



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 *SOFT SENSOR* BASEADO EM REDES NEURAS ARTIFICIAIS E
RANDOM FOREST PARA PREDIÇÃO EM TEMPO REAL DO TEOR DE FERRO NO
CONCENTRADO DA FLOTAÇÃO DE MINÉRIO DE FERRO**

Érick Victor de Oliveira Guedes

Ouro Preto

Minas Gerais, Brasil

2020

Érick Victor de Oliveira Guedes

**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 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. José Alberto Naves Cocota Júnior, D.Sc.

Coorientador: Prof. Agnaldo José da Rocha Reis, D.Sc.

Ouro Preto

2020

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

G924a Guedes, Érick Victor De Oliveira .
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. [manuscrito] / Érick Victor
De Oliveira Guedes. - 2020.
75 f.: il.: color., gráf., tab..

Orientador: Prof. Dr. José Alberto Naves Cocota Júnior.
Coorientador: Prof. Dr. Agnaldo José da Rocha Reis.
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. Flotação. 2. Soft sensor(software). 3. Redes neurais (Computação)
. 4. Algoritmos computacionais - Random forest. I. Cocota Júnior, José
Alberto Naves. II. Reis, Agnaldo José da Rocha. 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
DEPARTAMENTO DE ENGENHARIA CONTROLE E AUTOMACAO

**FOLHA DE APROVAÇÃO****Érick Victor de Oliveira Guedes****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**

Membros da banca

José Alberto Naves Cocota Júnior - D.Sc. - UFOP
Agnaldo José da Rocha Reis - D.Sc. - UFOP
José Aurélio Medeiros da Luz - D.Sc. - UFOP
Sílvia Grasiella Moreira Almeida - D.Sc. - IFMG

Versão final

Aprovado em 09 de setembro de 2020

De acordo

Professor (a) Orientador (a) José Alberto Naves Cocota Júnior



Documento assinado eletronicamente por **Jose Alberto Naves Cocota Junior**, **PROFESSOR DE MAGISTERIO SUPERIOR**, em 09/09/2020, às 17:48, 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 **0081108** e o código CRC **A938C124**.

Referência: Caso responda este documento, indicar expressamente o Processo nº 23109.006373/2020-33

SEI nº 0081108

R. Diogo de Vasconcelos, 122, - Bairro Pilar Ouro Preto/MG, CEP 35400-000
Telefone: 3135591533 - www.ufop.br

Agradecimentos

A conclusão de um curso de pós-graduação, como o mestrado, é mais que o fechamento de mais um ciclo de estudos. Mais do que a obtenção de mais um título. E muito mais que o final de um trabalho realizado. A conclusão do mestrado é um marco em nossa vida e em nossa pesquisa. Pode ser o final de uma pesquisa, sua continuação em outro formato ou modalidade, ou até o início de um novo projeto, mais ambicioso e desafiador; ou apenas diferente. Nesse marco, podemos dar uma olhada de relance para trás, ver o caminho traçado para chegar até aqui e que, acima de tudo, valeu a pena todo o esforço dedicado em aprender o que era necessário. A experiência de passar por esse caminho é gratificante e nos memora, a cada passo, o quão pouco conhecemos e a vastidão do que ainda podemos explorar, descobrir e nos maravilhar.

Como em toda trajetória repleta de desafios, esta não poderia deixar de contar com auxílios vindos de toda parte. Primeiramente, preciso agradecer ao meus pais, que me apoiaram e incentivaram a trilhar esse caminho desde o começo. Agradeço aos amigos e familiares que apoiaram nos momentos de dificuldades, compreenderam os momentos de ausência e comemoraram as pequenas conquistas. Em especial, agradeço ao tio Álvaro Guedes, que foi, e ainda é, um conselheiro, guia e exemplo que teve participação ímpar no curso percorrido. Agradeço ao Tiago Caixeta pelas sugestões e explicações e ao Pedro Couto pela cooperação e auxílio com detalhamento do processo. Por fim, agradeço ao Instituto Tecnológico Vale (ITV), à Escola de Minas (EM), à Universidade Federal de Ouro Preto (UFOP) e a seus membros, por proporcionarem que tantos sonhos se tornem realidade, além de fomentar o desenvolvimento da ciência e dar suporte ao crescimento do país.

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); e da Vale SA.

*“Remember you are like some
other people and like no other
person.”
(Brian Little)*

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 *SOFT SENSOR* BASEADO EM REDES NEURAIIS ARTIFICIAIS E *RANDOM FOREST* PARA PREDIÇÃO EM TEMPO REAL DO TEOR DE FERRO NO CONCENTRADO DA FLOTAÇÃO DE MINÉRIO DE FERRO

Érick Victor de Oliveira Guedes

Setembro/2020

Orientadores: José Alberto Naves Cocota Júnior

Agnaldo José da Rocha Reis

A flotação surgiu para a concentração onde propriedades classicamente utilizadas (densidade, suscetibilidade magnética, eletrocondutividade) possuem contraste insuficiente ou que apresentem faixas granulométricas muito finas, e passou a ser muito utilizada em áreas como a mineração. Este processo consiste da captação de partículas hidrofóbicas por bolhas de ar e sua flotação, fazendo a separação de tais partículas das hidrofílicas, que se mantêm em suspensão. Para permitir a separação de materiais que naturalmente não apresentam diferenças

em suas propriedades de hidrofobicidade, são utilizados reagentes para induzir as características desejadas. Como a medição do teor e obtenção da recuperação do mineral, objetivos primários do processo, acontece por meio de análise laboratorial que demanda um período de 2 horas, é proposto o desenvolvimento de um *soft sensor* para sua obtenção em tempo real. Um *soft sensor* utiliza um modelo desenvolvido para reunir diversas variáveis relevantes a um processo e inferir uma medição para as variáveis de saída desejadas. Os dados coletados de uma linha de flotação passaram por dois métodos de redução de dimensionalidade paralelamente, o *RReliefF* e o *Principal Component Analysis (PCA)*, e foram utilizados para treinamento, validação e teste dos modelos propostos. Para a modelagem do *soft sensor* foram utilizadas as técnicas de Redes Neurais Artificiais (RNA) e de *Random Forest (RF)*. Para o treinamento das RNA, foram usados os algoritmos de Levenberg-Marquardt (LM) e o *Scaled Conjugate Gradient (SCG)*. Os desempenhos dos métodos utilizados foram comparados e são apresentados neste texto.

Palavras-chave: flotação, *soft sensor*, redes neurais, *random forest*

Macrotema: Usina; **Linha de Pesquisa:** Instrumentação no Processamento de Minérios; **Tema:** Monitoramento On-line de Sites.

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.)

ARTIFICIAL NEURAL NETWORKS AND RANDOM FOREST BASED SOFT SENSOR APPLIED TO THE REAL TIME PREDICTION OF IRON ORE CONTENT IN THE IRON ORE FROTH FLOTATION CONCENTRATE

Érick Victor de Oliveira Guedes

September/2020

Advisors: José Alberto Naves Cocota Júnior

Agnaldo José da Rocha Reis

The flotation process emerged to be used where properties classically used (density, magnetic susceptibility, electroconductivity) have insufficient contrast or present very fine particle sizes, and became largely used in areas such as mining. This process consists of the capture and flotation of hydrophobic particles by air bubbles, separating these particles from the hydrophilic ones, which remain in suspension. In order to allow the separation of minerals that naturally do not exhibit differences in its hydrophobic properties, reagents are added to induce the desired characteristics. Given that the measurement of the

process' primary objectives, concentrate grade and recovery, is realized through laboratory analysis which demands a 2 hours period, this work proposes the development of a soft sensor to infer these data in real time. A soft sensor uses a model to gather several variables relevant to the process and infer a measurement to the desired output variables. Two methods of dimensionality reduction were applied in parallel to the data collected from a flotation line, the RReliefF and the principal component analysis (PCA), before the data was used for training, validation and test of the proposed models. Artificial neural networks (ANN) and Random Forests (RF) were used to model the soft sensor. For training the ANN, it was used the Levenberg-Marquardt (LM) algorithm and the Scaled Conjugate Gradient (SCG). The performance of the methods used were compared and are presented in this text.

Keywords: flotation, soft sensor, neural networks, random forest

Macrotheme: Plant; **Research Line:** Instrumentation in Ore Processing;

Theme: Online Site Monitoring.

Lista de Figuras

Figura 2.1	Circuito simplificado do beneficiamento por flotação	20
Figura 2.2	Esquema representando as partes de uma coluna de flotação	22
Figura 2.3	Esquema representando as partes de uma célula de flotação em vista explodida	23
Figura 2.4	Máquina de flotação modelo Galigher	23
Figura 2.5	Célula de flotação <i>TankCell</i> , que possui um tanque cilíndrico	24
Figura 2.6	Esquemáticos de diferentes tipos de rotores e estatores	24
Figura 2.7	Fluxograma de procedimentos para construção de um <i>soft sensor</i>	26
Figura 2.8	(a) Função de ativação degrau bipolar; (b) Função de ativação de Heavi- side ou degrau unitário	28
Figura 2.9	(a) Função de ativação linear; (b) Função de ativação logística; (c) Fun- ção de ativação tangente hiperbólica	29
Figura 2.10	Representação de uma Rede Neural Artificial do tipo <i>single-layer feed- forward</i>	30
Figura 2.11	Representação de uma Rede Neural Artificial do tipo <i>multilayer feed- forward</i>	30
Figura 2.12	Representação de uma Rede Neural Artificial recorrente	31
Figura 2.13	Representação de uma Rede Neural Artificial <i>mesh</i>	31
Figura 2.14	Representação de uma árvore de decisão	33
Figura 3.1	Representação da linha de flotação de onde os dados foram coletados	40
Figura 3.2	Média, mediana e intervalos interquartis da densidade da polpa na caixa de alimentação a cada período de duas horas	41
Figura 3.3	Etapas realizadas para redução de dimensionalidade	43
Figura 3.4	Mapa de calor das correlações entre os dados disponíveis ao final da segunda etapa de redução de variáveis	45
Figura 4.1	Peso das variáveis de entrada por ordem decrescente, onde as variáveis com peso positivo foram selecionadas e as com peso negativo foram removidas	51
Figura 4.2	Percentual de variância explicada por cada componente principal e pela somatória dos componentes	53
Figura 4.3	Representação da Rede Neural gerada	54

Figura 4.4	Resultado da regressão do treinamento com maior R no conjunto de testes da RNA com os dados do <i>RReliefF</i> usando o algoritmo de LM	56
Figura 4.5	Resultado da regressão do treinamento com menor MSE da RNA com os dados do <i>RReliefF</i> usando o algoritmo de LM	57
Figura 4.6	Resultado da regressão do treinamento com maior R no conjunto de testes da RF com os dados do PCA	59
Figura 4.7	Resultado da regressão do treinamento com menor MSE no conjunto de testes da RF com os dados do <i>RReliefF</i>	61
Figura 4.8	Teor de ferro em % no mês utilizado para treinamento obtida em laboratório e obtida pelo <i>soft sensor</i> usando RNA, com MAE de 0,6132	62
Figura 4.9	Teor de ferro em % nos meses de janeiro, fevereiro, abril e maio obtida em laboratório e obtida pelo <i>soft sensor</i> usando RNA	63
Figura 4.10	Teor de ferro em % no mês utilizado para treinamento obtida em laboratório e obtida pelo <i>soft sensor</i> usando RF, com MAE de 0,5904	64
Figura 4.11	Teor de ferro em % nos meses de janeiro, fevereiro, abril e maio obtida em laboratório e obtida pelo <i>soft sensor</i> usando RF	65

Lista de Tabelas

Tabela 3.1	Relação entre período avaliado e resultado da análise laboratorial	41
Tabela 3.2	Variáveis mantidas e removidas na segunda etapa da redução de variáveis pela identificação qualitativa de redundância	44
Tabela 3.3	Variáveis mantidas e removidas na terceira etapa da redução de variáveis pela identificação quantitativa de redundância	46
Tabela 4.1	Maiores valores de R para o conjunto de testes das RNA utilizando os algoritmos LM e SCG sobre os dados do <i>RReliefF</i> e do PCA	55
Tabela 4.2	Quantidade de resultados acima dos valores definidos para R nas RNA utilizando os algoritmos LM e SCG sobre os dados do <i>RReliefF</i> e do PCA	55
Tabela 4.3	Menores valores do MSE para RNA utilizando os algoritmos LM e SCG sobre os dados do <i>RReliefF</i> e do PCA	56
Tabela 4.4	Quantidade de resultados abaixo dos valores definidos para MSE nas RNA utilizando os algoritmos LM e SCG sobre os dados do <i>RReliefF</i> e do PCA	57
Tabela 4.5	Maiores valores de R para o conjunto de testes das RF utilizando os dados do <i>RReliefF</i> e do PCA	58
Tabela 4.6	Quantidade de resultados acima dos valores definidos para R nas RF utilizando os dados do <i>RReliefF</i> e do PCA	59
Tabela 4.7	Menores valores do MSE para RF utilizando os dados do <i>RReliefF</i> e do PCA	60
Tabela 4.8	Quantidade de resultados abaixo dos valores definidos para MSE nas RF utilizando os dados do <i>RReliefF</i> e do PCA	60

Lista de Siglas e Abreviaturas

ANN Artificial Neural Networks

EEG Eletroencefalograma

HEVC High Efficiency Video Coding

ICA Independent Component Analysis

LDA Linear Discriminant Analysis

LDPE Low-density Polyethylene

LLE Locally Linear Embedding

LM Levenberg-Marquardt

MAE Mean Absolute Error

MI Melt Index

MLP Multilayer Perceptron

MSE Mean Squared Error

NDFS Nonnegative Discriminative Feature Selection

PCA Principal Component Analysis

PCR Principal Component Regression

PIMS Plant Information Management System

PLS Partial Least Squares

RF Random Forest

RNA Redes Neurais Artificiais

SCG Scaled Conjugate Gradient

SVR Support Vector Regression

Sumário

1	INTRODUÇÃO	15
1.1	Objetivo Geral	16
1.2	Objetivos Específicos	16
1.3	Organização do Texto	16
2	REFERENCIAL TEÓRICO	18
2.1	Flotação de Minério	18
2.2	Flotação em Coluna	21
2.3	Flotação em Célula Mecânica	22
2.4	<i>Soft Sensors</i>	25
2.5	Redes Neurais Artificiais	27
2.6	<i>Random Forest</i>	32
2.7	Redução de Dimensionalidade	35
2.7.1	Extração de Características	35
2.7.2	Seleção de Características	36
3	MATERIAIS E MÉTODOS	38
3.1	Dados	38
3.2	Redução de Dimensionalidade	42
3.3	Rede Neural Artificial	46
3.3.1	Parâmetros de Treinamento da RNA	47
3.4	<i>Random Forest</i>	48
4	RESULTADOS E DISCUSSÃO	50
4.1	<i>RReliefF</i>	50
4.2	PCA	53
4.3	RNA	54
4.3.1	Coefficiente de Correlação Linear	54
4.3.2	MSE	56
4.4	<i>Random Forest</i>	58
4.4.1	Coefficiente de Correlação Linear	58
4.4.2	MSE	59

4.5	Aplicação do <i>Soft Sensor</i>	61
4.5.1	RNA	61
4.5.2	<i>Random Forest</i>	63
5	CONCLUSÕES	66
6	TRABALHOS FUTUROS	68
	REFERÊNCIAS BIBLIOGRÁFICAS	69

1. Introdução

Via de regra, para atingir os requisitos de concentração de venda do minério, os minerais precisam ser beneficiados. Uma parte deste beneficiamento envolve a separação, que pode ser por tamanho (*sizing*, em inglês) ou por espécie (*sorting*, em inglês). A liberação dos elementos constituintes do material, alcançada por meio da cominuição, deve ter um grau adequado para que a separação, mais especificamente por espécie, seja feita de forma eficiente. A moagem, muitas vezes, precisa ser realizada até que as partículas estejam pequenas o suficiente para que atinjam o grau de liberação necessário, ainda que a sobre-cominuição seja prejudicial à eficiência dos processos de separação por espécie. A flotação é então usada na separação de finos. Neste contexto, existem duas alternativas classicamente aplicáveis ao minério de ferro: uma com alto custo de implementação e baixos custos operacionais, a concentração por alto gradiente de campo magnético, e outra com investimento inicial menor e custos operacionais maiores, a concentração por flotação, abordada neste trabalho (LUZ, 2016).

Os objetivos primários do processo de flotação são a recuperação e o teor do concentrado (MOHANTY, 2009), o que pode ser estimado com o uso de equipamentos baseados na fluorescência de raios X. Tais equipamentos têm custo elevado, de forma que uma tentativa de diluir seu custo é fazer uma multiplexação de vários fluxos de processo para um equipamento de fluorescência de raios X, o que reduz sua disponibilidade para cada processo individualmente. Além disso, a precisão e confiabilidade da estimativa ainda são fortemente dependentes da qualidade da calibração e da manutenção do sistema como um todo. Dadas as dificuldades de aplicação e altos custos de implementação e manutenção, a realização do controle comumente não faz uso destas variáveis (BERGH e YIANATOS, 2003, 2011).

Visto os custos e dificuldades de observação dos objetivos primários, o controle da flotação é feito usando objetivos secundários, sendo os valores dos objetivos primários obtidos por meio de análise laboratorial. Para proporcionar sua obtenção em tempo real, podem ser desenvolvidos *soft sensors*, *softwares* que utilizam dados disponíveis para inferir uma variável não mensurada, para fazer uma estimativa com base nos parâmetros de processo atuais e nas variáveis mensuradas. A predição em tempo real é vantajosa por reduzir o tempo de resposta de sistemas especialistas, como o de dosagem de amina e o de dosagem de amido, auxiliando na atribuição dos valores de referência do processo e em seu controle. Parâmetros de entrada manual também se beneficiam da maior agilidade na obtenção de resultados, já que o operador

pode contar com a disponibilidade das informações para definir variações nos *setpoints* dos reagentes e do nível de interface de espuma, por exemplo. A rapidez no ajuste dos parâmetros de processo evita desperdícios com material fora das especificações, excessos nas dosagens de reagentes e, conseqüentemente, pode gerar uma economia no processo.

Diferentes algoritmos de aprendizado de máquina podem ser usados para o desenvolvimento de um *soft sensor*. As Redes Neurais Artificiais (RNA) são bastante utilizadas para este fim, por isso foi um dos métodos escolhidos no presente trabalho. Outro algoritmo utilizado neste trabalho foi o da *Random Forest* (RF). Os dois algoritmos selecionados são amplamente conhecidos entre pesquisadores e profissionais da área de aprendizado de máquina e serão descritos ao longo dos próximos capítulos.

1.1. Objetivo Geral

Este trabalho propõe o desenvolvimento de um *soft sensor* para predição em tempo real do teor de ferro no concentrado do minério de ferro resultante do processo de flotação.

1.2. Objetivos Específicos

Para que o objetivo geral seja alcançado, são buscados os seguintes objetivos específicos:

- Treinar e validar uma Rede Neural para predizer os resultados a partir dos dados de entrada;
- Treinar e validar uma *Random Forest* para predizer os resultados a partir dos dados de entrada;
- Propor o uso do *soft sensor* desenvolvido para a predição em tempo real na indústria.

1.3. Organização do Texto

O presente trabalho contém uma breve introdução sobre o tema, abordando seus objetivos, no Capítulo 1. O referencial teórico é feito no Capítulo 2. No Capítulo 3 são apresentados os materiais utilizados e os métodos para desenvolvimento do trabalho. Os resultados são apre-

sentados no Capítulo 4, juntamente com a discussão a seu respeito. Por fim, as conclusões são expostas no Capítulo 5 e propostas para trabalhos futuros são apresentadas no Capítulo 6.

2. Referencial Teórico

A flotação do minério pode ser feita em colunas de flotação, em células mecânicas ou células pneumáticas de flotação. O uso de *soft sensors* pode contribuir na melhoria do processo de flotação. Para propiciar um melhor entendimento em relação a estes temas, um levantamento teórico sobre eles é feito nas próximas seções.

2.1. Flotação de Minério

O processo da flotação tem como fundamento as propriedades de hidrofobicidade e hidrofiliabilidade das partículas que constituem o material tratado, também chamado de “molhabilidade” das partículas. Estas propriedades estão relacionadas à polaridade das partículas, sendo as hidrofóbicas essencialmente não-polares, de forma que não reagem com moléculas de água e apresentam maior afinidade pelo ar, enquanto as hidrofílicas são polares ou iônicas e se hidratam rapidamente. A partir destas diferenças, a flotação realiza a separação dos materiais, capturando as partículas hidrofóbicas por meio de bolhas de ar e levando-as até a camada de espuma, no topo da célula de flotação, enquanto as partículas hidrofílicas permanecem em suspensão e são removidas de acordo com o tipo de equipamento utilizado, formando o afundado (CHAVES *et al.*, 2018; FARROKHPAY, 2011; LUZ, 2016; SANTOS, 2014). Desta forma, o ponto central do processo de separação por flotação é a adesão bolha-partícula, que inclui diversos fenômenos físico-químicos e hidrodinâmicos, como tamanho de bolha, tamanho de partícula e temperatura (CAPPONI, 2005; LUZ, 2016; NGUYEN e EVANS, 2004; PAIVA, 2012).

O controle da hidrofobicidade dos materiais é feito com a adição de reagentes na preparação da polpa. De acordo com Luz (2016), a grande maioria dos minerais são naturalmente hidrofílicos, de forma que é necessária a adição de um coletor para poder realizar a flotação destes minerais. Os coletores são reagentes que induzem a hidrofobicidade ao mineral, fazendo com que ele tenha mais afinidade com o ar e possa ser flotado. Coletores são compostos orgânicos heteropolares, apresentando em sua estrutura uma parte iônica de caráter polar, que adsorve na superfície do mineral, e uma parte covalente, que dá a característica hidrofóbica ao mineral (LUZ, 2016; SANTOS, 2014).

Existem casos em que os coletores não funcionam de forma eficiente, necessitando de reagentes adicionais. Os ativadores são usados quando o coletor não adsorve em nenhum dos

minerais presentes. Eles ativam seletivamente o mineral desejado, fazendo com que ele se torne atrativo ao coletor. Já no caso em que o coletor adsorve em todas as partículas presentes, de forma a não haver seletividade, são usados depressores seletivos, que se adsorvem ao mineral que deverá compor o afundado e impedem a ação do coletor (SANTOS, 2014).

Além destes, existem também os reguladores de pH e os espumantes. Os reguladores de pH são importantes por influenciar na adsorção ou hidratação das partículas minerais; e os espumantes auxiliam no controle do tamanho de bolhas (FARROKHPAY, 2011; SANTOS, 2014).

Na flotação de minério de ferro (do tipo catiônica reversa), normalmente os coletores utilizados são as aminas, para deixar a ganga hidrofóbica, e o depressor dos minerais de ferro é o amido. As aminas são compostos orgânicos que possuem um grupo polar e uma cadeia hidrocarbônica não-polar, de forma que adsorvem os minerais naturalmente hidrofílicos deixando-os hidrofóbicos. O amido é um polímero de alta massa molecular encontrado em vários produtos de origem vegetal, como arroz, batata, milho, mandioca e trigo. Destes, o amido de milho é o mais utilizado na depressão de óxidos de ferro, ainda que amidos de mandioca apresentem maior viscosidade, indicando massa molecular maior, e teor de óleo desprezível, o que reduz a inibição da ação espumante, já que o óleo inibe a formação de espuma (MAMEDE, 2016; MAZON, 2006). Os espumantes são reagentes tensoativos que diminuem a tensão superficial da água, ajudando a manter as bolhas razoavelmente estáveis e formando a espuma. Na flotação, são utilizados como espumantes compostos não-iônicos, geralmente da classe dos álcoois ou dos éteres (PERES, 2003 *apud* MAMEDE, 2016).

A flotação pode ser realizada em células ou colunas. As células mecânicas, método convencional, são formadas por uma câmara de flotação com um sistema de rotação de alta velocidade, que cisalha o ar usualmente auto aspirado formando bolhas. Elas apresentam um regime hidrodinâmico turbulento e alta cinética de flotação, o que representa um tempo de retenção pequeno. Já a coluna de flotação, técnica patenteada por Boutin e Tremblay em 1964, tem se desenvolvido rapidamente e se tornou uma técnica amplamente aceita na indústria da mineração (REIS, 2015; ZHANG *et al.*, 2013). A coluna apresenta elevada relação altura/diâmetro, regime hidrodinâmico pouco turbulento, geração de bolhas por borbulhadores em sua base e não possui partes móveis (PAIVA, 2012; ZHANG *et al.*, 2013). Geralmente, uma razão mínima de 3:1 é usada para a relação altura/diâmetro de uma coluna, sendo 10:1 uma razão mais comumente encontrada (LUZ, 2016). Assim, os dois métodos se diferenciam, basicamente, por

sua geometria, adição de água de lavagem, ausência de agitação mecânica e sistema de geração de bolhas, sendo que a coluna apresenta uma operação mais simples, resultando em economias nos custos de operação (AQUINO *et al.*, 2010; REIS, 2015). Já nas células pneumáticas, um método mais recentemente introduzido, a polpa é bombeada a altas pressões (entre 2,5 e 3,0 bar), o que promove a sucção de ar e formação de pequenas bolhas. A colisão e captura das partículas hidrofóbicas pelas bolhas acontece, então, antes da polpa chegar na célula, reduzindo o tempo de residência requerido. Além disso, a célula pneumática gera rejeitos com baixa concentração do material de interesse e possui operação simples, já que também não possui partes móveis (LIMA *et al.*, 2018).

Dada a complexidade do processo e as especificidades das condições para a ocorrência da flotação de uma partícula, a recuperação e o teor desejados são difíceis de serem obtidos em apenas uma etapa. Por isso, é feito um circuito de beneficiamento, onde a polpa é alimentada inicialmente em uma coluna denominada *rougher*. O concentrado dessa primeira etapa ainda é pobre, enquanto o rejeito ainda contém material útil. Desta forma, o concentrado passa por uma nova flotação, numa etapa chamada de *cleaner*, gerando o concentrado final e um rejeito com alto teor, que ainda pode ser beneficiado. O rejeito do *rougher* também passa por uma nova flotação, esta chamada de *scavenger*, que gera um rejeito pobre, considerado como rejeito final, e um concentrado que não está rico o suficiente para ser considerado final. Este concentrado, assim como o rejeito da etapa *cleaner*, voltam, então, para a etapa *rougher*. Em muitos casos, no entanto, apenas uma etapa *cleaner* não é suficiente para gerar um concentrado com as características requisitadas, sendo usadas, então, etapas *cleaner* adicionais, chamadas *re-cleaner* (CHAVES *et al.*, 2018). Na Figura 2.1 tem-se a ilustração do circuito de beneficiamento com uma etapa *cleaner*.

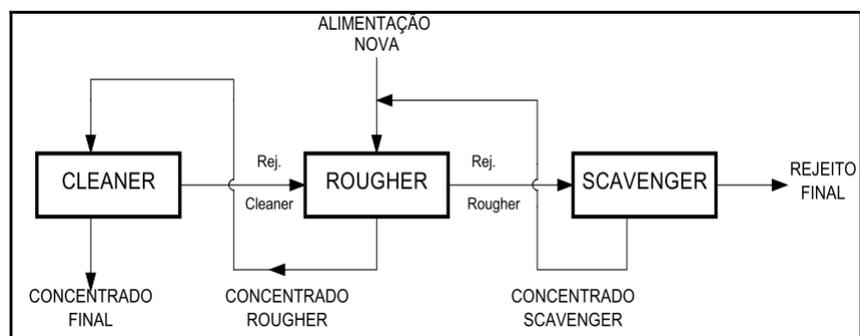


Figura 2.1: Circuito simplificado do beneficiamento por flotação
 Fonte: CHAVES *et al.*, 2018

2.2. Flotação em Coluna

Apesar de sua relativa simplicidade de operação e da ausência de partes móveis, a coluna de flotação apresenta uma complexa fluidodinâmica, o que gera grandes possibilidades de estudos e investigações para sua compreensão (REIS, 2015).

A concentração por espuma em coluna de flotação pode ser feita de diferentes formas. A flotação direta é aquela em que os materiais de interesse são flotados na espuma e a ganga, minerais que irão compor o rejeito, segue o fluxo da polpa. Na flotação reversa, a ganga é flotada e os materiais de interesse seguem o fluxo da polpa. A flotação seletiva é quando apenas uma espécie é flotada, enquanto a flotação coletiva é quando diferentes minerais, que apresentem características similares, são flotados em conjunto (CHAVES *et al.*, 2018). A flotação direta é usada em situações onde o minério tem teor muito baixo, podendo se mostrar economicamente lucrativa, além de reduzir o impacto ambiental (LUZ, 2016). No Brasil, é utilizada a flotação reversa para o beneficiamento de minério de ferro, composto majoritariamente por quartzo e hematita. Ela é usada pois a hematita, minério de interesse, é mais abundante e mais pesada, o que dificultaria sua remoção na espuma e se faria necessário ser moída até uma granulometria menor para poder ser flotada. Desta forma, a ganga, composta pelo quartzo, é flotada e a hematita é deprimida (CHAVES *et al.*, 2018).

Na Figura 2.2 tem-se a ilustração de um esquema de uma coluna de flotação apresentando seus principais componentes. Sua seção transversal pode apresentar variados formatos, por exemplo, quadrada, retangular ou circular, sendo que o diâmetro das colunas industriais varia entre 0,3m e 5,0m. Sua altura varia entre 10m e 15m, dependendo das características operacionais requeridas (AQUINO *et al.*, 2010). As partículas hidrofílicas, que constituem o não-flotado, são removidas pela base da coluna. Ainda na parte inferior da coluna, fica o aerador, responsável pela geração das bolhas que farão a coleta das partículas a serem flotadas. A alimentação de polpa é feita a aproximadamente dois terços de sua altura total, devendo ficar abaixo da interface polpa-espuma. A região compreendida entre esta interface e o aerador é chamada de zona de coleta ou de recuperação. A região compreendida desde a interface polpa-espuma até o topo da coluna, o transbordo, é chamada de zona de limpeza. Na zona de coleta as partículas hidrofóbicas são coletadas pelas bolhas de ar, em fluxo contra-corrente, e levadas até a zona de limpeza, enquanto as partículas hidrofílicas seguem o fluxo e formam o não-flotado. Na parte superior, entra a água de lavagem, por meio de um dispersor. Suas funções são, básica-

mente, a remoção de partículas hidrofílicas que estejam sendo arrastadas pelo fluxo ascendente, aumentar a altura e a estabilidade da espuma e reduzir a coalescência das bolhas (ALVES, 2015; AQUINO *et al.*, 2010; REIS, 2015; SANTOS, 2014).

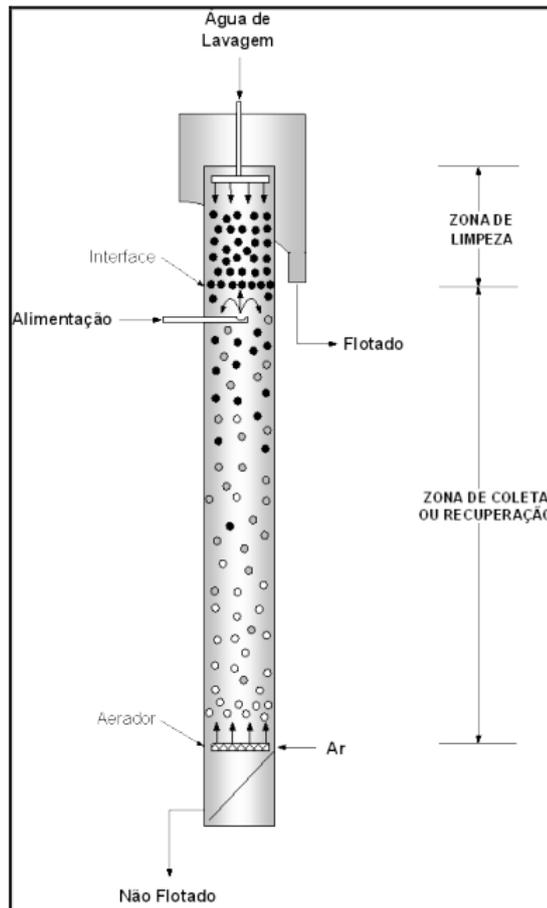


Figura 2.2: Esquema representando as partes de uma coluna de flotação
Fonte: AQUINO *et al.*, 2010

2.3. Flotação em Célula Mecânica

Assim como na coluna de flotação, a complexa fluidodinâmica do processo de flotação em célula mecânica oferece grande espaço para exploração.

Na célula mecânica, a flotação também pode ser direta ou reversa, apresentando os mesmos conceitos da flotação em coluna. Da mesma maneira, a flotação reversa é utilizada para o beneficiamento de minério de ferro no Brasil (MAMEDE, 2016).

A célula mecânica de flotação é uma máquina agitada mecanicamente, onde as bolhas são geradas como resultado da turbulência do agitador. Como mostrado na Figura 2.3, o compartimento de alimentação da célula é instalado em uma de suas extremidades e possui um

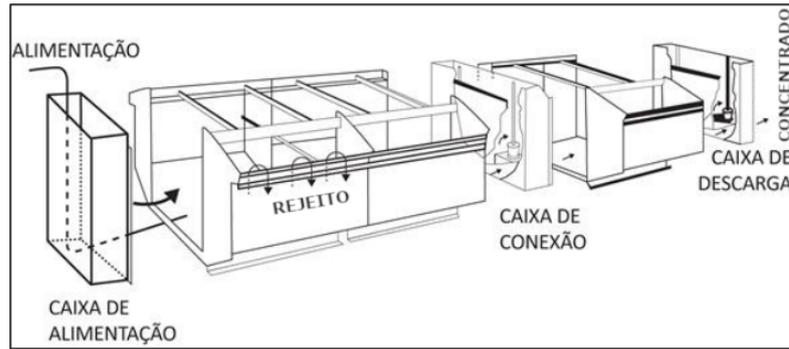


Figura 2.3: Esquema representando as partes de uma célula de flotação em vista explodida
Adaptado de: CHAVES *et al.*, 2018

dispositivo para a regulagem do nível da polpa dentro das células. A espuma formada é descarregada pela frente, transbordada sobre calhas ao longo da célula, enquanto a parte não-flotada sai pelo fundo e é passada para a próxima célula ou é descarregada na caixa de descarga, localizada na outra extremidade do conjunto. Existem diferentes tipos de máquinas de flotação, que são instaladas dentro da célula. Essas máquinas são compostas, basicamente, de um rotor suspenso por um eixo conectado a um acionamento. Ele é responsável por agitar a polpa e mantê-la em suspensão. Em alguns modelos, a pressão negativa gerada pelo movimento rotacional é suficiente para a geração de bolhas de ar. Em outros casos, é necessária a utilização de ar comprimido e é feita a instalação de um estator em torno do rotor, para fragmentar as bolhas de ar (CHAVES *et al.*, 2018; LUZ, 2016).

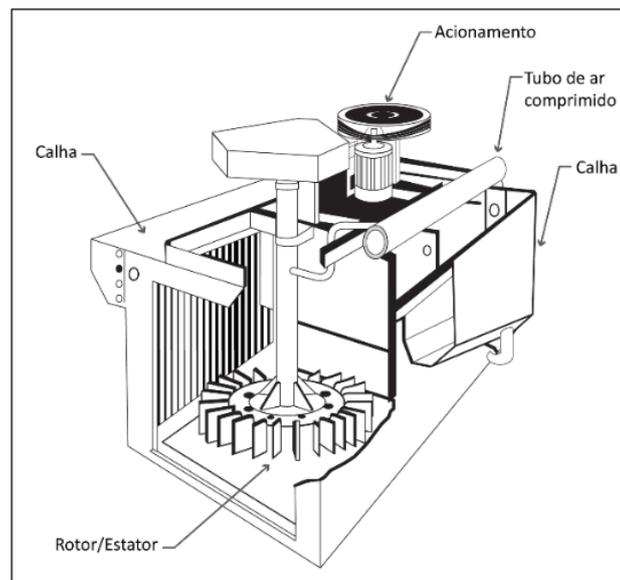


Figura 2.4: Máquina de flotação modelo Galigher
Fonte: CHAVES *et al.*, 2018

Existem variados projetos de células mecânicas de flotação, com diferentes formatos de células, projetos de rotores e estatores, entre outros aspectos mecânicos. Um desenho mais moderno das células é o cilíndrico, que possui maior facilidade de construção e ainda fornece maior eficiência para o contato entre as partículas e as bolhas de ar (CHAVES, 2006 *apud* CHAVES *et al.*, 2018; YOUNG, 1982 *apud* CHAVES *et al.*, 2018).

Dois exemplos de tipos de células de flotação são o modelo Galigher, mostrado na Figura 2.4, e o TankCell, mostrado na Figura 2.5. Diferentes desenhos de rotores e estatores também podem ser vistos na Figura 2.6. O *TankCell* é uma célula com tanque cilíndrico, que consegue emular o comportamento de uma coluna de flotação, um formato bastante utilizado atualmente.

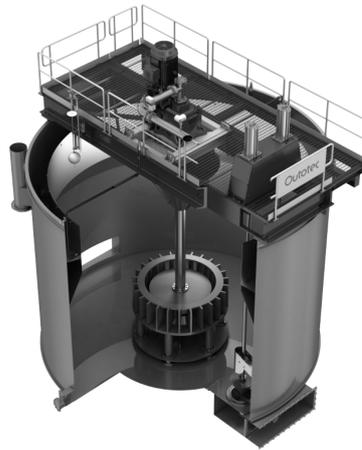


Figura 2.5: Célula de flotação *TankCell*, que possui um tanque cilíndrico
Adaptado de: OUTOTEC, 2017

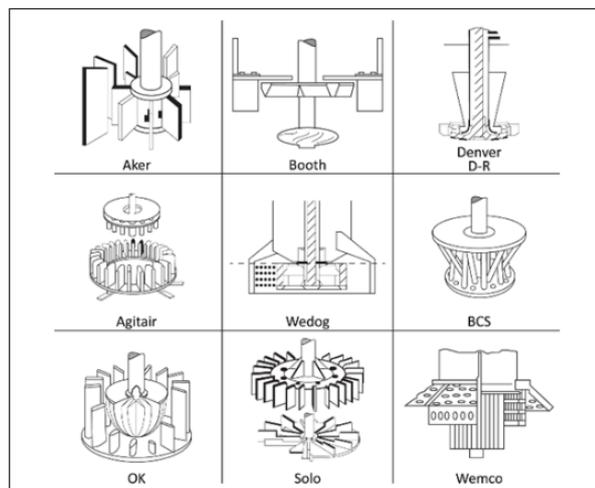


Figura 2.6: Esquemáticos de diferentes tipos de rotores e estatores
Fonte: YOUNG, 1982 *apud* CHAVES *et al.*, 2018

2.4. *Soft Sensors*

Soft sensor ou sensor virtual, são nomes usados para denominar um software que processa diversas medições e infere o valor de uma variável não mensurada a partir das entradas disponíveis. Com o aumento do poder computacional e a crescente disponibilidade de dados de processo, faz-se possível o desenvolvimento de *soft sensors*, que reúnem diferentes medições e variáveis relevantes a um processo, utilizam um modelo para calcular ou estimar outras variáveis e têm como saída a inferência do valor de variáveis difíceis de se mensurar ou que não podem ser medidas em tempo real (GONZAGA *et al.*, 2009; KADLEC *et al.*, 2011; RALLO *et al.*, 2002).

Usualmente, são descritas três formas de modelar um *soft sensor*. Os modelos denominados caixa-branca, ou *model-driven*, são construídos quando se possui completo conhecimento da natureza e do comportamento do sistema, com informações detalhadas e baseados na modelagem matemática do processo. Já os modelos denominados caixa-preta, ou *data-driven*, são construídos sem o conhecimento detalhado do sistema, usando seus dados para criar modelos estatísticos, utilizando técnicas de regressão para aprender as dinâmicas e não-linearidades do sistema. Os modelos chamados de caixa-cinza, ou híbridos, são construídos por meio de uma combinação entre os dois anteriores, utilizando as modelagens matemáticas disponíveis incorporadas aos dados do processo. Uma vantagem da abordagem caixa-preta em relação à caixa-branca é a maior facilidade em ser desenvolvida quando se tem dados representativos do sistema (GONZALEZ, 1999; KULCSÁR, 2016; LOTUFO e GARCIA, 2008; SUSANA, 2015).

Os procedimentos básicos para a construção de um *soft sensor*, ilustrados na Figura 2.7, podem ser descritos como: coleta, seleção e análise de dados, quando são identificados os dados a serem coletados e analisados aspectos como sua variação no tempo e faixas de medição; pré-processamento de dados, para detecção e tratamento de *outliers*, dados faltantes e realização da normalização; seleção, treinamento e validação do modelo, etapa essencial para o desenvolvimento do *soft sensor*; e manutenção do *soft sensor*, necessária para que o sensor continue apresentando boa resposta mesmo após variações decorridas no processo (KADLEC *et al.*, 2009; KULCSÁR, 2016; SUSANA, 2015).

Na indústria, o *soft sensor* pode ser usado para diferentes finalidades, desde a simples obtenção do dado para monitoramento até o seu uso no controle do sistema. Todo dispositivo de medição está sujeito a falhas, de modo que precisam ser desligados e substituídos, seja tem-

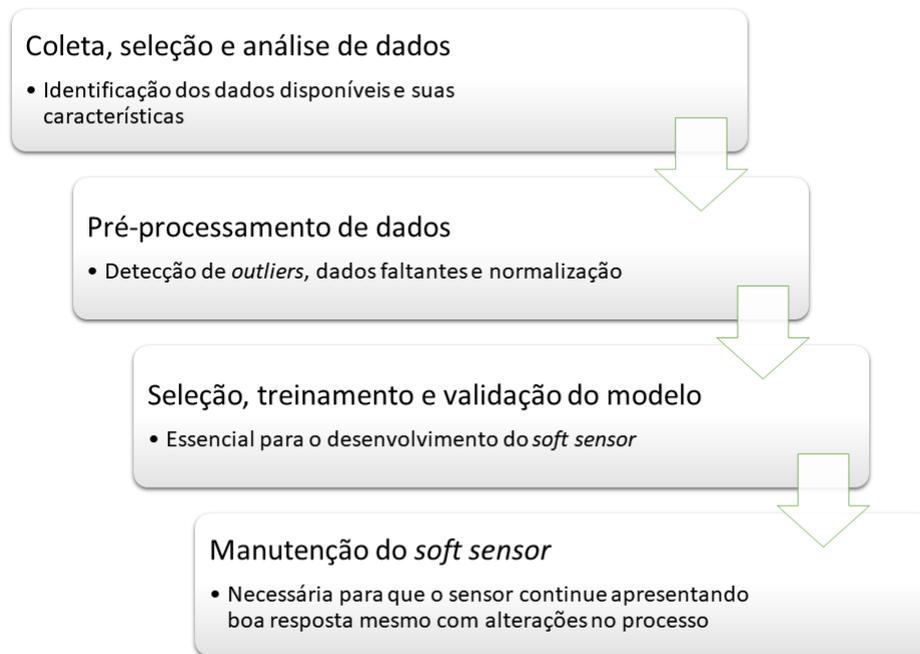


Figura 2.7: Fluxograma de procedimentos para construção de um *soft sensor*
Fonte: o autor

porariamente, para manutenção, ou definitivamente, por um novo. Nestes casos, o *soft sensor* pode substituir temporariamente o dispositivo desligado, funcionando como um *backup* e evitando perdas de performance no processo e aumento de custos. Ele também pode ser usado como um substituto definitivo, reduzindo custos com *hardware*. Contudo, dada a falta de redundância, isto deve ser feito com cautela, sendo realizada uma análise crítica da performance do modelo e avaliações periódicas para verificação e recalibração, no caso de processos dinâmicos. Deve-se levar em conta que uma nova calibração deve ser feita quando os parâmetros do processo são alterados. Sendo usado como uma forma de redundância, o *soft sensor* pode funcionar como uma validação do sensor real, detector de falhas e ferramenta de diagnóstico, indicando falhas quando os dados gerados apresentam discrepância em relação aos dados do *hardware* (FORTUNA *et al.*, 2007).

O foco deste trabalho, no entanto, está no uso do *soft sensor* para a estimação em tempo real de variáveis de processo. Uma das vantagens da estimação em tempo real é a possibilidade de evitar o uso de *hardware* de alta complexidade e elevados custos. Exemplos de *hardwares* deste tipo são equipamentos que utilizam raios X, que demandam cuidados especiais no manuseio, manutenção e descarte. Outra vantagem é a redução do tempo de espera para a obtenção do resultado da medição, muitas vezes obtidos apenas por meio de análise laboratorial, o que

toma algum tempo e atrasa possíveis tomadas de decisão. Em casos de alto tempo de espera para obtenção dos dados de medição, estes valores se tornam inviáveis para aplicação no controle do processo, o que passa a ser possível com o uso de um *soft sensor* bem construído e calibrado (FORTUNA *et al.*, 2007; GONZAGA *et al.*, 2009). Lin *et al.* (2007) desenvolvem dois *soft sensors*, um para predição de cal livre e outro para a predição de emissão de óxidos de nitrogênio (NOx), ambos usados para avaliação da qualidade do cimento produzido na indústria de cimentos. Foram utilizados o *principal component regression* (PCR) e o *partial least squares* (PLS) com resultados razoavelmente precisos, mostrando potencial para utilização no controle de qualidade do produto. Em Rallo *et al.* (2002), RNA são usadas no desenvolvimento de um *soft sensor* para predição em tempo real do índice de fusão (MI), indicador da qualidade do produto da polimerização de etileno para produção de LDPE (*low-density polyethylene*). O *soft sensor* desenvolvido alcançou erro médio relativo de aproximadamente 4%. A obtenção do MI normalmente é feita em laboratório e leva entre 2h e 4h. Já Susana (2015) fez a implementação de 13 algoritmos para desenvolvimento de *soft sensor* que poderia reduzir pela metade a frequência de análises laboratoriais para medição da quantidade de flúor no efluente de uma estação de tratamento de águas residuais.

Duas metodologias amplamente utilizadas no desenvolvimento de *soft sensors*, que são apresentadas a seguir, são as Redes Neurais Artificiais, como nos trabalhos de Rallo *et al.* (2002), Ko e Shang (2011), Wang *et al.* (2014), Yan *et al.* (2018), e Drumond *et al.* (2018), e as *Random Forests*, como nos trabalhos de Pal (2005), Marais (2010), Wang *et al.* (2016) e Drumond *et al.* (2018), o que pode ser explicado por sua confiabilidade na estimação de valores e elevada capacidade de reconhecimento de padrões em sistemas complexos, não lineares e com atrasos (ZANATA, 2005).

2.5. Redes Neurais Artificiais

As RNA foram desenvolvidas para tentar levar a programação de um computador digital a um patamar mais próximo do processamento do cérebro humano, que é um sistema de processamento altamente complexo e não-linear. Para isso, as RNA são compostas de duas ou mais camadas formadas por unidades de processamento simples que se interconectam, chamadas de neurônios. As duas camadas básicas são a de entrada e de saída. Como o nome indica, a camada de entrada corresponde aos sinais de entrada da RNA, e a camada de saída corres-

ponde aos resultados dos processamentos da RNA. Outras camadas podem compor a interface entre as camadas de entrada e saída, são as chamadas camadas ocultas. Estas são opcionais e responsáveis por extrair padrões associados ao processo analisado.

Cada uma das entradas do neurônio é formada pelo produto do sinal de entrada pelo peso atribuído à conexão entre o sinal de entrada e o neurônio. Todo neurônio possui uma função somatória, que efetua a soma de todas as suas entradas, e uma função de ativação, que recebe a soma das entradas e limita a amplitude do sinal de saída do neurônio a um intervalo especificado, normalmente $[0, 1]$ ou $[-1, 1]$ (HAYKIN, 2005; SILVA *et al.*, 2017).

Várias funções de ativação podem ser utilizadas, sendo divididas entre funções com descontinuidade e funções contínuas diferenciáveis. Na Figura 2.8 são mostradas as funções degrau bipolar (a) e de Heaviside ou degrau unitário (b), que possuem um ponto de descontinuidade cada. Na função degrau bipolar, entradas com valores positivos assumirão o valor de saída positivo igual 1, entradas com valores negativos assumirão o valor de saída negativo igual a -1, e as entradas nulas terão saídas nulas. Na função de Heaviside a saída será igual a 1 para todo valor maior ou igual a 0, enquanto qualquer entrada negativa terá saída nula (SILVA *et al.*, 2017).

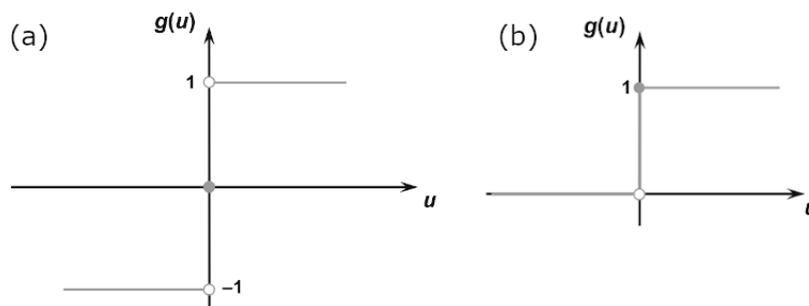


Figura 2.8: (a) Função de ativação degrau bipolar; (b) Função de ativação de Heaviside ou degrau unitário

Adaptado de: SILVA *et al.*, 2017

As funções de ativação linear, logística e tangente hiperbólica são diferenciáveis em todo seu domínio e mostradas na Figura 2.9. Na função linear, o resultado de saída é igual ao valor de entrada. O resultado de saída da função logística assumirá sempre um valor real entre 0 e 1, sendo possível ajustar a inclinação da curva no ponto de inflexão. Já o resultado de saída da função tangente hiperbólica assumirá sempre um valor real entre -1 e 1, também sendo possível ajustar a inclinação da curva no ponto de inflexão. É importante ressaltar que tanto a função logística quanto a tangente hiperbólica são exemplos de função sigmoide (SILVA *et al.*, 2017).

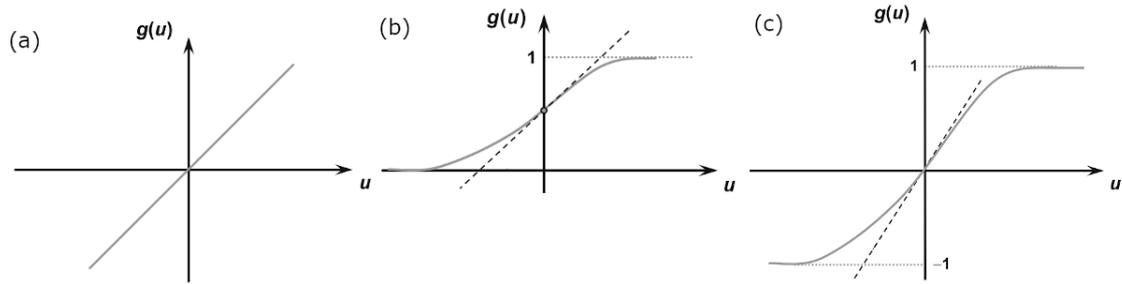


Figura 2.9: (a) Função de ativação linear; (b) Função de ativação logística; (c) Função de ativação tangente hiperbólica

Adaptado de: SILVA *et al.*, 2017

As principais arquiteturas de RNA são a *single-layer feedforward* (diretas com uma camada), *multilayer feedforward* (direta com múltiplas camadas), recorrentes ou *mesh*. Nas redes *feedforward*, ou diretas, o fluxo de informações segue em apenas um sentido, da camada de entrada para a de saída.

A arquitetura *single-layer feedforward*, é dita de uma única camada desconsiderando os neurônios de entrada, logo possui apenas a camada de saída. Na Figura 2.10 está representado um exemplo de *single-layer feedforward* com 3 entradas e 4 neurônios na camada de saída. Essa arquitetura normalmente é usada em problemas de classificação e filtragem linear. Já a arquitetura *multilayer feedforward* possui uma ou mais camadas ocultas e é utilizada para problemas como o de aproximação de função, classificação, identificação de sistemas, entre outros. Um dos principais modelos de RNA que usa a arquitetura *multilayer feedforward* é o *multilayer perceptron*, MLP. Na Figura 2.11 está representado um esquema de uma rede *multilayer feedforward* com 3 entradas, 1 camada oculta composta por 6 neurônios, e pela camada de saída, composta por 1 neurônio. O número de camadas ocultas e suas quantidades de neurônios não é predeterminado e depende da natureza e complexidade do problema abordado, assim como da quantidade e qualidade de dados disponíveis (SILVA *et al.*, 2017).

O MLP possui fluxo de informações direto, da entrada, passando pelas camadas ocultas e indo até a camada de saída. Seu treinamento é do tipo supervisionado, onde estão disponíveis os resultados de saída esperados, e chamado de *backpropagation*, pois os erros (diferença entre o resultado obtido e o esperado) são calculados e propagados no sentido contrário, realizando ajustes nos pesos e *bias* para que tais erros sejam reduzidos. Uma das variantes do *backpropagation* é o método de Levenberg-Marquardt (LM), um método quasi-Newton, mais rápido que métodos de Newton (que usam a matriz Hessiana) pois utiliza uma aproximação da matriz

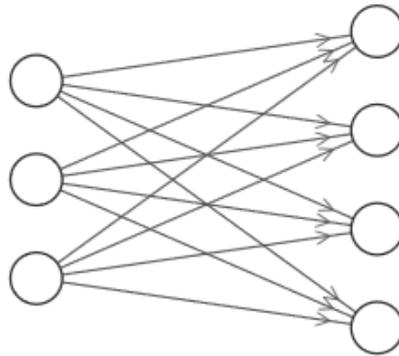


Figura 2.10: Representação de uma Rede Neural Artificial do tipo *single-layer feedforward*
 Fonte: o autor

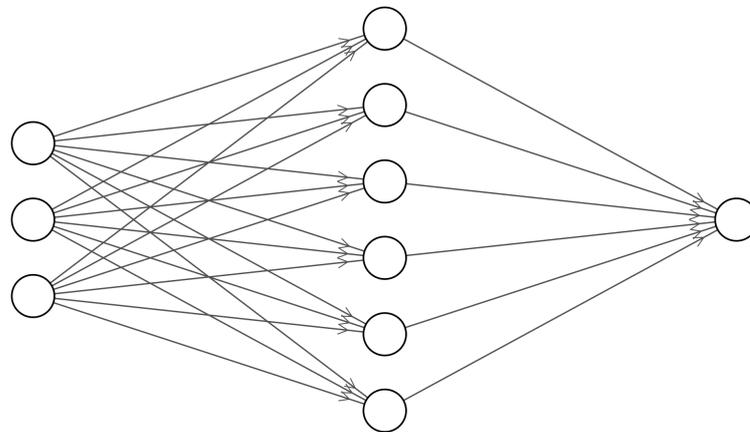


Figura 2.11: Representação de uma Rede Neural Artificial do tipo *multilayer feedforward*
 Fonte: o autor

Hessiana para fazer o ajuste dos pesos e *bias* da rede e possui um parâmetro que ajusta a taxa de convergência do algoritmo, indo de um comportamento próximo à descida gradiente e se aproximando do método de Newton quando a solução corrente está próxima de um ótimo local. Normalmente o método de LM é o mais rápido e recomendado como primeira opção. Outro método variante é o *Scaled Conjugate Gradient* (SCG), ou gradiente conjugado em escala, que pode se comportar como gradiente conjugado clássico, descida gradiente, quasi-Newton ou uma combinação de quasi-Newton com gradiente conjugado, conforme seus parâmetros são definidos. Esse método pode apresentar melhor desempenho que LM em regressões onde o número de pesos e bias são elevados (ANDREI, 2007; HAYKIN, 2005; SILVA *et al.*, 2017; ZANATA, 2005).

As redes recorrentes usam a saída de pelo menos um neurônio da camada de saída como

entrada de neurônios das camadas anteriores. Essa característica permite o uso dessa arquitetura em sistemas variáveis no tempo, como predição de séries temporais, identificação de sistemas, controle de processos, entre outros. A Figura 2.12 mostra um exemplo de rede recorrente onde um sinal de saída realimenta a camada central. Já as redes com arquitetura *mesh* levam em consideração a posição espacial dos neurônios, diretamente relacionada ao ajuste dos pesos e *bias*, e podem ser usadas para aplicações como clusterização de dados, reconhecimento de padrões, otimização de sistema, entre outros. Um exemplo de rede com neurônios arranjados em um espaço bidimensional é mostrado na Figura 2.13.

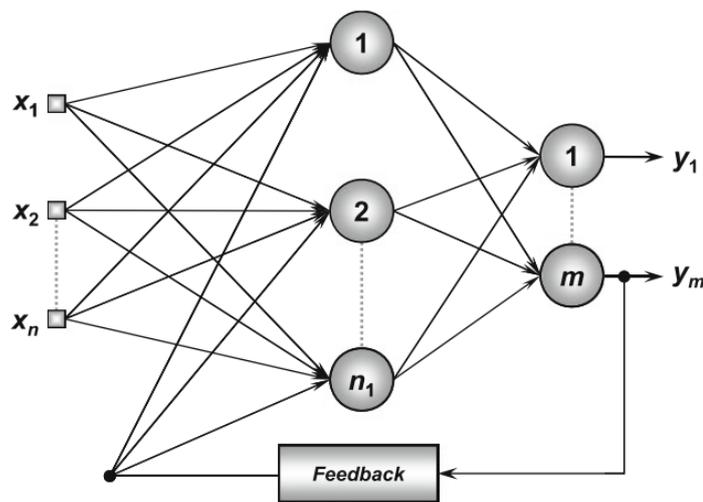


Figura 2.12: Representação de uma Rede Neural Artificial recorrente
 Fonte: SILVA *et al.*, 2017

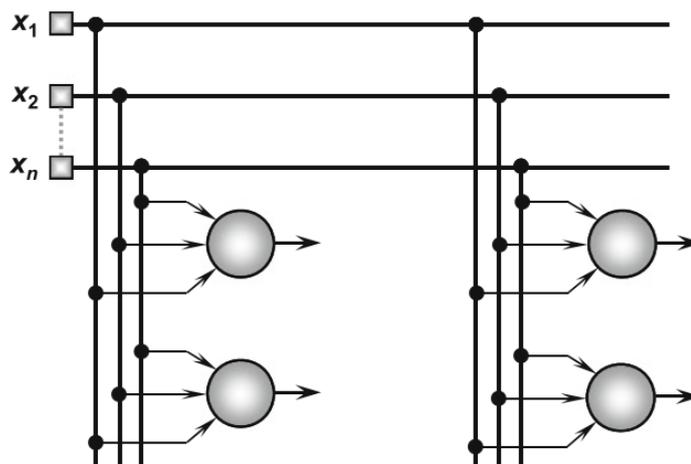


Figura 2.13: Representação de uma Rede Neural Artificial *mesh*
 Fonte: SILVA *et al.*, 2017

Uma importante aplicação de RNA é o desenvolvimento de *soft sensors*. No processo da flotação, assim como em vários outros processos na mineração de forma geral, existe uma

grande dificuldade de se fazer uma modelagem matemática, devido à sua alta complexidade, já que é um processo multivariável e uma variável tem grande influência sobre as outras. Neste contexto, o uso de RNA para o desenvolvimento de *soft sensors* se mostra adequado por utilizar parâmetros relevantes ao processo para a inferência das variáveis desejadas e por seu alto poder de aprendizagem (RALLO *et al.*, 2002). Esta adequação pode ser vista em trabalhos como Ko e Shang (2011), que utiliza análise de imagem para a estimação da distribuição do tamanho de partículas na moagem de minério e obtém valores de R^2 que ultrapassam 99%; Wang *et al.* (2014), que utiliza análise de imagem para a predição do teor do concentrado no processo de flotação de ferro e alcança resultados com percentagem de erro médio absoluto de até 0,099% após o uso do algoritmo de busca *shuffled cuckoo* para otimização dos parâmetros de aprendizado da RNA; Yan *et al.* (2018), que utiliza RNA com *backpropagation* para a predição do teor do concentrado no processo de flotação de cobre, obtendo 87% dos resultados com erro entre -2% e 2%; e Drumond *et al.* (2018), que utilizou tanto RNA quanto RF na predição do teor do concentrado no processo de flotação de ferro, alcançando um erro médio absoluto de 0,80% com utilização de RNA e 1,15% com RF.

2.6. *Random Forest*

Árvore de decisão é uma técnica que se tornou popular devido à sua facilidade de uso e interpretação, porém possui como desvantagens a falta de robustez e performance subótima. Para lidar com tais deficiências, uma das opções é o uso de agrupamentos de árvores seguido pela média das predições de uma certa quantidade de árvores, método chamado de *Random Forest* (AHMAD *et al.*, 2017; LARIVIÈRE e VAN DEN POEL, 2005).

Uma árvore de decisão é formada a partir de um conjunto inicial de dados, que compõe o nó raiz. Esse nó é dividido em dois novos nós, de acordo com critérios predeterminados baseados em uma variável e um ponto de divisão, que são também divididos sucessivamente até que não seja mais possível dividi-los ou que um critério de parada tenha sido atingido. Nós que não foram divididos são chamados de folhas e representam a predição da árvore. Na Figura 2.14 está ilustrado um esquema de uma árvore de decisão, onde o conjunto de dados inicial se encontra no nó raiz, representado por uma elipse. Os nós que são subdivididos são chamados de nós internos e estão representados na figura pelos círculos. Já as folhas, ou nós terminais, estão representadas pelos retângulos.

Ao fim do processo de divisão, chamado de crescimento, uma árvore excessivamente grande pode ter sido gerada, podendo apresentar uma falta de generalização. Após essa etapa é comum a aplicação da poda, em que é avaliada a melhoria da performance da árvore com cada divisão realizada. Caso uma divisão não alcance uma melhoria mínima, a árvore é podada naquele ponto e o nó interno passa a ser uma folha, reduzindo a complexidade da árvore sem abrir mão de uma boa performance (YANG *et al.*, 2017). A etapa da poda pode também ser realizada durante o crescimento da árvore, apresentando uma redução no seu custo computacional de desenvolvimento, porém isso pode levar à perda prematura de um ramo que poderia gerar melhoria da performance.

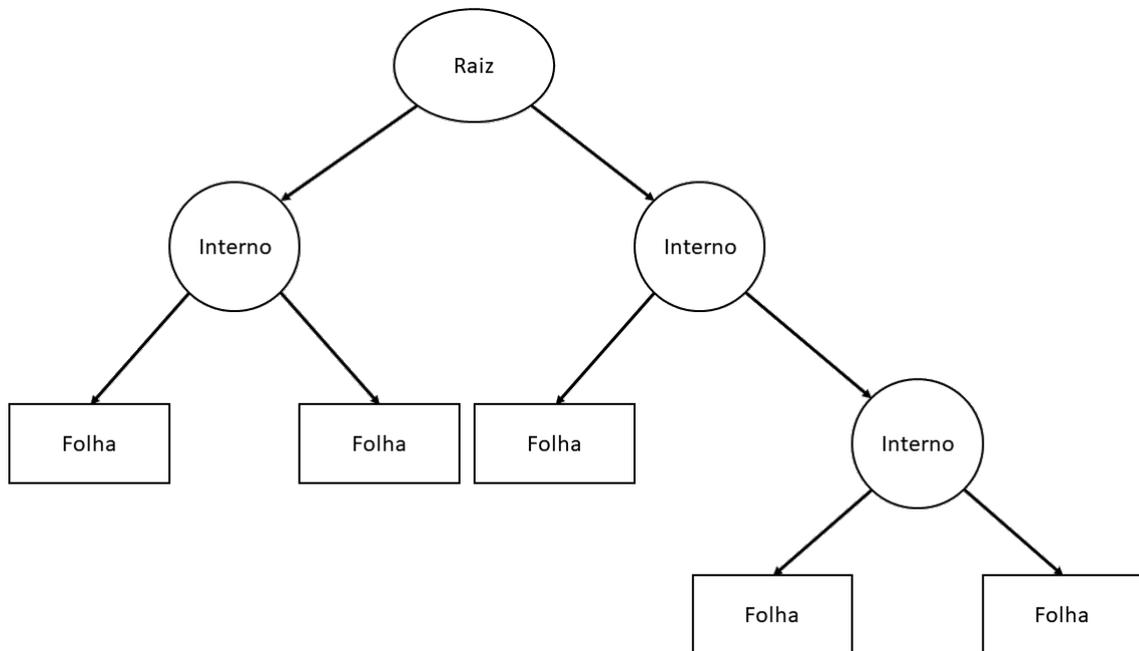


Figura 2.14: Representação de uma árvore de decisão

Fonte: o autor

Uma forma de melhorar o desempenho de um algoritmo isolado, como uma árvore de decisão, é usar um método do tipo *ensemble*. *Ensemble* é o nome dado a um conjunto de algoritmos de aprendizado de máquina que têm seus resultados combinados de alguma maneira, como pela média, média ponderada, moda, entre outros. Em geral, métodos *ensemble* são mais precisos do que os algoritmos que os geraram (DIETTERICH, 2000).

Dentre as técnicas para desenvolvimento de algoritmos *ensemble* é possível destacar *bagging*, *boosting* e *stacking*. *Bagging* vem do termo em inglês *bootstrap aggregation*, que seleciona aleatoriamente vários subconjuntos de treino do conjunto inicial. Os subconjuntos têm mesmo tamanho e podem ser escolhidos sem reposição ou com reposição, de forma que

uma mesma amostra pode ser escolhida em diferentes subconjuntos ou mais de uma vez em um mesmo subconjunto. *Boosting* é uma técnica iterativa que se concentra nos exemplos mal classificados pelo modelo anterior. Para isso, exemplos considerados mais difíceis são incluídos mais vezes nos conjuntos de treino, além de atribuir pesos aos modelos gerados, sendo pesos maiores para aqueles que apresentam erro pequeno e pesos menores para aqueles que apresentam erro grande. Já o *stacking*, diferentemente do *bagging* e do *boosting*, utiliza vários algoritmos diferentes em sua construção, sendo considerado um *ensemble* heterogêneo. De forma simplificada, o *stacking* faz o treinamento de diferentes modelos que fazem previsões intermediárias, uma para cada modelo treinado. Tais previsões intermediárias são usadas como entrada para o modelo final, que fará a previsão final. O modelo final é dito estar empilhado (*stacked*) sobre os iniciais, por isso o nome da técnica (DIETTERICH, 2000; MURRUGARRA-LLERENA, 2011).

A RF é uma derivação do *bagging*, proposta por Breiman (2001), composta por um agrupamento de árvores de decisão. Nela, além da seleção dos subconjuntos ser feita de forma aleatória, a seleção da divisão dos nós também é feita aleatoriamente para cada nó de cada árvore, com o resultado da previsão sendo obtido pela média das previsões das árvores geradas. Nessa abordagem, todas as árvores são desenvolvidas completamente, significando que o processo de crescimento é feito até que os nós terminais atinjam o tamanho mínimo definido (BIAU e SCORNET, 2016; BREIMAN, 2001).

As RF também são amplamente utilizadas, sendo possível encontrar diversos trabalhos com esse método. Pal (2005) utilizou RF para fazer a identificação de 7 tipos de plantações usando imagens de satélite em período de colheita, obtendo acurácia de 88,37%. Também com a utilização de imagens de satélite, Wang *et al.* (2016) usa RF para fazer uma estimativa da biomassa em plantações de trigo em 3 de suas etapas de desenvolvimento, sendo a biomassa um dos indicadores mais úteis do desenvolvimento e saúde de uma plantação. O modelo de Wang *et al.* (2016) alcançou R^2 de 0,533, 0,721 e 0,790 para cada uma das etapas (inicial, intermediária e final) superando resultados encontrados com a utilização de RNA e *support vector regression* (SVR). Marais (2010) aplicou a RF para fazer a previsão do teor do concentrado no processo de flotação de platina baseado em características extraídas de imagens da espuma. O modelo utilizado foi de classificação, dividido em 3 classes de faixas de teores, e obteve uma acurácia de 85%.

2.7. Redução de Dimensionalidade

O termo dimensionalidade é dado ao número de características de uma representação, ou seja, a quantidade de variáveis usadas (CAMPOS, 2001). Com a grande disponibilidade de informações, grandes capacidades de armazenamento digital e melhorias constantes na transmissão de dados, o volume de dados coletados pode ser enorme. No entanto, muitos métodos estatísticos perdem poder quando aplicados a uma base de dados de alta dimensionalidade. Ao reduzir a dimensionalidade dos dados, é possível evitar esse fenômeno, obtendo uma representação que reduz a presença de ruídos e preserva os sinais de maior influência no processo (NGUYEN e HOLMES, 2019).

A redução de dimensionalidade pode ser realizada por métodos que são divididos em duas categorias: extração de características (*feature extraction*) e seleção de variáveis (*feature selection*).

2.7.1. Extração de Características

A extração de variáveis, tradução do termo em inglês *feature extraction*, é um processo que extrai um conjunto de novas variáveis a partir das variáveis originais por meio de alguma função de mapeamento.

Alguns exemplos de métodos de extração de características podem ser citados, como o *independent component analysis* (ICA), usado em Artoni *et al.* (2019) para apresentar uma coleção de dados de eletroencefalogramas (EEG) de pessoas realizando tarefas de memória visual, e o *linear discriminant analysis* (LDA), usado em Kenobi *et al.* (2017) para classificar a arquitetura da raiz de plantas de trigo entre alta ou baixa captação de nitrogênio e entre alto e baixo nitrato no meio de cultura. Outro exemplo é o *locally linear embedding* (LLE), usado em Monteiro *et al.* (2016) para aumentar a eficiência na codificação de imagens do formato HEVC (*High Efficiency Video Coding*).

Por fim, deve ser citado o PCA (*principal component analysis*), um método de redução de dimensionalidade linear, que significa que ele reduz o número de variáveis por meio de uma combinação linear dos dados em um subespaço linear de dimensionalidade inferior. O PCA é o mais popularmente utilizado dos métodos de extração de características linear (VAN DER MAATEN *et al.*, 2009).

O PCA faz uma projeção das amostras de dados originais por meio de uma transformação linear e atribuição de pesos de contribuição das variáveis originais em cada um dos componentes principais. O primeiro componente principal é o responsável pela maior variabilidade nos dados e cada componente seguinte tem a maior variabilidade sob a restrição de ser ortogonal aos componentes anteriores. A redução é alcançada com a seleção dos componentes de maior contribuição na explicação da variância nos dados originais, escolhendo tantos quanto desejado (POPLI, 2017).

2.7.2. Seleção de Características

A seleção de características, tradução do termo em inglês *feature selection*, é definida por Yu e Liu (2003) como “um subconjunto de variáveis (ou características) originais tal que o espaço de variáveis é otimamente reduzido de acordo com um certo critério de avaliação”, ou seja, as variáveis são selecionadas sem passar por nenhuma transformação, como é o caso da extração de características.

Os métodos de seleção de características podem ainda ser classificados em “*wrapper*” ou “filtro”. Métodos classificados como filtro utilizam ranqueamento e técnicas estatísticas para avaliação da qualidade das variáveis, sem o uso de algoritmos de aprendizado de máquina. O ranqueamento filtra as variáveis, resultando na seleção das características mais relevantes. Além disso, tais métodos demandam menos poder computacional e são mais rápidos que os do tipo *wrapper*. Métodos classificados como *wrapper* utilizam um algoritmo de aprendizado de máquina e a performance é empregada como critério de avaliação. Uma vantagem desses métodos é a interação entre a seleção de características e a seleção do modelo, porém possuem maior risco de sobre-ajuste (*overfitting*) que os métodos de filtro, além de serem computacionalmente mais caros (VISALAKSHI e RADHA, 2015).

Exemplos de métodos do tipo *wrapper* são o *forward selection*, em que variáveis são adicionadas sequencialmente a um modelo vazio na ordem em que mais contribuem com a melhoria do modelo, e o *backward selection*, em que variáveis mais irrelevantes são removidas uma a uma de um modelo inicialmente completo. O *forward selection* é conhecido pela sua velocidade, no entanto em algumas etapas pode incluir uma variável que torna outras redundantes. Isso é evitado no *backward selection*, que tem a desvantagem de ser computacionalmente mais caro e poder ser aplicado apenas quando a quantidade de amostras é muito superior ao número

de variáveis (NGUYEN, 2019).

Dentre os exemplos de métodos de seleção de características do tipo filtro estão o *Laplacian Score*, que pode ser utilizado de forma supervisionada ou não-supervisionada e que avalia cada variável individualmente de acordo com sua capacidade de manter a localidade da amostra de acordo com o gráfico de vizinhança (HE *et al.*, 2005; MIAO e NIU, 2016), e o SPEC, como é chamado o *framework* que pode ser utilizado de forma supervisionada ou não-supervisionada, que usa decomposição do espectro da matriz Laplaciana normalizada (*SPECTrum decomposition of \mathcal{L}*) e engloba algoritmos como *ReliefF* e *Laplacian Score* como casos especiais de sua aplicação (ZHAO e LIU, 2007). Outro exemplo é o NDFS (*nonnegative discriminative feature selection*), que explora as informações discriminativas em um cenário não-supervisionado por meio de uma clusterização em espectro para aprendizado das classificações dos *clusters* enquanto realiza a seleção de características de forma simultânea, permitindo a seleção das características mais discriminativas (LI *et al.*, 2012).

Por fim, é citada a família de algoritmos *Relief*. O algoritmo de treinamento supervisionado *Relief* estima a relevância (qualidade) das variáveis de acordo com quão bem ela distingue duas classes. Para isso, utiliza a diferença de valores entre duas instâncias para atribuir pesos para cada variável, indicando qual tem maior relevância na distinção entre as duas classes. A seleção é feita por meio de um valor de corte, em que as variáveis com peso médio de relevância superior ao valor de corte são selecionadas (KIRA e RENDELL, 1992). Porém, o *Relief* não lida com dados incompletos e está limitado a problemas de duas classes, limitações que são resolvidas com o uso de uma de suas extensões, o *ReliefF* (KONONENKO *et al.*, 1997). O algoritmo *Relief* e o *ReliefF*, contudo, são aplicáveis a problemas de classificação. Para a aplicação em problemas de regressão foi desenvolvida a extensão usada neste trabalho, a *RReliefF* (ROBNIK-ŠIKONJA e KONONENKO, 1997).

3. Materiais e Métodos

Foram coletados dados de entrada e saída do processo de flotação em célula mecânica de uma usina de beneficiamento de minério de ferro. Os dados são armazenados em um sistema de gerenciamento de dados, PIMS (*plant information management system*), de onde são retirados e analisados sob demanda. Para o desenvolvimento do *soft sensor*, esses dados foram organizados, filtrados e tratados para possibilitar seu uso, a identificação de relações entre eles e o desenvolvimento do *soft sensor* em si.

Neste capítulo os dados são descritos e são apresentados os métodos utilizados para seu tratamento e o desenvolvimento do *soft sensor* com a utilização do *software* MATLAB, que conta com diversas ferramentas auxiliares que facilitam a construção das RNA, das RF e a visualização dos resultados.

3.1. Dados

Para o desenvolvimento do *soft sensor* no processo de flotação, foi utilizado um conjunto de dados retirado de um circuito de flotação de minério de ferro composto de 11 células mecânicas de flotação. Os dados coletados compreendem o período entre as 0h do dia 1º de janeiro de 2019 até as 23h59min do dia 31 de maio de 2019.

Foram disponibilizadas um total de 73 variáveis obtidas do processo da flotação, sendo oito resultantes de análises laboratoriais. As análises de laboratório são feitas a partir de amostras coletadas e demandam duas horas para a liberação dos resultados. Duas dessas variáveis representam dados de saída, que são o teor de ferro e o teor de sílica no produto da flotação. Neste trabalho, o teor de ferro foi a saída escolhida para ser estimada e o teor de sílica não foi utilizada. As outras 6 variáveis obtidas por análise laboratorial (teor de ferro e teor de sílica na alimentação, na carga circulante e no rejeito) também não foram utilizadas, pois, assim como o teor de ferro, são obtidas apenas em intervalos de 2 horas, logo não possuem disponibilidade adequada para serem utilizadas na predição em tempo real.

A linha de flotação objeto de estudo é formada pelas células, uma caixa de alimentação (condicionadora) e uma caixa de recirculação, que são representadas no diagrama da Figura 3.1. As variáveis obtidas, separadas por etapa, são:

- **Caixa de Alimentação:** Nível, Densidade, Vazão Volumétrica, Vazão Mássica, Vazão de

Amina, Vazão de Amido;

- ***Rougher 1***: Nível de Espuma, Vazão de Ar, Abertura Válvula Dardo 1, Abertura Válvula Dardo 2, Abertura Válvula de Ar;
- ***Rougher 2***: Nível de Espuma, Vazão de Ar, Abertura Válvula Dardo 1, Abertura Válvula Dardo 2, Abertura Válvula de Ar;
- ***Cleaner 1***: Nível de Espuma, Vazão de Ar, Abertura Válvula Dardo 1, Abertura Válvula Dardo 2, Abertura Válvula de Ar;
- ***Cleaner 2***: Nível de Espuma, Vazão de Ar, Abertura Válvula Dardo 1, Abertura Válvula Dardo 2, Abertura Válvula de Ar;
- ***Recleaner***: Nível de Espuma, Vazão de Ar, Abertura Válvula Dardo 1, Abertura Válvula Dardo 2, Abertura Válvula de Ar;
- **Caixa de Recirculação**: Nível, Densidade, Vazão Volumétrica, Vazão Mássica;
- ***Scavenger 1 Banco 1***: Nível de Espuma, Vazão de Ar, Abertura Válvula Dardo 1, Abertura Válvula Dardo 2, Abertura Válvula de Ar;
- ***Scavenger 1 Banco 2***: Nível de Espuma, Vazão de Ar, Abertura Válvula Dardo 1, Abertura Válvula Dardo 2, Abertura Válvula de Ar;
- ***Scavenger 1 Banco 3***: Nível de Espuma, Vazão de Ar, Abertura Válvula Dardo 1, Abertura Válvula Dardo 2, Abertura Válvula de Ar;
- ***Scavenger 2 Banco 1***: Nível de Espuma, Vazão de Ar, Abertura Válvula Dardo 1, Abertura Válvula Dardo 2, Abertura Válvula de Ar;
- ***Scavenger 2 Banco 2***: Nível de Espuma, Vazão de Ar, Abertura Válvula Dardo 1, Abertura Válvula Dardo 2, Abertura Válvula de Ar;
- ***Scavenger 2 Banco 3***: Nível de Espuma, Vazão de Ar, Abertura Válvula Dardo 1, Abertura Válvula Dardo 2, Abertura Válvula de Ar;

Os dados do processo são coletados com diferentes taxas de amostragem, enquanto os resultados do laboratório são fornecidos a cada 02h. Para a modelagem do *soft sensor*, os dados

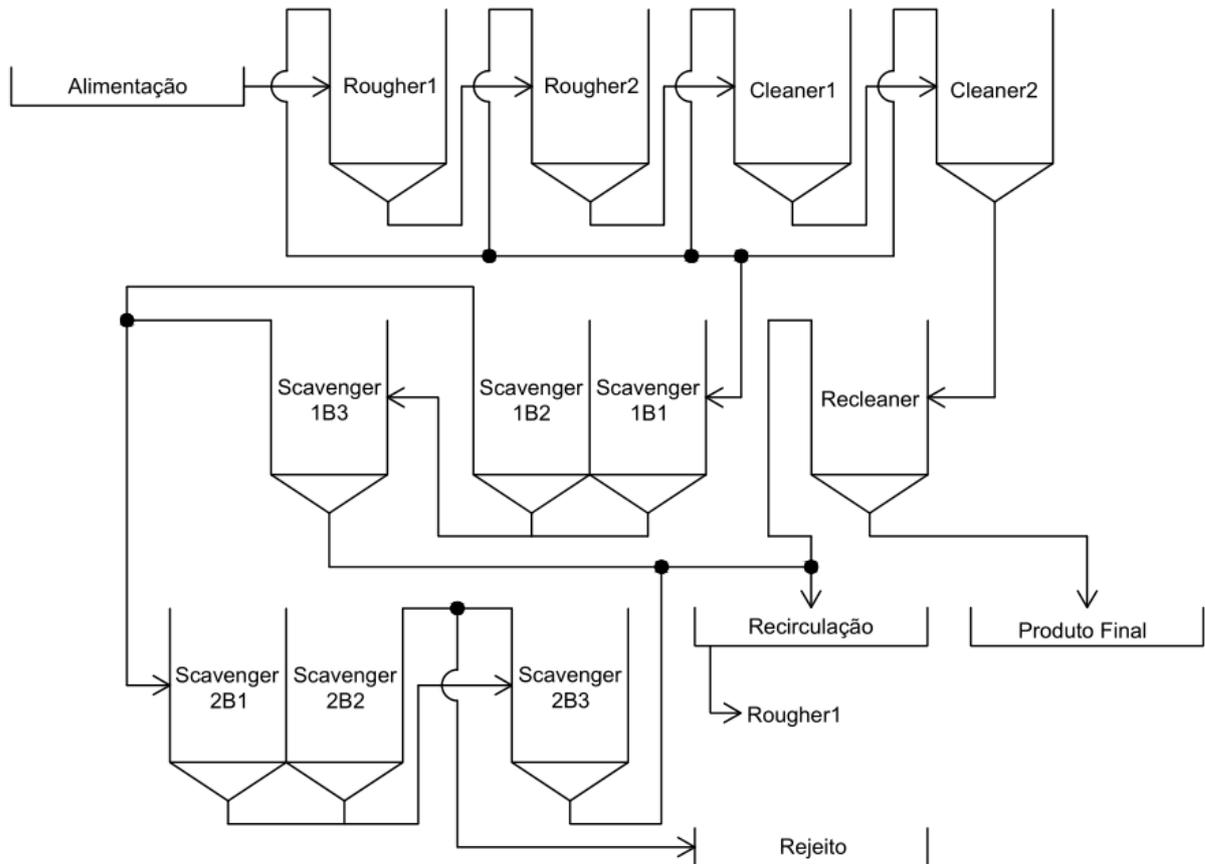


Figura 3.1: Representação da linha de flotação de onde os dados foram coletados
 Fonte: o autor

do processo foram agrupados em intervalos de duas horas e associados ao resultado da análise laboratorial obtido após decorrido o período supracitado. Essa associação está exemplificada na Tabela 3.1, onde os dados de processo entre 0h e 01h59min são associados ao resultado da amostra que foi coletada às 02h e liberado às 04h. Os dados de processo entre 02h e 03h59min são associados ao resultado da amostra que foi coletada às 04h e liberado às 06h, e assim por diante.

Cada variável possui diferentes tempos de amostragem, motivo pelo qual cada uma possui uma quantidade diferente de amostras no mesmo intervalo de tempo. Com isso e visto que a cada duas horas apenas um resultado de análise laboratorial é obtido, foi utilizada a mediana para resumir o conjunto de dados em cada intervalo de duas horas para cada entrada. Dessa forma, cada variável de entrada foi associada a uma saída. A mediana foi escolhida por sofrer menos interferências decorrentes de curtas variações no processo em comparação com a média simples, outra opção considerada. Para exemplificar tais interferências, é mostrada na Figura 3.2 a variação da média e da mediana da densidade da polpa na caixa de alimentação

Tabela 3.1: Relação entre período avaliado e resultado da análise laboratorial

Dados PIMS	Resultado Laboratorial				
	04:00	06:00	08:00	10:00	12:00
00:00 - 01:59					
02:00 - 03:59					
04:00 - 05:59					
06:00 - 07:59					
08:00 - 09:59					

Fonte: o autor

a cada duas horas e os intervalos interquartis destes agrupamentos. Na figura, é possível observar que em certo período de tempo, a densidade diminui consideravelmente devido a algum distúrbio no processo, porém por um intervalo de tempo relativamente curto, já que a maior parte dos dados permanece aproximadamente dentro do mesmo intervalo dos outros agrupamentos. Nesse período, a média diminui acompanhando a breve queda da densidade, enquanto a mediana acompanha a tendência dos dados, mantendo o comportamento esperado no período.

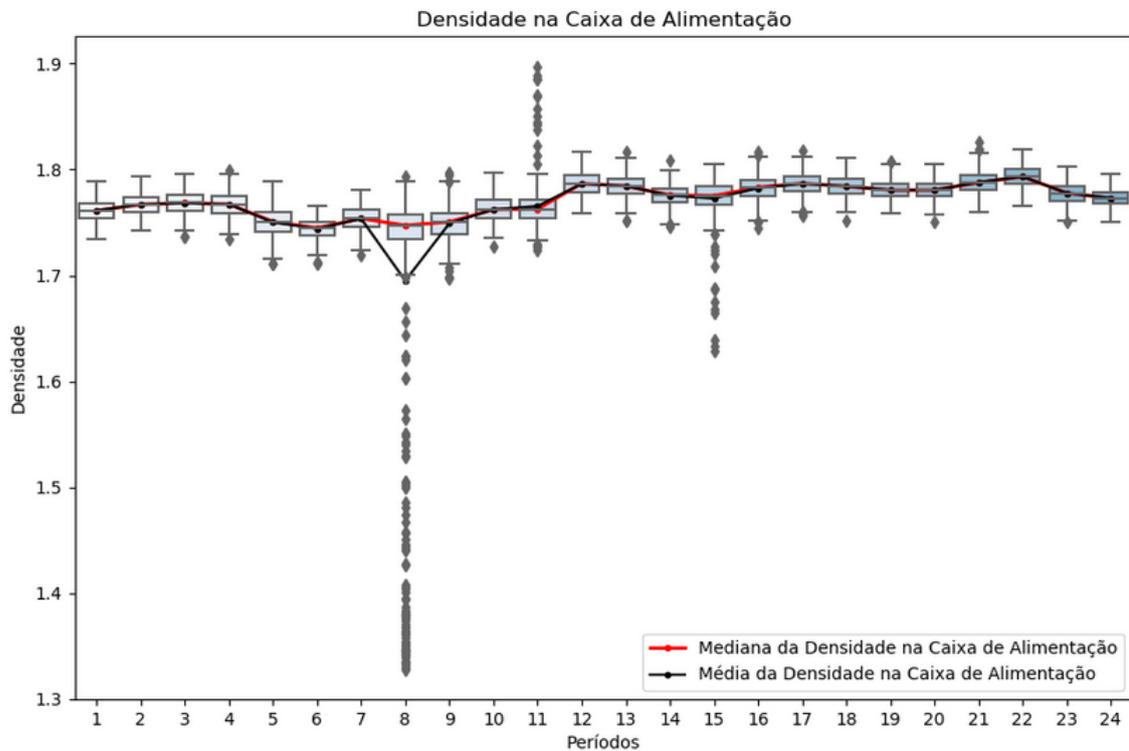


Figura 3.2: Média, mediana e intervalos interquartis da densidade da polpa na caixa de alimentação a cada período de duas horas

Fonte: o autor

Além dos tempos de amostragem, a escala das variáveis também é diferente entre si.

Para permitir que as variáveis fossem comparadas de forma igualitária, foi feita a normalização dos dados entre 0 e 1 com a utilização da fórmula descrita na Equação 3.1.

$$x' = \frac{(x - \text{antigoMenor}) \times (\text{novoMaior} - \text{novoMenor})}{\text{antigoMaior} - \text{antigoMenor}} + \text{novoMenor} \quad (3.1)$$

Sendo,

- x → valor antes da normalização
- x' → valor normalizado
- antigoMenor → menor valor antes da normalização
- antigoMaior → maior valor antes da normalização
- novoMenor → menor valor depois da normalização, definido como 0
- novoMaior → maior valor depois da normalização, definido como 1

Para realizar a normalização dos dados entre 0 e 1, a Equação 3.1 pode ser simplificada para a Equação 3.2.

$$x' = \frac{x - \text{antigoMenor}}{\text{antigoMaior} - \text{antigoMenor}} \quad (3.2)$$

A normalização dos dados tem o objetivo de facilitar o treinamento dos modelos de predição, reduzindo o efeito de grandes diferenças nas ordens de grandeza dos dados e ajudando na convergência de resultados assertivos de forma mais rápida.

3.2. Redução de Dimensionalidade

Um grande volume de dados é uma vantagem no desenvolvimento de uma pesquisa. Porém, a grande quantidade de variáveis dificulta o desenvolvimento de vários métodos estatísticos. Como exemplo, é possível citar a necessidade da disponibilidade de todas as variáveis na hora de usar uma Rede Neural treinada e o aumento das chances de a rede se tornar muito específica, sendo pouco generalista. Além disso, trabalhos como os de Kohavi e John (1997), Kohavi e Sommerfield (1995) e Hall (1999) mostram que variáveis irrelevantes e/ou redundantes atrapalham o desempenho de algoritmos de aprendizado de máquina e, portanto, devem ser removidas.

Para realizar a redução de dimensionalidade e, posteriormente, a modelagem do *soft sensor*, os dados foram agrupados por mês. Os dados que compõem o mês de março, que se encontram no centro do intervalo disponível, foram escolhidos para a continuidade dos procedimentos, de forma que é possível testar o modelo criado em dados anteriores e posteriores ao período utilizado no treinamento. Os critérios de avaliação utilizados neste trabalho formaram uma sequência de quatro etapas para alcançar uma redução de variáveis eficaz, resumidas na Figura 3.3. A primeira etapa consistiu de uma seleção preliminar dos dados com base em sua disponibilidade. Depois de realizado o agrupamento dos dados, conforme descrito na Seção 3.1, as variáveis com mais de 15% de dados faltantes foram consideradas impróprias para uso, pois uma grande quantidade de dados faltantes indica uma inconstância na disponibilidade dos dados, o que prejudica a aplicação do modelo criado caso essas variáveis não estejam sendo alimentadas. Essa foi a etapa com a maior redução de variáveis, sendo removidas 33 e restando 32. A maioria das variáveis removidas foram de Abertura de Válvula Dardo 1, Abertura de Válvula Dardo 2 ou Abertura de Válvula de Ar.

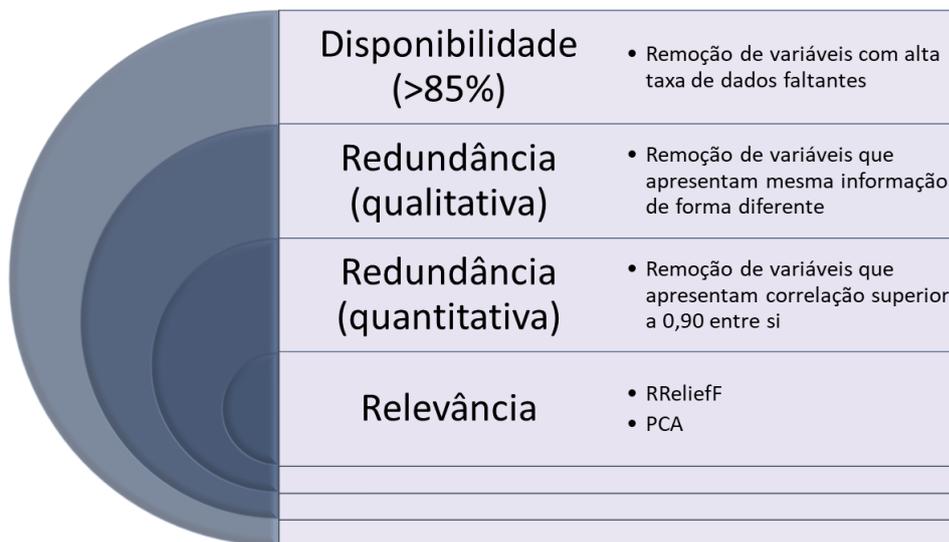


Figura 3.3: Etapas realizadas para redução de dimensionalidade

Fonte: o autor

A segunda etapa foi a identificação qualitativa de variáveis redundantes, que apresentam as mesmas informações de forma diferente, como Vazão Mássica e Vazão Volumétrica. No caso das variáveis Vazão Mássica e Vazão Volumétrica, o critério adotado para a escolha de qual deveria permanecer foi o fato de que a Vazão Volumétrica é medida por um instrumento e a Vazão Mássica é calculada a partir da Vazão Volumétrica e da Densidade do material, prevalecendo a

medição direta. As válvulas dardo 1 e válvulas dardo 2 são posicionadas lado a lado na saída da célula de flotação e respondem a um mesmo valor de *setpoint* de abertura, consequentemente também apresentam as mesmas informações e foram consideradas redundantes. Como padrão, as válvulas dardo 1 foram escolhidas para permanecerem no modelo. Outro conjunto de variáveis redundantes encontrado foi o de Vazões de Ar com Aberturas de Válvulas de Ar. Nesse caso nenhuma variável foi removida, pois todas as Aberturas de Válvulas de Ar já haviam sido desconsideradas na primeira etapa. A lista dos pares de variáveis que ainda estavam presentes, foram identificadas como redundantes e removidas é exibida na Tabela 3.2

Tabela 3.2: Variáveis mantidas e removidas na segunda etapa da redução de variáveis pela identificação qualitativa de redundância

Variável Mantida	Variável Redundante Removida
Vazão Volumétrica (Caixa de Alimentação)	Vazão Mássica (Caixa de Alimentação)
Vazão Volumétrica (Caixa de Recirculação)	Vazão Mássica (Caixa de Recirculação)
Abertura Válvula Dardo 1 (Cleaner 1)	Abertura Válvula Dardo 2 (Cleaner 1)
Abertura Válvula Dardo 1 (Cleaner 2)	Abertura Válvula Dardo 2 (Cleaner 2)
Abertura Válvula Dardo 1 (Recleaner)	Abertura Válvula Dardo 2 (Recleaner)
Abertura Válvula Dardo 1 (Scavenger 1 Banco 2)	Abertura Válvula Dardo 2 (Scavenger 1 Banco 2)

Fonte: o autor

Na terceira etapa foi feita uma avaliação quantitativa da redundância entre as variáveis de entrada, onde as variáveis redundantes são aquelas que apresentam alta correlação entre si e, por apresentarem informações duplicadas, acabam por prejudicar o aprendizado de máquina (HALL, 1999). De acordo com Yu e Liu (2003), a definição da correlação considerada suficientemente alta pode tomar como referência a maior correlação entre uma variável de entrada e a variável a ser estimada, sendo o valor de corte maior ou igual a tal correlação. Neste trabalho, a maior correlação encontrada com a concentração final de Ferro foi de 0,32. Esse e os demais valores de correlações encontrados podem ser vistos na Figura 3.4, onde é apresentado um mapa de calor de tais correlações. Como esse valor pode ser considerado baixo, foi adotado um valor mínimo de 0,90 para que um par de variáveis de entrada fosse considerado redundante. Na Tabela 3.3 são mostrados os pares de variáveis com alta correlação, indicando quais foram removidas e qual a correlação encontrada.

Com a remoção de 6 variáveis de entrada na segunda etapa e 5 na terceira, restaram 21 variáveis para a quarta etapa, que consiste da ordenação das variáveis por nível de relevância. Para isso, essa etapa foi realizada sob duas abordagens, uma utilizando o *RReliefF* para fazer uma seleção de variáveis e outra utilizando o PCA para fazer uma extração de variáveis (carac-



Figura 3.4: Mapa de calor das correlações entre os dados disponíveis ao final da segunda etapa de redução de variáveis
 Fonte: o autor

Tabela 3.3: Variáveis mantidas e removidas na terceira etapa da redução de variáveis pela identificação quantitativa de redundância

Variável Mantida	Variável Redundante Removida	Correlação
Vazão de Ar (Rougher 1)	Vazão de Ar (Rougher 2)	0,98
Nível de Espuma (Cleaner 1)	Nível de Espuma (Cleaner 2)	0,92
Vazão de Ar (Cleaner 1)	Vazão de Ar (Cleaner 2)	0,98
Abertura Válvula Dardo 1 (Cleaner 1)	Abertura Válvula Dardo 1 (Cleaner 2)	0,93
Abertura Válvula Dardo 1 (Cleaner 1)	Abertura Válvula Dardo 1 (Recleaner)	0,92

Fonte: o autor

terísticas). O algoritmo *RReliefF* não assume a existência de independência entre as variáveis e pode estimar corretamente a qualidade das variáveis em problemas com forte dependência entre parâmetros (ROBNIK-SIKONJA e KONONENKO, 2003) e por isso foi a abordagem escolhida utilizando a seleção de variáveis. Feita a ordenação com o *RReliefF*, os pesos atribuídos a cada variável foram utilizados para determinar as que seriam e as que não seriam usadas na modelagem do *soft sensor*. Já o PCA, que é um eficiente método e o mais popularmente utilizado dentre os métodos de extração de características, foi o escolhido para a segunda abordagem. Com a identificação dos principais componentes, foram selecionados os que, juntos, são responsáveis por mais de 90% da variabilidade dos dados.

3.3. Rede Neural Artificial

Usando a ferramenta *nntool*, do MATLAB, foi criada uma RNA MLP, do tipo *feedforward* com *backpropagation*. Isso significa dizer que os sinais de entrada da RNA são propagados camada a camada até a camada de saída, daí o nome *feedforward*, e a diferença entre o resultado da predição e uma parcela do resultado observado é propagada no sentido contrário até os sinais de entrada, sendo feitos ajustes nos pesos para o resultado se aproximar cada vez mais do observado, o que é chamado de *backpropagation*. A rede foi inicialmente criada com um neurônio na camada de saída, já que o objetivo é a predição de um único valor, a concentração de ferro, e uma camada oculta com 10 neurônios, valor padrão da ferramenta utilizada. Para essa determinação, parte-se do princípio de que as RNA MLP do tipo *feedforward* com 1 camada oculta são capazes de aproximar qualquer função contínua, sendo usadas camadas adicionais para obtenção de redundância de neurônios e maior precisão (ZANATA, 2005). Além disso, não existe um número definido para a quantidade de camadas ocultas e neurônios que

devem ser usados, variando conforme a natureza e complexidade do problema abordado, da quantidade e da qualidade dos dados disponíveis (SILVA *et al.*, 2017).

A função de ativação descreve a relação entre a entrada e a saída associada ao neurônio, e pode limitar o valor da saída a valores predefinidos (HAYKIN, 2005; SILVA *et al.*, 2017). Foi usada como função de ativação a função tangente hiperbólica, *tansig*, uma função comumente utilizada no desenvolvimento de Redes Neurais e definida pela Equação 3.3.

$$\varphi(u) = \frac{2}{1 + e^{-2u}} - 1 \quad (3.3)$$

3.3.1. Parâmetros de Treinamento da RNA

Para o treinamento da RNA, foram utilizados dois algoritmos supervisionados. O primeiro foi feito através da função *trainlm*, que faz a atualização dos pesos e *bias* por meio do método de otimização de Levenberg-Marquardt. Esse é chamado de quasi-Newton e é mais rápido que métodos de Newton, que usam a matriz Hessiana (que possui cálculos muito complexos), pois utiliza uma aproximação dessa matriz para fazer o ajuste dos pesos e *bias* da rede.

Como meio de comparação, outra função utilizada para o treinamento da RNA foi a *trainscg*, que faz a atualização dos pesos e *bias* por meio do método do Gradiente Conjugado em Escala (SCG). Essa função apresenta melhor desempenho que a de LM em regressões onde o número de pesos e *bias* são elevados.

Para a avaliação da performance da RNA foi usada a função MSE (*mean squared error*), que utiliza a média dos quadrados dos erros, indicando uma melhor performance quando o valor de MSE se aproxima de zero. A função MSE é definida pela Equação 3.4, onde n é o número de amostras tomadas, t_i é o resultado observado da amostra i , e y_i é o resultado da predição para a amostra i . A diferença $t_i - y_i$ é o erro resultante da predição.

$$MSE = \frac{1}{n} \sum_{i=1}^n (t_i - y_i)^2 \quad (3.4)$$

No desenvolvimento da RNA os dados são divididos em 3 subconjuntos para realização da validação cruzada. O conjunto de treinamento é usado para o treinamento da RNA; o conjunto de validação é utilizado para avaliação do modelo gerado e como um dos critérios de parada do treinamento; e o conjunto de testes é usado para avaliar a generalização do modelo, sendo importante para a identificação de *overfitting*, ou sobre-ajuste. Os dados obtidos no mês

de março, período tomado para o desenvolvimento, somaram 358 conjuntos de amostras, que foram divididas em 70% para o conjunto de treinamento, 15% para validação e 15% para o conjunto de testes.

A divisão dos conjuntos de treinamento, validação e testes foi feita de forma aleatória a cada 5 ciclos de treinamento. O treinamento de uma RNA usando o algoritmo de LM e uma RNA usando o algoritmo do SCG é chamado de um ciclo. Para cada divisão aleatória dos dados são realizados 5 ciclos, sendo então feita uma nova divisão aleatória.

Para iniciar o treinamento da RNA são definidos parâmetros para determinar a interrupção da modelagem. Foram definidos o limite de 1.000 iterações, um valor arbitrariamente alto; 6 iterações sem melhorias, valor padrão do método no MATLAB e um critério importante para evitar sobre-ajuste; um objetivo de performance igual a 0, já que quanto mais próximo de 0 a média dos quadrados dos erros está, melhor é a performance do modelo; e gradiente mínimo de 10^{-9} , valor padrão do algoritmo no MATLAB. O tempo de duração do treinamento não foi usado como critério de parada, para que mesmo um treinamento demorado pudesse ser avaliado e interrompido pelos mesmos critérios de um treinamento mais rápido.

O número de 10 neurônios na camada oculta também é o valor padrão da ferramenta utilizada, porém essa quantidade foi alterada após o experimento descrito a seguir. Foi definida uma divisão fixa dos conjuntos de treinamento, validação e testes, com a qual foram realizados repetidos treinamentos com diferentes quantidades de neurônios na camada oculta. Visto que os pesos iniciais são aleatórios e exercem forte influência sobre a performance da RNA treinada, os resultados foram variados. Para a determinação da quantidade de neurônios a serem utilizados, foi usado como parâmetro de comparação o coeficiente de correlação linear encontrado para o conjunto de testes, que indica a performance da rede em um conjunto com dados novos. Após o procedimento, ficou definido em 20 o número de neurônios na camada oculta a ser utilizado no desenvolvimento do *soft sensor*.

3.4. *Random Forest*

Para modelar a RF foi usada a ferramenta *regressionLearner*, do MATLAB. Nessa ferramenta são definidos o tamanho mínimo dos nós folha e a quantidade de árvores formadas. Depois de passar os dados de entrada juntamente com os resultados observados e treinar o modelo pela primeira vez, o algoritmo gerado é exportado para a área de trabalho do MATLAB,

onde o treinamento pode ser refeito, assim como a predição de novas amostras. A função de avaliação da performance da RF foi a mesma utilizada para a RNA, a MSE, definida na Equação 3.4.

Os dados utilizados para o treinamento das RF foram formados pela união dos conjuntos de treinamento e validação utilizados no desenvolvimento das RNA, enquanto o conjunto de testes utilizado foi o mesmo utilizado no desenvolvimento das RNA. Para uma comparação mais equiparável, cada divisão de dados entre os três conjuntos foi utilizada para o treinamento de um mesmo número de RNA e de RF.

O valor padrão definido pela ferramenta para o tamanho mínimo dos nós folha é 5. Kruppa *et al.* (2013) ressaltam que nós finais menores tendem a explorar mais as informações disponíveis, com o prejuízo de ter um custo computacional maior. Além disso, árvores com folhas de tamanho 1 tendem a apresentar *overfit*. Diante disto, foi escolhido o tamanho mínimo igual a 3 para o modelo gerado, que corresponde a aproximadamente 1% das amostras, já que o conjunto de testes é composto por 85% do total de amostras para o mês de março (358). Já para o número de árvores formadas é esperado que o uso de um número superior ao necessário não seja prejudicial ao modelo, porém aumenta consideravelmente o custo computacional. Dessa forma, foi definida a criação de 500 árvores, a exemplo de outros trabalhos, como Kruppa *et al.* (2013), Biau e Scornet (2016) e Noi e Kappas (2018).

4. Resultados e Discussão

Neste capítulo são apresentados os resultados encontrados na redução de dimensionalidade e os desempenhos dos *soft sensors* desenvolvidos com RNA e com RF a partir dos dados obtidos pela seleção de variáveis e pela extração de características. Em seguida, são apresentados os resultados da aplicação dos *soft sensors* na predição da concentração de ferro para os dados referentes a outros meses.

4.1. *RReliefF*

O algoritmo *RReliefF* recebe como parâmetros de entrada as variáveis de entrada, as variáveis de saída e um parâmetro k , que indica quantas instâncias vizinhas serão avaliadas para a determinação do peso da variável de entrada. Robnik-Sikonja e Kononenko (2003) enfatizam que esse parâmetro é dependente do problema e especialmente relacionado à sua complexidade. Uma das recomendações é o uso de um valor inicial igual a 10, sendo ajustado de acordo com os resultados encontrados. Seguindo essa orientação, chegou-se a um valor de $k = 53$. O valor de corte do peso de importância do preditor entre as variáveis utilizadas e não utilizadas deve ser definido durante a implementação do algoritmo, sendo um valor em que seja possível observar uma diferença entre os dois grupos. O valor definido como corte foi 0. Na Figura 4.1 estão apresentados os pesos das 21 variáveis de entrada, selecionadas na Seção 3.2, em ordem decrescente. Nela, é possível definir uma separação entre as variáveis com pesos positivos daquelas com pesos negativos, sendo esse o critério de separação escolhido para a seleção das variáveis utilizadas. Com isso, 15 das 65 variáveis de entrada iniciais foram selecionadas para utilização na modelagem do *soft sensor*, identificadas abaixo na ordem decrescente de pesos:

1. Densidade da Caixa de Recirculação;
2. Vazão de Amina;
3. Vazão Volumétrica da Caixa de Recirculação;
4. Densidade da Polpa na Caixa de Alimentação;
5. Nível de Espuma do Recleaner;
6. Vazão de Amido;

7. Nível de Espuma do Rougher 2;
8. Nível da Caixa de Alimentação;
9. Abertura da Válvula Dardo 1 do Cleaner 1;
10. Nível de Espuma do Scavenger 1 Banco 2;
11. Nível de Espuma do Rougher 1;
12. Nível de Espuma do Scavenger 1 Banco 3;
13. Vazão Volumétrica da Caixa de Alimentação;
14. Vazão de Ar do Scavenger 2 Banco 3;
15. Nível de Espuma do Cleaner 1.

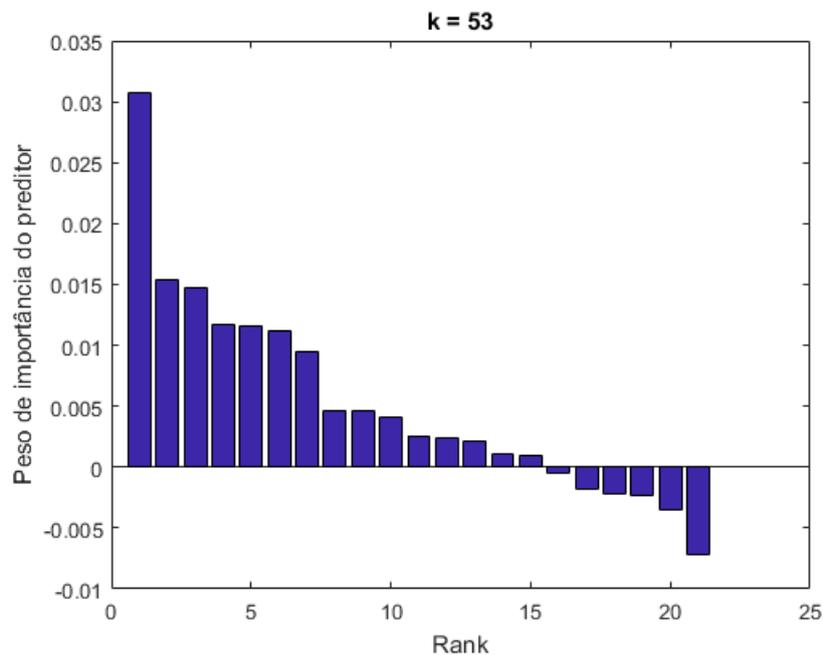


Figura 4.1: Peso das variáveis de entrada por ordem decrescente, onde as variáveis com peso positivo foram selecionadas e as com peso negativo foram removidas

Fonte: o autor

O conjunto de variáveis selecionadas foi enviado para validação da equipe responsável por sistemas de controle da flotação da usina, tendo na resposta os seguintes destaques, apresentados em conjunto com as réplicas enviadas:

- Todas as variáveis da alimentação são de grande importância e não devem ser desconsideradas. A exceção é a vazão mássica, que é resultado de um cálculo envolvendo a vazão volumétrica e a densidade;
 - Com a metodologia adotada, apenas a vazão mássica foi removida dentre as variáveis da alimentação e da recirculação.
- As válvulas dardo são importantes por terem relação direta com o nível de espuma;
 - Devido a essa relação direta, algumas variáveis de válvulas dardo não foram selecionadas pelo algoritmo, o que pode ser explicado pelo nível de espuma ser suficiente para representar a variabilidade. Outra razão da remoção de variáveis de válvulas dardo foi a quantidade de dados faltantes. No entanto, o nível de espuma continuou presente nestes casos, de forma que a informação não foi perdida.
- Seria recomendada a avaliação de alguma variável de entrada de ar no Rougher 2;
 - A vazão de ar no Rougher 2 não foi utilizada por apresentar uma grande correlação (0,98) com a vazão de ar do Rougher 1.
- Não descartaria um banco (célula de flotação) inteiro, como no caso do Cleaner 2;
 - As variáveis do Cleaner 2 não foram incluídas por apresentarem correlação muito forte com as respectivas variáveis do Cleaner 1, de forma que sua inclusão estaria apenas acrescentando dados redundantes. O mesmo pode ser dito para as células Scavenger que não possuem variáveis selecionadas.

Foi notada uma preocupação do engenheiro responsável por sistemas de controle da usina sobre o modelo estar considerando certas variáveis ou células como irrelevantes para o processo. No entanto, o *soft sensor* precisa das variáveis cuja variação permite uma melhor predição do resultado final, fato que não implica em maior ou menor importância no processo frente às demais variáveis e/ou etapas. Além disso, variáveis que contribuem agregando o mesmo tipo de informação, as variáveis altamente correlacionadas, não melhoram o desempenho do *soft sensor*, podendo inclusive piorá-lo.

Com os devidos esclarecimentos e as dúvidas sanadas, a seleção das variáveis foi validada pela equipe de controle supracitada e os experimentos foram continuados.

4.2. PCA

A função de aplicação do PCA recebe como entrada apenas os dados que formarão a entrada do *soft sensor*, sem a necessidade dos dados de saída. É possível especificar também o número de componentes principais retornados, porém essa opção não foi utilizada para que toda a variância dos dados originais pudesse ser explorada, com o PCA retornando 21 componentes principais.

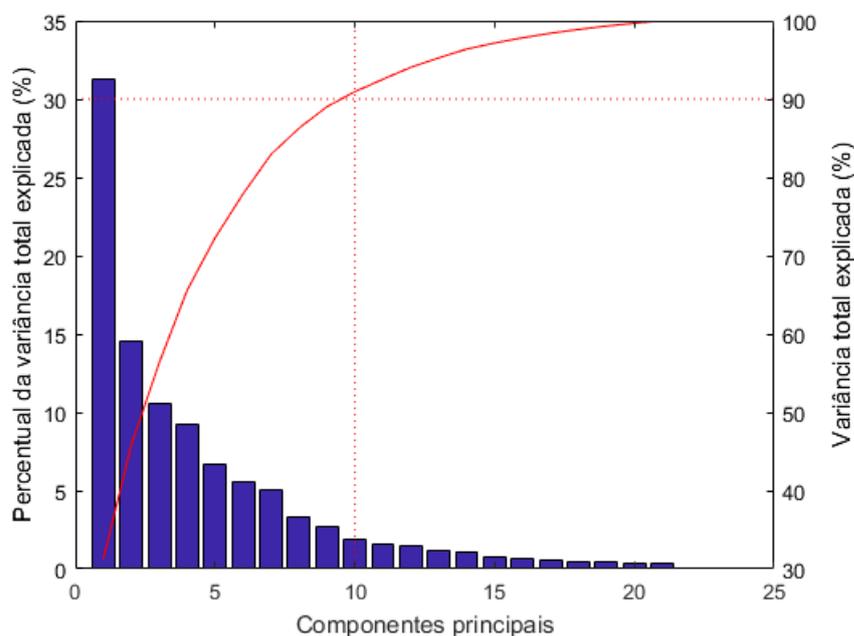


Figura 4.2: Percentual de variância explicada por cada componente principal e pela somatória dos componentes

Fonte: o autor

Os componentes principais buscam explicar a variabilidade dos dados originais e são ordenados em ordem decrescente de variância, sendo o primeiro componente principal responsável pela maior variabilidade nos dados e cada componente seguinte tem a maior variabilidade sob a restrição de ser ortogonal aos componentes anteriores. Na Figura 4.2 estão apresentados os percentuais da variância total explicada por cada componente principal. Na figura, o primeiro componente principal explica pouco mais de 30% da variância total, o segundo, explica quase 15% e assim por diante. Seus valores estão indicados no eixo vertical à esquerda. A linha sólida vermelha representa o somatório da variância explicada pelos componentes à esquerda de cada um de seus pontos. Dessa forma, em seu primeiro ponto, ela representa o valor correspondente ao primeiro componente principal, enquanto em seu segundo ponto, ela representa a soma dos

valores do primeiro e do segundo componente, sendo cerca de 46%. Seus valores estão indicados no eixo vertical à direita. Foram selecionados para o desenvolvimento do *soft sensor* os 10 primeiros componentes principais, que conseguem explicar mais de 90% da variância dos dados originais. Como tais componentes são resultado da transformação dos dados originais, não é possível fazer uma associação direta entre os componentes selecionados e as variáveis do processo da flotação.

4.3. RNA

A Rede Neural gerada pode ser representada pela Figura 4.3, que possui 15 nós de entrada (determinados na Seção 4.1), 20 neurônios na camada oculta e 1 neurônio na camada de saída (definidos na Seção 3.3.1).

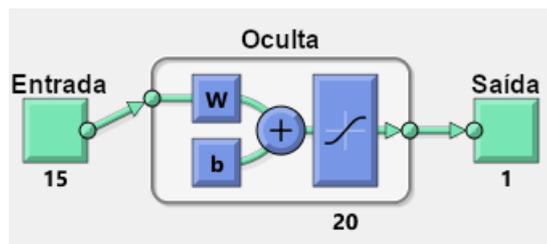


Figura 4.3: Representação da Rede Neural gerada

Fonte: o autor

Foram realizados 1.000 treinamentos da RNA utilizando o algoritmo de LM e 1.000 utilizando o SCG tanto para os dados obtidos pelo *RReliefF* quanto para os dados obtidos pelo PCA. Para comparar os métodos, foram utilizados o MSE e o coeficiente de correlação linear R para o conjunto de testes, onde a rede treinada é aplicada para a predição de um conjunto de dados ao qual ela não foi exposta. Num contexto geral, a rede treinada com o algoritmo de LM teve um desempenho melhor que a treinada com o SCG.

4.3.1. Coeficiente de Correlação Linear

Na Tabela 4.1 são mostrados os 5 melhores resultados do valor de R para o conjunto de testes com cada um dos algoritmos, onde um valor mais próximo de 1 representa uma correlação linear maior, conseqüentemente uma melhor aproximação entre o valor esperado e o valor da predição. É possível notar que a rede treinada com LM obteve valores de R mais próximos de 1 que os da rede com SCG, tanto nos dados obtidos pelo *RReliefF* quanto nos obtidos pelo PCA.

Tabela 4.1: Maiores valores de R para o conjunto de testes das RNA utilizando os algoritmos LM e SCG sobre os dados do *RReliefF* e do PCA

	<i>RReliefF</i>		PCA	
	LM	SCG	LM	SCG
1	0,773509	0,669	0,715647	0,553992
2	0,772578	0,648349	0,707473	0,53162
3	0,771459	0,640266	0,69202	0,510408
4	0,770209	0,63914	0,688717	0,507486
5	0,750907	0,617288	0,685685	0,500885

Fonte: o autor

Na Tabela 4.2 são mostradas as quantidades de resultados obtidos com R maior que 0,7 e maior que 0,6 para cada um dos algoritmos de treinamento. Essas marcas de referência foram escolhidas pelo autor para fins de comparação e análise dos resultados. É possível notar uma quantidade maior de resultados melhores (R mais próximo de 1) utilizando o algoritmo LM em relação ao SCG entre os treinamentos realizados tanto nos dados obtidos pelo *RReliefF* quanto nos obtidos pelo PCA. Vale destacar, ainda, que os resultados obtidos usando os dados provenientes do *RReliefF* tiveram desempenho levemente superior aos obtidos com os dados gerados pelo PCA. Isso é visto mais claramente pelos resultados dos treinamentos utilizando SCG, que não foram capazes de alcançar níveis de desempenho equivalentes aos demais resultados quando usando os dados do PCA.

Tabela 4.2: Quantidade de resultados acima dos valores definidos para R nas RNA utilizando os algoritmos LM e SCG sobre os dados do *RReliefF* e do PCA

	<i>RReliefF</i>		PCA	
	LM	SCG	LM	SCG
R maior que 0,7	13	0	2	0
R maior que 0,6	50	8	15	0

Fonte: o autor

Na Figura 4.4 podem ser vistos os resultados das regressões no conjunto de treinamento, validação, testes e em todos os dados do treinamento da rede com maior R no conjunto de testes da RNA, que foi realizado com os dados do *RReliefF* usando o algoritmo de LM. Nessa figura, é possível observar o ajuste das predições nos diferentes conjuntos de dados. No conjunto de treinamento, o coeficiente de correlação linear foi igual a 0,88569. No conjunto de validação, foi de 0,21095 e no conjunto de testes, 0,77021. Quando todos os dados são analisados em

conjunto, o coeficiente foi $R = 0,78218$. Essa RNA obteve MSE igual a 1,2395.

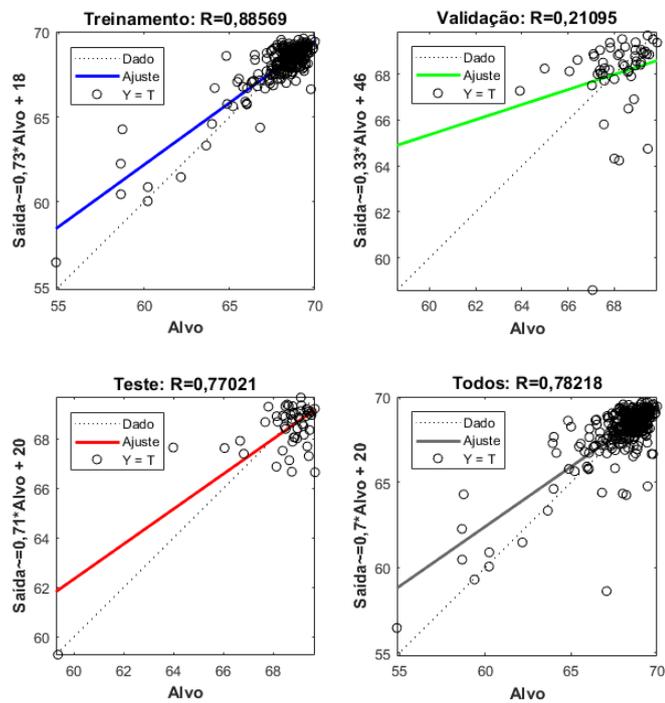


Figura 4.4: Resultado da regressão do treinamento com maior R no conjunto de testes da RNA com os dados do *RReliefF* usando o algoritmo de LM

Fonte: o autor

4.3.2. MSE

Na Tabela 4.3, são mostrados os 5 melhores resultados do MSE, onde um valor mais próximo de 0 representa uma distância (ou erro) menor entre o valor esperado e o valor da predição.

Tabela 4.3: Menores valores do MSE para RNA utilizando os algoritmos LM e SCG sobre os dados do *RReliefF* e do PCA

	<i>RReliefF</i>		PCA	
	LM	SCG	LM	SCG
1	0,749783	1,577013	0,926836	2,260513
2	0,857617	1,795156	1,358046	2,278756
3	0,957877	1,857298	1,401384	2,295989
4	1,018666	1,869756	1,439002	2,364107
5	1,020206	1,90809	1,475373	2,3946

Fonte: o autor

Na Tabela 4.4 são mostradas as quantidades de resultados obtidos com MSE menor que 1 e menor que 2 para cada um dos algoritmos de treinamento. Essas marcas de referência foram escolhidas pelo autor para fins de comparação e análise dos resultados.

Tabela 4.4: Quantidade de resultados abaixo dos valores definidos para MSE nas RNA utilizando os algoritmos LM e SCG sobre os dados do *RReliefF* e do PCA

	<i>RReliefF</i>		PCA	
	LM	SCG	LM	SCG
MSE menor que 1	3	0	1	0
MSE menor que 2	43	6	14	0

Fonte: o autor

Na Figura 4.5 são mostradas as regressões para o treinamento com menor MSE, que foi realizado com os dados do *RReliefF* usando o algoritmo de LM. No conjunto de treinamento, o coeficiente de correlação linear foi igual a 0,9436. No conjunto de validação, foi de 0,38838 e no conjunto de testes, 0,60117. Quando todos os dados são analisados em conjunto, o coeficiente foi $R = 0,87122$. Essa RNA obteve MSE igual a 0,749783.

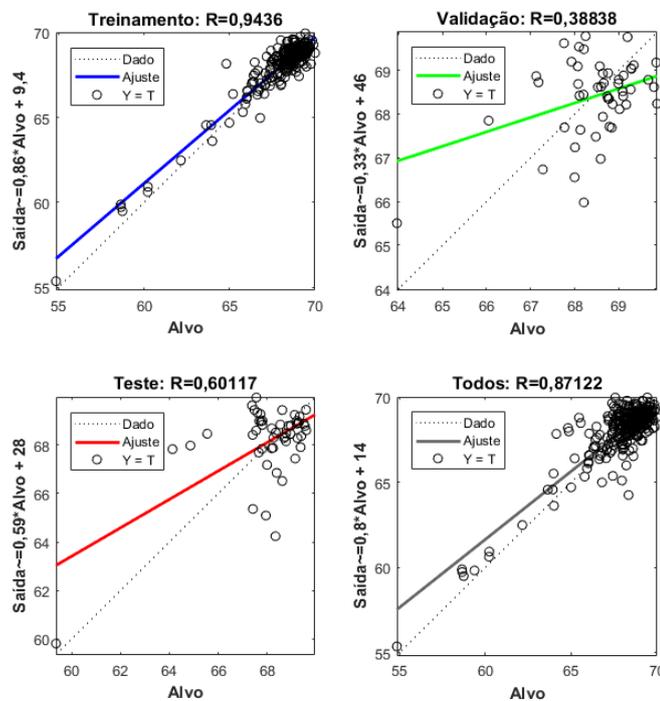


Figura 4.5: Resultado da regressão do treinamento com menor MSE da RNA com os dados do *RReliefF* usando o algoritmo de LM

Fonte: o autor

4.4. *Random Forest*

A RF gerada é formada por um conjunto de 500 árvores de decisão onde cada folha possui o tamanho mínimo igual a 3. Foram realizados 1.000 treinamentos da RF utilizando os dados obtidos pelo *RReliefF* e 1.000 utilizando os dados obtidos pelo PCA. Para comparar os métodos, foram utilizados o MSE e o coeficiente de correlação linear R para o conjunto de testes, onde a RF treinada é aplicada para a predição de um conjunto de dados ao qual ela não foi exposta. Num contexto geral, a RF treinada com os dados do PCA e a treinada com os dados do *RReliefF* tiveram desempenhos muito similares.

4.4.1. Coeficiente de Correlação Linear

Na Tabela 4.5 são mostrados os 5 melhores resultados do valor de R para o conjunto de testes com cada um dos métodos de redução de variáveis, onde um valor mais próximo de 1 representa uma correlação linear maior, conseqüentemente uma melhor aproximação entre o valor esperado e o valor da predição.

Tabela 4.5: Maiores valores de R para o conjunto de testes das RF utilizando os dados do *RReliefF* e do PCA

	<i>RReliefF</i>	PCA
1	0,74358	0,770967
2	0,740575	0,753743
3	0,734437	0,752114
4	0,731023	0,738834
5	0,726223	0,73813

Fonte: o autor

Já na Tabela 4.6 são mostradas as quantidades de resultados obtidos com R maior que 0,7 e com R maior que 0,6 para cada um dos métodos de redução de variáveis. É possível notar a proximidade dos resultados da RF usando os dois métodos. Enquanto a quantidade de resultados com R superior a 0,7 foi maior usando os dados do PCA, a quantidade de resultados superiores a 0,6 foi maior usando os dados do *RReliefF*.

Na Figura 4.6 podem ser vistos os resultados das regressões no conjunto de treinamento, testes e em todos os dados do treinamento com maior R no conjunto de testes da RF, que foi realizado com os dados do PCA. No conjunto de treinamento, o coeficiente de correlação

Tabela 4.6: Quantidade de resultados acima dos valores definidos para R nas RF utilizando os dados do *RReliefF* e do PCA

	<i>RReliefF</i>	PCA
R maior que 0,7	10	19
R maior que 0,6	77	74

Fonte: o autor

linear foi igual a 0,88099. No conjunto de testes, foi de 0,77097. Quando todos os dados são analisados em conjunto, o coeficiente foi $R = 0,86341$. Essa RF obteve MSE igual a 1,2382.

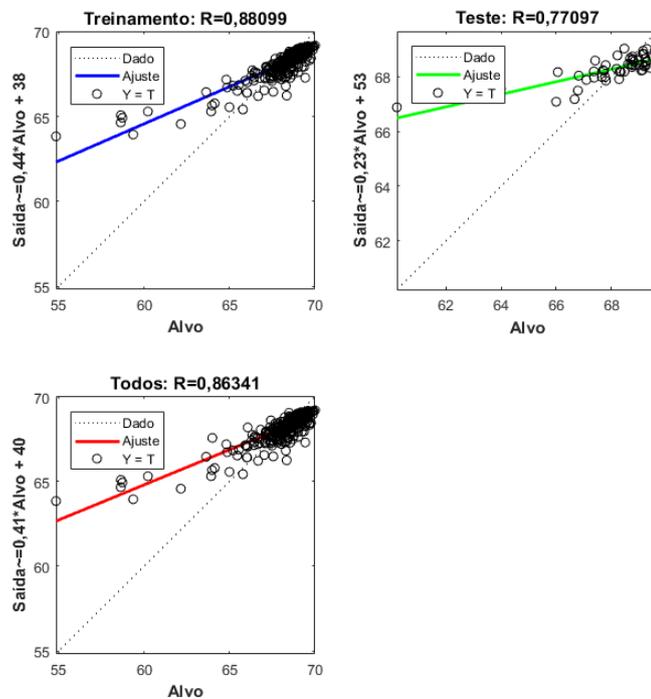


Figura 4.6: Resultado da regressão do treinamento com maior R no conjunto de testes da RF com os dados do PCA

Fonte: o autor

4.4.2. MSE

Na Tabela 4.7, são mostrados os 5 melhores resultados do MSE, onde um valor mais próximo de 0 representa uma distância (ou erro) menor entre o valor esperado e o valor da predição. É possível notar que os treinamentos usando os dados do PCA obtiveram uma correlação linear ligeiramente superior às obtidas usando os dados do *RReliefF*. Em contrapartida,

os treinamentos usando os dados do *RReliefF* alcançaram um MSE ligeiramente menor que os alcançados usando os dados do PCA.

Tabela 4.7: Menores valores do MSE para RF utilizando os dados do *RReliefF* e do PCA

	<i>RReliefF</i>	PCA
1	1,018065	1,082903
2	1,020555	1,092677
3	1,023036	1,092994
4	1,024267	1,094296
5	1,0245	1,106025

Fonte: o autor

Já na Tabela 4.8 são mostradas as quantidades de resultados obtidos com MSE menor que 1 e com MSE menor que 2 para cada um dos métodos. É possível notar a proximidade dos resultados da RF usando os dois métodos de redução de variáveis. Enquanto a quantidade de resultados com R superior a 0,7 foi maior usando os dados do PCA, a quantidade de resultados superiores a 0,6 foi maior usando os dados do *RReliefF*. A constância da abordagem da RF também pode ser observada na Tabela 4.8, em que não houve nenhum resultado com MSE menor que 1, no entanto todos os resultados alcançaram um MSE menor que 2.

Tabela 4.8: Quantidade de resultados abaixo dos valores definidos para MSE nas RF utilizando os dados do *RReliefF* e do PCA

	<i>RReliefF</i>	PCA
MSE menor que 1	0	0
MSE menor que 2	1000	1000

Fonte: o autor

Na Figura 4.7 são mostradas as regressões para o treinamento com menor MSE, que foi realizado com os dados do *RReliefF*. No conjunto de treinamento, o coeficiente de correlação linear foi igual a 0,89788. No conjunto de testes, foi de 0,43437. Quando todos os dados são analisados em conjunto, o coeficiente foi $R = 0,87808$. Essa RF obteve MSE igual a 1,0181.

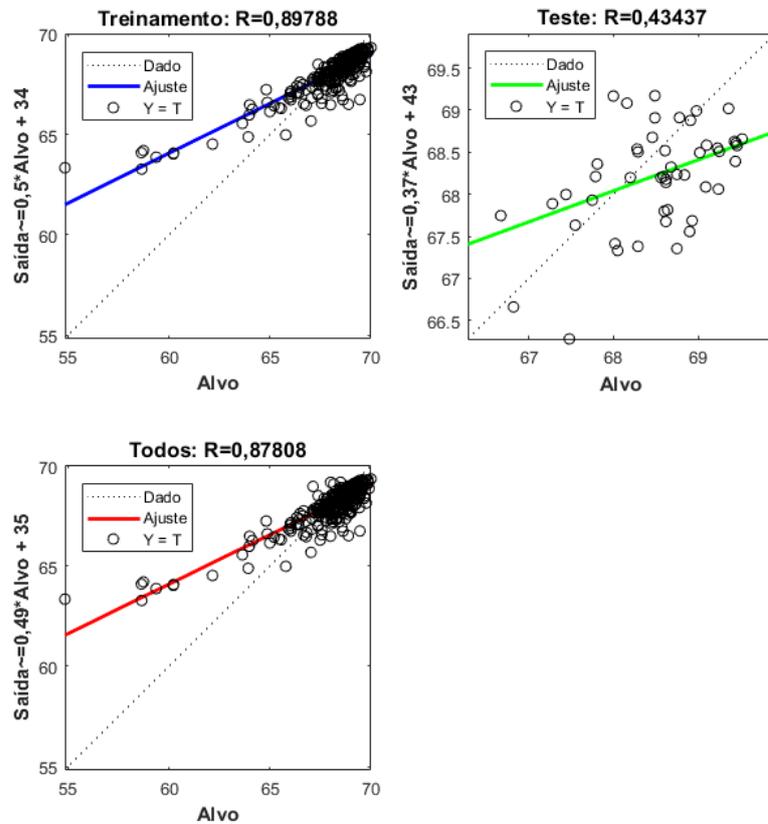


Figura 4.7: Resultado da regressão do treinamento com menor MSE no conjunto de testes da RF com os dados do *RReliefF*

Fonte: o autor

4.5. Aplicação do *Soft Sensor*

Os melhores modelos gerados usando RNA e RF foram usados na predição da concentração de ferro em outros conjuntos de dados históricos, como se segue.

4.5.1. RNA

As RNA representadas nas Figuras 4.4 e 4.5 foram utilizadas para realizar a predição nos dados dos meses de janeiro a maio, pois apresentaram o maior R no conjunto de testes e o menor MSE, respectivamente. Ambas as RNA usaram dados provenientes do *RReliefF*. Considerando esse período, a RNA que obteve menor erro médio absoluto (MAE) foi a da Figura 4.5 (menor MSE), sendo então o modelo adotado pelo *soft sensor* usando RNA. Para a avaliação do desempenho dos modelos na aplicação do *soft sensor* em novos dados foi utilizado o MAE, dada a sua facilidade e rapidez de compreensão por estar na mesma escala e unidade da

predição. Na Figura 4.8 estão representados os dados previstos pelo *soft sensor* usando a RNA e os dados obtidos em laboratório para março, mês usado no treinamento da rede e em que o MAE obtido foi de 0,6132.

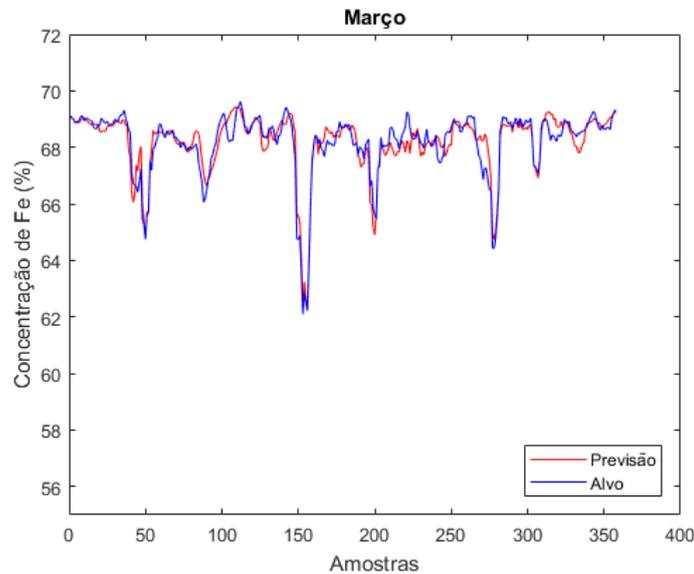


Figura 4.8: Teor de ferro em % no mês utilizado para treinamento obtida em laboratório e obtida pelo *soft sensor* usando RNA, com MAE de 0,6132

Fonte: o autor

Os dados obtidos em laboratório, confrontados com as previsões realizadas pelo *soft sensor* usando RNA, são apresentados para os meses de janeiro, com MAE de 1,8890, fevereiro, MAE de 2,3933, abril, MAE de 2,1705, e maio, com MAE de 2,1016, na Figura 4.9. Como esperado, o MAE em dados novos é maior que o obtido no conjunto de treinamento, porém ainda se encontram em uma faixa relativamente baixa e aceitável em uma versão inicial do *soft sensor*. Ainda que tenham aparecido picos na previsão, o erro médio continuou baixo e uma indicação da tendência de subida ou descida da concentração de ferro podem ser observadas nos gráficos, o que fornecem uma referência para a operação.

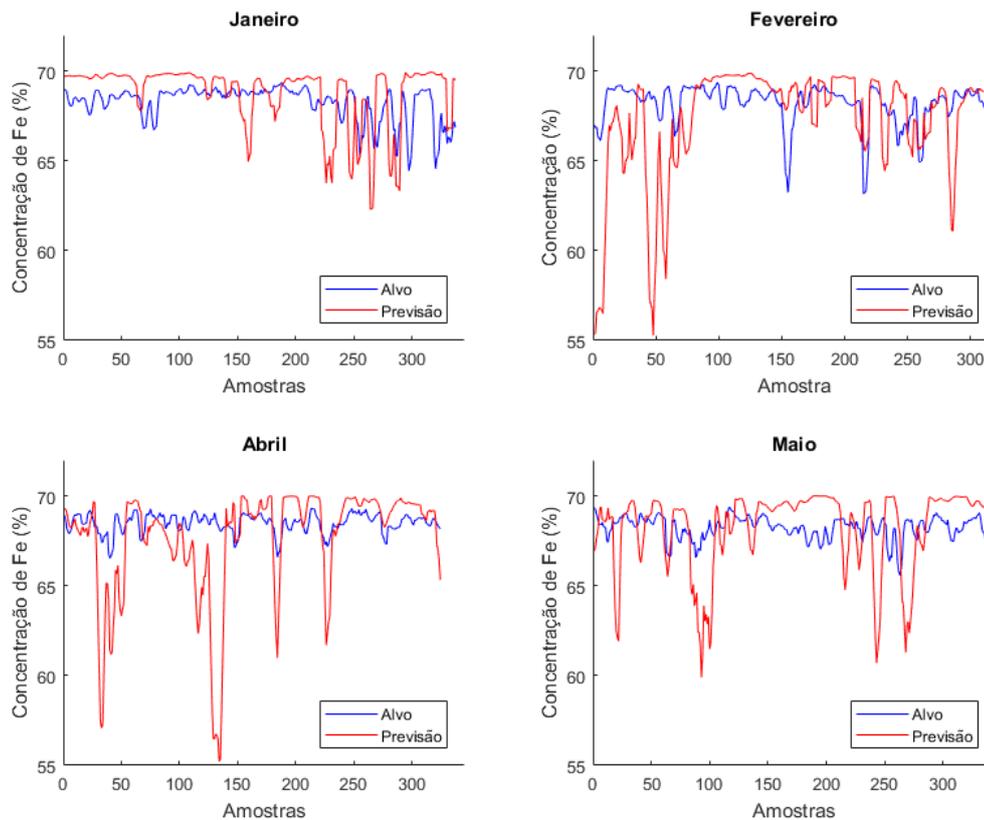


Figura 4.9: Teor de ferro em % nos meses de janeiro, fevereiro, abril e maio obtida em laboratório e obtida pelo *soft sensor* usando RNA
 Fonte: o autor

4.5.2. *Random Forest*

As RF representadas nas Figuras 4.6 e 4.7 foram usadas para realizar a predição nos dados dos meses de janeiro a maio, pois apresentaram o maior R no conjunto de testes e o menor MSE, respectivamente. A primeira RF utilizou os dados obtidos pelo PCA, enquanto a segunda utilizou dados selecionados pelo *RReliefF*. Considerando esse período, as RF obtiveram resultados extremamente parecidos, porém o MAE da RF da Figura 4.6 (maior R), que usou os dados resultantes do PCA, foi ligeiramente menor, sendo então o modelo adotado pelo *soft sensor* usando RF. Na Figura 4.10 estão representados os dados previstos pelo *soft sensor* usando RF e os dados obtidos em laboratório para o mês de março, usados no treinamento da RF, mês em que o MAE obtido foi de 0,5904.

Os dados obtidos em laboratório, confrontados com as previsões realizadas pelo *soft sensor* usando RF, são apresentados para os meses de janeiro, com MAE de 0,9645, fevereiro, MAE de 1,0625, abril, MAE de 1,3007, e maio, com MAE de 1,1816, na Figura 4.11. Da

mesma forma que na RNA, o MAE em dados novos é maior que o obtido no conjunto de treinamento, porém ainda pode ser considerado baixo, além de ter sido aproximadamente a metade do obtido utilizando RNA. Ademais, não foram gerados picos de previsão da concentração de ferro, com previsões permanecendo mais constantes e alinhadas aos resultados laboratoriais. Ainda que picos súbitos na observação da concentração não tenham sido previstos pelo *soft sensor*, a previsão foi bem próxima ao alvo, de um modo geral.

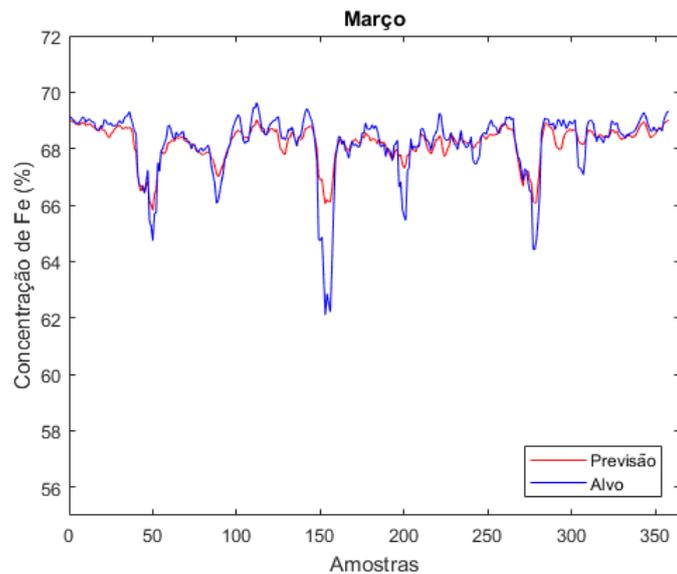


Figura 4.10: Teor de ferro em % no mês utilizado para treinamento obtida em laboratório e obtida pelo *soft sensor* usando RF, com MAE de 0,5904

Fonte: o autor

Vale ressaltar que a etapa da manutenção do *soft sensor* não foi realizada, pois é esperado que seja necessária apenas no decorrer do tempo, conforme as características do processo se alterem e a performance do *soft sensor* sofra prejuízos, para que sua calibração seja ajustada e a sua performance se mantenha nos níveis esperados no momento de sua implantação.

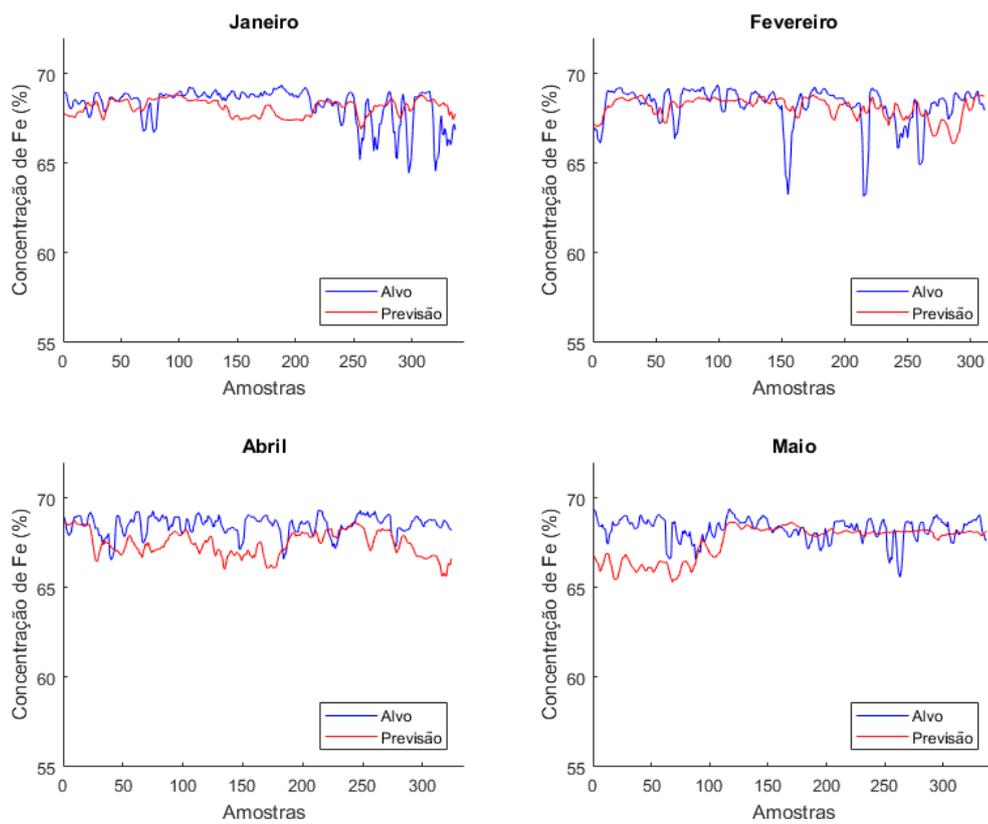


Figura 4.11: Teor de ferro em % nos meses de janeiro, fevereiro, abril e maio obtida em laboratório e obtida pelo *soft sensor* usando RF

Fonte: o autor

5. Conclusões

Em parceria com a Vale, foram coletados dados de uma das linhas de flotação de minério de ferro da usina de Conceição, em Itabira, Minas Gerais. Os dados foram tratados e os que foram selecionados para o desenvolvimento do *soft sensor* passaram por duas técnicas distintas de redução de dimensionalidade. O *RReliefF* foi usado para a técnica de seleção de variáveis e o PCA foi usado para a técnica de extração de características. De acordo com os critérios utilizados, o PCA obteve um conjunto de dados de dimensão menor, com 10 variáveis contra 15 do *RReliefF*. No entanto, os resultados do desenvolvimento sobre ambas as metodologias foram bastante similares, de forma que a escolha do método de redução de dimensionalidade não deve ser um fator decisivo na construção desse *soft sensor*. Vale ressaltar que o PCA realiza uma transformação linear dos dados para determinar os principais componentes e essa transformação deve ser aplicada a novos dados antes que estes possam ser usados para predição no modelo desenvolvido. Dessa forma, é preciso ter disponíveis todas as variáveis utilizadas na primeira transformação, 21 no caso deste trabalho. Por outro lado, o *RReliefF* é aplicado apenas uma vez, necessitando a disponibilidade apenas das variáveis selecionadas, que nesse trabalho foram 15.

Utilizando cada um dos conjuntos de dados resultantes, foram desenvolvidos modelos de RNA com o algoritmo de LM e com o algoritmo SCG. Num contexto geral, a RNA treinada com o algoritmo de LM teve um desempenho melhor que a treinada com o SCG. Usando os resultados com o LM para comparar os conjuntos de dados, os modelos que utilizaram os dados do *RReliefF* foram levemente superiores, porém os resultados mostram que o PCA não deixa de ser uma opção viável para essa aplicação.

No desenvolvimento do modelo utilizando *Random Forest* os resultados foram mais homogêneos, sendo muito similares independentemente do uso do conjunto de dados obtido pelo *RReliefF* ou PCA. Essa constância pode ser observada pela quantidade de resultados com MSE menor que dois. Enquanto nos modelos de RNA foram encontrados menos de 50 modelos com MSE menor que dois, nos modelos de RF todos os 1000 modelos, tanto com dados do *RReliefF* quanto do PCA, apresentaram resultados nessa faixa.

Dentre as RNA selecionadas (a que apresentou maior R no conjunto de testes e a que apresentou menor MSE), a que obteve melhor desempenho nos novos dados foi a que apresentou menor MSE no treinamento. Como esperado, o MAE nos novos dados foi superior ao

encontrado nos dados de treinamento. Nos dados de janeiro, fevereiro, abril e maio, o MAE variou entre 1,8 e 2,4, contra 0,61 nos dados de março. Analogamente, dentre as RF selecionadas, a que obteve melhor desempenho nos novos dados foi a que apresentou maior R no treinamento e o MAE nos novos dados variou entre 0,9 e 1,2, contra 0,59 nos dados de março.

O engenheiro responsável por sistemas de controle da usina destacou a ciência da dificuldade de previsão em um processo tão complexo e dinâmico. Por isso, optou por não definir um objetivo específico de MAE para o *soft sensor*, além de que seja o menor possível. Dessa forma, os resultados do *soft sensor*, tanto com RNA quanto com RF, foram aceitos como bons resultados para um desenvolvimento inicial.

Observando os valores de MAE encontrados, é possível indicar a adoção do *soft sensor* com utilização de RF, pois chegou a erros menores e foi percebida uma certa facilidade no treinamento do modelo devido à repetibilidade de seus resultados.

Os resultados indicam que uma correlação entre os dados disponíveis em tempo real no processo da flotação e os resultados da análise laboratorial em intervalos de 2 horas podem ser explorados com o uso de um *soft sensor*. Uma redução no tempo de aquisição do teor de concentrado de ferro possibilita uma resposta mais rápida para correções nos parâmetros de processo da flotação, o que evita a obtenção de polpa fora da faixa especificada e, conseqüentemente, reduz desperdícios e impacto ambiental.

6. Trabalhos Futuros

A próxima etapa a ser realizada consiste da aplicação do *soft sensor* em uma linha de flotação para que o modelo possa ser avaliado na prática e identificadas possíveis vantagens e dificuldades decorrentes de sua utilização na indústria. Uma vez implementado, é necessário que se mantenha o *soft sensor* sob observação para que a etapa de manutenção possa ser realizada caso sua performance venha a cair. Em seguida, pode ser realizada uma proposta de controle dos parâmetros de processo em tempo real com a utilização do *soft sensor*.

Para outros trabalhos, é sugerida a coleta de dados sob observação direta de alterações na qualidade do minério lavrado, para que se obtenha uma melhor estratificação dos dados nesse sentido. Quando a frente de lavra é alterada, não há informações que indiquem tal alteração no conjunto de dados da flotação, ainda que isso influencie nos parâmetros de processo.

Outra possibilidade é a inserção de atraso nos dados de processo, podendo ser definidos diferentes atrasos de acordo com o posicionamento da medição em relação à caixa de produto final, visto que uma variação no início da linha pode demorar um tempo para apresentar variações no resultado do teor de ferro no concentrado.

É sugerido também o cálculo do índice de seletividade de Gaudin, que utiliza a recuperação de útil e de ganga no concentrado para avaliar o desempenho de operações de concentração e representa de forma mais abrangente a qualidade do processo (COSTA, 2009). Em conversa com o engenheiro responsável por sistemas de controle da usina, foi demonstrado interesse no índice, que relaciona qualidade e produtividade, ainda que atualmente uma avaliação semelhante não seja utilizada na planta.

Referências Bibliográficas

- 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. ISSN: 03787788. doi: 10.1016/j.enbuild.2017.04.038.
- ALVES, L. P. *Implantação de Coluna de Flotação no Laboratório de Tratamento de Minérios do CEFET/MG, Unidade Araxá e Realização de Estudo Exploratório*. Monografia, Centro Federal de Educação Tecnológica de Minas Gerais - CEFET, 2015.
- ANDREI, N. “Scaled conjugate gradient algorithms for unconstrained optimization”, *Computational Optimization and Applications*, v. 38, n. 3, pp. 401–416, 2007. ISSN: 09266003. doi: 10.1007/s10589-007-9055-7.
- AQUINO, J. A. D., OLIVEIRA, M. L. M. D., FERNANDES, M. D. “Flotação em Coluna”. Em: Luz, A. B., Sampaio, J. A., de França, S. C. A. (Eds.), *Tratamento de Minérios*, 5 ed., CETEM, pp. 515–556, Rio de Janeiro, 2010. ISBN: 8572272046.
- ARTONI, F., DELORME, A., MAKEIG, S. “A visual working memory dataset collection with bootstrap Independent Component Analysis for comparison of electroencephalographic preprocessing pipelines”, *Data in Brief*, v. 22, pp. 787–793, 2019. ISSN: 23523409. doi: 10.1016/j.dib.2018.12.022. Disponível em: <<https://doi.org/10.1016/j.dib.2018.12.022>>.
- BERGH, L. G., YIANATOS, J. B. “Flotation column automation: state of the art”, *Control Engineering Practice*, v. 11, n. 1, pp. 67–72, 2003. ISSN: 09670661. doi: 10.1016/S0967-0661(02)00093-X.
- BERGH, L. G., YIANATOS, J. B. “The long way toward multivariate predictive control of flotation processes”, *Journal of Process Control*, v. 21, n. 2, pp. 226–234, 2011. ISSN: 0959-1524. doi: 10.1016/j.jprocont.2010.11.001. Disponível em: <<http://dx.doi.org/10.1016/j.jprocont.2010.11.001>>.
- BIAU, G., SCORNET, E. “A random forest guided tour”, *Test*, v. 25, n. 2, pp. 197–227, 2016. ISSN: 11330686. doi: 10.1007/s11749-016-0481-7.

- BOUTIN, P., TREMBLAY, R. J. “Method and apparatus for the separation and recovery of ores”, *Canadian patent*, n. 694547, 1964.
- BREIMAN, L. “Random Forests”, *Machine Learning*, v. 45, pp. 5–32, 2001. doi: 10.1007/978-3-662-56776-0_10.
- CAMPOS, T. E. D. *Técnicas de Seleção de Características com Aplicações em Reconhecimento de Faces*. Dissertação de Mestrado, Universidade de São Paulo, 2001.
- CAPPONI, F. *Avanços na Recuperação de Finos de Minérios pelo Processo de Flotação “Extensora”, O Caso da Mina de Chuquicamata*. Dissertação de Mestrado, Universidade Federal do Rio Grande do Sul, 2005.
- CHAVES, A. P. “Máquinas de Flotação”. Em: Chaves, A. P. (Ed.), *Flotação: O Estado da Arte no Brasil*, Signus Editora, São Paulo, 2006.
- CHAVES, A. P., LEAL FILHO, L. D. S., BRAGA, P. F. A. “Flotação”. Em: da Luz, A. B., França, S. C. A., Braga, P. F. A. (Eds.), *Tratamento de Minérios*, 6 ed., CETEM/MC-TIC, cap. 10, pp. 403–459, Rio de Janeiro, 2018.
- COSTA, J. L. *Seletividade na Flotação de Minérios Ferríferos Dolomíticos*. Dissertação de Mestrado, Universidade Federal de Ouro Preto, 2009.
- DIETTERICH, T. G. “Ensemble Methods in Machine Learning”. Em: *Multiple Classifier Systems*, pp. 1–15, Berlin, Heidelberg, 2000. Springer Berlin Heidelberg. ISBN: 978-3-540-45014-6.
- DRUMOND, P., KAPPES, D., DE MORAES, C., et al.. “Soft Sensor: Machine Learning Tradicional Ou Deep Learning?” Em: *Seminário de Automação e TI*, pp. 231–242, São Paulo, 2018. doi: 10.5151/2237-0234-31915.
- FARROKHPAY, S. “The significance of froth stability in mineral flotation - A review”, *Advances in Colloid and Interface Science*, v. 166, n. 1-2, pp. 1–7, 2011. ISSN: 00018686. doi: 10.1016/j.cis.2011.03.001. Disponível em: <<http://dx.doi.org/10.1016/j.cis.2011.03.001>>.
- FORTUNA, L., GRAZIANI, S., RIZZO, A., et al.. *Soft Sensors for Monitoring and Control of Industrial Processes*. Londres, Springer, 2007. ISBN: 978-1-84628-480-9.
- GONZAGA, J. C. B., MELEIRO, L. A. C., KIANG, C., et al.. “ANN-based soft-sensor for real-time process monitoring and control of an industrial polymerization process”, *Computers and Chemical Engineering*, v. 33, pp. 43–49, 2009. doi: 10.1016/j.compchemeng.2008.05.019.

- GONZALEZ, G. D. “Soft sensors for processing plants”, *Proceedings of the 2nd International Conference on Intelligent Processing and Manufacturing of Materials, IPMM 1999*, v. 1, pp. 59–69, 1999. doi: 10.1109/IPMM.1999.792454.
- HALL, M. A. *Correlation-based Feature Selection for Machine Learning*. Tese de Doutorado, The University of Waikato, 1999.
- HAYKIN, S. *Neural Networks - A Comprehensive Foundation*. 9 ed. Hamilton, Pearson Education, 2005. ISBN: 81-7808-300-0.
- HE, X., CAI, D., NIYOGI, P. “Laplacian Score for feature selection”, *Advances in Neural Information Processing Systems*, pp. 507–514, 2005. ISSN: 10495258.
- KADLEC, P., GABRYS, B., STRANDT, S. “Data-driven soft sensors in the process industry”, *Computers and Chemical Engineering*, v. 33, n. 4, pp. 795–814, 2009. ISSN: 00981354. doi: 10.1016/j.compchemeng.2008.12.012.
- KADLEC, P., GRBIĆ, R., GABRYS, B. “Review of adaptation mechanisms for data-driven soft sensors”, *Computers and Chemical Engineering*, v. 35, n. 1, pp. 1–24, 2011. ISSN: 00981354. doi: 10.1016/j.compchemeng.2010.07.034.
- KENOBI, K., ATKINSON, J. A., WELLS, D. M., et al. “Linear discriminant analysis reveals differences in root architecture in wheat seedlings related to nitrogen uptake efficiency”, *Journal of Experimental Botany*, v. 68, n. 17, pp. 4969–4981, 2017. ISSN: 14602431. doi: 10.1093/jxb/erx300.
- KIRA, K., RENDELL, L. A. “A Practical Approach to Feature Selection”. Em: *Machine Learning Proceedings 1992*, Morgan Kaufmann Publishers, Inc., pp. 249–256, Aberdeen, Scotland, 1992. doi: 10.1016/b978-1-55860-247-2.50037-1. Disponível em: <<http://dx.doi.org/10.1016/B978-1-55860-247-2.50037-1>>.
- KO, Y. D., SHANG, H. “A neural network-based soft sensor for particle size distribution using image analysis”, *Powder Technology*, v. 212, n. 2, pp. 359–366, 2011. ISSN: 00325910. doi: 10.1016/j.powtec.2011.06.013.
- KOHAVI, R., JOHN, G. H. “Wrappers for feature subset selection”, *Artificial Intelligence*, n. 97, pp. 273–324, 1997. ISSN: 03029743. doi: 10.1007/978-3-642-39038-8-27.
- KOHAVI, R., SOMMERFIELD, D. “Feature subset selection using the wrapper method: Overfitting and dynamic search space topology”, *First International Conference on Knowledge Discovery and Data Mining*, n. June, pp. 192–197, 1995. Disponível em: <<http://www.aaai.org/Papers/KDD/1995/KDD95-049.pdf>>.

- KONONENKO, I., ŠIMEC, E., ROBNIK-ŠIKONJA, M. “Overcoming the Myopia of Inductive Learning Algorithms with RELIEFF”, *Applied Intelligence*, v. 7, n. 1, pp. 39–55, 1997. ISSN: 0924669X. doi: 10.1023/A:1008280620621.
- KRUPPA, J., SCHWARZ, A., ARMINGER, G., et al.. “Consumer credit risk: Individual probability estimates using machine learning”, *Expert Systems with Applications*, v. 40, n. 13, pp. 5125–5131, 2013. ISSN: 09574174. doi: 10.1016/j.eswa.2013.03.019. Disponível em: <<http://dx.doi.org/10.1016/j.eswa.2013.03.019>>.
- KULCSÁR, T. *Data mining and machine learning algorithms for soft sensor development*. Tese de Doutorado, University of Pannonia, 2016.
- LARIVIÈRE, B., VAN DEN POEL, D. “Predicting customer retention and profitability by using random forests and regression forests techniques”, *Expert Systems with Applications*, v. 29, n. 2, pp. 472–484, 2005. ISSN: 09574174. doi: 10.1016/j.eswa.2005.04.043.
- LI, Z., YANG, Y., LIU, J., et al.. “Unsupervised feature selection using nonnegative spectral analysis”, *Proceedings of the National Conference on Artificial Intelligence*, v. 2, pp. 1026–1032, 2012.
- LIMA, N. P., PERES, A. E. C., GONÇALVES, T. A. R. “Comparative evaluation between mechanical and pneumatic cells for quartz flotation in the iron ore industry”, *REM - International Engineering Journal*, v. 71, n. 3, pp. 437–442, 2018. doi: 10.1590/0370-44672016710179.
- LIN, B., RECKE, B., KNUDSEN, J. K., et al.. “A systematic approach for soft sensor development”, *Computers and Chemical Engineering*, v. 31, n. 5-6, pp. 419–425, 2007. ISSN: 00981354. doi: 10.1016/j.compchemeng.2006.05.030.
- LOTUFO, F., GARCIA, C. “Sensores Virtuais ou Soft Sensors: Uma Introdução”, *7 th Brazilian Conference on Dynamics, Control and Applications*, p. 9, 2008.
- LUZ, J. A. M. D. “Flotation of Iron Ore”. Em: Totten, G. E., Colas, R. (Eds.), *Encyclopedia of Iron, Steel, and Their Alloys*, 1 ed., CRC Press, pp. 1249–1288, Boca Raton, 2016. doi: 10.1081/E-EISA-120049089. Disponível em: <<http://www.crcnetbase.com/doi/10.1081/E-EISA-120049089>>.
- MAMEDE, C. R. S. *Mitigação dos Efeitos de Cátions Ca²⁺ na Flotação Catiônica Reversa de Minério de Ferro*. Dissertação de Mestrado, Universidade Federal de Minas Gerais, 2016.

- MARAIS, C. *Estimation of concentrate grade in platinum flotation based on froth image analysis*. Dissertação de Mestrado, University of Stellenbosch, 2010. Disponível em: <<http://scholar.sun.ac.za/handle/10019.1/5346>>.
- MAZON, R. *Efeito do tempo de residência sobre a recuperação metálica da flotação colunar reversa de minério de ferro*. Tese de Doutorado, Universidade Federal de Ouro Preto, 2006.
- MIAO, J., NIU, L. “A Survey on Feature Selection”. Em: *Procedia Computer Science*, v. 91, 2016. doi: 10.1016/j.procs.2016.07.111.
- MOHANTY, S. “Artificial neural network based system identification and model predictive control of a flotation column”, *Journal of Process Control*, v. 19, n. 6, pp. 991–999, 2009. ISSN: 09591524. doi: 10.1016/j.jprocont.2009.01.001. Disponível em: <<http://dx.doi.org/10.1016/j.jprocont.2009.01.001>>.
- MONTEIRO, R., LUCAS, L., CONTI, C., et al.. “Light field HEVC-based image coding using locally linear embedding and self-similarity compensated prediction”, *2016 IEEE International Conference on Multimedia and Expo Workshop, ICMEW 2016*, pp. 1–4, 2016. doi: 10.1109/ICMEW.2016.7574670.
- MURRUGARRA-LLERENA, N. E. *Ensembles na classificação relacional*. Dissertação de Mestrado, Universidade de São Paulo, 2011.
- NGUYEN, A. V., EVANS, G. M. “Movement of fine particles on an air bubble surface studied using high-speed video microscopy”, *Journal of Colloid and Interface Science*, v. 273, n. 1, pp. 271–277, 2004. ISSN: 00219797. doi: 10.1016/j.jcis.2003.12.066.
- NGUYEN, L. H., HOLMES, S. “Ten quick tips for effective dimensionality reduction”, *PLoS Computational Biology*, v. 15, n. 6, pp. 19, 2019. doi: 10.1371/journal.pcbi.1006907.
- NGUYEN, T. “Dropping forward-backward algorithms for feature selection”, *arXiv preprint arXiv:1910.08007*, 2019.
- NOI, P. T., KAPPAS, M. “Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery”, *Sensors (Switzerland)*, v. 18, 2018. ISSN: 14248220. doi: 10.3390/s18010018.
- OUTOTEC. “Outotec Flotation Technologies”. 2017.
- PAIVA, M. P. D. *Caracterização e Flotação de Minério de Fosfato da Mina Sofia - Chile*. Dissertação de Mestrado, Universidade Federal do Rio Grande do Sul, 2012.

- PAL, M. “Random forest classifier for remote sensing classification”, *International Journal of Remote Sensing*, v. 26, n. 1, pp. 217–222, 2005. ISSN: 01431161. doi: 10.1080/01431160412331269698.
- PERES, A. E. C. *Apostila de Flotação, Curso de Pós-Graduação em Engenharia Metalúrgica e de Minas*. PPGEM - UFMG, 2003.
- POPLI, K. *Real-time Process Monitoring for Froth Flotation Processes Using Image Processing and Dynamic Fundamental Models*. Tese de Doutorado, University of Alberta, 2017.
- RALLO, R., FERRE-GINÉ, J., ARENAS, A., et al.. “Neural virtual sensor for the inferential prediction of product quality from process variables”, *Computers and Chemical Engineering*, v. 26, 2002. ISSN: 0098-1354. doi: 10.1016/S0098-1354(02)00148-5.
- REIS, A. D. S. *Estudo da Geração de Bolhas de Diversos Tamanhos em Coluna de Flotação*. Dissertação de Mestrado, Universidade Federal de Uberlândia, 2015.
- ROBNIK-SIKONJA, M., KONONENKO, I. “Theoretical and Empirical Analysis of Relief and RReliefF”, *Machine Learning*, v. 53, pp. 23–69, 2003. doi: 10.1023/A:1025667309714.
- ROBNIK-ŠIKONJA, M., KONONENKO, I. “An adaptation of Relief for attribute estimation in regression”. Em: *Machine Learning: Proceedings of the Fourteenth International Conference (ICML'97)*, v. 5, pp. 296–304, 1997.
- SANTOS, M. A. D. *A Flotação por Ar Dissolvido como Alternativa ao Tratamento de Efluente Mineral Visando ao Reuso da Água e à Melhoria do Processo de Flotação de Apatita*. Tese de Doutorado, Universidade Federal de Uberlândia, 2014.
- SILVA, I. N. D., SPATTI, D. H., FLAUZINO, R. A., et al.. *Artificial Neural Networks: A Practical Course*. São Carlos, Springer, 2017. ISBN: 9783319431628. doi: 10.1007/978-3-319-43162-8.
- SUSANA, P. R. P. *Sensores Virtuais Usando Aprendizagem Online para Processos Industriais*. Dissertação de Mestrado, Universidade de Coimbra, 2015.
- VAN DER MAATEN, L. J. P., POSTMA, E. O., VAN DEN HERIK, H. J. “Dimensionality Reduction: A Comparative Review”, *Journal of Machine Learning Research*, v. 10, pp. 1–41, 2009. ISSN: 0169328X. doi: 10.1080/13506280444000102.
- VISALAKSHI, S., RADHA, V. “A literature review of feature selection techniques and applications: Review of feature selection in data mining”. Em: *2014 IEEE International Conference on Computational Intelligence and Computing Research, IEEE ICCIC 2014*, 2015. ISBN: 9781479939725. doi: 10.1109/ICCIC.2014.7238499.

- WANG, J. S., HAN, S., SHEN, N. N., et al.. “Features extraction of flotation froth images and BP neural network soft-sensor model of concentrate grade optimized by shuffled cuckoo searching algorithm”, *Scientific World Journal*, v. 2014, 2014. ISSN: 1537744X. doi: 10.1155/2014/208094.
- WANG, L., ZHOU, X., ZHU, X., et al.. “Estimation of biomass in wheat using random forest regression algorithm and remote sensing data”, *Crop Journal*, v. 4, n. 3, pp. 212–219, 2016. ISSN: 22145141. doi: 10.1016/j.cj.2016.01.008. Disponível em: <<http://dx.doi.org/10.1016/j.cj.2016.01.008>>.
- YAN, H., WANG, Q., WANG, Z., et al.. “Estimation of copper concentrate grade for copper flotation”, *IFAC-PapersOnLine*, v. 51, n. 21, pp. 94–98, 2018. ISSN: 24058963. doi: 10.1016/j.ifacol.2018.09.398.
- YANG, L., LIU, S., TSOKA, S., et al.. “A regression tree approach using mathematical programming”, *Expert Systems with Applications*, v. 78, pp. 347–357, 2017. ISSN: 09574174. doi: 10.1016/j.eswa.2017.02.013. Disponível em: <<http://dx.doi.org/10.1016/j.eswa.2017.02.013>>.
- YOUNG, P. “Flotation machines”, *Mining Magazine*, v. 145, pp. 35, 1982.
- YU, L., LIU, H. “Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution”. Em: *Proceedings of the Twentieth International Conference on Machine Learning*, v. 2, pp. 856–863, Washington, DC, 2003.
- ZANATA, D. R. P. *Desenvolvimento de Sensor Virtual Empregando Redes Neurais Para Medição da Composição em uma Coluna de Destilação*. Dissertação de Mestrado, Universidade de São Paulo, 2005.
- ZHANG, H., LIU, J., WANG, Y., et al.. “Cyclonic-static micro-bubble flotation column”, *Minerals Engineering*, v. 45, pp. 1–3, 2013. ISSN: 08926875. doi: 10.1016/j.mineng.2013.01.006. Disponível em: <<http://dx.doi.org/10.1016/j.mineng.2013.01.006>>.
- ZHAO, Z., LIU, H. “Spectral feature selection for supervised and unsupervised learning”, *ACM International Conference Proceeding Series*, v. 227, pp. 1151–1157, 2007. doi: 10.1145/1273496.1273641.