The Twelfth International Conference on
Advanced Geographic Information Systems,
Applications, and Services

A Tool for Spatially Based Prediction of Consumer Lawsuits against Electric Power Companies

Domingos Junior, Johnatan Souza, João Diniz, Geraldo Junior, João Almeida and **Anselmo Paiva**
E-mail: paiva@nca.ufma.br

Context

The main purpose of an energy company is to provide electricity services to the final consumer

- Equatorial Energy Group is responsible for electricity distribution in four states in Brazil, with over 5 million customers

Customer dissatisfaction and lawsuits

- Millions of dollars are spent on lawsuits yearly

Main causes

- Poor quality of service
- Power interruption
- Equipment maintenance

Most frequent cause of lawsuits

Unregistered Power Consumption (UPC)

- Caused when energy theft is detected at the customer-connected facility
- Common in Brazil

Purpose

Construct an intelligent method to identify clients that may be related to unregistered power consumption using customers history

- Extreme Gradient Boosting and Balanced Bagging Classifier

Provide spatial visualization and analysis of lawsuit predictions

Proposed Method

The method consists of five steps:

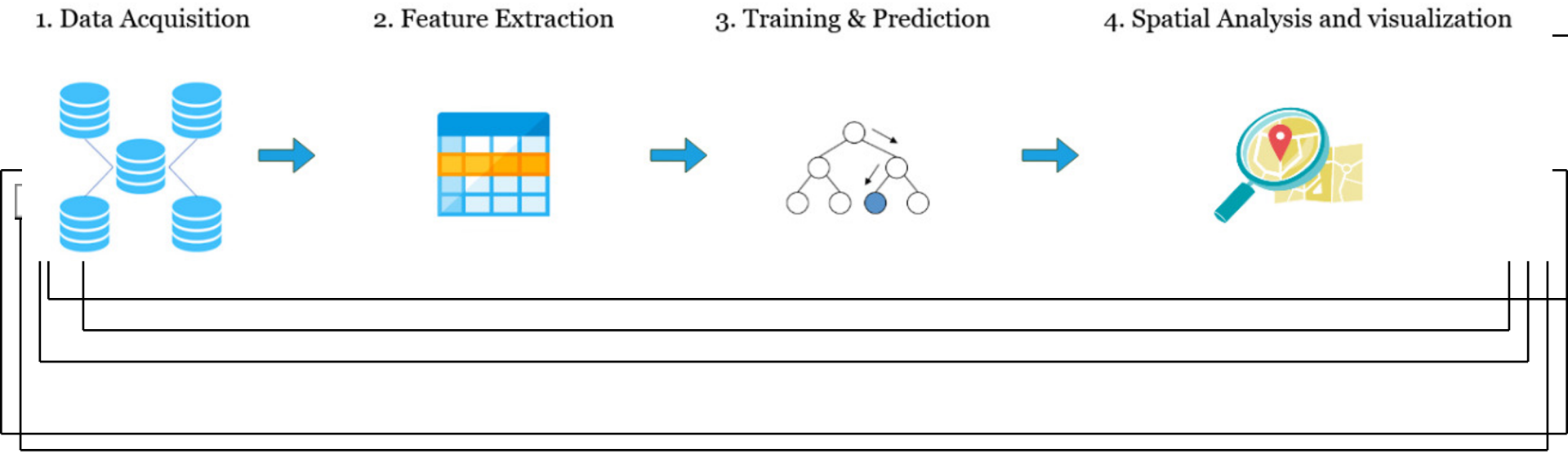

Figure 1. The proposed method

Data Acquisition

Private dataset from Equatorial Energy Group, Maranhao

- **Historical data**, acquired by several sectors of the company
  - consumption history
  - power supply discontinuity information
  - financial information
  - customer complaints
  - equipment information

Feature Extraction

Features are extracted from lawsuits and customer history

- General Information (spatial location, neighborhood, type of client)
- Power consumption
- Power Loss
- Invoice information
- Financial history
- Lawsuit history

Feature Extraction

Transformation and Encoding step

Pre-process:

1. **Normalization**
2. Encoding Categorical Variables using **one-hot encoding**

Feature generation

1. **Encoding Time:** last 3 months, last 18 months, all time
2. Generate **statistics** over these time intervals:
    - Counter, mean, standard deviation, kurtosis and skew

Training

Extreme Gradient Boosting

- Tree-based machine learning algorithm that implements an optimization of the Gradient Boosting method, which is a technique that produces a predictive model from the union of less robust classifiers, showing better performance than if they were used in isolation.

Balanced Bagging Classifier

- It consists in an unbalanced data handling technique, which selects by default 10 subsets of the initial data and performs random sampling.

Results

Experiment

- Unbalanced dataset
- Balanced Bagging

TABLE I. THE PROPORTION OF CLIENTS WITH UNREGISTERED CONSUMPTION SUBJECT IN TRAINING AND TEST DATASET.

Dataset	Consumer with UPC	Consumer without UPC	Total
Train	8.560	1.476.042	1.484.602
Test	5.714	998.442	1.004.156
Total	14.274	2.474.484	2.488.758

Results

TABLE II. RESULT OF PREDICTION STEP USING BALANCED BAGGING.

Subject	Acc (%)	Sen (%)	Spec (%)
UPC	91.86	96.52	91.83

Spatial Analysis and visualization

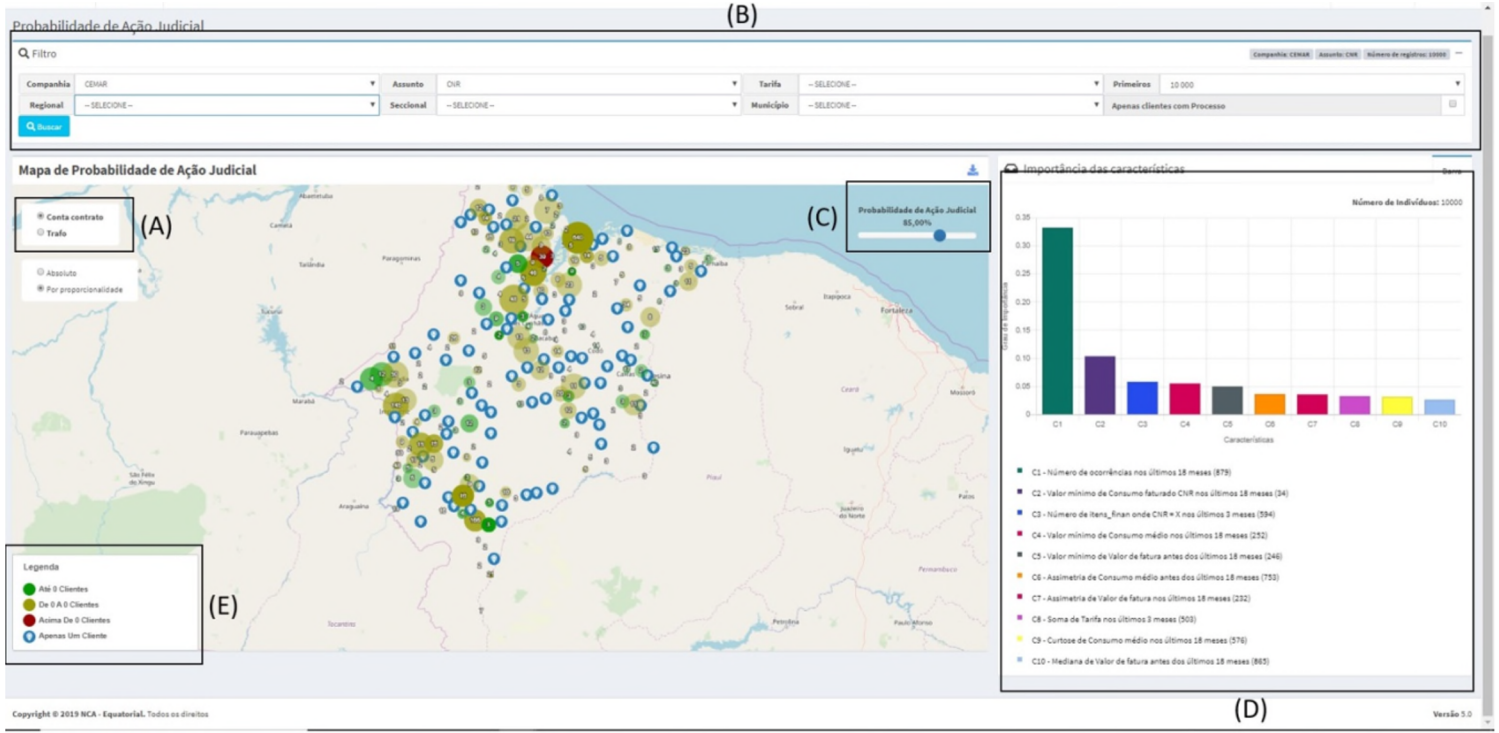

Figure 2. Spatial analysis and visualization tool for customer lawsuits prediction.

Spatial Analysis and visualization

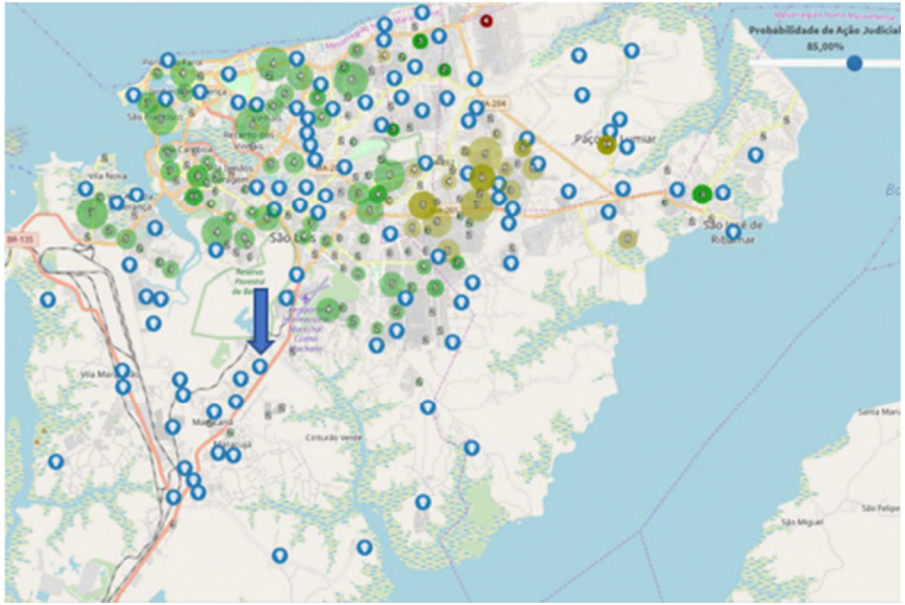

Figure 3. Selecting an individual consumer, blue marker

Spatial Analysis and visualization

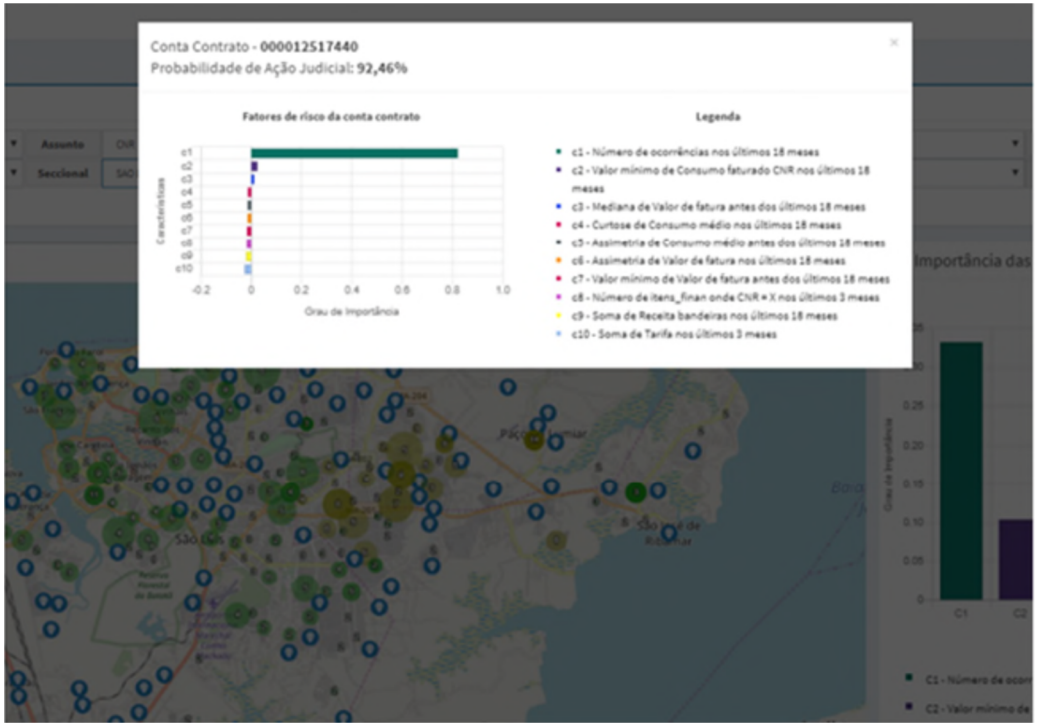

Figure 4. Display of variables that influence that consumer to file a lawsuit

Conclusion

Promising results

- Divided randomly in train/test dataset with UPC

Factors:

- Client geographic location is the most dominant feature regarding the issue of new lawsuits
- Other: consumption history, type of service , invoices issued to the consumer, number of previously issued UPC invoices

Future Works:

- new lawsuit subjects (not only in UPC)
- more classifiers in the training and prediction step
- generating other forms of data visualization such as heatmaps

Questions?

Anselmo Paiva
paiva@nca.ufma.br