ANALYTICS USE CASE | Utility

Predictive Analytics Applied to Manual Exception Process

At a publicly traded gas utility company, predictive analytics helped reduce manual workload by 89%.

Each month, utilities create customer bills in multiple cycles. For a utility with one million or more meters, it is common practice to break up the territory into pre-defined routes to collect the meter customer usage data. This is necessary to calculate a customer's usage and generate a bill. For each cycle processed, it is not uncommon for the utility to receive meter readings that indicate zero consumption, i.e. no gas was used by the customer that month. Many months of the year the number of non-registering meters can be in the thousands.

The utility must evaluate the non-registering meters, or NONRs, to determine if the meter is malfunctioning, and, possibly dispatch a field specialist to investigate. The challenge? Rolling a truck to investigate is expensive. How can you determine if the NONR indicates that a meter is faulty? They don't typically malfunction. Compounding the issue, it could be related to normal customer behavior. Many customers may not use gas for a period of time when cold months are particularly warm, while they vacation, or when they relocate to their alternate residence.

Over the years the utility had developed a team that studies NONRs to make the determination whether to roll a truck to investigate. They rely on their judgement, the customer's past usage patterns, and what they are observing with other customers in that area based on weather. It is an imperfect approach, and, at times the team may be overwhelmed with a large volume of NONRs that they can't fully review them all.

The recently formed Corporate Analytics team was looking for an opportunity and business partner to try out their new Big Data tools. The Use Case seemed solid. With 5,000 NONRs to be reviewed and a dedicated team to review them, there was real savings opportunity. After taking time to study the business process and all the data associated with the process, the team developed a predictive modeling solution that was able to categorize the NONR exceptions as either: "review for potential meter failure" or "do not review". The model was 91% accurate in its determination in a test environment.

The next step was to gain the confidence of the team and to deploy the model in the production environment. A "supervisory" dashboard was developed to monitor the model's performance and prioritize the work, but the group continued to work the exceptions manually. Within a month the model was tweaked and consistently achieved 98% accuracy. The team was seeking approval to let the model autonomously generate an investigate work order, based on whether the NONR was determined to require field review. This is what we refer to as "narrow AI", a machine learning model that can make an autonomous decision for a narrowly defined business function.

Additional Background

The advanced analytics and Data Science lifecycle encountered several challenges along the way and surfaced valuable insights.

Coordinating with the very busy Customer Billing support team and getting on their schedule to extract the data took time. Since the science lifecycle is iterative, they didn't get all the data they needed and had to go make additional requests, requiring additional time.

While studying the current state process, which had a simple database and front end screen for the business to track their exception reviews, a few flaws in the process surfaced.

- Items that were manually determined to be reviewed generated an investigate work order; however, the results of that investigation did not feed back into their process to improve learning over time.
- To reduce and manage workload, business process rules were developed that missed the opportunity to catch a faulty meter. For example, items that had been reviewed, regardless of the manual determination, were flagged as not requiring another review for 12 months. The database software was deleting them from the queue for consideration. While researching the data the team found meters that had been intermittently generating zero reads over a period of months.

  Another rule the team used was to check the meter's usage for that month, in the previous year, to see if it was also zero, indicating that the customer may not be home. While researching the data, the team discovered several instances where the meter had not registered consumption from the day it was installed. In a few cases, meters had not been registering consumption for over 5 years.

The team spent a significant amount of time preparing the data so that it could populate a random forest statistical classification model, designed to perform the prediction. Several discussions were required with IT and the business to understand what the data elements were and how they were used in the system. At times the IT experts had a different understanding than the business users who performed the process. This surfaced potential issues with the software design and other monthly billing procedures.

To improve the model, the team wanted to use as many years of customer consumption history as was available. While reviewing historical data, several of the data fields had changed over time due to changes in the software and business process. Time was required to reconcile any discrepancies. Some of these discrepancies were given to the software support team to investigate the behavior of the production system.

Next, the team wanted to add additional data, such as weather history and results data from leak investigation work orders to improve the predictive model's ability to identify a faulty meter condition.

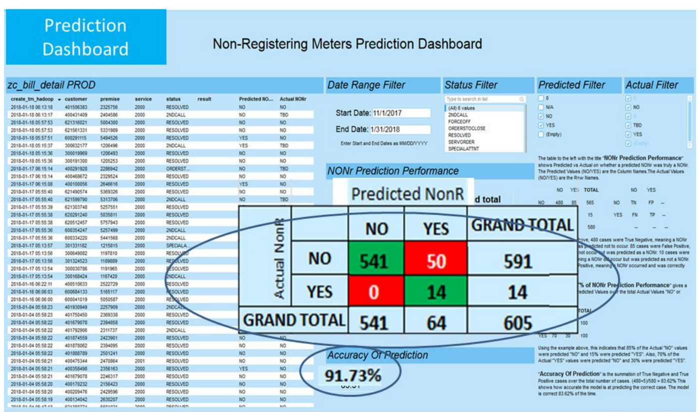
A major learning from this effort was that over 80% of the time spent developing this predictive model was spent accessing, cleaning, and preparing data to make it fit for use.

Because this was the team's first predictive analytics model, a significant amount of time was spent developing the Data Science workflows and methods to build and manage the model.

The team prepared a test dataset with actual NONR exception cases and results data that connected back to the original exception case. From this dataset they extracted “training data” used to populate the model and generate the array used for prediction. They also set aside a sample subset of data to process through the model and validate the prediction accuracy.

The next major step was to design how to integrate and deploy the model into the enterprise architecture. This required following traditional IT development and test procedures. Working with the IT team that developed the current process and technology to extract the billing data and direct it to the NONR exception database, the team integrated a feedback loop from the investigation work order results. The model was fully tested in a development environment, followed by a user acceptance environment.

The design called for the model to run in parallel for a period of time so that the business could monitor and validate predictions. This required the analytics team to develop a Supervisory Dashboard to monitor the predictive model. This dashboard provided statistics on the number of records presented to the model, the number that could be predicted, and a prediction accuracy index. This led to some design iterations to improve the model’s accuracy from 91% to 98%.



Once everyone was in agreement on the model’s performance, it was scheduled with the normal IT release management group and deployed to production.

The analytics team also developed the source code, version control, and documentation procedures to ensure the statistical methods used as well as operation of the model could be understood for future development, which was likely to occur. This was also important to trace the evolution of the model as it migrated from the research environment, to the test environment, and ultimately into production.

The total development timeframe from start to finish, working with a new team and new tools, was ten months. The first seven months focused on data access, cleansing, and preparation as well as the development of the predictive model script. The next three months focused on creating the Data Science environment to manage the model and integrate it with the enterprise architecture.

The Data Science environment was designed to support the development and deployment of complementary models that could act as a “champion/challenger” to the original model. The team also anticipated how to design the environment to support a multiple model “ensemble” approach which uses different models (using different statistical methods) to average the results of all of them. You probably have heard television meteorologists refer to the American and European "ensemble" used to predict severe weather patterns such as a hurricane's trajectory.

Deployed at month ten, the model predicted with 92% accuracy. In one additional month, making adjustments to the model's "confusion matrix," which allows the developer to apply different tolerances, the model performed with 98% accuracy.

The technologies used were:

- Tibco EAI middleware to extract customer data and load it to Hadoop
- Spark and other open source tools to bring data through the Hadoop distribution environment
- R scripting language and Apache Spark to develop the predictive script
- RStudio for model management and deployment
- Git for version control
- Tibco iProcess to develop the business process flow to accommodate the model in the enterprise architecture
- Tibco Spotfire to create the Supervisory dashboard

Summary Key Points:

- Finding a team performing a highly manual exception review afforded the interest and sponsorship to develop this application.
- 80% of the time was accessing, cleansing, and preparing the data to make it fit for use
- Developing the code to present the data to the model was over 1000 lines. The actual prediction script was 7 lines.
- Conclusion? Data Science was the easy part. It was 10% inspiration and 90% perspiration!

Keller Schroeder's Data Strategy Group

At Keller Schroeder, we absolutely subscribe to the idea that Data Science, Machine Learning, and Artificial Intelligence are skills that every organization should have, and, in a connected, social media, *Internet of Everything* world, are vital to your company's future. **We think it is that simple - not easy, but simple.**

Keller Schroeder's Data Strategy Consulting Services are available to help guide you through the Framework and Technical resources to assist you with the data lifecycle management activities, on an ongoing basis or until you are comfortable.

Our Data Strategy Group also has access to a Trusted Partner network that can provide the platform, tools, and skills you will need if don't already have them. For larger organizations, we also have partnerships with data lifecycle management practitioners and can scale resources to meet your needs.

Prepared By
Tom Vargo
Director, Data Strategy Group
812-492-7341 | tvargo@kellerschroeder.com