

Optimizing Water Pipeline Renewal

A Water Distribution Company Uses RapidMiner to Decide Where to Invest in Pipeline Rehabilitation & Replacement

Executive Summary

- A water distribution company needed to predict which parts of its pipeline had the greatest likelihood of failure (LoF)
- [Expert Infrastructure Solutions \(EIS\)](#), a consultancy working with the company's Asset Integrity Management department, used data science to predict pipeline leaks and breaks
- By overlaying the potential consequences of each predicted failure, EIS helped the company adopt a monetized risk approach to managing the pipeline
- As a result, the company's pipeline operations are now safer, while spend decision-making is more efficient

About the Company and the RapidMiner User Group

The focus of this case study is a water distribution company in the United States. The company is privately held, and provides water and wastewater services to consumers and businesses in several regions in the country - much like municipal water systems do in other areas. The company buys water from a variety of sources and delivers and sells it to customers through a pipeline distribution network.

The water pipeline network is a critical part of the company's operations. The company's network is typical of water distribution systems, in terms of its age and health. A large portion of the nation's one million miles of water distribution pipelines was built before World War II, and is likely to reach the end of useful life within roughly twenty five years.

The company has an Asset Integrity Management department, which is responsible for ensuring the well being of the pipeline network, including identifying sections of the network needing repair and overseeing repairs when they happen. To aid in this effort, the company has contracted with Expert Infrastructure Solutions (EIS), Inc., out of Denver, Colorado, to lend its data science and machine learning expertise and help the company increase the precision and effectiveness of its asset management process. EIS is a certified partner of RapidMiner, and the EIS team members are the primary RapidMiner users on behalf of the company.

The Company's Need

Better predict likelihood of failure (LoF)

The company hired EIS to use its data science expertise to help it better identify the parts of the pipeline that represented the highest overall risk, in terms of likelihood of failure (LoF) and potential consequences should a leak or break occur in that location. The company needed to optimize its repair investments, as the company operates on razor thin margins, given how little consumers are accustomed to paying for water. The company also often applies for government grants and loans to subsidize pipeline renewal, and needs to justify its request for funds and how and where they will be used, usually in terms of LoF.

“Water distribution companies have an obligation to their customers and their communities to ensure safe, clean and reliable drinking water.” said Michael Gloven, managing partner at EIS. “At the same time, these companies generally have limited renewal budgets. Progressive water distribution companies are starting to use everything at their disposal, including advanced data science and machine learning techniques, to predict LoF, so they can spend wisely.”

Small leaks in water pipelines are common, especially considering how old some of the infrastructure is. In fact, the water distribution industry finds and repairs hundred of thousands of leaks per year. Even more leaks are never found, resulting in billions of gallons of water lost.

Despite the company's best efforts to ensure pipeline integrity, the risk of not finding or fixing a critical leak or break remains high. The ramifications of failure to proactively address a pipeline network issue can range from merely inconvenient to disastrous. Customers have little patience for lack of water service, understandably. Beyond that, leaks or breaks in the wrong places can result in significant property damage, as anyone who has witnessed a water main break knows. And there's always the risk of a leak or break resulting in contamination of water being delivered to homes and businesses. While not attributed to a leak, per se, the fallout from the calamity in Flint, Michigan looms large over any water distribution company's operations.

The Solution

RapidMiner enables building multiple predictive models and weighing the unique, pipeline-related trade-offs of each

The company has a large volume of data available for use in building predictive models. Data about the pipeline itself is the starting point - the specific design, the diameter of pipes, pipe wall thickness, the metallurgy of the pipes, how exactly they were welded, and much more. Geo-spatial data is also important, as it describes the areas the pipeline networks run through, including buildings, houses, roads, waterways, etc. This not only impacts the likelihood of failure, but also the damage that might be done should a leak or break occur. Data on the performance of the pipeline rounds out the set, covering leak surveys, and corrosion surveys. This data also includes the history of pipeline failures, which serves as the independent variable when building models.

“There’s a lot of data available, but it’s not always ready for us to use to build models,” said Gloven. “In some cases, the company collects data but doesn’t know exactly how best to use it, so it just sits there. We use RapidMiner’s data prep functions to get the data in shape for modeling.”

Once the data has been prepped, RapidMiner puts over a hundred different modelling approaches at the fingertips of Gloven and his team. They can quickly build multiple models and compare results and predictive power. The EIS team makes heavy use of RapidMiner’s confusion matrices to provide transparency into the performance of the different models, which is important given the unique nature of the work EIS is doing. When trying to predict potential leaks and breaks, the company needs to understand and make very deliberate choices when weighing the trade-offs between false positives, which might result in digging a hole to find a leak that isn’t there (costly, but not a big deal), and predictive misses which could result in not anticipating and fixing a leak (which could cause damage or contamination).

“RapidMiner plays a vital role in our work to predict pipeline leaks and breaks,” said Gloven. “It’s the perfect tool for either a subject matter or business domain expert who wants to leverage the value of machine learning through an intuitive interface, easy-to-get results and and the ability to share insights with others. It also integrates easily with all our data sources and visualization tools, such as Tableau & ArcMap.”

The Results

RapidMiner has helped pinpoint the right repairs to make proactively

The impact of EIS’s work using RapidMiner on the company’s asset integrity management is substantial. Better identifying the sections of the pipeline needing repair increases the chance of fixing leaks or breaks before they happen. This reduces the risk of significant property damage and, most importantly, negatively impacting consumers’ health.

But the impact of data science goes beyond these top-line concerns. It also helps the company with more day-to-day issues, such as optimizing the allocation of repair budget to the parts of the pipeline that need it most. This allows the company’s asset integrity management department to have a bigger impact without needing additional resources.

“Water distribution companies need to look beyond routine pipeline repair, such as saying ‘we’ll just replace 1% of the network each year,’” said Gloven. “They need instead to adopt a monetized risk approach, applying data science to achieve more efficient repair operations. It’s a win on every front. Every company with a network of any kind - other kinds of fuel pipelines, power grids, telecommunications networks, even roads and bridges - should take note, as they could use data science in the same way with similar benefits.”