



**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

**APLICAÇÃO DE SENSORES VIRTUAIS BASEADOS EM APRENDIZADO DE
MÁQUINA PARA ESTIMATIVA DA VAZÃO MÁSSICA DE MINÉRIO DE FERRO
EM CORREIAS TRANSPORTADORAS**

Sandro Geraldo Alves Sobreira

**Ouro Preto
Minas Gerais, Brasil
2021**

Sandro Geraldo Alves Sobreira

APLICAÇÃO DE SENSORES VIRTUAIS BASEADOS EM APRENDIZADO DE MÁQUINA PARA ESTIMATIVA DA VAZÃO MÁSSICA DE MINÉRIO DE FERRO EM CORREIAS TRANSPORTADORAS

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. Gustavo Pessin, D.Sc.

Coorientador: Prof. Geraldo Pereira Rocha Filho, D.Sc.

Coorientador: Prof. Pedro Henrique Gomes, Ph.D.

Ouro Preto

2021

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

S677a Sobreira, Sandro Geraldo Alves.

Aplicação de sensores virtuais baseados em aprendizado de máquina para estimativa da vazão mássica de minério de ferro em Correias Transportadoras. [manuscrito] / Sandro Geraldo Alves Sobreira. - 2021. 57 f.: il.: color., gráf., tab..

Orientador: Prof. Dr. Gustavo Pessin.

Coorientadores: Prof. Dr. Geraldo Pereira Rocha Filho, Dr. Pedro Henrique Gomes.

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 Mineraiis.

1. Aprendizado de Máquina. 2. Random Forest. 3. Sensores virtuais. I. Gomes, Pedro Henrique. II. Pessin, Gustavo. III. Rocha Filho, Geraldo Pereira. IV. Universidade Federal de Ouro Preto. V. Título.

CDU 681.5:622.2

Bibliotecário(a) Responsável: Angela Maria Raimundo - SIAPE: 1.644.803



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

Sandro Geraldo Alves Sobreira

Aplicação de Sensores Virtuais Baseados em Aprendizado de Máquina para Estimativa da Vazão Mássica de Minério de Ferro em Correias Transportadoras

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 08 de novembro de 2021

Membros da banca

Orientador - Gustavo Pessin - Orientador - Instituto Tecnológico Vale
Doutor - Geraldo Pereira Rocha Filho - Universidade de Brasília
Doutor - Jodelson Aguilar Sabino - Vale
Doutor - Heitor de Freitas Vieira - Sidia Intituto de Ciência e Tecnologia

Gustavo Pessin, orientador do trabalho, aprovou a versão final e autorizou seu depósito no Repositório Institucional da UFOP em 18/01/2022

Documento assinado eletronicamente por **Bruno Nazário Coelho, COORDENADOR(A) DE CURSO DE PÓS-GRADUAÇÃO EM INST. CONTROLE AUTOMAÇÃO DE PROCESSOS DE MINERAÇÃO**, em 31/01/2022,



às 16:48, 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 **0273579** e o código CRC **84C422A1**.

Referência: Caso responda este documento, indicar expressamente o Processo nº 23109.001012/2022-62

SEI nº 0273579

R. Diogo de Vasconcelos, 122, - Bairro Pilar Ouro Preto/MG, CEP 35400-000
Telefone: - www.ufop.br

Agradecimentos

Agradeço primeiramente aos meus pais, Emiliana e Sílvio, por todo o esforço para proporcionarem o melhor para mim e para meus irmãos, e a toda a minha família por todo o apoio durante todas as etapas da minha vida.

Gostaria de agradecer ainda à Universidade Federal de Ouro Preto, ao Instituto Tecnológico Vale e à Vale pela oportunidade de participar deste programa de mestrado. Aos meus amigos da equipe de Automação do S11D, do laboratório do ITV e da 4ª turma do PROFICAM, em especial ao Matheus, Pedro e Robson, sou grato pela colaboração durante o desenvolvimento deste trabalho.

Meus agradecimentos também ao Alexandre, André, Clauder, Dalton, Daniel, Eduardo, Krysthian, Lucas e todos os meus outros amigos de longa que sempre estiveram ao meu lado. Agradeço ainda a todas as pessoas que de alguma forma contribuíram de alguma forma para que eu chegasse até aqui.

Por fim, deixo um agradecimento especial aos meus orientadores Geraldo Pereira Rocha Filho, Gustavo Pessin e Pedro Henrique Gomes, por todos os ensinamentos durante o desenvolvimento deste trabalho.

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 S.A.

*“A vida é como andar de bicicleta.
Para manter o equilíbrio, você
deve continuar se movendo”
(Albert Einstein).*

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.)

APLICAÇÃO DE SENSORES VIRTUAIS BASEADOS EM APRENDIZADO DE MÁQUINA PARA ESTIMATIVA DA VAZÃO MÁSSICA DE MINÉRIO DE FERRO EM CORREIAS TRANSPORTADORAS

Sandro Geraldo Alves Sobreira

Novembro/2021

Orientadores: Gustavo Pessin
Geraldo Pereira Rocha Filho
Pedro Henrique Gomes

O transporte de minério de ferro através de correias transportadoras é amplamente utilizado na indústria da mineração. A medição de vazão mássica desse minério é fundamental para fins de proteção aos equipamentos, evitando que as correias sofram sobrecarga de minério, e para fins de controle operacional, ao se contabilizar a produção de um equipamento durante determinado período. Os equipamentos mais utilizados na medição da vazão mássica de minério são as balanças integradoras instaladas nas correias transportadoras. Por apresentarem um elevado custo de aquisição e necessitarem de uma equipe especializada para manutenção, a instalação dessas balanças são limitadas apenas a alguns pontos chave de uma planta de exploração mineral. O presente trabalho propõe a utilização de técnicas de aprendizado de máquina, mais especificamente M5P, *Random Forest* e REPTree, para o desenvolvimento de sensores virtuais para estimar a vazão mássica de minério em um equipamento que não é equipado com uma balança integradora. Os sensores foram desenvolvidos a partir de dados de corrente, torque e velocidade do motor de uma correia transportadora e das medições de vazão de minério de ferro de uma balança integradora instalada em outro transportador. Dois dos sensores virtuais propostos foram implementados em um controlador lógico programável de um transportador de correia no complexo S11D, no estado do Pará, onde foi possível verificar desempenho dos sensores virtuais em uma situação real. Como resultado, os sensores virtuais propostos foram capazes de medir a vazão de minério com uma taxa de erro aceitável em comparação com uma

balança integradora. A contabilização da produção dos sensores propostos também se mostrou próxima a totalização de uma balança.

Palavras-chave: Aprendizado de Máquina, M5P, *Random Forest*, REPTree, Sensores virtuais.

Macrotema: Mina; **Linha de Pesquisa:** Instrumentação no Processamento de Minérios;

Tema: Transportadores de Correia; **Área Relacionada da Vale:** S11D.

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.)

APPLICATION OF MACHINE LEARNING-BASED VIRTUAL SENSORS FOR ESTIMATING IRON ORE MASS FLOW IN CONVEYOR BELTS

Sandro Geraldo Alves Sobreira

November/2021

Advisors: Gustavo Pessin

Geraldo Pereira Rocha Filho

Pedro Henrique Gomes

Transporting iron ore by conveyor belts is widely used in the mining industry. The measurement of the mass flow of this ore is essential for equipment protection purposes, preventing the belts from being overloaded with ore, and for operational control purposes, when accounting for the production of equipment during a certain period. The equipment most used in measuring the mass flow of ore are integrator scales installed on conveyor belts. As they present a high acquisition cost and require a specialized team for maintenance, the installation of these scales is limited to just a few key points of a mineral exploration plant. The present work proposes the use of machine learning techniques, more specifically M5P, Random Forest, and REPTree, for the development of virtual sensors to estimate the mass flow of ore in equipment that is not equipped with an integrator scale. The sensors were developed from current, torque, and motor speed data from a conveyor belt and from iron ore flow measurements from an integrator scale installed on another conveyor. Two of the proposed virtual sensors were implemented in a programmable logic controller of a belt conveyor in the S11D complex, in the state of Pará, where it was possible to verify the performance of the virtual sensors in a real situation. As a result, the proposed virtual sensors were able to measure the ore flow with an acceptable error rate compared to an integrator scale. The accounting of the production of the proposed sensors also proved to be close to the totalization of a scale.

Keywords: Machine Learning, M5P, Random Forest, REPTree, Soft sensors.

Macrotheme: Mine; **Research Line:** Instrumentation in Ore Processing; **Theme:** Conveyor Belts; **Related Area of Vale:** S11D.

Lista de Figuras

Figura 1.1	Representação do processo de transporte de minério no complexo S11D.	18
Figura 1.2	Comparação entre a balança integradora e a balança virtual existente.	19
Figura 2.1	Extração de minério no Complexo S11D.	22
Figura 2.2	Representação esquemática de uma correia transportadora inclinada.	26
Figura 2.3	Análise da regressão linear da Equação 2.2.	27
Figura 3.1	Modelos dos Sensores Virtuais Desenvolvidos.	28
Figura 3.2	Correlação Cruzada e deslocamento entre a corrente do <i>Hopper</i> e a Balança Real.	29
Figura 3.3	Correlação Cruzada e deslocamento entre o torque do <i>Hopper</i> e a Balança Real.	30
Figura 3.4	Exemplo de <i>3-fold cross-validation</i>	31
Figura 3.5	Interface do Weka: (a) tela inicial do Weka e (b) página principal do do Weka <i>Explorer</i>	33
Figura 3.6	Interface do Weka: (a) lista para seleção do algoritmo a ser utilizado e (b) resultados obtidos com a implementação de um algoritmo.	34
Figura 3.7	(a) Configurações utilizadas no <i>Random Forest</i> (<i>soft sensors</i> 1 a 4). (b) Configurações utilizadas no <i>Reduced-Error Pruning Tree</i> (REPTree).	35
Figura 3.8	Configurações utilizadas no M5P.	36
Figura 4.1	Medições da balança integradora e estimativas realizadas pelos <i>soft sensors</i> 1 a 4.	38
Figura 4.2	Comparativo entre os <i>soft sensors</i> desenvolvidos e a balança integradora.	40
Figura 4.3	Parâmetros utilizados no desenvolvimento do <i>soft sensor</i> 5.	41
Figura 4.4	Comparativo entre a balança integradora, o <i>soft sensor</i> 5 e a balança virtual existente.	43
Figura 4.5	Comparativo entre a balança integradora, o REPTree (fase de testes) e a balança virtual existente.	44
Figura 4.6	Comparativo entre a balança integradora, o M5P e a balança virtual existente.	45
Figura 4.7	Comparativo entre a balança integradora, o M5P, do REPTree, do SS5 (<i>Random Forest</i>) e da balança virtual existente.	47

Figura 4.8	Vazão instantânea de minério obtida do sistema de controle da planta. . . .	51
Figura 4.9	Produção contabilizada pela balança integradora, pelo M5P, pelo REPTree e pela balança virtual existente.	52

Lista de Tabelas

Tabela 2.1	Comparação entre caminhões fora de estrada e correias transportadoras.	21
Tabela 2.2	Comparação entre sensores para medição de vazão mássica.	26
Tabela 4.1	Resultados da fase de treinamento dos <i>soft sensors</i> 1 a 4.	38
Tabela 4.2	Resultados da fase de teste dos <i>soft sensors</i> 1 a 4 e da balança virtual existente.	39
Tabela 4.3	Resultados da fase de treinamento dos <i>soft sensors</i> 4 e 5.	41
Tabela 4.4	Resultados da fase de teste dos <i>soft sensors</i> 4 e 5 e da balança virtual existente.	42
Tabela 4.5	Resultados das fases de treinamento e testes do REPTree e da balança virtual existente.	43
Tabela 4.6	Resultados das fases de treinamento e testes do M5P (fase de testes) e da balança virtual existente.	45
Tabela 4.7	Resultados da fase de testes do M5P, do REPTree, do SS5 (<i>Random Forest</i>) e da balança virtual existente.	46
Tabela 4.8	Resultados do M5P, do REPTree e da balança virtual existente coletados do sistema de controle da planta.	50
Tabela 4.9	Vazão média de minério.	51

Lista de Siglas e Abreviaturas

BVE Balança Virtual Existente

CLP Controlador Lógico Programável

GB Gigabyte

KNN *K-Nearest Neighbour*

MAE Erro Médio Absoluto (*Mean Absolute Error*)

min Minutos

MPa Megapascal

R² Coeficiente de Determinação

REP *Reduced-Error Pruning*

REPTree *Reduced-Error Pruning Tree*

RMSE Raiz do Erro Quadrático Médio (*Root Mean Square Error*)

SS Sensor Virtual (*Soft Sensor*)

ST Texto Estruturado (Structured Text)

t Toneladas

t/h Toneladas Por Hora

Weka *Waikato Environment for Knowledge Analysis*

Lista de Símbolos

I Corrente

τ Torque

\dot{m} Vazão Mássica

v Velocidade

Sumário

1	Introdução	17
1.1	Motivação	17
1.2	Objetivos	19
1.2.1	Objetivos Gerais	19
1.2.2	Objetivos Específicos	19
1.3	Estrutura do Trabalho	20
2	Referencial Teórico e Fundamentação Científica	21
2.1	Correias Transportadoras	21
2.2	<i>Soft Sensors</i>	22
2.3	<i>Reduced-Error Pruning Tree</i> (REPTree)	23
2.4	<i>Random Forest</i>	23
2.5	M5P	24
2.6	Medição de Vazão Mássica em Correias Transportadoras	25
3	Materiais e Métodos	28
3.1	Coleta de Dados	28
3.2	CrITÉrios para Avaliação	31
3.3	Weka	33
4	Resultados e Discussão	37
4.1	<i>Random Forest</i>	37
4.1.1	Treino dos <i>Soft Sensors</i> 1 a 4	37
4.1.2	Teste dos <i>Soft Sensors</i> 1 a 4	39
4.1.3	<i>Soft Sensor</i> 5: Corrente, Torque e Velocidade	40
4.2	REPTree	43
4.3	M5P	44
4.4	Comparativo Entre o M5P, o <i>Random Forest</i> e o REPTree	46
4.5	Implementações no CLP	47
4.5.1	Comparativo Entre as Estimativas em Tempo Real	50

5 Conclusão	53
6 Trabalhos Futuros	54
Referências Bibliográficas	55

1. Introdução

De acordo com a Agência Nacional de Mineração (2020), no ano de 2019 a produção de minério de ferro bruto no Brasil foi de aproximadamente 510 milhões de toneladas, sendo a mineradora Vale S.A. responsável por aproximadamente 71,90 % desta produção. O processo de transporte do minério extraído geralmente é realizado por caminhões, rodoviários ou fora de estrada, e também por correias transportadoras.

No caso das correias transportadoras, a vazão mássica de minério geralmente é medida com a utilização de balanças integradoras. Estas balanças possuem alta acurácia, mas em contra partida, seu custo de aquisição é elevado e também demandam manutenções periódicas para se manterem em boas condições operacionais (MOREIRA *et al.*, 2017). Para correias transportadoras que não são equipadas com balanças integradoras, pode-se estimar a vazão de material sobre a correia utilizando outros métodos, tais como o que utiliza a potência do motor de um transportador de correia inclinado, desenvolvido por Hulthén (2010), ou os que utilizam um transdutor de potência, um perfilômetro a laser, um sensor ultrassônico ou um strain gauge, testados por Väyrynen *et al.* (2013).

Um método amplamente utilizado em problemas de predição é o Aprendizado de Máquina, no qual um conjunto de dados é utilizado para realizar o treinamento de algoritmos que são programados para aprender através do conjunto de amostras fornecido (CARVALHO *et al.*, 2011). Entre suas aplicações, podemos destacar: a predição do fluxo de veículos em uma rodovia Crosby *et al.* (2016) ou de bicicletas em uma cidade (HOLMGREN *et al.*, 2017), a predição do tempo para liberação de vias após um acidente de trânsito (ZHAN *et al.*, 2011), a predição do consumo de energia em um hotel (AHMAD *et al.*, 2017) e a estimativa do teor de ferro na flotação de minério (GUEDES, 2020).

1.1 Motivação

Neste trabalho, chamaremos de *Hopper* um equipamento composto por uma moega de alimentação e um transportador de correia, destinado a receber minério transportado por caminhões fora de estrada. Ele tem a função de permitir que o minério das frentes de lavra que não estejam próximas ao britador principal do sistema seja incluído no circuito de transporte da mina. A Figura 1.1 mostra uma representação esquemática de um sistema de transporte de minério no Complexo S11D.

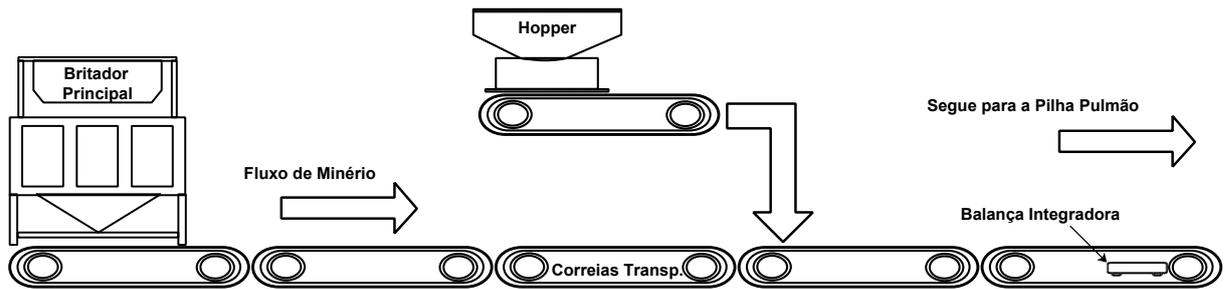


Figura 1.1: Representação do processo de transporte de minério no complexo S11D. Fonte: O autor.

Na concepção de projeto deste *Hopper*, não foi prevista a instalação de uma balança física para mensurar a vazão de minério que ele injeta no sistema. Desta maneira, fez-se necessário a criação de uma balança virtual para que esta vazão pudesse ser estimada. Atualmente, a Vazão Mássica (\dot{m}), em t/h, do transportador do *Hopper* é estimada através da Equação 1.1.

$$\dot{m} = \begin{cases} 0, & \text{se } I < 85A \\ 15,81 \times v + 142, & \text{se } I \geq 85A \end{cases} \quad (1.1)$$

Onde,

I é a corrente do motor do transportador, em Ampères;

v é a velocidade do motor do transportador, em percentagem da velocidade nominal do motor.

Como a Equação 1.1 leva em conta apenas a velocidade do transportador, a taxa da balança é fixa sempre que o transportador estiver rodando com uma velocidade constante. Este fato faz com que as variações de carga no transportador devido a quantidade de material existente na moega sejam desconsideradas. A Figura 1.2 mostra uma comparação entre a balança integradora e a balança virtual com dados coletados a cada 9 segundos durante aproximadamente 3 horas. É possível perceber que a balança virtual existente não consegue acompanhar as variações captadas pela balança integradora.

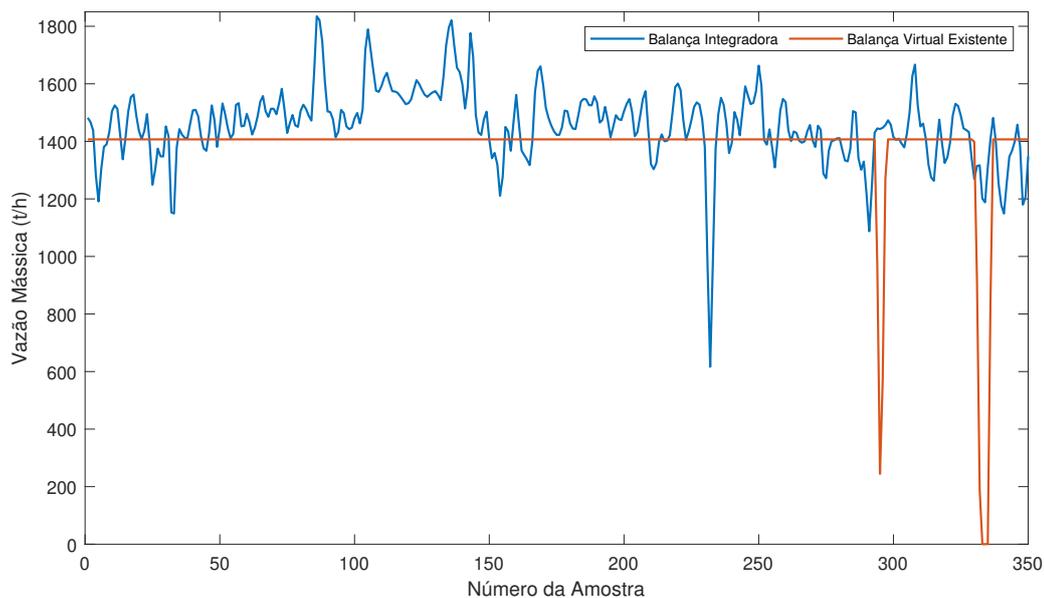


Figura 1.2: Comparação entre a balança integradora e a balança virtual existente. Fonte: O autor.

1.2 Objetivos

1.2.1 Objetivos Gerais

Este trabalho propõe a utilização de sensores e algoritmos de aprendizado de máquina (*machine learning*) para estimar a vazão mássica de minério que passa por uma correia transportadora.

1.2.2 Objetivos Específicos

Os objetivos específicos deste trabalho são:

- Desenvolver e investigar *soft sensors* para medição da vazão mássica de minério em uma correia transportadora utilizando três algoritmos de aprendizado de máquina: M5P, *Random Forest* e REPTree.
- Utilizar o *Random Forest* para avaliar, através de simulações, o comportamento do *soft sensor* de acordo com diferentes entradas, (1) apenas a corrente, (2) apenas o torque, (3) a corrente e o torque, e (4) corrente, torque e a velocidade.
- Validar o funcionamento *soft sensors* com base no M5P e no REPTree, ao embarcá-los em um CLP da planta.

1.3 Estrutura do Trabalho

O restante deste documento está organizado da seguinte maneira. O Capítulo 2 expõe os conceitos relacionados a correias transportadoras e suas balanças, além dos trabalhos relacionados. O Capítulo 3 detalha o método utilizado para o desenvolvimento do trabalho. O Capítulo 4 apresenta resultados parciais e discussões. No Capítulo 5 são apresentadas as conclusões. A dissertação é encerrada com o Capítulo 6, que apresenta as perspectivas para trabalhos futuros.

2. Referencial Teórico e Fundamentação Científica

Este capítulo apresenta os conceitos necessários para a fundamentação teórica necessária para o desenvolvimento da pesquisa, descrevendo os algoritmos utilizados e alguns métodos para medição de vazão mássica em correias transportadoras.

2.1 Correias Transportadoras

Com as crescentes preocupações ambientais e financeiras na indústria da mineração, a utilização de correias transportadoras vem crescendo em detrimento da utilização de caminhões fora de estrada. Segundo Osanloo e Paricheh (2020), durante o ciclo de operação de um caminhão fora de estrada, apenas 40% da energia consumida é gasta para movimentar a carga e os 60% restantes são gastos para movimentar apenas o caminhão. Nas correias transportadoras, a situação se inverte e 70% da energia consumida é utilizada para movimentar o material, com apenas 30% sendo consumida pela movimentação da correia.

Ribeiro *et al.* (2016) analisaram a viabilidade econômica da utilização de correias transportadoras em substituição aos caminhões fora de estrada, normalmente utilizados em plantas de extrações de minério. Assim como Osanloo e Paricheh (2020), os autores concluem que a utilização de correias transportadoras se mostra economicamente vantajosa em médio e longo prazo. A tabela 2.1 apresenta as vantagens e desvantagens de cada método de transporte de minério.

Tabela 2.1: Comparação entre caminhões fora de estrada e correias transportadoras. Fonte: Adaptada de Ribeiro *et al.* (2016).

Caminhões Fora de Estrada	Correias Transportadoras
Vantagens	
Menor custo de aquisição	Operam em longas distâncias
Maior flexibilidade	Maior capacidade
	Ecologicamente correta
	Produção contínua
Desvantagens	
Alto custo de operação e manutenção	Possui restrições de tamanho de material a ser transportado
Alto número de equipamentos para movimentar grandes quantidades de material	
Alta taxa de manutenção de estradas	

Este trabalho foi desenvolvido no Complexo S11D Eliezer Batista, de propriedade da Vale S.A., que foi projetado para ser operado utilizando o menor número de caminhões fora

de estrada possível. Desta maneira, o complexo conta com aproximadamente 37 quilômetros de correias transportadoras que são as responsáveis pela movimentação do minério desde o seu ponto de extração dentro da área de mina (SOUZA *et al.*, 2020), conforme mostrado na Figura 2.1.



Figura 2.1: Extração de minério no Complexo S11D. Fonte: Souza *et al.* (2020).

2.2 *Soft Sensors*

O aumento na quantidade de dados gerados e armazenados levou ao desenvolvimento de sensores que são baseados em softwares, os Sensores Virtuais (*Soft Sensors*). De acordo com Lotufo e Garcia (2008), *soft sensors* são modelos que permitem deduzir o valor de uma variável com base em outras informações do processo, sem que seja necessário realizar a medição dessa variável diretamente. Existem três métodos para modelar um *soft sensor* (FORTUNA *et al.*, 2007; LIN *et al.*, 2007; LOTUFO & GARCIA, 2008):

- *Model-driven*, ou caixa branca, na qual existe um modelo que descreve satisfatoriamente o processo ao qual o *soft sensor* será aplicado;
- *Data-driven*, ou caixa preta, na qual o *soft sensor* é desenvolvido apenas com base em dados históricos do processo. Nesse caso, não é necessário modelar o processo;
- *Gray-box*, ou caixa cinza, que é uma combinação dos dois métodos anteriores. Nele, as informações disponíveis a respeito do processo são utilizadas para refinar um modelo

data-driven.

Segundo Fortuna *et al.* (2007), os principais benefícios dos *soft sensors* são:

- Representam uma alternativa de baixo custo quando comparado aos sensores físicos, possibilitando mais pontos de monitoramento;
- Podem trabalhar em conjunto com sensores físicos, inclusive para identificar quando eles apresentam falhas;
- Podem ser implementados em dispositivos existentes;
- Fornecem estimativas em tempo real, se mostrando uma opção para medições onde sensores físicos dependem de análises demoradas

2.3 *Reduced-Error Pruning Tree (REPTree)*

Proposto por Quinlan (1987), o *Reduced-Error Pruning* (REP) é um método para realizar a poda de árvores de classificação e regressão. Considerando uma árvore completa T_{max} , para cada um de seus nós t é realizada uma comparação entre os erros de classificação (ou regressão) do conjunto de poda em duas situações: quando a subárvore T_t é mantida e quando o nó t é transformado em uma folha. Nos casos em que a árvore simplificada apresenta erros menores, é realizada a poda de T_t . Este passo é repetido até que a simplificação da árvore faça com que o erro aumente (ESPOSITO *et al.*, 1997).

Crosby *et al.* (2016) analisaram o desempenho do REPTree, do *K-Nearest Neighbour* (KNN) e da regressão linear para prever o fluxo de veículos nas rodovias do Reino Unido. Para desenvolver seu estudo, eles utilizaram uma base de dados com mais de 116 mil dados de medições de tráfego coletados através de 8 mil sensores instalados na Inglaterra e no País de Gales. O REPTree apresentou a maior acurácia entre os três, com 88,2%, quando comparado a 74,0% do e KNN e 41,0% da regressão linear.

Holmgren *et al.* (2017) realizaram uma comparação entre diversos algoritmos para realizar a predição da quantidade de bicicletas que trafegam em uma região do centro da cidade de *Malmö* na Suécia. Os autores utilizaram como base os dados coletados por um contador de bicicletas instalado na região. Entre os algoritmos testados, o REPTree e o *Support Vector Regression* apresentaram os melhores resultados.

2.4 *Random Forest*

O algoritmo de aprendizado de máquina utilizado neste trabalho é *Random Forest*, que foi introduzido por Breiman (2001). Segundo o autor, as *Random Forests* são uma combinação de várias árvores de decisão, onde cada uma das árvores influencia na classificação final do

método. Seu princípio de funcionamento é baseado no conceito de "dividir e conquistar", a base de dados é dividida em pequenas amostras, que são classificadas utilizando as árvores de decisão e ao final, o resultado de todas elas são agrupados (BIAU & SCORNET, 2016).

Guedes (2020) utilizou dados de ao menos 15 variáveis para desenvolver dois *soft sensors* para a predição do teor de ferro no concentrado da flotação de minério de ferro. Um deles foi desenvolvido utilizando Redes Neurais Artificiais (RNA) enquanto o outro utilizou *Random Forest*. A *Random Forest* utilizada por ele possui 500 árvores de decisão e cada folha possui um tamanho mínimo igual a 3. Em seus resultados, o autor destaca que o *soft sensor* que utilizou *Random Forest* obteve os melhores resultados, apresentando menores taxas de erros.

Adusumilli *et al.* (2013) propuseram um sistema de navegação que utiliza *Random Forest* para realizar a integração entre um Sistema de Navegação Inercial (INS, *Inertial Navigation System*) e o Sistema de Posicionamento Global (GPS, *Global Positioning System*). O *Random Forest* foi utilizado para fazer a regressão dos dados de velocidade e azimute coletados enquanto o GPS está em funcionamento, para prever as coordenadas de GPS quando não houvesse conexão com satélites. Os autores simularam 4 situações de perda de sinal do GPS e verificaram que o método proposto por eles proporcionou a reduções entre 24,06% e 55,75% no erro de posicionamento, quando comparado com uma solução já existente.

Ahmad *et al.* (2017) compararam o desempenho do *Random Forest* e de uma Rede Neural Artificial para predizer o consumo de energia em um hotel de Madrid na Espanha. Para isso, os autores coletaram dados de consumo de energia do sistema de climatização do hotel e também da temperatura externa, da velocidade do vento e da umidade relativa do ar de uma estação meteorológica localizada próximo ao hotel. Além disso, eles levaram em consideração o horário do dia, o dia da semana, o número de quartos reservados e o número de hóspedes do hotel. Para este estudo, a Rede Neural Artificial apresentou um desempenho melhor do que o *Random Forest*, com valores de erro quadrático médio de 4,97kWh e 6,10 kWh, respectivamente. Mesmo com o *Random Forest* se saindo ligeiramente pior, os autores destacam que ele foi capaz de prever o consumo de energia com um erro aceitável para a aplicação.

Couronné *et al.* (2018) realizaram um extenso comparativo sobre o desempenho para classificação entre o *Random Forest* e a regressão logística utilizando 243 conjuntos de dados reais. Em cerca de 69% dos conjuntos de dados estudados o *Random Forest* se mostrou mais preciso que a regressão logística. Os autores ainda destacam que o *Random Forest* foi implementado com parâmetros padrão, possibilitando que haja uma melhora de desempenho com a realização de um ajuste nos parâmetros.

2.5 M5P

O M5 é um método proposto por Quinlan (1992) que utiliza o conceito de árvores de modelos. Neste modelo, as árvores geradas possuem modelos lineares em suas folhas, ao invés de valores numéricos como em árvores de regressão. O autor destaca que o M5 é capaz de lidar

com centenas de atributos e que as árvores geradas são, geralmente, menores do que as geradas por árvores de regressão. Wang e Witten (1997) modificaram o M5 original para aumentar a capacidade do algoritmo lidar com valores ausentes dos conjuntos de dados, dando origem ao M5P.

Zhan *et al.* (2011) coletaram dados de 2535 acidentes de trânsito com bloqueios de pista de rolamento, entre abril de 2006 e março de 2007 na cidade de *Fort Lauderdale* no estado da Flórida (EUA). Então, eles desenvolveram dois modelos para prever o tempo de liberação da pista após um acidente, um com base no M5P e outro utilizando regressão linear. Na etapa de testes dos modelos, foram utilizados dados de 50 acidentes entre abril e maio de 2007. O modelo baseado no M5P apresentou melhores resultados comparado com o baseado na regressão linear, com Erro Médio Absoluto (*Mean Absolute Error*) (MAE) de 34,01 min e 37,70 min, respectivamente. Os autores destacam ainda, que a grande vantagem do M5P quando comparado com outros algoritmos tradicionais é sua capacidade de lidar tanto com variáveis contínuas quanto com variáveis categóricas.

Behnood *et al.* (2017) utilizaram o M5P para prever a resistência a compressão de concretos normais e de alto desempenho. Em seus resultados, os autores destacam que o M5P obteve um desempenho muito similar nas fases de treinamento e de testes. Enquanto na fase de treinamento o coeficiente de correlação e o MAE foram de 0,962 e 3,928 MPa, nos testes eles foram de 0,962 e 4,485 MPa, respectivamente. Por fim, eles concluem que o algoritmo baseado em M5P se mostrou capaz de produzir boas previsões da resistência a compressão tanto em concretos normais quanto nos de alto desempenho.

2.6 Medição de Vazão Mássica em Correias Transportadoras

Na indústria da mineração, a medição da vazão mássica do minério de ferro transportado por correias transportadoras geralmente é feita utilizando balanças integradoras. O custo de aquisição de uma balança integradora é de aproximadamente R\$ 95.000,00 (noventa e cinco mil reais) e a este valor ainda devem ser somados os custos de infraestrutura elétrica, cabeamento e mão de obra para instalação. Por este motivo, em uma planta de exploração de minério, as balanças são instaladas apenas em algumas correias transportadoras.(MOREIRA *et al.*, 2017).

Väyrynen *et al.* (2013) compararam diferentes métodos para medição de vazão mássica de material transportado por correias. Os sensores estudados pelos autores foram um transdutor de potência, um perfilômetro a laser, um sensor ultrassônico e um *strain gauge*. Eles realizaram experimentos e utilizaram regressão linear para analisar o comportamento de cada um dos sensores estudados. O Coeficiente de Determinação (R^2) de cada sensor é mostrado na Tabela 2.2. Como pode ser visto, o transdutor de potência apresentou o maior R^2 e todos os outros ficaram acima de 0,96, indicando que os modelos apresentam alta capacidade de prever a vazão de material a partir das medições.

Tabela 2.2: Comparação entre sensores para medição de vazão mássica. Fonte: Adaptada de Väyrynen *et al.* (2013).

<i>Sensor</i>	Coefficiente de Determinação (R^2)
Transdutor de Potência	$0,984 \pm 0,011$
Perfilômetro a Laser	$0,975 \pm 0,001$
Sensor Ultrassônico	$0,961 \pm 0,009$
<i>Strain Gauge</i>	$0,964 \pm 0,023$

Hulthén (2010) desenvolveu um método para medir a vazão mássica de material transportado sobre correias transportadoras inclinadas que trabalhem transportando material de um ponto mais baixo para um mais alto. No desenvolvimento do seu método, os autores se basearam no princípio que uma correia inclinada trabalha convertendo energia elétrica em energia potencial gravitacional. A Figura 2.2 mostra a representação esquemática de uma correia transportadora inclinada.

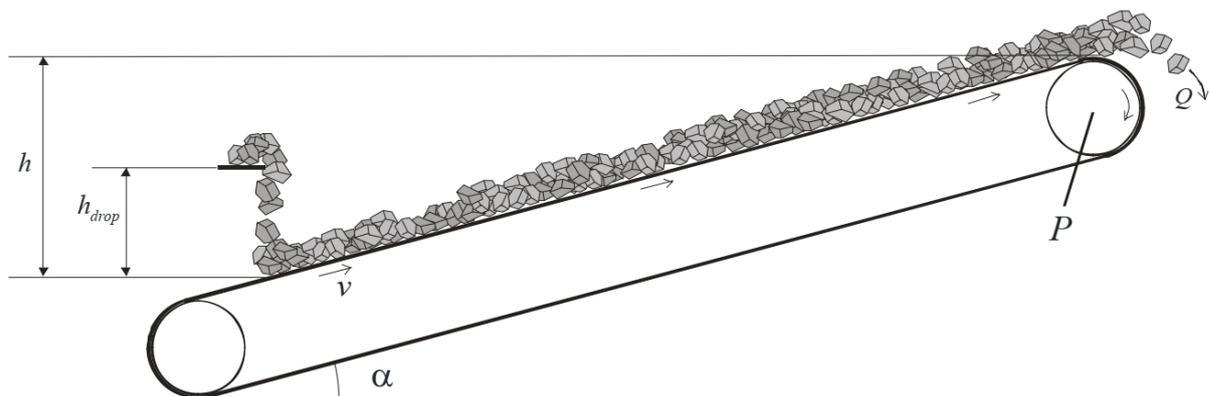


Figura 2.2: Representação esquemática de uma correia transportadora inclinada. Fonte: Hulthén (2010).

Partindo desses conceitos, os autores afirmam que a vazão mássica (\dot{m}) de uma correia transportadora pode ser calculada a partir da Equação 2.1.

$$\dot{m} = \frac{(P_{electrical} - P_{idle}) \eta_{tot}}{gh + v^2 + v \sqrt{2gh_{drop}} \sin \alpha} \quad (2.1)$$

Onde,

$P_{electrical}$ é a energia consumida pela correia com carga;

P_{idle} é a energia consumida pela correia sem carga;

η_{tot} é o coeficiente de eficiência da correia;

g é a aceleração da gravidade;

v é a velocidade da correia;

h é a diferença de altura entre os pontos carga e descarga da correia;

h_{drop} é a altura da queda do material no ponto de carga;

α é o ângulo de inclinação da correia.

Moreira *et al.* (2017) aplicaram o método desenvolvido por Hulthén (2010) em uma correia transportadora intermediária de uma usina de beneficiamento de minério. Os autores coletaram dados minuto a minuto da potência do motor e da balança integradora de uma correia anterior à que eles desejavam medir a vazão. Com base nos dados, aos autores chegaram na equação 2.2, que descreve a vazão mássica do material sobre a correia. Quando considerado a média mensal, os resultados obtidos tiveram um erro médio na faixa de $\pm 4\%$, se mostrando um ótimo resultado para a aplicação proposta.

$$\dot{m} = 12,75 P_{medida} - 286,86 \quad (2.2)$$

Onde,

P_{medida} é a energia consumida pela correia com carga.

Conforme mostrado na Figura 2.3, o R^2 da equação encontrada foi igual a 0,7803. Este valor é menor do que o encontrado por Väyrynen *et al.* (2013), 0,984, mas ainda assim mostra que a equação descreve de forma satisfatória a vazão mássica de minério de ferro no transportador estudado.

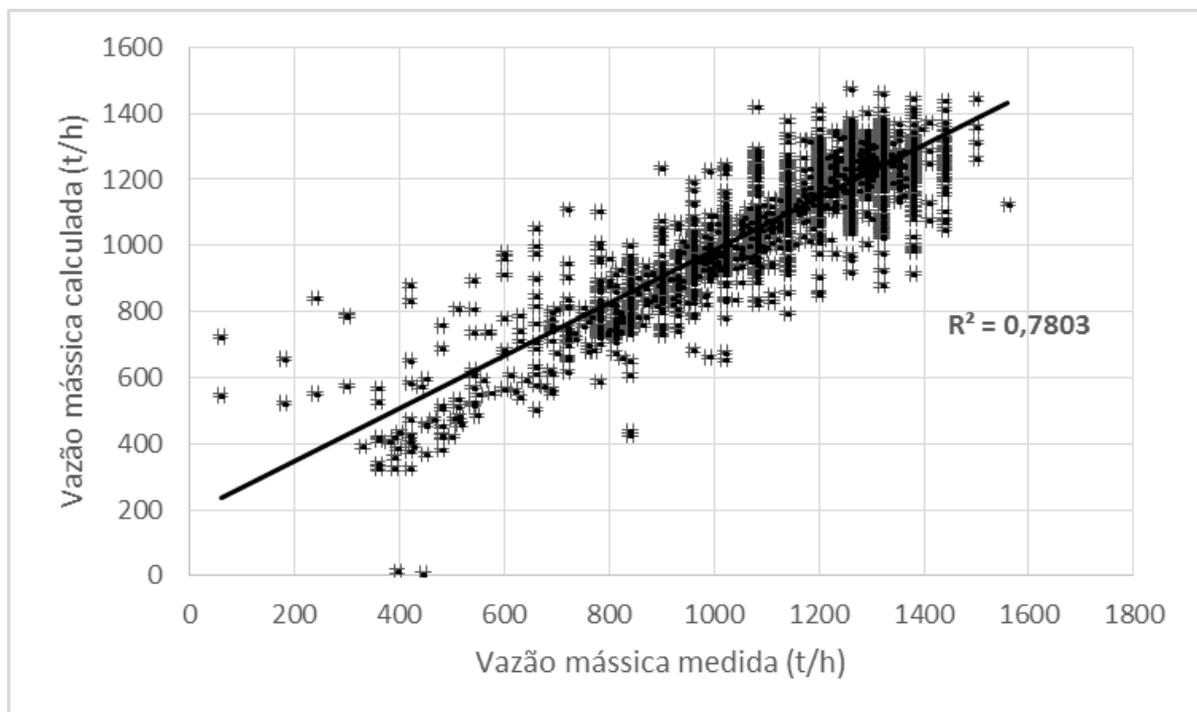


Figura 2.3: Análise da regressão linear da Equação 2.2. Fonte: Moreira *et al.* (2017).

3. Materiais e Métodos

Para atingir o objetivo de estimar a vazão mássica de minério em uma correia transportadora, optou-se por utilizar o método da pesquisa experimental para desenvolver os sensores virtuais com base em modelos gerados por três algoritmos de aprendizagem de máquina: M5P, *Random Forest* e REPTree. O ponto de partida do trabalho foi a realização da coleta de dados de corrente, torque e velocidade do motor do *Hopper* e da vazão de minério da balança integradora. Os dados coletados foram utilizados para alimentar o software Weka, que é apresentado no subcapítulo 3.3, e gerar os modelos com base no três algoritmos estudados. A Figura 3.1 apresenta o modelo que foi utilizado nos sensores virtuais desenvolvidos, variando apenas os parâmetros de entrada (corrente, torque ou velocidade, ou uma combinação entre eles) e também o algoritmo utilizado (M5P, *Random Forest* ou REPTree, apenas um por *soft sensor*).



Figura 3.1: Modelos dos Sensores Virtuais Desenvolvidos. Fonte: O autor.

Em uma segunda etapa do trabalho, alguns dos modelos gerados pelo Weka foram implementados em um CLP do sistema de controle do S11D para validar como os *soft sensors* se comportam para estimar a vazão de minério em tempo real.

3.1 Coleta de Dados

Para chegar até a primeira balança do circuito, o minério que sai do *Hopper* precisa passar por dois transportadores de correia. Devido a este fato, primeiramente é necessário calcular o deslocamento no tempo entre as variações dos parâmetros do motor do *Hopper* (torque e corrente) e a balança do transportador de correia. Para realizar esse cálculo, foram coletados 1202 valores para cada uma das variáveis: corrente e torque do *Hopper*, taxa da balança real da correia transportadora. Os 1202 valores coletados correspondem a cerca de 4 horas e 10 minutos de operação, o que fez com que o intervalo no tempo entre cada medição fosse de, aproximadamente, 12,5 segundos.

Ingle e Proakis (2011) definem a correlação cruzada como a medida em que dois vetores são similares. A correlação cruzada, $r_{x,y}(\ell)$, entre dois vetores, $x(n)$ e $y(n)$, pode ser calculada através da Equação 3.1. Quanto maior o valor de $r_{x,y}(\ell)$, maior a correlação entre os dois vetores.

$$r_{x,y}(\ell) = \sum_{n=-\infty}^{\infty} x(n)y(n-\ell) \quad (3.1)$$

Onde,

ℓ é o deslocamento entre os vetores.

Para calcular a correlação cruzada entre os dados coletados, foi utilizado o software MATLAB. Ele possui a função *xcorr* que calcula automaticamente o coeficiente de correlação cruzada entre dois vetores e também o deslocamento entre eles (THE MATHWORKS, INC., 2021). Para melhorar a visualização da correlação entre os sinais, a função *xcorr* foi utilizada com a opção de normalização ativada, fazendo com que a correlação máxima entre dois sinais fosse igual a 1.

Como pode ser visto na Figura 3.2, a corrente do *Hopper* e a taxa na balança apresentam um coeficiente de correlação igual a 0,8999 para um deslocamento de -43 posições, sendo este o maior coeficiente encontrado. Para deslocamentos de -44 e -42 posições, os coeficientes de correlação foram de 0,8994 e 0,8991, respectivamente.

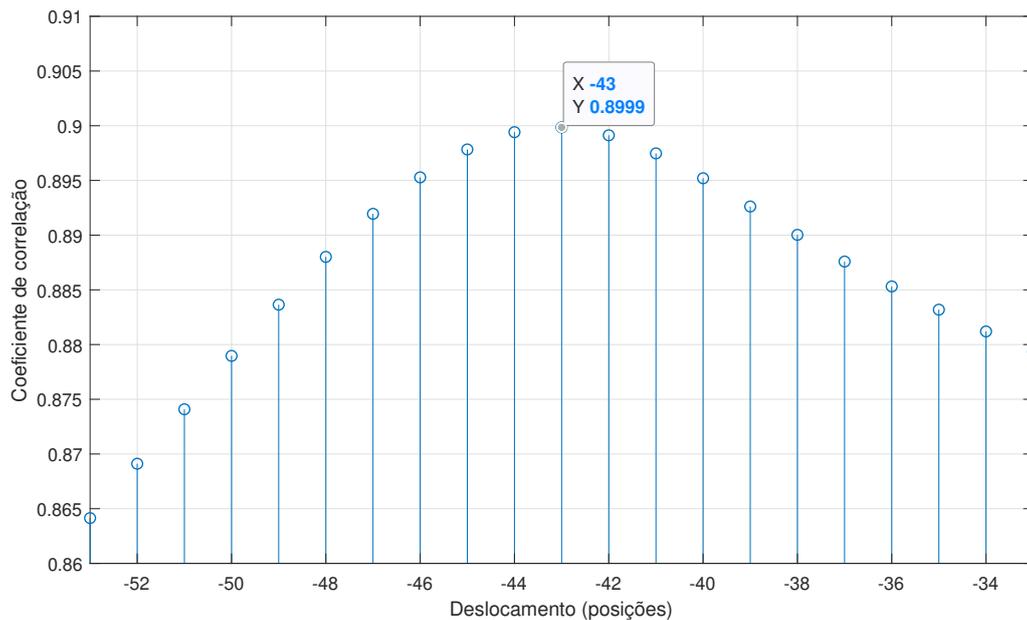


Figura 3.2: Correlação Cruzada e deslocamento entre a corrente do *Hopper* e a Balança Real. Fonte: O autor.

Ao analisar a variação da balança em relação à variação do torque, o maior coeficiente de correlação encontrado foi de 0,9579 para um deslocamento de -44 posições, conforme mostrado na Figura 3.3. Para deslocamentos de -45 e -43 posições, os coeficientes de correlação foram de 0,9557 e 0,9571, respectivamente.

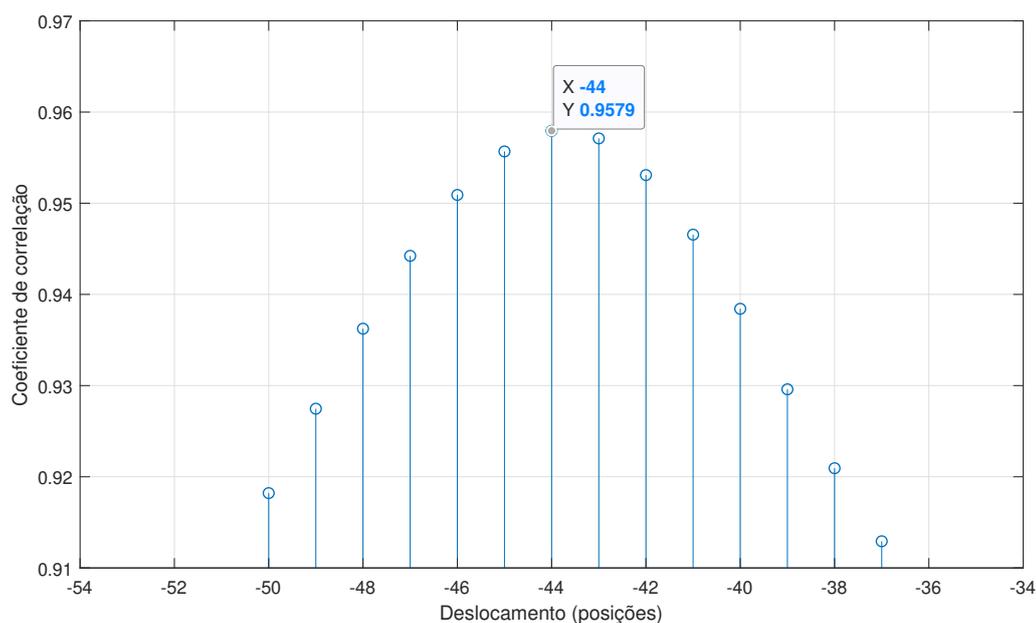


Figura 3.3: Correlação Cruzada e deslocamento entre o torque do *Hopper* e a Balança Real.
Fonte: O autor.

Como o torque apresentou o maior coeficiente de correlação no conjunto analisado, definiu-se que, para o desenvolvimento do *soft sensor*, a coleta de dados seria realizada com os dados da balança estando 9 minutos e 10 segundos atrasados em relação aos dados do *Hopper*.

Desta maneira, foi realizada uma nova coleta de dados entre os dias 27/07/2021 e 15/08/2021. Foram coletados 19.200 amostras que totalizam aproximadamente 32 horas de operação do equipamento. Por ser um equipamento recém instalado na planta, não há uma grande base de dados histórica para ser utilizada. Outro fator que limitou a obtenção de dados é que para evitar interferências do minério que vem do britador principal deste sistema na taxa da balança real, as coletas só foram realizadas durante em períodos em que o britador principal estava parado.

Após a coleta dos dados, foi feita uma análise para verificar a necessidade de pré-processar os dados coletados antes que eles pudessem ser utilizados. O pré-processamento de dados compreende todas as ações realizadas com os dados coletados antes de se iniciar a análise, tais como ordenação, seleção, filtragem, edição e amostragem dos dados (FAMILI *et al.*, 1997). A principal razão para se realizar o pré-processamento de dados é que, geralmente, os dados coletados estão incompletos, apresentam ruídos ou valores fora da escala de trabalho (GARCÍA *et al.*, 2015). Para este trabalho o pré-processamento realizado no conjunto de dados coletados consistiu de:

- Remoção dos dados em que a velocidade do transportador do *Hopper* fosse menor que

5% da velocidade nominal do motor. Esta ação foi tomada para retirar os valores de picos de torque e corrente que ocorrem na partida da correia transportadora;

- Quando o transportador estava rodando sem carga, foram identificados momentos em que a balança integradora estava medindo valores negativos para a vazão de minério. Todos os valores negativos de vazão encontrados foram substituídos pelo valor "0".
- Para a redução dos ruídos da medição da balança integradora, foi calculada a média por minuto das medições de vazão e, por consequência, da corrente, torque e velocidade do motor do transportador de correia.

3.2 Critérios para Avaliação

Para avaliar o desempenho dos *soft sensors* durante a fase de treinamento, foi utilizado o método de validação cruzada *k-fold* (*k-fold cross-validation*). Neste método, o conjunto de dados é dividido em *k* partes iguais (ou quase iguais), e então *k-1* partes são utilizadas para o treinamento e 1 parte é utilizada para o teste. O processo se repete por *k* rodadas, onde em cada rodada uma parte diferente é utilizada para teste. A Figura 3.4 mostra um exemplo de *k-fold cross-validation*, com *k* igual a 3. De acordo com Refaeilzadeh *et al.* (2016), em aprendizado de máquina, o mais comum é utilizar o *10-fold cross-validation* ($k = 10$), que é o valor que será adotado nesta dissertação. O desempenho do algoritmo pode ser calculado a partir da média do desempenho das rodadas de testes (CARVALHO *et al.*, 2011).

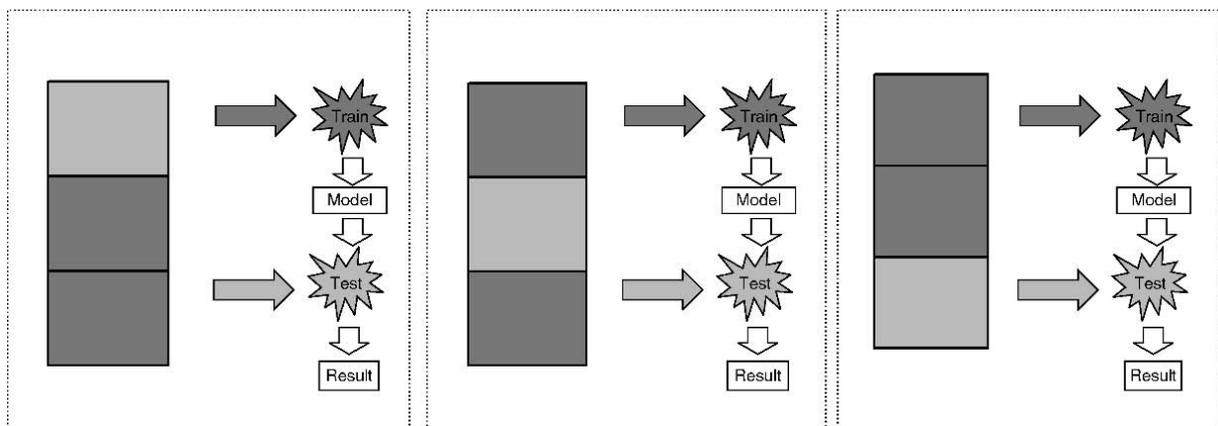


Figura 3.4: Exemplo de *3-fold cross-validation*, os conjuntos escuros são utilizados para o treinamento e o conjunto claro é utilizado para teste. Fonte: Refaeilzadeh *et al.* (2016).

Na fase de testes dos *soft sensors*, foram escolhidas métricas que são amplamente utilizadas para comparar para avaliar o desempenho de algoritmos: o Erro Médio Absoluto (*Mean Absolute Error*) (MAE), a Raiz do Erro Quadrático Médio (*Root Mean Square Error*)

(RMSE) e o Coeficiente de Correlação. O MAE pode ser definido como a média do módulo dos erros, podendo ser calculado com base na equação 3.2 (WILLMOTT & MATSUURA, 2005; WITTEN *et al.*, 2011). Segundo Willmott e Matsuura (2005), o MAE é uma medida natural para a magnitude do erro e pode ser utilizada como métrica de comparação entre diferentes modelos de predição.

$$MAE = \frac{1}{n} \sum_{i=1}^n |e_i| \quad (3.2)$$

Conforme Chai e Draxler (2014) destacam em seu trabalho, o RMSE, que pode ser calculado com base na Equação 3.3, também deve ser levado em consideração na avaliação de modelos de predição. O MAE dá o mesmo peso para todos os erros independente do valor absoluto, enquanto o RMSE penaliza a maior variância dos erros, dando maior peso aos erros com maiores valores absolutos (CHAI & DRAXLER, 2014).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2} \quad (3.3)$$

De acordo com Witten *et al.* (2011), o coeficiente de correlação representa a correlação estatística entre os valores preditos e os atuais, podendo ser calculado pela Equação 3.4. Resultados que fossem perfeitamente correlatados obteriam um coeficiente de correlação igual a 1, enquanto resultados que não possuem nenhuma correlação obteriam um coeficiente igual a zero.

$$\text{Coeficiente de Correlação} = \frac{S_{PA}}{\sqrt{S_P S_A}} \quad (3.4)$$

Onde,

$$S_{PA} = \frac{\sum_{i=1}^n (p_i - \bar{p})(a_i - \bar{a})}{n - 1}$$

$$S_P = \frac{\sum_{i=1}^n (p_i - \bar{p})^2}{n - 1}$$

$$S_A = \frac{\sum_{i=1}^n (a_i - \bar{a})^2}{n - 1}$$

Onde,

a_i é i-ésimo o valor atual da variável;

\bar{a} é a média dos valores atuais;

p_i é i-ésimo o valor predito da variável;

\bar{p} é a média dos valores preditos;

i é a instância atual;

n é o número total de instâncias dos dados do conjunto.

3.3 Weka

Para implementar os algoritmos de aprendizado de máquina, foi utilizado o software Weka¹ (*Waikato Environment for Knowledge Analysis*). Desenvolvido pela Universidade de Waikato, na Nova Zelândia, ele conta com diversas implementações de algoritmos de aprendizado de máquina, além de várias ferramentas para o pré-processamento dos dados (WITTEN *et al.*, 2011). Desde seu lançamento, em 1996, ele vem sendo amplamente utilizado para a análise de dados tanto na academia quanto na indústria (HALL *et al.*, 2009). A versão utilizada neste trabalho é a 3.8.5 e foi instalada no Windows 10 64 bits, rodando em uma máquina com as seguintes especificações: processador Intel Core i5-8300H, 8 GB de memória RAM, placa de vídeo GeForce GTX 1050 4GB e um SSD Samsung 970 EVO Plus de 250 GB.

A Figura 3.5 mostra a tela principal do Weka *Explorer*. Através dela podemos acessar todas as funcionalidades disponíveis no Weka através das abas. Na aba *Preprocess*, estão disponíveis os filtros que podem ser aplicados para o pré-processamento dos dados.

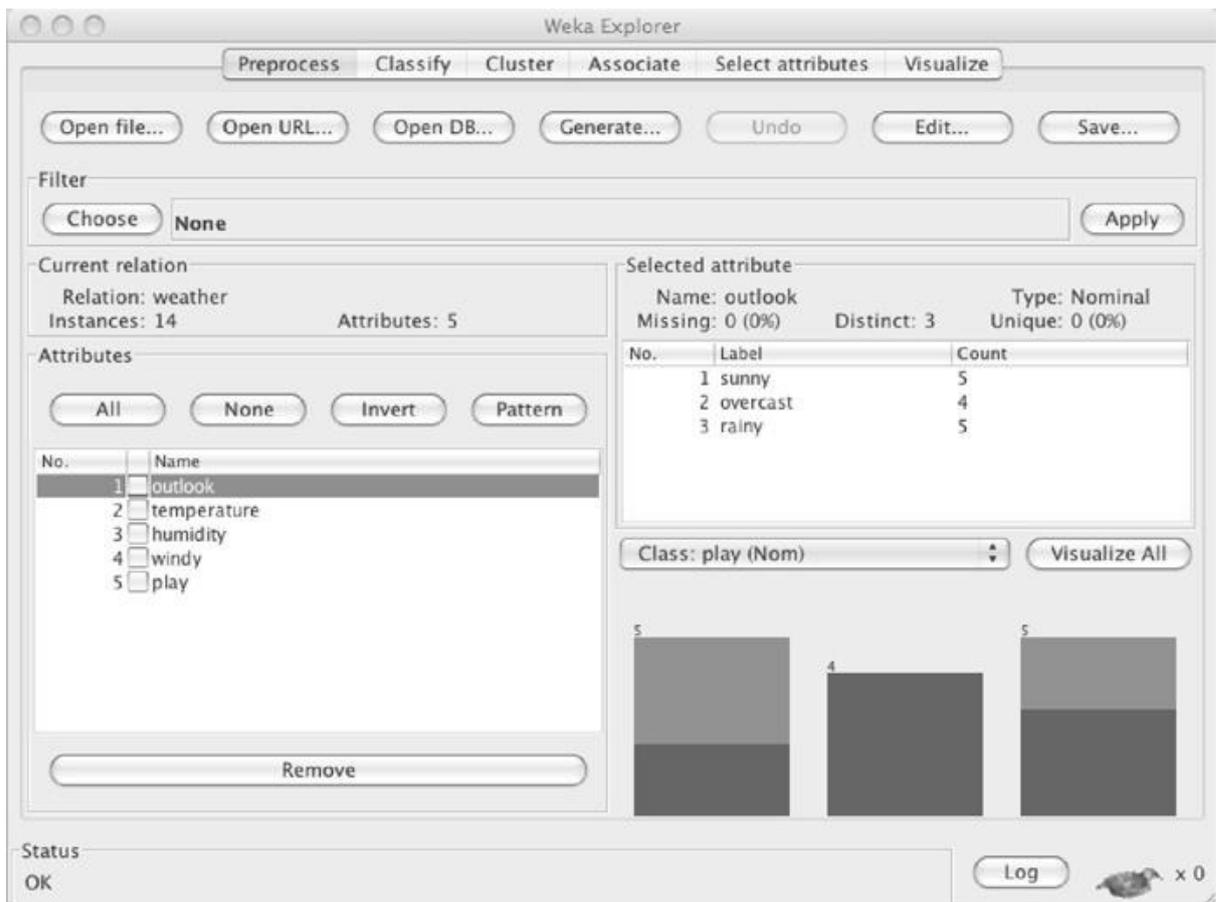


Figura 3.5: Interface do Weka: (a) tela inicial do Weka e (b) página principal do do Weka *Explorer*. Fonte: Witten *et al.* (2011).

¹<https://www.cs.waikato.ac.nz/ml/weka/>

Na aba *Classify*, todos os algoritmos implementados no Weka estão disponíveis para a seleção, permitindo que eles sejam facilmente aplicados ao conjunto de dados. Na Figura 3.6 (a), são mostrados alguns dos algoritmos disponíveis para a seleção. Enquanto na Figura 3.6 (b) é mostrado um exemplo do resultado obtido ao se aplicar um algoritmo a um conjunto de dados.

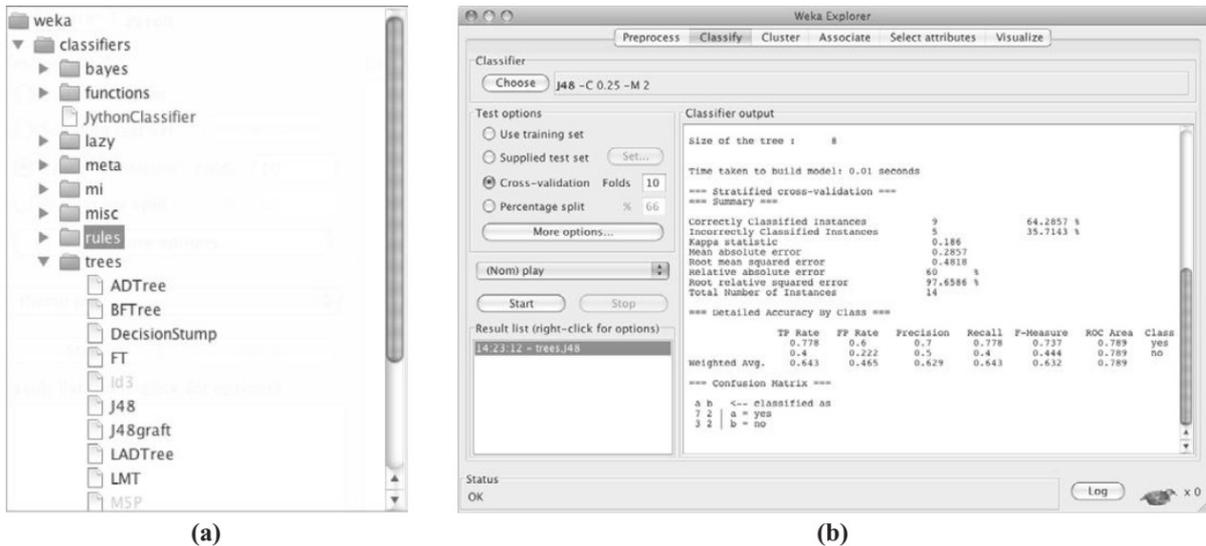


Figura 3.6: Interface do Weka: (a) lista para seleção do algoritmo a ser utilizado e (b) resultados obtidos com a implementação de um algoritmo. Fonte: Witten *et al.* (2011).

A Figura 3.7 (a) mostra as configurações que foram utilizadas na implementação do *Random Forest*, a 3.7 (b) mostra as configurações utilizadas no REPTree. As configurações utilizadas no M5P estão na Figura 3.8.

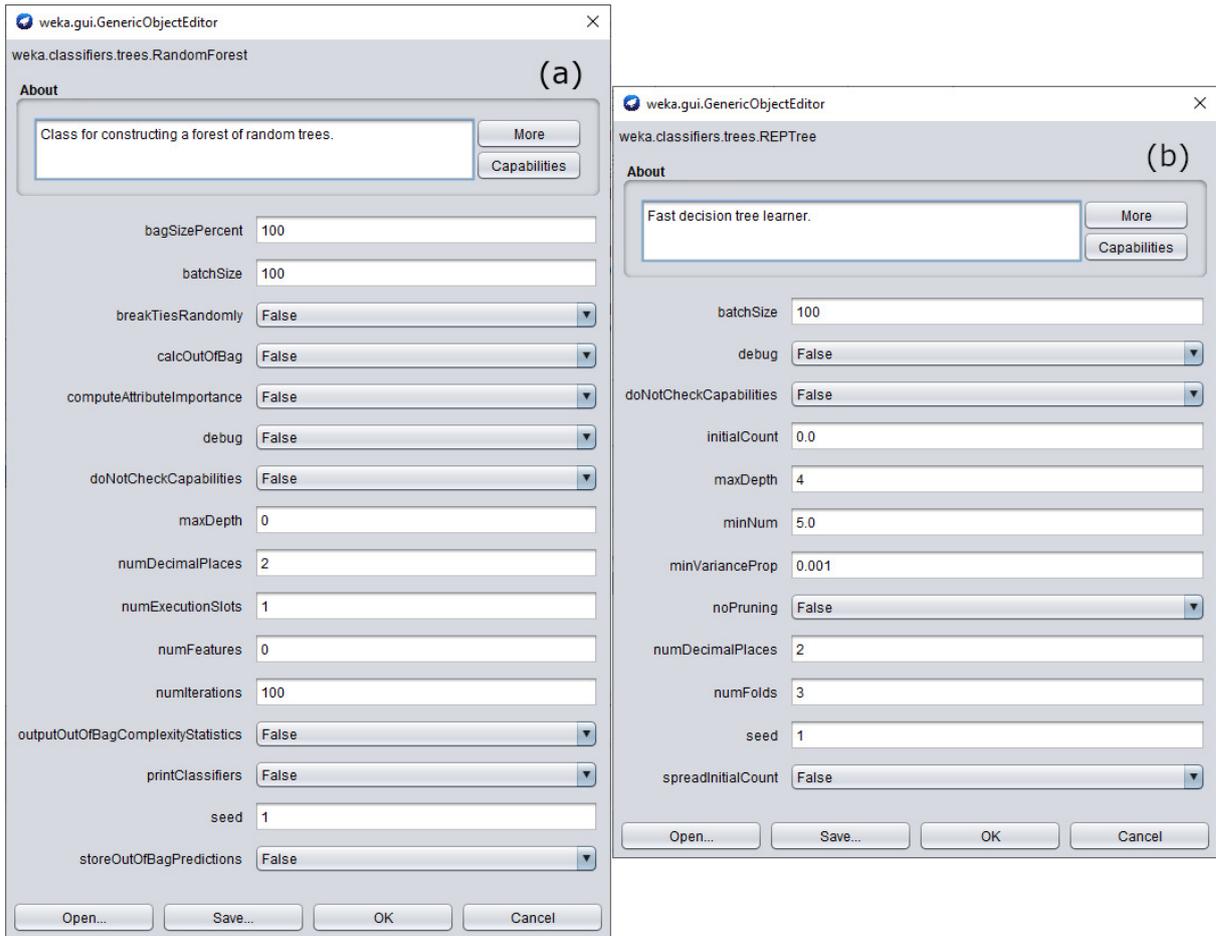


Figura 3.7: (a) Configurações utilizadas no *Random Forest* (soft sensors 1 a 4). (b) Configurações utilizadas no REPTree. Fonte: Captura de tela realizada pelo autor.

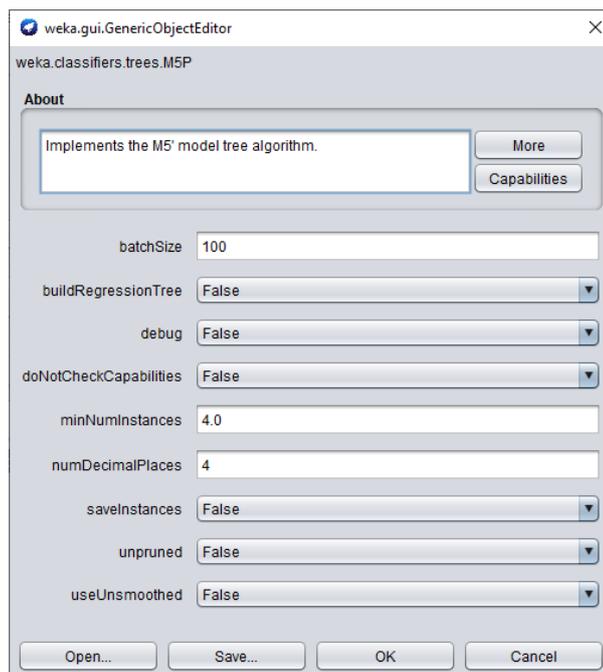


Figura 3.8: Configurações utilizadas no M5P. Fonte: Captura de tela realizada pelo autor.

4. Resultados e Discussão

4.1 *Random Forest*

Para verificar a importância dos atributos coletados (corrente, torque e velocidade), o primeiro passo foi calcular a correlação entre cada atributo e a balança integradora. Diferentemente da Seção 3.1, onde era necessário investigar o deslocamento entre os dados do motor e da balança integradora, aqui o deslocamento já era conhecido: as variações da balança estão 9 minutos e 10 segundos atrasadas em relação as variações dos dados do motor. Por este motivo, para analisar a correlação entre as variáveis do conjunto de dados do treinamento, foi utilizado o algoritmo *CorrelationAttributeEval* do Weka. Os coeficientes de correlação obtidos foram:

- Corrente: 0,71;
- Torque: 0,837;
- Velocidade 0,296.

Dentre os atributos coletados, o torque foi o que apresentou a maior correlação com a balança integradora, 0,837. O segundo maior fator de foi o da corrente, com 0,71. A velocidade apresentou o menor fator, de apenas 0,296. Deste modo, para verificar como cada atributo influencia nos resultados obtidos, foi estabelecido o teste com quatro combinações de parâmetros:

- *Soft Sensor 1*: utilizando apenas a corrente;
- *Soft Sensor 2*: utilizando apenas o torque;
- *Soft Sensor 3*: utilizando a corrente e o torque;
- *Soft Sensor 4*: utilizando a corrente, o torque e a velocidade;

Os quatro *soft sensors* foram desenvolvidos utilizando o *Random Forest* e a mesma base de dados, a única diferença é que os atributos não utilizados pelo *soft sensor* foram removidos. Conforme explicado na seção 3.1, todas as comparações dos atributos do motor foram feitas com os valores de vazão da balança integradora de 9 minutos e 10 segundos depois. Como a velocidade apresentou um coeficiente de correlação baixo quando comparados aos outros atributos, optou-se por não desenvolver um *soft sensor* com base apenas nela.

4.1.1 Treinamento dos *Soft Sensors 1 a 4*

A Tabela 4.1 apresenta os valores do coeficiente de correlação, o MAE e do RMSE obtidos na etapa de treinamento dos quatro *soft sensors*. O SS1 (corrente) foi o que apresentou os piores resultados, com coeficiente de correlação igual a 0,8930, MAE de 154,49 t/h e RMSE

de 221,95 t/h. O SS2 (torque) apresentou resultados próximos ao do SS1: 0,9000, 146,52 t/h e 214,55 t/h para o coeficiente de correlação, MAE e RMSE, respectivamente.

Os resultados do SS3 (corrente e torque) foram significativamente melhores do que os apresentados pelos *soft sensors* 1 e 2. Os valores do coeficiente de correlação, do MAE e do RMSE, foram respectivamente 0,9337, 110,26 t/h e 174,74 t/h. Por fim, o SS4 (corrente, torque e velocidade) apresentou os melhores resultados entre os quatro desenvolvidos: coeficiente de correlação igual a 0,9559, MAE igual a 87,84 t/h e RMSE de 143,03 t/h.

Tabela 4.1: Resultados da fase de treinamento dos *soft sensors* 1 a 4. Fonte: O autor.

	SS1	SS2	SS3	SS4
Coeficiente de Correlação	0,8930	0,9000	0,9337	0,9559
MAE (t/h)	154,49	146,52	110,26	87,84
RMSE (t/h)	221,95	214,55	174,78	143,03

O comportamento das estimativas dos *soft sensors* em comparação com as medições realizadas pela balança integradora é mostrado na Figura 4.1. Como pode ser visto, na etapa de treinamento, todos os quatro *soft sensors* conseguem acompanhar de maneira aproximada o comportamento da balança integradora.

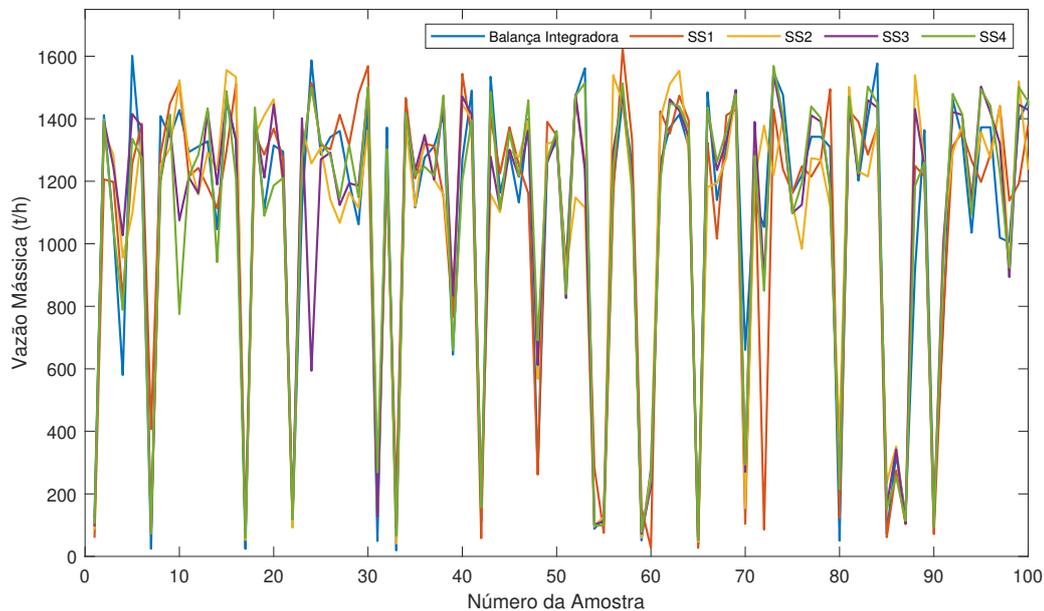


Figura 4.1: Medições da balança integradora e estimativas realizadas pelos *soft sensors* 1 a 4. Fonte: O autor.

4.1.2 Teste dos *Soft Sensors* 1 a 4

Após a fase de treinamento, todos os *soft sensors* foram submetidos a um novo conjunto de dados para testar o seu desempenho em testes. O novo conjunto de dados é composto de 7116 elementos de cada atributo e foi obtido entre os dias 19 e 25 de agosto de 2021. Durante a etapa de testes, diferentemente do que foi feito na etapa de treinamento, não foi feita a média por minuto dos dados. A razão para esta decisão é verificar como os *soft sensors* se comportam para estimar a vazão mássica registrada pela balança em tempo real. Para este novo conjunto de dados, a Balança Virtual Existente (BVE) apresentou um coeficiente de correlação de 0,9167 com MAE igual a 123,25 t/h e RMSE de 183,90 t/h.

A Tabela 4.2, mostra o coeficiente de correlação, o MAE e o RMSE das estimativas dos *soft sensors* e da balança virtual existente utilizando a nova base de dados. Como esperado, todos os quatro *soft sensors* apresentaram desempenhos mais baixos quando comparados aos obtidos durante a fase de treinamento. No SS1, o coeficiente de correlação caiu de 0,8930 para 0,8352 enquanto o MAE aumentou de 154,49 para 190,01 t/h e o RMSE cresceu de 221,95 para 236,09 t/h. O SS2, por sua vez, viu o coeficiente de correlação ser reduzido de 0,9000 na fase de treinamento, para 0,8502 na fase de testes, enquanto o MAE e o RMSE aumentaram de 146,52 t/h para 177,85 t/h e de 214,55 t/h para 224,15, respectivamente. Nos testes do SS3, o coeficiente de correlação foi de 0,8849, o MAE foi de 153,266 t/h e o RMSE foi de 195,97 t/h, valores maiores do que os obtidos durante o treinamento onde eles foram de 0,9337, 110,26 t/h e 174,78 t/h, respectivamente. o SS4 obteve valores de coeficiente de correlação iguais a 0,9559 e 0,9286, de MAE iguais a 87,84 t/h e 114,30 t/h e RMSE iguais a 143,03 t/h e 150,13 t/h nas fases de treinamento e teste, respectivamente.

Tabela 4.2: Resultados da fase de teste dos *soft sensors* 1 a 4 e da balança virtual existente. Fonte: O autor.

	SS1	SS2	SS3	SS4	BVE
Coeficiente de Correlação	0,8352	0,8502	0,8849	0,9286	0,9167
MAE (t/h)	190,02	177,85	153,27	114,30	123,25
RMSE (t/h)	236,09	224,15	195,97	150,13	183,90

Assim como na fase de treinamento, durante os testes o *soft sensor* 4, desenvolvido utilizando uma combinação da corrente, do torque e da velocidade, apresentou os melhores resultados. Dentre os quatro *soft sensors* desenvolvidos, apenas o SS4 apresentou resultados melhores que a balança virtual existente. Enquanto na BVE o erro médio absoluto foi de 123,25 t/h, no SS4 ele foi de 114,30. Esta redução no erro de 8,95 t/h representa uma redução de aproximadamente 7,26%. Na Figura 4.2 são mostradas 150 predições da fase de testes de cada um dos *soft sensors* em comparação com a balança integradora. Nela, é possível perceber que o comportamento da curva dos *soft sensors* já não é tão próxima da curva da balança integradora.

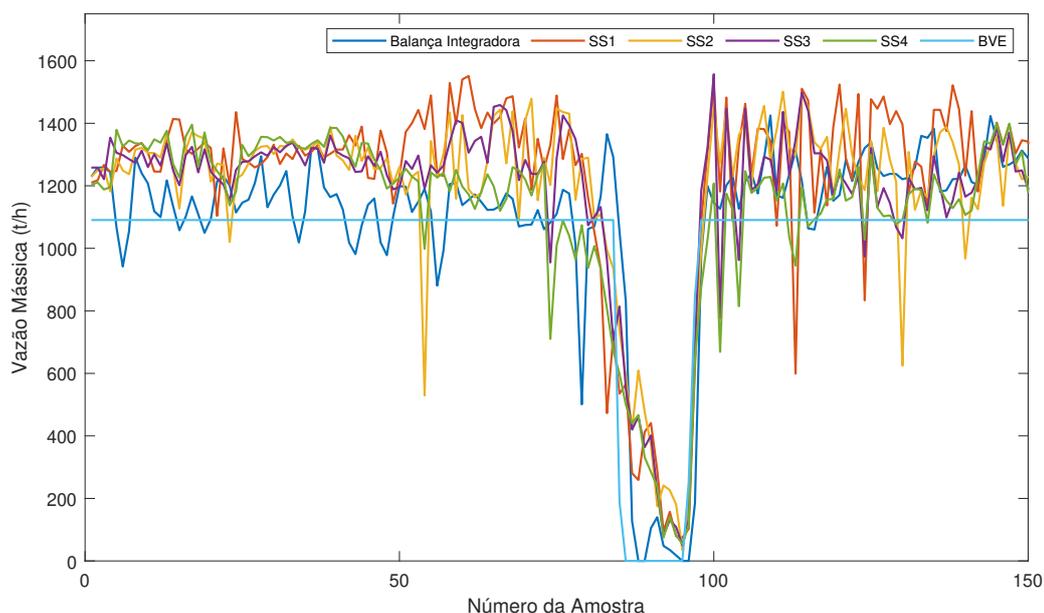


Figura 4.2: Comparativo entre os *soft sensors* desenvolvidos e a balança integradora. Fonte: O autor.

4.1.3 *Soft Sensor 5: Corrente, Torque e Velocidade*

O *soft sensor 4* foi o que apresentou os melhores resultados tanto na fase de testes quanto do treinamento. Mesmo assim, quando utilizado com o novo conjunto de dados para teste, o erro médio absoluto aumentou em mais de 30%, saltando de 87,84 t/h para 114,30 t/h, se mostrando apenas 7,26% melhor do que a balança virtual existente. Estes fatos motivaram a busca por novos ajustes nos parâmetros utilizados no *Random Forest* durante o treinamento para que se pudesse produzir um *soft sensor* que apresentasse maior robustez na predição dos novos dados.

Ao realizar alterações nos parâmetros de configuração do Weka, apenas dois deles apresentaram uma melhora significativa nos resultados: *bagSizePercent* e *maxDepth*. No SS4, o valor utilizado para o *bagSizePercent* foi de 100 e o *maxDepth* foi 0 (ilimitado). No SS5 por sua vez, os valores foram de 35 para o *bagSizePercent* e 5 para o *maxDepth*. Alterações nos valores de *batchSize* não resultaram em modificações significativas, enquanto a redução no *numIterations* resultaram em aumento nas taxas de erro. A Figura 4.3 mostra a configuração de todos parâmetros utilizados no SS5.

bagSizePercent	35
batchSize	100
breakTiesRandomly	False
calcOutOfBag	False
computeAttributeImportance	False
debug	False
doNotCheckCapabilities	False
maxDepth	5
numDecimalPlaces	2
numExecutionSlots	1
numFeatures	0
numIterations	100
outputOutOfBagComplexityStatistics	False
printClassifiers	False
seed	1
storeOutOfBagPredictions	False

Figura 4.3: Parâmetros utilizados no desenvolvimento do *soft sensor* 5. Fonte: Captura de tela realizada pelo autor.

O treinamento do SS5 foi feito utilizando o mesmo conjunto de dados que foi utilizado para o treinamento dos outros *soft sensors* e ele também foi avaliado utilizando a validação cruzada *k-fold*, com *k* igual a 10. Na etapa de treinamento, o SS5 obteve os seguintes resultados mostrados na Tabela 4.3. Em comparação com o SS4, o SS5 apresentou reduções de 4,26% e 2,53% no valores do MAE e RMSE, respectivamente, além de um pequeno aumento, 0,23%, no coeficiente de correlação.

Tabela 4.3: Resultados da fase de treinamento dos *soft sensors* 4 e 5. Fonte: O autor.

	SS4	SS5
Coefficiente de Correlação	0,9559	0,9581
MAE (t/h)	87,84	83,58
RMSE (t/h)	143,03	139,41

Após o treinamento do SS5, o modelo gerado foi aplicado ao conjunto de dados que já foi utilizado nos testes dos outros *soft sensors*, mas ainda inédito para o SS5. Na fase de testes do SS5, coeficiente de correlação foi de 0,9488, o MAE de 93,61 t/h e o RMSE foi de 124,16

t/h. A Tabela 4.4 mostra os valores do coeficiente de correlação, do erro médio absoluto e da raiz do erro quadrático médio da fase de testes para o *soft sensor* 4, para o *soft sensor* 5 e para a balança virtual existente.

Tabela 4.4: Resultados da fase de teste dos *soft sensors* 4 e 5 e da balança virtual existente.

Fonte: O autor.

	SS4	SS5	BVE
Coeficiente de Correlação	0,9286	0,9488	0,9167
MAE (t/h)	114,30	93,61	123,25
RMSE (t/h)	150,13	124,16	183,90

Ao comparar os resultados do SS5 com os obtidos pelo SS4, percebe-se que MAE foi reduzido em mais de 18,10%, saindo de 114,30 t/h para 93,61 t/h. Quando comparado com a balança virtual existente, que apresentou um MAE de 123,25 t/h, a redução próxima de 24,05%. Além disso, em uma comparação entre os resultados obtidos nas fases de treinamento e testes do SS5, mostra que houve um aumento de 12% no MAE na fase de testes. O RMSE obtido na fase de testes, 124,16 t/h, foi 10,94% menor que o da fase de treinamento, 139,41 t/h, fato que não havia ocorrido em nenhum dos outros quatro *soft sensors* desenvolvidos com o *Random Forest*. Essa melhora no desempenho do SS5 mostra que o ajuste realizado nos parâmetros se mostrou eficaz na redução do MAE e também do RMSE. A Figura 4.4 mostra o comportamento da curva de 350 estimativas do *soft sensor* 5 e da balança virtual existente em comparação com as medições da balança integradora.

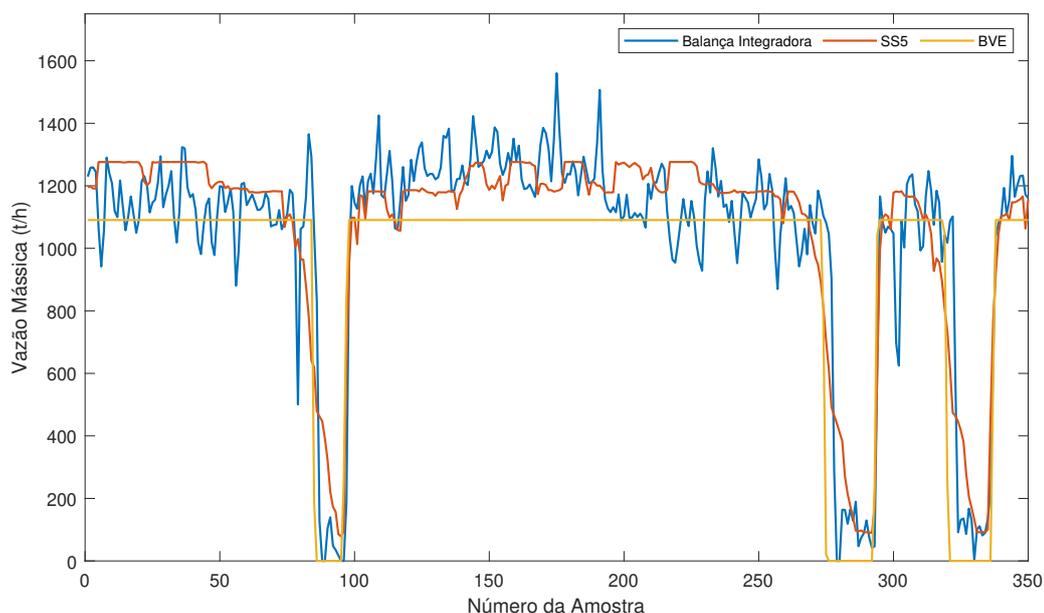


Figura 4.4: Comparativo entre a balança integradora, o *soft sensor 5* e a balança virtual existente. Fonte: O autor.

4.2 REPTree

O *soft sensor* baseado no REPTree foi desenvolvido com base na mesma base de dados utilizada para o treinamento dos *soft sensors* baseados no *Random Forest* e utilizando as configurações mostradas na Figura 3.7 (b). Na fase de treinamento, o REPTree obteve os seguintes resultados: coeficiente de correlação igual a 0,9541, MAE igual a 91,44 t/h e RMSE de 145,96 t/h, como pode ser visto na Tabela 4.5.

Tabela 4.5: Resultados das fases de treinamento e testes do REPTree e da balança virtual existente. Fonte: O autor.

	REPTree (Treinamento)	REPTree (Testes)	BVE
Coefficiente de Correlação	0,9541	0,9427	0,9167
MAE (t/h)	91,44	94,04	123,25
RMSE (t/h)	145,86	128,16	183,90

Ao submeter o modelo gerado pelo REPTree ao conjunto de testes, houve uma pequena redução do coeficiente de correlação, que foi de 0,9427. O MAE apresentou um leve aumento, atingindo 94,04 t/h na fase de testes. Assim como ocorreu no SS5 (*Random Forest*), o RMSE da

fase de testes, 128,16 t/h, foi menor do que o obtido na fase de treinamento do *soft sensor*. Em todas as três métricas avaliadas, os resultados do REPTree se mostraram melhores do que os da balança virtual existente utilizando o mesmo conjunto de dados. Na Figura 4.5, são mostradas as curvas das medições da balança integradora, das estimativas da fase de testes do REPTree e também da balança virtual existente.

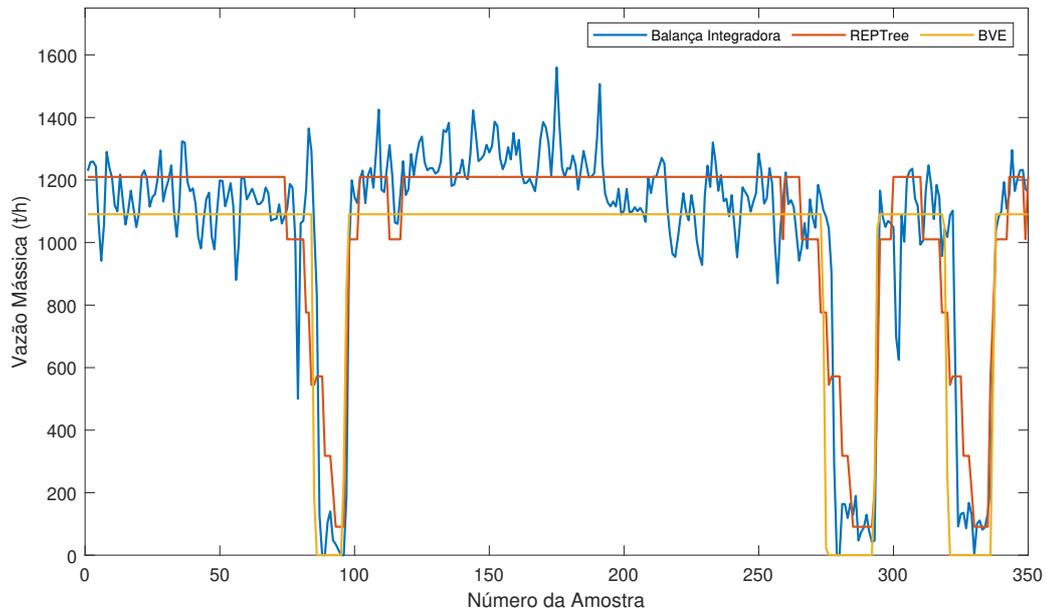


Figura 4.5: Comparativo entre a balança integradora, o REPTree (fase de testes) e a balança virtual existente. Fonte: O autor.

4.3 M5P

Gerado a partir da mesma base de dados utilizada no treinamento do *Random Forest* e do REPTree, o *soft sensor* baseado no M5P mostrou um desempenho próximo ao SS5 na fase de treinamento. Ele obteve o maior valor para o coeficiente de correlação, 0,9587, o menor valor de RMSE, 138,53 t/h, e um MAE de 96,61 t/h. No SS5, esses valores foram de 0,9581, 83,58 t/h e 139,41 t/h, respectivamente. Na tabela 4.6, são mostrados os resultados obtidos pelo M5P nas fases de treinamento e teste, além dos resultados da balança virtual existente.

Tabela 4.6: Resultados das fases de treinamento e testes do M5P (fase de testes) e da balança virtual existente. Fonte: O autor.

	M5P (Treinamento)	M5P (Testes)	BVE
Coefficiente de Correlação	0,9587	0,9553	0,9167
MAE (t/h)	83,82	84,06	123,25
RMSE (t/h)	138,53	113,17	183,90

Em uma comparação entre as fases de treinamento e testes do M5P, verifica-se que o coeficiente de correlação foi de 0,9553, valor apenas 0,35% menor que o da fase de treinamento. O aumento do MAE na fase de treinamento também foi muito discreto, apenas 0,29%, atingindo 84,06 t/h. No RMSE, houve uma redução considerável, 18,31%, saindo de 138,53 t/h no treinamento para 113,17 t/h nos testes. Quando comparado aos resultados obtidos pela balança virtual existente, houve um aumento de 4,21% no coeficiente de correlação e reduções de 31,80% e 38,46% no MAE e RMSE, respectivamente. As curvas das estimativas da fase de testes do M5P, das medições da balança integradora e das estimativas da balança virtual existente podem ser vistas na Figura 4.6.

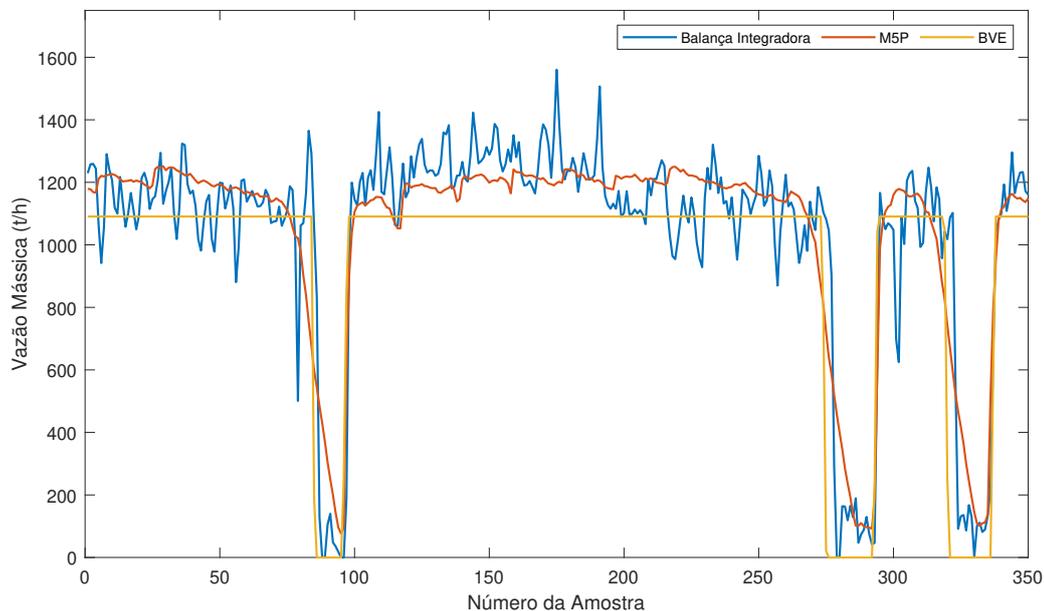


Figura 4.6: Comparativo entre a balança integradora, o M5P e a balança virtual existente. Fonte: O autor.

4.4 Comparativo Entre o M5P, o *Random Forest* e o REPTree

Com base nos dados apresentados em 4.1.3, 4.2 e 4.3, foi elaborada a Tabela 4.7 onde estão os valores do coeficiente de correlação, do MAE e do RMSE obtidos na fase de testes dos *soft sensors* desenvolvidos e também da balança virtual existente. O coeficiente de correlação de todos os *soft sensors* desenvolvidos ficaram acima de 0,94 com o M5P apresentando o maior valor, 0,9553, e o REPTree o menor, 0,9427.

Para o MAE, os valores variaram entre 84,06 t/h, do M5P, e 94,04 t/h, do REPTree. O SS5 (*Random Forest*) obteve um MAE de 93,61 t/h, valor apenas 0,46% menor do que o apresentado pelo REPTree. A situação se repete para o RMSE, onde o M5P apresentou o menor valor, 113,17 t/h, o REPTree o maior, 128,16 t/h, e o SS5 um valor próximo ao do REPTree, 124,16 t/h.

Tabela 4.7: Resultados da fase de testes do M5P, do REPTree, do SS5 (*Random Forest*) e da balança virtual existente. Fonte: O autor.

	M5P	REPTree	SS5	BVE
Coeficiente de Correlação	0,9553	0,9427	0,9488	0,9167
MAE (t/h)	84,06	94,04	93,61	123,25
RMSE (t/h)	113,17	128,16	124,16	183,90

Desta maneira, pode-se afirmar que o M5P apresentou o melhor desempenho entre os *soft sensors* propostos, já que obteve os melhores valores em todas as métricas analisadas. Os *soft sensors* baseados no *Random Forest* e no REPTree apresentaram valores muito próximos em todos os quesitos, com uma pequena vantagem para o *Random Forest*. Destaca-se ainda, o fato de que todos os *soft sensors* propostos apresentaram desempenho consideravelmente superior ao da balança virtual existente. Na Figura 4.7, são mostrados os comportamentos das medições da balança integradora, das estimativas dos *soft sensors* propostos e também da balança virtual existente.

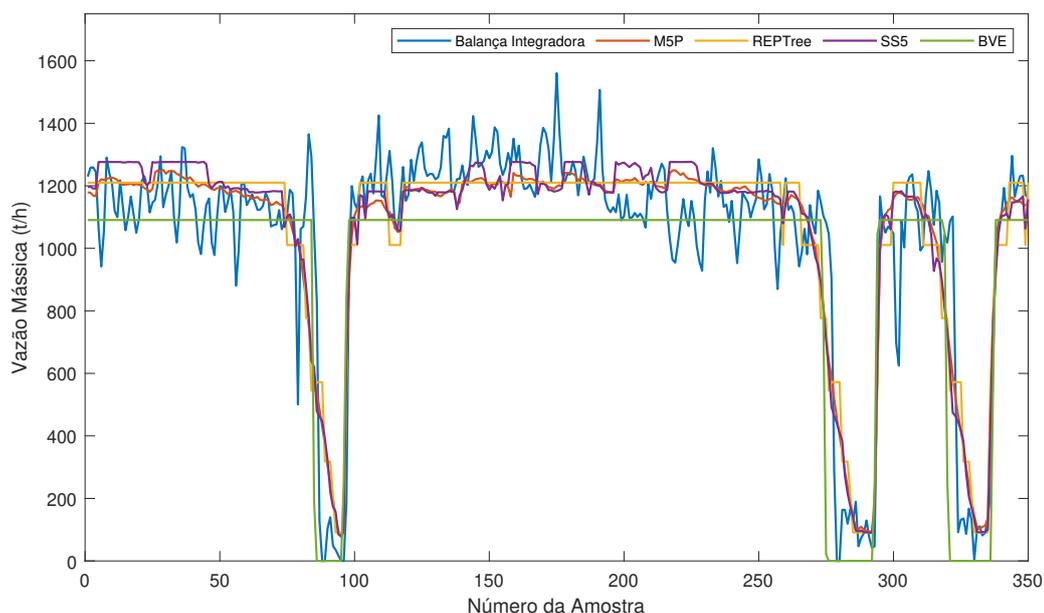


Figura 4.7: Comparativo entre a balança integradora, o M5P, do REPTree, do SS5 (*Random Forest*) e da balança virtual existente. Fonte: O autor.

4.5 Implementações no CLP

No Weka, é possível obter as árvores geradas pelos modelos obtidos na fase de treinamento dos algoritmos. A partir dessa árvores, foram gerados os códigos em Texto Estruturado (Structured Text) (ST) para que pudessem ser embarcados no CLP. O ST é uma linguagem de programação de alto nível para CLP que é definida pela norma IEC 61131-3 (JOHN & TIEGELKAMP, 2001).

Conforme demonstrado no subcapítulo 4.4, as taxas de MAE e RMSE obtidos pelo *Random Forest* foram próximas as obtidas pelo REPTree. Desta maneira, definiu-se que seriam implementados no CLP apenas os códigos gerados a partir do M5P, que obteve os melhores resultados, e também do REPTree, que apresenta uma árvore de menor complexidade quando comparada com as 100 árvores geradas pelo *Random Forest*. O código do M5P implementado no CLP é mostrado no Algoritmo 1, enquanto o do REPTree é mostrado no Algoritmo 2.

Algoritmo 1: Código gerado utilizando o M5P que foi implementado no CLP.

```
if (( $I < 1$ ) or ( $v < 5$ ) or ( $\tau < 1$ )) then
|    $\dot{m} = 0$ ;
else
|   if ( $\tau \leq 21.514$ ) then
|   |    $\dot{m} = 5.1216 \times v + 42.6358 \times \tau - 8.4641 \times I + 102.5496$ ;
|   else
|   |   if ( $\tau \leq 32.905$ ) then
|   |   |    $\dot{m} = 14.9127 \times v + 48.8824 \times \tau - 17.705 \times I + 134.8287$ ;
|   |   else
|   |   |    $\dot{m} = 16.9951 \times v + 24.5788 \times \tau - 14.5306 \times I + 512.6288$ ;
|   |   end
|   end
end
end
```

Onde,

\dot{m} é a vazão mássica instantânea de minério de ferro em toneladas por hora;

I é a corrente do motor do Hopper, em Ampères;

v é a velocidade do motor do transportador, em percentagem da velocidade nominal do motor;

τ é o torque do motor do transportador, em percentagem do torque nominal do motor.

Algoritmo 2: Código gerado utilizando o REPTree que foi implementado no CLP.

```
if (( $\tau > 1$ ) and ( $v > 5$ ) and ( $I > 1$ )) then
  if ( $\tau < 21.5$ ) then
    if ( $\tau < 11.19$ ) then
      if ( $\tau < 9.65$ ) then
        if ( $v < 2.84$ ) then
           $\dot{m} = 215.2$ ;
        else
           $\dot{m} = 90.87$ ;
        end
      else
         $\dot{m} = 209.05$ ;
      end
    else
      if ( $\tau < 15.58$ ) then
        if ( $I < 82.24$ ) then
           $\dot{m} = 317.83$ ;
        else
           $\dot{m} = 445.74$ ;
        end
      else
         $\dot{m} = 572.12$ ;
      end
    end
  else
    if ( $\tau < 29.59$ ) then
      if ( $v < 69.21$ ) then
        if ( $\tau < 24.15$ ) then
           $\dot{m} = 544.79$ ;
        else
           $\dot{m} = 776.51$ ;
        end
      else
        if ( $\tau < 27.21$ ) then
           $\dot{m} = 938.62$ ;
        else
           $\dot{m} = 1104.04$ ;
        end
      end
    else
      if ( $v < 80$ ) then
        if ( $\tau < 45.39$ ) then
           $\dot{m} = 1010.52$ ;
        else
           $\dot{m} = 1210.05$ ;
        end
      else
        if ( $\tau < 37.19$ ) then
           $\dot{m} = 1263.03$ ;
        else
           $\dot{m} = 1454.53$ ;
        end
      end
    end
  end
else
   $\dot{m} = 0$ ;
end
```

Onde,

\dot{m} é a vazão mássica instantânea de minério de ferro em toneladas por hora;

I é a corrente do motor do Hopper, em Ampères;

v é a velocidade do motor do transportador, em percentagem da velocidade nominal do motor;

τ é o torque do motor do transportador, em percentagem do torque nominal do motor.

4.5.1 Comparativo Entre as Estimativas em Tempo Real

Para verificar o desempenho dos algoritmos implementados, foi realizada uma coleta de dados de aproximadamente 2 horas e 18 minutos no dia 09/10/2021, entre 00:12 e 02:30, totalizando 2395 amostras. Conforme explicado no subcapítulo 3.1, as coletas de dados só podem ser realizadas nos momentos em que o britador principal do sistema não está operando, para que a vazão medida pela balança integradora seja referente ao minério que é transportado pelo *Hopper*. A partir destas amostras, foram calculados os valores do coeficiente de correlação, do MAE e do RMSE mostrados na Tabela 4.8.

Tabela 4.8: Resultados do M5P, do REPTree e da balança virtual existente coletados do sistema de controle da planta. Fonte: O autor.

	M5P	REPTree	BVE
Coefficiente de Correlação	0,9335	0,9264	0,9060
MAE (t/h)	125,48	132,11	124,18
RMSE (t/h)	181,13	190,93	223,37

Tanto o *soft sensor* baseado no M5P quanto o baseado no REPTree apresentaram coeficientes de correlação maiores em comparação com a balança virtual existente, indicando que as estimativas feitas pelos soft sensors apresentam maior correlação com as medições feitas pela balança integradora. A BVE obteve o menor valor de MAE, com um valor muito próximo ao do M5P que apresentou o segundo melhor resultado. Os valores de RMSE dos dois soft sensors desenvolvidos foram menores do que o da BVE, indicando que há uma maior variância nos erros da BVE.

A Figura 4.8 apresenta o comportamento das curvas de 350 amostras de vazão instantânea das estimativas do M5P, do REPTree e da balança virtual existente em comparação com as medições da balança integradora. É possível perceber que tanto o M5P quanto o REPTree conseguem reproduzir de forma aproximada o comportamento da balança integradora.

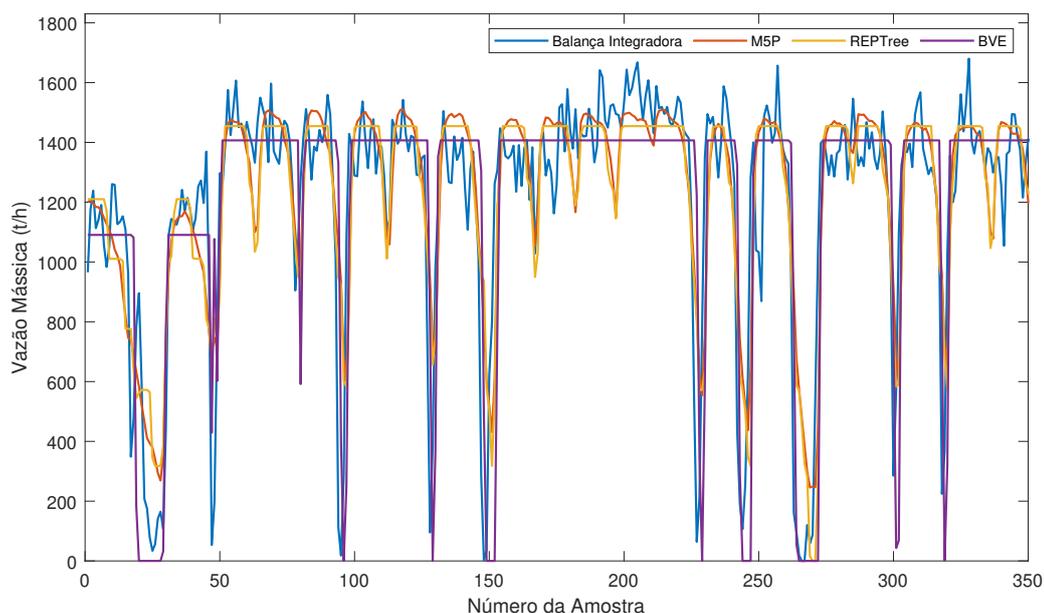


Figura 4.8: Vazão instantânea de minério obtida do sistema de controle da planta. Fonte: O autor.

Ao calcular a média de todas as amostras coletadas, a vazão média de minério da balança integradora ficou em 1.027,60 t/h. Como pode ser visto na Tabela 4.9, o REPTree obteve o resultado mais próximo da balança integradora, com uma vazão média de 1.036,43 t/h, uma diferença de apenas 0,86%. O MSP, por sua vez, registrou uma vazão média de 1.046,08 t/h, valor 1,80% maior que o registrado pela balança integradora. Com uma vazão média de 992,04 t/h, balança virtual existente apresentou a maior diferença para a balança integradora, 35,57 t/h, ou 3,46%.

Tabela 4.9: Vazão média de minério. Fonte: O autor.

	Balança Integradora	MSP	REPTree	BVE
Vazão média (t/h)	1.027,60	1.046,08	1.036,43	992,04
Diferença para a balança int. (t/h)	-	18,48	8,82	-35,57
Diferença para a balança int. (%)	-	1,80%	0,86%	-3,46%

A contabilização da produção do *Hopper* utilizando os sensores virtuais desenvolvidos se mostrou bem próxima a contabilização realizada pela balança integradora. Durante o período analisado a produção total registrada pela a balança integradora foi de 2.416 t de minério de ferro. O MSP contabilizou 2.455 t, 39 t a mais que a balança integradora, o que representa uma diferença de 1,61%. A produção aferida pelo REPTree foi ainda mais próxima da balança

integradora, marcando 2.433 t, apenas 17 t, ou 0,70%, maior. A balança virtual existente totalizou a menor produção entre todas as balanças, contabilizando 2351 t, valor 65 t, ou 2,65%, menor que a balança integradora. A Figura 4.9 mostra a totalização da produção no período analisado.

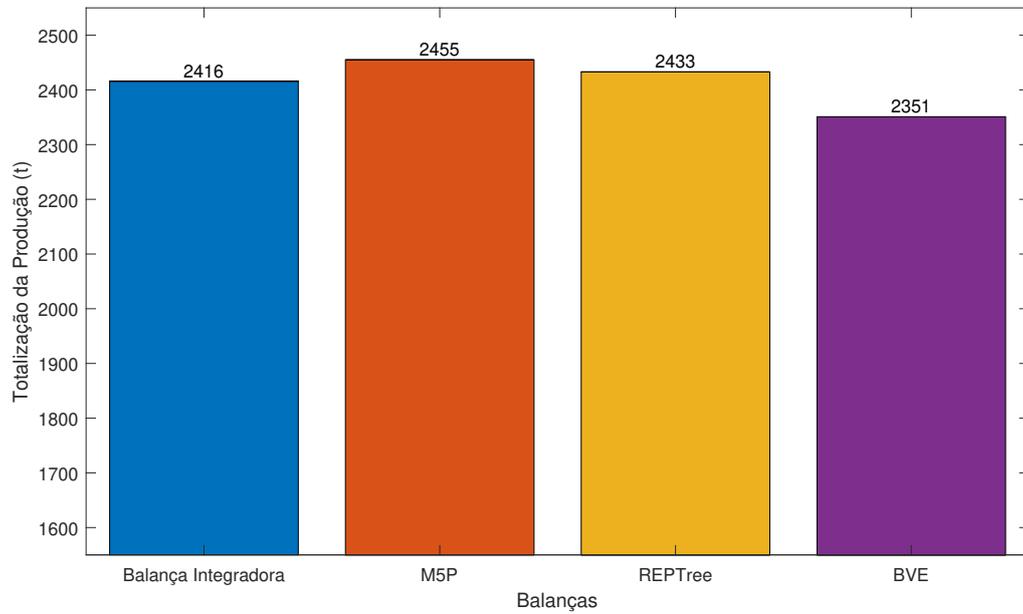


Figura 4.9: Produção contabilizada pela balança integradora, pelo M5P, pelo REPTree e pela balança virtual existente. Fonte: O autor.

5. Conclusão

Balanças integradoras para medição de vazão de minério de ferro em correias transportadoras são equipamentos com alto custo de aquisição, que demandam mão de obra especializada para manutenção e, como todos os sensores físicos, estão sujeitos a falhas, tais como o rompimento de cabos, mal contato em conexões, entre outras.

Este trabalho demonstrou o desenvolvimento de sensores virtuais, utilizando técnicas de aprendizado de máquina, que pudessem estimar a vazão mássica de minério em um transportador de correia que não é equipado com uma balança integradora. Com relação aos dados utilizados como entrada dos sensores virtuais, o sensor que utilizou uma combinação da corrente, do torque e da velocidade do motor do transportador apresentou o melhor desempenho quando comparado aos que não utilizaram os três parâmetros como entrada.

Nas simulações realizadas utilizando o Weka, os sensores virtuais apresentaram maiores valores para o coeficiente de correlação e menores taxas de MAE e RMSE quando comparados a balança virtual existente, indicando que seria possível estimar de maneira aproximada a vazão instantânea de minério de ferro transportado por uma correia transportadora.

Na última etapa no desenvolvimento desta dissertação, os códigos gerados a partir do M5P e no REPTree, foram embarcados em um CLP do sistema de controle do complexo S11D, para que fosse possível avaliar na prática o comportamento dos *soft sensors* desenvolvidos. Nesses testes, os sensores virtuais embarcados estimaram vazões médias ligeiramente maiores às registradas pela balança integradora, com o M5P estimando uma vazão média 1,80% maior e o REPTree 0,86% maior. Durante o período observado, os valores de totalização da produção através das estimativas dos sensores embarcados apresentaram uma diferença, em comparação com a totalização da balança integradora, de 0,70% no caso do REPTree e de 1,61% utilizando o M5P. Desta maneira, pode-se afirmar que os *soft sensors* desenvolvidos mostraram-se eficazes para estimar a vazão de minério em um transportador de correia que não é equipado com uma balança integradora.

6. Trabalhos Futuros

As próximas etapas deste trabalho consistem em:

- Investigar, através de novos experimentos, qual será o comportamento do *soft sensor* ao se utilizar como base a potência elétrica consumida pelo motor do transportador;
- Desenvolver novos *soft sensors* para medição da vazão de minério em outros transportadores de correia que não são equipados com balanças integradoras.

Referências Bibliográficas

- ADUSUMILLI, S., BHATT, D., WANG, H., BHATTACHARYA, P., DEVABHAKTUNI, V. “A low-cost INS/GPS integration methodology based on random forest regression”, **Expert Systems with Applications**, v. 40, n. 11, pp. 4653–4659, 2013.
- AGÊNCIA NACIONAL DE MINERAÇÃO. **Anuário Mineral Brasileiro: principais substâncias metálicas**. Brasília, ANM, 2020.
- AHMAD, M. W., MOURSHED, M., REZGUI, Y. “Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption”, **Energy and Buildings**, v. 147, pp. 77–89, 2017.
- BEHNOOD, A., BEHNOOD, V., GHAREHVERAN, M. M., ALYAMAC, K. E. “Prediction of the compressive strength of normal and high-performance concretes using M5P model tree algorithm”, **Construction and Building Materials**, v. 142, pp. 199–207, 2017.
- BIAU, G., SCORNET, E. “A random forest guided tour”, **Test**, v. 25, n. 2, pp. 197–227, 2016.
- BREIMAN, L. “Random forests”, **Machine learning**, v. 45, n. 1, pp. 5–32, 2001.
- CARVALHO, A., FACELI, K., LORENA, A., GAMA, J. **Inteligência Artificial—uma abordagem de aprendizado de máquina**. Rio de Janeiro, LTC, 2011.
- CHAI, T., DRAXLER, R. R. “Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature”, **Geoscientific model development**, v. 7, n. 3, pp. 1247–1250, 2014.
- COURONNÉ, R., PROBST, P., BOULESTEIX, A.-L. “Random forest versus logistic regression: a large-scale benchmark experiment”, **BMC bioinformatics**, v. 19, n. 1, pp. 1–14, 2018.
- CROSBY, H., DAVIS, P., JARVIS, S. A. “Spatially-intensive decision tree prediction of traffic flow across the entire UK road network”. In: **2016 IEEE/ACM 20th International Symposium on Distributed Simulation and Real Time Applications (DS-RT)**, pp. 116–119. IEEE, 2016.

- ESPOSITO, F., MALERBA, D., SEMERARO, G., KAY, J. “A comparative analysis of methods for pruning decision trees”, **IEEE transactions on pattern analysis and machine intelligence**, v. 19, n. 5, pp. 476–491, 1997.
- FAMILI, A., SHEN, W.-M., WEBER, R., SIMOUDIS, E. “Data preprocessing and intelligent data analysis”, **Intelligent data analysis**, v. 1, n. 1, pp. 3–23, 1997.
- FORTUNA, L., GRAZIANI, S., RIZZO, A., XIBILIA, M. G., OTHERS. **Soft sensors for monitoring and control of industrial processes**, v. 22. Springer, 2007.
- GARCÍA, S., LUENGO, J., HERRERA, F. **Data preprocessing in data mining**, v. 72. Springer, 2015.
- GUEDES, E. V. O. **Aplicação de soft sensor baseado em redes neurais artificiais e Random Forest para predição em tempo real do teor de ferro no concentrado da flotação de minério de ferro**. Dissertação (Mestrado), Escola de Minas, Universidade Federal de Ouro Preto, Ouro Preto, 2020.
- HALL, M., FRANK, E., HOLMES, G., PFAHRINGER, B., REUTEMANN, P., WITTEN, I. H. “The WEKA data mining software: an update”, **ACM SIGKDD explorations newsletter**, v. 11, n. 1, pp. 10–18, 2009.
- HOLMGREN, J., ASPEGREN, S., DAHLSTRÖMA, J. “Prediction of bicycle counter data using regression”, **Procedia Computer Science**, v. 113, pp. 502–507, 2017. The 8th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2017) / The 7th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH-2017) / Affiliated Workshops.
- HULTHÉN, E. **Real-time optimization of cone crushers**. Tese de Doutorado, Chalmers University of Technology, 2010.
- INGLE, V. K., PROAKIS, J. G. **Digital Signal Processing Using MATLAB**. Cengage Learning, 2011.
- JOHN, K.-H., TIEGELKAMP, M. “**The New Programming Languages of IEC 61131-3**”. In: IEC 61131-3: Programming Industrial Automation Systems: Concepts and Programming Languages, Requirements for Programming Systems, Aids to Decision-Making Tools, pp. 95–200, Berlin, Heidelberg, Springer Berlin Heidelberg, 2001. ISBN: 978-3-662-07847-1. doi: 10.1007/978-3-662-07847-1_4. Disponível em: <https://doi.org/10.1007/978-3-662-07847-1_4>.

- LIN, B., RECKE, B., KNUDSEN, J. K., JØRGENSEN, S. B. “A systematic approach for soft sensor development”, **Computers & chemical engineering**, v. 31, n. 5-6, pp. 419–425, 2007.
- LOTUFO, F. A., GARCIA, C. “Sensores virtuais ou soft sensors: Uma introdução”. In: **7th Brazilian Conference on Dynamics, Control and Applications, Presidente Prudente**, 2008.
- MOREIRA, V. D. S., EUZÉBIO, T. A. M., MIOLA, W. “Aplicação de soft sensor de baixo custo para medição de vazão mássica de minério em transportador de correia”. In: **Anais do XIII Simpósio Brasileiro de Automação Inteligente - SBAI**, pp. 1905–1910, 2017.
- OSANLOO, M., PARICHEH, M. “In-pit crushing and conveying technology in open-pit mining operations: a literature review and research agenda”, **International Journal of Mining, Reclamation and Environment**, v. 34, n. 6, pp. 430–457, 2020.
- QUINLAN, J. R. “Simplifying decision trees”, **International journal of man-machine studies**, v. 27, n. 3, pp. 221–234, 1987.
- QUINLAN, J. R. “Learning with Continuous Classes”. In: **5th Australian Joint Conference on Artificial Intelligence**, pp. 343–348, Singapore, 1992. World Scientific.
- REFAEILZADEH, P., TANG, L., LIU, H. “**Cross-Validation**”. In: *Encyclopedia of Database Systems*, pp. 677–683, New York, NY, Springer New York, 2016. ISBN: 978-1-4899-7993-3.
- RIBEIRO, B. G. C., SOUSA, W. T. D., LUZ, J. A. M. D. “Feasibility project for implementation of conveyor belts in an iron ore mine. Study case: Fabrica Mine in Minas Gerais State, Brazil”, **Rem: Revista Escola de Minas**, v. 69, pp. 79–83, 2016.
- SOUZA, M., FRIGUETTO, E., SOUZA, W., CASTRO, A. “An overview of the slope monitoring program at the Serra Sul Iron Ore Project, S11D, northern Brazil”. In: **Proceedings of the 2020 International Symposium on Slope Stability in Open Pit Mining and Civil Engineering**, pp. 165–176. Australian Centre for Geomechanics, 2020.
- THE MATHWORKS, INC. “Cross-correlation - MATLAB xcorr”. 2021. Disponível em: <<https://www.mathworks.com/help/matlab/ref/xcorr.html>>. Acesso em: 18 de jul. de 2021.
- VÄYRYNEN, T., ITÄVUO, P., VILKKO, M., JAATINEN, A., PELTONEN, M. “Mass-flow estimation in mineral-processing applications”, **IFAC Proceedings Volumes**, v. 46, n. 16, pp. 271–276, 2013.

- WANG, Y., WITTEN, I. H. “Induction of model trees for predicting continuous classes”. In: **Poster papers of the 9th European Conference on Machine Learning**. Springer, 1997.
- WILLMOTT, C. J., MATSUURA, K. “Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance”, **Climate research**, v. 30, n. 1, pp. 79–82, 2005.
- WITTEN, I. H., FRANK, E., HALL, M. A. **Data Mining: Practical Machine Learning Tools and Techniques (Third Edition)**. The Morgan Kaufmann Series in Data Management Systems. Third edition ed. Boston, Morgan Kaufmann, 2011.
- ZHAN, C., GAN, A., HADI, M. “Prediction of lane clearance time of freeway incidents using the M5P tree algorithm”, **IEEE Transactions on Intelligent Transportation Systems**, v. 12, n. 4, pp. 1549–1557, 2011.