Programa de Pós-Graduação em Instrumentação, Controle e Automação de Processos de Mineração - PROFICAM

Universidade Federal De Ouro Preto - Escola de Minas

Associação Instituto Tecnológico Vale

APLICAÇÃO DE MÉTODOS DE APRENDIZADO DE MÁQUINA PARA OTIMIZAÇÃO DO DESEMPENHO DA PRENSA DE ROLOS NO PROCESSO DE PELOTIZAÇÃO

Thiago Nicoli de Abreu

Ouro Preto

Minas Gerais, Brasil

2021

Thiago Nicoli de Abreu

APLICAÇÃO DE MÉTODOS DE APRENDIZADO DE MÁQUINA PARA OTIMIZAÇÃO DO DESEMPENHO DA PRENSA DE ROLOS NO PROCESSO DE PELOTIZAÇÃO

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. Saul Emanuel Delabrida Silva, D.Sc.

Coorientador: Prof. Andrea G. Campos Bianchi, D.Sc.

Ouro Preto

2021

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

A162a Abreu, Thiago Nicoli De .
Aplicação de métodos de aprendizado de máquina para otimização do desempenho da prensa de rolos no processo de pelotização. [manuscrito] / Thiago Nicoli De Abreu. - 2021.
84 f.: il.: color., tab..

Orientador: Prof. Dr. Saul Emanuel Delabrida Silva. Coorientadora: Profa. Dra. Andrea G. Campos Bianchi. Dissertação (Mestrado Profissional). Universidade Federal de Ouro Preto. Programa de Mestrado Profissional em Instrumentação, Controle e Automação de Processos de Mineração. Programa de Pós-Graduação em Instrumentação, Controle e Automação de Processos de Mineração.
Área de Concentração: Engenharia de Controle e Automação de Processos Minerais.

1. Aprendizado do computador. 2. Mineração de dados (Computação). 3. Banco de Dados. 4. Prensa de Rolos. 5. Pelotização. I. Bianchi, Andrea G. Campos. II. Silva, Saul Emanuel Delabrida. III. Universidade Federal deOuro 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
INSTITUTO DE CIENCIAS EXATAS E BIOLOGICAS
DEPARTAMENTO DE COMPUTACAO

FOLHA DE APROVAÇÃO

Thiago Nicoli de Abreu

Aplicação de Métodos de Aprendizado de Máquina para Otimização do Desempenho da Prensa de Rolos no Processo de Pelotização

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

Aprovada em 09 de setembro de 2021

Membros da banca

Doutor - Saul Emanuel Delabrida Silva - Orientador - Universidade Federal de Ouro Preto
Doutor - Andrea Gomes Campos Bianchi - Co-Orientadora - Universidade Federal de Ouro Preto
Doutor - Alan Kardek Rêgo Segundo - Universidade Federal de Ouro Preto
Doutor - Sérgio de Oliveira - Universidade Federal de São João del-Rei

Saul Emanuel Delabrida Silva, orientador do trabalho, aprovou a versão final e autorizou seu depósito no Repositório Institucional da UFOP em 01/12/2021

Documento assinado eletronicamente por **Saul Emanuel Delabrida Silva**, **PROFESSOR DE MAGISTERIO SUPERIOR**, em 28/12/2021, às 16:52, 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 **0261310** e o código CRC **263778F0**.

Referência: Caso responda este documento, indicar expressamente o Processo nº 23109.013422/2021-75 SEI nº 0261310

R. Diogo de Vasconcelos, 122, - Bairro Pilar Ouro Preto/MG, CEP 35400-000
Telefone: 3135591692 - www.ufop.br

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brasil (CAPES), Código de Financiamento 001; do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq); da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG); e da Vale SA.

Resumo

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

APLICAÇÃO DE MÉTODOS DE APRENDIZADO DE MÁQUINA PARA OTIMIZAÇÃO DO DESEMPENHO DA PRENSA DE ROLOS NO PROCESSO DE PELOTIZAÇÃO

Thiago Nicoli de Abreu

Setembro/2021

Orientadores: Prof. Saul Emanuel Delabrida Silva, D.Sc

Prof. Andrea G. Campos Bianchi, D.Sc.

A tecnologia da prensa de rolos é útil nos processos de pelotização para cominuição do *pellet feed* e para o aumento da superfície específica do minério de ferro, o que impacta em ganhos de produtividade e qualidade na pelotização. O aumento da eficiência da prensa depende de um grande número de variáveis. Este trabalho identifica as variáveis de maior importância no ganho da superfície específica, desenvolve um modelo de classificação para determinar regras de configurações ótimas de operação e apresenta um modelo de regressão para predição da variável de superfície específica. As variáveis de maior influência foram ranqueadas e os *setups* ótimos de operação foram determinados, fatores estes que suportam a tomada de decisão pelos operadores e pela engenharia de processo. Os resultados deste trabalho agilizam e automatizam o diagnóstico do desempenho da prensa de rolos em tempo real.

Palavras-chave: Modelamento, Aprendizado de Máquina, Mineração de dados, Banco de Dados, Prensa de Rolos, Pelotização.

Macrotema: Usina; **Linha de Pesquisa:** Tecnologias de Informação, Comunicação e Automação Industrial; **Tema:** Aumento de Produtividade na Usina/Pelotização.

Abstract

Abstract of Master Thesis presented to the Graduate Program on Instrumentation, Control and Automation of Mining Process of the Escola de Minas/UFOP and ITV as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

APPLICATION OF MACHINE LEARNING METHODS FOR OPTIMIZING ROLLER PRESS PERFORMANCE IN THE PELLETIZING PROCESS

Thiago Nicoli de Abreu

September/2021

Advisors: Prof. Saul Emanuel Delabrida Silva, D.Sc

Prof. Andrea G. Campos Bianchi, D.Sc.

The roller press technology is useful in pelletizing processes for comminution of pellet feed and for increasing the specific surface of the iron ore, which impacts productivity and quality gains in pelletizing. The increase in press efficiency depends on a large number of variables. This work identifies the most important variables in the gain of the specific surface, determines a classification model to determine rules for optimal operating configurations and presents a regression model to predict the specific surface variable. The most influential variables were ranked and the optimal operating setups were determined, factors that support decision making by operators and process engineering. The results of this work streamline and automate the diagnosis of roller press performance in real time.

Keywords: Modeling, Machine Learning, Data Mining, Database, Roller Press, Pelletizing.

Macrotheme: Plant; **Research Line:** Information Technology, Communication and Industrial Automation; **Theme:** Increased Productivity at the Plant/Pelletizing.

Lista de figuras

Lista de tabelas

Sumário

1 INTRODUÇÃO

O processo de pelotização de minério de ferro é um processo de aglomeração de finos de minério (*pellet feed* - tamanho inferior a 0,15 mm) em esferas com granulometria e qualidade adequadas, chamadas pelotas (Figura 1). Estas são diretamente utilizadas no processo siderúrgico.

Figura 1: Pelotas de minério de ferro.
Fonte: Vale S.A.[1]

O processo de pelotização é composto de várias etapas (Figura 2):

1. Empilhamento e recuperação do *pellet feed* proveniente da extração da mina e estocado no pátio de alimentação;
2. Moagem para redução da granulometria com duas etapas, cominuição do minério de ferro e classificação através da ciclonagem (processo presente nos chamados "circuitos fechados");
3. Espessamento, cujas finalidades são recuperar a água para ser reutilizada no processo e aumentar a porcentagem de sólidos para etapas posteriores;
4. Homogeneização da polpa de minério, controlando a densidade da polpa de minério e a adição de antracito no processo;
5. Filtragem, cuja finalidade é reduzir a umidade da polpa proveniente das etapas de espessamento e homogeneização;
6. Prensagem, cuja finalidade é aumentar a superfície específica do minério através do processo de cominuição (em algumas plantas esse processo está localizado na etapa de pré-moagem);

[1] Disponível em: <http://www.vale.com/brasil/pt/aboutvale/news/paginas/entenda-funciona-processo-pelotizacao-usinas.aspx>. Acesso em: 02 dez. 2019.

7. Mistura, em que aditivos são adicionados com a finalidade de controlar as taxas de pelotamento e assegurar que as pelotas cruas permaneçam unidas até que sejam endurecidas na etapa de queima, além de possibilitar que as pelotas adquiram certas propriedades no final do processo;

8. Pelotamento, quando ocorre a formação de esferas, denominadas pelotas cruas;

9. Peneiramento, cuja função é classificar as pelotas verdes (pelotas que ainda não passaram pelo processo de queima) com relação a sua granulometria, pois a seleção e distribuição do tamanho das pelotas é fundamental para a sua qualidade;

10. Queima, que envolve o endurecimento das pelotas por meio de tecnologias de fornos de grelha móvel, forno rotativo e forno vertical;

11. Nova etapa de peneiramento, que classifica as pelotas queimadas para o processo de empilhamento e embarque.

Ao final, o percentual de finos gerados retorna ao processo.

O sucesso na produção de pelotas depende do sucesso em cada uma das etapas citadas. Um erro na etapa precedente não é completamente corrigido nas etapas posteriores.

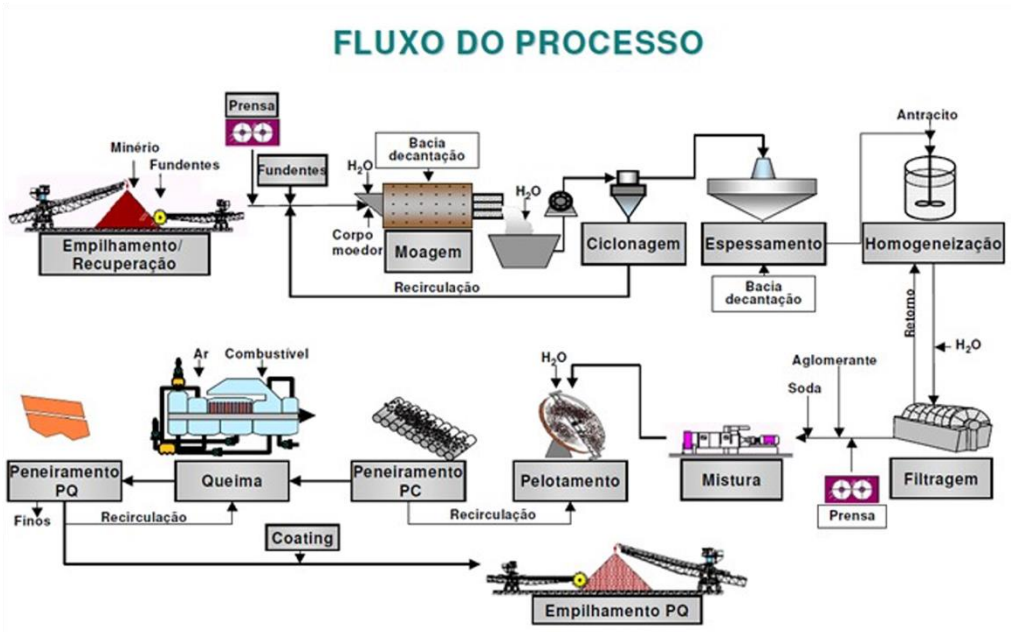

Figura 2: Fluxo do processo de Pelotização.
Fonte: Vale S.A.[2]

[2] Disponível em: <http://www.vale.com/brasil/pt/aboutvale/news/paginas/entenda-funciona-processo-pelotizacao-usinas.aspx>. Acesso em: 27 nov. 2019.

No início da década de 1990, uma série de plantas industriais de pelotização começaram a implantar a prensa de rolos em circuitos industriais de prensagem de minério de ferro e pellet feed. Este foi um avanço importante na área de cominuição e processamento mineral (BARRIOS, 2014).

Um dos maiores benefícios da prensa de rolos no processo de pelotização é o aumento do ganho específico de superfície, propriedade que contribui diretamente para a melhoria das propriedades físicas e mecânicas das pelotas e da qualidade do produto acabado. Quando um material possui uma superfície específica elevada, maior é a capilaridade dos vasos, resultando em uma pelota mais compacta, com maior acabamento e com melhor resistência mecânica (SILVA, 2008).

A ineficiência nas etapas anteriores ao processo de prensagem, possíveis problemas com o ajuste dos parâmetros operacionais e a indisponibilidade operacional deste equipamento (devido a falhas) são alguns dos fatores que impactam o seu desempenho.

A baixa eficiência da prensa acarreta distúrbios para as próximas etapas do processo, principalmente na etapa de pelotização (formação das pelotas), ocasionando perdas na produção e até paralisações da planta. Uma mineradora pode incorrer em perdas correspondentes a milhões de dólares em poucas horas devido à paralisação de sua produção. Além disso, um produto fabricado de forma inadequada pode levar a retrabalho ou não conformidade com os requisitos de qualidade do cliente.

Para o período de dados avaliados para este trabalho, o levantamento de dados de paradas e perdas de produção diretamente relacionados à prensa de rolos em cinco usinas de pelotização de uma indústria brasileira entre 01 de janeiro de 2018 a 01 de dezembro de 2019 mostram um total de 290,95 horas (3,94% relacionado a paradas das usinas e 96,06% a perdas de produção (Figura 3).

Figura 3: Soma total de horas de paradas e perdas de produção relacionados à prensa de rolos.
Fonte: O autor.

A distribuição de horas também pode ser categorizada pelo impacto em cada uma das cinco usinas de pelotização (Figura 4).

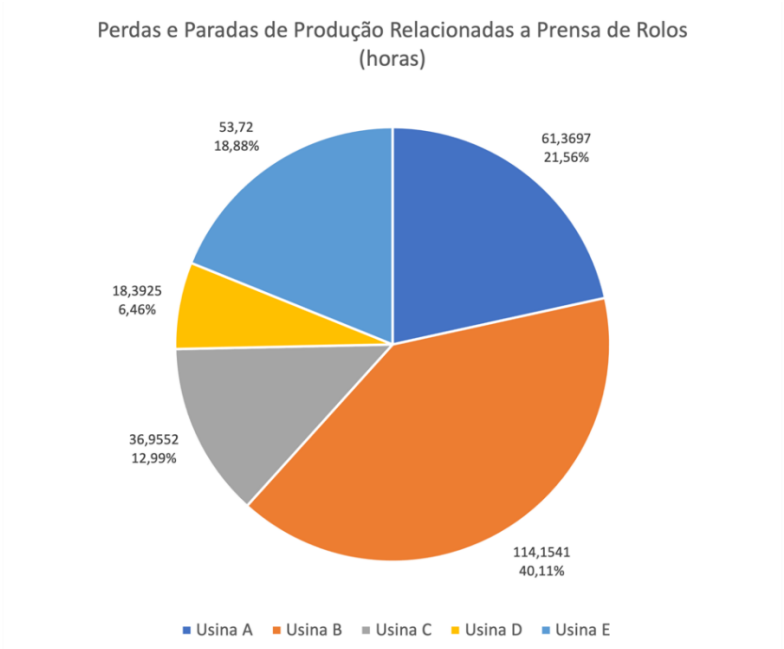

Figura 4: Soma de horas de paradas e perdas de produção relacionados à prensa de rolos de cada uma das cinco usinas de pelotização de uma indústria brasileira.
Fonte: O autor.

Percebe-se, portanto, que falhas da prensa de rolos no processo de pelotização produz alta perda de produção e que as usinas de pelotização B, E e A sofrem, respectivamente, os maiores impactos.

Diante desta análise, a planta em estudo obteve 11,22 horas de paralisação e 273,37 horas de perda ou redução de produção. Considerando o valor médio da produção da planta em 700 t/h e o preço comercial da tonelada de minério de ferro em US $ 100,00/t, o valor estimado da perda econômica foi de aproximadamente US $ 785.000,00. A principal causa relacionada à perda de produção está relacionada aos baixos valores de superfície específica após o processo de prensagem

Na Tabela 1 é possível verificar alguns exemplos das causas que relacionam os impactos de paradas e perdas nas usinas.

Tabela 1: Exemplos das causas de problemas da prensa de rolos e as consequências no processo das usinas de pelotização.

Causas de parada e perda de produção em função da Prensa de Rolos	**Consequência**
Baixo nível dos silos de pelotamento durante manutenção preventiva da prensa.	**Parada de Produção**
Variações do material no início do dia devido à estabilização da prensa pós parada.	**Perda de Produção**
Produção reduzida devido a manutenção preventiva da prensa de rolos por 216h.	
Redução de produção, devido manutenção preventiva da prensa de rolos por 216h.	
Elevação do retorno do pelotamento devido a queda de superfície específica resultante da prensa de rolos.	
Baixa granulometria da polpa prensada (prensa trabalhando com controle de nível em modo virtual). Para ter condição de operação a pressão de trabalho foi alterada de 85bar para 65bar.	
Prensa operando pelo controle dos rolos pela balança (Controle virtual), ocasionando baixo ganho de superfície específica.	
Variação do material no pelotamento, superfície específica baixa da polpa prensada 1877 cm^2/g com ganho de 197g/cm^2, prensa com baixa eficiência.	
Retorno de produção após manutenção na prensa de rolos e parada por falta de material. Desmontagem de andaimes no interior do chute de alimentação da prensa.	
Redução de produção devido baixos valores de superfície específica de polpa prensada, gerando excesso de finos no pelotamento, elevando as taxas de retorno.	
Redução de produção devido baixos valores de superfície específica de polpa prensada, gerando excesso de finos no pelotamento, elevando as taxas de retorno.	
Redução de produção para preparação de parada programada de usina.	
Superfície específica abaixo da meta. Baixo rendimento da prensa. Média de ganho de 166g/cm^2.	
Descontrole no pelotamento, provocando elevação do retorno. Prensa não atingiu o *set point* de pressão e torque após ajustes no *gap* zero, foi necessário intervenção da instrumentação para ajustes nos sensores.	

Fonte: Vale S.A.

Assim, o conhecimento para operação da prensa de rolos é um fator preponderante no cumprimento de seu objetivo, que é o aumento da superfície específica do minério de ferro. A grande quantidade de variáveis envolvidas para parametrização (pressão de trabalho dos rolos, *gap* entre os rolos, vazão de alimentação, nível da calha da prensa, umidade do minério prensado, velocidade dos rolos, torque e corrente dos motores, dentre outras) e *inputs* de controle aumentam a complexidade em sua operação.

Adicionalmente, outro ponto que acentua essa complexidade está relacionado com as variações das características do processo produtivo (variações nas etapas anteriores e posteriores, características físico-químicas do minério, restrições de processo etc.) que introduzem novos cenários durante a operação.

Este cenário de múltiplas variáveis a serem controladas enfatiza a necessidade de um estudo de métodos computacionais inteligentes que possam cooperar nos ajustes e na melhor tomada de decisão em seu funcionamento. Portanto, o conhecimento para operar a prensa de rolos é um fator importante para o cumprimento de seu objetivo, que é aumentar a superfície específica do minério de ferro. Depender da ação humana apenas para controlar e manter um padrão operacional diante de toda a complexidade para análise é praticamente impossível.

Diante de todos os impactos da prensa de rolos no processo, assim como os problemas enfrentados, faz-se necessário responder às seguintes perguntas: quais os principais fatores que interferem direta e indiretamente na eficiência deste equipamento? Quais os possíveis *setups* (configuração de valores de operação) ótimos para operação visando o melhor ganho no processo?

A resposta para estas perguntas permite a melhoria do processo para aumento da produtividade (volume de produção) e do ganho de superfície específica, além de possibilitar a antecipação de falhas com a consequente redução de perdas por manutenção corretiva e anomalias no processo.

Outro fator de motivação está relacionado com a medição da superfície específica da polpa prensada. Atualmente essa medição é realizada em laboratório, no intervalo de 4 horas, através da coleta realizada pelo amostrador (equipamento para coleta de amostras) pós-circuito de prensagem. Sendo assim, a predição da classificação desta medição de forma *online* seria uma maneira mais ágil para avaliar o desempenho do equipamento e do processo.

Este trabalho propõe uma análise da prensa de rolos no processo de prensagem em uma das usinas de pelotização, através de técnicas de inteligência artificial para identificação das variáveis de maior influência para a determinação do ganho da prensa. Por fim, desenvolve um modelo de classificação para determinar *setups* ótimos de operação da prensa de rolos. Este modelo é utilizado para predição da variável de superfície específica da polpa prensada, buscando ganhos na determinação do desempenho deste ativo.

Este trabalho restringe-se em utilizar dados do processo para identificação das variáveis por meio de simulação. Logo, não faz parte do escopo realizar instalações e intervenções nos equipamentos da planta.

2 OBJETIVOS

2.1 Objetivo geral

Utilizar técnicas de inteligência artificial para identificar as principais variáveis que impactam diretamente no desempenho da prensa de rolos e desenvolver um modelo de classificação da superfície específica da prensa.

2.2 Objetivos específicos

Os objetivos específicos deste trabalho consistem em:

- Identificar as variáveis de maior influência no processo de prensagem.
- Desenvolver um modelo de classificação da superfície específica da prensa, identificando as regras para tomada de decisão utilizando diferentes valores de *setups* ótimos de operação da prensa de rolos.
- Desenvolver um modelo de predição do valor da superfície específica da prensa.

O restante deste documento está organizado da seguinte forma: O capítulo 3 descreve o referencial teórico e a fundamentação científica relacionados ao uso de técnicas de inteligência computacional para o processo de prensagem. O capítulo 4 descreve os materiais e métodos que foram usados para desenvolver este trabalho. O capítulo 5 descreve os resultados obtidos e as discussões realizadas com o uso de técnicas de aprendizado de máquina para: determinar as variáveis mais relevantes, obter as regras de setups ótimos e o modelo de predição da superfície específica da prensa. O capítulo 6 descreve as conclusões após a análise realizada.

3 REFERENCIAL TEÓRICO E FUNDAMENTAÇÃO CIENTÍFICA

Estudos realizados em processos de pelotização, área de prensagem e aplicações de modelos matemáticos em processos de mineração são fundamentais como referência e desenvolvimento deste trabalho.

Ao longo das últimas duas décadas, foram propostos diferentes modelos matemáticos empíricos e fenomenológicos com o objetivo de prever a capacidade da prensa de rolos, a potência demandada durante a prensagem, a distribuição granulométrica do produto e também avaliar o comportamento das variáveis de desempenho da prensa quando aplicada em diferentes materiais e condições de operação.

Daniel (2002) e Barrios (2015) afirmam que os modelos desenvolvidos não consideram a dinâmica dos parâmetros chave durante a operação da prensa para realizar a previsão das variáveis de desempenho do equipamento.

A utilização de modelos para previsibilidade em processos de mineração estão sendo cada vez mais utilizados. Monteiro (2003), apresenta o desenvolvimento da modelagem por redes neurais dos tipos MLP (*Multilayer Perceptron*) e RBF (*Radial Basics Function*) no software Matlab para previsão da qualidade física das pelotas queimadas da usina de pelotização da Ferteco Mineração.

Bastos (2015) faz uma análise comparativa entre métodos convencionais e o uso de software para a seleção de britadores e peneiras, concluindo que os resultados obtidos se mostraram bastante congruentes, sendo que o dimensionamento através de simulador resulta em uma resposta bem mais detalhada e proporciona melhorias para verificação de erros e estudos de novos cenários na utilização destes equipamentos.

Campos (2018) discute o modelo matemático fenomenológico desenvolvido por torres e Cassali (2009) capaz de prever a capacidade, potência e distribuição granulométrica do produto gerado no equipamento e abordam uma série de testes em diferentes escalas para prensagens de pellet feed e minério de ferro.

Vyhmeister *et al.* (2019) apresenta um estudo de modelamento para prensa de rolos baseado em modelo de controle preditivo (MPC), mostrando que a crescente necessidade de análises em controles multivariáveis para processos complexos exige estratégias cada vez mais robustas e avançadas. Seu modelo considera a energia total consumida como uma das

principais variáveis controladas, utilizando como variáveis de saída do modelo: a dinâmica da capacidade de tratamento, a distribuição granulométrica do produto, a energia de compressão e a energia de laminação. Por fim utiliza este modelo para produzir um esquema de controle de múltiplas entradas e saídas, observando que o modelo tem uma representação correta dos fenômenos envolvidos e que a velocidade e pressão periféricas utilizadas na prensa de rolos são variáveis manipuladas úteis para controlar a energia consumida pelo equipamento em um esquema de MPC.

Hasanzadeh e Farzanegan (2011) aplicam um método baseado em algoritmos genéticos para estimativa dos parâmetros de modelos matemáticos desenvolvidos para prensa de rolos, tendo como base o modelo explicado por Torres e Cassali (2009). Eles desenvolveram um algoritmo de simulação e implementaram no Matlab para testar e demonstrar a aplicação deles para calibração e obtenção dos valores ótimos dos parâmetros dos modelos, validando-os em um conjunto de dados experimentais e prevendo a distribuição do tamanho do produto da prensa de rolos sob várias condições operacionais.

Portanto, fica evidente a necessidade do desenvolvimento de modelos precisos de previsão para melhoria das operações das unidades de prensa de rolos e que a maioria das representações destes equipamentos está baseada em modelos de estado estacionário para projetos e otimização *offline*, tornando-os inadequados para o controle de processo e otimização *online*.

Não foram encontrados trabalhos acadêmicos relacionados com técnicas de aprendizado de máquina sendo utilizadas para determinar as principais variáveis que possuem maior influência no processo de prensagem de minério de ferro em usinas de pelotização bem como o desenvolvimento de um classificador para a meta estabelecida do resultado de superfície específica do minério prensado. Portanto, este trabalho mostra que existe um grande potencial ainda a ser explorado na utilização da área de inteligência artificial junto aos processos de prensagem de minério de ferro na indústria.

3.1 Processo de prensagem na pelotização

A prensa de rolos constituiu um importante avanço dentro da área de cominuição e de processamento mineral nas últimas décadas (BARRIOS, 2015).

A prensagem de *pellet feed* tem mostrado uma série de benefícios para o processo de pelotização, como:

- Redução do consumo energético através da introdução de microfissuras, que ajudam a aumentar a taxa de produção e a redução da granulometria do produto.
- Geração de uma alta proporção de ultrafinos, principalmente pela atrição e cisalhamento das superfícies das partículas no leito de material úmido, resultando em aumento da área superficial específica da ordem de 300 a 600 cm^2/g.
- Geração de formato de partícula mais angular em comparação com produtos de moinho de bolas, o que permite que o produto seja adequado para pelotização, presumindo um melhor empacotamento e densidade das pelotas.
- Vantagens no processo de pelotamento em termos de uniformidade no tamanho das pelotas verdes, maior resistência, menor consumo de aditivos e menor taxa de recirculação na grelha (equipamento utilizado no processo de queima).

Enfatizando a propriedade de superfície específica da pelota, quanto menor o tamanho das partículas, maior sua área superficial (OLIVEIRA, 2010). No caso do *pellet feed*, quando uma maior fração granulométrica é próxima de 45 μm, a superfície específica (SE) aumenta consideravelmente, o que favorece a etapa de pelotamento e qualidade das pelotas cruas (SOLÉ; WENDLING, 2014).

Pal *et al.* (2015) verificaram que um *pellet feed* hematítico com índice Blaine de 1628 cm2/g é suficiente para produzir pelotas cruas com propriedades físicas e mecânicas desejadas, tendo as pelotas queimadas, a resistência à compressão melhorada quando esse índice aumenta. Isso ocorre porque quando um material possui alta SE, maior será a capilaridade dos vasos, o que resulta numa pelota mais compacta, mais acabada e de melhor resistência mecânica (SILVA, 2008). O maior adensamento de partículas durante o pelotamento pode resultar em menor umidade presente na pelota (MOURÃO, 2017).

A Prensa de Rolos consiste em dois rolos girando em sentidos opostos, sendo um rolo denominado fixo, gira sobre um eixo fixo, e o outro como rolo móvel, que gira sobre um eixo móvel que executa um movimento de translação em direção ao rolo fixo através de um sistema hidráulico (Figura 3). Esse sistema permite uma variação da força específica de compressão exercida sobre o leito de partículas entre os rolos.

A alimentação de material é introduzida na abertura entre rolos, onde a cominuição das partículas ocorre pelo mecanismo de quebra interparticular (DANIEL, 2002).

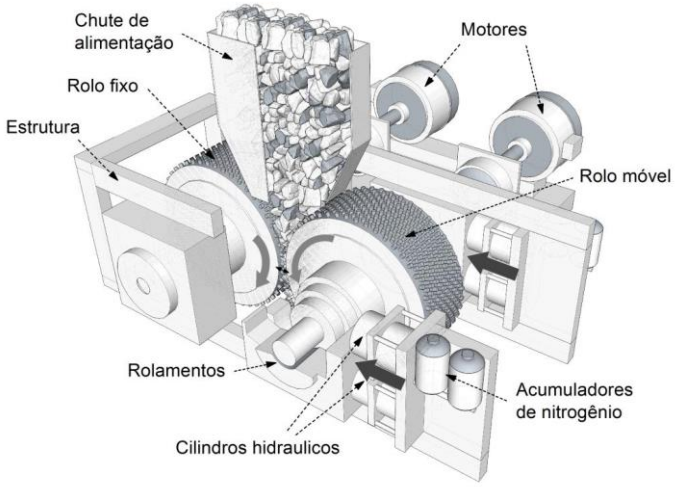

Figura 5: Esquema da prensa de rolos.
Fonte: BARRIOS, 2015.

Impulsionado pelos benefícios provindos deste equipamento, a Vale implementou a prensagem de minério de ferro em prensa de rolos combinada com a moagem em moinhos de bolas para produzir *pellet feed* finos em várias das usinas integradas de moagem e pelotização do Complexo de Tubarão, seja nas etapas de pré-moagem ou remoagem de descarga do moinho de bolas (VAN DER MEER, 1997).

3.2 Inteligência artificial

Durante muitos anos tem-se procurado entender como os seres humanos pensam, como uma matéria presente nestes seres pode produzir sentimentos, percepções, compreensão e avaliação de um ambiente ou situações durante os instantes de seu viver.

A afirmação de Russel e Norvig (2009) corrobora o avanço desse entendimento, “a conclusão verdadeiramente espantosa é que uma coleção de células simples pode levar ao pensamento, à ação e a consciência” ou, nas palavras incisivas de John Searle (1992), “os cérebros geram mentes.”

Desde o primeiro trabalho reconhecido como inteligência artificial (IA) em 1943 realizado por Mcculloch e Pitts (1943), baseando-se no conhecimento fisiológico e função dos neurônios, lógica proposicional e teoria da computação, até os dias de hoje, sua aplicação e desenvolvimento são em áreas diversas, tais como exemplo: reconhecimento de voz, veículos

robóticos, planejamento autônomo e escalonamento, combate a *spam*, planejamento logístico, robótica e tradução automática.

Portanto o campo da inteligência artificial tem sido um dos mais recentes na ciência e engenharia. O desenvolvimento de técnicas e sistemas computacionais que imitam aspectos humanos, tais como percepção, raciocínio, aprendizado, evolução e adaptação estão sendo largamente estudados.

Segundo estudo publicado pela Gartner (empresa americana de pesquisa e consultoria em tecnologia da informação) até 2022 90% das estratégias corporativas mencionaram explicitamente as informações como um ativo crítico da empresa e as análises como uma competência essencial.

Na mineração por exemplo, a utilização de IA para detecção de falhas tem sido amplamente estudada e implementada. Análises de defeitos em transportadores de correias com modelos de análise de imagem por rede neural tem apresentados resultados satisfatórios e possibilidade de implementação nos ambientes produtivos (KLIPPEL *et al.*, 2021).

A detecção de rolos defeituosos em tempo real utilizando modelos de detecção de objetos em uma arquitetura de aprendizado profundo (*deep learning*) também tem tido grandes avanços (D'ANGELO *et al.*, 2019). Modelos de predição de falha utilizando algoritmos de aprendizado de máquina como *Random Forest e Multilayer Perception* para análise de ruídos de falha típico em rolos de correias também demostram o avanço e a aplicabilidade desta ciência no ambiente industrial (ERICEIRA *et al.*, 2020).

3.3 Processo KDD (*Knowledge Discovery in Databases*)

KDD é um processo de descoberta de conhecimento em bases de dados que tem como objetivo principal extrair conhecimento a partir de grandes bases de dados. Para isto ele envolve diversas áreas do conhecimento, tais como: estatística, matemática, bancos de dados, inteligência artificial, visualização de dados e reconhecimento de padrões. São utilizadas técnicas, em seus diversos algoritmos, oriundas dessas áreas.

O conhecimento ou processo de KDD (Figura 6) é um processo que busca identificar potenciais padrões úteis, que estejam embutidos nos dados e, tornando-os compreensíveis para um determinado contexto (FAYYAD *et al.*, 1996).

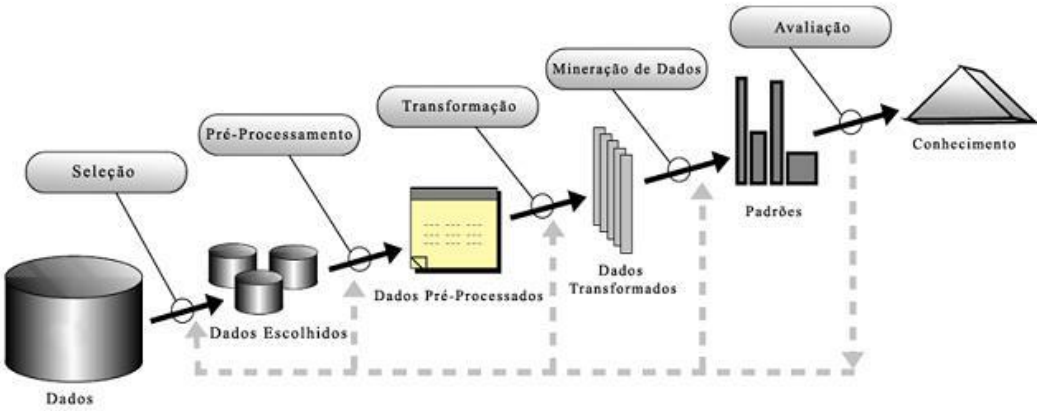

Figura 6: Processo de KDD. Etapas do processo que foram seguidas e utilizadas na metodologia deste trabalho. Fonte: FAYYAD *et al.*, 1996.

Ainda segundo Fayyad *et al.* (1996), este processo foi proposto em 1989 para referir-se às etapas que produzem conhecimento a partir dos dados. Dentro deste processo, a etapa de mineração de dados é a fase que transforma dados em informação.

Fayyad *et al.* (1996) publicaram o trabalho "*From Data Mining to Knowledge Discovery in Databases*" no qual descrevem como são relacionadas a mineração de dados e o KDD em um banco de dados, como em seus campos relacionados – estatística e aprendizagem de máquina.

Neste trabalho é conceituado que KDD é todo o processo de descoberta de conhecimento e a mineração de dados refere-se a apenas uma fase deste processo. No trabalho são relatadas técnicas específicas para mineração de dados tais como árvore de decisão, regressão não linear e modelos de aprendizagem relacional. É discutido que não existe um método mais eficiente que sirva para todas as aplicações. A escolha do método vai variar de acordo com o objetivo da mineração de dados.

Portanto, seu objetivo principal é extrair conhecimento de grandes bases de dados. A descoberta do conhecimento envolve uma sequência de fases que devem ser obedecidas, iniciando-se com a coleta de informações, passando pelo tratamento e, por fim, a apresentação do resultado final da extração do conhecimento.

A sequência de etapas do processo de KDD devem ser executadas sequencialmente, pois ao final de cada etapa, o resultado obtido serve de auxílio para a etapa seguinte, podendo repetir etapas anteriores sempre que necessário.

Sobre o processo KDD, as etapas são apresentadas a seguir.

3.3.1 Seleção de dados

Etapa de agrupamento de forma organizada dos dados. Conhecer o tipo dos dados com o qual se irá trabalhar também é fundamental para a escolha do(s) método(s) mais adequado(s). Pode-se categorizar os dados em dois tipos: quantitativos e qualitativos. Os dados quantitativos são representados por valores numéricos. Eles ainda podem ser discretos e contínuos. Já os dados qualitativos contêm os valores nominais e ordinais (categóricos). Em geral, antes de se aplicar os algoritmos de mineração é necessário explorar, conhecer e preparar os dados.

Segundo Olson e Delen (2008), o processo de preparação dos dados na maioria dos projetos de mineração, compreende até 50% de todo o processo. Para Mccue (2014), está etapa pode compreender até 80%.

Segundo Shedroff (1999), dados “são o produto da descoberta, pesquisa, coleta e criação. É a matéria-prima que encontramos ou criamos que usamos para construir nossas comunicações” e informação “torna os dados significativos para o público, porque requer a criação de relacionamentos e padrões entre os dados”.

3.3.2 Pré-processamento dos dados

Etapa da limpeza dos dados visando adequá-los aos algoritmos que serão utilizados. O processo de preparação dos dados para a mineração, também chamado de pré- processamento (Figura 7), segundo Han e Kamber (2006b), consiste principalmente em:

- Limpeza dos dados

Frequentemente, os dados são encontrados com diversas inconsistências, tais como registros incompletos, valores errados e dados inconsistentes. A etapa de limpeza dos dados visa eliminar estes problemas de modo que eles não influam no resultado dos algoritmos usados. As técnicas usadas nesta etapa vão desde a remoção do registro com problemas, passando pela atribuição de valores padrões, até a aplicação de técnicas de agrupamento para auxiliar na descoberta dos melhores valores. Devido ao grande esforço exigido nesta etapa, Han e Kamber (2006b) propõem o uso de um processo específico para a limpeza dos dados.

- Integração dos dados

É comum obter-se os dados a serem minerados de diversas fontes: banco de dados, arquivos textos, planilhas, vídeos, imagens, entre outras. Surge então a necessidade da integração destes dados de forma a termos um repositório único e consistente. Para isto, é necessária uma análise aprofundada observando redundâncias, dependências entre as variáveis e valores conflitantes (categorias diferentes para os mesmos valores, chaves divergentes, regras diferentes para os mesmos dados, entre outros).

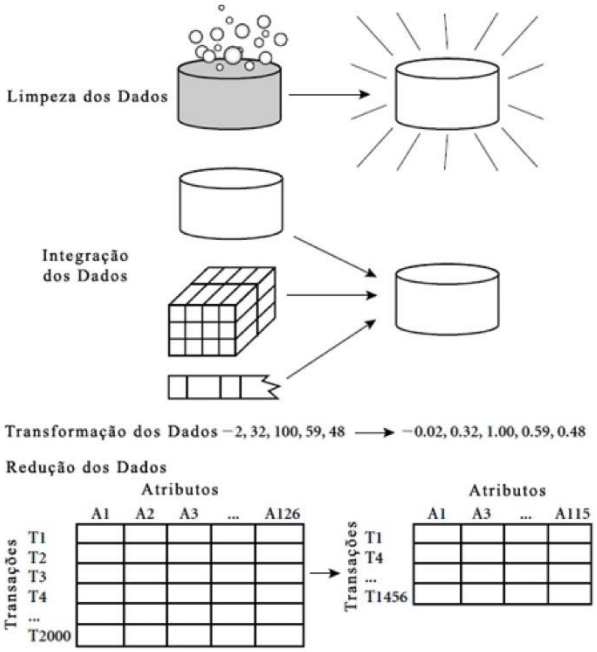

Figura 7: Atividades do pré-processamento.
Fonte: HAN E KAMBER, 2006b.

Na maioria das grandes plantas industriais estes dados provindos de diversas fontes são coordenados para o armazenamento nos sistemas de gerenciamento das atividades de produção (MES – *Manufacturing Execution System*) e que estabelecem uma ligação direta entre o planejamento e o chão de fábrica, interligando os dados de cada área operacional para o fornecimento de informações em tempo real do processo produtivo.

3.3.3 Transformação dos dados

Para facilitar o uso das técnicas de mineração de dados, os dados ainda podem passar por uma transformação que os armazena adequadamente em arquivos para serem lidos pelos algoritmos.

3.3.4 Mineração dos dados

É um processo em que métodos inteligentes são utilizados a fim de extrair padrões de dados. Esta é a etapa em que se inicia a fase de mineração de dados especificamente, começando com a escolha das ferramentas (algoritmos) a serem utilizadas.

A mineração de dados tornou-se uma ferramenta de apoio com papel fundamental na gestão da informação dentro das organizações. A manipulação dos dados e a análise das informações de maneira tradicional tornou-se inviável devido ao grande volume de dados (coletados diariamente e armazenados em bases históricas). Descobrir padrões implícitos e relacionamentos em repositórios que contém um grande volume de dados de forma manual, deixou de ser uma opção. As técnicas de mineração passaram a estar presentes no dia a dia.

Os dados são considerados hoje como o principal ativo de um projeto de software. Isso se deve, além da redução nos custos de aquisição de hardware e software, ao desenvolvimento de técnicas capazes de extrair, de forma otimizada, a informação contida, e muitas vezes implícita, nestes dados.

Apesar dos bons resultados obtidos com a aplicação da mineração de dados, os desafios ainda são muitos. Diversos problemas relativos ao uso da mineração (tais como a segurança dos dados e a privacidade dos indivíduos), juntamente com o aumento na complexidade das estruturas de armazenamento, criam cenários complexos e desafiadores.

3.3.5 Avaliação dos dados

Onde técnicas de visualização e representação de conhecimento são utilizadas para apresentar o conhecimento extraído para o usuário. De acordo com o algoritmo utilizado será gerado um arquivo de descobertas (que pode ser um relatório ou um gráfico, por exemplo). Este arquivo deve ser interpretado, gerando as conclusões que fornecem o conhecimento da base de dados estudada.

3.4 Aprendizado de máquina (AM)

Aprendizado de máquina é uma área de Inteligência Artificial que visa o desenvolvimento de técnicas computacionais capazes de adquirir conhecimento de forma automática. Um sistema de aprendizado é um algoritmo que toma decisões baseado em

experiências acumuladas por meio da solução bem-sucedida de problemas anteriores (WEISS; KULIKOWSKI, 1991).

Uma definição mais simples é que as máquinas podem detectar padrões e criar conexões entre dados, por meio de *Big Data* e algoritmos sofisticados, para aprenderem sozinhas a executar uma tarefa.

Pode-se subdividir a aprendizagem de máquina em várias categorias diferentes:

- Aprendizagem Supervisionada

O aprendizado supervisionado requer um programador que ofereça exemplos de quais entradas se alinham com os resultados ou saídas. Então, um algoritmo de aprendizagem supervisionado tentaria usar explicitamente essa informação para no futuro ser hábil para classificar ou predizer os resultados para novas entradas. Segundo Russel e Norvig (2009), através da observação dos pares de entrada e saída, é concebida uma função que realiza o mapeamento da entrada para saída.

- Aprendizagem Não-Supervisionada

A aprendizagem não supervisionada exige que o sistema desenvolva suas próprias conclusões a partir de um determinado conjunto de dados. Neste caso, a aprendizagem deverá encontrar associações em grupos com algum grau de similaridade para classificar os dados de entrada. Não há uma predefinição de classificação dos dados de entrada. Segundo Russel e Norvig (2009), é realizado o aprendizado dos padrões de entrada, mesmo que não seja fornecido nenhuma resposta.

- Aprendizagem Semi-Supervisionada

A aprendizagem semi-supervisionada é uma combinação de aprendizado supervisionado e não supervisionado. Um sistema de aprendizagem semi-supervisionado utiliza dados rotulados para fazer inferências sobre quais dados não foram rotulados. Segundo Russel e Norvig (2009), existe um ruído aleatório nos dados, com imprecisões sistemáticas e descobri-las envolve um aprendizado não supervisionado, fazendo com que tanto os ruídos como a falta de rótulos criem um ciclo contínuo para melhoria dos dados e dos resultados.

- Aprendizagem Por Reforço

O aprendizado por reforço envolve um sistema que recebe feedback análogo a punições e recompensas. Um exemplo clássico de aprendizagem de reforço (como se aplica à aprendizagem de máquina) é um agente aprendendo a jogar um game. O objetivo é vencer o game e o agente vai sendo recompensado ou punido de acordo com seus erros e acertos, até atingir seu objetivo.

Os sistemas de aprendizado podem ser divididos ainda em simbólicos e não-simbólicos. Os métodos simbólicos ou orientados ao conhecimento desenvolvem representações simbólicas do conhecimento, as quais são, geralmente, facilmente interpretadas por seres humanos. São exemplos de métodos simbólicos as árvores de decisão e conjuntos de regras.

Os métodos não-simbólicos ou caixa-preta, por sua vez, são caracterizados pelo desenvolvimento de representações próprias do conhecimento, as quais, geralmente, não são facilmente interpretadas por seres humanos. Como exemplos de métodos não-simbólicos, podemos citar as Redes Neurais Artificiais, KNN e *Naive Nayes*.

Os problemas de aprendizado de máquina são basicamente divididos em três subáreas principais:

- Classificação: baseia-se em prever a categoria de uma observação fornecida. Neste caso, procura-se estimar um "classificador" que gera como saída a classificação qualitativa de um dado com base em dados de entrada (que abrangem observações com classificações já definidas). Portanto, quando o valor da classe desejada for um valor discreto, a tarefa é denominada classificação (DUDA *et al.*, 2001).
- Regressão: de forma similar a classificação, utiliza dados de entrada (preditores) já observados para prever uma resposta. A grande diferença é que, neste caso, procura-se estimar um valor numérico e não uma classificação de uma observação. Portanto, quando o valor da classe desejada for um valor contínuo, a tarefa é denominada regressão (DUDA *et al.*, 2001).
- Agrupamento: também conhecido como "*Clustering*", tem como objetivo agrupar observações em grupos conhecidos como "*clusters*". Essas observações apresentam similaridades dentro de seu *cluster* e diferenças em relação aos demais *clusters* formados.

Diferente da classificação, não é realizada a rotulação dos *clusters*, fazendo com que não exista uma clusterização errada ou certa. A clusterização utilizada resulta em diferentes tipos de *clusters*, e a escolha dessas técnicas deve ser previamente analisada pelo pesquisador.

Neste trabalho serão desenvolvidas as tarefas de classificação e regressão. Algumas das técnicas relacionadas neste estudo são:

- Árvores de Decisão (*Decision Trees*)

O método de classificação por árvore de decisão, funciona como um fluxograma em forma de árvore, onde cada nó (não folha) indica um teste feito sobre um valor. As ligações entre os nós representam os valores possíveis do teste do nó superior, e as folhas indicam a classe (categoria) a qual o registro pertence. Após a árvore de decisão montada, para classificar um novo registro, basta seguir o fluxo na árvore (mediante os testes nos nós não-folhas) começando no nó raiz até chegar a uma folha. Pela estrutura que formam, as árvores de decisões podem ser convertidas em regras de classificação. O sucesso das árvores de decisão, deve-se ao fato de ser uma técnica extremamente simples, não necessita de parâmetros de configuração e geralmente tem um bom grau de assertividade. Apesar de ser uma técnica extremamente poderosa, é necessário uma análise detalhada dos dados que serão usados para garantir bons resultados. Yang e Pedersen (1997) apresentam um algoritmo para extrair regras acionáveis, ou seja, regras que são realmente úteis para a tomada de decisões. Um exemplo de árvore de decisão pode ser visto na Figura 8.

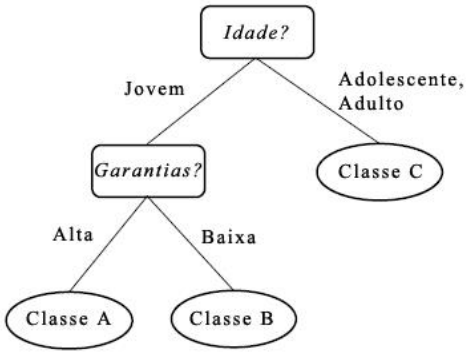

Figura 8: Exemplo de árvore de decisão.
Fonte: HAN E KAMBER, 2006a.

- *Random Forest* (Floresta Aleatória)

É uma técnica que utiliza uma combinação de múltiplas árvores de decisão. Dado um conjunto de dados, são amostradas aleatoriamente n instâncias, com possibilidade de

repetição, sendo n o número de instâncias do conjunto de dados original. Cada árvore é induzida a partir de um desses subconjuntos, porém cada nó da árvore utiliza m atributos da quantidade total f de atributos, sendo m=1 quando f=1, e$m < f$ quando$f > 2$. Os m atributos são escolhidos aleatoriamente e com repetição.

É necessário estabelecer a quantidade desejada de árvores de decisão presentes na *Random Forest*. Através do uso combinado dessa variedade de árvores de decisão é possível convergir o valor de erro para um valor que não sofreu *overfit*, ou sobre-ajuste, em relação ao conjunto de dados fornecido.

O conjunto de dados inicial é dividido em vários subconjuntos aleatórios e com repetição que, por sua vez, servirão para realizar a indução de cada árvore de decisão correspondente. A classificação final da *Random Forest* é feita através de votação, onde cada árvore de decisão providencia uma classificação, e a classe que for maioria entre os votos é escolhida como a classificação final da *Random Forest*.

- Classificação Bayesiana (*Bayesian Classification*)

É uma técnica estatística (probabilidade condicional) baseada no teorema de Thomas Bayes. Segundo o teorema de *Bayes*, é possível encontrar a probabilidade de um certo evento ocorrer, dada a probabilidade de um outro evento que já ocorreu: Probabilidade (B dado A) = Probabilidade (A e B)/Probabilidade (A). Comparativos mostram que os algoritmos Bayesianos, chamados de *Naive Bayes*, obtiveram resultados compatíveis com os métodos de árvore de decisão e redes neurais. Devido a sua simplicidade e o alto poder preditivo, é um dos algoritmos mais utilizados. O algoritmo de *Naive Bayes* parte do princípio que não exista relação de dependência entre os atributos. No entanto, nem sempre isto é possível. Nestes casos, uma variação conhecida como *Bayesian Belief Networks*, ou *Bayesian Networks*, deve ser utilizada.

- Redes Neurais (*Neural Networks*)

É uma técnica que tem origem na psicologia e na neurobiologia. Consiste basicamente em simular o comportamento dos neurônios. De maneira geral, uma rede neural pode ser vista como um conjunto de unidades de entrada e saída conectadas por camadas intermediárias e cada ligação possui um peso associado. Durante o processo de aprendizado, a rede ajusta estes pesos para conseguir classificar corretamente um objeto. É uma técnica que necessita de um

longo período de treinamento, ajustes finos dos parâmetros e é de difícil interpretação, não sendo possível identificar de forma clara a relação entre a entrada e a saída. Em contrapartida, as redes neurais conseguem trabalhar de forma que não sofram com valores errados e também podem identificar padrões para os quais nunca foram treinados. Um dos algoritmos mais conhecidos de redes neurais é o *backpropagation*, popularizado na década de 80, que realiza o aprendizado pela correção de erros. Na Figura 9 observa-se um exemplo de uma rede neural.

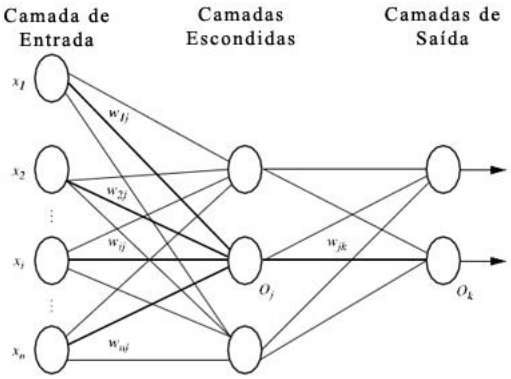

Figura 9: Exemplo de uma rede neural.
Fonte: HAN E KAMBER, 2006a.

- SVM (*Support Vector Machines*)

Apesar de ser uma técnica nova, tem chamado muita atenção pelos seus resultados: obtém altos índices de assertividade, permite modelar situações não-lineares complexas gerando modelos de simples interpretação, pode ser usada para relações lineares e não-lineares, entre outros.

As máquinas de vetores suporte (support vector machines, mais conhecida como SVMs), desenvolvidas por Boser *et al.* (1992), têm a capacidade de resolver problemas de classificação e regressão, adquirindo com o aprendizado na etapa de treinamento a capacidade de generalização. Considerando um problema binário, o objetivo da SVM é separar as instâncias das duas classes através de uma função que será obtida a partir dos exemplos conhecidos na fase de treinamento. O objetivo é produzir um classificador que funcione de forma adequada com exemplos não conhecidos, ou seja, exemplos que não foram aplicados durante o treinamento, adquirindo assim a capacidade de predizer as saídas de futuras novas entradas.

As funções de kernel têm a finalidade de projetar os vetores de características de entrada em um espaço de características de alta dimensão para classificação de problemas que

se encontram em espaços não linearmente separáveis. Isso é feito, pois à medida que se aumenta o espaço da dimensão do problema, aumenta também a probabilidade desse problema se tornar linearmente separável em relação a um espaço de baixa dimensão. Entretanto, para obter uma boa distribuição para esse tipo de problema é necessário um conjunto de treinamento com um elevado número de instâncias.

Existem vários tipos de kernels que podem ser usados, porém as mais conhecidas são: Polinomial, Gaussiano (ou RBF) e Sigmoidal.

- KNN (*K-Nearest Neighbors*)

É um classificador onde o aprendizado é baseado na analogia. O conjunto de treinamento é formado por vetores n-dimensionais e cada elemento deste conjunto representa um ponto no espaço n-dimensional. Para determinar a classe de um elemento que não pertença ao conjunto de treinamento, o classificador KNN procura K elementos do conjunto de treinamento que estejam mais próximos deste elemento desconhecido, ou seja, que tenham a menor distância.

Estes K elementos são chamados de K-vizinhos mais próximos. Verifica-se quais são as classes desses K-vizinhos e a classe mais frequente será atribuída à classe do elemento desconhecido. Existem algumas métricas mais comuns para o cálculo de distância entre dois pontos, sendo que a mais utilizada é a distância euclidiana. É um processo de classificação que pode ser computacionalmente exaustivo se considerado um conjunto com muitos dados. Para determinadas aplicações, no entanto, o processo é bem aceitável.

- AdaBoost

Para falar desta técnica, é necessário abordar as técnicas de *Bagging e Boosting*. O *Bagging (Bootstrap Aggregating)* é um método proposto por Breiman em 1996, tendo um conjunto de dados por amostragem bootstrap dos dados originais. O conjunto de dados gera um conjunto de modelos utilizando um algoritmo de aprendizagem simples por meio da combinação por votos para classificação. O seu uso é particularmente atraente quando a informação disponível é de tamanho limitado. No *Bagging* os classificadores são treinados de forma independente por diferentes conjuntos de treinamento através do método de inicialização. Para construí-los é necessário montar k conjuntos de treinamento idênticos e replicar esses dados de treinamento de forma aleatória para construir k redes independentes

por re-amostragem com reposição. Em seguida, deve-se agregar as k redes através de um método de combinação apropriada, tal como a maioria de votos (BREIMAN, 1996).

Já no *Boosting*, de forma semelhante ao *Bagging*, cada classificador é treinado usando um conjunto de treinamento diferente. A abordagem por *Boosting* original foi proposta por Schapire em 1990. A principal diferença em relação ao *Bagging* é que os conjuntos de dados re-amostrados são construídos especificamente para gerar aprendizados complementares e a importância do voto é ponderado com base no desempenho de cada modelo, em vez da atribuição de mesmo peso para todos os votos. Essencialmente, esse procedimento permite aumentar o desempenho de um limiar arbitrário simplesmente adicionando aprendizes mais fracos.

Dada a utilidade desse achado, *Boosting* é considerado uma das descobertas mais significativas em aprendizado de máquina (LANTZ, 2013). Já o *AdaBoost*, "*Adaptive Boosting*", é uma combinação das ideias de *Bagging e Boosting* e não exige um grande conjunto de treinamento como o *Boosting*. Inicialmente, cada exemplo de formação de um determinado conjunto de treinamento tem o mesmo peso. Para treinar o k-ésimo classificador como um "modelo de aprendizagem fraca", n conjuntos de amostras de treinamento (n <m) entre S são usadas para treinar o k-ésimo classificador. Em seguida, o classificador treinado é avaliado para identificar os exemplos de treinamento que não foram classificados corretamente (TSAI *et al.*, 2014). A rede k+1 é então treinada por um conjunto treinado modificado que reforça a importância desses exemplos classificados incorretamente.

Este procedimento de amostragem será repetido até que k amostras de treinamento sejam construídas para a construção da k-ésima rede. Portanto, a decisão final baseia-se na votação ponderada dos classificadores individuais.

- Regressão logística

É uma técnica estatística que produz, a partir de uma série de variáveis explicativas, um modelo que permita a predição de valores tomados por uma variável dependente categórica. Assim através de um modelo de regressão, é possível calcular a probabilidade de ocorrência de um evento, através da função de ligação, conforme descrita na Equação 1:

$$\pi(x)=\frac{e\ (\beta 0 +\beta 1\ x1 +\beta 2\ x2 +...\beta i\ xi\ )}{1 + e(\beta 0+\beta 1x1+\beta 2x2+...\beta ixi)} \qquad (1)$$

O $\pi(x)$ é a probabilidade de sucesso quando o valor da variável preditiva é x, $\beta 0$ é uma constante usada para ajuste e βi são os coeficientes das variáveis preditivas. Para encontrar a estimativa dos coeficientes β na Equação 1, é usada a técnica de máxima verossimilhança que maximiza a probabilidade de obter o grupo observado de dados, através do modelo estimado.

O modelo de regressão logística binária é um caso especial de modelo linear generalizado, que usa a função de ligação logit para obter as estimativas dos coeficientes da Equação 1. Após encontrar os coeficientes, é possível encontrar a probabilidade de sucesso, que nessa pesquisa é a probabilidade de fraude, aplicando na Equação 1 os valores das estimativas dos coeficientes encontrados.

3.4.1 Conceitos básicos

Nesta seção, serão apresentados alguns conceitos básicos de aprendizado de máquina utilizados durante o desenvolvimento deste projeto.

3.4.1.1 Medidas de desempenho – Modelos de Classificação

O primeiro conceito a ser apresentado para utilização de medidas de desempenho está relacionado à matriz de confusão, uma ferramenta muito usada para avaliações de modelos de classificação em aprendizado de máquina.

A matriz de confusão é uma matriz cuja dimensão corresponde ao número de classes existentes em um determinado conjunto de exemplos. A sua diagonal principal corresponde ao número de acertos de cada classe e os elementos fora da diagonal principal correspondem ao número de erros. Na Tabela 2 é mostrado um exemplo de uma matriz de confusão referente a um conjunto de exemplos com duas classes geralmente denominadas como positiva e negativa.

Tabela 2: Exemplo de uma matriz de confusão.

	Predição Positiva	Predição Negativa
Classe Positiva	Verdadeiro Positivo (VP)	Falso Negativo (FN)
Classe Negativa	Falso Positivo (FP)	Verdadeiro Negativo (VN)

Fonte: OSHIRO, 2013.

A matriz de confusão estendida (Figura 10) permite identificar algumas medidas de desempenho que são utilizadas nas análises de modelos, tais como: acurácia (accuracy), precisão (precision), revocação (recall), área abaixo da curva ROC (AUC - *Area Under the ROC Curve*) e F1 *score*.

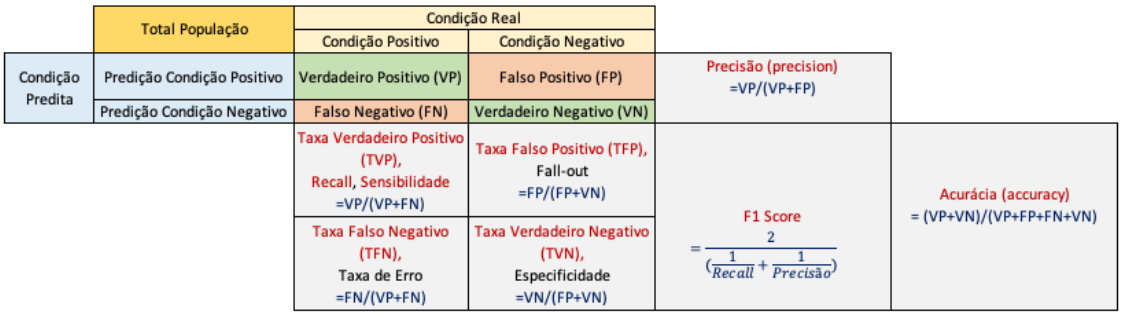

Figura 10: Matriz de desempenho estendida, identificando algumas métricas de desempenho.
Fonte: O autor.

Portanto, a partir da matriz de confusão são apresentadas as equações destas métricas:

- Acurácia (*accuracy*) (Equação 2): em uma classificação binária representa o quão correto as instâncias foram classificadas com sucesso, ou seja, indica uma performance geral do modelo. Dentre todas as classificações, quantas o modelo classificou corretamente:

$$\text{Acurácia} = VP + VN/(VP + VN + FP + FN) \quad (2)$$

Precisão (*precision*) (Equação 3): representa a porcentagem de instâncias que dentre todas as classificações de classe Positivo que o modelo fez, quantas estão corretas:

$$\text{Precisão} = VP/(VP + FP) \quad (3)$$

Revocação (*recall*) (Equação 4): representa a porcentagem do quão efetivo o classificador foi em detectar instâncias com insulto, ou seja, dentre todas as situações de classe Positivo como valor esperado, quantas estão corretas:

$$\text{Revocação} = VP/(VP + FN) \quad (4)$$

AUC (Equação 5): representa a porcentagem do quão efetivo o classificador foi em evitar falsas classificações:

$$\text{AUC} = ½ * (\text{revocação} + VN/(VN+FP)) \quad (5)$$

O cálculo dessa métrica através da análise ROC (*Receiver Operating Characteristic*), método gráfico para avaliação, organização e seleção de sistemas de diagnóstico e/ou predição, funcionam de forma muito semelhante à matriz de confusão. Ao plotar o eixo de ordenada a métrica de sensibilidade (taxa de verdadeiros positivos) e a o eixo de abscissa sendo os valores correspondentes a 1-especificidade (taxa de falsos positivos) obtém-se a curva ROC (Figura 11).

Prati *et al.* (2008) demonstram e explanam a importância da análise de gráficos ROC para avaliação de algoritmos de aprendizado de máquina.

Logo, a área sob essa curva é tida como uma medida de qualidade do classificador, pois quanto maior a área, melhor o desempenho do classificador. A Tabela 3 fornece um mapa para através do valor da área sob a curva ROC para definir o poder de classificação de um modelo:

Tabela 3: Poder de classificação de um modelo através do valor da área sob a curva ROC.

Valor da Área Sob a Curva ROC	Poder de Classificação
0,5	Não há
$0{,}7 \leq ROC < 0{,}8$	Aceitável
$0{,}8 \leq ROC < 0{,}9$	Muito Bom
$ROC \geq 0{,}9$	Excelente

Fonte: MATSUBARA, 2008.

Baseado na divisão por regiões abordado por Matsubara (2008), na Figura 11 é ilustrado um gráfico ROC identificando quatro regiões importantes do gráfico, descritas a seguir, e uma linha diagonal que representa classificadores aleatório:

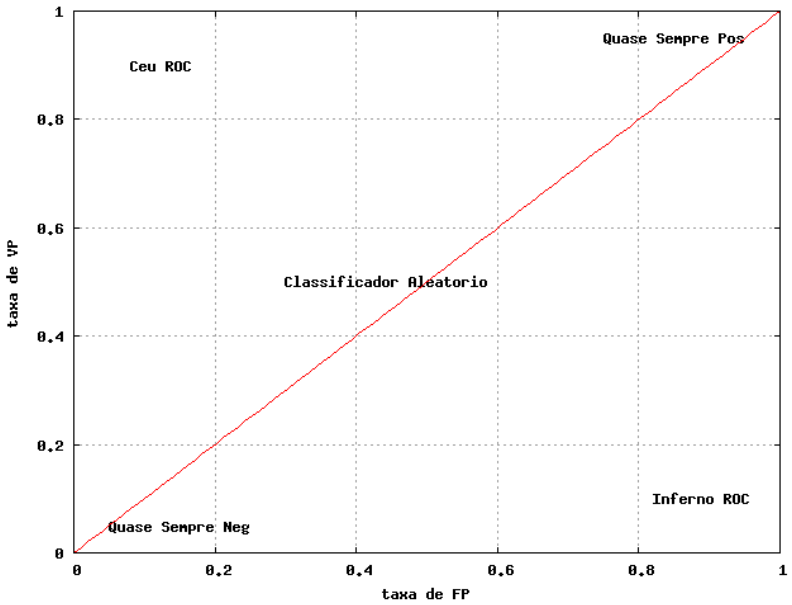

Figura 11: Exemplo de gráfico ROC demonstrando as regiões de classificação importantes quanto ao valor da área sob a curva ROC de classificadores.
Fonte: MATSUBARA, 2008.

As regiões destacadas na Figura 11 são as seguintes:

- Céu ROC: O ponto (0,1) representa uma classificação perfeita, na qual todos os exemplos positivos e negativos são rotulados corretamente. Na região Céu ROC encontram-se os pontos mais próximos da classificação perfeita e representam bons resultados.
- Inferno ROC: região localizada no lado oposto ao Céu, pode ser considerada uma região na qual são encontrados os resultados "ruins". No entanto, classificadores representados nessa região possuem informações com capacidade de distinguir as classes, mas não as utilizam corretamente (FLACH; WU, 2005). Observe que, ao inverter os rótulos, o ponto que era (x,y) passa a ser (y,x), passa do inferno para o céu ROC.
- Quase Sempre Neg: classificadores que são representados nessa região rotulam quase sempre os exemplos como negativos. Assim, o número de exemplos negativos rotulados errados normalmente é baixo (TFP próximo de 0) e número de exemplos positivos rotulados corretamente também é baixo (TVP próximo de 0).
- Quase Sempre Pos: classificadores que são representados nessa região rotulam quase sempre os exemplos como positivos. Assim, quase todos os exemplos positivos são rotulados corretamente (TVP próximo de 1), e quase todos os exemplos negativos incorretamente (TFP próximo de 1).

Normalmente, os classificadores representados por pontos próximos a linha diagonal são considerados classificadores aleatórios, não possuem informação sobre a classe alvo do modelo.

- F1 *score* (Equação 6): combina a precisão e a revocação, e pode ser interpretada como a média de ambas as medidas. Seu melhor valor é 1, e o pior 0.

$$\mathrm{F1} = 2 \times (precisão \times revocação)^{*}\, precisão + revocação \tag{6}$$

3.4.1.2 Medidas de desempenho – Modelos de Regressão

Para os modelos de regressão, a técnica de porcentagem de acerto não pode ser levada em consideração, pois dificilmente um modelo irá acertar o valor corretamente, com precisão. Entretanto, os modelos de predições obterão um valor aproximado do valor real, sendo assim, é necessário verificar o quão distante os valores preditos estão dos valores reais.

Para resolver este problema, os modelos de predição possuem algumas métricas próprias de avaliação. Entre diversas métricas, as que mais se destacam são: MSE, RMSE, MAE e o R-Quadrado.

- MSE (*Mean Squared Error*): O erro quadrático médio é comumente usado para verificar a acurácia de modelos e dá um maior peso aos maiores erros, já que, ao ser calculado, cada erro *e* (valor predito - valor real) é elevado ao quadrado individualmente e, após isso, a média desses erros quadráticos é calculada. Por conta do expoente ao quadrado que o erro assume, essa métrica é bastante sensível a *outliers* (valores discrepantes) e, caso tenha muitos erros significativos em sua análise, essa métrica poderá ser extrapolada. Um valor menor indica um modelo de qualidade superior.

A Equação 7 descreve o cálculo utilizado para obter a assertividade desta métrica.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n} e_i^2 \tag{7}$$

- RMSE (*Root Mean Square Error*): A raiz do erro quadrático médio resume-se em calcular o somatório dos erros *e* (valor predito - valor real) ao quadrado, realizar a média dos erros e então calcular a raiz quadrada. Este modelo permite uma variação do modelo

MSE, com a possibilidade de apresentar os valores do erro na mesma dimensão das variáveis analisadas (HALLAK; FILHO, 2011). A Equação 8 descreve o cálculo utilizado para obter a assertividade desta métrica. Um valor menor indica um modelo de qualidade superior.

$$RMSE= \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2} \tag{8}$$

- MAE (*Mean Absolute Error*): O erro médio absoluto é a média do somatório do módulo do erro. Este modelo permite que o valor final não seja tão afetado por valores anormais, outliers. A Equação 9 descreve o cálculo utilizado para obter a assertividade desta métrica. Um valor menor indica um modelo de qualidade superior.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |e_i| \tag{9}$$

- R-*Squared* (R^2): O valor do R-quadrado descreve a fração da variância total nos dados observados que pode ser explicada pelo modelo. É dado por 1 menos o resultado da divisão entre o valor do somatório dos erros ao quadrado e do valor do somatório dos valores reais menos a média dos valores reais ao quadrado (MILES, 2014). A Equação 10 descreve o cálculo utilizado para obter a assertividade desta métrica.

$$R^2=1-\frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} (y_i-\overline{y})} \tag{10}$$

Sendo $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, a média dos dados observados.

Em geral, a métrica R-quadrado indica que quanto maior for o seu valor, mais aderente é o modelo. Entretanto esta não é uma regra, visto que o R-quadrado individualmente não indica se um modelo de regressão é adequado ou não. É possível ter um valor baixo de R-quadrado para um bom modelo ou um valor alto de R-quadrado para um modelo que não se encaixa nos dados. Portano, é importante que ele seja avaliado em conjunto com outras de outras de avaliação estatística.

3.4.1.3 Técnicas de Seleção de Características

A relevância dos atributos pode ser determinada por diferentes técnicas de seleção. O conceito de entropia é utilizado em grande parte destas. A entropia de um espaço, utilizando o conceito de entropia de Shannon (SHANNON; WEAVER, 1949), é caracterizada pelo menor número de bits necessários para codificar uma informação. Quanto maior a quantidade de bits necessários para codificar uma informação maior a entropia desta (YANG; PEDERSEN, 1997).

A definição formal de entropia pode ser aplicada no contexto de classificação, onde a distribuição de instâncias ao longo das classes é tratada como a informação em questão. Assim, se as instâncias são distribuídas ao longo das classes, o número de bits para codificar a informação será alto, porque será necessário enumerar cada instância. Já se todas as instâncias estiverem em uma única classe, a entropia do espaço amostral é baixa, porque com um bit seria possível descrever se todas as instâncias estão na primeira classe ou não.

Assim, podemos dizer que um atributo, ou mesmo um espaço amostral, possui maior entropia se os dados estão espalhados ao longo das classes.

A partir desse conceito, destacam-se as técnicas de seleção a seguir:

1) *Information Gain*

O método *Information Gain* é supervisionado e se baseia em filtragem e com a avaliação univariada de atributos baseada em ranking. O mérito de um atributo é calculado através do seu ganho de informação para o modelo, baseando-se no conceito de entropia. O ganho de informação de um atributo é determinado pela redução da sua entropia, apresentada matematicamente pela Equação 11, que descreve o ganho de informação de um atributo x.

$$\text{InfoGain}(x) = H(C) - H(C|x) \tag{11}$$

$H(C)$ é a entropia da classe e $H(C|x)$ é a entropia da classe relativa ao atributo x, calculado pela Equação 12.

$$H(C|x) = \sum_{j=1}^{m} \left(\frac{|x_j|}{m}\right) H\left(C\middle|x=x_j\right) = p(x,c) \times \log\left(p(c|x)\right) \tag{12}$$

$H(C|x = x_j)$ é a entropia relativa ao subconjunto de instâncias que tem um valor x_j para o atributo x. Se x é um bom descritor para a classe, cada valor de x terá uma baixa entropia distribuída entre as classes, ou seja, cada valor deve estar predominantemente em uma classe.

2) *Gain Ratio*

A técnica de seleção de atributos *Gain Ratio* é uma versão ponderada da técnica *Information Gain*, que visa solucionar uma limitação da técnica de *Information Gain*, a qual tende a selecionar atributos com maior número de valores distintos. Um exemplo clássico seria um atributo com um identificador sequencial de cada instância. Se esse identificador fosse usado como um atributo, ele teria um ótimo ganho de informação e seria selecionado, uma vez que todas as instâncias com um determinado valor estão na mesma classe.

Para tratar essa deficiência, a técnica *Gain Ratio* procura selecionar atributos que maximizam o ganho de informação, enquanto minimizam o número de valores de um atributo. Para isso, divide-se o ganho de informação, pela entropia do atributo. Essa adaptação é apresentada na Equação 13.

$$Gain\ Ratio\ (x) = H(C) - H(C|x)/H(x) \tag{13}$$

3) Coeficiente de Gini (Gini *index*)

O coeficiente de Gini, criado pelo matemático italiano Conrado Gini, é um instrumento estatístico para medir a desigualdade de uma distribuição. O coeficiente de Gini é definido por uma razão com valores entre 0 e 1: o numerador é a área entre a curva de distribuição de Lorenz e a linha de distribuição uniforme (área a mostrada na Figura 12); o denominador é a área sob a linha de distribuição uniforme (área a + b mostrada na Figura 12). O índice de Gini é o coeficiente de Gini expresso em porcentagem, e é igual ao coeficiente de Gini multiplicado por 100.

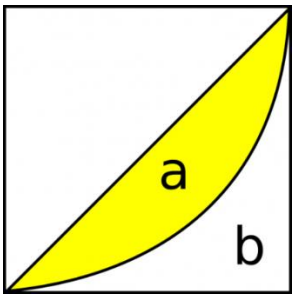

Figura 12: Coeficiente de Gini. O eixo horizontal representa a percentagem de pessoas, e o eixo vertical, a porcentagem da renda. A diagonal representa a igualdade perfeita de renda, o coeficiente de Gini = a / (a + b). Fonte: WIKIPEDIA[3].

4) O índice ANOVA (Análise de Variância)

ANOVA é a técnica estatística que permite avaliar afirmações sobre as médias de populações. O objetivo da técnica é analisar se existe uma diferença significativa de um ou mais fatores (também chamados de variáveis de entrada, ou variáveis X) comparando as médias das variáveis de resposta em diferentes níveis dos fatores.

A hipótese nula afirma que todas as médias das populações (médias dos níveis dos fatores) são iguais, enquanto a hipótese alternativa afirma que pelo menos uma é diferente.

Um procedimento estatístico que compara, dentro de um experimento, a variação devida aos tratamentos com a variação devida ao acaso. A hipótese testada na ANOVA é a de que as médias dos tratamentos não diferem entre si, a determinado nível de significância, a qual pode ser representada pela Equação 14 da forma:

$$H_o := \mu_1 = \mu_2 = \ldots = \mu_\alpha, \qquad (14)$$

α representa o número total de tratamentos do ensaio e μi representa a média populacional do i-ésimo tratamento, i = 1, ..., α .

O teste estatístico utilizado para comparar as médias dos tratamentos na ANOVA foi proposto por Ronald Aylmer Fisher, e é chamado teste z de Fisher. Atualmente, o teste foi suprido pelo seu equivalente teste F de Snedecor, ou simplesmente teste F. Caso o teste F seja significativo, sendo os tratamentos qualitativos, então a aplicação de testes de médias é feita, com o objetivo de investigar eventuais diferenças entre pares de médias específicas ou combinações lineares dessas médias.

[3] Disponível em:<https://en.wikipedia.org/wiki/Gini_coefficient>. Acesso em 20 de fev. 2020

5) A métrica do chi-quadrado (χ^2)

O teste qui-quadrado proposto pelo estatístico Karl Person em 1900 serve para comprovar se existem diferenças significativas entre duas distribuições quaisquer, sendo um dos principais testes para associação. Permite calcular o total de desvios entre o número de ocorrências observadas e o de esperadas, e observa sua probabilidade de ocorrência segundo uma distribuição χ^2 com número de graus de liberdade adquiridos da estrutura de contingência da forma: gl = (l-1) (c-1). Dessa maneira ele é adequado para testar a hipótese nula se não há relação entre as categorias.

Então, as hipóteses são dadas por:

- H_0: Não existe associação entre as categorias.
- H_1: Existe alguma associação entre as categorias.

A estatística usada para o teste é dada pela Equação 15:

$$\chi^2 = \sum_{i=1}^{l} \sum_{j=1}^{c} \left(\frac{O_{ij} - E_{ij)}^{\ 2}}{E_{ij}} \right. \qquad (15)$$

O_{ij} = número de casos observados na linha i da coluna j.

E_{ij} = número de casos esperados, sob H_0, na linha i da coluna j.

Sob a hipótese, a estatística dada tem distribuição Qui-quiadrado com (l-1) (c- 1) graus de liberdade. A frequência esperada para a célula (nij) é dada pela Equação 16:

$$E_{ij} = \frac{\eta_i * \eta_j}{\eta} \qquad (16)$$

As exigências para se aplicar o teste de qui-quadrado é que a amostra estudada tenha no mínimo 20 observações.

6) ReliefF

É a capacidade de um atributo distinguir entre classes em instâncias de dados semelhantes. O algoritmo Relief (KIRA; RENDELL, 1992) trabalha por meio da amostragem aleatória de exemplos do conjunto de dados e da localização do vizinho mais próximo da mesma classe e do vizinho mais próximo da classe oposta. Os valores dos atributos dos

vizinhos mais próximos são comparados aos da classe amostrada e utilizados para atualizar os pesos de relevância de cada atributo em relação à classe. Portanto o Relief foi proposto com o objetivo de avaliar a qualidade dos atributos, analisando a capacidade de distinguir as instâncias entre suas vizinhas mais próximas.

Esse processo é repetido um número *m* de vezes. A idéia do Relief é que atributos importantes devem diferenciar exemplos de classes diferentes e possuir valores similares para exemplos da mesma classe. A proposta original do algoritmo Relief, a qual permitia trabalhar com duas classes, foi posteriormente estendida no algoritmo ReliefF para lidar com ruído e conjuntos de dados contendo múltiplas classes (KONONENKO, 1994).

No ReliefF, a influência de ruído nos dados é amenizada por meio da distribuição da contribuição dos k vizinhos mais próximos da mesma classe do exemplo correntemente considerado e de k vizinhos mais próximos de cada uma das classes diferentes do exemplo amostrado, ao invés de considerar apenas um único vizinho mais próximo. É interessante notar que quanto maior o valor de *m*, o número de exemplos amostrados a partir do conjunto de dados, mais confiáveis são as estimativas fornecidas pelo algoritmo ReliefF, embora aumentar *m* signifique aumentar o tempo necessário para a execução desse algoritmo.

ReliefF apresenta uma complexidade de tempo de O ($m \cdot N \cdot M$), onde N é a quantidade de exemplos do conjunto de dados, M é o número de atributos desse conjunto de dados e *m*, como mencionado anteriormente, o número de vezes que o algoritmo procura por exemplos no conjunto de dados para calcular os pesos para os atributos (ROBNIK-SIKONJA; KONONENKO, 2003).

7) FCBF (Filtro Rápido de Correlação Baseada)

O algoritmo FCBF (YU; LIU, 2004) realiza a seleção de atributos em duas etapas: primeiramente, os atributos são analisados para determinar o subconjunto de atributos relevantes em relação à classe, removendo os atributos irrelevantes; na segunda etapa, por meio da análise de redundância, são determinados e removidos os atributos redundantes a partir do subconjunto que contém apenas os atributos relevantes, produzindo o subconjunto final de atributos selecionados.

Nesse algoritmo é utilizada a medida *Symmetrical Uncertainty* — SU (PRESS *et al.*, 1992) como a medida de correlação para aproximar tanto a análise de relevância quanto a análise de redundância.

Assim, na primeira etapa, a medida SU entre cada atributo e a classe é calculada para todos os atributos, os quais são classificados de acordo com sua relevância em relação à classe. Apenas os atributos que possuírem um valor SU maior que um limiar mínimo, que determina quão relevantes os atributos devem ser para serem considerados, são analisados na próxima etapa. Na segunda etapa, os atributos são avaliados na ordem em que foram classificados na etapa anterior, de acordo com a redundância de uns em relação aos outros, produzindo um subconjunto final contendo apenas os atributos relevantes e não redundantes. É importante notar que no algoritmo FCBF os atributos numéricos são discretizados utilizando o algoritmo para discretização de atributos *Minimum Description Length* (MDL) proposto por Fayyad e Irani (1993).

O FCBF apresenta a vantagem, sobre as abordagens tradicionais para avaliação de subconjuntos de atributos, de que por meio da separação das tarefas de análise de relevância e de redundância, ele evita o alto custo da busca por subconjuntos. Esse algoritmo apresenta uma complexidade de tempo de $O(M^2)$, M é o número de atributos desse conjunto de dados (YU; LIU, 2004).

3.5 *Orange data mining*

Uma das ferramentas que está em acentudada utilização na área da mineração de dados, o *Orange Data Mining* é uma ferramenta de código aberto (*open source*) para classificação, regressão e tarefas descritivas de dados, sendo compatível com as principais plataformas e sistemas operativos existentes no mercado, como por exemplo a utilização de bibliotecas de *Python* comuns para computação científica, como *numpy*, *scipy* e *scikit-learn*.

O fluxo de execução da ferramenta é dado por uma estrutura de nós, em que cada nó adicionado ao campo de fluxo executa uma determinada tarefa. Também apresenta um sistema de retorno ao usuário (*feedback*) para cada método, retornando-lhe os dados de entrada e de saída de cada método. Com interface gráfica bastante intuitíva (funções arrastar e soltar – *drag and drop*), permite que os usuários se concentrem na análise exploratória de dados, em vez de codificação. Também possui componentes para *machine learning* e

complementos de mineração de dados de fontes externas para execução de processamento de linguagem natural, mineração de texto, bioinformática, análise de rede e mineração de regras de associação.

Segundo Viterbo *et al.* (2016), dentre 5 ferramentas de mineração de dados avaliadas em seu estudo com estudantes, a *Orange Canvas* é a que apresenta a interface onde é mais fácil encontrar as informações e elementos desejados.

Matos *et al.* (2019) realizou um estudo utilizando técnicas e mineração de texto com *Orange Canvas* para encontrar padrões nos relatos das gestantes contidos nos documentos de planos de parto, no domínio da saúde obstétrica.

4 MATERIAIS E MÉTODOS

Este trabalho é composto do estudo de métodos matemáticos, análise de dados, técnicas de mineração de dados e interface com o processo produtivo em tempo real.

O processo KDD foi utilizado como metodologia aplicada para este estudo, explorando em cada uma de suas etapas ferramentas e métodos de forma a extrair e preparar as informações para as etapas seguintes até que o resultado seja atingido.

As ferramentas utilizadas foram o Microsoft Excel 2016, o software *Plant Information Management System* (PIMS) da empresa brasileira em estudo e o *Orange Canvas* 3.23.

Foram realizadas as seguintes etapas do processo de KDD: seleção, pré-processamento e limpeza, transformação, mineração, interpretação e avaliação dos dados (FAYYAD *et al.*, 1996).

Neste estudo, duas análises foram feitas para alcançar os objetivos propostos. A primeira foi feita para ranquear as variáveis de maior influência do processo e determinar os *setups* ótimos para operação da prensa. Já a segunda análise permitiu determinar o modelo de predição do valor da variável de superfície específica da prensa.

4.1 Ranqueamento e *setups* ótimos de operação

4.1.1 Seleção dos dados

O método utilizado para exploração dos dados neste trabalho é denominado de extração dos dados. Foram definidas a(s) fonte(s) que se relacionam com o domínio para a extração dos dados apropriados para o contexto.

Para esta etapa os dados foram extraídos da fonte de dados temporal do sistema PIMS através da seleção de todas as variáveis disponíveis para o domínio da área de prensagem, um total de 40 variáveis. Dentre estas foram selecionadas as que se relacionam diretamente com o processo de prensagem, excluindo aqueles referentes à proteção do equipamento da prensa (medições de vibração e temperatura), totalizando 16 variáveis para este estudo.

O intervalo de tempo de coleta dos dados foi de 10 minutos, durante o período de 12/02/2019 a 01/10/2019, sendo gerado um total de 32.998 registros (linhas) por 16 variáveis (colunas).

Para esta etapa foram utilizados os softwares AspenTech – Infoplus como fonte de dados e o Microsoft Excel 2016 para o devido tratamento e manipulação dos dados.

A Figura 13 mostra a relação entre as 16 variáveis selecionadas de entrada (dados coletados), a variável de meta de predição, o desenvolvimento do modelo utilizando técnicas de aprendizado de máquina e os resultados esperados (saída) para este trabalho.

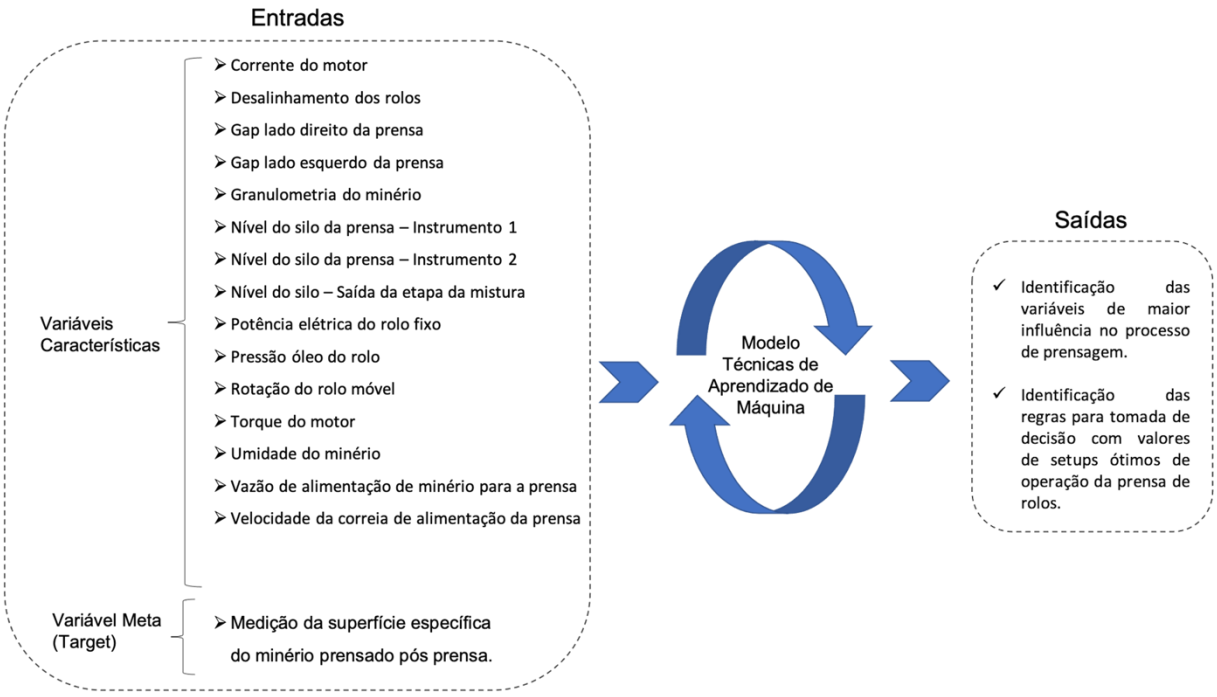

Figura 13: Esquema de entrada de dados e modelo de aprendizagem para análise de ranqueamento das variáveis de maior influência do processo e determinação de *setups* ótimos para operação da prensa.
Fonte: O autor.

4.1.2 Pré-processamento e limpeza

Após o subconjunto selecionado dos dados, foram identificados alguns erros, como dados ausentes, dados com erro, registros duplicados e ruídos, tornando-se necessário tratá-los por um processo de integração, padronização e limpeza, sendo gerado um subconjunto que represente com mais exatidão o domínio de estudo.

A limpeza dos dados foi realizada utilizando-se a ferramenta de filtro do software Excel conforme as seguintes regras:

- Seleção de dados relacionados ao período de funcionamento da prensa de rolos: realizado filtro da coluna binária da variável contendo o valor do estado de

funcionamento (célula de valor igual a 1) e de não funcionamento (célula de valor igual a zero).

- Seleção de dados relacionados ao período da vazão da balança de alimentação de minério da prensa de rolos: realizado filtro na coluna de medição da vazão maior do que 100 t/h.

Após a limpeza dos dados, a exclusão de valores discrepantes, dados cujos valores se distanciam dos demais, foi realizada utilizando o componente *outlier* do Orange, um estimador de detecção externa que tenta ajustar as regiões onde os dados são os mais concentrados, ignorando as observações divergentes.

O método de detecção utilizado neste componente foi o *One-Class SVM with non-linear kernel*, um algoritmo não supervisionado que aprende uma função de decisão para a detecção de novidades, classificando novos dados como semelhantes ou diferentes do conjunto de treinamento. O SVM é um classificador que utiliza o princípio de maximização da margem. O princípio da margem máxima coloca a superfície de decisão exatamente entre o limite das duas classes e maximiza a distância do limite das classes. Segundo Hamel (2009), essa abordagem reduz a probabilidade de erro de classificação.

Atualmente não existe um método universal para guiar a seleção de parâmetros do kernel do SVM (BONESSO, 2013). Campos (2015) utilizou para a classe *outlier* uma subamostragem com valores de 20%, 10%, 5% e 2%. Já Keller *et al.* (2012) subamostram para 10% da quantidade original dos dados para produzir a classe *outlier*. Portanto, neste estudo foi considerado uma subamostragem de 10% para cada conjunto de dados.

Essa proporção é o valor dado ao parâmetro *nu* do *One-Class SVM* e o parâmetro de contaminação dos outros algoritmos de detecção de *outlier*.

Para isso utilizou-se o componente de cálculo *outlier* do software Orange, parametrizado da seguinte maneira:

a) Método de detecção: *Support Vector Machine* (SVM) de uma classe não linear com função kernel de base radial (RBF).

b) Parâmetro Nu: Limite superior da fração de erros de treinamento e um limite inferior da fração de vetores de suporte, com valor parametrizado de 10%.

c) O coeficiente do kernel: é um parâmetro gama, que especifica a influência que um único dado tem nesta instância, com valor parametrizado de 0,01.

4.1.3 Transformação dos dados

Após a limpeza dos dados foram realizados dois processos de transformação:

a) Discretização da variável meta: o valor analógico da variável de superfície específica foi discretizada em função do valor de referência de meta. Caso o valor do dado seja maior de 2.100 g/cm³, diz-se que o dado $\in$ ao conjunto "Meta" (verdadeiro ou igual a 1: valor dentro da meta estabelecida), caso contrário o dado $\notin$ ao conjunto "Meta" (falso ou valor igual a zero: valor abaixo da meta estabelecida).

b) Normalização de todos os dados com o propósito de minimizar os problemas oriundos do uso de unidades e dispersões distintas entre as variáveis coletadas. Para isso foi utilizado a normalização segundo a amplitude do dado através da Equação 17.

$$y=\frac{x\text{-}Xmin}{Xmax\text{-}Xmin} \tag{17}$$

Nessa equação, Y é o resultado do valor normalizado, x é o valor da variável medida, Xmax é o valor máximo e Xmin o valor mínimo dessa variável.

Campos (2015) atribui a importância da normalização dos dados em casos em que ocorra uma dominância artificial de alguns atributos meramente por variações grandes de escala.

A classificação de cada dado da variável de superfície específica e o cálculo para a normalização de cada dado foi realizado implementando-se a Equação 17 em um código programado na linguagem do *Visual Basic for Applications* (VBA) no software Microsoft Excel versão 2016 (Anexo A).

4.1.4 Mineração de dados

Nesta etapa, os dados foram explorados e analisados de forma supervisionada, com o intuito de perceber padrões em grandes fontes de dados e assim obter informações relevantes.

Para a execução desta etapa foi utilizado o software Orange a fim de explorar os componentes matemáticos visando atingir os resultados.

Com os dados transformados e salvos em arquivo .xlsx (extensão de arquivo do software Microsoft Excel 2016), este foi importado no software *Orange*. Todas as variáveis foram categorizadas como numéricas, exceto a variável de saída (superfície específica), categorizada como categórica.

A variável de saída foi categorizada da seguinte maneira:

- Valor da variável de superfície específica >= 2.100 g/cm^3: variável de saída denominada Valor_Meta ∈ "Meta" (valor igual a 1).
- Valor da variável de superfície específica < 2.100 g/cm^3: variável de saída denominada Valor_Meta ∈ "AbaixoMeta" (valor igual a 0).

Essa classificação permitiu a discretização desta variável para utilização nos modelos de predição descritos mais adiante.

Em seguida, o componente *Select Columns* foi utilizado para selecionar as variáveis denominadas *Features* (características), as variáveis de entrada do modelo, e a variável *Target*, variável a ser obtida na classificação do modelo.

O componente o*utlier* foi inserido e configurado conforme descrito na etapa de pré-processamento dos dados. Como resultado deste componente, foram fornecidos dois grupos de variáveis: grupos das variáveis *inliers* e *outliers.*

4.1.5 Interpretação e avaliação

Essa etapa utilizou a interpretação e avaliação dos padrões encontrados pela etapa de mineração de dados. Foram utilizadas diversas ferramentas com funcionalidades estatísticas e de visualização para validarem ou julgarem um padrão irrelevante.

Utilizando o componente *Rank* do Orange, responsável por calcular as métricas de avaliação e ranqueamento para os métodos de classificação, foram atribuídas como entradas os dados *Inliers* resultante da etapa de mineração de dados. A partir da execução deste componente foi obtido o resultado das principais variáveis de maior influência e correlação para o ganho de superfície específica do processo de prensagem (variável meta deste estudo).

A partir do resultado obtido no ranqueamento das variáveis, a modelagem foi realizada por meio de 8 métodos de classificação do Orange: 1-Árvore de decisão (*Tree*), 2-SVM, 3-

Naive Bayes, 4-*kNN*, 5-Rede Neural (*Neural Network*), 6-*AdaBoost*, 7-Regressão Logística (*Logistic Regression*) e 8-*Random Forest* (Figura 14).

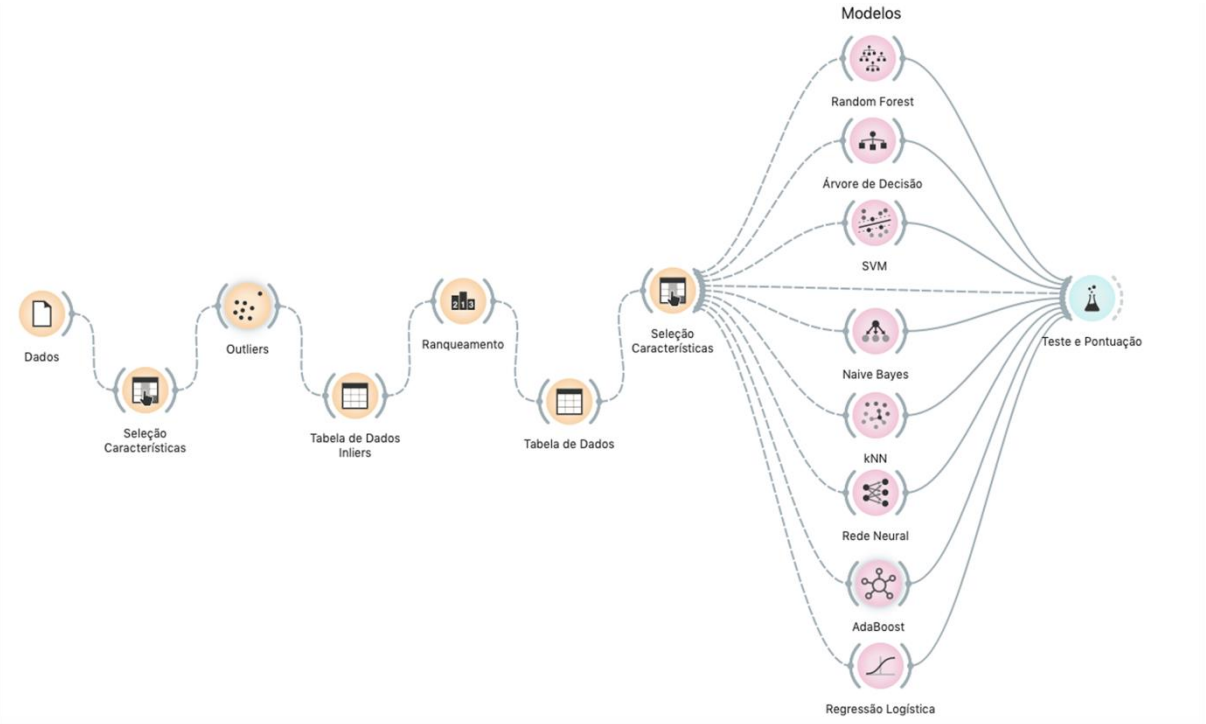

Figura 14: Etapas do processo KDD modeladas no *Orange* para obter o modelo de classificação. Alocação e aplicação dos componentes de aprendizado de máquina: *Random forest*, Árvore de decisão (*Tree*), SVM, *Naive bayes*, *kNN*, Rede neural (*Neural Network*), *AdaBoost* e Regressão logística (*Logistic Regression*).
Fonte: O autor.

Para cada modelo foi utilizado o seu respectivo componente no software *Orange* e seus resultados foram avaliados pelo componente *Test and Score*, utilizando as medidas de desempenho para escolha do melhor método a ser utilizado.

Este componente é responsável pelo teste e validação da amostragem dos dados de entrada, tendo como saída o resultado dos testes dos algoritmos de classificação que estão sendo utilizados. Para este trabalho adotou-se a utilização do método de amostragem *Cross-validation* para aprendizado e treinamento e teste da base de dados.

O método de *Cross-validation* é uma técnica para medir como os resultados de uma análise estatística serão generalizados para um conjunto de dados independente. A cada rodada do *cross-validation* é realizado o particionamento (*folds*) de uma amostra de dados em subconjuntos complementares, executando a análise de um subconjunto (chamado de conjunto de treinamento), e validando a análise em outro subconjunto (chamado de conjunto de validação ou teste). Para reduzir a variabilidade, múltiplas rodadas do *cross-validation* são

executadas usando diferentes e aleatórias partições, e os resultados de validação são a média de todas as rodadas. Neste trabalho utilizou-se 10 *folds cross-validation* para treinamento e validação dos modelos de aprendizagem de máquina: árvore de decisão (*Tree*), SVM, *Naive Bayes*, kNN, Rede Neural (*Neural Network*), *AdaBoost*, Regressão Logística (*Logistic Regression*) e *Random Forest.*

4.2 Predição do valor de superfície específica

4.2.1 Seleção dos dados

Para esta etapa foram utilizados o mesmo intervalo de dados e ferramentas para coleta e seleção descritos em 4.1.1.

Entretanto não foram selecionadas características referentes ao processo, mas utilizadas todas as 40 variáveis disponíveis na base de dados coletadas, totalizando 32.998 registros (linhas) por 40 variáveis (colunas). O intervalo de tempo de coleta dos dados foi de 10 minutos, durante o período de 12/02/2019 a 01/10/2019.

A Figura 15 mostra a relação entre as 40 variáveis selecionadas de entrada (dados coletados), a variável de predição, o desenvolvimento do modelo utilizando técnicas de aprendizado de máquina e os resultados esperados (saída) para este trabalho.

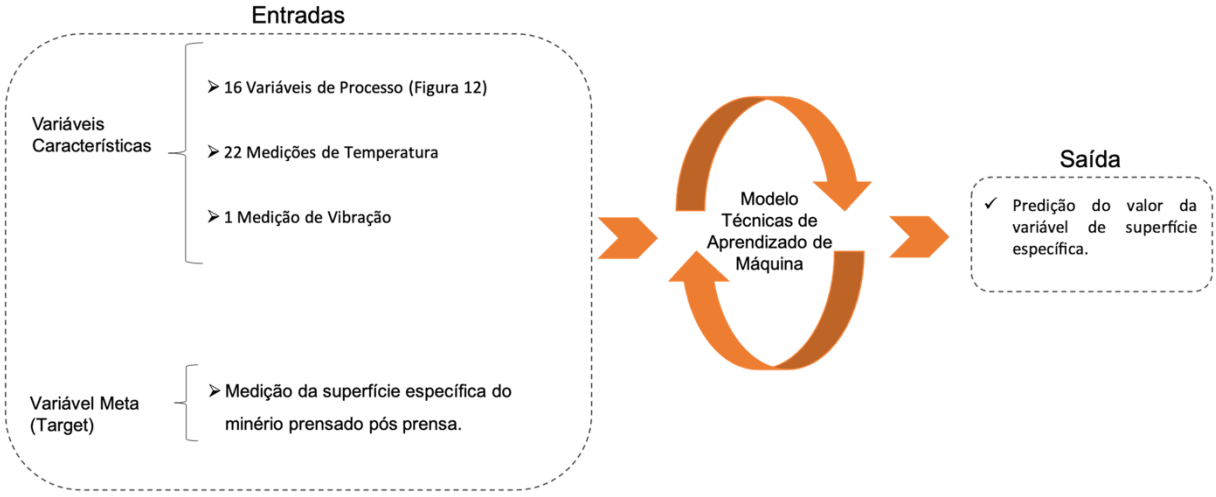

Figura 15: Esquema de entrada de dados e modelo de aprendizagem para determinação do modelo de predição do valor da variável de superfície específica da prensa.
Fonte: O autor.

4.2.2 Pré-processamento e limpeza

Para o desenvolvimento desta etapa foi utilizada a mesma metodologia do item 4.1.2.

4.2.3 Transformação dos dados

Para o desenvolvimento desta etapa foi utilizada a mesma metodologia do item 4.1.3.

4.2.4 Mineração de dados

Para o desenvolvimento desta etapa foi utilizada a mesma metodologia, parametrização e utilização dos componentes *Select Columns* e *Outliers* do item 4.1.4.

Entretanto a variável de saída (superfície específica) foi categorizada como numérica e *Target* do modelo, não sendo necessário sua discretização visto o objetivo ser a predição do valor dessa variável pelo modelo desenvolvido.

4.2.5 Interpretação e avaliação

Essa etapa utilizou a interpretação e avaliação dos padrões encontrados pela etapa de mineração de dados. Foram utilizadas diversas ferramentas com funcionalidades estatísticas e de visualização para validarem ou julgarem um padrão irrelevante.

Devido ao grande número de variáveis, foi utilizado o componente *Rank* para selecionar as 12 principais variáveis na determinação do modelo de predição.

A partir do resultado obtido no ranqueamento das variáveis, a modelagem foi realizada por meio de 5 métodos de classificação do *Orange*: 1 - SVM, 2 - *KNN*, 3 - Rede Neural (*Neural Network*), 4 - *AdaBoost*, 5 - *Random Forest* (Figura 16).

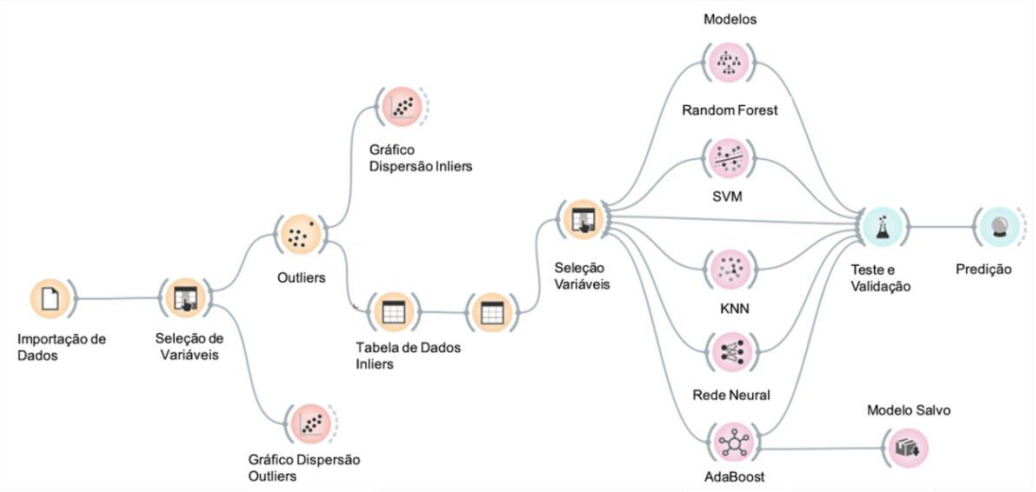

Figura 16: Etapas do processo KDD modeladas no *Orange* para obter o modelo de regressão. Alocação e aplicação dos componentes de aprendizado de máquina: *Random Forest*, SVM, *KNN*, Rede Neural (*Neural Network*) e *AdaBoost*.
Fonte: O autor.

Os componentes de teste, avaliação, validação e método de amostragem bem como as configurações dos componentes foram utilizados conforme descrito em 4.1.5.

5 RESULTADOS E DISCUSSÕES

Nesta seção são apresentados os resultados das duas análises realizadas visando alcançar os objetivos propostos.

Para determinar o ranqueamento das variáveis de maior influência e os *setups* ótimos de operação foram utilizados a base de dados com seleção das 16 características de processo, com isso espera-se que esses resultados possam cooperar no suporte à tomada de decisão pelos operadores e engenharia de processo.

Já para determinar a predição da variável de superfície específica, foram utilizadas 40 variáveis, sem atribuição de seleção de característica, sendo este resultado de grande valia para cooperar na avaliação do desempenho do processo de prensagem.

Adiante são descritos os resultados e discussões em cada processo realizado.

5.1 Ranqueamento e *setups* ótimos de operação

5.1.1 Pré-processamento e limpeza dos dados

Na etapa de pré-processamento e limpeza dos dados, obteve-se para cada uma das 16 variáveis, a seleção dos conjuntos de dados *inliers* e *outliers.* Os dados *outliers* foram excluídos da amostra e somente os dados *inliers* foram utilizados nas etapas posteriores.

Comparando-se as variáveis Corrente do Motor e Nível do Silo da Prensa, por exemplo, observa-se que os dados *inliers* resultantes da execução do algoritmo *One-Class* SVM com *kernel* de detecção de *outliers* não linear apresentam uma melhoria considerável para o conjunto de dados quando comparados aos dados originais do banco de dados (Figura 17), demonstrando os dados mais concentrados e selecionados em torno da média de valores normalizados entre as faixas 0,4 e 0,8. Os valores distantes dessa média de valores, aproximadamente menores do que 0,4 no eixo da abscissa, bem como a densidade de valores entre 0 e 0,5 no eixo da coordenada sofreram grande redução, mostrando que os valores de maiores distâncias da faixa de valores média foram expurgados e detectados como *outliers*.

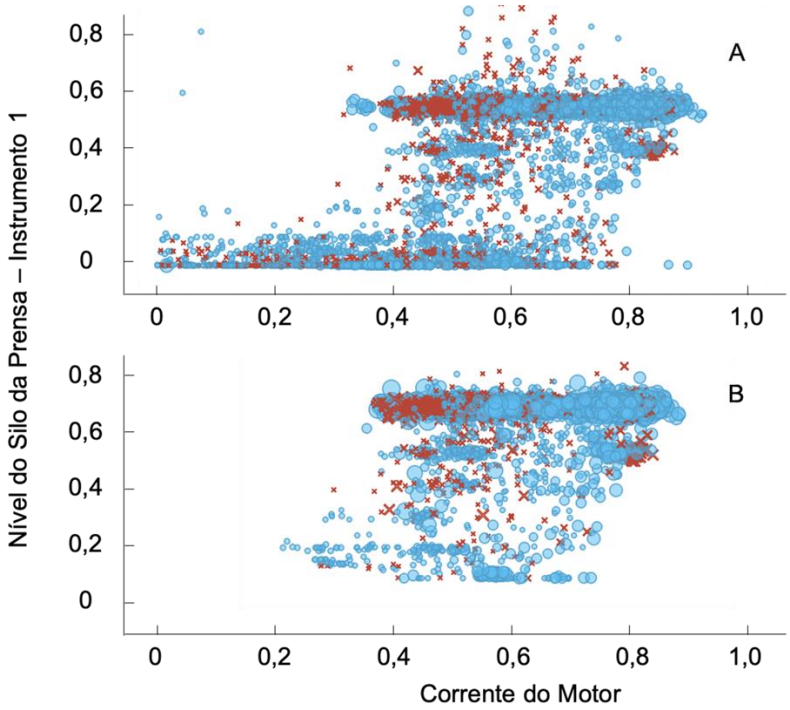

Figura 17: Exclusão de *outliers* do banco de dados da prensa de rolos. Valores e densidade da variável "Nível do Silo da Prensa - Instrumento 1" em relação à variável "Corrente do Motor". A: Sem tratamento de dados para exclusão dos *outliers*. B: Com tratamento de dados para exclusão de *outliers*. Em azul são mostrados os valores classificados como “AbaixoMeta” e em vermelho os valores classificados como “Meta”.
Fonte: O autor.

A função de densidade de probabilidade para os dados originais da variável contínua Corrente do Motor obteve média μ=0,656337 e desvio padrão α = 0,134441, já para os dados *inliers* provenientes do algoritmo de detecção de *outliers* para esta mesma variável, obteve valores de média μ=0,667163 e desvio padrão α = 0,115963. Novamente, comparando estes resultados, ratifica-se a redução do desvio padrão dos valores após a execução do algoritmo e a elevação do valor da média devido a concentração dos dados resultantes, conforme objetivo proposto para esta etapa.

Após a execução do componente *Outlier*, 3299 dados foram classificados com *outliers* e 29699 como *inliers*. Os dados *inliers* foram, portanto, utilizados para as etapas seguintes, permitindo que os modelos e resultados fossem obtidos com a nova base de dados, mais consistente e agrupada em regiões de maior significância.

5.1.2 Identificação das variáveis de maior influência no processo de prensagem

As variáveis mais significativas para o aumento do valor da variável de superfície específica (variável de saída nesta análise) foram identificadas utilizando o componente *Rank*

do Orange com avaliação dos métodos de seleção *Information Gain*, *Information Gain Ratio*, *Gini*, *Anova*, Qui-quadrado (X^2), *ReliefF* e FCBF conforme Tabela 4.

Sharma e Dey (2012), com relação aos métodos de seleção de recursos, o *Gain Ratio* apresenta o melhor desempenho em quase todos os métodos de aprendizado de máquina. Dağ *et al.* (2012) utilizaram os métodos de *Information Gain* e *Gain Ratio* com opção de benchmarks para a seleção de atributos.

Utilizando os conhecimentos da engenharia de processo e os resultados dos métodos de seleção, percebeu-se uma maior coerência do ranqueamento obtido das variáveis para os métodos *Information Gain*, *Gain Ratio* e *Gini* quando comparado com os demais métodos.

Desta maneira, neste trabalho, o método *Gain Ratio* foi utilizado para o determinar o ranqueamento das variáveis.

Portanto os resultados do ranqueamento por ordem de significância são mostrados na Tabela 4.

Tabela 4: Resultado do ranqueamento das variáveis ordenadas pelo método *Gain Ratio*.

Variáveis	**Método de Seleção**						
	Information Gain	**Gain Ratio**	**Gini**	**Anova**	**X^2**	**ReliefF**	**FCBF**
Umidade	0,055	0,028	0,033	105,461	48,920	0,003	0,040
Corrente do Motor	0,020	0,010	0,011	470,812	367,736	0,013	0,000
Torque do Motor	0,020	0,010	0,011	467,472	361,514	0,012	0,000
Desalinhamento dos Rolos	0,011	0,006	0,006	42,330	121,978	0,005	0,000
Nível do Silo - Saída da etapa da mistura	0,009	0,005	0,005	338,071	209,876	0,023	0,000
Vazão de Minério	0,008	0,004	0,004	320,527	254,451	0,020	0,000
***Gap* Lado Esquerdo**	0,005	0,003	0,003	12,097	71,752	0,008	0,000
Pressão de Óleo do Rolo	0,005	0,002	0,003	27,322	27,964	0,017	0,000
Velocidade da Correia de Alimentação	0,002	0,001	0,001	77,219	46,901	0,018	0,000
***Gap* Lado Direito**	0,002	0,001	0,001	6,159	13,855	0,011	0,000
Rotação Rolo Móvel	0,001	0,000	0,000	105,758	18,405	0,012	0,000
Potência Elétrica	0,001	0,000	0,000	28,581	18,110	0,010	0,000
Nível da Silo da Prensa - Instrumento 1	0,000	0,000	0,000	32,640	3,198	0,001	0,000
Nível da Silo da Prensa - Instrumento 2	0,000	0,000	0,000	31,841	2,014	0,002	0,000
Granulometria do Minério	0,000	0,000	0,000	-7,580	NA	0,000	0,000

Fonte: O autor.

A Figura 18 mostra, dentre as 16 variáveis que estão relacionadas ao processo de prensagem, aquelas mais significativas para o resultado da superfície específica da prensa.

Figura 18: Classificação das variáveis mais significativas para o resultado específico da prensa de rolos.
Fonte: O autor.

Estes resultados mostram que a umidade do minério é a variável de maior interferência neste processo. Segundo Saramak e Kleiv (2013), há uma faixa de umidade ótima para cada distribuição granulométrica da alimentação em conjunto, sob uma determinada condição operacional, que interfere diretamente na cominuição do minério na prensa de rolos.

A importância da medição de corrente do motor, ranqueada em segundo lugar por este método, abre um horizonte de discussão já que atualmente não existe nenhuma referência direta desta no processo, ou seja, não lhe é atribuída a criticidade de impacto no processo como foi mostrado no resultado deste trabalho. Isso poderá permitir um estudo mais profundo sobre seus impactos no desempenho da prensa, principalmente por ser uma grandeza elétrica com alto dinamismo para controle e diagnóstico.

O torque do motor já era uma variável esperada no topo de ranqueamento da prensa, devido ao fato de ser a principal variável de controle no processo de prensagem atualmente (sistema especialista baseado em regras *Fuzzy* para otimização do desempenho da prensa).

Portanto, esse resultado sustenta a validação do controle utilizado no processo produtivo. Para dimensionar a possível correlação do torque com a corrente elétrica do motor, foi realizada a correlação entre ambas com resultados de + 0,605 (coeficiente de correlação de Pearson), ou seja, embora demonstre certa correlação, ainda se faz necessário uma avaliação individual destas variáveis para futuras análises e comportamento destas no processo.

A variável de desalinhamento dos rolos está relacionada com a dispersão granulométrica durante o processo de cominuição, podendo ocorrer na aplicação de uma força desproporcional ao longo dos rolos devido ao desalinhamento. Esse fato interfere diretamente no processo de quebra dos grãos e consequentemente no ganho de superfície específica da prensa. Atualmente, essa variável atua somente como proteção do equipamento. A partir deste resultado, observa-se a possibilidade de uma discussão mais profunda sobre a influência desta medida devido ao grau de significância encontrado.

Outra variável importante é o nível do silo, resultante da etapa anterior ao processo de prensagem, etapa da filtragem. Este nível tem impacto direto no nível de alimentação do silo da prensa, que segundo Oliveira (2016), o chute de alimentação da prensa não pode limitar o fluxo de material à zona de compressão, o que impacta diretamente na capacidade e desempenho deste equipamento. Portanto, manter um nível constante de abastecimento demonstra um forte indício e relevância no processo. Ainda segundo Oliveira (2016), a capacidade específica é um dos principais aspectos de desempenho da prensa, tendo a distribuição granulométrica como um de seus fatores de grande importância. A abertura operacional (diretamente proporcional ao *gap* entre os rolos) tem alto grau de influência na distribuição granulométrica, sendo identificado nas variáveis de distâncias de gap dos lados esquerdo e direito do resultado obtido. O fato de estarem ranqueados em posições diferentes abre uma discussão sobre a possibilidade de a prensa estar trabalhando em estado de desalinhamento.

A pressão de óleo que é aplicada ao rolo móvel é outra variável sinalizada com alto grau de importância, sendo esta a responsável pelo ajuste de incremento/decremento de pressão do rolo móvel sobre o minério.

A velocidade de rotação é utilizada atualmente como a variável manipulada (variável de saída) do controlador PID (proporcional-integral-derivativo) existente para controle do nível do silo de alimentação da prensa. Devido a complexidade para medição deste nível,

foram instalados dois instrumentos distintos (medidor radar e outro de célula capacitiva), visando manter a estabilidade da medição para o processo de prensagem de minério. Ambas as variáveis não foram ranqueadas como as mais importantes para o processo, demonstrando que possivelmente, devido a estabilidade do controle PID, não foi possível verificar alta interferência destas variáveis do ganho direto da prensa. O fato de ambas as medições de nível serem ranqueadas com o mesmo peso sugere que o modelo é coerente na análise de classificação e ranqueamento.

O resultado de classificação da potência elétrica constata a baixa importância desta grandeza, assim como verificado no processo atualmente.

Por fim, o resultado de classificação da variável de granulometria do minério parece não estar de acordo com o esperado, pois esta é de grande importância no desempenho do processo. Segundo Campos *et al.* (2017), uma série de desafios envolve o processo de prensagem de *pellet feed* do ponto de vista da granulometria do produto e, em alguns casos, da sua alta umidade. Fatos estes que dificultam o aumento da área superficial do material e acarretam uma maior dificuldade no controle de processo. Diante disso, conclui-se que a base de dados para esta variável deve ser verificada e devido ao fato de serem valores obtidos por medições externas ao processo (informações adquiridas em uma frequência muito alta), faz-se necessário um estudo em instrumentação para possibilitar uma medição que permita obter dados em intervalos de tempo mais curtos e assertivos.

Portanto, o resultado obtido indica coerência com as implicações do processo de prensagem e o conhecimento da classificação de criticidade das variáveis pode permitir a tomada de decisão mais assertiva e otimizada para melhoria deste processo. Além disso, variáveis que anteriormente não eram apontadas como críticas podem ser objetos de estudo e análise para o melhor desempenho da prensa. Isso indica que a aplicação de modelos de ranqueamento em aprendizado de máquina podem cooperar para otimizar o ganho da prensa de rolos.

5.1.3 Modelo de classificação da superfície específica da prensa

Os modelos de aprendizado de máquina mostrados no Capítulo 4 foram testados e avaliados conforme as medidas de desempenho: acurácia (CA – *classification accuracy*) e AUC.

Fornecendo ao componente *Test & Score* do *Orange* o conjunto de dados e os oito algoritmos de aprendizado citados (4.1.5), o modelo classificador *Random Forest* utilizando como parâmetro inicial de 5 quantidades de árvores apresentou os melhores resultados quanto à precisão medidos pelas métricas CA=0,875 e AUC=0,913 comparando todos os oito modelos de aprendizado de máquina avaliados (Tabela 5). No segundo e terceiro lugares foram apontados os modelos KNN e *AdaBoost*, respectivamente.

Tabela 5: Resultados obtidos após os testes e validação dos modelos propostos para classificação da variável meta.

Modelos	AUC	CA	F1	Precision	Recall
Random Forest	0,913	0,875	0,871	0,872	0,875
kNN	0,896	0,845	0,839	0,841	0,845
AdaBoost	0,826	0,859	0,859	0,859	0,859
Árvore de Decisão	0,807	0,860	0,859	0,858	0,860
Rede Neural	0,804	0,791	0,769	0,784	0,791
Naive Bayes	0,652	0,727	0,712	0,707	0,727
Regressão Logística	0,647	0,720	0,630	0,680	0,720
SVM	0,443	0,626	0,592	0,571	0,626

Fonte: O autor.

A partir do primeiro resultado, foram realizados vários testes para acréscimo do número de árvores na floresta do *Random Forest*, objetivando a possibilidade de melhoria nos seus resultados (Tabela 6).

Tabela 6: Resultado testes variando a quantidade de árvores do modelo de classificação *Random Forest*.

Quantidade de Árvores (*Random Forest*)	Métodos de Avaliação	
	Valores de AUC	Valores de CA
5 Árvores	0,912	0,875
10 Árvores	0,936	0,888
15 Árvores	0,944	0,894
20 Árvores	0,949	0,895
25 Árvores	0,951	0,897
30 Árvores	0,954	0,899
35 Árvores	0,953	0,899
40 Árvores	0,954	0,899
45 Árvores	0,956	0,901
50 Árvores	0,957	0,901
55 Árvores	0,957	0,901
60 Árvores	0,957	0,901

Fonte: O autor.

À medida que se aumenta a quantidade de árvores, percebe-se a melhoria de desempenho do modelo através das métricas CA e AUC. Após atingir a quantidade de cinquenta árvores pode-se observar que há uma estabilidade nestes valores, sendo CA=0,901 e AUC=0,957 (Tabela 7). Portanto, frente ao melhor desempenho encontrado nestes testes, o

modelo final encontrado para este trabalho é um *Random Forest* com cinquenta árvores de decisão.

Tabela 7: Resultado da avaliação dos métodos de aprendizagem de máquina, utilizando a quantidade de 50 árvores para o modelo *Random Forest*.

Modelos	AUC	CA	F1	Precision	Recall
Random Forest	0,957	0,901	0,897	0,901	0,901

Fonte: O autor.

Estes valores demonstram um ótimo poder de discriminação do modelo resultante deste trabalho, indicando a possibilidade de grande eficácia na predição do resultado esperado dentro da meta de ganho (valor da variável de superfície específica >= 2.100 g/cm^3).

Vários testes também foram realizados para os modelos KNN e *AdaBoost* em busca de melhores ajustes destes modelos, entretanto não foram obtidos maiores resultados em comparação aos obtidos pelo modelo *Random Forest*.

A matriz de confusão (Tabela 8) mostra o resultado obtido para o método mais bem avaliado, *Random Forest*, com percentual de assertividade de predição na classificação da variável de superfície específica como “AbaixoMeta” de 89,8% e como “Meta” de 90,9%.

Tabela 8: Componente *Confusion Matrix*. Matriz de confusão do método mais bem avaliado, *Random Forest*, com percentual de assertividade de predição na classificação da variável de superfície específica como “AbaixoMeta” de 89,8% e como “Meta” de 90,9%.

Real	Predição	
	Abaixo Meta	**Meta**
Abaixo Meta	89,80%	9,10%
Meta	10,20%	90,90%

Fonte: O autor.

O gráfico da curva ROC mostrado na Figura 19 apresenta o resultado obtido pelos oito modelos de predição avaliados. O método *Random Forest* apresenta a curva mais próxima da área “Céu ROC”, detalhado na Seção 3.4.1.1, enfatizando seu ótimo desempenho conforme mencionado na discussão acima.

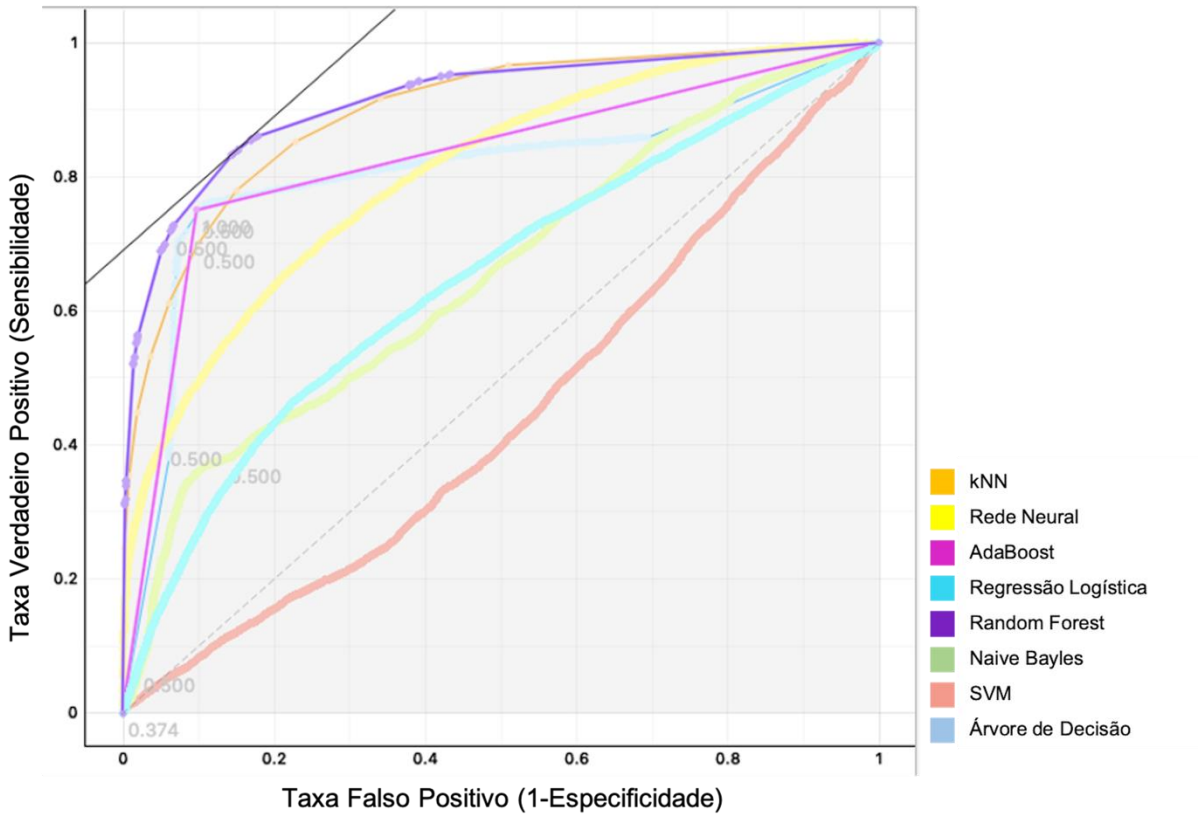

Figura 19: Gráfico da curva ROC resultado da avaliação dos oito modelos de predição.
Fonte: O autor.

A validação deste modelo é de grande valia pois, identificando o desempenho da prensa pelo atendimento à meta especificada, e quanto maior for o valor atingido para o ganho de superfície específica após o processo de prensagem, pode-se diminuir o consumo energético (KWh) na etapa da moagem, visto ser a etapa de maior consumo energético para cominuição do *pellet feed* no processo de pelotização.

5.1.4 Validação do modelo de classificação da superfície específica da prensa

Após a obtenção do modelo *Random Forest*, utilizou-se uma nova base de dados para validar a aplicação deste modelo. Os novos dados utilizados são oriundos do processo entre os dias 30/08/2018 e 06/09/2018, totalizando 1015 registros.

A Figura 20 mostra o diagrama desenvolvido para utilização desses dados para predição da classificação da superfície específica (se predição resultante do modelo $\in$ classe "Meta" ou $\in$ classe "AbaixoMeta"). O resultado de predição foi exportado para análise dos resultados.

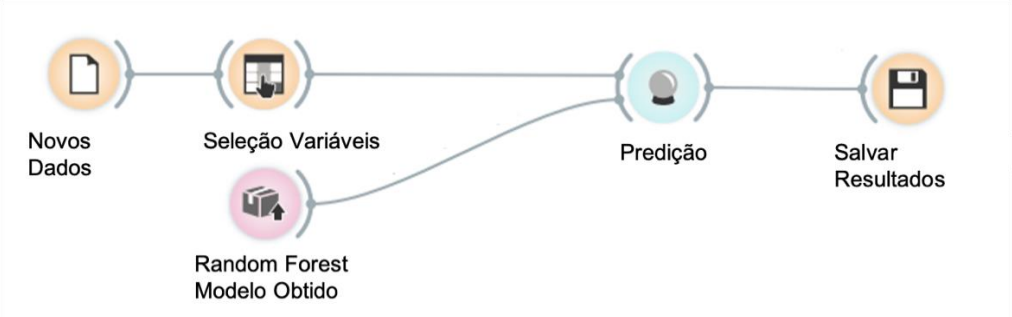

Figura 20: Diagrama de configuração de utilização dos novos dados junto ao modelo *Random Forest* obtido na etapa de aprendizado e validação.
Fonte: O autor.

O comparativo do resultado de predição é mostrado no gráfico da Figura 21.

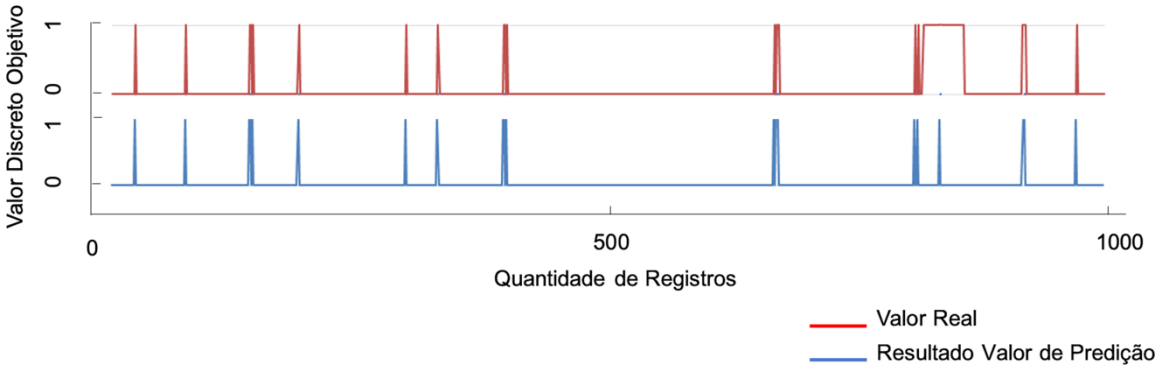

Figura 21: Gráfico demonstrativo de comparação dos dados reais com o resultado obtido pelo modelo de predição *Random Forest.*
Fonte: O autor.

Percebe-se que a sobreposição entre os valores reais e preditos pelo modelo (linha vermelha: dados reais do processo; linha azul: dados preditos pelo modelo) resulta em 93,69% de assertividade (951 dados correlatos / 1015 dados avaliados).

Portanto, é possível inferir que este modelo atinge o objetivo de utilização para uma predição do desempenho da prensa no processo com alta assertividade, possibilitando sua aplicação para medição e ajustes de forma mais ágil no processo de prensagem e, consequentemente, o aprimoramento para maior eficácia operacional.

5.1.5 Identificação dos valores de *setups* ótimos para operação da prensa

Analisando os resultados obtidos do modelo *Random Forest*, pode-se dizer que extrair os *setups* ótimos de operação do processo de prensagem é algo mais amplo. O que existe agora é a possibilidade de se obter regras que associam várias condições e, em cada uma destas, *setups* para cada variável pertencente a mesma. Portanto, os resultados mostram um

conjunto de regras como saída do modelo referenciado, permitindo a tomada de decisão de forma mais complexa e ampla do que uma única tabela de valores para melhor operação do sistema. Para corroborar com essa discussão, os resultados são mostrados a seguir.

A partir do componente do *Orange Pythagorean Forest* e *Tree Viewer* obteve-se os resultados que estabelecem várias regras para atingir a meta do valor de superfície específica (valores >= 2.100 g/cm^3).

O componente *Pythagorean Forest* é uma floresta pitagórica que mostra todos os modelos de árvore de decisão aprendidos no modelo resultante *Random Forest*. A partir da execução do algoritmo de aprendizagem deste modelo, as quantidades de árvores utilizadas para tal são mostradas como árvores pitagóricas.

A Figura 22 mostra 8 das 50 árvores geradas pelo modelo. Em destaque, uma das árvores selecionadas, a priori aquela com os galhos mais curtos e com cores mais fortes (significando que poucos atributos dividem bem os ramos da árvore) para extração das informações ou regras para atingir a meta de predição de classificação da superfície específica.

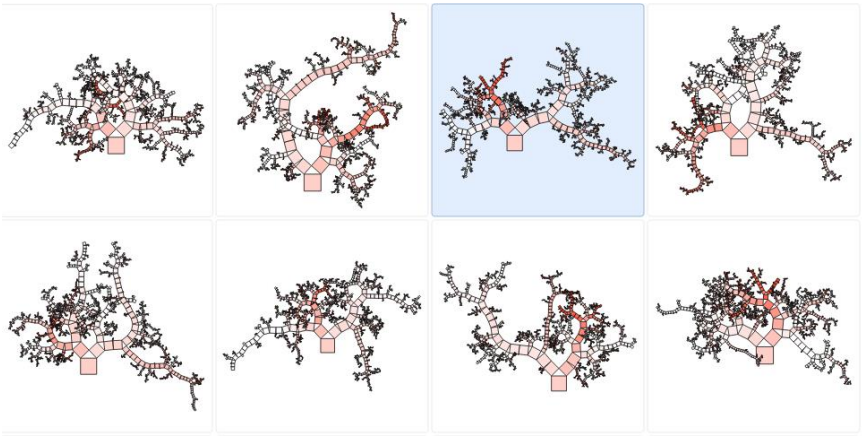

Figura 22: Algumas árvores pitagóricas resultantes do modelo *Random Forest*. Em destaque a árvore selecionada para extração das informações no componente *Tree Viewer*, sendo as cores mais escuras os ramos que orientam para atingir a meta do modelo de predição.
Fonte: O autor.

Após isso, o componente *Tree Viewer* permite visualizar esta árvore de decisão utilizando um gráfico de representação de conhecimento em formato de árvore para demonstrar visualmente as condições e as probabilidades para se chegar aos resultados.

O resultado obtido é uma árvore com as variáveis que são utilizadas para tomada de decisão, contendo os resultados variados e conclusivos para atingir o objetivo do modelo.

As 50 árvores geradas produzem inúmeras regras de decisão, variando-se a quantidade de regras para cada uma das árvores. O modelo *Random Forest* possui esta característica, pois ao invés de se criar uma única árvore de decisão com todas as características ao mesmo tempo, são criadas várias miniárvores menores de decisão selecionando subconjuntos aleatórios das características de forma a formar uma floresta de árvores que compõe a solução global.

A Figura 23 mostra o exemplo de uma das regras em formato de miniárvore de decisão, com a desnormalização dos dados, englobando 7 características mapeadas no modelo:

- Corrente Motor > 179,2 A
- 0,13 mm < Desalinhamento da Prensa < 0,7 mm
- Gap do Lado Direito > 5,4 mm
- 604,0 t/h < Vazão Alimentação da Prensa < 633,8 t/h
- Pressão Óleo do Rolo > 94,0 kgf/cm^2
- Nível Silo da Etapa da Mistura > 31,4 %
- 82,0 % < Torque do Motor > 82,4 %

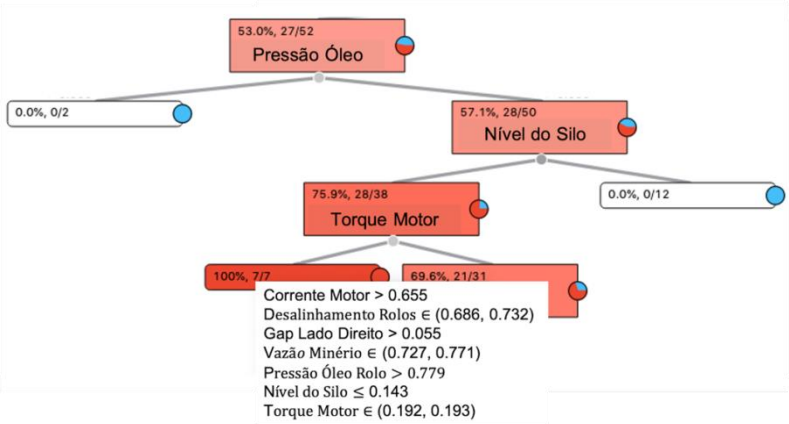

Figura 23: Amostra de uma miniárvore obtida pelo modelo *Random Forest*. A janela em destaque evidencia uma das regras para atingir a meta do modelo, mostrando os valores normalizados de cada variável a ser controlado para obter o resultado esperado para o processo.
Fonte: O autor.

Esse resultado mostra os valores a serem avaliados durante a operação da prensa, permitindo conhecer alguns dos melhores *setups* para cada condição verificada. A priori, a regra extraída do modelo pode permitir um controle da variável de pressão de óleo quando ocorrer um desalinhamento dos rolos maior do que 0,13 mm e menor do que 0,70 mm ou

quando o *gap* do lado esquerdo da prensa ultrapassar o valor de 5,44 mm. Assim para ambas as condições acima, o incremento de pressão de óleo deve ser realizado até atingir um valor maior do que 94,0 kgf/cm^2 condicionado ao incremento do torque (maior do que 82,0 % e menor do que 82,4 %) e corrente do motor (maior do que 179,2 A), isso quando a vazão de alimentação de minério para o processo de prensagem for maior do que 604,0 t/h e menor do que 633,8 t/h.

Portanto, a partir da discussão desta regra percebe-se o poder de análise que é fornecido pelo modelamento de aprendizado de máquina. A utilização destes modelos demonstra a capacidade de tomada de decisão frente às variadas condições de processo e correlação entre as variáveis mais significantes, permitindo ganhos com ajustes que direcionam a otimização do resultado esperado da prensa.

5.2 Predição da variável de superfície específica da prensa

5.2.1 Modelo de regressão da superfície específica da prensa

A predição da variável de superfície específica é de grande importância para a determinação do desempenho da prensa de rolos. Essa variável determina, de maneira geral, qual o ganho do processo de prensagem na qualidade do minério no processo de formação das pelotas.

Os 5 modelos de regressão foram testados e avaliados pelo componente *Test & Score* do *Orange* (Figura 16) conforme as medidas de desempenho: MSE, RMSE, MAE e R^2. O modelo *AdaBoost* apresentou os melhores resultados, com MSE = 96,392, RMSE = 9,818, MAE = 3,580 e R^2 = 0,989 (Tabela 9).

Tabela 9: Resultados obtidos após os testes e validação dos modelos de regressão propostos para predição da variável de superfície específica da prensa.

Modelos	Métodos de Avaliação			
	MSE	RMSE	MAE	R^2
AdaBoost	96,392	9,818	3,580	0,989
kNN	313,107	17,695	5,982	0,964
Random Forest	550,144	23,455	14,740	0,937
Rede Neural	3168,636	56,291	42,926	0,637
SVM	5983,792	77,355	60,489	0,315

Fonte: O autor.

Observa-se que todos os métodos de avaliação foram coerentes quanto aos resultados obtidos, ou seja, para valores de R^2 próximos de 1 tem-se valores menores de MSE, RMSE e MAE.

Atualmente, como já descrito, o valor dessa variável é obtido em laboratório em uma frequência de 4 horas, sendo uma medição lenta quando comparada a necessidade de tomada de decisão do processo. Não existem atualmente instrumentos instalados capazes de realizar essa medição.

Portanto esse resultado possibilita a obtenção do valor da superfície específica da prensa de maneira ágil, podendo ser calculada em uma frequência na unidade de tempo de segundos/minutos, permitindo auxiliar a tomada de decisão de processo ou fornecendo o valor medido para ser utilizado em ferramentas de sistemas de otimização. Atualmente, em virtude do tempo de amostragem desta variável ser alta, torna-se inviável sua utilização nos softwares de otimização para efeitos de comparação e ajustes em tempo real.

Esse resultado demonstra, portanto, a possibilidade de criação de um sensor virtual (*soft sensor*) para esta medição, fortalecendo a análise operacional quanto ao ganho do processo de prensagem do minério e consequentemente do desempenho da prensa de rolos, antecipação de falhas, anomalias, redução de custos (consumo de energia elétrica, custos de manutenção do equipamento por exemplo) e otimização.

5.2.2 Validação do modelo de regressão da superfície específica da prensa

Após a obtenção do modelo *AdaBoost*, foram utilizadas 3 (três) novas bases de dados para validar a aplicação deste modelo. Os novos dados utilizados são oriundos do processo entre os dias 05/12/2018 à 26/12/2018, 07/03/2018 à 13/04/2018, 06/02/2018 à 11/03/2018, totalizando 3001, 6033, 5914 registros respectivamente.

A Figura 24 demonstra o diagrama desenvolvido para utilização dos dados dos 3 períodos para validação do modelo de regressão da superfície específica. O resultado de predição foi exportado para análise dos resultados.

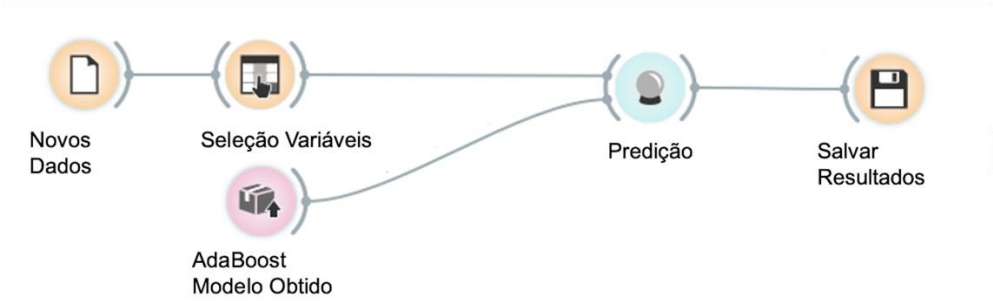

Figura 24: Diagrama de configuração de utilização dos novos dados junto ao modelo *AdaBoost* obtido na etapa de aprendizado e validação do modelo de regressão.
Fonte: O autor.

O comparativo dos resultados de predição pelo modelo de regressão obtido para os períodos citados acima é mostrado na Tabela 10.

Tabela 10: Resultados obtidos pela utilização do modelo de regressão para predição da superfície específica da prensa para 3 diferentes períodos de dados.

Período de Dados	Resultados		
	Erro Médio	RMSE	MAE
05/12/2018 à 26/12/2018	3,830%	89,740	73,860
07/03/2018 à 13/04/2018	3,398%	95,287	72,430
06/02/2018 à 11/03/2018	3,056%	69,790	59,252

Fonte: O autor.

As principais métricas utilizadas na avaliação deste resultado foram os valores de erro médio, RMSE e MAE. Isso porque o objetivo é validar a variabilidade da resposta do modelo frente à incerteza do valor predito da superfície específica, ou seja, qual é o impacto da variação desta resposta quando comparado à faixa de medição desta variável no processo.

Percebe-se que os melhores resultados foram para os dados do período de 06/02/2018 a 11/03/2018, sendo estes menores do que nos demais períodos. Pode-se inferir que 3,056% de erro médio ou 59,252 de erro médio absoluto seriam valores aceitáveis e toleráveis para uma predição de uma variável de processo, visto que os próprios instrumentos utilizados para as medições das variáveis de entrada do modelo (temperatura, pressão, corrente, nível, vazão etc.) detém de sua incerteza de medição assim como as variações do processo potencializam esse aspecto.

Portanto pode-se entender que existe a possibilidade de utilização do modelo de regressão para predição da variável de superfície específica, entretanto é necessário ponderar a necessidade de revalidação dos dados utilizados para que se possa realizar treinamentos

frequentes do modelo em virtude das variações dos cenários de produção, ou seja, este modelo pode não ser aplicável para determinados cenários e variações de processo (alteração das propriedades físico-química do minério, ajustes de equipamentos e patamares de produção, dentre outros).

6 CONCLUSÕES

Os resultados deste trabalho permitem agilizar a análise preditiva do desempenho da prensa de rolos, automatizando a correlação das informações dos vários sistemas disponíveis e possibilitando o diagnóstico do desempenho da prensa em tempo real, visto que atualmente é necessário realizar uma análise em laboratório para se obter as informações de seu desempenho em um intervalo de 4 horas.

A aplicabilidade na indústria bem como sua escalabilidade são altamente possíveis, visto que a possibilidade de implantação pode ser aplicada e customizada para outras prensas de rolos existentes, para os demais equipamentos do processo de pelotização (como moinho de bolas, filtros, discos de pelotização e outros) e até processos diferentes, desde que avaliados para cada necessidade e peculiaridade.

A determinação das variáveis mais significativas, a obtenção dos modelos de classificação *Random Forest* e de regressão *AdaBoost* no processo de prensagem demonstram a eficácia na aplicação de métodos de inteligência artificial na indústria.

Existe a possibilidade de utilização, após a validação dos resultados do modelo de regressão, da variável de superfície específica da prensa nos modelos do software de otimização. O *feedback* desta grandeza calculada em tempo real e utilizada com entrada no modelo de otimização pode aprimorar os estudos, conhecimento e parametrização existentes. Ainda, pode contribuir em novas conclusões e tomada de decisões sobre o processo, visto a possibilidade de medição do desempenho do processo de forma mais ágil em relação a frequência atual desta medição.

Ainda, a utilização do conhecimento das variáveis de maior influência no processo agregando às regras de *setups* ótimos e aos modelos obtidos permitem também a criação de um ambiente de simulação deste processo, sendo possível avançar em novas abordagens de sistemas cibernéticos para auxílio da tomada de decisão, alinhado ao avanço tecnológico da indústria 4.0.

A metodologia pode ser aplicada para o desenvolvimento de novos modelos, considerando novos dados de processo, a variação das grandezas físico-químicas do minério, as mudanças do *modus operandi* do processo e a condição dos equipamentos ao longo do tempo.

A previsão do desempenho do processo pode abrir uma ampla discussão e possibilidade de estudo para a previsão da vida útil deste equipamento, adotando as diversas técnicas de aprendizado de máquina como, por exemplo, a previsão da vida útil restante do rolamento com base em rede neural ou prever seu estado ao longo do estágio de vida do ativo.

Portanto, percebe-se que a escalabilidade da aplicação deste trabalho é alta, o que pode possibilitar a replicação e melhoria do gerenciamento não só da prensa de rolos no processo de pelotização na mineração, mas como visto e discutido em trabalhos recentes, em vários equipamentos e processos de produção na indústria.

Os resultados deste trabalho renderam uma importante publicação na *23rd International Conference on Enterprise Information Systems* (ABREU *et al.*, 2021).

REFERÊNCIAS BIBLIOGRÁFICAS

ABREU, T. N. *et al.* Application of Machine Learning Methods to Improve of the Roller Press Performance in the Pelletizing Process. *In*: **Proceedings of the 23rd International Conference on Enterprise Information Systems**: SciTePress, 2021. v. 1, p. 677-684.

BARRIOS, G., Tavares, M., Pérez-Prim, J. High pressure grinding rolls simulation using the discrete element method dynamic coupling interface. *In*: **XXVII International Mineral Processing Congress,** 2014, Santiago, Chile.

BARRIOS, G. K. P. **Modelagem da prensa de rolos usando o método dos elementos discretos com acoplamento dinâmico e o modelo de substituição de partículas**. 2015. Universidade Federal do Rio de Janeiro.

BASTOS, P. C. **Análise comparativa entre o uso de métodos convencionais e o uso de softwares para a seleção de britadores e peneiras.** 2015. -, Universidade Federal de Minas Gerais.

BONESSO, D. **Estimação dos Parâmetros do Kernel em um Classificador SVM na Classificação de Imagens Hiperespectrais em uma Abordagem Multiclasse**. 2013. Universidade Federal do Rio Grande do Sul, Porto Alegre.

BOSER, B. *et al.* A training algorithm for optimal margin classifiers. *In*: **Proceedings of the Fifth Annual Workshop on Computational Learning Theory**, 1992, Pittsburgh.

BREIMAN, L. Bagging Predictors. **Machine Learning**, 24, p. 123–140, 1996.

CAMPOS, G. O. **Estudo, avaliação e comparação de técnicas de detecção não supervisionada de outliers**. 2015. Instituto de Ciências Matemáticas e de Computação, USP, São Carlos - SP.

CAMPOS, T. l. M. **Modelagem matemática da prensa de rolos aplicada à cominuição de minério de ferro**. 2018. Universidade Federal do Rio de Janeiro, Rio de Janeiro.

CAMPOS, T. M. *et al.* Desafios na modelagem da capacidade e potência consumida da prensa de rolos. *In*: **XXVII Encontro Nacional de Tratamento de Minérios e Metalurgia Extrativa**, 2017, Belém-PA. p. 1743-1753.

D'ANGELO, T. *et al.* Deep Learning-Based Object Detection for Digital Inspection in the Mining Industry. *In*: **18th IEEE International Conference On Machine Learning And Applications (ICMLA)**, 2019, p. 633-640. DOI: 10.1109/ICMLA.2019.00116.

DAĞ, H. *et al.* Comparison of Feature Selection Algorithms for Medical Data. *In*: **IEEE**, 2012. DOI: 978-1-4673-1448-0.

DANIEL, M. J. **HPGR model verification and scale-up.** 2002. Julius Kruttschnitt Mineral Research Centre, University of Queensland.

DUDA, R. O. *et al.* **Pattern Classification**. *In*: Wiley (Ed.). 2001.

ERICEIRA, D. R. *et al.* Early Failure Detection of Belt Conveyor Idlers by Means of Ultrasonic Sensing. *In*: **International Joint Conference on Neural Networks (IJCNN)**, 2020, p. 1-8. DOI: 10.1109/IJCNN48605.2020.9207646.

FAYYAD, U. M.; IRANI, K. B., 1993, **Multi-interval discretisation of continuous-valued attributes.** 1022–1027.

FAYYAD, U. M. *et al.* Knowledge discovery and data mining: towards a unifying framework. *In*: **KDD**, 1996. v. 96, p. 82-88.

FLACH, P. A.; WU, S. Repairing concavities in ROC curves. *In*: IJCAI' 05: **Proceeding of the Nineteenth International Joint Conference on Artificial Intelligence**, 2005, p. 702-707.

HALLAK, R.; FILHO, A. J. P. Metodology for performance analysis of simulations of convective systems in the metropolitan area of são paulo with the arps model: sensitivity to variations with the advection and the data assimilation schemes. **Revista Brasileira de Meteorologia**. 26**:** 591-608 p. 2011.

HAMEL, L. Knowledge Discovery with Support Vector Machines. **John Wiley & Sons**, New Jersey, 2009.

HAN, J.; KAMBER, M. **Data Mining: Concepts and Techniques**. 2006a.

HAN, J.; KAMBER, M. Data Mining: Concepts and Techniques. **Morgan Kaufmann**, 2006b.

HASANZADEH, V.; FARZANEGAN, A. Robust HPGR model calibration using genetic algorithms. **Minerals Engineering**, 24, p. 424-432, 2011.

KELLER, F. *et al.* **Hics**: high contrast subspaces for density-based outlier ranking. Washington, DC 2012.

KIRA, K.; RENDELL, L., 1992, **A Practical Approach to Feature Selection**. 249-256.

KLIPPEL, E. *et al.* Embedded Edge Artificial Intelligence for Longitudinal Rip Detection in Conveyor Belt Applied at the Industrial Mining Environment. *In*: **23rd International Conference on Enterprise Information Systems (ICEIS 2021)**, 2021, **1**. p. 496-505. DOI: 10.5220/0010447204960505.

KONONENKO, I., 1994, **Estimating Attributes: Analysis and Extension of Relief**. 171-182.

LANTZ, B. **Machine Learning with R**. Packt Publishing Ltd. 2013.

MATOS, F. F. *et al.* Análise de Dados de Saúde: Mineração de Texto com a Utilização do Orange Canvas para Exploração da Informação. *In*: **XX Encontro Nacional de Pesquisa em Ciência da Informação – ENANCIB**, 2019, Florianópolis – SC.

MATSUBARA, E. T. **Relações entre Ranking, Análise ROC e Calibração em Aprendizado de Máquina**. 2008. Instituto de Ciências Ma- temáticas e de Computação, USP - São Carlos, São Carlos.

MCCUE, C. **Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis**. Butterworth-Heinemann, 2014. 379 p.

MCCULLOCH, W. S.; PITTS, W. H. A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics**, 5, p. 115-133, 1943.

MILES, J. R. **Squared, adjusted r squared**. *In*: Online, W. S. S. R. (Ed.). 2014. p. 1-3.

MONTEIRO, A. M. **Modelagem neural de um processo de produção de pelotas de minério de ferro**. Universidade Federal de Minas Gerais, 2003.

MOURÃO, J. M. **Aspectos Concentuais Relativos à Pelotização de Minérios de Ferro.** 2017. Vitória-ES.

OLIVEIRA, R. N. M. d. **Análise de desempenho do HRC HPGR em circuito piloto**. 2016. Universidade de São Paulo, São Paulo.

OLIVEIRA, V. M. **Estudo da porosidade de pelotas de minério de ferro para altos-fornos através de adsorção física.** 2010. Universidade Federal de Minas Gerais.

OLSON, D. L.; DELEN, D. **Advanced Data Mining Techniques**. Springer VerlagBerlinHeidelberg, 2008. 177 p.

OSHIRO, T. M. **Uma abordagem para construção de uma única a partir de uma Random Forest para classificação de bases de expressão gênica**. 2013. Universidade de São Paulo, Ribeirão Preto.

PAL, J. *et al.* Effect of Blaine Fineness on the quality of hematite iron ore pellets for blat furnace. **Mineral Processing and Extractive Metallurgy Review: An International Journal**, 36, p. 83-91, 2015.

PRATI, R. C. *et al.* **Curvas ROC para avaliação de classificadores**. 2008.

PRESS, W. H. *et al.* **Numerical Recipes in C: The Art of Scientific Computing.** 1992. Cambridge University Press, New York.

ROBNIK-SIKONJA, M.; KONONENKO, I. Theoretical and empirical analysis of ReliefF and RReliefF. **Mach. Learn.**, 53(1-2):23–69, 2003.

RUSSEL, S. J.; NORVIG, P. **Artificial Intelligence: A Modern Approach** Prentice Hall, 2009. 1152 p.

SARAMAK, D.; KLEIV, R. A. The effect of feed moisture on the comminution efficiency of HPGR circuits. . **Minerals Engineering**, 43-44, p. 105-106, 2013.

SHANNON, C. E.; WEAVER, W. **The mathematical theory of communication**. 1949.

SHARMA, A.; DEY, S. **A comparative study of feature selection and machine learning techniques for sentiment analysis**. 2012.

SHEDROFF, N. Information Interaction Design: A Unified Field Theory in Design. *In*: JACOBSON, R. (Ed.). **Information Design**: The MIT Press, 1999. p. 267-292.

SILVA, L. **Influência da umidade no processo de pelotização**. Serra, ES: Faculdade do Centro Leste, 2008.

SOLÉ, R. A. L.; WENDLING, F. **Curso de Pelotização**. : Fundação Gorceix – Departamento de Pesquisa e Educação Continuada, 2014.

TORRES, M.; CASSALI, A. A novel approach for the modelling of high-pressure grinding rolls. **Minerals Engineering**, 22, p. 1137–1146, 2009.

TSAI, C.-F. *et al.* A comparative study of classifier ensembles for bankruptcy prediction. **Applied Soft Computing**, 24, p. 977-984, 2014.

VAN DER MEER, F. P. Roller Press Grinding of Pellet Feed. Experiences of KHD in the Iron Ore Industry. *In*: **AusIMM Conference on Iron Ore Resources and Reserves Estimation**, 1997, p. 1-15.

VITERBO, J. *et al.* Avaliação de Ferramentas de Apoio ao Ensino de Técnicas de Mineração de Dados em Cursos de Graduação. *In*: **XXXVI Congresso da Sociedade Brasileira de Computação**, 2016, p. 11-20. DOI: 10.5753.

VYHMEISTER, E. *et al.* Modeling and energy-based model predictive control of high pressure grinding roll. **Minerals Engineering**, 134, 2019.

WEISS, S. M.; KULIKOWSKI, C. A. **Computer systems that learn:classification and prediction methods from statistics, neural nets, machine learning, and expert systems**. *In*: Morgan Kaufmann Publishers, 1991.

YANG, Y.; PEDERSEN, J. O. A comparative study on feature selection in text categorization. **ICML**, 97, p. 412-420, 1997.

YU, L.; LIU, H. Efficient feature selection via analysis of relevance and redun- dancy. **Journal of Machine Learning Research**, 5, p. 1205–1224, 2004.

ANEXO

ANEXO A

1) Código para normalização dos dados

```
Sub Macro1()
'
' Macro1 Macro
Dim i, j
Dim num, nmCell1, nmCell2
Planilha1.Activate
'Quantidade de Registros
num = WorksheetFunction.Count(Range("A:A"))
nmCell1 = Planilha1.Cells(2, 1).Address
nmCell2 = Planilha1.Cells(num - 2, 1).Address
'1)  Calcula valores Max e Min de cada variavel
'Quantidade de Variaveis a serem calculadas
For j = 0 To 50
'Valores Max
Planilha1.Cells(2, j + 59).Select
   ActiveCell.Value = WorksheetFunction.Max(Range(nmCell1 & ":" &
nmCell2).Offset(Columnoffset:=(j)))
'Valores Min
Planilha1.Cells(3, j + 59).Select
   ActiveCell.Value = WorksheetFunction.Min(Range(nmCell1 & ":" &
nmCell2).Offset(Columnoffset:=(j)))
Next j
Dim a
```

```
a = 1
If a Then
'2) Gera Planilha de Dados Normalizados
Planilha2.Activate
For j = 4 To 54
   For i = 3 To num + 2
      '3) Classifica qualidade Superficie Especifica (Se > 2100->Meta, Senão->AbaixoMeta)
      If (Planilha1.Cells(i, 54).Value >= 2100) Then
         Planilha1.Cells(i, 55).Select
            ActiveCell.Value = "Meta"
         Planilha2.Cells(i, 55).Select
            ActiveCell.Value = "Meta”
      Else
         Planilha1.Cells(i, 55).Select
            ActiveCell.Value = "AbaixoMeta"
         Planilha2.Cells(i, 55).Select
            ActiveCell.Value = "AbaixoMeta"
      End If
      'Normalizacao y=(x-min)/(max-min)
      If (Planilha1.Cells(2, j + 55).Value - Planilha1.Cells(3, j + 55)) <> 0 And
Planilha1.Cells(i, j).Value <> "" Then
         Planilha2.Cells(i, j).Select
            ActiveCell.Value = (Planilha1.Cells(i, j).Value - Planilha1.Cells(3, j + 55)) /
(Planilha1.Cells(2, j + 55).Value - Planilha1.Cells(3, j + 55))
      End If
```

```
    Next i
Next j
End Sub
```