Artificial Intelligence and Social Computing, Vol. 163, 2025, 53–58
https://doi.org/10.54941/ahfe1006035

AI Optimization of Resolution Strategy in Utility Billing and Revenue Assurance

Faraz E. Ahangar, Lauren R. Huang, and Zining Yang

Southern California Edison, 2244 Walnut Grove Avenue, Rosemead, CA 91770, USA

ABSTRACT

Sustainable profitability for utility companies hinges on the reliability of their billing and revenue collection processes. While the majority of billing operations are efficiently managed through Robotic Process Automation (RPA), there remains a segment that eludes automation and will be delayed. This portion of the billing requires manual intervention to complete the billing process. The timely resolution of these bills is especially important for Southern California Edison (SCE) since they might be subject to Tariff Rule 17 and result in permanent lost revenue. Unresolved bills also affect customer satisfaction adversely. Ensuring that these manual processes are handled promptly and accurately is crucial in maintaining the financial health of the company and fostering customer trust. Efficiently addressing these challenges can enhance operational efficiency and support the long-term growth of utility companies as well as excellence and continuous improvement. In this study, we explored the delayed bills accounts to identify patterns and trends. We combined our findings with machine learning models, such as Density-Based Spatial Clustering of Applications with Noise (DBSCAN) model, to enhance the process of addressing these delayed bills. This method selectively targeted accounts for a more efficient resolution, reduced lost revenue and brought in greater profitability. Moreover, we expanded this analysis by utilizing predictive models to detect future accounts that are likely to encounter repeated issues. This proactive approach contrasts with the current reactive measures, providing opportunities for improving the efficiency and effectiveness of bill resolution.

Keywords: Billing, Delayed billing, Revenue assurance, Operation optimization, Clustering models, Predictive models

INTRODUCTION

In recent years, the landscape of power systems has evolved significantly. The rise of distributed renewable generation, the shift to electric vehicles, and the new risks from climate change are just a few examples of new challenges utility companies are facing. The utility companies' operation relies on effective billing management. The conventional method for recovering the costs associated with providing electricity involves billing customers monthly for their previous consumption. If the bill is not collected for any reason (e.g. the inability of the customer to pay, grid issues, and processing errors) the companies face not only loss of revenue but also disruptions in the overall service model, potentially compromising the reliability of the electricity supply system. The implementation of Robotic Process Automation (RPA) through these companies has been helping to increase efficiency in various processes at utility companies including billing recovery (Lacity and Willcocks, 2016). While most of the billing can be managed through RPA systems, there remains a portion of it that cannot be collected on due to potential errors in the billing process and needs human intervention. Reducing these errors or addressing them in a more efficient way can result in increased efficiency for companies (Ogura, 2013).

If the errors in the billing are not addressed in time, the bill cannot be processed, and the customer does not receive a bill from the utility company. These are called delayed bills and can cause disruption in the company revenue. Poor billing quality can further cause customer dissatisfaction and breed nonpayers which causes additional damages for the company (Mugabi et al., 2007). These delayed bills also require more resources to be resolved since they need to be addressed manually. Delayed billing issues become more important for the SCE utility company as for most residential and small business customers, the customer will only be responsible for paying the most recent three cycles portion of delayed bills. This causes a loss of revenue for any older bills (https://docs.cpuc.ca.gov/published/FINAL_DECISION/80244-02.htm, accessed in February 2024).

In this work, we analysed the trends in delayed billing across the SCE territory to find the patterns that can help enhance the resolution procedure. Currently, these bills are addressed and prioritized based on value and number of delayed days. Based on our findings, we introduced additional parameters for prioritization and potential proactive solutions.

PROBLEM STATEMENT

The delayed bills are calculated by monitoring the last readings from customer's meters. If a customer account has not been billed by the utility company for longer than a certain time, it will be categorized as a delayed account. Our work aims to understand these delayed accounts better and enhance the resolution procedure.

Regression and clustering analysis were done for these delayed accounts. We applied temporal and spatial analysis for various features in the data. We also looked at other environmental features like seasonality, demographic, geographical, etc.

We did not observe any meaningful relationship between demographic, climate, or seasonal features and the delayed accounts. However, we noticed the portion of delayed accounts increases and decreases at different zip code levels for different time periods. These results showed that some delayed accounts tend to form clusters at different time periods and motivated us to apply a clustering model over the delayed accounts which is discussed in the next section.

We also noticed that about 15% of the delayed accounts tend to become delayed again within the same year. These accounts generally tended to reappear within the next two billing cycles, usually with similar issue. This showed that the issues for this account were persistent and needed more attention. We developed a predictive model for these repeated accounts to provide a proactive view for the agent resolution.

MODELING FRAMEWORK

Clustering Model

Clustering model was applied through a Density-based spatial clustering of applications with noise (DBSCAN) (Ester et al., 1996) algorithm to all the delayed account locations in the SCE territory.

Predictive Model

The predictive model aims to find the delayed accounts that get repeated in the future after resolution. As mentioned, an important attribute of this data is the imbalanced classes. Approximately 15% of bills are repeatedly delayed (i.e. repeated delayed or persistent accounts), and the other majority are not delayed after resolution (uniquely delayed). Imbalanced data can result in inaccurate predictions in favor of the majority class. One of the most common solutions is to adjust the class imbalance ratio through resampling methods (Ali et al., 2013). We explored various resampling methods of classes; Specifically, different ratios of down sampling. Eventually we decided on the 1:1 ratio, as it produced the most consistent prediction results.

To choose an appropriate model, we focused on exploring various classification models, including Logistic Regression, tree-based, and Support Vector Machine models (Cortes, 1995; Menard, 2001). We chose the Random Forest model (Breiman, 200) because it had the best performance out of all the models tested. Because the delayed billing data is chronological, a random train-test split of the data could not be implemented, as it would cause data leaking. Instead, 3 months of data are used to train, and the model makes predictions on the delayed bill status during the upcoming month. Random search cross-validation (RandomizedSearchCV) technique was applied for optimizing the hyperparameters in the model.

In addition to generating predictions on delayed bills, a key goal is to track the accuracy of the model and understand the underlying contributing factors of a delayed bill. The vast number of predictors in the data makes it extremely challenging to distinguish the features of greater importance from features of less importance without an analysis procedure. As such, imposing feature selection and parameter tuning on the model is a crucial step.

Feature selection consists of two stages, manual selection, and algorithm-based selection. Before putting the data through the model, we exclude features such as labels, dates, and redundant variables, anything that should not be considered when making predictions.

Once the model outputs predictions, we transform the results into a useful format. The model calculates the probability that a bill will be delayed and uses a threshold to assign a binary prediction. This allows for prioritization of the effort given to resolve bills that are at a higher risk of being delayed again.

RESULTS AND DISCUSSION

Clustering Analysis

The clustering model was applied to all the delayed accounts on the SCE territory. We noticed that about 20% of the delayed accounts fall within a cluster. We also distinguished among the accounts with the exact same location as "co-located" and the rest of the clusters since the resolution process can differ as the field agent needs to only visit one location if needed.

Random Forest Model

Random Forest impurity-based feature importance analysis showed that the most important predictors in the model were location data Figure 1. This shows the clustering of delayed accounts. Practically, delayed accounts tend to get repeated in locations where other accounts persist in becoming delayed. Other important factors include rate, usage, number of days the account is delayed, delay category, invoice type, and meter type.

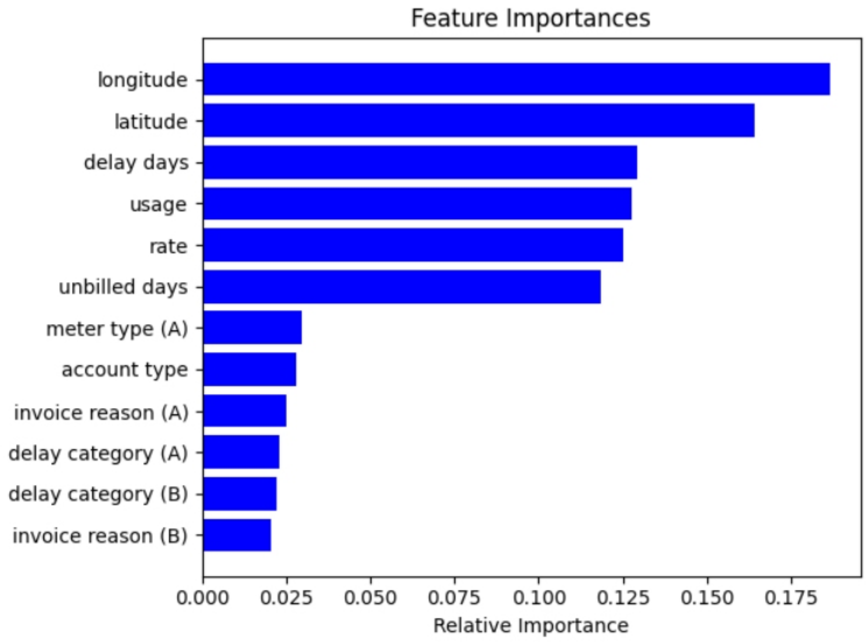

Figure 1. Top 12 features in the predictive delayed billing model.

The performance of the model was evaluated using Receiver Operating Characteristic Area Under the Curve (ROC AUC) along with the recall and precision scores (Provost et al., 1998). ROC AUC was 0.7 with recall of 0.6 and precision of 0.5. We noticed that the performance of the model deteriorates with time as it depends on the location of the previous persistent delayed bills. This means that the model needs to be retrained when the performance falls below a threshold.

DEVELOPING THE HUMAN SYSTEM INTEGRATION

We implemented our models and findings as an interactive live dashboard to help billing specialist agents operate more efficiently. The framework for implementing the dashboard is shown in Figure 2.

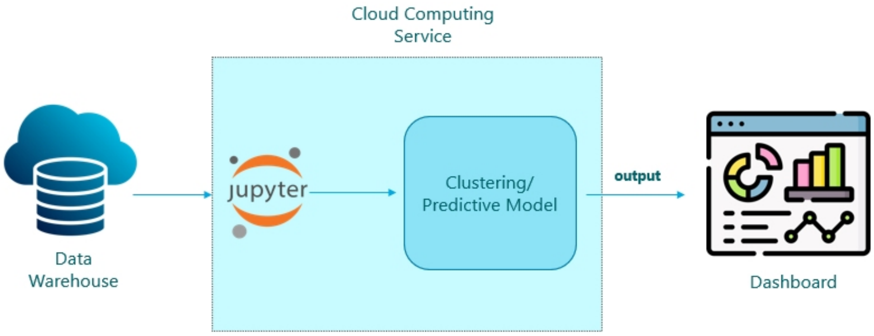

Figure 2. Model and dashboard implementation framework.

The dashboard provides an overview of the clustered accounts across SCE territory as shown in Figure 3. It also flags the accounts which have been delayed multiple times or might be delayed again in the future so the agents can proactively focus on resolving potential future delayed account. This dashboard has resulted in a change in the process for resolving the delayed accounts. Instead of conventionally relying on billing amount and number of delayed days, additional priority is given to the account within a cluster and same agents are assigned to resolve the whole cluster of accounts together and more efficiently. This solution went even beyond its original scope as some persisting delayed accounts were found which could not be resolved without the clustering view.

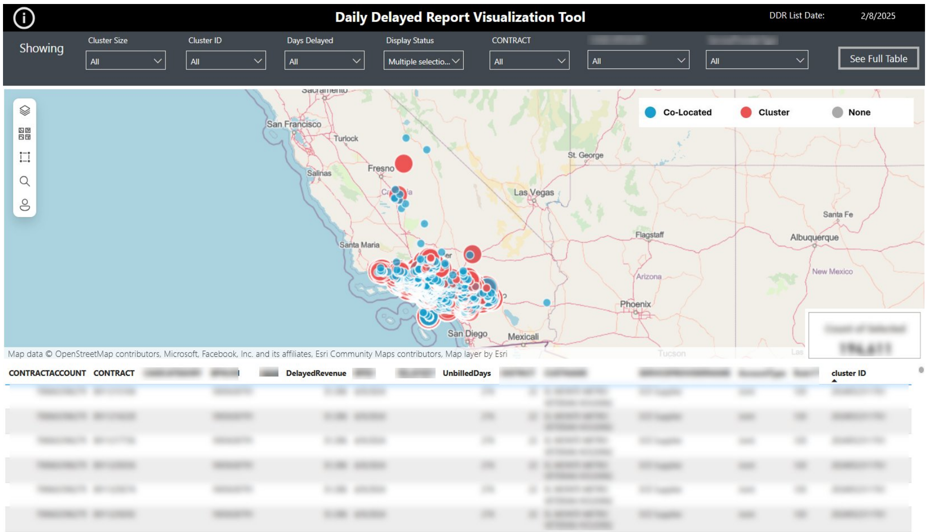

Figure 3. A view of the visualization tool for the agents.

CONCLUSION

In this work, we addressed some of the issues in billing collection process in a utility company, Southern California Edison, by applying machine learning models. We utilized clustering and predictive models to provide additional information about the nature of the accounts with unprocessed bills, referred to as delayed bills. We noticed that about 20% of the delayed bills tend to be in clusters and about 15% of them are persistent, becoming delayed again after being resolved. We developed a model to predict if the accounts with delayed bills get repeated or not. These findings were implemented through an interactive dashboard, helping agents resolve delayed billing issues more efficiently, save resources, and avoid loss of revenue. While this work has been applied for one utility case, it has the potential to be used for other agencies facing similar issues across different fields.

ACKNOWLEDGMENT

The authors would like to acknowledge the support of the Southern California Edison Digital & Process Transformation and Customer Service Operations Departments. The views expressed in this document are solely those of the authors and do not necessarily reflect those of the Southern California Edison.

REFERENCES

Ali, A., Shamsuddin, S. M. and Ralescu, A. L., 2013. Classification with class imbalance problem. Int. J. Advance Soft Compu. Appl, 5(3), pp. 176–204.

Breiman, L., 2001. Random forests. Machine learning, 45, pp. 5–32.

Cortes, C., 1995. Support-Vector Networks. Machine Learning.

Ester, M., Kriegel, H. P., Sander, J. and Xu, X., 1996, August. Density-based spatial clustering of applications with noise. In Int. Conf. knowledge discovery and data mining (Vol. 240, No. 6).

Lacity, M. C. and Willcocks, L. P., 2016. A new approach to automating services. MIT Sloan Management Review, 58(1), pp. 41–49

Menard, S., 2001. Applied logistic regression analysis. SAGE publications.

Mugabi, J., Kayaga, S. and Smout, I., 2007. Why water utility customers don't pay their bills promptly.

Ogura, N., 2013. A systems approach to reducing utility billing errors (Doctoral dissertation, Massachusetts Institute of Technology).

Provost, F. J., Fawcett, T. and Kohavi, R., 1998, July. The case against accuracy estimation for comparing induction algorithms. In ICML (Vol. 98, pp. 445–453).