	A developer designs logic or algorithms to solve a problem.
Find rules using existing examples.	Rule-based approach
Knowledge Driven	Data Driven
No labelled Data	Labelled Data for Training
Producing same results every time.	Programs evolves to improve results.

In order to forecast a categorical result that indicates whether a consumer has problems paying an increasing energy bill based on historical data, a machine learning approach is more practical in this scenario. This method of machine learning is supervised learning. Additionally, manual rule formulation is required in traditional programming, which is ineffective for handling this kind of issue [4]. Traditional programming requires manual rule formulation, which is inefficient for dealing with this kind of issue [4].

Type Of Predictive Task

Predicting if a customer would have trouble paying their energy bill is a classification-type machine learning issue, and the challenge is supported by a labelled dataset with the intended class. Since this is a supervised learning activity, labelled training examples are needed [1]. The solution to the issue can be classified as either True or False (True when the customer encounters a problem, False when they do not). In order to forecast the targeted class, the manager gives past data on the customer's capacity to make payments or not.

Informative Features

Different direct link features are used in the task to forecast the target class. For instance, the effectiveness of household appliances and the strength of the heating system, as well as the nature of the individual, the make-up of the family, the weather, how much time is spent indoors, and several other aspects. Among these, the following are effective classification predictors:

Family composition: Domestic energy price increases with more the number of family members, electricity usage increases and that will affect overall expenditure of the family.

Weather: Climate is another important feature that affect energy bill increase. During winter, usually the cost is 40% higher than normal time.

Frequency of Usage: It depend on the amount time people spent in house.

Nature of Individual: Some people always don't care about energy usage.

Learning Procedures

The choice of a learning process is influenced by the nature of the problem, the type and quantity of the data, and the number of features and rows in the dataset. According on this data, the following learning method [2] is suggested to predict the categorical result:

Decision Tree

- Support Vector Machine (SVM)

- K-Nearest Neighbour
- Logistic Regression

There are some reasons to choose a particular algorithm for this task.

- The number of data points and features: Methods like SVM is best suitable to predict a model with more features.
- Data Size: Speed of execution is affected by size of data.
- Training and Prediction time: Higher the accuracy, higher the training time
- Type of Data: If the response variable is categorical, Classification is used. If it is continuous regression is used

Performance Evaluation

The effectiveness of a machine learning model on a dataset that it has never seen before is evaluated in machine learning performance evaluation, and matrices play a crucial role in this process. There are several performance evaluations measures available to evaluate the performance of machine learning models for classification and regression. Accuracy, precision, recall, and confusion are used to assess classification-type issues, as well as R Square and Mean Square Error for regression type problems. Matrix evaluation is used to assess classification-related issues [3]. Additionally, visualisation is a useful tool for evaluating how well a certain learning process is working. Deeper understanding of the data is provided through plots and graphs such the pair plot, bar plot, and histogram.

Conclusion

This research discusses approaches and techniques for picking a machine learning algorithm to determine whether a consumer will struggle to pay an increasing energy bill. However, it is difficult to determine which machine learning method is best for this issue without first analysing and assessing how well each algorithm performs on the data.

References

1. <https://www.linkedin.com/learning/machine-learning-with-python-foundations/what-is-supervised-learning?autoplay=true&resume=false&u=51088249>
2. <https://www.kaggle.com/general/253858>
3. <https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234>
4. <https://www.avenga.com/magazine/machine-learning-programming/#:~:text=In%20Traditional%20programming%2C%20we%20write,solve%20the%20problem%20by%20itself.>