Project title:

ELEKTRILEVI-2: Cables, transformers, poles

Project slide page:

https://goo.gl/1oPwF5

Project Members

- Muhammed Bial Shahid (bilal.shahid@ut.ee)
- Bilawal Hussain (bilawal.hussain@ut.ee)
- Hippolyto Fayol (hippolyto.fayol@ut.ee)

Background

Our client, an electric supply companay called ELEKTRILEVI, supplies electricity to almost every home and business in Estonia. Also they maintain and repair all the electric poles, transformers and electric lines by themselves. They collect data about poles, transformers and electric lines regularly and store them for making future decisions. We will be using that data to help the company in making decision on various feilds for example when to change the pole, get the transformers with best performance and reduce electricity outage that is caused because of electric line failuers.

Business goals

The business goals for this project are as follow

- Identify cable reliablity i-e when to install new cable, put joints on it etc
- Identify types of defects that are likely be seen in different kind of transformers as their age increases so that the risk of outage can be reduced
- Identify durability of poles of each type and when to replace them with new poles

Business success criteria

The system should predict the reliablity of the electric line (cable), types of defects in transformers and what are the main causes of them and durability of the poles and what type of pole to be used whenever the need arrives to make new poles. The system will be considered successfull if the system can give clear cut answers to the above suggested predictions. For example the cable that is going to be installed has 'x' reliability, can handle 'y' amounts of joints and minimum age of the cable before first joint can be installed.

Assessing your situation

Inventory of resources

- **People**
  - Data Minners:
    - Muhammed Bial Shahid (bilal.shahid@ut.ee)
    - Bilawal Hussain (bilawal.hussain@ut.ee)

    * Hippolyto Fayol (hippolyto.fayol@ut.ee)
  - Contact Person: Shahin Abbaszadeh (shahin.abbaszadeh@elektrilevi.ee)
  - Supervisor: Meelis Kull (meelis.kull@ut.ee)

- **Data**
  - poleLifeSpan.csv
  - cableReliability.csv
  - defectsTrafo.csv

- **Hardware**
  - Hp Elite books (given by university of Tartu)

- **Software**
  - R studio

Requirements, assumptions, and constraints

The project should be finished before the poster session held at 8 Jan, 2017. Any data that is given to us by the companay or will be provided to us by the companay if asked should remain confidential and will not be shared in public. And we assume that the project will be considered successfull if it meets the Business success criteria.

Risks and contingencies

If assistance is required from the contact person of the given company and is delayed then it might throw off the completion date of the project. The contingency plan for this will be to get assistance from the supervisor and follow his instruction so the project can be completed in timely manner.

Terminology

The glossary of the terminology will be provided soon.

Costs and benefits

As this is a academic project so it costs nothing. No need to buy any hardware or software and all the participants working on this project are students so there is no cost for their compensation. The benifit of this project is that it may help the company to make decisions like what type of transformers they should buy for future, what kind of poles are more cost effective in term of durability.

Defining your data-mining goals

Data-mining goals

- Determine which type of transformer can last longer than others.
- Report which cable line can survive with maximum number of joints.
- Find out which type of pole need minimum observation.

Data-mining success criteria

- Success will be conformed, when the system can find out all Data mining goals correctly.

Gathering data

We have access to the desired data and that data is compatible with our data mining tool that will be used for this project. The data is in csv format and have 3 csv files named poleLifeSpan.csv, cableReliability.csv and defectsTrafo.csv. These files contains the data that will be used in this project

Outline data requirements

The data that is required for this project is available in the following files

- poleLifeSpan.csv
- cableReliability.csv
- defectsTrafo.csv

Verify data availability

Yes the data that is required for this project exists and available for us to use.

Define selection criteria

For this project we were given very relevant data. We have all the data that is needed for this project and all these three files will be used in this project. All the feilds from those csv files will also be used in the project.

Describing data

The source of the given data is ELEKTRILEVI company and the format of the given data is in csv format. The fields are as follow:

- poleLifeSpan.csv
  - **ID**: ID of poles, which is unique for each pole
  - **OBJECTID**: Object ID of the pole
  - **CLASSID**: Class ID of the pole
  - **AGE**: age of the pole, how long before that pole was installed.
  - **OBSYEAR**: when the last time pole was observed
- cableReliability.csv
  - **JID**: Joint ID, which is unique for each pole
  - **CID**: Cable ID
  - **CLASSID**: calss id of the pole
  - **CABLENO**: The number of the cable
  - **CLASSNAME**: The name of the cable class
  - **FEEDERID**: From which feeder the cable is coming
  - **AREA**: Area in which the cable is. The possible values for area are 1-5 (area codes)
  - **JYEAR**: Joint year, When the joint was installed in the cabel
  - **CYEAR**: cabel year, When the cable was installed
- defectsTrafo.csv
  - **ID**: ID of transformer, which is unique for each transformer
  - **JAOTUSALAJAAM**: Name of the distribution area(?)
  - **REPLDATE**: when the transformer where replaced
  - **CONYEAR**: year when the connection was established
  - **CONNECTIONDATE**: date when the connection was established
  - **REGION**: area where transformer was placed
  - **OWNER**: who owned the transformer

- **MANUFTYPE**: They manufucture type of the transformer
- **OBSID**: observation ID, when the pole was observed and saved in the database.
- **OBSERVATIONDATE**: when the transformer was observed
- **NAME**: transformer name
- **DESCR**: Description of the condition of the pole (example dirty, ok, dusty etc)
- **URGENCY**: How crucial the transformer is. For example how much impact it will have if break down

Exploring data

By exploring data more deeply, we have concluded some initial findings

- By given information in JID, JYEAR, CYEAR, CID and AREA fields, we can find out that which cable most likely need to be repair and in which area.
- By given information in JAOTUSALAJAAM, REPLDATE, MANUFTYPE and OBSERVATIONDATE fields, we can find out that which type of transformer will need a repair earlier than the other type of transformer.
- By given information in ID, CLASSID, AGE and OBSYEAR fields, we can find out that which type of pole can last longer than the other type of poles.

Verifying data quality

We have our data and it's good enough to support our goals. There are no such quality issues in our data set. Just in some fields there are "NA" values given which can be replaced with desired values to achieve the goal. Otherwise, data is perfect and ready to use.

Project Plan

- Clean the data, i-e replace the NA, delete/remove some values, Change Column names to easy to understanable format(from Estonian to English)
- Study the data to find out patterns in it. For example when first joint was added in a certian cable, when the cable was replaced etc. Basically fully understand the data and try to find out interesting facts that can help in solving the project.
- For Transformers, First find out the minimum and maximun age. Check if those values are frequent or not (to check if there are no vague max and min values).
- Then find out what kind of defects transformers can have. Then classify them with the manufucture type and see which type is most likely to get defected
- Then find out what is the relation between defects and age of transformer.
- Then predict which transformer is most likely to break down
- For poles, classify them according to their pole type and then find out the minimum and maximum age of each pole type.
- Then predict which pole is most likely to be replaced in near future. (form all four types of poles)
- For Cables, Find out the minimum age required to add a Joint in it. Also classify cables according to Area so we can predict which area will be needing a patch soon
- Find out what type poles are best in term so cable durability(maximum age to add a Joint it)
- Predict when to add a Joint in certian cable and in whicg area.

- After finishing the core functionality of project, start working on Poster
- Present all the finding in visual format so they are easy to understand