Implementation of Possible Machine Learning Models on Predicting Payment Difficulties for Increasing Electricity Costs:
A Proposal for AENERGY

Prepared by: Onyemelonu Nneoma Charity [2207938]

[Word Count: 605]

Overview

With the incessant cost of electricity bills, there are imaginative likely hoods that customers of AENERGY may likely be affected, resulting to an inability to keep up with payments of energy. However, a conclusion on this will be dependent on the customers and some other factors surrounding the history of the customer. In order to keep the Company in the know and curb any future financial loss that may emanate as a result of this inflation, a machine learning model will be built to help predict the percentage number of AENERGY customers this hike in price will affect.

Objective

As slightly stated above, the objective of this proposal is to assist AENERGY in predicting if a Customer is going to encounter difficulties in paying the increasing cost with respect to some features relatable to the Customer, such as; demography, age, family composition, etc.

Data / Informative Features

Having a historical record of some of the information of AENERGY customers, AENERGY will provide numerical data of certain factors or features that may likely influence this prediction. These numerical data could be:

Average Age of Household, Number of Households, Total square footage of home, Number of rooms in the home, Number of appliances, Average monthly electricity bill, Energy consumption history, Monthly income, Percentage of income spent on energy bills, Credit score, Number of years with AENERGY, Number of years at current residence, Access to government assistance, Previous payment history, Number of late payments in the past year, Average Monthly payment made, Homeownership status, Number of complaints to AENERGY, Average Daily energy consumption,

Number of energy-efficient upgrades and Home energy efficiency rating. These data will make up a training dataset (with a Class column) and test datasets (without a Class column) to enable a proper confirmation of the best-suited model.

Type of Task

AENERGY's interest in using machine learning methods to classify customers, as either likely to struggle with paying their energy bills or not, is a supervised binary classification problem [1]. Using all the provided features for each customer, the prediction will be a binary output of either a 1 (True -"likely to struggle") or 0 (False - "not likely to struggle") with paying energy bills.

Methodology

The data as provided by AENERGY will be pre-processed to handle any missing values, and outliers, and then placed within a range to enable the machine learning model understand and predict almost accurately

The Binary Classification models [2]; Decision Tree Classifier, Random Forest Classifier and Logistic Regression will be experimented on by fitting the train data individually and then predicting using the test data.

Evaluation

The performance of the models will be evaluated using metrics such as accuracy, precision, recall and f1-score. A confusion matrix will be printed too, to demonstrate the True Positive, False Positive, True Negative, and False Negative of the model. Once the model has been trained and evaluated on the training data, it will then be tested on the test data to evaluate its generalization performance. The model's performance on this test set will provide a more realistic estimate of how well the model performed.

Conclusion

The proposed solution of using machine learning models to predict customers who may encounter difficulties in paying the increasing electricity cost will enable AENERGY to consider and prepare any pre-caution measures, and as well, do every needful in retaining customers and improving customers' experience The evaluation

of the models will provide insights into the factors that are most important in predicting payment difficulties.

References

1. Saurabh Kumar Srivastava "Machine Learning: A Review on Bi□nary Classification", International Journal of Computer Applications (0975 – 8887) Volume 160 – No 7, February 2017

2. Patil, Tina R., and S. S. Sherekar. ''Performance analysis of Naive Bayes and J48 classification algorithm for data classification.'' International journal of computer science and applications 6, no. 2 256-261. 2013.