challenge within your industry that could be addressed through machine learning.

To complete this activity, first briefly summarise a unique challenge in your industry (for example, ‘calculating risk for new insurance policies’ or ‘optimising yields for agricultural output’). Then answer the following questions:

1. What is the problem you are trying to solve?
2. Why do you think machine learning is a good fit for addressing this problem?

Finally, list any questions you have about your chosen application for machine learning and its viability. When reading other participants’ posts, provide any advice or suggestions you have to their questions.

Clearly and succinctly summarises a unique industry problem and suggests ways it could potentially be addressed through machine learning. Provides some questions on the potential application of machine learning that other participants can respond to.

* Problem

Energy industry is currently undre significant pressure due to current economic context. This is reflected worldwide, across all industries and economies. One main area of focus is residant5ial properties - Drivong the price increases and pressure especially on the households fromhouseld wiyth financial burden.

One simple, yet effective machine learning model is detecting the vulnerable customers. A vulnerable customer would be households not able to afford electricity/gas due to financial constraints.

Usability:

Having such information can help companies tailor marketing approach, or email campaigning around specific products that help the households, wile keeping them informed (e.g. introducoing smart meters, energy-saving tips such as discounted rates).

* Increase certain emergency electricity usage, knowing the person in under financial burden
* Monitor any changes in behaviour that can be a good indicator of ‘afordability’
* Need for ML

Machine learning is a good method of solving this problem as most people might not report they might have financial pressure that impacts their ability to pay. Hovever, given the consumption and payment information can be modelled to create clusters and

* Questions on the application?
* What other methods can be potentially used?
* If unsupervised methods are used, what features do you believe could help in addressing this problem?

Now that you have identified a potential machine learning application for your industry and received feedback from your peers, you’ll now define more of the details of your specific problem. To complete this activity, provide a brief overview of the problem. Then, answer the following questions:

1. What data would you use?

Primarly, I’d use customer data : demographic and transactional. Depending on what is available features such as: payment, usage, billing/debt information, tenure

1. What are your key input and output variables? (Review section 2.4.)

Input:

Output: Since I see this as an unsupervised ML problem, the output variable would be the cluster number. Output is a categorical data type, but the problem is not a classification problem as the classes are not pre-defined.

1. What type of machine learning problem is this? (Review section 2.5.)

As stated above, although this is very close to classification, when it comes to output, this is a clustering method (unsupervised).

1. What steps would you take to solve this problem through machine learning? (Review section 2.6.)
2. I’s follow the 10 steps outlined in the section 2.6, although the implementation wouldn’t be so linear.
3. **Define the purpose of the ML project**

**This would start with the need for identifying the vulnerable customers, and use this information foir decision making**

Take decision with a client or a sponsor within the company whether a project method is a one-off effort or an ongoing procedure. This will have implications for users that apply a method or for those who are going to interpret the result of the method.

**2. Obtain the data set for the analysis**

This is most likely the internal customer database, ghaving both transactional and demoghaphic information to a certain degree. If there are other external data sources, I’d investigate this.

*Data can be sourced internally, such as customer database and purchase databases, and externally, such as credit rating databases.*

***3. Explore, clean and pre process the data***

I’d perform EDA and inform business on the main fainding. After, I’d decide to deal with missing dataor any outliers. Following this, depending on the data type, I’d normalise/ standardise the dataset.

***4. Dimension reduction and feature engineering***

Aftert performing EDA, rinning correlation or feature imjportance it is important to give the important variable, as well as dimensionality redu=ction to get an idea of what features might not be needed.

*Remove input variables that will not be available while employing the ML technique to forecast or infer data. Also, remove input variables that are irrelevant for the analysis and not correlated to the output variable. Few of the input variables can also be transformed to numeric values or vice versa. Page* ***5*** *of* ***7***

***5. Determine the ML task at hand***

Because there are no=pre defined classes and the output variable highlights a categorical cluster, the problem is an unsupervised machine learning problem.

1. ***Partition the data (if supervised ML)***

*The clusters are formed to optimize an objective partitioning*

*criterion*

*¤ Objects within a cluster are similar*

*¤ Objects of different clusters are dissimilar*

CLARA (Clustering Large Applications) uses a samplingbased method to deal with large data sets

*For a supervised ML problem, partition the data into training data set, validation data set and test data set.*

***7. Choose the ML technique(s)***

*Select an appropriate ML technique for the task under consideration – either a single technique or multiple techniques. A few examples of the techniques used are regression and classification techniques, clustering techniques, k – nearest neighbour technique etc.*

***8. Use the ML technique(s)***

*Apply the technique(s) to the task in hand and wait for results.*

***9. Interpret the results***

*Compare the algorithms against each other and against simple benchmarking strategies.*

***10. Deploy the ML technique (optional)***

*Once the results are analysed, use the chosen algorithm, and deploy it in the task.*

1. What might cause missing data in your data set? Which approach outlined in the lecture materials do you think would be most suitable for dealing with missing data, and why?

For the initial model iteration, I’d choose to inspect missing data along with the subject matter experts as of why data might be missing. Checking for any patterns/distributions of the observations where missing data occurs. Is it a system/data quality issue or there are other reasons? For this reason, I do not see imputation a good method as it can introduce bias. Simply eliminating the recods can provide robust enough information.