

PREVENDO CONDIÇÕES CRÍTICAS E INDESEJADAS DE PLANTA COM ANALÍTICA AVANÇADA*

Daniel Codo de Faria¹
Vinícius Hudson Alves²
Fábio Barros de Carvalho³

Resumo

O desafio de se extrair conhecimento prático das enormes massas de dados disponibilizadas pelos sistemas de informação e automação nas indústrias já vem sendo discutido e enfrentado há anos. Neste contexto, a aplicação de técnicas de analítica avançada se apresenta como uma solução poderosa, capaz de mudar a operação dos processos industriais. Especificamente, a área de analítica preditiva é constantemente considerada como uma nova revolução para a indústria. Combinando técnicas como análise estatística, mineração de dados, “big data”, aprendizado de máquina e inteligência artificial, a analítica preditiva gera modelos a partir de dados históricos com o objetivo de prever condições futuras. Tal previsão permite que ações de correção ou otimização possam ser definidas e executadas antecipadamente e com maior precisão, aumentando significativamente a eficiência dos processos, seja em relação a produtividade, custo ou qualidade. Este trabalho apresenta os fundamentos da analítica preditiva e uma aplicação prática na indústria. As técnicas utilizadas podem ser diretamente replicadas nas mais diversas áreas da indústria e para os mais diferentes problemas, evidenciando o forte potencial de expansão esperado para os próximos anos.

Palavras-chave: Analítica preditiva; Analítica avançada; Big data; Aprendizado de máquina; Inteligência artificial; Modelagem; Otimização de processos industriais.

PREDICTING CRITICAL AND UNDESIRABLE CONDITIONS IN PLANT WITH ADVANCED ANALYTICS

Abstract

The challenge of extracting practical knowledge from huge data sets of industrial information and automation systems has been discussed and faced for years. In this context, the application of advanced analytics' techniques appears as a powerful solution, capable of changing the operation of industrial processes. Specifically, predictive analytics is constantly considered as a new revolution for industry. Combining techniques such as statistical analysis, data mining, big data, machine learning and artificial intelligence, predictive analytics generates models from historical data for predicting future conditions. Such prediction allows corrective or optimization actions to be defined and executed in advance and with higher precision, significantly increasing process efficiency on productivity, cost or quality. This work presents the fundamentals of predictive analytics its application on the industry. The techniques applied can be directly replicated to any industrial area and several different problems, showing the strong potential for expansion expected for the next years.

Keywords: Predictive analytics; Advanced analytics; Big data; Machine learning; Artificial intelligence; Modeling; Industrial process optimization.

¹ Engenheiro de Controle e Automação, Mestre em Ciência da Computação, Gerente de Automação da Accenture Asset and Operations Services, Nova Lima, MG, Brasil.

² Engenheiro de Controle e Automação, Analista de Automação da Accenture Asset and Operations Services, Nova Lima, MG, Brasil.

³ Accenture Asset and Operations Services, Nova Lima, MG, Brasil.

1 INTRODUÇÃO

Centenas de variáveis de processo (como vazões, pressões, temperaturas, densidades, etc) historiadas a cada segundo durante anos de operação das plantas industriais, criam imensas bases de dados com um potencial incalculável para geração de conhecimento útil para otimização dos processos. A aquisição e historiamento destes dados através de PIMS (*Plant Information Management System*) já pode ser considerada uma prática consolidada e implementada pela grande maioria das empresas há vários anos. Entretanto, o uso efetivo destes dados ainda é, em geral, muito superficial. Telas e relatórios para visualização de indicadores chave de desempenho (KPIs), gráficos de tendência de variáveis e correlações básicas são exemplos de aplicações simples frequentemente encontradas. Apesar de proporcionarem um primeiro nível de visibilidade e análise, tais funcionalidades não automatizam a geração de conhecimento a partir da informação disponibilizada, deixando este desafio exclusivamente a cargo do engenheiro ou especialista envolvido nas análises.

Neste contexto, a prática da Analítica Avançada, já amplamente aplicada em áreas como marketing, finanças e medicina [1], pode ser vista como a próxima revolução na forma de gestão e operação de processos industriais complexos, substituindo incertezas por probabilidades e permitindo ações de correção e otimização mais precisas e eficazes [2]. Através do uso de técnicas sofisticadas de análise de dados (ex: estatística multivariada, modelagem, mineração de dados, inteligência artificial, simulação, otimização), a analítica avançada evolui os tradicionais conceitos de *business intelligence*, utilizando fatos históricos não apenas para entender o passado das operações, mas para identificar automaticamente padrões ocultos de comportamento e guiar a tomada de decisão [3].

A Analítica Preditiva é considerada uma das diversas sub-áreas da analítica avançada, focada em antecipar condições ou comportamentos futuros (com um nível aceitável de confiabilidade) a partir de modelos estatísticos gerados com dados históricos e presentes. Os chamados “modelos preditivos” são construídos utilizando-se técnicas computacionais avançadas de regressão ou aprendizado de máquina (como árvores de classificação, redes neurais, máquinas de vetores de suporte, etc) e conseguem explicar o impacto de vários fatores, conhecidos e desconhecidos e com os mais diversos níveis de correlação entre si, no comportamento de uma ou mais variáveis de interesse (como qualidade, custo ou produção).

Especificamente, quando aplicada em bases de dados muito grandes (tanto estruturados quanto não estruturados) e de processos produtivos muito complexos, a analítica preditiva se consolida como uma ferramenta poderosa para aumentar a eficiência operacional dos mais diversos setores da indústria. A antecipação de situações indesejadas e de alta criticidade para a operação permite a mobilização de esforços para evitar impactos negativos nos KPIs de negócio mais críticos da produção. Aumento da qualidade do produto final, redução de custos de manutenção, redução de consumo de energia e aumento de produtividade são apenas alguns dos resultados práticos possíveis de serem obtidos com o uso da analítica preditiva.

Este trabalho apresenta uma aplicação prática da analítica preditiva na indústria da mineração, discutindo a metodologia básica para o seu desenvolvimento, resumindo os resultados (já obtidos e potenciais) e evidenciando o vasto campo para replicação das técnicas em outras áreas e problemas do processo.

2 CONCEITOS BÁSICOS

A aplicação prática da analítica avançada pode ser feita através do uso de diversas técnicas e algoritmos, dependendo dos objetivos almejados, do processo envolvido ou da característica dos dados disponíveis. Nesta seção, os principais conceitos particularmente relevantes para a aplicação a ser apresentada neste trabalho são brevemente descritos, com o intuito de embasar as discussões e conclusões subsequentes.

2.1 Análise Multivariável

De forma geral um processo pode ser considerado bem controlado quando suas variáveis conseguem se manter dentro de certos limites de operação ótimos. Sendo atendidos esses limites, é atingida uma produção de boa qualidade. Porém muitas vezes distúrbios desconhecidos podem criar ocorrências nas quais mesmo com as variáveis sob controle, a relação entre elas é quebrada e por isso o resultado final não é satisfatório.

Ilustrando o caso acima temos a Figura 1, onde ambas as variáveis estão dentro de seus limites operacionais, porém ao se verificar a relação entre elas é possível perceber que os três últimos pontos se relacionam de maneira diferente que os demais.

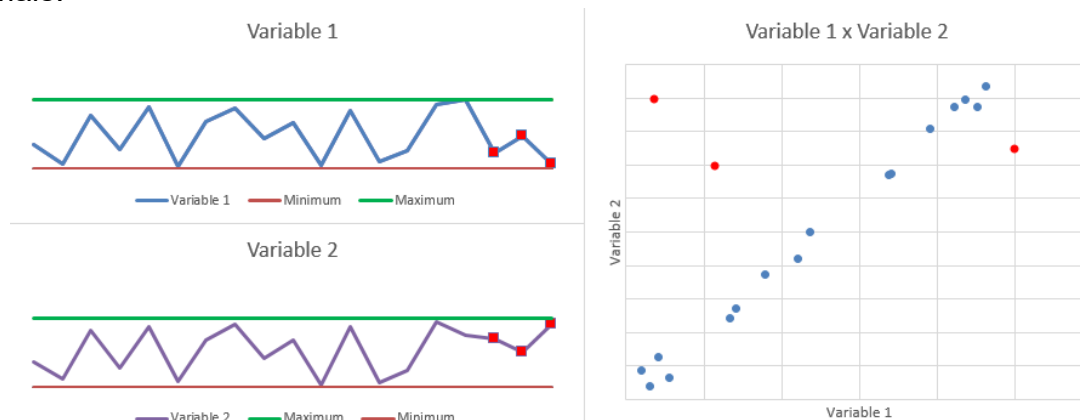


Figura 1. Distribuição por tipo de variável no modelo

É fácil imaginar esse processo para o monitoramento de algumas poucas variáveis, porém isso é ainda mais impactante em processos complexos nos quais existem milhares de medições sendo feitas a cada segundo. Até mesmo a verificação univariável fica inviável pois é provável que muitos alarmes sejam disparados ao mesmo tempo. E mesmo que algumas variáveis estejam fora de controle, pode ser que a relação multivariável que as governa ainda seja válida, deixando o processo dentro de sua operação normal.

2.2 Classificação de Padrões

Uma vez determinada qual a situação normal do processo, é possível dividir o comportamento multivariável do mesmo em pelo menos duas classes, “sob controle” e “fora de controle”, por exemplo. Como mostrado na Figura 1, essa situação seria quando a relação linear entre as variáveis é mantida.

Um modelo simples para esse processo seria medir a distância entre os pontos formados pelas coordenadas dadas por cada variável e a reta formada pela relação

esperada entre as duas. A partir do momento que essa distância passe de um limiar pré-determinado se considera que o processo saiu de controle e alguma ação precisa ser tomada para retomá-lo.

A abordagem descrita envolve um certo conhecimento de processo para determinar a relação esperada entre as variáveis, e não seria possível aplicá-la caso essa seja desconhecida ou não tão bem definida.

A fim de endereçar esse problema existem vários algoritmos que utilizam dados históricos para estimar qual seria o comportamento multivariável esperado para o processo em cada condição operacional definida. Esses algoritmos não esperam que os dados possuam uma relação específica e, portanto, conseguem dar bons resultados até mesmo para casos onde existem interações não lineares.

2.3 Árvores de Classificação e Florestas Aleatórias

Uma técnica simples para realizar a classificação de padrões é a árvore de classificação. Como o nome sugere, o algoritmo se baseia em uma árvore de decisão onde a cada passo é feita uma consideração sobre uma variável. Basicamente o algoritmo inicia dividindo o espaço total em dois através de um corte simples em uma das dimensões dos dados, e então continua a fazer isso com os subespaços restantes até que os critérios de parada sejam satisfeitos. A Figura 2 ilustra esse processo para dados em duas dimensões e pertencentes a três classes diferentes.

Para treinar uma árvore de classificação, são necessários os dados de treinamento juntamente com parâmetros que definem o critério de parada para o crescimento da árvore. Os parâmetros mais comuns são a profundidade máxima da árvore (quantas condições podem existir ao máximo) e o número mínimo de dados em cada folha (para garantir que a árvore não crie uma partição para cada dado individual).

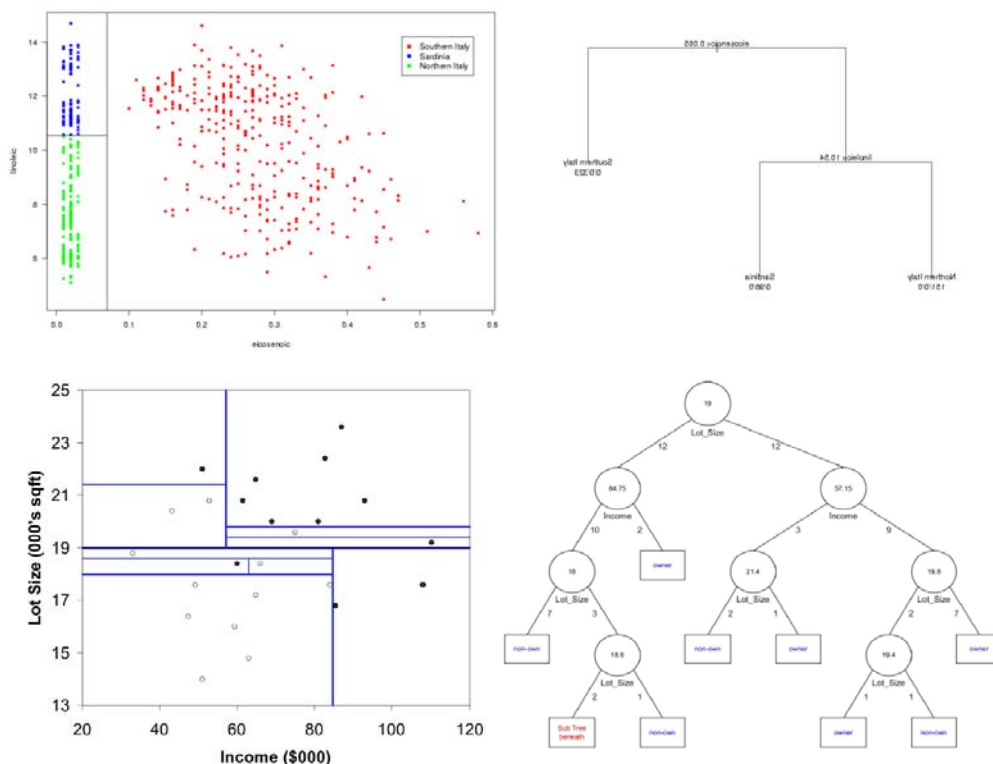


Figura 2. À esquerda, representação bidimensional de dados com três classes diferentes e sua divisão por uma árvore de classificação. À direita, a árvore resultante.

A fim de utilizar as vantagens da simplicidade desse modelo e minimizar algumas de suas limitações (como alta dependência da ordem das variáveis de corte ou problemas de ajustamento), uma técnica empregada e com bons resultados na prática é a combinação de múltiplas árvores com características selecionadas aleatoriamente. Esse tipo de abordagem é chamado de florestas aleatórias.

Uma floresta aleatória, portanto, contém um número muito alto de árvores treinadas utilizando um subgrupo de características e dados escolhidos ao acaso. A predição final é dada por uma votação, na qual caso o percentual de árvores que tenha como resultado uma predição específica acima de certo limite de corte, é decidido que a predição do modelo é essa classe.

Na Figura 3 temos uma ilustração de um modelo de floresta aleatória.

No modelo de florestas aleatórias os parâmetros mais comuns para serem ajustados são o número de árvores utilizadas, o número de características e dados sorteados para o treinamento de cada árvore, o parâmetro de corte para considerar pertencente a uma classe e todos os outros parâmetros necessários para treinar as árvores de classificação.

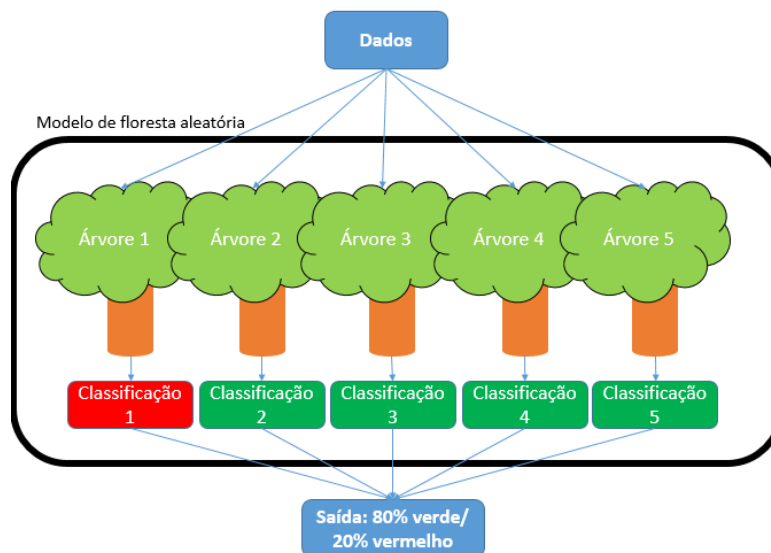


Figura 3. Ilustração do funcionamento de um modelo de árvore aleatória

2.4 Variáveis de Processo X Características do Modelo

Um aspecto do trabalho de análise de dados tão importante quanto a seleção de um modelo adequado é a seleção das características que serão analisadas pelo modelo. Essas características irão compor o espaço multidimensional sobre o qual os algoritmos de treinamento e predição trabalharão. Portanto, elas precisam incorporar as informações determinantes para separar as classes a serem identificadas.

Para explicar os conceitos utilizados no trabalho a seguir, uma importante distinção é entre variável de processo e característica do modelo. Uma variável de processo é uma grandeza medida que interfere no comportamento do processo, por exemplo o nível ou a vazão de entrada de um tanque. Para aplicar as técnicas de reconhecimento de padrão é necessário definir características dessas variáveis a fim de alimentar o modelo. Uma característica pode ser, por exemplo, o valor atual da medição, ou ainda um cálculo como média ou desvio padrão dentro de um determinado período de tempo. Essas características servirão de entrada para o modelo de aprendizagem de máquina prever o comportamento do sistema, portanto a sua correta definição é de suma importância para o sucesso da aplicação.

3 APLICAÇÃO PRÁTICA NA MINERAÇÃO – PREDIÇÃO DE PERDA DE QUALIDADE NA FLOTAÇÃO

3.1 Contexto e Objetivo

O processo de flotação é parte essencial do processo de beneficiamento do minério de ferro, visando garantir que níveis de impurezas estejam dentro das tolerâncias e os teores de minerais de interesse sejam maximizados. Na saída deste processo, é desejado que o percentual de silicatos seja o mais baixo possível, normalmente menor que 1% para cumprir com a especificação para os produtos de maior valor. Sua eficiência está associada ao controle de variáveis de nível, densidade, pH e adição de reagentes a fim de garantir as condições ótimas para a separação do minério e seu rejeito. Além disso existem outras variáveis de perturbação que possuem conhecido impacto no seu desempenho, como a granulometria e composição química do minério de entrada. A utilização de diferentes técnicas de controle pode ajudar a melhorar a estabilidade e qualidade do minério beneficiado. Porém dada a presença de todas as variáveis internas e externas ao processo que possuem influência no resultado final, é difícil garantir que não haverá pequenas falhas que podem passar despercebidas e levar a grandes problemas com perda de qualidade, gerando impactos técnicos e financeiros significativos. Outro fator desafiador é que a determinação da qualidade do minério beneficiado depende de análises laboratoriais, e estas normalmente possuem um atraso que pode ser de algumas horas. Por este motivo, a equipe de operação somente possui a confirmação de que o processo perdeu o controle após um considerável período de tempo. Sendo assim, uma ferramenta que seja capaz de alertar sobre eventuais problemas com antecedência, possibilitando que uma checagem seja feita e ações preventivas sejam tomadas com antecedência tem grande valor para a indústria.

3.2 Metodologia e Desenvolvimento

Utilizando dados históricos de um processo de flotação real, armazenados no PIMS, foi possível obter uma extensa base de dados para geração do modelo preditivo. Foram selecionadas 895 variáveis, distribuídas por todo o processo. Entraram no modelo estados dos equipamentos presentes no processo, correntes, velocidades e vazões de bombas, níveis de tanques e células de flotação, posições de válvulas e medições provenientes de laboratório como composição química da polpa em pontos intermediários do processo (Figura 4).



Figura 4. Distribuição por tipo de variável no modelo

Utilizando os conceitos apresentados anteriormente, as características para as variáveis de processo foram definidas como o valor mediano em janelas de 15 minutos que iniciam no tempo corrente e vão até duas horas no passado. Para as variáveis de qualidade, as características foram definidas como o valor medido a cada hora entre os períodos de duas e cinco horas no passado (devido ao atraso dos resultados de laboratório). A Figura 5 ilustra essa definição. Ao final, o modelo incluiu mais de 4500 características como entrada.

Dados de sensores	Dados de laboratório	Objetivo de predição
Mediana dos valores entre [t, t-15m]		Nível de SiO ₂ dentro do limite em t
Mediana dos valores entre [t-15m, t-30m]		
Mediana dos valores entre [t-30m, t-45m]		
Mediana dos valores entre [t-45m, t-60m]		
Mediana dos valores entre [t-60m, t-75m]		
Mediana dos valores entre [t-75m, t-90m]		
Mediana dos valores entre [t-90m, t-105m]		
Mediana dos valores entre [t-105m, t-120m]		
	Valores em t-2h	
	Valores em t-3h	
	Valores em t-4h	
	Valores em t-5h	

Figura 5. Definição das características de acordo com as variáveis medidas

Um modelo de florestas aleatórias foi então treinado utilizando-se 6 meses de dados, sendo os primeiros 3,5 meses para o treinamento propriamente dito, 1,2 mês para validação e 1,2 mês para teste. Os parâmetros que geraram os melhores resultados foram identificados e o modelo foi implementado para realizar as predições em tempo real no processo. Após um período de aproximadamente 53 dias em operação, seu desempenho foi avaliado. Finalmente, de posse do alarme adiantado foi possível então desenvolver um plano de ação juntamente com a equipe de operação da planta, com o objetivo de atuar antecipadamente no processo de maneira a fazer com que o impacto final na produção seja reduzido.

3.3 Resultados

A validação do modelo desenvolvido mostrou que os melhores resultados foram alcançados com a utilização de 20.000 árvores utilizando 10 características. Esse procedimento também serviu para definir o parâmetro inicial de corte de 0,56, significando que apenas quando menos de 56% das 20.000 árvores utilizadas decidirem que a predição é de piora na qualidade, um alarme deve ser gerado. No total foram avaliadas 1284 predições (equivalente a 1284 horas) de eventos de fora de controle de qualidade feitos pelo modelo com as respectivas medições de laboratório. Quando a medição de laboratório ultrapassou o limite estabelecido, de 2,7% de conteúdo de SiO₂, foi computado um evento fora de controle. Quando a saída do modelo foi menor que o limite de corte estabelecido, de 0,56, foi computada uma previsão de que a qualidade sairia de controle. O resultado da execução em tempo real está mostrado na Figura 6 abaixo, que mostra todos os pontos amostrados durante a execução.

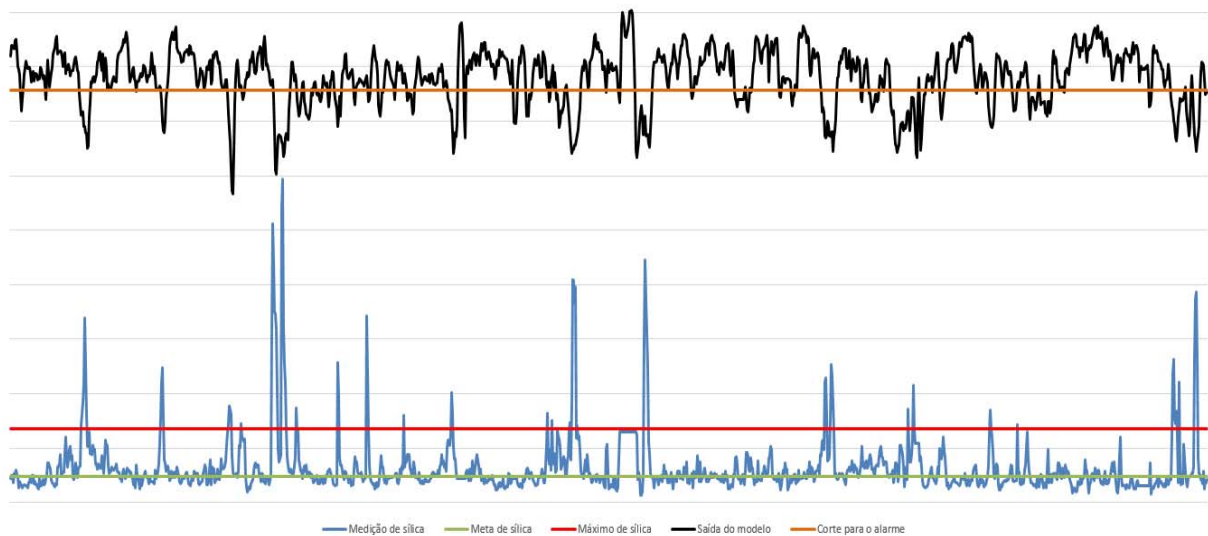


Figura 6. Resultados da execução do modelo em tempo real. Em azul está o percentual de impurezas no produto final, em vermelho o limite máximo usado pelo modelo para gerar alarmes, em verde a meta de qualidade utilizada pela operação, em preto a saída do modelo e em laranja o limite de corte do modelo definido para essa etapa do teste.

A tabela abaixo resume os resultados obtidos. Dentre as 1.283 horas de dados, durante 71 horas a saída do processo ficou com o índice de impurezas acima do limite máximo definido, e o modelo acertou esse comportamento em 60 ocorrências. Durante o tempo restante (1212 horas), o processo não violou o limite de qualidade. O modelo acertou essa previsão 923 vezes.

Tabela 1. Resultados obtidos no período de testes

	Predição: fora de controle	Predição: dentro de controle	Percentual de acertos do modelo na condição
Real: fora de controle	60	11	84,5%
Real: dentro de controle	289	923	76,15%

Os resultados corroboram que o evento analisado é raro, ocorrendo apenas em torno de 5,53% do tempo total (71 de 1283 horas). Além disso, analisando a série temporal é possível perceber que uma vez que ocorre um primeiro evento de qualidade fora da faixa, outros normalmente se seguem. Para a operação da planta um número mais importante que o percentual de tempo que o modelo está certo ou errado, é quantas vezes ele consegue detectar corretamente que o processo que estava sob controle saiu da especificação.

Na Figura 7 abaixo, é possível verificar um exemplo de perda do controle. Após sair de controle (segundo ponto após a linha vermelha vertical), o processo permanece nesse estado por mais seis horas. Neste caso, o modelo foi capaz de antecipar que o problema ocorreria quatro horas antes da confirmação de laboratório (duas horas antes do evento real), o que possibilita a execução de ações preditivas por parte da operação da planta para evitar ou reduzir os impactos do desvio.

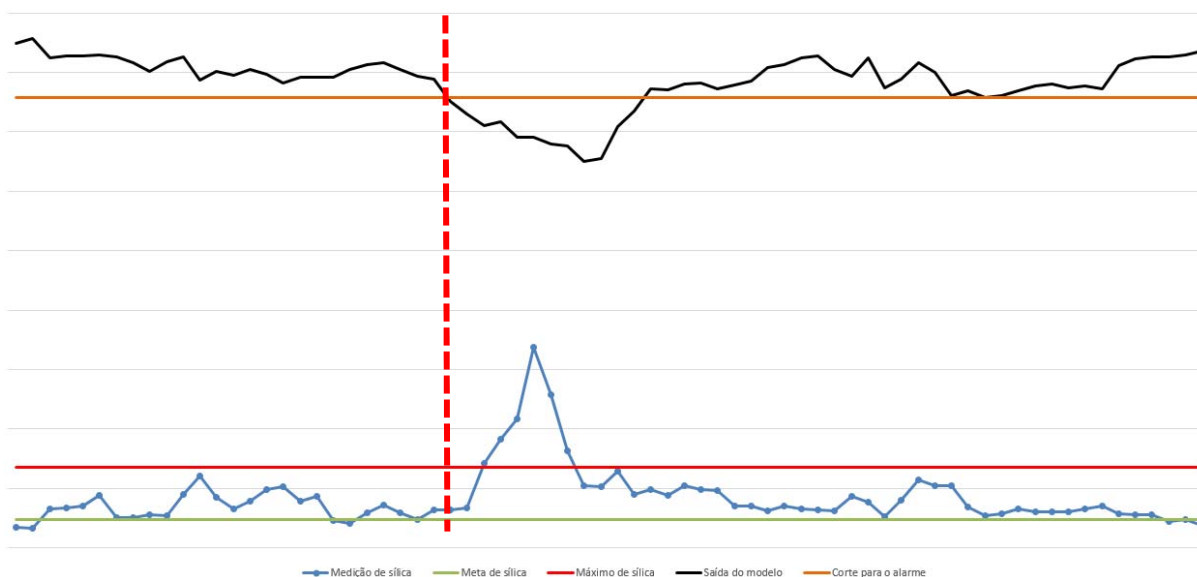


Figura 7. Exemplo de um pico detectado com mais de duas horas de antecedência. Em azul está a qualidade, em verde a saída do modelo e vermelho o limite de corte do modelo. A linha vertical vermelha mostra o momento no qual o alarme foi gerado.

Os resultados dessa análise estão na Tabela 2 abaixo, onde o número de eventos considera apenas a contagem da transição “sob controle” para “fora de controle”, e não quanto tempo o processo ficou fora de controle.

Tabela 2. Resultados obtidos de acordo com o número de picos

	Número de eventos	Percentual
Eventos fora de controle	21	100%
Eventos detectados com antecedência (mais de 1 hora)	16	76,2%
Detectado apenas com a confirmação de laboratório	5	23,8%

Podemos verificar que em aproximadamente 76% dos casos o modelo de floresta aleatória foi capaz de prever corretamente que a qualidade do produto sairia da especificação. Isso significa que com a utilização desse alarme os operadores poderiam agir antecipadamente na maioria das ocorrências e conseqüentemente diminuir a duração das mesmas ou até mesmo evitá-las totalmente.

Outro aspecto importante a ser analisado é o número de alarmes falsos (ou falsos positivos) gerados. Ou seja, momentos em que o modelo previu que algo aconteceria, mas a confirmação do laboratório mostrou que a qualidade ainda estava no limite especificado. Como o evento de interesse é raro (aproximadamente 5% do tempo), alguns falsos positivos são admissíveis. Verificaram-se 22 casos de prováveis falsos positivos, sendo que em 36,4% (8 eventos) houve uma queda repentina na qualidade medida (mas sem violar o limite). Estes eventos são considerados como prováveis pelo fato da equipe de operação da planta atuar constantemente para corrigir desvios no processo, o que pode ter ocorrido em alguns destes momentos.

Os resultados obtidos são muito dependentes da escolha do parâmetro de corte no qual o alarme será gerado, e existe uma relação de compromisso entre o número de alarmes falsos e detecções corretas gerados (Figura 8). Foi escolhido um parâmetro mais conservador (0,56), porém, tanto o horizonte de predição quanto a quantidade

de predições corretas poderiam ser alterados com uma política de escolha do parâmetro de corte diferente.

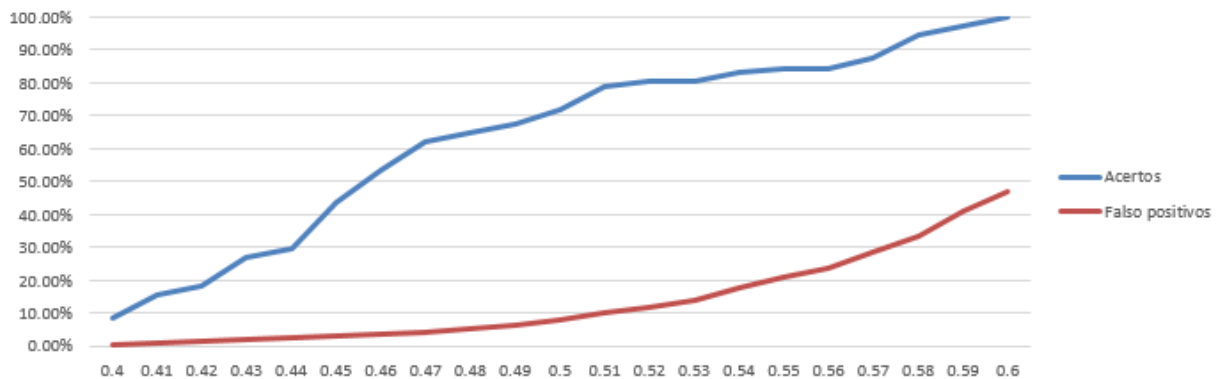


Figura 4. Compromisso entre número de acertos e alarmes falsos com a escolha do parâmetro de corte

Um plano de ação para redução dos eventos de baixa qualidade ou, pelo menos, de sua duração, a partir dos alertas antecipados foi discutido com a operação da planta, e será colocado em prática. Uma tela, desenvolvida no próprio PIMS, para monitoramento dos alertas e registro das ações efetivamente realizadas também será disponibilizada na sala de controle. Ao final deste período de operação assistida, as novas ocorrências de problemas de qualidade serão comparadas a períodos anteriores com o objetivo de comprovar o valor do sistema.

4 CONCLUSÃO

Extraír conhecimento útil das centenas ou milhares de variáveis de processo armazenadas ao longo de anos de operação de uma planta industrial é um desafio reconhecido e enfrentado constantemente por engenheiros e gestores. O uso de técnicas avançadas de análise de dados promete ser uma nova revolução neste contexto e tem potencial para trazer um nível de visibilidade de produção nunca antes observado pela indústria. A aplicação prática de analítica preditiva apresentada neste trabalho, focada na detecção antecipada de problemas de qualidade no processo de flotação no beneficiamento de minério de ferro, demonstrou tal potencial. Com um índice de assertividade de mais de 80% e alarmes com até 4 horas de antecedência em relação à medição de laboratório, o sistema desenvolvido apresenta fortes indícios de seu valor. A antecipação das ações, já conhecidas, de correção ou prevenção de problemas de qualidade deverá trazer benefícios visíveis para o processo. Finalmente, considera-se que a metodologia e técnicas utilizadas neste trabalho podem ser replicadas para os mais diferentes problemas, nas mais diversas indústrias.

REFERÊNCIAS

- 1 Siegel, Eric. How Predictive Analytics Reinvents These Six Industries. Information Management. 2015.
- 2 Global Industry Analysts, Inc. Press Release MCP-7902. Predictive Analytics – A Global Strategic Business Report. 2015.
- 3 Franks, Bill; Vanvalkenburg, Scott. Moving from Business Intelligence to Advanced Analytics. Information Management. 2010.