

AENERGY

PILOT STUDY PROPOSAL

STUDENT ID: 2202396, NAME: ABDUL



The purpose of this proposal is purely study based for the assignment

It should not be considered for any other aspects

OVERVIEW



This section gives overview on the problem of AENERGY company for predicting if the clients will be able to pay bill or not

With the increase of consumption and the huge prices of bills it has become more and more difficult to exactly if the clients are able to pay for the cost of the bill. I am grateful to represent the ways we can tackle the situation and also, how we will be most efficiently and accurately predicting if the customers will be able to pay the electric bill or not.

The Objective



The objective of this proposal is to give an overview on what steps can be taken to overcome the problem at hand.

- **Need 1: Method Selection**
- **Need 2: Attribute Selection**
- **Need 3: Model Selection**
- **Need 4: Evaluation**

The Solution



These are the steps that can be taken in order to solve the problem at hand

- **Recommendation 1: Use Classification Methods**
- **Recommendation 2: If Applicable get information on customers income and other monthly expenses**
 - **Monthly Income**
 - **Monthly Total Expense**
 - **Number of Appliances**
 - **Monthly Power Consumption**
 - **Number of Rooms**
- **Recommendation 3: This problem can be solved with two preferable models**
 - **Decision Tree Classifier**
 - **Random Forest Classifier**
- **Recommendation 4: There are different methods to Evaluate that can be implemented**
 - **Confusion Matrix**
 - **Confusion Report**
 - **ROC & AUC**

OUR PROPOSAL



This section contains the reasoning behind selecting the solutions.

The solution required is to predict if the customer can pay for the bill or not and this is generally categorized a classification problem, hence we need to apply one of the methods. In this case many algorithms can be applied which can give satisfactory results but all depends on the type of the data. First thing I presume that will be required is data Preprocessing as there might be some missing data which will be needed to be removed by looking at the dataset itself and figuring out the pattern. This will also depend on the amount of data as we need to make sure if we have limited amount of data, we will have to fill in the columns which can also become a but prone to biasness. On the other hand if we have large amounts of data we can either discard the entries or just fill in by data science techniques.

The attributes that would be best in order to get accuracy would be related to the finance of the customer. For example, if we can have a general idea of Monthly income and other expenditures of a particular and along with the total approximate consumption they have, we will have better idea. Most importantly, we need to keep in mind while calculating the information, we consider the consumption can vary with the season or weather. As instance if it's really cold weather outside. I'll have to increase the heating in the house or if it's really warm outside, I'll have

Word Count: 754

to increase the cooling. This also puts an importance on keeping number of rooms in mind because it will increase the cost drastically.

The model selection is the most crucial part as different models can give different results generally depending on the type of dataset for example, SVM (Support Vector Machine) gives us really good accuracy when it comes to small datasets, but it has worst accuracy on larger datasets. Again, if the dataset is less than let's say 8000 entries e.g. We can apply SVM but if that's not the case, then I believe Decision Tree and Random Forest are the best bet on the matter as both are very good at handling large amounts of data.

For the Evaluation the confusion matrix and ROC and AUC work just perfectly fine. One gives information on right and wrong choices made, the other one gives a line or chart to indicate performance of the model. If you want some extra information, we can also use confusion report which gives us detailed information on efficiency of the model

CONCLUSION



Final Summary of what's defined above

We can utilize classification models to solve this problem mainly, Decision Tree and Random Forest. The other attributes like monthly income, expenditure, consumption, and number of rooms will be really good additional predictors. And we can use Confusion Matrix and ROC & AUC curve to evaluate our model