



INSTITUTO TECNOLÓGICO VALE



**Programa de Pós-Graduação em Instrumentação, Controle e
Automação de Processos de Mineração (PROFICAM)
Escola de Minas, Universidade Federal de Ouro Preto (UFOP)
Associação Instituto Tecnológico Vale (ITV)**

Dissertação

**PREVISÃO DE FALTA DE EQUIPAMENTO DE CARGA PARA FROTA DE
TRANSPORTE COM BASE EM APRENDIZADO DE MÁQUINA E CENÁRIOS DE
DESPACHO**

Guilherme Gouveia Cordeiro

**Ouro Preto
Minas Gerais, Brasil
2022**

Guilherme Gouveia Cordeiro

**PREVISÃO DE FALTA DE EQUIPAMENTO DE CARGA PARA FROTA DE
TRANSPORTE COM BASE EM APRENDIZADO DE MÁQUINA E CENÁRIOS DE
DESPACHO**

Dissertação apresentada ao Programa de Pós-Graduação em Instrumentação, Controle e Automação de Processos de Mineração da Universidade Federal de Ouro Preto e do Instituto Tecnológico Vale, como parte dos requisitos para obtenção do título de Mestre em Engenharia de Controle e Automação.

Orientador: Prof. Frederico Gadelha Guimarães, Ph.D.

Coorientador: Prof. Marcone Jamilson Freitas Souza, Ph.D.

Ouro Preto
2022

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

C794p Cordeiro, Guilherme Gouveia.

Previsão de falta de equipamento de carga para frota de transporte com base em aprendizado de máquina e cenários de despacho. [manuscrito] / Guilherme Gouveia Cordeiro. - 2022.

43 f.: il.: color., gráf., tab..

Orientador: Prof. Dr. Frederico Guimarães.

Coorientador: Prof. Dr. Marccone Souza.

Dissertação (Mestrado Profissional). Universidade Federal de Ouro Preto. Programa de Mestrado Profissional em Instrumentação, Controle e Automação de Processos de Mineração. Programa de Pós-Graduação em Instrumentação, Controle e Automação de Processos de Mineração.

Área de Concentração: Engenharia de Controle e Automação de Processos Minerais.

1. Minas e Mineração. 2. Indústria mineral. 3. Processos de fabricação - Automação. 4. Teoria da previsão. 5. Aprendizado do computador. I. Guimarães, Frederico. II. Souza, Marccone. III. Universidade Federal de Ouro Preto. IV. Título.

CDU 681.5:622.2

Bibliotecário(a) Responsável: Maristela Sanches Lima Mesquita - CRB-1716



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE OURO PRETO
REITORIA
ESCOLA DE MINAS
PROGR. POS GRAD. PROF. INST. CONT. E AUT.
PROCESSOS DE MIN.



FOLHA DE APROVAÇÃO

Guilherme Gouveia Cordeiro

Previsão de falta de equipamento de carga para frota de transporte com base em aprendizado de máquina

Dissertação apresentada ao Programa de Pós-Graduação em Instrumentação, Controle e Automação de Processos de Mineração (PROFICAM), Convênio Universidade Federal de Ouro Preto/Associação Instituto Tecnológico Vale - UFOP/ITV, como requisito parcial para obtenção do título de Mestre em Engenharia de Controle e Automação na área de concentração em Instrumentação, Controle e Automação de Processos de Mineração.

Aprovada em 06 de junho de 2022

Membros da banca

Doutor - Frederico Gadelha Guimarães - Orientador - Universidade Federal de Ouro Preto
Doutor - Marcone Jamilson Freitas Souza - Universidade Federal de Ouro Preto
Doutor - Agnaldo José da Rocha Reis - Universidade Federal de Ouro Preto
Doutor - Michel Bessani - Universidade Federal de Minas Gerais

Frederico Gadelha Guimarães, orientador do trabalho, aprovou a versão final e autorizou seu depósito no Repositório Institucional da UFOP em 05/09/2022



Documento assinado eletronicamente por **Bruno Nazário Coelho, COORDENADOR(A) DE CURSO DE PÓS-GRADUAÇÃO EM INST. CONTROLE AUTOMAÇÃO DE PROCESSOS DE MINERAÇÃO**, em 08/09/2022, às 12:29, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0394923** e o código CRC **46D870EB**.

Agradecimentos

A Deus que me guia.

Ao apoio de minha família que me fortalece mesmo em períodos de distância.

Aos orientadores Prof. Marcone Jamilson Freitas Souza e Prof. Frederico Gadelha Guimarães, exemplos de pessoa e educador.

A Vale, pela oportunidade de desenvolvimento profissional e pessoal.

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); e da Vale SA.

*“O resultado é o que muitas
pessoas chamam de frutos da
genialidade, enquanto, na verdade,
são os frutos do estudo e do
trabalho”
(Alexander Hamilton).*

Resumo

Resumo da Dissertação apresentada ao Programa de Pós Graduação em Instrumentação, Controle e Automação de Processos de Mineração como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

PREVISÃO DE FALTA DE EQUIPAMENTO DE CARGA PARA FROTA DE TRANSPORTE COM BASE EM APRENDIZADO DE MÁQUINA E CENÁRIOS DE DESPACHO

Guilherme Gouveia Cordeiro

Junho/2022

Orientadores: Frederico Gadelha Guimarães

Marcone Jamilson Freitas Souza

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 internos (como quebra de equipamentos), externos ao processo (como intempéries) e de interface (como parada de britagem). Este trabalho trata do desenvolvimento de uma abordagem de aprendizado de máquina para previsão da quantidade de horas por dia do evento Falta de Equipamento de Carga para a frota de transporte utilizando dados de despacho. O modelo aplicado foi o *Random Forest* variando 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 coeficiente de correlação (R). Dados de operação real foram utilizados para construção do modelo contendo ao final do trabalho 11,7 milhões de registros em 2 anos de coleta. Os resultados apontaram através do teste de *Shapiro-Wilk* e *T-Welch* que os modelos de 20, 50 e 100 árvores eram equivalentes, os coeficientes de correlação médio foram respectivamente 0,768, 0,773 e 0,773 que após a ampliação da massa de dados de 860 mil para 11,7 milhões de registros chegaram a 0,809, resultados melhores do que o valor de 0,35 obtido com o método regressão linear, atualmente utilizado pela área. 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.

Palavras-chave: Mineração, Indústria, Processo, Predição, Aprendizado de Máquina.

Macrotema: Mina; **Linha de Pesquisa:** Tecnologias da Informação, Comunicação e Automação Industrial; **Tema:** Aumento de Eficiência na Cadeia, do Desmonte à Britagem; **Área Relacionada da Vale:** Operação de Mina.

Abstract

Abstract of Dissertation presented to the Graduate Program on Instrumentation, Control and Automation of Mining Process as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

LACK OF LOAD EQUIPMENT FOR THE TRANSPORT FLEET PREDICTION USING MACHINE LEARNING AND DISPATCH DATA

Guilherme Gouveia Cordeiro

June/2022

Advisors: Frederico Gadelha Guimarães

Marcone Jamilson Freitas Souza

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 amount of hours per day in 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 11,700,000 records gathered for 2 years. The results of the Shapiro-Wilk and T-Welch tests showed that the models of 20, 50, and 100 trees were equivalent, the average correlation coefficients were respectively 0.768, 0.773, and 0.773 that after expand the data mass from 860 thousands to 11.7 million records reached 0.809; better than the value of 0.35 obtained with the linear regression method currently used by the area. Thus, the use of machine learning through Random Forest showed good results and applicability for predictions and classifications of events in fleet management in the mining industry.

Keywords: Mining, Industry, Process, Prediction, Machine Learning.

Macrotheme: Mine; **Research Line:** Information Technology, Communication and Industrial

Automation Technologies; **Theme:** Increased Efficiency in the Mining Process, from Blasting to Milling; **Related Area of Vale:** Mining Operations.

Lista de Figuras

Figura 1.1	Interações entre processos e equipamentos predecessores à etapa de beneficiamento. A falta de equipamento de carga ocorre pela interação do processo de Transporte (6) e Carregamento (7) apresentado acima Fonte: VALE (2019)	14
Figura 1.2	Distribuição de Ciclos e Eventos Fonte: Autor.	15
Figura 1.3	Desdobramento do evento Falta de Equipamento de Carga a partir do indicador <i>Overall Equipment Effectiveness</i> - OEE Fonte: Autor	15
Figura 3.1	Ilustração de funcionamento do Classificador <i>Random Forest</i> . Fonte: Traduzido de Primartha e Adhi Tama (2017)	23
Figura 4.1	Processo de <i>BackWard Selection</i> para definição dos parâmetros a manter e a descartar. Nas barras de rachura tracejada temos o incremento do coeficiente de correlação pela redução dos parâmetros assinalados no eixo, Nas barras sem preenchimento temos o decréscimo do coeficiente de correlação pela redução dos parâmetros assinalados no eixo, após o final do estudo foram retirados portanto os 15 parâmetros das barras rachuradas resultando no modelo final de 5 parâmetros Fonte: Autor	25
Figura 4.2	Utilização dos 5 parâmetros de entrada para predição da Falta de Equipamento de Carga. Fonte: Autor	27
Figura 4.3	Avaliação temporal entre o valor predito e executado no período para o evento Falta de Equipamento de Carga Fonte: Autor	27
Figura 4.4	Atributos de importância relacionados ao número de nós que os utilizam. Fonte: Autor	28
Figura 4.5	BoxPlot dos coeficientes de correlação para cada tipo de Random Forest. Fonte: Autor.	31
Figura 4.6	Histograma dos coeficientes de correlação para cada tipo de <i>Random Forest</i> . Fonte: Autor.	31
Figura 4.7	Comparativo entre os métodos com relação ao índice de correlação. Fonte: Autor.	32
Figura 4.8	Comparativo entre o valor predito e realizado aumentando a frequência de coleta dos dados, antes 0,773 no horizonte 24hrs agora 0,802 no horizonte 6hrs. Fonte: Autor.	33

Figura 4.9	Densidade de Kernel em 2D do Valor Predito x Real do Modelo gerado pelo	
	aprendizado no período de Mar/20-Out/20 pelo método RF50. Fonte: Autor.	35
Figura 4.10	Densidade de Kernel em 2D do Valor Predito x Real do Modelo gerado pelo	
	aprendizado no período de Nov/20-Mar/22 pelo método RF50. Fonte: Autor.	36
Figura 4.11	Representação do Teste Cruzado, testando o modelo no <i>DataSet</i> comple-	
	mentar ao utilizado para seu aprendizado. Fonte: Autor	36

Lista de Tabelas

Tabela 4.1	Resultados do teste de Shapiro-Wilk	29
Tabela 4.2	Resultados do teste de Welch Two Sample t-test	29
Tabela 4.3	Valores do coeficiente de correlação para cada tipo de <i>Random Forest</i> por Seed	30
Tabela 4.4	Valores do coeficiente de correlação para cada tipo de frequência de coleta por Seed	34
Tabela 4.5	Valores do coeficiente de correlação para cada intervalo de coleta alterando o período para teste	37
Tabela 4.6	Valores do coeficiente de correlação para cada intervalo de coleta por Seed, testados por 10-Fold Cross Validation.	37

Sumário

1	Introdução	14
1.1	Objetivos	16
1.2	Contribuições	17
1.3	Organização do Trabalho	17
2	Trabalhos Relacionados	18
3	Materiais e Métodos	20
4	Resultados e Discussões	24
4.1	Preparação dos dados	24
4.2	Pré-Processamento e Análise	24
4.3	Entrada de parâmetros e geração do modelo	26
4.4	Comparativo entre os modelos <i>Random Forest</i>	29
4.5	Aumento da frequência do <i>Dataset</i> de aprendizado	33
4.6	Aumento do intervalo de amostragem do <i>Dataset</i> de aprendizado	35
5	Conclusões	38
5.1	Trabalhos Futuros	39
	Referências Bibliográficas	40

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 2,5 p.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 processos. 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 e 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 e Choi (2021), há uma escassez de trabalhos com este escopo na indústria de mineração e, em especial, na gestão de frota.

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*, sendo a viabilidade econômica de um empreendimento minerário sensível à boa performance tanto desta etapa quanto do beneficiamento.

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 a interface entre a mina e a etapa de beneficiamento, conforme mostra a Figura 1.1.

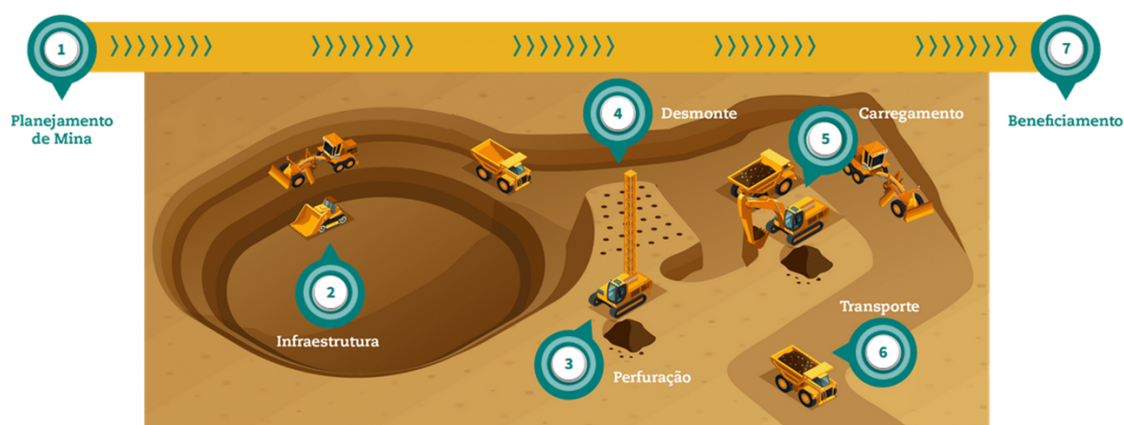


Figura 1.1: Interações entre processos e equipamentos predecessores à etapa de beneficiamento. A falta de equipamento de carga ocorre pela interação do processo de Transporte (6) e Carregamento (7) apresentado acima

Fonte: VALE (2019)

O controle destes equipamentos é realizado a nível de “ciclos” e “eventos” (Figura 1.2). O ciclo é o tempo de horas trabalhadas em que o caminhão realiza uma viagem, a partir da análise deste é possível identificar a performance do processo, tal como viabilizar os algoritmos

Horas Calendário (24/equipamento)		
Horas Disponíveis		Eventos de Horas de Manutenção
Ciclos de Transporte ou Eventos de Horas Trabalhadas*	Eventos de Horas Ociosas	

Figura 1.2: Distribuição de Ciclos e Eventos

Fonte: Autor.

de decisão de alocação de equipamentos (Despacho); O Ciclo é analisado pelo Indicador “produtividade” (*quantas toneladas de material movimentado realizo por cada hora trabalhada*). Por sua vez, os eventos permitem identificar as restrições de operação, seja devido a alguma hora ociosa (como o caminhão parado por Falta de Equipamento de Carga) ou a alguma hora de manutenção (como o caminhão parado por manutenção corretiva) representado pelo Indicador “Rendimento” (*quantas horas trabalhadas eu realizo durante minhas horas de calendário*). A integração destes indicadores resulta no OEE (*Overall Equipment Effectiveness*), calculado de acordo com a Equação (1.1).

$$OEE = FQ \times \frac{Prod_{Real}}{Prod_{Ref}} \times RO \quad (1.1)$$

Na Equação (1.1), FQ representa o fator de qualidade (adimensional, definido *a priori*, e utilizado o valor “1” normalmente), $Prod_{Real}$ é a produtividade realizada (em toneladas por hora trabalhada), $Prod_{Ref}$ é a produtividade de referência (em toneladas por hora trabalhada), enquanto RO representa o rendimento operacional executado, em percentagem (dividindo horas trabalhadas por horas calendário).

Assim, desvios de alimentação de usina podem ser identificados através da análise do OEE de transporte, o qual, por sua vez, norteia os desvios na mina através de seu desdobramento. A Figura 1.3 ilustra o desdobramento do indicador OEE até o nível de parcela de evento “Falta de Equipamento de Carga”.

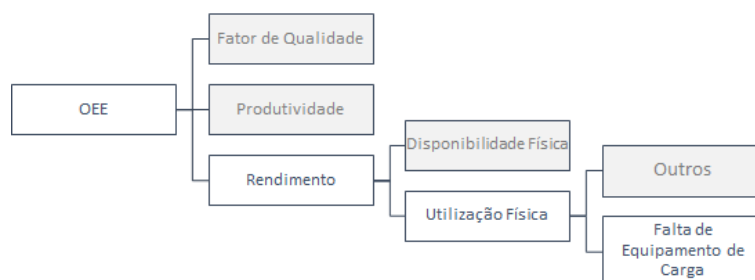


Figura 1.3: Desdobramento do evento Falta de Equipamento de Carga a partir do indicador *Overall Equipment Effectiveness* - OEE

Fonte: Autor

Um dos problemas conhecidos neste processo é a “Falta de Equipamento de Carga” para a frota de transporte. Este indicador evidencia o desequilíbrio entre a capacidade de carrega-

mento existente e a capacidade de transporte da frota, 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 entre a mina e seu 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 1.3. 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, percebemos que há uma carência de trabalhos científicos que tratem aplicações de métodos de inteligência computacional na indústria de mineração. Em particular, de nosso conhecimento, não encontramos nenhuma 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 um coeficiente de correlação de 0,77 entre o valor predito e o real na amostra relativa ao período compreendido entre 01/05/2020 e 01/10/2020, valor muito superior à correlação linear de 0,35 obtida por métodos de regressão linear.

1.1. Objetivos

O objetivo geral consiste no aumento de performance da frota de transporte através da viabilização de estratégias ao prever a quantidade de horas por turno em que os caminhões ficarão ociosos por não haver equipamento para carrega-los (Apropriados no evento “Falta de Equipamento de Carga” no sistema de despacho), utilizando o método de aprendizado de máquina *Random Forest*. Para tanto, utilizaremos a base de dados disponibilizada pelo sistema de despacho de mina; O objetivo de tal abordagem consiste no uso de um sistema amplamente utilizado na indústria de mineração, minimizando custos de implantação de sistemas de instrumentação para coleta de dados e viabilizando a abrangência do trabalho para outras operações mineiras.

Como critério de avaliação dos resultados, comparam-se os resultados do modelo pelo coeficiente de correlação obtidos pelo software Weka.

1.2. Contribuições

Para a área, este trabalho identificou a relação de novos parâmetros com o evento de “Falta de Equipamento de Carga” na frota de transporte de mina como: “Atrasos Operacionais”, “Distância Média de Transporte” e a “Disponibilidade de equipamentos de Transporte (Analisado como Horas de Manutenção de equipamentos de Transporte)”. Viabilizando assim o mapeamento destes como forma de controle e mitigação das horas ociosas no evento de falta de equipamento de carga (estratégias como, por exemplo, viabilizar o aumento da distância de transporte para um cenário de baixa oferta de máquina de carga mitigando a ocorrência do evento).

Como contribuição acadêmica, o trabalho foi incluído no XV Simpósio Brasileiro de Automação Inteligente (SBAI), 17 a 20 de outubro de 2021, com o título de “Previsão de Falta de Equipamento de Carga para Frota de Transporte com base em Aprendizado de Máquinas e Cenários de Despacho”.

1.3. Organização do Trabalho

O restante deste trabalho está organizado como segue. O Capítulo 2 apresenta uma breve investigação quanto aos trabalhos relacionados, abrangendo aplicações do método *Random Forest* na indústria. O Capítulo 3 descreve o método desenvolvido para a aplicação considerada através do software Weka. O Capítulo 4 apresenta os resultados obtidos. O Capítulo 5 apresenta as conclusões advindas do trabalho.

2. Trabalhos Relacionados

Para subsidiar este trabalho, realizou-se investigação de artigos que abrangem a aplicação de aprendizado de máquina nas operações de lavra mineira.

A seguir, serão apresentados os trabalhos na sequência do “geral ao mais específico”, iniciando com o trabalho de Jung e Choi (2021) que apresenta um panorama geral dos trabalhos de aprendizado de máquina no setor mineral até o ano de 2021. A seguir o trabalho de Liaw e Wiener (2001) fundamentando o método de “Random Forest”. Temos então os trabalhos de Dong *et al.* (2013), Maheswari *et al.* (2020), Ouanan e Abdelwahed (2019), Kozielski *et al.* (2015) e Zelinska (2020) com aplicações similares de aprendizado de máquina no setor mineral.

Ademais, não foi identificado até o ano de 2021 na literatura menção de trabalhos utilizando o aprendizado de máquina para predição e controle de eventos de horas ociosas de transporte nas operações de gestão de frota no setor mineral.

Jung e 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 e Wiener (2001) apresentam 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. Os autores apresentam algumas considerações para trabalhos práticos com *Random Forest* através de interpretações de resultados pela quantidade de árvores.

Dong *et al.* (2013) apresentam uma aplicação prática do método *Random Forest* para predição de eventos geomecânicos no setor de mineração, mostrando a eficácia do aprendizado de máquina para melhora de segurança das operações mineiras. Os autores realizam um comparativo do resultado entre os métodos de *Random Forest*, *artificial neural network* e *support vector machine*, desmonstrando a vantagem do método *Random Forest*. Uma desvantagem do estudo é seu escopo centralizado em eventos geomecânicos sem abrangência para outros temas de 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 infraestrutura de rede para coleta de dados entre servidores 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 e 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 previsões 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.

3. Materiais e Métodos

Os parâmetros utilizados para predição do tempo dos caminhões parados por “Falta de Equipamento de Carga (min)” 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 basculo de transporte (min), tempo de manobra no basculo de transporte (min), tempo durante o basculo de transporte (min), tempo aguardando o basculo 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).

Os métodos para aprendizado de máquina podem ser classificados principalmente como “Aprendizado Supervisionado” e “Aprendizado não 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, a metodologia é implementada gerando o aprendizado da massa de dados sem suas respectivas “Saídas” almejadas. Esta ausência de direcionamento no aprendizado pode algumas vezes ser vantajoso, pois permite que o algoritmo identifique padrões que podem não ter sido previamente considerados (SATHYA e ABRAHAM, 2013).

Pelo software Weka é possível realizar o “K-fold Cross Validation”, que consiste em dividir os dados aleatoriamente em “k” subconjuntos. A cada teste, um conjunto formado por “k-1” subconjuntos são utilizados para treino enquanto os valores restantes são utilizados para avaliação do coeficiente de correlação. Neste trabalho, utilizado o “10-fold Cross Validation”.

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 de saída do *Random Forest* para o aprendizado de máquinas. Tal estratégia permitiu dar profundidade aos estudos em andamento e gerar maior confiança nos resultados obtidos por parte da equipe operacional. Para evitar vieses de análises, iniciamos o estudo com todas as informações disponíveis no despacho de equipamentos, reduzindo gradativamente na etapa de seleção de atributos.

Iniciamos com a estruturação da base a ser utilizada para o aprendizado do método supervisionado. No primeiro momento, realizamos 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 da rede de coleta e consulta aos dados tende a ser elevada na indústria de mineração pelo uso de equipamentos de fornecedores e sistemas distintos. Realizamos portanto, em caráter de estudo, consultas via SQL dos servidores e criamos uma chave primária temporal para integração das bases.

Por termos dados em períodos de coletas distintos, normalizamos todos na escala diária (aplicando ou média ponderada, ou somatória a depender do parâmetro), uma vez que todos os sistemas utilizados realizam o fechamento contábil no dia. A chave primária para mesclagem das bases de dados foi a “data” no horizonte diário.

Para a coleta dos dados, particular a esta operação mineira, o mesmo servidor em que se gera os relatórios também é responsável pela operacionalização dos caminhões em modo autônomo na mina. Portanto para não impactar a performance da frota optou-se por gerar pequenas consultas aos relatórios com uma massa de dados equivalente a 15 dias nos horários de menor demanda (03:00AM).

O agrupamento das informações para a massa de dados de 730 dias tal como a mesclagem entre os relatórios foi realizado através do “Powe Query” de tal forma que o incremento dos relatórios possa ser automatizado.

Buscando minimizar a variação da coleta entre os turnos, realizou-se um “Workshop” com a equipe responsável pela apropriação dos eventos no sistema de despacho padronizando assim os critérios de parada dos equipamentos de transporte. Também realizou-se a operação assistida nos turnos de trabalho no período de Mar/20 a Out/20 tal como a conferência diária dos eventos de “Falta de Equipamento de Carga” e seus critérios na rotina diária de Confiabilidade Operacional da Gerência de Operação e Infraestrutura de Mina de Brucutu.

Em seguida, realizamos 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 de mina.

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. Para o estudo, foi utilizado o coeficiente de correlação, por ser exportação nativa do sistema WEKA e podendo ser facilmente obtido o coeficiente de determinação elevando-se o coeficiente de correlação ao quadrado.

Há diversos métodos de seleção de atributos apresentados na literatura, como Seleção *Forward*, Eliminação *Backward*, Seleção Bidirecional e seleção Randômica (LEE, 2005).

Dentre estes, a eliminação *Backward* apresenta-se vantajoso quando há grande oferta de parâmetros disponíveis sem custos de processamento, pois quando realizamos o método de seleção *Forward* podemos ter o efeito supressor, em que um parâmetro somente se torna significativo quando outro está presente (GUIMARÃES e ALVES, 2008). De maneira simplificada, 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 fizemos uma análise de correlação entre as variáveis, uma vez que a permanência de atributos correlacionados pode potencializar a relevância de determinado parâmetro no modelo além do esperado. Posteriormente, aplicamos a redução gradativa dos parâmetros conforme a ordem temporal dos eventos em conformidade com o despacho de mina. Este ciclo se inicia no tempo de deslocamento com a bascula vazia, tempo de fila para carregar, tempo de manobra para carregar, tempo de aguardo a ser carregado, tempo de carregamento, tempo de deslocamento com a bascula cheia, tempo de fila para despejo, tempo de manobra para despejo, tempo de aguardo do despejo e tempo do despejo.

Ao final do processo, mapeamos os atributos cuja combinação apresentou melhor desempenho no coeficiente de determinação. Para fins de comparação entre a predição através de aprendizado de máquina, comparamos o coeficiente de determinação obtido pelo *backward elimination* com o coeficiente de determinação 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 *ensemble learning* pois a combinação das árvores de decisão são a base de sua classificação (KUNCHEVA, 2004).

A Figura 3.1 ilustra o modelo de predição via RF, aonde a classe final de predição é obtida através da votação da predição do resultado de cada árvore de decisão individualmente (daí o nome *forest*, ou floresta em português) (PRIMARTHA e ADHI TAMA, 2017). De acordo com Kuncheva (2004), os dois parâmetros comuns 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 os resultados dos modelos com 20 árvores (RF20), 50 árvores (RF50) e 100 árvores (RF100).

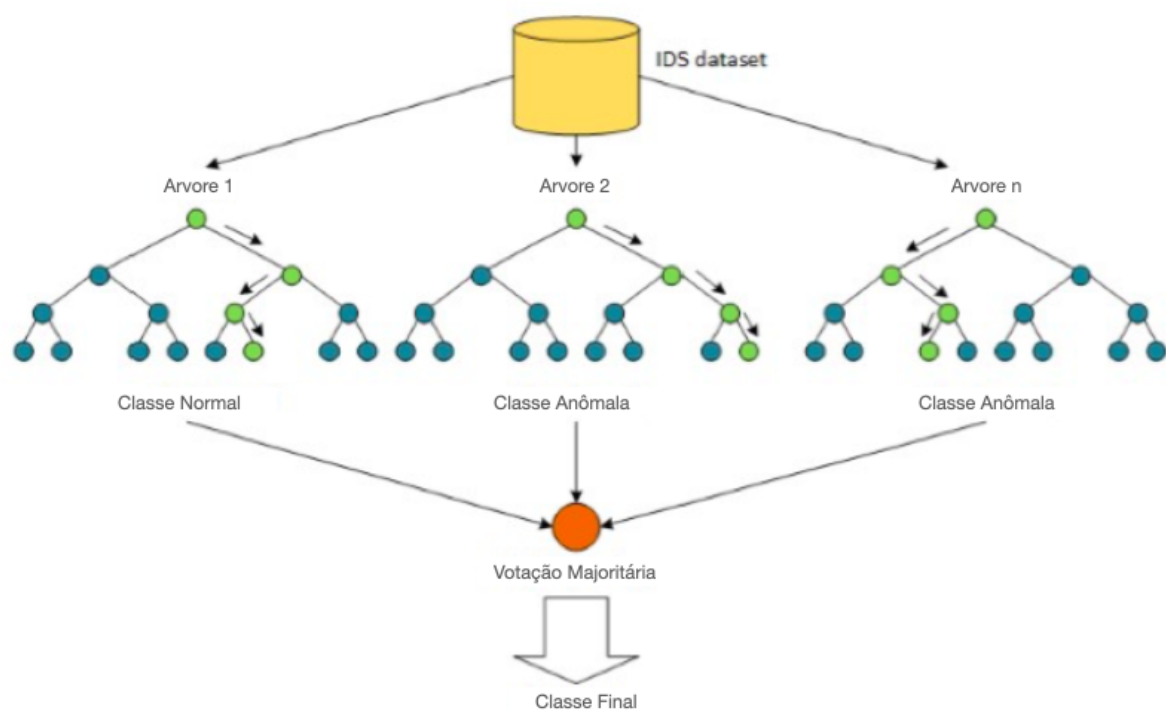


Figura 3.1: Ilustração de funcionamento do Classificador *Random Forest*.
 Fonte: Traduzido de [Primartha e Adhi Tama \(2017\)](#)

4. Resultados e Discussões

Este capítulo está organizado como segue. Na Seção 4.1, é apresentado o processo de preparação dos dados incluindo método de coleta tal como parâmetros iniciais. Na Seção 4.2 é abordada a etapa de pré-processamento e análise para a seleção de atributos. Na Seção 4.3 trata da geração do modelo através do software Weka. A Seção 4.4 apresenta o comparativo entre os modelos *Random Forest* gerados. Na seção 4.5 é apresentado os efeitos no aumento da frequência de coleta dos dados utilizados no aprendizado do modelo. Por fim, a seção 4.6 explora os efeitos ao incrementar o período de coleta dos dados para aprendizagem do modelo.

4.1. Preparação dos dados

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 Basculo 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 (devido ao método da operação em estudo ser “Lavra em Céu Aberto”, aonde chuvas e secas alteram a condição de trafegabilidade e performance de equipamentos) pelo índice pluviométrico, apurado diariamente.

4.2. Pré-Processamento e Análise

Iniciamos o estudo com 20 parâmetros e, a partir do agrupamento deles, realizamos o processo de *backward elimination*, analisando o indicador coeficiente de correlação entre o valor predito e o realizado em cada etapa pelo Weka, conforme mostra a Figura 4.1.

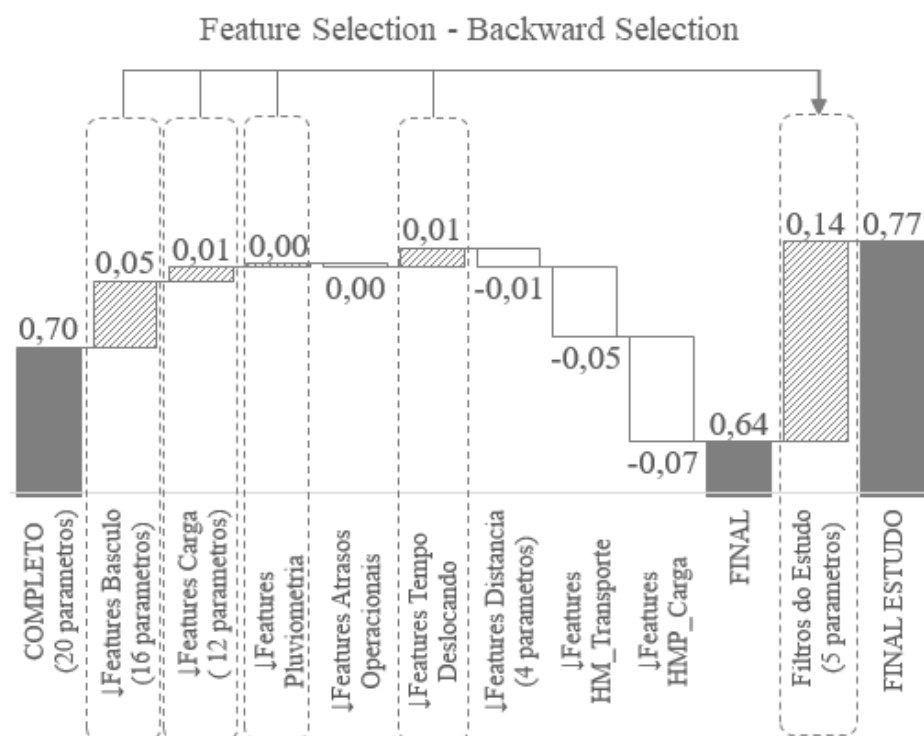


Figura 4.1: Processo de *BackWard Selection* para definição dos parâmetros a manter e a descartar. Nas barras de rachura tracejada temos o incremento do coeficiente de correlação pela redução dos parâmetros assinalados no eixo, Nas barras sem preenchimento temos o decréscimo do coeficiente de correlação pela redução dos parâmetros assinalados no eixo, após o final do estudo foram retirados portanto os 15 parâmetros das barras rachuradas resultando no modelo final de 5 parâmetros

Fonte: Autor

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).

Iniciamos o modelo com um coeficiente de correlação de 0,7 com 20 parâmetros. Após a exclusão dos parâmetros “tempo de basculamento”, “tempo de carregamento”, “tempo de deslocamento” e “pluviometria”, o coeficiente de correlação entre o predito e executado 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 4.2.

4.3. Entrada de parâmetros e geração do modelo

Utilizou-se o pacote de softwares WEKA 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 4.3). 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 relevância de cada atributo para a predição (Figura 4.4), 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.

A manutenção geral (HM) engloba tanto as ocorrências de manutenção preventiva (HMP) quanto corretiva (HMC). Identificou-se, portanto, que não basta somente o equipamento ter uma manutenção corretiva para a ocorrência do evento “Falta de Equipamento de Carga”, e

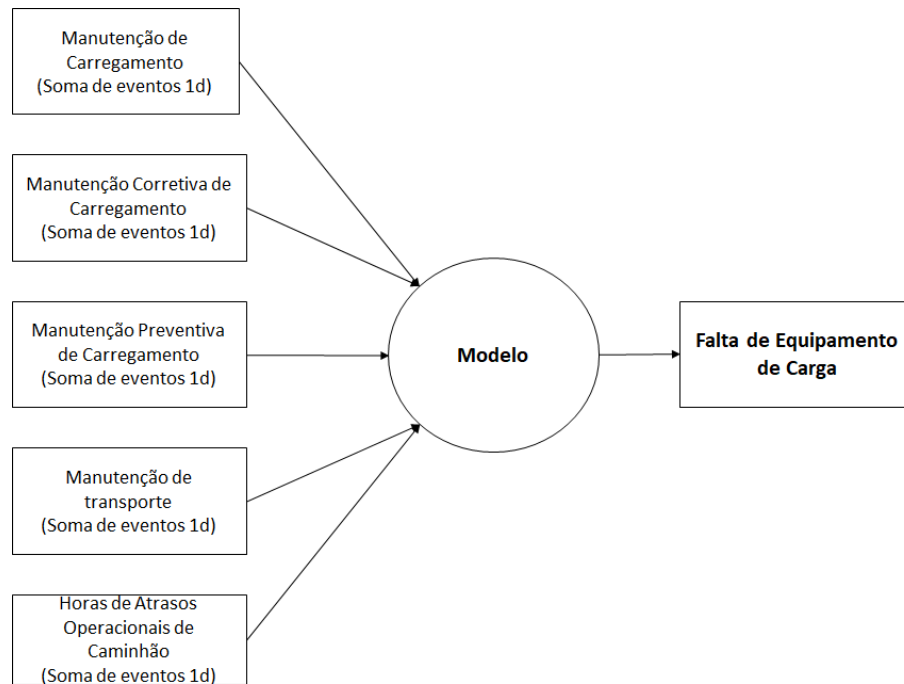


Figura 4.2: Utilização dos 5 parâmetros de entrada para predição da Falta de Equipamento de Carga.

Fonte: Autor

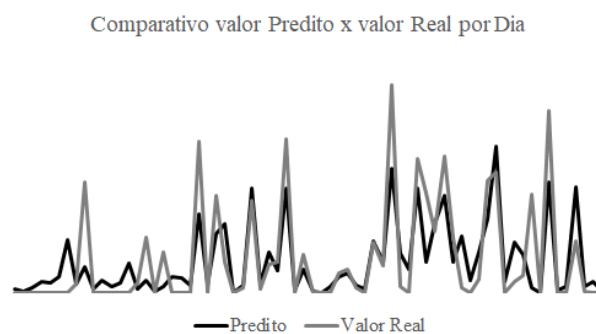


Figura 4.3: Avaliação temporal entre o valor predito e executado no período para o evento Falta de Equipamento de Carga

Fonte: Autor

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 durante determinado período sem a capacidade de absorção pela frota de carregamento.

```
RandomForest
Bagging with 50 iterations and base learner
weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 9 -do-not-check-capabilities
Attribute importance based on average impurity decrease (and number of nodes using that attribute)
3246991.9 ( 399) HM_Carga
3220546.27 ( 181) HMC_Carga
1967095.67 ( 128) HMP_Carga
1141048.37 ( 13) HAO
1040763.38 ( 216) DMT_Ch
579259.67 ( 284) HM_Transporte
```

Figura 4.4: Atributos de importância relacionados ao número de nós que os utilizam.
Fonte: Autor

4.4. Comparativo entre os modelos *Random Forest*

Para verificar se há diferença estatística entre os métodos RF20, RF50 e RF100 *Random Forest*, Primeiramente foi realizado o teste de normalidade Shapiro-Wilk, identificando se a distribuição do erro pode ser considerado uma distribuição normal, paramétrica ou não normal. Os resultados deste teste estão reportados na Tabela 4.1.

Tabela 4.1: 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

Fonte: Autor

Como os valores de *p-value* foram maiores do que 0,05, podemos admitir que os três métodos seguem uma distribuição normal. Desta maneira, foi viabilizada a realização do *Welch Two Sample t-test*, cujos resultados estão apresentados na Tabela 4.2.

Tabela 4.2: 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

Fonte: Autor

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

Assim, estudamos o comportamento do coeficiente de correlação de cada modelo, fixando em 30 *seeds*. A Tabela 4.3 reporta os resultados obtidos.

Fonte: Autor

A Figura 4.5 mostra os gráficos *box plot* de cada um dos métodos *Random Forest* com relação ao coeficiente de correlação.

Por sua vez, a Figura 4.6 mostra os histogramas dos métodos RF20, RF50 e RF100, em 30 *seeds*.

Pela Tabela 4.3 e Figuras 4.5 e 4.6, é possível notar que, apesar de todos os métodos possuírem média próxima, o método RF100 é o que tem a média e a mediana mais alta com relação ao coeficiente de correlação, e é o que tem a 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

Tabela 4.3: 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

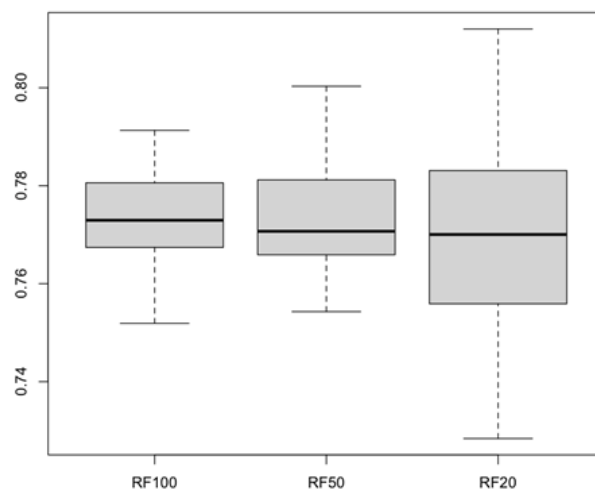


Figura 4.5: BoxPlot dos coeficientes de correlação para cada tipo de Random Forest.
Fonte: Autor.

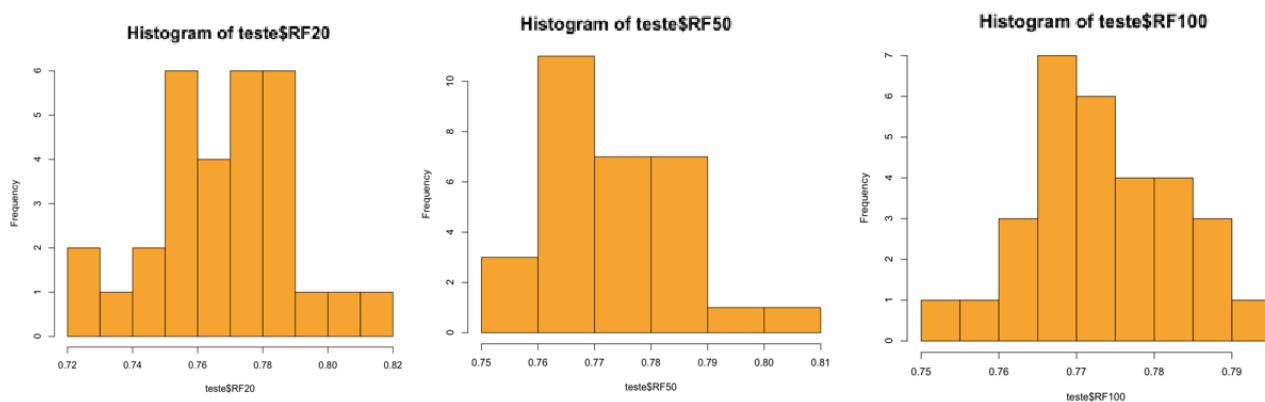


Figura 4.6: Histograma dos coeficientes de correlação para cada tipo de *Random Forest*.
Fonte: Autor.

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 4.7), 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).

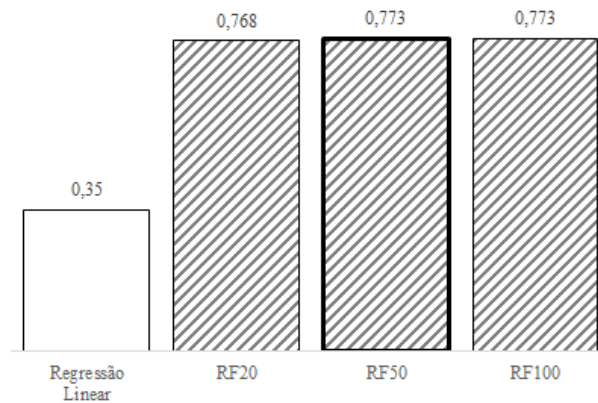


Figura 4.7: Comparativo entre os métodos com relação ao índice de correlação.
Fonte: Autor.

4.5. Aumento da frequência do *Dataset* de aprendizado

Com o objetivo da melhora da capacidade da predição utilizando o mesmo *Dataset* do modelo de predição, aumentamos a frequência de coleta de dados antes no horizonte diário (24 horas de coleta), agora no horizonte turno-a-turno (6 horas de coleta). Na prática tratamos o mesmo período porém com 4 vezes mais informações saindo de 860 mil para 3,44 milhões de registros. O resultado foi o incremento do coeficiente de correlação de 0,773 para 0,802 pelo método RF50 como apresentado na Tabela 4.6 e Figura 4.8. Tornamos então o modelo mais eficaz; Tanto em sua acurácia pela melhora de aderência entre o valor predito e o realizado, quanto em tempestividade, pois sua aplicação agora é aplicável para predição do próximo turno do trabalho enquanto antes dependia da conclusão dos 4 turnos que resultam no dia.

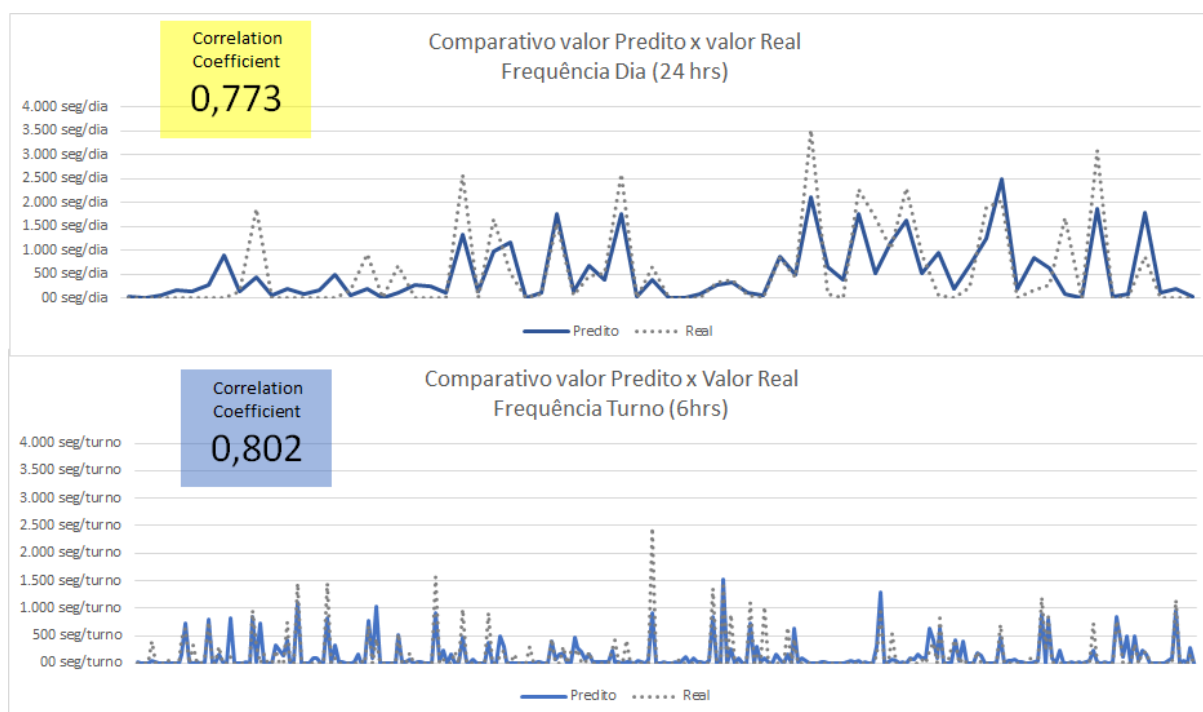


Figura 4.8: Comparativo entre o valor predito e realizado aumentando a frequência de coleta dos dados, antes 0,773 no horizonte 24hrs agora 0,802 no horizonte 6hrs.

Fonte: Autor.

Fonte: Autor.

O efeito colateral do aumento da frequência consiste na perda de alguns parâmetros que possuem seu monitoramento limitado a 24 horas, como a pluviometria. Este efeito não compromete o modelo em estudo pois após a etapa de seleção de atributos, os parâmetros que possuíam esta restrições não foram classificados como relevantes a este modelo.

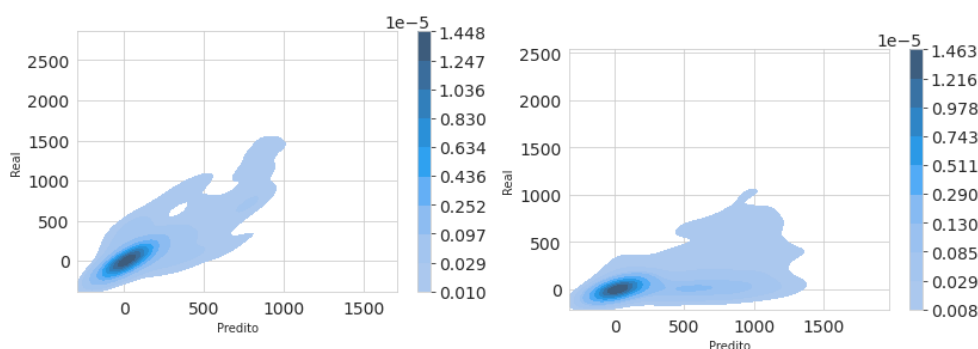
Tabela 4.4: Valores do coeficiente de correlação para cada tipo de frequência de coleta por Seed

Seed	Dia (24h)	Turno (6h)
1	0,786	0,805
2	0,754	0,801
3	0,759	0,808
4	0,767	0,798
5	0,766	0,803
6	0,779	0,798
7	0,782	0,803
8	0,781	0,801
9	0,800	0,799
10	0,767	0,805
11	0,757	0,798
12	0,771	0,795
13	0,780	0,806
14	0,761	0,800
15	0,789	0,818
16	0,761	0,807
17	0,771	0,791
18	0,770	0,802
19	0,789	0,794
20	0,769	0,801
21	0,766	0,815
22	0,772	0,810
23	0,768	0,809
24	0,780	0,805
25	0,782	0,796
26	0,776	0,793
27	0,776	0,798
28	0,795	0,796
29	0,761	0,812
30	0,767	0,801
Média	0,773	0,802
Desvio	0,009	0,005

4.6. Aumento do intervalo de amostragem do *Dataset* de aprendizado

Com o objetivo da melhora da capacidade da predição ampliando o intervalo de coleta do *Dataset* de aprendizado; Ampliamos a coleta para 01/11/2020 até 31/03/2022 e mantendo a frequência turno-a-turno (que apresentou maior eficácia), resultando em 11,7 milhões de registros. O resultado foi a redução do coeficiente de correlação utilizando o método RF50 de 0,805 para 0,691. Para avaliar a eficácia, realizamos um “teste cruzado” conforme apresentado na Figura 4.11 que consiste em utilizar o *Dataset* complementar ao utilizado na geração do modelo para seu teste, podendo assim comparar a capacidade de predição do modelo para um mesmo conjunto de dados de teste. Ao todo foram 4 testes como descrito abaixo:

- Aprendizado em Mar/20-Out/20 com teste por 10-Fold Cross Validation em Mar/20-Out/20 - Figura 4.9a
- Aprendizado em Mar/20-Out/20 com teste pelo dataset de nov/20-mar/22 - Figura 4.9b
- Aprendizado em nov/20-mar/22 com teste por 10-Fold Cross Validation em nov/20-mar/22 - Figura 4.10b
- Aprendizado em nov/20-mar/22 com teste pelo dataset de mar/20-out/20 - Figura 4.10a



(a) Teste por 10-Fold Cross Validation em Mar/20-Out/20

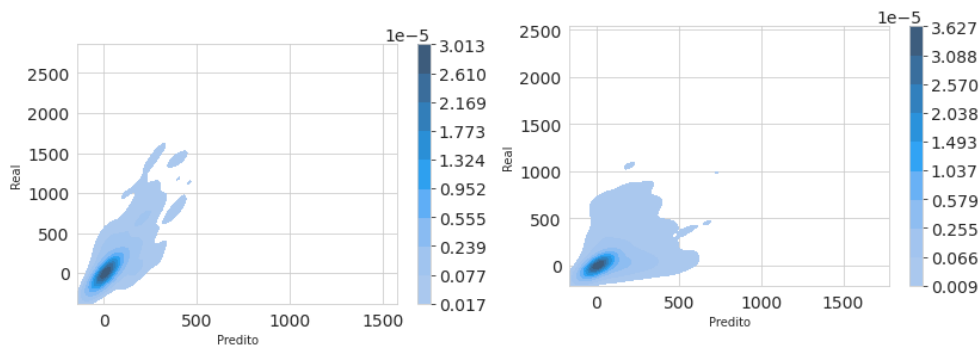
(b) Teste pelo dataset de nov/20-mar/22

Figura 4.9: Densidade de Kernel em 2D do Valor Predito x Real do Modelo gerado pelo aprendizado no período de Mar/20-Out/20 pelo método RF50.

Fonte: Autor.

Os valores do coeficiente de correlação dos 4 testes estão resumidos na Tabela 4.5. Constatamos que o aprendizado no intervalo de coleta superior (Nov/20 - Mar/22) torna o modelo mais eficaz pois eleva sua capacidade de predição para ambos os períodos para teste, Mar/20-Out/20 elevando de 0,805 para 0,809 e Nov/20-Mar/22 elevando de 0,585 para 0,691.

O motivo da redução do coeficiente de correlação ao expandir o período de amostragem foi coincidente com o *turnover* da equipe responsável pela apropriação, reforçando a



(a) Teste pelo dataset de Mar/20-Out/20 (b) Teste por 10-Fold Cross Validation em nov/20-mar/22

Figura 4.10: Densidade de Kernel em 2D do Valor Predito x Real do Modelo gerado pelo aprendizado no período de Nov/20-Mar/22 pelo método RF50.

Fonte: Autor.

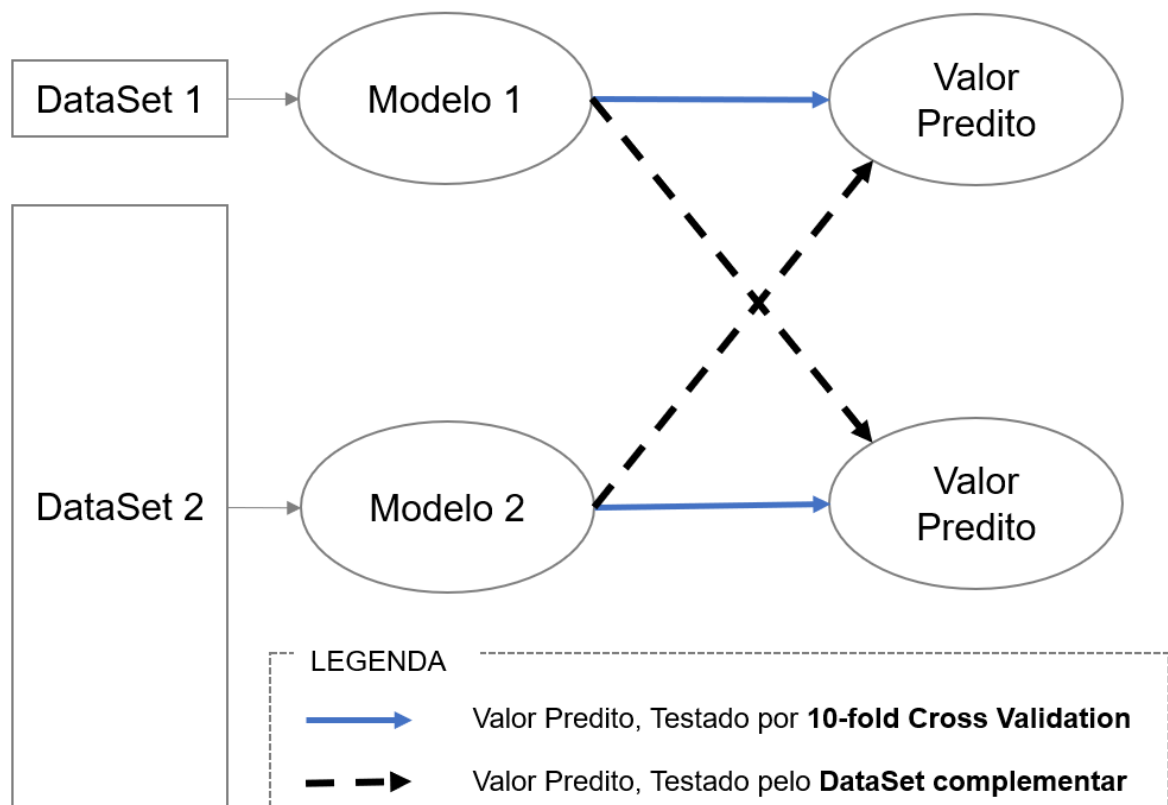


Figura 4.11: Representação do Teste Cruzado, testando o modelo no *DataSet* complementar ao utilizado para seu aprendizado.

Fonte: Autor

dependência de um bom padrão de apropriação para minimizar o efeito do viés humano no modelo.

Tabela 4.5: Valores do coeficiente de correlação para cada intervalo de coleta alterando o período para teste

Seed	Aprendizado em Mar/20-Out/20	Aprendizado em Nov/20-Mar/22
Teste em Mar/20-Out/20	0,802	0,809
Teste em Nov/20-Mar/22	0,585	0,691

Fonte: Autor.

Tabela 4.6: Valores do coeficiente de correlação para cada intervalo de coleta por Seed, testados por 10-Fold Cross Validation.

Seed	Mai/20-Out/20	Nov/20-Mar/22
1	0,805	0,692
2	0,801	0,687
3	0,808	0,687
4	0,798	0,694
5	0,803	0,694
6	0,798	0,696
7	0,803	0,690
8	0,801	0,684
9	0,799	0,689
10	0,805	0,692
11	0,798	0,693
12	0,795	0,690
13	0,806	0,693
14	0,800	0,690
15	0,818	0,687
16	0,807	0,699
17	0,791	0,693
18	0,802	0,693
19	0,794	0,689
20	0,801	0,688
21	0,815	0,696
22	0,810	0,689
23	0,809	0,690
24	0,805	0,689
25	0,796	0,686
26	0,793	0,693
27	0,798	0,694
28	0,796	0,695
29	0,812	0,686
30	0,801	0,689
Média	0,802	0,691
Desvio	0,005	0,003

Fonte: Autor.

5. Conclusões

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, realizamos a consolidação do *data-set* do despacho, desenvolvemos uma etapa de seleção de atributos a partir da metodologia “*Backward Selection*”, reduzindo de 20 para 5 parâmetros e, por fim, testamos 3 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.

A análise da importância de atributos identificou aqueles que possuem maior influência no modelo predito, identificando atributos além da manutenção corretiva vista inicialmente e otimizando o uso de recursos computacionais através da exclusão de parâmetros não considerados relevantes, como a pluviometria. Isto permitiu ampliar o escopo de atuação de controle do evento “Falta de Equipamento de Carga” de somente “Manutenção corretiva de Equipamentos de Carga” para “Manutenção Geral de equipamentos de Carregamento”, “Distância a ser percorrida pelos equipamentos de Transporte”, “Quantidade de atrasos operacionais no ciclo de transporte” e “Manutenção da frota de transporte”.

O aumento da frequência de coleta de dados para aprendizagem de 24 horas para 6 horas surtiu um efeito positivo na eficácia do modelo elevando o coeficiente de correlação de 0,773 para 0,802.

O aumento do período de coleta de dados para aprendizagem de mar/20-ou20 para nov/20-mar/22 surtiu um efeito positivo na eficácia do modelo quando comparado ao mesmo período de teste, elevando o coeficiente de correlação de 0,802 para 809.

A quantidade em horas por dia do evento de falta de equipamento de carga foi predita através do *Random Forest* com uma correlação de 0,809 no melhor cenário, superior à regressão linear antes considerada entre as “horas de manutenção de equipamentos de carga” e “horas de falta de equipamento de carga” com correlação de 0,35. Com isto, o uso do aprendizado de máquina para a predição na indústria de mineração se revela uma importante ferramenta de apoio à decisão, visto que esses métodos detectam a influência de parâmetros no processo nem sempre detectados em métodos estatísticos tradicionais.

5.1. Trabalhos Futuros

Para trabalhos futuros é recomendado:

- Reduzir a variabilidade advinda do fator humano através do treinamento da equipe e auditorias constantes, uma vez que os dados utilizados para o aprendizado provém de apontamentos manuais que estão sujeitos a um grau de subjetividade.
- Utilizar da capacidade de “Classificação” do aprendizado de máquina para ao invés de realizar a predição do evento, classificar baseado nos parâmetros mapeados no trabalho se o evento do caminhão ocioso se trata de “Falta de Equipamento de Carga” tal como outros eventos, automatizando assim o processo de apropriação de eventos no despacho.
- Implantar uma arquitetura de rede que permita a integração dos dados em tempo real, tornando a solução *online* frente a elaboração deste estudo que necessitou da consolidação manual de 160 relatórios.
- Avaliar o uso de outros métodos de aprendizado de máquina como *Support Vector Machine (SVM)* e *Deep Learning*, que de acordo com o mapeamento dos trabalhos científicos realizado por Jung e Choi (2021) também foram aplicados com sucesso no setor mineral.

Referências Bibliográficas

- DONG, L., BING LI, X., PENG, K. “Prediction of rockburst classification using Random Forest”, *Transactions of Nonferrous Metals Society of China*, v. 23, pp. 472–477, 2013. doi: 10.1016/S1003-6326(13)62487-5.
- GUIMARÃES, A. L. D. S., ALVES, W. D. O. “UM MODELO DE REGRESSÃO LOGÍSTICA PARA A PREVISÃO DE INSOLVÊNCIA DE OPERADORAS DE PLANOS DE SAÚDE”, *XXXII Encontro da ANPAD*, 2008.
- JUNG, D., CHOI, Y. “Systematic Review of Machine Learning Applications in Mining: Exploration, Exploitation, and Reclamation”, *Minerals*, v. 11, n. 2, 2021. ISSN: 2075-163X. doi: 10.3390/min11020148.
- KOZIELSKI, M., SIKORA, M., WRÓBEL, L. “DISESOR - decision support system for mining industry”. Em: *2015 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pp. 67–74, 2015. doi: 10.15439/2015F168.
- KUNCHEVA, L. I. *Combining Pattern Classifiers: Methods and Algorithms*. New York, Wiley-Interscience, 2004.
- LEE, H. D. *Seleção de atributos importantes para a extração de conhecimento de bases de dados*. Tese de doutorado, ICMC-USP, São Carlos, 2005.
- LIAW, A., WIENER, M. “Classification and Regression by randomForest”, *Forest*, v. 23, pp. 18–22, 2001. ISSN: 1609-3631.
- MAHESWARI, C., PRIYANKA, E. B., THANGAVEL, S., et al.. “Multiple regression analysis for the prediction of extraction efficiency in mining industry with industrial IoT”, *Production Engineering*, v. 14, n. 4, pp. 457–471, 2020. ISSN: 1863-7353. doi: 10.1007/s11740-020-00970-z. Disponível em: <<https://doi.org/10.1007/s11740-020-00970-z>>.
- MME. “Desempenho do setor mineral em 2020 supera expectativas”. 2021. Ministério de Minas e Energia. Disponível em encurtador.com.br/yABHY, acesso em 17/05/2021.

- OUANAN, H., ABDELWAHED, E. H. “Image processing and machine learning applications in mining industry: Mine 4.0”. Em: *2019 International Conference on Intelligent Systems and Advanced Computing Sciences (ISACS)*, pp. 1–5, Taza, Morocco, 2019. doi: 10.1109/ISACS48493.2019.9068884.
- PRIMARTHA, R., ADHI TAMA, B. “Anomaly detection using random forest: A performance revisited”. Em: *International Conference on Data and Software Engineering (ICoDSE)*, pp. 1–6, Palembang, Indonesia, 2017. IEEE.
- SATHYA, R., ABRAHAM, A. “Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification”, *International Journal of Advanced Research in Artificial Intelligence*, 2013.
- VALE. “Apresentação sobre Operações Mineiras”. 2019.
- ZELINSKA, S. “Machine learning: technologies and potential application at mining companies”. Em: *The International Conference on Sustainable Futures: Environmental, Technological, Social and Economic Matters (ICSF 2020)*, v. 166, p. 03007, Ukraine, May 20-22 2020. EDP Sciences.