

Previsão de Falta de Equipamento de Carga para Frota de Transporte com base em Aprendizado de Máquina e Cenários de Despacho

Guilherme Gouveia Cordeiro* Marcone Jamilson Freitas Souza**
Frederico Gadelha Guimarães***

* *Programa de Pós-Graduação em Instrumentação, Controle e Automação de Processos de Mineração, Universidade Federal de Ouro Preto e Instituto Tecnológico Vale; Vale S.A,*
(e-mail: guilherme.gouveia.cordeiro@aluno.itv.org).

** *Departamento de Computação, Universidade Federal de Ouro Preto (UFOP) (e-mail: marcone@ufop.edu.br)*

*** *Machine Intelligence and Data Science (MINDS) Lab, Departamento de Engenharia Elétrica, Universidade Federal de Minas Gerais (UFMG) (e-mail: fredericoguimaraes@ufmg.br)*

Abstract: A current challenge in fleet management in the mining industry is the prediction of process failure events, due to the high number of internal influence parameters (such as equipment breakage), external to the process (such as bad weather) and interface (such as shutdown of crushing). This paper deals with the development of a machine learning approach to predict the Lack of Load Equipment event for the transport fleet using dispatch data. The model applied was the Random Forest varying the number of trees between 20, 50, and 100. Backward Elimination was performed to reduce the number of attributes from 20 to 5. The metric used to evaluate the models was the correlation factor (R). Real operation data was used to build the model containing 860,000 records gathered for four months. Through the Shapiro-Wilk and T-Welch tests, it was verified that the models of 20, 50, and 100 trees were equivalent. The mean correlation coefficients of these models were 0.768, 0.773, and 0.773, respectively, which are better than the value of 0.35 obtained with the linear regression method currently used by the company's decision-maker. Thus, machine learning through Random Forest showed good results and applicability for predictions and classifications of events in fleet management in the mining industry.

Resumo: Um desafio vigente na gestão de frota do setor de mineração é a predição de eventos de falhas de processo, devido à alta quantidade de parâmetros de influência interna (como quebra de equipamentos), externa ao processo (como intempéries) e de interface (como parada de britagem). Este artigo trata do desenvolvimento de uma abordagem de aprendizado de máquina para previsão do evento Falta de Equipamento de Carga para a frota de transporte utilizando dados de despacho. O modelo aplicado foi o *Random Forest* variando-se o número de árvores entre 20, 50 e 100. Realizou-se a seleção de atributos por meio de *Backward Elimination* para a redução de 20 para 5 parâmetros. A métrica utilizada para avaliação dos modelos foi o fator de correlação (R). Dados de operação real foram utilizados para construção do modelo contendo 860 mil registros em quatro meses de coleta. Verificou-se, através dos testes de *Shapiro-Wilk* e *T-Welch*, que os modelos de 20, 50 e 100 árvores eram equivalentes. Os resultados das médias dos coeficientes de correlação desses modelos foram 0,768, 0,773 e 0,773, respectivamente, os quais são melhores do que o valor de 0,35 obtido com o método de regressão linear, atualmente utilizado pelo tomador de decisão da empresa. Assim, o uso do aprendizado de máquina através do *Random Forest* apresentou bons resultados e aplicabilidade para predições e classificações de eventos da gestão de frota na indústria de mineração.

Keywords: Mining; Industry; Process; Hang time; Haul truck; Random Forest.

Palavras-chaves: Mineração; Indústria; Processo; Tempo ocioso; Caminhão-fora-de-estrada; Random Forest.

1. INTRODUÇÃO

A indústria de mineração apresenta um papel importante na economia e na geração de milhares de empregos no Brasil. De acordo com dados do governo (MME, 2021), em 2020 o setor mineral sozinho foi responsável por 2,5p.p do PIB brasileiro, garantindo tributação de R\$ 6,1 bilhões somente em impostos no mesmo período. Esta representatividade também fez com que o setor se movimentasse em prol de iniciativas de governança ambiental, social e corporativa e demandasse o uso de novas tecnologias e otimizações de processo. Dentre as tecnologias aplicáveis à mineração, destaca-se o aprendizado de máquina. De fato, dado seu sucesso, houve um crescimento vertiginoso de aplicações desses métodos na última década (Jung and Choi, 2021). Porém, a construção dos modelos demanda a contribuição não somente de especialistas em inteligência computacional, mas, também, de especialistas no tema da aplicação. De acordo com Jung and Choi (2021), há uma escassez de trabalhos com este escopo na indústria de mineração e, em especial, na gestão de frota.

A lacuna de aplicação de métodos de aprendizagem de máquina apresentada por Jung and Choi (2021), juntamente com a pressão exercida pelo mercado para operação de um processo de alta confiabilidade, motivou o presente trabalho.

No processo de lavra a céu aberto, o controle da interação entre os equipamentos nas etapas predecessoras ao beneficiamento é comumente chamada de *gestão de frota*. Essa gestão é importante para a viabilidade econômica de um empreendimento minerário porque ela representa mais da metade dos gastos da fase mina do minério de ferro.

Dentre os equipamentos alocados no processo de gestão de frota, destacam-se os caminhões para transporte de materiais. Por constituírem o processo predecessor à alimentação da usina, eles atuam como interface entre a mina e a etapa de beneficiamento, conforme mostra a Figura 1.

O controle destes equipamentos é realizado por meio de controle de “ciclos” e “eventos”. O ciclo permite avaliar a performance do processo, tal como viabilizar os algoritmos de decisão de alocação de equipamentos (Despacho), representado pelo Indicador “produtividade”. Por sua vez, os eventos permitem identificar as restrições de operação, representado pelo indicador “Rendimento”. A integração destes indicadores resulta no OEE (*Overall Equipment Effectiveness*), calculado de acordo com a Equação (1).

* O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), do Instituto Tecnológico Vale (ITV), da Universidade Federal de Ouro Preto (UFOP) e da Universidade Federal de Minas Gerais (UFMG).

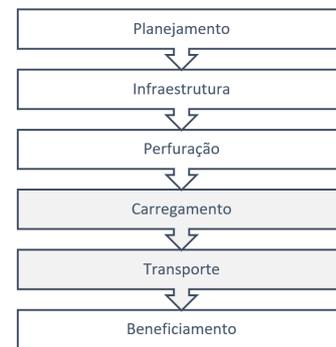


Figura 1. Interações entre processos e equipamentos predecessores à etapa de beneficiamento. A falta de equipamento de carga ocorre pela interação do processo Transporte e Carregamento destacado acima.

$$OEE = FQ \times \frac{Prod_{Real}}{Prod_{Ref}} \times RO \quad (1)$$

Na Equação (1), FQ representa o fator de qualidade (adimensional), $Prod_{Real}$ é a produtividade realizada (em toneladas por hora), $Prod_{Ref}$ é a produtividade de referência (em toneladas por hora), enquanto RO representa o rendimento operacional executado, em porcentagem.

Assim, desvios entre a programação e o realizado da alimentação de usina podem ser identificados pela análise do OEE de transporte, o qual, por sua vez, norteia os desvios entre a programação e o realizado na mina. A Figura 2 ilustra o desdobramento do indicador OEE até o evento “Falta de Equipamento de Carga”.

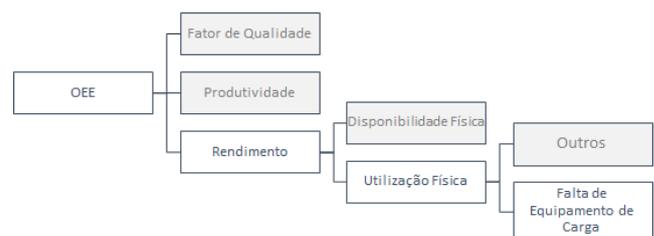


Figura 2. Desdobramento do evento Falta de Equipamento de Carga a partir do indicador *Overall Equipment Effectiveness* - OEE

Um dos problemas conhecidos neste processo é a “Falta de Equipamento de Carga” para a frota de transporte. Este evento evidencia o desequilíbrio entre a capacidade da frota de carregamento e transporte, por vezes devido a um acréscimo da capacidade de transporte ou a um decréscimo da capacidade de carregamento.

Tal problema é potencializado quando há uma redução de equipamentos no processo produtivo, o que aumenta a sensibilidade da oscilação da capacidade das frotas. A redução de frota pode ocorrer abruptamente, como por exemplo, por determinações judiciais, necessidade de manutenção na

usina, chuvas intensas ou neblinas e variações no modelo geológico ou geotécnico.

A mina em estudo apresentou uma redução abrupta em 2020 no número de equipamentos, com um aumento do evento “Falta de Equipamento de Carga”, conforme destacado na Figura 2. Apesar de sua ocorrência ser difícil de ser dimensionada e controlada, outros parâmetros podem afetar esse evento, como a manutenção de equipamentos, o aumento de trajeto de transporte e a pluviometria.

Pela revisão de literatura feita, nota-se uma carência de trabalhos científicos que tratam aplicações de métodos de inteligência computacional na indústria de mineração. Em particular não foi identificada a aplicação dessas técnicas para prever a falta de equipamentos de carga na mineração.

Portanto, este trabalho visa preencher essa lacuna e contribuir com a apresentação de um método de aprendizado de máquina, no caso, o *Random Forest* (RF), para a predição do evento “Falta de Equipamento de Carga” na indústria de mineração.

O método RF usado para a predição utiliza dados de despacho comumente encontrados em operações mineiras. Os resultados do aprendizado de máquina apontam uma correlação entre o valor predito e o real superior aos métodos previamente utilizados pela área operacional.

Os parâmetros utilizados foram obtidos através do sistema de despacho de mina e estão descritos a seguir: Tempo de manutenção corretiva dos equipamentos de carga (min), tempo de manutenção total dos equipamentos de carga (min), tempo de manutenção preventiva dos equipamentos de carga (min), tempo de manutenção dos equipamentos de transporte (min), tempo de atraso operacional da frota de transporte (min), pluviometria (mm), pluviometria do dia anterior (mm), tempo de fila no despejo de transporte (min), tempo de manobra no despejo de transporte (min), tempo durante o despejo de transporte (min), tempo aguardando o despejo de transporte (min), tempo de fila na carga do transporte (min), tempo de manobra na carga do transporte (min), tempo aguardando a carga no transporte (min), tempo em locomoção do transporte carregado (min), tempo em locomoção do transporte vazio (min) e distância média de transporte (m).

O restante deste trabalho está organizado como segue. A Seção 2 apresenta uma breve investigação quanto aos trabalhos relacionados, abrangendo aplicações do método *Random Forest* na indústria. A Seção 3 descreve o método desenvolvido para a aplicação considerada. A Seção 4 apresenta os resultados obtidos através do software Weka. Finalmente, a Seção 5 conclui o trabalho e aponta perspectivas para trabalhos futuros.

2. TRABALHOS RELACIONADOS

Jung and Choi (2021) avaliam publicações da última década que tratam de técnicas de aprendizado de máquina no setor mineral. Dos 109 trabalhos analisados, somente cinco deles aplicam o método *Random Forest* (RF) para a gestão de frota. Entretanto, este foi o método mais usado dentre os 14 artigos que abrangem gestão de frota. Do ponto de vista de eficácia, os autores ressaltam que

métodos RF apresentam a melhor performance quando avaliado o coeficiente de determinação.

Liaw and Wiener (2001) apresenta um embasamento para a metodologia de classificação e regressão através de *Random Forest*, por meio do detalhamento de sua implementação através da linguagem R, servindo de referência a trabalhos que venham a utilizar este método.

Dong et al. (2013) apresentam uma aplicação prática do método RF para predição de eventos geomecânicos no setor de mineração, mostrando a eficácia do aprendizado de máquina para melhora de segurança das operações mineiras.

Maheswari et al. (2020) aplicam métodos de aprendizado de máquina para predição de modelos geológicos na indústria de mineração. Os autores dão ênfase maior à arquitetura de rede e apresentam as dificuldades do uso de inteligência computacional quando há escassez ou dificuldade para obtenção dos dados de aprendizado. Eles reforçam a relevância do uso de sistemas de despachos como aliados na implantação de ferramentas de aprendizado de máquina no setor mineral, uma vez que estes já possuem um longo histórico de uso na indústria de mineração conforme também apontado por Kozielski et al. (2015).

Ouanan and Abdelwahed (2019) apresentam uma aplicação bem sucedida do aprendizado de máquina como complemento de processamento de imagens na etapa de beneficiamento da indústria de mineração. A aplicação permite o controle de novos parâmetros de forma contínua, como o tamanho de bolhas no processo de flotação para melhora da recuperação do beneficiamento, isto é, da relação entre o produto e a característica visual do material durante a flotação.

Kozielski et al. (2015) desenvolvem um módulo de um sistema de despacho capaz de realizar predições por meio de técnicas de aprendizado de máquina. Entretanto, para tornar o trabalho mais abrangente, os autores não mencionam a etapa de seleção de atributos para o aprendizado de máquina supervisionado, uma vez que esta etapa apresenta particularidades que tornam difícil sua generalização.

Zelinska (2020) apresenta a dificuldade na tomada de decisões na rotina operacional de controle de mina dado o cenário com um número expressivo de variáveis a controlar. Para mitigar esta dificuldade, ele usa uma série de parâmetros provenientes de despacho de mineração, com potencial de uso em métodos de aprendizado de máquina. Notou-se que o autor não realizou menção a parâmetros ambientais para controle dos equipamentos de mina tal como pluviometria, distância média de transporte ou perfil de trajeto que também possuem potencial de aplicabilidade em aprendizado de máquina na indústria de mineração.

Comparando-se com os trabalhos relacionados, este busca a aplicação em gestão de frota similar à de Zelinska (2020). Porém, são acrescentados parâmetros ambientais referentes ao perfil de lavra, expresso por “Distância Média de Transporte (DMT)”, e pluviometria, expresso pelo parâmetro “Milímetros de chuva”, devido à interferência destes atributos na performance de frotas que operam em lavra a céu-aberto. No viés de aplicabilidade, o presente trabalho

assemelha-se ao de Kozielski et al. (2015) por também utilizar a base de dados de despacho; porém, difere-se dele ao realizar a etapa de seleção de atributos para otimização do modelo e redução de esforço computacional.

3. MÉTODOS

Os métodos para aprendizado de máquina podem ser classificados como: Aprendizado Supervisionado, Aprendizado não supervisionado e aprendizado semi-supervisionado.

No aprendizado supervisionado, a metodologia é implementada através da seleção dos parâmetros de entrada com sua respectiva saída. Seu objetivo é induzir uma hipótese (modelo) que seja capaz de relacionar os parâmetros de entrada com a saída. Algoritmos deste tipo utilizam o histórico responsável pelo treinamento do aprendizado para interpretar quaisquer novos dados. Muitos modelos de aprendizado são baseados em uma transição genérica dos dados utilizados no treinamento para a nova amostra através de similaridades entre os dados (Zelinska, 2020).

No aprendizado não-supervisionado assume-se que os rótulos, nomes ou descrições dos parâmetros não são relevantes para o aprendizado, ficando a cargo somente do aprendizado a seleção dos parâmetros que melhor relacionam com a saída objetivada. Tais algoritmos são utilizados quando é desejado encontrar padrões escondidos na base de dados (Zelinska, 2020).

Por sua vez, no aprendizado semi-supervisionado, os dados de entrada são utilizados no aprendizado com algumas restrições definidas, mantendo-se certo grau de controle quanto ao uso dos parâmetros durante o aprendizado (Zelinska, 2020).

A seleção do método a ser aplicado é particular a cada aplicação. Dada a disponibilidade de especialistas da operação de mina, o método adotado foi o *Random Forest* supervisionado, ou seja, com a determinação dos parâmetros prioritariamente ao aprendizado de máquinas.

O trabalho foi iniciado com a estruturação da base a ser utilizada para o aprendizado do método supervisionado. No primeiro momento, realizou-se o mapeamento das fontes de informações disponíveis para estudo, tal como sua frequência de medição disponível por minuto, hora, dia, turno, semana e mês. Esta etapa é tida como crítica, pois conforme exposto por Kozielski et al. (2015), a complexidade da arquitetura de rede tende a ser elevada na indústria de mineração pelo uso de equipamentos de fornecedores e sistemas distintos. Realizado portanto, em caráter de estudo, consultas via SQL dos servidores e criada uma chave primária temporal para integração das bases.

Os dados foram extraídos do sistema SAP®/BO, juntamente com as medições manuais de pluviômetros dentro da área de influência meteorológica da mina no período compreendido entre 01/05/2020 a 01/10/2020, resultando em 860 mil registros. Foram compilados 20 parâmetros agrupados nos universos expostos na lista a seguir:

- 4 parâmetros de Despejo de Transporte
- 4 parâmetros de Carregamento de Transporte
- 4 parâmetros de Pluviometria/Condição Climática

- 3 parâmetros de Tempo de Manutenção da Frota de Carregamento
- 2 parâmetros de tempo de deslocamento
- 1 parâmetro de Atrasos Operacionais
- 1 parâmetro de Distância Percorrida pelo transporte
- 1 parâmetro de Tempo de Manutenção da Frota de Transporte

Quanto à frequência dos dados, esta foi diária, pois apesar de grande parte dos atributos possuírem a granularidade em segundos, um parâmetro considerado importante na hipótese inicial é a sazonalidade climática descrita pelo índice pluviométrico, apurado diariamente.

Em seguida, realiza-se uma etapa de seleção de atributos, uma vez que a otimização do uso dos atributos, além de gerar ganhos em tempo de processamento, aumenta a viabilidade da aplicação do método na rotina por determinar uma quantidade ideal de parâmetros a controlar pela equipe operacional.

As métricas de avaliação são utilizadas para comparar os resultados obtidos pelos modelos de aprendizado de máquina com diversas técnicas estatísticas. De acordo com Kozielski et al. (2015), na indústria de mineração, modelos associados com classificação (ex: ocorrência de falha de desmonte com explosivo) são comumente analisados pela *área sob a curva* ou *Matriz de confusão*, enquanto modelos associados com regressão (ex: estimativa do custo de operação) são geralmente avaliados através do erro quadrático médio, erro absoluto médio, ou coeficiente de determinação.

Há diversos métodos de seleção de atributos apresentados na literatura, como *Percent Missing Values*, *Amount of Variance*, *Correlation (with the target)*, *Pairwise Correlation*, *Multicollinearity*, *Principal Component Analysis*, *Cluster Analysis*, *Forward Selection*, *Backward Elimination*, *Stepwise Selection*, *LASSO* e *Tree-based methods* (Patel, 2016).

Dentre estes, o *Backward Elimination* foi o escolhido. A justificativa é que a consulta à base de dados de despacho retorna todos os parâmetros utilizados, tornando mais simples realizar uma única exportação dos dados em sua totalidade com remoção dos parâmetros na modelagem de dados. No método *Backward Elimination*, o modelo é iniciado com todas as variáveis e, a partir de critérios lógicos, realiza-se a exclusão gradativa deles, mantendo-se como parâmetro de acompanhamento o coeficiente de determinação. Em cada etapa, realiza-se no mínimo 10 *seeds*.

Para redução de parâmetros, primeiramente foi realizada uma análise de correlação entre os parâmetros, uma vez que a permanência de parâmetros correlacionados podem potencializar a relevância deste no modelo. Posteriormente, foi aplicada a redução gradativa dos parâmetros conforme a ordem cronológica dos eventos em conformidade com o despacho de mina. O ciclo de transporte se inicia no tempo de deslocamento com a balança vazia, seguido de: tempo de fila para carregar, tempo de manobra para carregar, tempo de aguardo a ser carregado, tempo de carregamento, tempo de deslocamento com a balança cheia, tempo de fila para despejo, tempo de manobra para

despejo, tempo de aguardo do despejo e finalizando com o tempo do despejo.

Ao final do processo, selecionam-se os atributos cuja combinação apresenta melhor desempenho no coeficiente de correlação. Foi comparado o coeficiente de correlação obtido pelo *Backward Elimination* com o coeficiente de correlação da regressão linear comumente utilizada pela área operacional. Esta regressão é obtida através da análise entre as horas em manutenção corretiva da máquina de carga com as horas do evento *Falta de equipamento de Carga* pelo caminhão.

Após a etapa de pré-processamento do *Random Forest* (RF), é necessária a avaliação de qual modelo apresenta a melhor resposta frente ao problema. O RF é considerado um método de *essemble learning*, em que as árvores de decisão são consideradas a base de sua classificação (Kuncheva, 2004). A Figura 3 ilustra o modelo de predição via RF, na qual a classe final de predição é obtida através da “votação” da predição de cada “árvore” individual (daí o nome *forest*) (Primartha and Adhi Tama, 2017). De acordo com Kuncheva (2004), os dois principais parâmetros possíveis de serem ajustados são: o número de variáveis escolhidas em cada nó da árvore de decisão, o que é normalmente fixado em todos os nós, e o número de árvores que irão compor a floresta. Assim, foram avaliados o resultado do modelo com 20 árvores (RF20), 50 árvores (RF50) e 100 árvores (RF100).

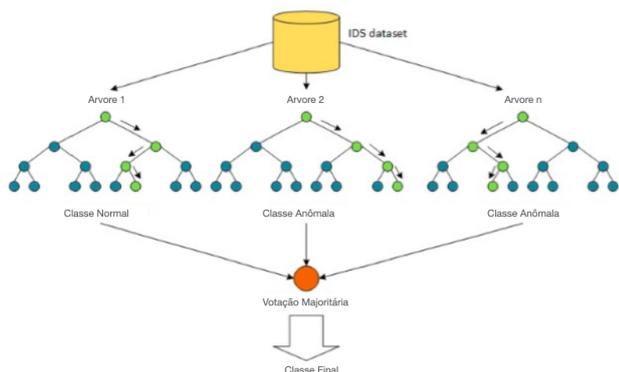


Figura 3. Ilustração de Classificador *Random Forest*, adaptado de Primartha and Adhi Tama (2017)

4. RESULTADOS

4.1 Pré-Processamento e Análise

O estudo foi iniciado com 20 parâmetros e, a partir do agrupamento deles, realizamos o processo de *backward elimination*, analisando o indicador coeficiente de correlação em cada etapa, conforme mostra a Figura 4.

Trabalhos desenvolvidos pela área operacional atribuem como principal parâmetro de correlação com o evento “falta de equipamento de carga”, as “horas em manutenção corretiva de equipamentos de carga”, que para o mesmo período apresentou coeficiente de correlação de 0,35 em sua regressão linear. Desta maneira foi possível comparar a evolução do processo de *backward elimination* frente à correlação obtida pelos métodos convencionais aplicados pela equipe (Regressão Linear).

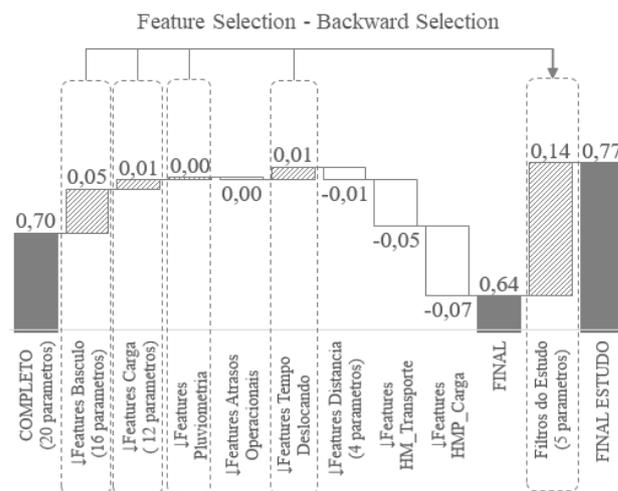


Figura 4. Processo de *Backward Selection* para definição dos parâmetros a manter e a descartar

Ao início do processo, o modelo apresentava um coeficiente de correlação de 0,7 com 20 parâmetros. Após a exclusão dos parâmetros “tempo de despejo”, “tempo de carregamento”, “tempo de deslocamento” e “pluviometria”, o coeficiente de correlação passou para 0,77 com 5 parâmetros, sendo estes:

- Soma de Horas de Manutenção de equipamentos de carga por dia (HM Carga)
- Soma de Horas de Manutenção corretiva de equipamentos de carga por dia (HMC Carga)
- Soma de Horas de manutenção preventiva de equipamentos de carga por dia (HMP Carga)
- Soma de Horas de Manutenção de equipamentos de transporte por dia (HM Transporte)
- Soma de Horas de Atrasos Operacionais de equipamentos de transporte por dia (HAO Transporte)

O fluxo do sistema, com os parâmetros selecionados, é apresentado na Figura 5.

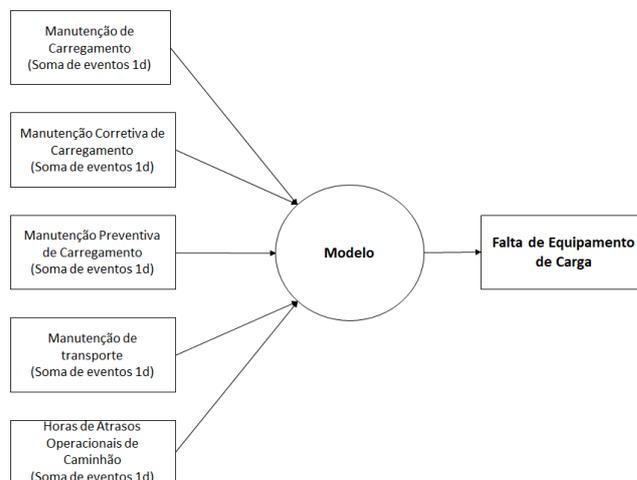


Figura 5. Utilização dos 5 parâmetros de entrada para predição da Falta de Equipamento de Carga.

Comparativo valor Predito x valor Real por Dia

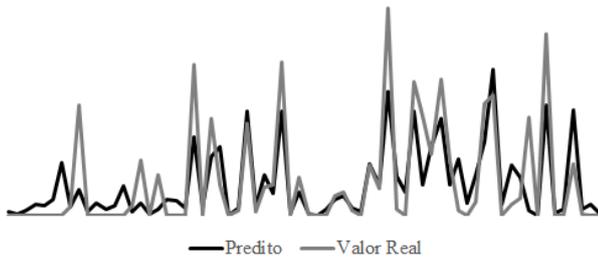


Figura 6. Avaliação temporal entre o valor predito e executado no período para o evento Falta de Equipamento de Carga

4.2 Entrada de parâmetros e geração do modelo

O pacote de softwares WEKA foi utilizado para geração do modelo. As testagens foram realizadas com *cross-validation* de 10 *folds* e 30 *seeds* distintos, alterando o número de árvores em 20, 50 e 100.

O valor predito apresentou coeficiente de correlação de 0,773 com o real (Figura 6). É oportuno expor que tal quantificação do evento possui elevado viés humano, visto que a decisão de quando e como registrar o evento compete à equipe presente na sala de controle no momento do evento.

Quanto à importância de atributos (Figura 7), a “Manutenção total dos equipamentos de Carga” (HM Carga) apresentou uma importância superior à hipótese inicial de “eventos de manutenção corretiva” (HMC Carga), o que indica que o atributo acompanhado anteriormente não foi o de maior importância para o modelo e, sim, a manutenção geral dos equipamentos, resultado da soma das corretivas com as preventivas.

Identificou-se, portanto, que não basta somente o equipamento ter uma manutenção corretiva para a ocorrência do evento “Falta de Equipamento de Carga”, e sim haver uma manutenção corretiva durante a manutenção preventiva de outro equipamento de carga, ocasionando a parada dos equipamentos de transporte em “Falta de Equipamento de Carga”. Os parâmetros “Atraso Operacional” (HAO) e “Distância a Percorrer” (DMT CH) também foram considerados importantes, sendo estes passíveis de serem controlados via despacho por integrarem os indicadores de performance da frota de transporte. Estratégias de aumento de variações de distância, tal como controle de atrasos operacionais influenciam no evento de falta de equipamento de carga através do modelo. O parâmetro de “manutenção de transporte” (HM Transporte) também é exposto como parâmetro de importância ao modelo, sendo portanto necessário controlar não somente a manutenção dos equipamentos de carga que causam a falta de equipamento de carga, como também a “escassez” de manutenção de transporte que geram o aumento da capacidade de transporte sem a capacidade de absorção pela frota de carregamento.

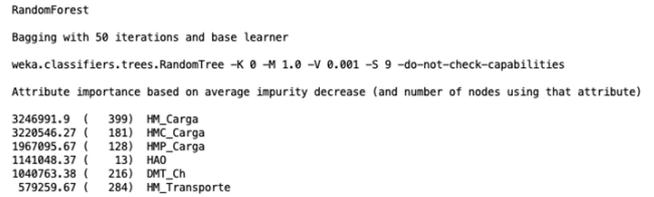


Figura 7. Atributos de importância relacionados ao número de nós que os utilizam

4.3 Comparativo entre os modelos Random Forest

Para verificação de diferença estatística entre os métodos *Random Forest* implementados, foi realizado o teste de normalidade Shapiro-Wilk. Os resultados deste teste estão reportados na Tabela 1.

Tabela 1. Resultados do teste de Shapiro-Wilk

Parâmetro	RF20	RF50	RF100
w	0,981	0,968	0,986
p-value	0,862	0,494	0,950

Como os valores de *p-value* foram maiores do que 0,05, pode-se admitir que os três métodos seguem distribuição normal. Desta maneira, é viabilizada a realização do *Welch Two Sample t-test*, cujos resultados estão apresentados na Tabela 2.

Tabela 2. Resultados do teste de Welch Two Sample t-test

Parâmetro	RF20 × RF100	RF50 × RF100	RF20 × RF50
t	-1,194	0,100	-1,209
df	40,245	55,734	45,127
p-value	0,240	0,921	0,233

Como em todos os testes, o *p-value* foi maior que 0,05, pode-se afirmar que estatisticamente as distribuições dos três métodos são equivalentes.

Assim, realizou-se o estudo do comportamento do coeficiente de correlação de cada modelo, fixando em 30 *seeds*. A Tabela 3 reporta os resultados obtidos.

A Figura 8 mostra os gráficos *box plot* de cada um dos métodos *Random Forest* com relação ao coeficiente de correlação.

Por sua vez, a Figura 9 mostra os histogramas dos métodos RF20, RF50 e RF100, em 30 *seeds*.

Pela Tabela 3 e Figuras 8 e 9, é possível notar que, apesar de todos os métodos serem estatisticamente equivalentes, o método RF100 apresenta média e mediana da distribuição dos coeficientes de correlação mais elevadas e com menor variabilidade. Por outro lado, o método com o menor coeficiente de correlação médio e maior variabilidade é o RF20, reforçando seu descarte como método de predição. Já o método RF50 também tem o mesmo coeficiente de correlação médio do método RF100, não apresentando grande variação quando comparado com o RF100. Por envolver menos árvores, sua complexidade é menor, o que o torna mais indicado como método de predição do evento “Falta de Equipamento de Carga”.

Em comparativo com o método de regressão linear normalmente utilizado pelas áreas operacionais (Figura 10),

Tabela 3. Valores do coeficiente de correlação para cada tipo de *Random Forest* por *Seed*

Seed	RF20	RF50	RF100
1	0,812	0,786	0,781
2	0,733	0,754	0,757
3	0,730	0,759	0,761
4	0,775	0,767	0,772
5	0,783	0,766	0,772
6	0,786	0,779	0,776
7	0,802	0,782	0,786
8	0,789	0,781	0,776
9	0,783	0,800	0,791
10	0,766	0,767	0,774
11	0,757	0,757	0,752
12	0,785	0,771	0,766
13	0,775	0,780	0,767
14	0,763	0,761	0,768
15	0,783	0,789	0,774
16	0,770	0,761	0,764
17	0,758	0,771	0,783
18	0,750	0,770	0,770
19	0,779	0,789	0,783
20	0,728	0,769	0,783
21	0,756	0,766	0,789
22	0,747	0,772	0,767
23	0,775	0,768	0,768
24	0,751	0,780	0,785
25	0,770	0,782	0,780
26	0,774	0,776	0,777
27	0,770	0,776	0,775
28	0,797	0,795	0,787
29	0,741	0,761	0,764
30	0,758	0,767	0,768
Média	0,768	0,773	0,773
Desvio	0,021	0,011	0,009

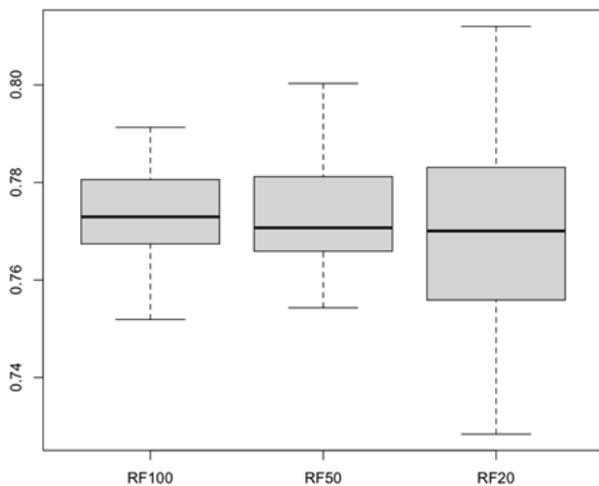


Figura 8. BoxPlot dos coeficientes de correlação para cada tipo de *Random Forest*

o aprendizado de máquina (RF20, RF50 e RF100) apresentou índice de correlação com valores superiores a 2x a “Regressão Linear” entre *Falta de Equipamento de Carga* (h) e *Manutenção Corretiva de Equipamento de Carga* (h).

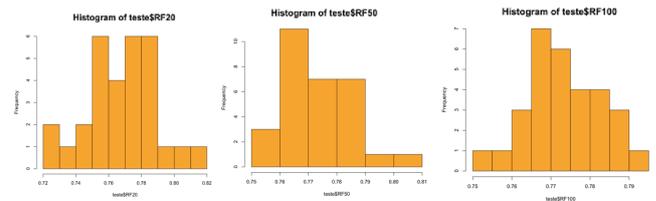


Figura 9. Histograma dos coeficientes de correlação para cada tipo de *Random Forest*

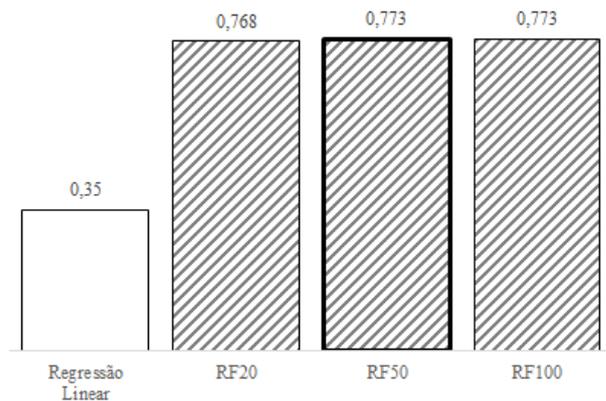


Figura 10. Comparativo entre os métodos com relação ao índice de correlação

5. CONCLUSÃO

Este trabalho tratou o problema de predição do evento de “falta de equipamento de carga” através do uso de dados de despacho para obter uma correlação superior ao método de Regressão Linear comumente aplicado na área operacional da empresa mineradora em estudo. Para resolvê-lo, realizou-se a consolidação do *data-set* do sistema de despacho, selecionando atributos pela metodologia *Backward Selection*. Nessa fase de seleção, reduziu-se de 20 para cinco parâmetros. Em seguida, foram testados três modelos de *Random Forest*, com 20 árvores (RF20), 50 árvores (RF50) e 100 árvores (RF100).

Com base em sua boa média de coeficiente de correlação frente aos demais, com desvio padrão moderado e equivalência quanto aos demais conforme o *t-test*, o modelo escolhido foi o RF50 para a predição do evento de “Falta de Equipamento de Carga” através dos parâmetros selecionados.

Por meio da análise da importância de atributos, retirou-se 15 dos 20 parâmetros iniciais do modelo, elevando o coeficiente de correlação de 0,7 para 0,77 juntamente com a redução do recurso computacional. Os 5 parâmetros remanescentes sugerem à área operacional a ampliação do escopo de controle de eventos para a “Falta de Equipamento de Carga”, predecessor ao trabalho controlado somente por “Horas de Manutenção Corretiva de Equipamentos de Carga (HMC Carga)”. Posteriormente ao trabalho, acrescido de “Horas de Manutenção de equipamentos de carga por dia (HM Carga)”, “Horas de Manutenção Preventiva de equipamentos de Carga (HMP Carga)”, “Horas de Manutenção de equipamentos de Transporte (HM Transporte)

e "Horas de Atrasos Operacionais de equipamentos de Transporte (HAO Transporte)".

A quantidade em horas por dia do evento de falta de equipamento de carga foi predita pelo *Random Forest*. A correlação obtida foi de 0,77, bem superior à do método de regressão linear utilizado pelo tomador de decisão da empresa, que retornava uma correlação de 0,35 considerada entre as "horas de manutenção de equipamentos de carga" e as "horas de falta de equipamento de carga".

Com este estudo, mostra-se que o uso do aprendizado de máquina para a predição na indústria de mineração se revela uma importante ferramenta de apoio à predição e classificação, visto que esses métodos detectam a influência de parâmetros no processo nem sempre percebidos em métodos estatísticos tradicionais.

Para trabalhos futuros, indica-se a redução da granularidade temporal, uma vez que o limitador (pluviometria), que poderia ser obtido somente em períodos diários devido a restrições da estação meteorológica, mostrou-se pouco relevante para a definição do modelo, viabilizando análises de turno ou horárias. Há, também, potencial para estratificar a frota de carregamento por equipamento de carga para avaliação de itens como confiabilidade e vida útil por equipamento a fim de aprimorar o modelo. Espera-se, além disso, explorar a distribuição dos valores para assim gerar a probabilidade de previsão do modelo.

REFERÊNCIAS

- Dong, L., bing LI, X., and Peng, K. (2013). Prediction of rockburst classification using random forest. *Transactions of Nonferrous Metals Society of China*, 23, 472–477. doi:10.1016/S1003-6326(13)62487-5.
- Jung, D. and Choi, Y. (2021). Systematic review of machine learning applications in mining: Exploration, exploitation, and reclamation. *Minerals*, 11(2). doi: 10.3390/min11020148.
- Kozielski, M., Sikora, M., and Wróbel, L. (2015). Disesor - decision support system for mining industry. In *2015 Federated Conference on Computer Science and Information Systems (FedCSIS)*, 67–74. doi:10.15439/2015F168.
- Kuncheva, L.I. (2004). *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, New York.
- Liaw, A. and Wiener, M. (2001). Classification and regression by randomforest. *Forest*, 23, 18–22.
- Maheswari, C., Priyanka, E.B., Thangavel, S., Vignesh, S.V.R., and Poongodi, C. (2020). Multiple regression analysis for the prediction of extraction efficiency in mining industry with industrial iot. *Production Engineering*, 14(4), 457–471. doi:10.1007/s11740-020-00970-z. URL <https://doi.org/10.1007/s11740-020-00970-z>.
- MME (2021). Desempenho do setor mineral em 2020 supera expectativas. Ministério de Minas e Energia. Disponível em encurtador.com.br/yABHY, acesso em 17/05/2021.
- Ouanan, H. and Abdelwahed, E.H. (2019). Image processing and machine learning applications in mining industry: Mine 4.0. In *2019 International Conference on Intelligent Systems and Advanced Computing Sciences (ISACS)*, 1–5. Taza, Morocco. doi:10.1109/ISACS48493.2019.9068884.
- Patel, V. (2016). A practical guide to dimensionality reduction techniques. PyData Conference 2016. Disponível em <https://pydata.org/dc2016/schedule/presentation/29/>, acesso em 12/10/2020.
- Primartha, R. and Adhi Tama, B. (2017). Anomaly detection using random forest: A performance revisited. In *International Conference on Data and Software Engineering (ICoDSE)*, 1–6. IEEE, Palembang, Indonesia.
- Zelinska, S. (2020). Machine learning: technologies and potential application at mining companies. In *The International Conference on Sustainable Futures: Environmental, Technological, Social and Economic Matters (ICSF 2020)*, volume 166, 03007. EDP Sciences, Ukraine.