Applying Predictive Analytics to Process Safety Leading Indicators

William R. Brokaw and A.W. Armstrong
Kestrel Management LLC
Author Email: wbrokaw@kestrelmanagement.com

In recent years, there has been significant research attention in developing and measuring leading indicators in the process safety industries. Leading indicators can be defined as safety-related variables that proactively measure organizational characteristics with the intention of predicting, and subsequently avoiding, process safety incidents. Conversely, lagging indicators are safety-related variables that have already occurred, such as total recordable incident rate (TRIR) or days away, restricted, or transferred duty rate (DART). Lagging indicators are helpful in documenting past trends in safety performance; investigating their occurrence can help prevent future incidents through the implementation of lessons learned.

Implementing lessons learned from lagging indicators, however, requires an incident to occur first. Leading indicators are preferable because, if measured and analyzed correctly, they can help prevent incidents from ever occurring. Leading indicators become especially powerful when combined with the use of advanced statistical methods, including predictive analytics.

Predictive analytics is a broad field encompassing aspects of various disciplines, including machine learning, artificial intelligence, statistics, and data mining. Classification algorithms, which predict future occurrences based on past data, could be particularly beneficial to the process industries. A classification algorithm, such as logistic regression, neural network, support vector machine, or naïve Bayes classifier, could be used to predict the timeframe and location of process incidents based on leading indicators. There are two main challenges associated with this method: (1) ensuring that leading indicators are actually predictive of incidents, and (2) measuring the leading indicators frequently enough for them to have predictive value.

A case study to illustrate this process will be discussed. The author developed a predictive analytical model for the transportation industry that we believe is applicable to the process industries. Using regularly updated inspection data, an incident predictive model was created using a logistic regression modified by Firth’s penalized likelihood method due to the low ratio of events to misses. The resulting model provides incident probabilities for each transportation segment over a six-month period (the approximate time between full data updates). Additionally, the model identifies the variables with the most predictive validity (i.e., those that are significantly contributing to incidents), thereby showing which factors are most likely to prevent incidents if they are addressed. During the validation period, the model showed that 7% of
transportation segments were at relatively high risk of having an incident, and 43% of the actual incidents occurred on those segments.

The same methodology could be used in the process industries to predict and prevent incidents, provided that organizations:
1. Identify leading indicators with predictive validity
2. Measure these indicators at regular intervals
3. Create a predictive model based on the measured indicators
4. Deploy the model whenever leading indicator measurements are taken to calculate predicted incident probabilities

**Keywords:** Continuous improvement, metrics, risk management