Study Committee D2

Information Systems and Telecommunication

Paper 10616_2022

IMPLEMENTATION OF A DECISION SUPPORT SYSTEM FOR UNACCOUNTED ELECTRICITY CONSUMPTION DETECTION USING MACHINE LEARNING METHODS

Voltov Ilya

R&D Centers at Rosseti PJSC

Motivation

- Unaccounted electricity consumption is a significant worldwide problem.
  Total global electricity losses due to electricity theft are about $96 billion per year.
- The main aims:
  - identifying points of unaccounted electricity consumption
  - reducing the percentage of non-technical losses
  - reducing the decision-making time
  - creating a data warehouse
- There is no similar decision in Russia.

Experimental setup & test results

- Analysis of foreign experience (including doctoral and grant programs).
- The data used for training, validating and testing the ML models was provided by PJSC «Rosseti Lenenergo».
- Size of dataset:
  - 5 years of hourly measurements
  - unique customers - 293 837
- Labeling of the dataset - about 1500 cases of theft recorded earlier (for the classification task).
- Choosing a ML model for each type of task (classification, regression, clustering).
- Personal inspection of addresses defined by ML algorithms.

Method/Approach

- Machine learning ensemble:

  data clustering
  - C-means

  classification task
  - CatBoost

  time series forecast
  - neural network – LSTM
- Classic approach:
  - statistical analysis of consumer data
  - calculation of the unbalance in the power grid

Discussion

- Results:
  - clustering of objects (legal entities and individuals) - grouping objects into identical groups, for example, by type of activity and area of the object
  - identification of the consumers who steal electricity
  - forecasting the breakdown of the metering device
- Benefits:
  - detection of several large legal entities (unaccounted consumption) - large monetary fine
  - detection of many medium- and small-sized legal entities (unaccounted consumption)

Objects of investigation

- The area of activity of one of the largest electric power companies in Russia - PJSC «Rosseti Lenenergo»:
  - 1432 - transformer substations
  - 264 381 – unique individuals
  - 29 456 – unique legal entities

Conclusion

- Built first of its kind the decision support system.
- ROI – 1 year (for the chosen area).
- Identified the best algorithms for detection of unaccounted consumption.

Study Committee D2
Information Systems and Telecommunication
Paper 10616_2022

IMPLEMENTATION OF A DECISION SUPPORT SYSTEM FOR UNACCOUNTED ELECTRICITY CONSUMPTION DETECTION USING MACHINE LEARNING METHODS continued

Creating a data warehouse

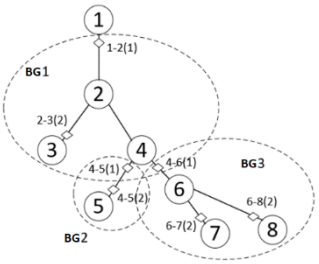
- Two DBMS are used:
  - PostgreSQL (for descriptive data)
  - Click House (for time series)
- Special platform for event streaming:
  - Apache Kafka
- All the tools used are open source.

Data sources (integrations)

- There are two data sources:
  - external
  - internal (customer)
- The external data (only open sources):
  - weather data
  - descriptive data about consumer objects
  - descriptive data on the activities of legal entities
- The internal data (electric power company data):
  - topology of the power grid
  - equipment characteristics
  - cases of theft recorded earlier
  - consumption volumes
  - readings of smart metering devices

How we built the topology

- Topology is the configuration of a graph.
- Depth First Search - algorithm for traversing or searching tree or graph data structures.
- Depth First Search builds balance groups.
- A balance group is a section of the electrical grid between two (or more) metering devices.



Why CatBoost, not other algorithms

- The quality is the same, the speed is greater:
  - benchmark – quality

CatBoost	LightGBM	XGBoost
0.39112	0.39749	0.39764

  - benchmark – learning speed

	CatBoost	LightGBM	XGBoost
CPU (Xeon E5-2660v4)	527 sec	4339 sec	1146 sec
GTX 1080Ti (11GB)	18 sec	890 sec	110 sec

Study Committee D2

Information Systems and Telecommunication

Paper 10616_2022

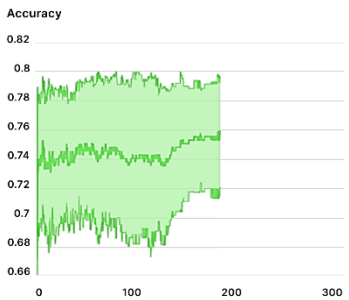
IMPLEMENTATION OF A DECISION SUPPORT SYSTEM FOR UNACCOUNTED ELECTRICITY CONSUMPTION DETECTION USING MACHINE LEARNING METHODS continued

CatBoost

- One of the original gradient boosting algorithm.
- The issue is solved:
  - classification of the consumer who steals electricity

CatBoost Accuracy

- Accuracy – 0.75

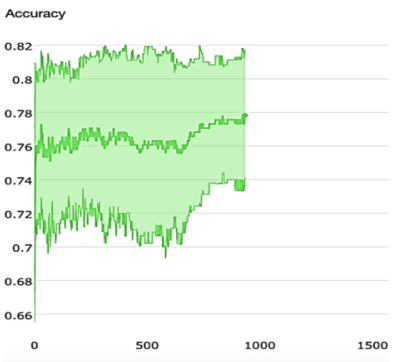


LSTM – long short-term memory

- An artificial recurrent neural network.
- The issue is solved:
  - forecasting the breakdown of the metering device

LSTM Accuracy

- Accuracy - 0.77



- The performance metric for unbalanced datasets is ROC - AUC.
- LSTM ROC – AUC – 0.72.

C-means

- A form of clustering in which each data point can belong to more than one cluster.
- The issue is solved:
  - grouping objects into identical groups

Conclusion

- Built first of its kind the decision support system.
- ROI – 1 year (for the chosen area).
- Identified the best algorithms for detection of unaccounted consumption.