



REDEMAT

REDE TEMÁTICA EM ENGENHARIA DE MATERIAIS

UFOP – UEMG



DISSERTAÇÃO DE MESTRADO

**“ANÁLISE MULTIVARIADA E REDES NEURAIS
APLICADAS NA PREDIÇÃO DE PROJEÇÃO NO
BOF”**



**Autora: Bruna Helena Malovini Loiola
Orientador: Prof. PhD. Carlos Antônio da Silva
Coorientador: Eng. Dr. Henrique Silva Furtado
Coorientador: Prof. Dr. Américo Tristão Bernardes**

Março de 2020



REDEMAT

REDE TEMÁTICA EM ENGENHARIA DE MATERIAIS

UFOP – UEMG

Bruna Helena Malovini Loiola

“Análise multivariada e redes neurais aplicadas na predição de projeção no BOF”

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia de Materiais da REDEMAT, como parte integrante dos requisitos para a obtenção do título de Mestra em Engenharia de Materiais.

Área de concentração: Processos de Fabricação
Orientador: Prof. PhD. Carlos Antônio da Silva
Coorientador: Eng. Dr. Henrique Silva Furtado
Coorientador: Prof. Dr. Américo Tristão Bernardes

Ouro Preto, março de 2020

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

L834a Loiola, Bruna Helena Malovini .
Análise multivariada e redes neurais aplicadas na predição de
projeção no BOF. [manuscrito] / Bruna Helena Malovini Loiola. - 2020.
66 f.: il.: color., gráf., tab..

Orientador: Prof. Dr. Carlos Antônio da Silva.

Coorientadores: Prof. Dr. Américo Tristão Bernardes, Dr. Henrique
Silva Furtado.

Dissertação (Mestrado Acadêmico). Universidade Federal de Ouro
Preto. Rede Temática em Engenharia de Materiais. Programa de
Engenharia de Materiais.

Área de Concentração: Processos de Fabricação.

1. Big Data. 2. Redes neurais (Computação) . 3. Python (Linguagem de
programação de computador). 4. Análise Multivariada. I. Bernardes,
Américo Tristão. II. Furtado, Henrique Silva. III. Silva, Carlos Antônio da.
IV. Universidade Federal de Ouro Preto. V. Título.

CDU 62-5

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



REDE TEMÁTICA EM ENGENHARIA DE MATERIAIS UFOP - UEMG

Pós-Graduação em Engenharia de Materiais



"Análise multivariada e redes neurais aplicadas na predição de projeção no BOF"

Autor(a): Bruna Helena Malovini Loiola

Dissertação defendida e aprovada, em 20 de março de 2020, pela banca
examinadora constituída pelos professores:

Prof. Carlos Antônio da Silva (PhD) - Orientador
Universidade Federal de Ouro Preto

Prof. Leandro Rocha Lemos (Doutor)
Universidade Federal de Minas Gerais

Prof. Johnê Jesus Mol Peixoto (Doutor)
Universidade Federal de Ouro Preto

Prof. Américo Tristão Bernardes (Doutor) - Coorientador
Universidade Federal de Ouro Preto

AGRADECIMENTOS

Agradeço a Deus e a Nossa Senhora Aparecida por todas as bênçãos, luz e proteção em minha vida.

À minha mãe Márcia, ao meu pai João Carlos e à minha irmã Monique por sempre acreditarem em mim e por serem a minha base familiar perfeita.

À minha avó Madalena, minha bisavó Herondina (*in Memoriam*) e toda minha família que estão sempre na torcida.

Agradeço ao professor PhD. Carlos Antônio da Silva por toda a orientação deste trabalho, por ter estado presente em todas as etapas, tirando dúvidas e ensinando de forma tão tranquila.

Ao professor Dr. Américo Tristão Bernardes, por toda atenção, disponibilidade e pela coorientação. Ao engenheiro metalurgista Dr. Henrique Furtado pelo incentivo para o início desde trabalho, por todas as ideias e coorientação.

Ao professor PhD. Paulo Santos Assis pelo estímulo a publicações internacionais; ao mestrando do Propemm Mateus Amaral pelas experiências trocadas sobre redes neurais em *Python*; aos amigos do laboratório de Pirometalurgia; e à secretária Ana.

Agradeço também aos amigos de estudos do início do mestrado (grupo Gelatina), a todos os amigos de longa data de Vila Velha/Vitória, e aos mais novos que Ouro Preto me deu, em especial, as meninas da República Tan Tan que foram como verdadeira família nesses últimos tempos.

Finalizo agradecendo aos demais docentes que de alguma forma contribuíram para realização deste trabalho; à CAPES pela bolsa e auxílios concedidos, à REDEMAT e à UFOP por me proporcionarem um mestrado de alta qualidade.

“Não há lugar para a sabedoria onde não há paciência. Sê humilde para evitar o orgulho, mas voa alto para alcançar a sabedoria”.

Santo Agostinho

SUMÁRIO

LISTA DE FIGURAS.....	VII
LISTA DE TABELAS.....	VIII
LISTA DE SÍMBOLOS	IX
LISTA DE SIGLAS.....	X
RESUMO.....	XI
ABSTRACT	XII
1 INTRODUÇÃO.....	1
2 OBJETIVOS.....	2
2.1 Objetivo Geral.....	2
2.2 Objetivos Específicos	2
3 REVISÃO BIBLIOGRÁFICA	3
3.1 Uma visão geral sobre siderurgia.....	3
3.2 A anomalia projeção	4
3.2.1 Formação da escória.....	7
3.2.2 Formação da emulsão.....	10
3.2.3 Causas da projeção	12
3.2.5 Controle de projeção	15
3.3 Análise Multivariada de dados.....	17
3.3.1 Análise de trilha	18
3.3.2 Análise de Componentes Principais.....	19
3.3.3 Análise de Fatores	20
3.3.4 Análise de <i>Cluster</i>	21
3.3.5 Análises de Regressões	22
3.3.6 Rede de Correlações.....	23
3.4 Redes Neurais Artificiais	24
3.4.1 Aprendizagem de uma rede neural.....	27
3.4.2 Classificador <i>Multilayer Perceptron</i>	29

3.4.3. Validação Cruzada <i>K-fold</i>	29
4 PROCEDIMENTO EXPERIMENTAL.....	30
4.1 Análises do Primeiro Grupo	31
4.2 Análises do Segundo Grupo	34
4.2.1 Determinação das variáveis de influência da projeção	35
4.2.2 Montagem do modelo matemático a partir de uma rede neural	36
5 RESULTADOS E DISCUSSÃO.....	38
5.1 Análises de <i>Cluster</i> , Fatores e CP.....	38
5.1.1 Regressão Linear no <i>NCSS</i>	40
5.2 Análises Finais	42
5.2.1 Análise de Trilha no Software <i>Genes</i>	43
5.2.2 Análise de Rede de Correlações no Software <i>RBio</i>	46
5.4 Redes Neurais Artificiais em Python no Software <i>Spyder</i>	47
6 CONCLUSÕES	51
7 SUGESTÕES PARA TRABALHOS FUTUROS	53
8 REFERÊNCIAS BIBLIOGRÁFICAS.....	54
9 ANEXO I – RESUMO DE ARTIGOS.....	60
10 ANEXO II – VARIÁVEIS DO PRIMEIRO E DO SEGUNDO GRUPO DE ANÁLISES	64

LISTA DE FIGURAS

Figura 3.1 – Esquema do sopro de oxigênio no aço em BOF.....	4
Figura 3.2 – Imagem de projeção ocorrendo em um convertedor.....	5
Figura 3.3 – Mudança na composição de escória durante o sopro em um BOF.....	10
Figura 3.4 – Exemplo de dendograma a partir de uma análise de <i>cluster</i>	22
Figura 3.5 – Exemplo do gráfico de rede de correlação entre 13 variáveis gerado no software Rbio.....	24
Figura 3.6 – Esquematização de um neurônio.....	26
Figura 3.7 – Estruturas de redes neurais em camadas (a) redes com apenas duas camadas de neurônios, (b) rede com uma camada de entradas, uma camada escondida, e uma camada de saída.....	27
Figura 4.1 – Planilha com parte dos dados cedidos para análise.....	30
Figura 4.2 – Esquema com a demonstração do passo a passo das análises do primeiro grupo.....	33
Figura 4.3 – Esquema com a demonstração do passo a passo das análises do segundo grupo.....	37
Figura 5.1 – Dendograma formado a partir da Análise de <i>Cluster</i>	38
Figura 5.2 – Rede de correlação gerada no Rbio das 19 variáveis com a variável projeção.....	47

LISTA DE TABELAS

Tabela 5.1 – Resultado gerado no <i>Statistica</i> com os números de autovalores, variância total, autovalor cumulativo e porcentagem cumulativa de variância para 7 CP.....	39
Tabela 5.2 – Tabela de contribuição das variáveis originais em cada fator.	39
Tabela 5.3 – Variáveis que mais influenciam nas projeções de acordo com análises prévias.....	40
Tabela 5.4 – Coeficientes gerados por RCP, Regressão por Mínimos Quadrados (RMQ), via <i>NCSS</i>	41
Tabela 5.5 – Principais valores encontrados para a variável projeção (Mpi).....	41
Tabela 5.6 – Resultado gerado no <i>Statistica</i> com os números de autovalores, variância total, autovalor cumulativo e porcentagem cumulativa de variância para 19 CP.....	42
Tabela 5.7 – Variáveis em ordem de maior influência nas projeções	44
Tabela 5.8 – Comparação dos erros dos diferentes testes da RN.	48
Tabela 5.9 – Valores dos erros das últimas 5 iterações da RN.....	49
Tabela 5.10 – Resultado do teste de precisão informado pela RN com $k = 8$	49

LISTA DE SÍMBOLOS

A	Área da seção transversal do convertedor (cm^2)
α	Fração de vazios
h_0	Altura da escória antes da formação de espuma (cm)
h	Altura da espuma (cm)
Δh	Varição da altura (cm)
L	Espessura da camada de espuma (cm)
Q_g	Vazão de gás (cm^3/s)
v_s^G	Velocidade superficial do gás (cm/s)
v_G	Velocidade real do gás (cm/s)
Σ	Índice de Formação de Espuma (s)
μ	Viscosidade (Pa.s)
ρ	Densidade (g/cm^3)
σ	Tensão interfacial (N/cm)
ε_{av}	Função de custo (medida de desempenho da rede neural)

LISTA DE SIGLAS

ACP	Análise de Componentes Principais
AMT	ArcelorMittal Tubarão
RCP	Regressão por Componentes Principais
BOF	Basic Oxygen Furnace
CP	Componentes Principais
EAF	Forno de Arco Elétrico
MLP	<i>Multilayer Perceptron</i>
NC	Número de Condição
RNA	Redes Neurais Artificiais
VIF	Fator de Inflação de Variância

RESUMO

A projeção de escória pode acontecer durante o processo de refino do ferro gusa no BOF devido à formação de um volume excessivo de emulsão gás-metal-escória, que, ao crescer além da capacidade do convertedor, expulsa grande porção da escória e metal. Assim, são gerados problemas recorrentes para a operação, como refino menos efetivo, menor rendimento metálico, danos ao equipamento, e principalmente, poluição ambiental devido aos gases que são emitidos no momento da projeção. O objetivo deste estudo foi criar um modelamento matemático para prever a ocorrência de projeção em corridas dos convertedores BOF. Este projeto se baseia em análise multivariada de *Big Data* a qual permitiu identificar as variáveis de operação que mais influenciam no fenômeno projeção, dentre as quais estão o peso de briquete misto e de dunito adicionados durante o sopro, e concentração de silício presente no gusa. Além disso, foi possível treinar redes neurais artificiais em linguagem *Python* o que gerou um modelo de predição com confiabilidade de 96%.

Palavras-chave: Projeção, Análise Multivariada, Big Data, Redes Neurais Artificiais, *Python*.

ABSTRACT

Slopping may occur during the hot metal refining process in the BOF due to the formation of an excessive volume of gas-metal-slag emulsion which, by growing beyond the capacity of the converter, expels a large portion of the slag and metal. This generates recurrent problems for the operation, such as reduction of refining effectiveness, lower metallic yield, damages to the equipment, and mainly, environmental pollution due to gases released during the slopping. The aim of this work was to create a mathematical model to predict the occurrence of slopping in BOF converters. It is based on multivariate Big Data analyzes in order to identify the variables that most influence the slopping phenomenon, such as the weight of mixed briquette and dunite added during blowing, and the silicon concentration in hot metal. In addition, it was possible to train artificial neural networks in Python language that generated a prediction model with 96% of reliability.

Keywords: Slopping, Multivariate Analysis, Big Data, Artificial Neural Networks, Python.

1 INTRODUÇÃO

Projeção de escória é um acontecimento presente em usinas siderúrgicas que utilizam o BOF (*Basic Oxygen Furnace*) como convertedor. Esse fenômeno é um evento de ocorrência comum em aciarias e pode ser observado a partir do momento em que a escória não é mais contida dentro do convertedor, e ocorre um transbordamento de material durante o processo de sopro a oxigênio.

As maiores consequências da projeção estão relacionadas ao meio ambiente, pois durante a ocorrência, há uma maior eliminação de vapores poluentes, a qual o sistema de captação de gases não é capaz de reter. Além disso, há grande perda metálica e perda de rendimento de fundentes e, há também uma redução da eficácia do refino, o que gera uma menor produtividade do aço, entaves de logística, danos ao equipamento, e riscos de acidentes.

O conhecimento do processo de formação e comportamento da emulsão metal-gás-escória, e a identificação das variáveis que estão presentes em todo o processo de refino são essenciais para o entendimento dessa anomalia, devido à sua relação com o controle do sopro.

Assim, existe um grande esforço visando à eliminação dos casos de projeção. Alguns sistemas de controle envolvem o uso de microfones, dispositivos sonoros nos convertedores, análise de imagem e botões de controle de vazão, entre outros.

A proposta deste trabalho foi utilizar métodos de análises multivariadas, denominadas análise de *cluster*, de fatores, de componentes principais, análise de trilha, e redes de correlações para efetuar uma análise de *Big Data* dos convertedores da ArcelorMittal Tubarão, de tal forma a identificar as variáveis de interferência nos episódios de projeções em BOF, além de criar modelos matemáticos para predição de projeção. Para este trabalho foram utilizadas propostas de técnicas estatísticas clássicas, bem como o uso de redes neurais em linguagem *Python*.

2 OBJETIVOS

2.1 Objetivo Geral

Desenvolver um modelo matemático para cálculo da predição de projeções em corridas de aço no BOF, a partir de análises multivariadas de *Big Data* e redes neurais, utilizando linguagem *Python*, de modo que haja a minimização dos impactos ambientais, dos riscos de acidentes, e das perdas metálicas.

2.2 Objetivos Específicos

- Identificar as variáveis que influenciam as causas de projeção de escória no BOF.
- Verificar as variáveis que mais influenciam na ocorrência de projeção a partir de análises multivariadas de dados.
- Treinar as redes neurais com as variáveis que atuam no processo de projeção e elaborar um modelo matemático que permita prever a ocorrência de projeção durante as corridas.

3 REVISÃO BIBLIOGRÁFICA

3.1 Uma visão geral sobre siderurgia

O aço é um material que está diretamente conectado com o desenvolvimento da indústria mundial, com aplicações em diversas áreas, desde a produção siderúrgica até a fabricação de equipamentos com os mais diferentes desígnios.

O Brasil possui 29 usinas siderúrgicas com capacidade total instalada de 50,4 milhões de toneladas, e segundo estatística preliminar do Instituto Aço Brasil (2019), em 2018, o país teve uma produção de aço bruto de 34,7 milhões de toneladas no ano. Esse valor representa uma expansão de 1,1% em comparação a 2017, quando o Brasil fechou o ranking como 10º maior exportador mundial com um saldo comercial de US\$ 5,8 bilhões (INSTITUTO AÇO BRASIL, 2018).

A maior parte da produção de aço no mundo é realizada em convertedores a oxigênio, alguns conhecidos como BOF (*Basic Oxygen Furnace*). No Brasil, esse tipo de convertedor recebe grande atenção, pois é responsável pela produção de 79% do aço bruto (INSTITUTO AÇO BRASIL, 2019).

Para o refino primário que acontece no BOF, esquematizado na Figura 3.1, utiliza-se uma lança de sopro de oxigênio puro em velocidade supersônica capaz de promover as reações químicas que geram a transformação do ferro gusa em aço.

A operação de um BOF é semicontínua. O ciclo completo consiste no carregamento de sucata e ferro gusa fundido, adição de fundentes e refrigerantes, sopro de oxigênio, amostragem e registro de temperatura, e vazamento do aço e escória. Em reatores modernos, até 380 toneladas de aço são produzidas em um intervalo de 30 a 40 minutos (REMUS *et al.*, 2013).

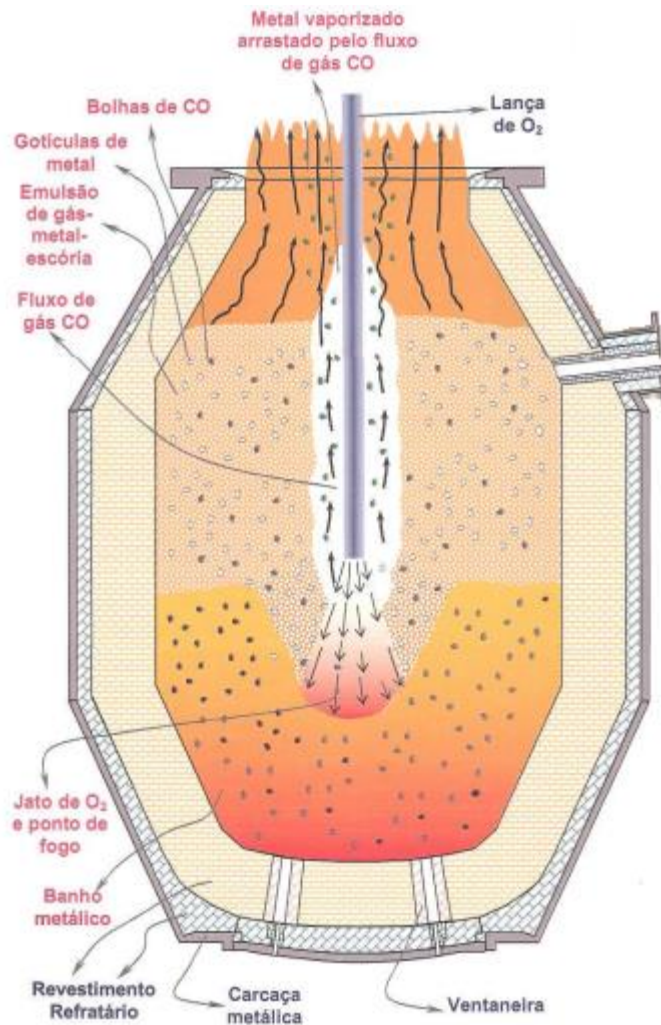


Figura 3.1 – Esquema do sopro de oxigênio no aço em um BOF (RIZZO, 2006).

3.2 A anomalia projeção

Durante o processo de refino primário de gusa no BOF ocorre a formação de uma emulsão gás-metal-escória que é essencial para as reações de refino e é importante para elevada produtividade neste tipo de convertedor (ALBANI, 2017). Este fenômeno é uma das principais preocupações durante produção de aço por oxigênio. Alguns sugerem que as projeções estão relacionadas a quantidades excessivas de gás geradas em uma escória altamente espumada (JUNG, 2000). A projeção está ligada com o volume da emulsão que, se aumentar além da capacidade do

convertedor, leva à expulsão de grande porção da escória e metal, como visto na Figura 3.2. Assim, ela pode ser prejudicial ao meio ambiente devido à poluição que é gerada pelos gases, causando também redução de eficiência de refino e gerando um menor rendimento metálico, além de danos ao equipamento (KOSTÚR *et al.*, 2006; BIRK *et al.*, 2001).



Figura 3.2 – Registro de projeção ocorrida em um convertedor durante visita em uma aciaria.

De acordo com Evestedt *et al.* (2007) a anomalia da projeção é caótica e imprevisível. A persistência desse problema em muitas empresas deu origem a estudos no intuito de se manter um volume de espuma adequado, evitando que a projeção. Assim, nas últimas décadas, ficou evidente que erradicar a projeção se tornou uma tarefa bastante desafiadora e os esforços para desenvolver um sistema de aviso para reduzir e aliviar a projeção são geralmente focados em três áreas principais:

- (1) Modelagem da projeção e seu potencial de ocorrência.
- (2) Dispositivos de medição que detectam o aparecimento de projeção.
- (3) Medidas de mitigação tomadas em tempo real para prevenir o desenvolvimento de eventos de projeção.

Considerando a projeção, portanto, um fenômeno complexo, Walker, *et al.* (2005) documentaram algumas variáveis de influência, que incluem:

- Viscosidade da escória,
- Tensão superficial da escória,
- Densidade de escória,
- Quantidade de partículas de segunda fase dentro da escória líquida,
- Tamanho das bolhas de gás geradas no processo de descarburização,
- Altura, volume e forma do revestimento do convertedor,
- Taxa de geração de gás,
- Efeito de resfriamento ou aquecimento de adições,
- Altura da lança acima do banho,
- Taxa de sopro de oxigênio através da lança,
- Taxa de fluxo da mistura de gás inerte na lança,
- Densidade da carga de sucata,
- Padrão de furos de lança,
- Desgaste do furo da lança,
- Penetração de jato de oxigênio e ângulo de dispersão,
- Química do gusa (P, Si, Ti),
- Química da sucata (Al, Si, Ti, S, P, Mn em particular),
- Instante de adição de fluxantes e minério,
- Velocidade de descarburização,
- Quantidade relativa de pós-combustão dentro ou perto da escória,
- Incrustações na lança,

- Pressão do gás perto da boca do convertedor,
- Introdução esporádica de materiais com química altamente variável e taxa de adição (sujeira na sucata, partículas refratárias).

É possível observar que o fenômeno de projeção é influenciado por muitos parâmetros de processo, altamente variáveis e imprevisíveis, o que torna o estudo de projeções, muitas vezes, difícil (WALKER *et al.*, 2005).

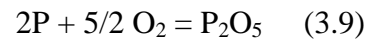
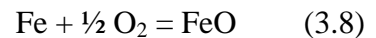
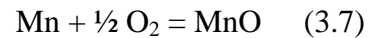
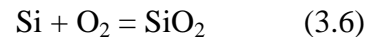
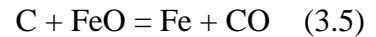
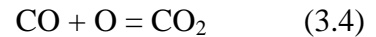
3.2.1 Formação da escória

A principal reação no convertedor é a descarburagem. Porém, o refino do aço também acontece para remoção de outros elementos não desejados que migram para a escória; como por exemplo, o silício, o magnésio e fósforo, ainda presentes no gusa líquido e também na sucata carregada no convertedor (GONÇALVEZ *et al.*, 2014).

De acordo com Rizzo (2006), durante a fabricação do aço, a escória pode exercer algumas funções como: absorver e incorporar os elementos indesejáveis do banho metálico; controlar o suprimento de oxigênio da atmosfera para o metal; agir como um reservatório de substâncias oxidantes; permitir a variação da atividade do oxidante; agir como um isolante físico dificultando a transferência de gases da atmosfera para o banho metálico; reduzir a projeção de metal durante o período de sopro; além disso, atuar na proteção do revestimento refratário do convertedor, depois de concluído o sopro.

Segundo Deo *et al.* (2013), a escória no convertedor é heterogênea e sempre contém bolhas de gás e material sólido não dissolvido ou precipitado. Em nenhum momento a escória é completamente líquida. Uma parte significativa das gotículas de metal se dispersa na escória semilíquida. Através deste mecanismo, as gotículas podem causar a formação de espuma na escória e projeção no BOF.

Dentro do convertedor as seguintes reações químicas principais (3.1 a 3.9) ocorrem (LIM, 2001):



Enquanto o carbono dissolvido no gusa escapa como CO e CO₂ no gás residual, existem, ainda, outros elementos, de acordo com Rizzo (2006), que são incorporados à escória, como Ca, Mg, Si, Al, Zr, Ti e B; elementos que se dividem entre aço e escória, como C, Mn, P, S, Cr, Nb e V; elementos incorporados pelo aço, como Cu, Ni, Sn Sb, Mo, Co, As e W; além dos elementos que deixam o forno na forma de gases ao se vaporizarem, como o Zn.

Uma pequena quantidade de ferro que é oxidada ou vaporizada escapa como fumaça de óxido. A quantidade total de ferro perdida dessa maneira é geralmente menor que 1% de aço para um processo operado com eficácia e de 2,5 a 3,0% de aço para um processo não eficiente, no qual a formação de escória não é adequadamente controlada (DEO *et al.*, 2013).

Para que se tenha uma maior eficácia de refino e uma escória adequada, com propriedades como viscosidade, densidade, tensão superficial e ponto de fusão controlados, durante o sopro acontece a adição de alguns fundentes. Assim, é o caso do CaO, que é imprescindível para que o Si seja escorificado na forma de óxido, e para que também ocorra a retirada do fósforo e enxofre do metal.

Conforme relatos de Silva (1998) a composição típica da escória de um convertedor varia de 10 a 20% de SiO₂, 40 a 50% de CaO, 10 a 25% de FeO, 4 a 10% de MgO, além de MnO e P₂O₅.

Sabe-se que mudanças na composição da escória durante o sopro de oxigênio dependem da conduta e prática do BOF, e que variam de acordo com cada siderúrgica. Levando em consideração, portanto, que a escória é formada por $\text{CaO} + \text{MgO} + \text{FeO} + \text{SiO}_2 = 100\%$, a Figura 3.3 exemplifica a evolução da composição da escória durante o sopro, de acordo com duas trajetórias. A trajetória I é avaliada como boa para se atingir uma taxa rápida de descarburização e mais adequada para baixo teor de enxofre e fósforo na carga do forno, além de conduzir a uma baixa dispersão de metal na escória e uma ínfima chance de projeção do banho. Já a trajetória II, entretanto, obtém-se uma escória líquida com elevada basicidade logo no início do sopro, o que resulta em uma baixa captação de MgO pela escória e uma taxa mais rápida de remoção de enxofre e fósforo o que exemplifica uma escória com altas chances de projeção (TURKDOGAN, 1996).

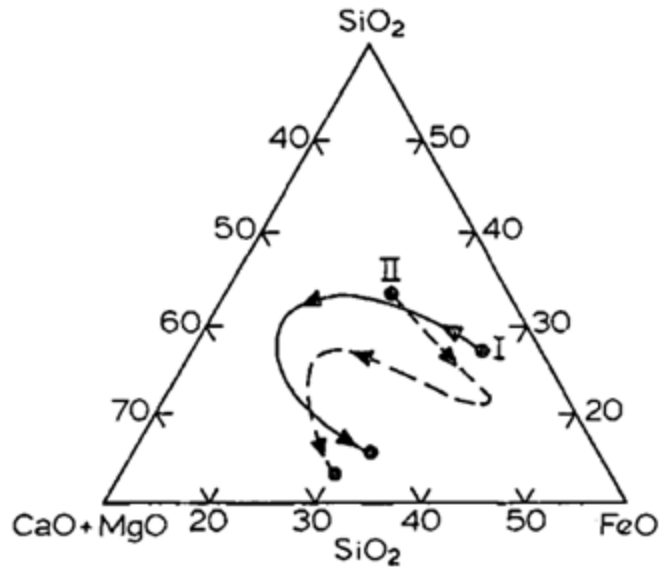


Figura 3.3 – Mudança na composição de escória durante o sopro em um BOF (TURKDOGAN, 1996).

3.2.2 Formação da emulsão

A formação de espuma pela escória é essencial não apenas para a produção de aço no BOF, mas também para a operação dos Fornos de Arcos Elétricos (EAF). Na fabricação de aço nos EAF, as práticas de formação de espuma estão sendo amplamente utilizadas para proteger os refratários da radiação do arco. A espuma também estabiliza o arco e protege o metal da atmosfera. O controle da altura da formação de espuma é necessário para se manter uma operação estável (ITO *et al.*, 1989).

No BOF, logo no início do processo de sopro as reações acontecem, sobretudo, na zona de impacto. Com o desenvolvimento suficiente de uma quantidade de escória, o jato de oxigênio provê a energia mecânica necessária ao início da emulsificação e as reações de oxidação de fragmentos metálicos, que estão lá presentes, ficam em evidência. A criação de bolhas de gás ou gotículas de líquido demandam o fornecimento de uma grande quantidade de energia, proporcional à área das novas superfícies formadas sendo, portanto, afetada pela tensão superficial. O desprendimento de bolhas de CO aumenta à medida que o teor de Si do metal diminui; a velocidade de descarburização então cresce, e os gases gerados no interior da escória

ajudam a estabilizar a emulsão metal-gás-escória. O volume aparente de escória aumenta continuamente com a formação de gases e sua altura supera o nível do bico da lança, que fica imersa na emulsão a partir de 1/4 do tempo de sopro (RIZZO, 2006).

A partir, ainda, do forte movimento que acontece no interior do convertedor ocorrem reações e também a transferência de energia para o banho e para a escória. Enquanto ocorre o sopro, os gases gerados das reações, o aço em forma de gotículas e a escória, como já citado, compõem a emulsão cujo volume é em torno de quatro vezes maior do que a parte de escória e metal não emulsificada (TURKDOGAN *et al.*, 1998).

A formação da emulsão é influenciada pelas substâncias que podem ser consideradas interfacialmente ativas; substâncias que levam à diminuição de energia livre interfacial preferencialmente se acumulam na interface. Quando uma bolha chega à superfície de um líquido, ela carrega consigo uma fina camada de líquido. De acordo com Lahiri *et al.* (2004), a superfície superior da bolha entra em contato com a superfície livre do líquido primeiro, então a concentração de surfactante é mais alta no topo do filme e, assim, a tensão superficial é mais baixa. No sentido de cima para baixo, ao longo do filme, a concentração de soluto gradualmente se aproxima da composição média, levando a um aumento da tensão superficial. Assim, um gradiente de tensão superficial é produzido no filme ao redor da bolha de modo que o mesmo implica no aparecimento de uma força ascendente. Essa força, chamada de força de Marangoni, impede a rápida drenagem do líquido do filme e faz com que a bolha permaneça na superfície por algum tempo. Uma espuma estável se forma se outras bolhas se juntarem à primeira bolha antes que ela se rompa devido à drenagem do líquido do filme. O tempo de drenagem do líquido, evidentemente, é insignificante na ausência da força de Marangoni, uma vez que não há resistência ao fluxo.

A extensão da vida da espuma e a sua velocidade de formação são normalmente determinadas pela taxa de drenagem da escória; a velocidade de formação de espuma é diretamente proporcional à viscosidade da escória e inversamente proporcional à densidade da escória e diâmetro da bolha. Quando uma pequena quantidade de um composto superficialmente ativo,

como por exemplo P_2O_5 , está presente na escória, a vida útil da espuma aumenta significativamente e a relação anterior não é obedecida. Os compostos fortemente ativos reduzem a taxa de drenagem do filme entre as bolhas adjacentes, de modo que a drenagem do filme controla a vida útil da espuma. A espessura do filme, a força de Marangoni, a viscosidade da escória e a difusão dos compostos ativos determinam a taxa de drenagem do filme. A interação destes fatores conduz a taxas de drenagem de película muito baixas ou elevadas e afeta significativamente a estabilidade do filme e da espuma (LAHIRI *et al.*, 2004).

3.2.3 Causas da projeção

De acordo com Cicutti *et al.* (2001), a reação de descarburização gera grandes quantidades de gás que aumentam o volume de escória e podem produzir projeção. Tal capacidade da escória para aumentar o seu volume é caracterizada pelo Índice de Formação de Espuma, ou *Foaming Index* (Σ), que indica o tempo médio de permanência do gás na camada de espuma, descrito a partir da Equação 3.10:

$$\Sigma = (h - h_0)/v_S^G \quad (3.10)$$

onde h_0 é a altura da escória antes da formação de espuma em centímetros, h é a altura da espuma em centímetros e v_S^G é a velocidade superficial do gás em centímetros por segundo. Segundo Ito e Fruehan (1989a) a velocidade superficial do gás pode ser definida pela Equação 3.11, e o Índice de Formação de Espuma (Σ) também pode ser definido pela Equação 3.12.

$$v_S^G = \frac{Q_g}{A} \quad (3.11)$$

$$\Sigma = \frac{\Delta h}{\Delta v_S^G} \quad (3.12)$$

onde Q_g (cm^3/s) é a vazão de gás, A é a área da seção transversal (cm^2), e Δh é a mudança de altura.

A velocidade superficial do gás (v_S^G) também está correlacionada com a fração de vazios (α), e a velocidade real do gás (v_G (cm/s)), dada pela Equação 3.13 abaixo:

$$v_s^G = \alpha v_G \quad (3.13)$$

Assim, o Índice de Formação de Espuma (Equação 3.14) foi proposto com o intuito de possibilitar um melhor entendimento do processo de formação da emulsão e conseqüentemente seu comportamento em relação à projeção. Ito *et al.*, (1988b) realizaram estudos adicionais e propuseram, em função de dados experimentais:

$$\Sigma = 5,7 \times 10^2 \frac{\mu}{\sqrt{\rho\sigma}} \quad (3.14)$$

Onde μ é a viscosidade da escória (Pa.s), ρ é a densidade da escória (g/cm^3) e σ é a tensão interfacial (N/cm). Caso as propriedades da escória não sejam bem controladas, pode haver projeção (transbordamento do reator) por espumação excessiva.

A partir de estudos baseados em dados industriais Cicutti *et al.* (2001) avaliaram a capacidade de formação de espuma das escórias em diferentes etapas do processo. Verificou-se que o Índice de Formação de Espuma é maior no início do sopro, especialmente devido a maior viscosidade da escória. Assim como a presença de partículas de cal não dissolvidas, a menor temperatura de escória também colabora para aumentar a viscosidade da escória nesse momento do processo. A taxa de fluxo de gás gerada pela reação de descarburização foi calculada usando os dados do teor de carbono medido ao longo de todo o processo. Combinando, portanto, os resultados da velocidade do gás e do Índice de Formação de Espuma, a evolução da altura da espuma pôde ser determinada. A altura máxima foi obtida na primeira metade do sopro, indicando que nesse momento do sopro é mais suscetível a ocorrer projeção.

As causas da projeção, ainda, podem estar relacionadas até mesmo diretamente ao projeto do convertedor, que deve receber um volume de gusa, sucata, fundentes e todas as outras adições, sem que haja sobrecarga. Além disso, existem considerações em relação ao tipo de lança e o bico, que são essenciais para um sopro adequado, pois bicos desgastados ou lanças com cascão originam um jato irregular com formação de ondulação e turbulência, afetando as condições de sopro e a diminuição da profundidade de penetração do jato de oxigênio no banho, o que gera uma má formação da escória com chances maiores de projeção (RIZZO, 2006).

De acordo com Brämning (2010), é comum acontecer projeção devido a uma combinação desfavorável de alguns fatores, tais como, por exemplo:

- Escória com alta viscosidade aparente e/ou tensão superficial reduzida;
- Elevado teor de óxido de ferro na escória;
- Aumento repentino na taxa de descarburização devido à adição de óxidos que podem gerar excesso de gás;
- Volume interno do convertedor menor do que o esperado;
- Ocorrência de balanço térmico inadequado;
- Alta basicidade da escória;
- Sopro duro durante grande parte do sopro;

O silício é um elemento importante no processo de refino. Sua oxidação produz parte do calor necessário ao processo e, além disso, determina o volume de escória formada. Dessa forma, se o teor de silício for considerado alto ($> 0,70\%$), o volume de escória será maior, pois será preciso utilizar uma maior quantidade de cal (CaO) para neutralizar a sílica formada. Sendo assim, com um maior volume de escória, a projeção é mais propícia de ocorrer (RIZZO, 2006).

Ainda, existe a influência do manganês. A oxidação do manganês não gera quantidades significativas de calor; é um elemento indispensável para se alcançar boas propriedades mecânicas do aço. Caso o teor de manganês no gusa seja baixo ($< 0,35\%$), a escória é um pouco menos fluida, diminuindo a tendência de projeções, uma vez que consegue-se evitar o excesso de espumação, que é gerado quando a escória está fluida em excesso; já para um teor elevado de Mn ($> 0,80\%$), há uma maior possibilidade de formação de cascões. Além disso, também é necessário analisar a relação entre o teor de Mn e de Si no gusa. Geralmente, se a relação for maior do que 1 existe uma maior probabilidade de projeções durante o sopro, uma vez que a escória fica mais fluida. Se a relação for menor que 1, pelo contrário, existe uma maior possibilidade de formação de cascão na lança de oxigênio. Assim, o teor de manganês no gusa é normalmente de 0,35 a 0,50% (RIZZO, 2006).

3.2.5 Controle de projeção

Existe uma grande preocupação em relação ao controle e/ou minimização dos casos de projeção em BOF. Algumas tecnologias de controle estão relacionadas à mudança das características do som emitido de dentro do convertedor, o que possibilita detectar quando a emulsão está ultrapassando o limite considerado normal do convertedor; outras estão relacionadas ao tipo de lança utilizada, ao uso de câmeras no topo do convertedor para análise de imagem, e ainda, botões de controle de vazão do sopro. Porém, estes são exemplos de tipos de controles que agem de maneira corretiva, enquanto a proposta de um modelo matemático de predição é, certamente, preventivo, uma vez que já antes do início do sopro se conhece a probabilidade de ocorrência de projeção.

Uma técnica difundida é a utilização de sensores sonoros para análise da altura da emulsão. Este instrumento se baseia na variação do nível de frequência do ruído do convertedor pelo envolvimento da lança pela emulsão metal-gás-escória, admitindo o ajuste, muitas vezes automático, de alguns parâmetros de processo, tais como: vazões de oxigênio e gás inerte, e altura de lança. De forma geral, uma baixa intensidade relativa do som indica que existe uma tendência de projeção durante a etapa de sopro (RIZZO, 2006).

Outro equipamento, denominado *Sonic Meter*, também é utilizado. Esse aparelho, igualmente conhecido como um medidor ultrassônico, é um tipo de equipamento de medição de vazão que verifica a velocidade de crescimento da emulsão dentro do convertedor, transmitindo ondas sonoras ultrassônicas. Se a frequência dessas ondas permanece inalterada de acordo com um padrão, denota-se que a altura da emulsão é apropriada; no entanto, se a altura começa a crescer, a frequência das ondas ultrassônicas mudará, e assim, se a velocidade de crescimento da emulsão é alta, a frequência da onda sonora varia linearmente.

Com o avanço da capacidade e disponibilidade de softwares para análise de imagens, permitiu-se o registro automático e o reconhecimento de ocorrências filmadas por câmeras com uma grande

área de visualização. Assim, com o monitoramento da área interna do convertedor, o sistema varre cada imagem da câmera (25 vezes por segundo) e conta o número de pixels de luz acima de um limite definido dentro de uma área designada do quadro da câmera. Tanto o limite quanto a área de imagem designada devem ser definidos de tal forma que qualquer iluminação adicional seja impedida. Assim, a chance de projeção é calculada continuamente como a porcentagem de pixels claros em relação ao número total de pixels na área de varredura designada. Para armazenamento de dados a longo prazo, um índice de dissipação de calor pode ser criado acumulando o valor do nível de projeção calculado durante o sopro completo. Para qualquer tipo de configuração de câmera e, especialmente, no caso de câmeras cobrindo a boca do convertedor se faz preciso aplicar algum padrão ou sistema de reconhecimento de imagem que possa distinguir faíscas, chamas e projeção (BRÄMMING, 2010).

Kobayashi *et al.* (1983) descreveram um método de predição e controle de projeção de escória no BOF. Um medidor de micro-ondas foi desenvolvido e testado em um reator de 150 toneladas. O resultado foi satisfatório, uma vez que acima de 80% do total de projeções pode ser prevista por este medidor de micro-ondas, pelo menos 30 segundos antes do acontecimento e a frequência da projeção foi reduzida para 1/5 das corridas não controlado pelo uso *on-line* deste medidor.

Shakirov *et al.* (2003) desenvolveram uma técnica que evita alterações nas entradas de material e mudanças operacionais antes mesmo que a emulsão comece a subir utilizando um modelo computacional para previsão e cálculo das alterações. Dessa forma, essa ferramenta não foi baseada na reação a um sinal, mas também em análise de dados. A etapa seguinte foi efetivar a tecnologia em operação rotineira com respostas automatizadas para ajustar os parâmetros de carga e sopro. A ideia foi implementada completamente na planta da Trinecke Zeleznary na República Tcheca, e a quantidade de corridas sem projeção cresceu 15,5%.

Um modo de controle de projeção também foi desenvolvido por Ni *et al.* (2016). A proposta compreende um sopro de gás no topo da superfície da espuma no convertedor. Os resultados dos experimentos físicos mostraram que a altura da espuma pode ser efetivamente diminuída pelo

sopro superior, e que, além disso, diferentes tipos de bicos produzem a uma grande diferença quanto à eficiência de destruição de espuma.

3.3 Análise Multivariada de dados

A análise multivariada faz referência a todos os métodos estatísticos que analisam simultaneamente múltiplas medidas em cada indivíduo ou objeto sob investigação, ou grande número de dados. As análises simultâneas de mais de duas variáveis são, de certa maneira, consideradas como análise multivariada. Em alguns testes, a análise multivariada é uma maneira de realizar em uma análise única o que demandaria múltiplas análises no caso univariado. Em outras situações, porém, nota-se que esse tipo de ferramenta foi projetado para casos específicos multivariados. Para alguns autores uma análise multivariada exige que todas as variáveis sejam aleatórias e inter-relacionadas, de modo que seus diversos efeitos não podem ser interpretados separadamente. Já para outros autores, a finalidade da análise multivariada é medir, explicar e prever o grau de combinação de variáveis (VIALI, 2005).

Embora a análise multivariada tenha surgido da estatística univariada e bivariada, o conceito dessa técnica introduz outras considerações e pontos de grande importância como a variável estatística, que é uma combinação linear de variáveis com pesos experimentalmente determinados; a escala de medidas, que envolve a medida de variação e mensuração em um conjunto de variáveis; e os erros de medida, que podem abranger efeitos menores de precisão (HAIR *et al.*, 2009).

Algumas das técnicas da análise multivariada são: análise de trilha, análise de agrupamento, análise de fatores, análise de componentes principais, análise de conglomerados, correlação canônica, regressão múltipla, dentre outras.

3.3.1 Análise de trilha

A análise de trilha é um método clássico de análise multivariada que envolve aplicações de estudos de causa e efeito, tendo sido descrita por Wright em 1921. Esse tipo de análise avalia o coeficiente de correlação entre as variáveis independentes (variáveis explicativas) e mede os efeitos diretos e indiretos sobre a variável principal, a chamada dependente e de maior importância para a seleção.

Essa técnica é muito utilizada na agricultura, por exemplo, em estudos sobre a genética de plantas, em que determinada característica (como tamanho, peso, diâmetro) é analisada de acordo com a influência de outras (como tipo de semente, casca, etc).

Uma das fundamentais vantagens da análise de trilha é permitir o conhecimento dos efeitos diretos e indiretos que variáveis explicativas exercem sobre uma variável principal, admitindo, assim, constituir qual estratégia será mais eficiente na seleção, para incrementar a melhora (SOUZA, 2013).

A execução da análise de trilha é em geral, segundo Souza (2013), realizada em duas partes:

1 – Formação de um diagrama de caminho que, embora não seja básico para a análise numérica, é muito favorável para exibir graficamente o padrão de hipótese das relações de causa e efeito entre um conjunto de variáveis, isto é, forma uma relação de causa e efeito entre variáveis,

2 – Decomposição das correlações analisadas em um conjunto que indica o efeito direto de uma variável hipoteticamente adotada como causa sobre uma variável abordada como efeito.

3.3.2 Análise de Componentes Principais

A análise estatística de componentes principais (ACP) é uma técnica que transforma sistematicamente um conjunto original de variáveis em outro menor com variáveis que representam a maior parte das informações fundamentais do grupo original sem grande perda. Fundamentalmente, esse novo grupo, chamado de componentes principais (ou fatores), é a combinação linear das variáveis originais, e, dessa forma, a interpretação dos dados fica mais simplificada (DUNTEMAN, 1987).

De acordo com Smith (2002), a ACP é uma técnica útil que encontra aplicação em campos como compressão de imagem e é comumente aplicada para encontrar padrões em dados de alta dimensão. É uma maneira de identificar padrões em dados e expressá-los de maneira a destacar suas semelhanças e diferenças (SMITH, 2002).

Dessa forma, a ACP é amplamente utilizada como procedimentos de redução de dimensão para uma dada coleção de variáveis, e identifica um pequeno conjunto de variáveis sintéticas, chamadas de autovetores, que explica a maior parte da variação total ou comum presente nas variáveis originais (BRYANT *et al.*, 2019).

A ACP considera a variância total e deriva fatores que contêm pequenas proporções de variância única e, em algumas situações, variância de erro. Todavia, os primeiros poucos fatores não contêm variância de erro ou única o suficiente para desviar a composição fatorial geral. Notadamente, com análise de componentes principais, unidades (valores de 1,0) são inseridas na diagonal da matriz de correlação, de modo que a variância completa é trazida à matriz fatorial (HAIR *et al.*, 2009).

Resumidamente, a ACP é indicada quando a redução de dados é um cuidado prioritário para fins de previsão, enfocando no número mínimo de fatores necessários para explicar a porção máxima da variância total representada no conjunto original de variáveis; além de ser apropriada também

no caso de informação anterior indicar que a variância específica e de erro representam uma proporção relativamente pequena da variância total (HAIR *et al.*, 2009).

Miletic *et al.* (2008) desenvolveram um trabalho que fornece uma sugestão sobre como formar matrizes de dados de uma maneira que possa ser aplicada ao processo de detecção e monitoramento de falhas na empresa ArcelorMittal Dofasco. Uma vez formada, a matriz de dados é decomposta usando vetores singulares e valores singulares derivados da análise de componentes principais porque esse método é bastante útil em processos difíceis de medir usando instrumentação tradicional. Ademais, os resultados mostraram que esse método de modelagem orientado por dados é útil para melhorar os processos de fabricação do aço. Em particular, a análise de componentes principais foi adequada para a implementação industrial, pois aborda questões práticas, como colinearidade e dados ausentes. Dessa forma, ficou claro que métodos multivariados foram considerados favoráveis para monitoramento, previsão e também outros procedimentos, como otimização de processo.

3.3.3 Análise de Fatores

Segundo Hair (2009), a análise de fatores é uma técnica simples em que um dos seus objetivos é decidir a estrutura inerente entre as variáveis na análise. Dessa forma, é uma ferramenta que oferece técnicas para avaliar a estrutura de correlações em um grande número de variáveis determinando os conjuntos que são intensamente inter-relacionadas, que são chamadas de fatores. Esses fatores são avaliados como representantes de dimensões dentro dos dados. A análise de fatores pode ser utilizada de forma de perspectiva confirmatória ou exploratória e ao se chegar a seus objetivos, é ajustada com quatro situações, sendo a especificação da unidade de análise, obtenção da redução e/ou resumo de dados, seleção de variáveis e uso do resultado da análise fatorial com outras técnicas multivariadas.

Vale ressaltar que o procedimento de decisão na análise fatorial compreende critérios para o número exato de fatores a considerar se extrair. Grande parte das vezes, a partir do mesmo número de variáveis é gerado o mesmo número de fatores já que há muitas combinações lineares

das variáveis originais. Todavia, embora exista o mesmo número de fatores que de variáveis, alguma dos primeiros fatores têm uma maior representatividade. A seguir, pode-se escolher somente um menor número de fatores.

3.3.4 Análise de *Cluster*

Em situações de análises de dados de grande volume de observações pode ser de interesse criar grupos. A análise de *cluster* é um artifício estatístico muito visual utilizado para classificar informações em grupos, de modo que as informações dentro de um mesmo grupo (*cluster*) sejam homogêneas, e informações em diferentes grupos sejam heterogêneas entre si (OLIVEIRA *et al.*, 2019; ILLUMINEO, 2019).

Nesse tipo de análise existe o método hierárquico, que resulta em um diagrama de árvore ou dendograma, como exemplo, mostrado na Figura 3.4. É possível aferir o grau de deformação provocado pela construção do dendograma através da Distância Euclidiana, do coeficiente de Pearson (r) e Distância Manhattan. O coeficiente de Pearson é calculado por meio de índices de similaridade da matriz original e os índices reconstituídos com base no dendograma. Assim, quanto maior for o coeficiente de Pearson, menor será a distorção (OLIVEIRA *et al.*, 2019).

Em suma, a interpretação do dendograma é simples. Se ele for de forma vertical, verifica-se de baixo para cima. Os grupos com maiores relações são incorporados e formam uma primeira ligação, com a distância em y (método Ward e Pearson r) menor possível. Depois disso, outros grupos podem ser formados e assim, outros indivíduos ou podem ir se ligando aos grupos, ou outros grupos podem ir se ligando aos primeiros. Se o gráfico estiver na forma horizontal, o método Ward e Pearson r se encontram na abcissa, e os grupos iniciais também são formados a partir do menor tamanho de x . Não há regra para o limite de divisão dos grandes grupos, ou seja, depende de cada pesquisador e de cada circunstância. Um modelo é o da Figura 3.4 que mostra cinco grupos, com um limite determinado baseado na maior diferença entre uma ligação e outra.

DENDOGRAMA

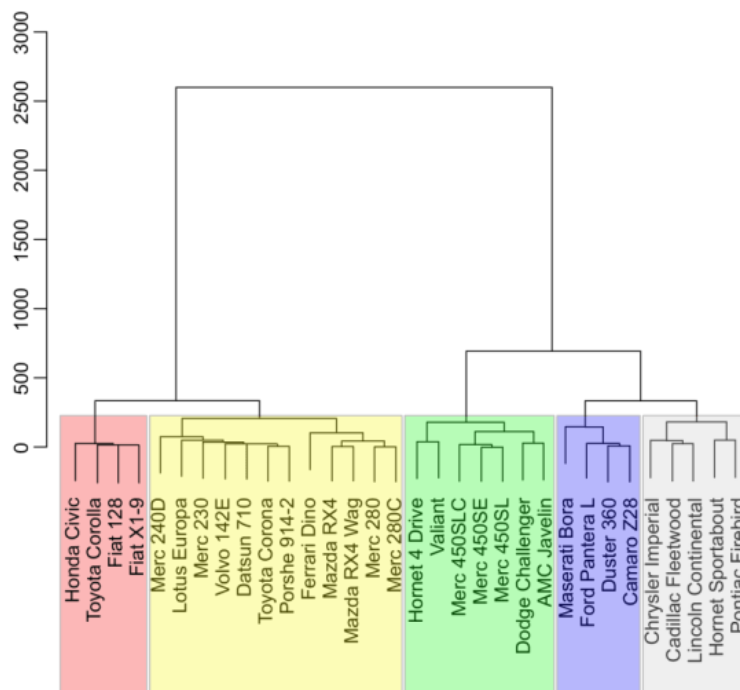


Figura 3.4 – Exemplo de dendrograma a partir de uma análise de *cluster* (OLIVEIRA *et al.*, 2019).

3.3.5 Análises de Regressões

Regressão é uma das práticas estatísticas mais utilizadas para buscar relação entre duas ou mais variáveis. Assim, é possível criar modelos matemáticos que descrevem relações entre variáveis explicativas de um determinado processo.

A regressão linear é o procedimento de traçar uma reta através dos dados em um diagrama de dispersão, e divide-se em simples e múltipla. A regressão linear simples é ideal para avaliar a relação linear entre duas variáveis, sendo uma resposta e uma preditora. Na regressão linear simples, a relação entre duas variáveis pode ser simulada por uma linha reta, com uma relação direta de causa e efeito. Deste modo, é possível prever os valores de uma variável dependente com base nos resultados da variável independente, como ocorre num gráfico de uma equação de

primeiro grau. Muitas vezes, uma única variável preditora não é capaz de explicar completamente a variável resposta, e então é necessário realizar uma regressão linear múltipla. Assim, a diferença entre a regressão linear simples e a múltipla é que nesta última são tratadas duas ou mais variáveis explicativas (PORTAL ACTION, 2019; MONTEIRO, 2019).

Já a análise de Regressão por Componentes Principais (RCP) é um método de regressão que é vastamente utilizada quando se tem muitas variáveis independentes ou há multicolinearidade existente entre os dados. Ela é realizada em duas partes, primeiro obtendo os componentes principais, e depois executando a análise de regressão nos componentes principais (CP). Suas características mais básicas são a redução de dimensionalidade e a remoção de multicolinearidade (MONTEIRO, 2018).

3.3.6 Rede de Correlações

A análise da rede de correlação é uma técnica multivariada que permite visualizar a associação padrão entre as variáveis investigadas, o que ajuda a interpretar conjuntos de dados complexos por meio da identificação de componentes-chave visuais. Assim, as redes formadas entre uma variável e outra são representadas como nodos e as relações entre elas são representadas por linhas verdes, quando positivas, e vermelhas quando negativas. Além disso, são visíveis as medidas de conectividade e proximidade de modo que despontem o quão cada variável conecta outros pares na rede e a intensidade de associações entre as demais variáveis. Ou seja, cada linha possui um peso que indica a força da correlação: quanto mais espessa é a linha de ligação, mais forte é a correlação entre duas variáveis, como representado na Figura 3.5 (COMPARINI *et al.*, 2017; ZUSHI *et al.*, 2015).

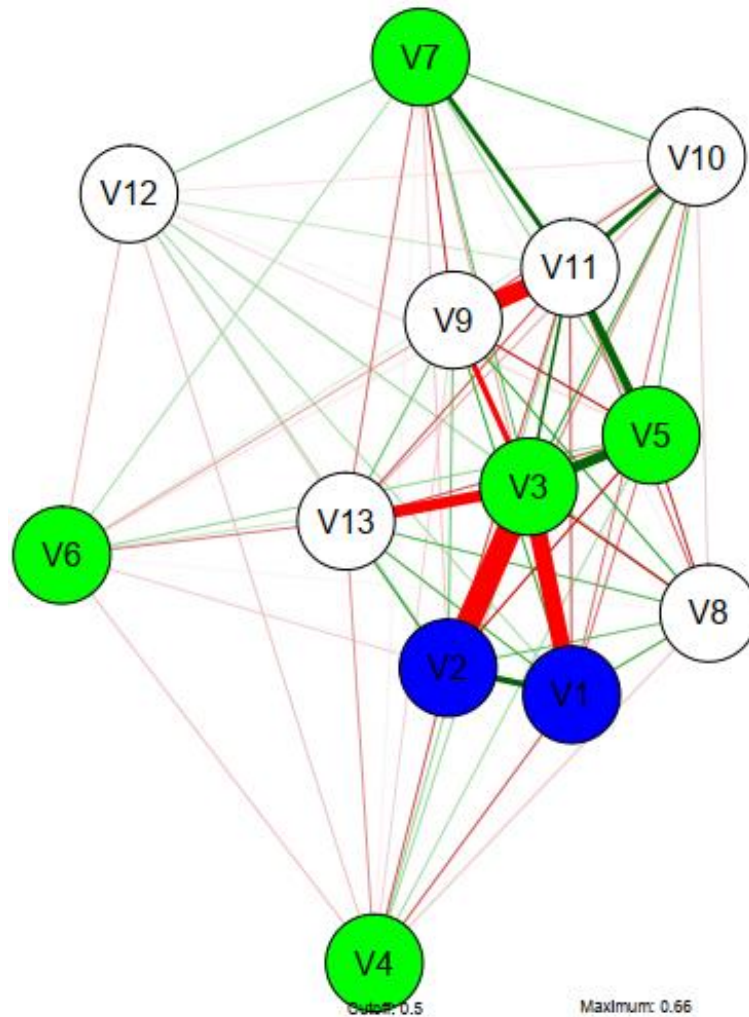


Figura 3.5 – Exemplo do gráfico de rede de correlações entre 13 variáveis gerado no software Rbio.

3.4 Redes Neurais Artificiais

Inteligência artificial envolve toda e qualquer técnica da ciência da computação que procura, por meio de algoritmos computacionais, construir mecanismos, técnicas ou dispositivos que simulem a habilidade de resolução de problemas e pensamentos. As redes neurais podem ser consideradas uma técnica de inteligência artificial.

As redes neurais artificiais (RNA) são modelos de processamento de dados que simulam redes de neurônios biológicos, capazes de realizar o aprendizado de máquina a partir de experiência

passada contida em dados acumulados, recuperar rapidamente arquivos de *Big Data*, e reconhecer padrões, fundamentados no treinamento. São sistemas que se adaptam a partir de uma abordagem de processamento distribuído, em que os neurônios se comunicam por meio de uma rede de elos interconectados. As redes neurais (RN) têm a capacidade de aprender e guardar o conhecimento adquirido, que são utilizados principalmente no reconhecimento de padrões baseados na experiência, tanto para fazer previsões quanto para realizar classificações (CORRAR, 2012).

Uma das razões para a abrangente área de aplicações das RNA é a capacidade que as redes neurais têm de generalização, auto-organização e processamento temporal que possibilita a resolução de diferentes problemas de alta complexidade (FIORIN *et al.*, 2011).

Viana (2018), por exemplo, utilizou dois modelos de redes neurais artificiais treinados tendo como base dados de uma empresa a partir de uma união de cálculos termodinâmicos para dissolução de ligas, adição de formadores de escória e reação de desoxidação no aço líquido com o intuito de prever a temperatura do aço líquido do final do sopro no BOF até a primeira medição de temperatura no refino secundário. Os resultados foram satisfatórios, gerando economias durante a operação e, ainda, mostraram que é possível utilizar RNA para modelar um processo industrial multivariado complexo.

David *et al.* (2014) utilizaram as redes neurais artificiais como meio para predição de teor de enxofre no ferro gusa no processo de alto-forno. As variáveis utilizadas no modelo de RNA foram significativas para prever o teor de enxofre, pois apresentaram boa correlação entre o conteúdo calculado e o real. Assim, diante à complexidade dos mecanismos envolvidos no processo de redução em alto-forno, as RNA foram utilizadas como uma ferramenta favorável.

Outro exemplo de uso de redes neurais em procedimentos siderúrgicos foi realizado por Rajesh *et al.* (2010) para a previsão do consumo final de oxigênio no convertedor LD usando um processo em duas etapas com o desenvolvimento de um modelo com alimentação em múltiplas

camadas. Na etapa um, a temperatura intermediária de parada foi prevista e, utilizando-a como uma entrada, o oxigênio final foi previsto. Em ambos os casos, duas camadas ocultas (ou escondidas) deram os melhores resultados em comparação com a rede neural de camada única. As temperaturas intermediárias e finais do sopro desempenharam um papel vital nas previsões de temperatura de parada de oxigênio e parada intermediária, enquanto o peso de gusa e sucata tornaram-se parâmetros secundários nas previsões do modelo. Como resultado, o modelo atua como guia para o operador e, assim, aumenta o rendimento do processo de fabricação de aço do convertedor.

De forma fundamental, as RNA são sistemas paralelos e distribuídos constituídos de unidades simples – os neurônios – que computam funções matemáticas. Os neurônios fazem parte do cérebro e são células que são compostas pelo corpo, dendritos e axônio, como mostrado na Figura 3.6. Os dendritos capturam os estímulos e os transmitem ao corpo do neurônio, onde são processados. Quando os estímulos atingem determinado limite, o corpo da célula envia novo impulso que passa pelo axônio e é transmitido até as células vizinhas por meio da sinapse. Este procedimento pode se repetir em várias camadas de neurônios. Por fim, a informação de entrada é processada, podendo levar o cérebro a comandar reações físicas (FERNEDA, 2006).

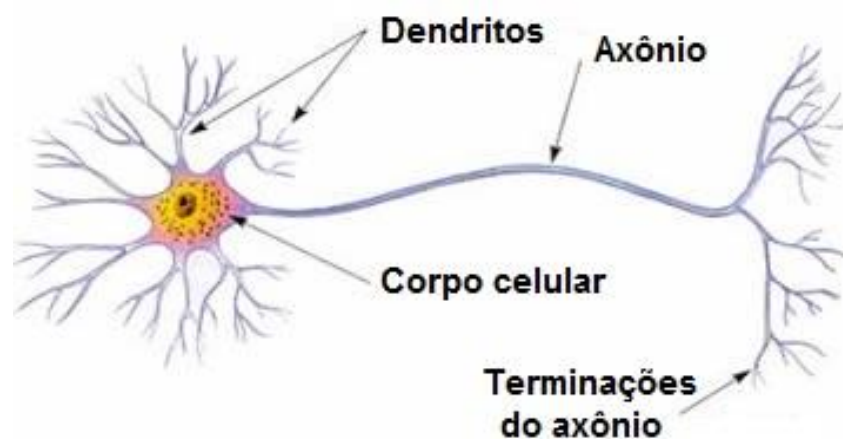


Figura 3.6 – Esquemática de um neurônio (VIRTUOUS TECNOLOGIA DA INFORMAÇÃO, 2008).

Resumidamente, as entradas dos dados são representadas pelos dendritos e a saída pelos axônios. Vários modelos de redes neurais são descritos na literatura, e o mais usual é o modelo chamado de multicamadas (com pelo menos uma camada escondida), representado da Figura 3.7 b, e o modelo simples, em 3.7 a.

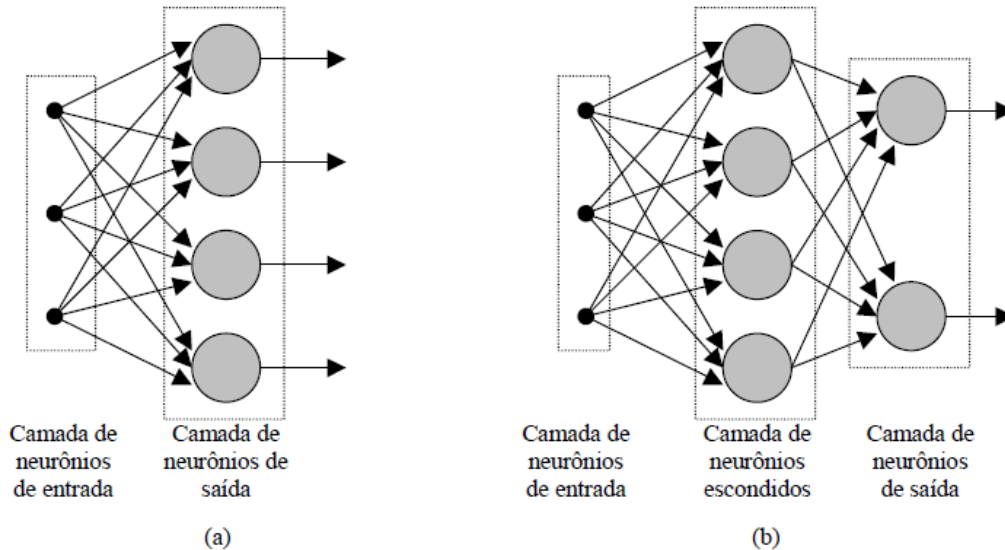


Figura 3.7 – Estruturas de redes neurais em camadas (a) rede com apenas duas camadas de neurônios, (b) rede com uma camada de entradas, uma camada escondida, e uma camada de saída (ZUBEN, 1993).

3.4.1 Aprendizagem de uma rede neural

Segundo Zuben (1993), existem alguns métodos aplicados na aprendizagem de uma rede neural, como o aprendizado supervisionado, que é baseado em um conjunto de exemplos de estímulo-resposta, ou em certo tipo de conhecimento, que represente o comportamento que deve ser apresentado pela rede neural; o aprendizado por reforço, que é quando o comportamento da rede é avaliado apenas com base em determinado critério numérico, fornecido em instantes espaçados de tempo; e o aprendizado não supervisionado, que é baseado apenas nos estímulos recebidos pela rede neural. De forma simples, a rede deve aprender a catalogar os estímulos.

De acordo com Corrar *et al.* (2012), nas redes que os métodos de aprendizado são por retropropagação, o resultado esperado deve ser exposto à rede para que esta possa aprender. O algoritmo de retropropagação trabalha com o modo de aprendizado conhecido como supervisionado, formado por três etapas:

1. A rede trabalha com os exemplos de treinamento e calcula as saídas (gerando respostas);
2. O algoritmo, logo, calcula o erro, a partir da diferença entre as saídas e o resultado esperado;
3. O erro é retroalimentado através da rede e os pesos são ajustados para minimizar o erro.

O foco desse algoritmo, portanto, é utilizar o valor do erro para ajustar os pesos de conexão entre os neurônios. O sinal de erro $e_j(n)$ na saída do neurônio j , segundo Zuben (1993), na iteração n (na apresentação do n -ésimo padrão de treinamento) é dada pela Equação 3.15:

$$e_j(n) = s_j(n) - y_j(n) \quad (3.15)$$

onde $s_j(n)$ é a resposta desejada para o neurônio j da camada de saída, sendo $y_j(n)$ a saída. O valor instantâneo do erro quadrático para o neurônio j é definido por $\frac{1}{2} e_j^2(n)$. A soma dos erros quadráticos instantânea da rede logo é definida pela Equação 3.16:

$$\varepsilon(n) = \frac{1}{2} \sum_{j \in C} e_j^2(n) \quad (3.16)$$

Em que C representa o conjunto que contém todos os neurônios da camada de saída da rede. Então, se N é o número total de padrões de treinamento contido no conjunto de treinamento o erro quadrático médio é dado como na Equação 3.17:

$$\varepsilon_{av} = \frac{1}{N} \sum_{n=1}^N \varepsilon(n) \quad (3.17)$$

É possível perceber que a soma dos erros quadráticos instantâneos, e, conseqüentemente o erro quadrático médio, é uma função de todos os pesos sinápticos da rede, já que são eles que definem a função de mapeamento ou transferência de entrada-saída da rede neural. Para um dado

conjunto de treinamento, ϵ_{av} representa uma função de custo, ou seja, uma medida de desempenho da rede neural. A finalidade do processo de treinamento é exatamente minimizar ϵ_{av} .

3.4.2 Classificador *Multilayer Perceptron*

O algoritmo *Perceptron* é simples e destinado a realizar a categorização que prevê se a entrada pertence a uma determinada categoria de interesse ou não, como por exemplo, positivo ou negativo; falso ou verdadeiro. É como um classificador linear que categoriza a entrada que separa duas divisões com reta. Originalmente, a entrada é um vetor de recursos – multiplicado por pesos – e adicionado a um viés (bias). Assim, o *Perceptron* produz uma única saída com base em várias entradas de valor real, formando uma combinação usando os pesos. O *Multilayer Perceptron* (MLP) é uma ferramenta utilizada em redes neurais formada por mais de um *Perceptron* que são arranjados por uma camada de entrada para receber o sinal, uma camada de saída que toma uma decisão ou previsão sobre a entrada, e entre estes, um número arbitrário de camadas ocultas. O MLP com uma camada oculta é preparado para aproximar qualquer função contínua e é constantemente aplicado a situações de aprendizagem supervisionada, isto é, treinam em um grupo de pares entrada/saída e aprendem a modelar a correlação entre estas. O treinamento envolve o ajuste dos parâmetros, ou os pesos e viés, do modelo para minimizar o erro (DATA SCIENCE ACADEMY, 2019).

3.4.3. Validação Cruzada *K-fold*

O validador cruzado *K-fold* fornece índices de treinamento e de testes dividindo os dados em conjuntos de treinamento/teste. Assim, essa técnica divide todos os dados em grupos de amostras, como dobras de tamanho K, de tamanhos iguais. A função de predição é aprendida usando as dobras (K-1) e a dobra deixada de fora é usada para teste. Assim, se o valor de K escolhido é igual a 20, então todos os dados serão divididos em grupos de 20, e então 19 serão treinadas, e a restante será utilizada para a validação e adequação ((KIT-LEARN DEVELOPERS BSD LICENSE, 2019; CORLETA *et al.*, 2017).

4 PROCEDIMENTO EXPERIMENTAL

Informações de Big Data de cerca de 60 mil corridas de aço dos três convertedores BOF da empresa ArcelorMittal Tubarão (AMT) foram cedidos para análise. Porém, para este trabalho, serão considerados apenas os dados do convertedor número 1 para análises mais profundas, uma vez que cada BOF possui suas particularidades, e os resultados não se distanciam, de acordo com análises prévias.

Os dados estão contidos em uma planilha no Excel, como exemplificado na Figura 4.1, com o código identificando as corridas do convertedor em cada linha na vertical e as informações, dadas como variáveis, na horizontal. As variáveis, que são descrições de cada corrida, com indicações de data e hora de início e fim de sopro, teores de Si, P, Mn finais, se houve ou não projeção, dentre outras, totalizam 160 variáveis.

	B	C	E	F	G	H	I	J	K	L	M	N	O	
	Código resumido da corrida	Data de vazamento	Letra do turno de operação	Peso de sucata de gusa lingotado (t)	Peso de sucata de aço tipo C (t)	Peso de sucata de gusa moldado (t)	Peso de sucata de gusa C (t)	Peso de sucata de gusa A (t)	Peso de sucata de gusa desclassificada (t)	Peso de sucata de aço tipo A (t)	Peso de sucata de gusa tipo b (t)	Peso de sucata de aço comprada tipo a (t)	Peso de sucata de aço importada (t)	Peso de sucata de gusa lingotado (t)
5	10037	2016-01-01 6:35	D	12	0	0	0	0	0	10.8	7.6	0	0	0
6	10048	2016-01-01 17:33	A	7.1	0	6.2	0	0	0	7.6	9.3	0	0	0
7	10051	2016-01-01 20:44	D	0	0	0	0	0	0	0	0	0	0	0
8	10052	2016-01-01 21:37	D	14.2	0	9.4	0	0	0	7.4	0	0	0	0
9	10053	2016-01-01 22:47	D	0	0	0	0	0	0	8.1	7.3	0	0	0
10	10054	2016-01-01 23:28	D	0	0	0	0	0	0	0	0	0	0	0
11	10055	2016-01-02 0:17	D	0	0	0	0	0	0	0	0	0	0	0
12	10056	2016-01-02 1:01	D	0	0	0	0	0	0	0	0	0	0	0
13	10057	2016-01-02 1:44	D	0	0	0	0	0	0	0	0	0	0	0
14	10058	2016-01-02 2:31	D	0	0	0	0	0	0	9.8	5.3	0	0	0
15	10059	2016-01-02 3:08	D	0	0	0	0	0	0	0	0	0	0	0

Figura 4.1 – Planilha com parte dos dados cedidos para análise.

Outros dados também foram cedidos pela AMT, por meio de uma planilha com informações do uso de um botão de controle, que é acionado toda vez que o operador percebe que há chances de

ocorrer projeção durante o sopro. Este botão age diretamente no padrão de sopro; uma vez que é acionado, ocorre a diminuição da vazão de oxigênio e controle da altura de lança, e assim o operador é capaz de evitar a projeção e/ou diminuir a intensidade que ela ocorreria.

A partir dessas planilhas, foi possível identificar as corridas com projeção, e assim, classificar as variáveis que mais se correlacionam com este fenômeno, a partir de análises multivariadas nos softwares Statistica, Genes e RBio; a regressão linear no software NCSS; e por fim, foram construídas redes neurais em linguagem Python utilizando o software Spyder para a construção do modelo matemático que prevê a ocorrência de projeção.

Assim, alguns testes prévios foram realizados até que fosse possível a identificação dos resultados mais confiáveis. Dessa forma, o primeiro grupo de dados seguiu a ordem de análises multivariadas e depois regressões; e o segundo grupo, seguiu a ordem de análises multivariadas e redes neurais artificiais.

4.1 Análises do Primeiro Grupo

Estudos anteriores dessas planilhas mostraram que alguns dados poderiam conter informações imprecisas. Assim, corridas de aço com dados de um ano foram delimitadas e utilizadas para análises, o que incluíram 7735 corridas, com informações referentes a 39 variáveis de interesse, listadas no Anexo II.

Para que não houvesse incompatibilidade na interpretação dos dados em relação às unidades diferenciadas (ex.: temperatura do gusa, peso do briquete, porcentagem de Mn no gusa), testes com os números sem a padronização foram realizados. Porém, os resultados dos erros foram altos. Dessa maneira, para a padronização dos números de 0 a 1, utilizou-se a equação 4.1:

$$Xa' = \frac{Xa - X_{mín}}{X_{máx} - X_{mín}} \quad (4.1)$$

onde Xa' é o valor padronizado, Xa é o valor da variável, $Xmín$ é o valor mínimo da coluna e $Xmáx$ o valor máximo da coluna.

Os dados foram transportados para uma matriz e seus valores padronizados. Em seguida, aplicaram-se as técnicas estatísticas de Análise de Cluster, Análise de Fatores (AF) e Análise de Componentes Principais (ACP), utilizando o software *Statistica* versão 7.0.

A Análise de *Cluster* permite agrupar as variáveis com base em critérios pré-definidos, gerando um dendograma, o qual possibilita a realização de uma análise de afinidade entre as variáveis independentes e destas com a variável dependente. Na AMT, ao atribuir valores a essa variável resposta, o operador da aciaria insere, com base em sua observação, um valor numérico representando se aconteceu ou não projeção na corrida. A planilha de dados analisada apresenta em duas de suas colunas dados da variável resposta: a primeira (neste caso: variável 25) avalia se a projeção ocorreu até a primeira metade do sopro e a segunda (neste caso: variável 26) avalia se a projeção ocorreu da metade até o final do processo. Essas duas variáveis determinam qual agrupamento deve ser selecionado para as etapas seguintes de análise. A técnica estatística da Análise de Fatores foi aplicada com a finalidade de reduzir o número de variáveis independentes do conjunto de dados a partir da criação de novas variáveis sintéticas, chamadas de fatores; e assim, determinar quantas das novas variáveis obtidas deveriam ser analisadas, bem como determinar qual seria a correlação entre as variáveis originais e os fatores, e quais os índices de contribuição das variáveis originais nos fatores. Por sua vez, a técnica de Análise de Componentes Principais foi aplicada para substituir as variáveis por outras sintéticas, o que permite indicar o número ideal de fatores a ser analisado.

As variáveis que mais contribuem para as CP foram consideradas para se obter um modelo que prevê projeção a partir de análises de regressão linear múltipla e de Regressão por Componentes Principais (RCP). As variáveis respostas do *Statistica* foram utilizadas nesta etapa porque permitem a simplificação da análise, ao excluir as variáveis que não possuem efeitos significativos sobre a projeção. Em consequência, foi utilizada uma matriz formada apenas com

os valores de variáveis que mais influenciam a projeção; além do mais, as duas variáveis de projeção foram transformadas em apenas uma, com valores assinalados como 0 para não projeção e 1 para projeção.

O programa *NCSS* aplicado a essa matriz gerou um relatório sobre a regressão que incluía diversas outras análises estatísticas com gráficos explicativos, porém, só o essencial para este estudo foi levado em consideração. Primeiro se fez a RCP que avalia dados de regressão múltipla que possuem multicolinearidade. Quando ocorre a multicolinearidade, as estimativas dos mínimos quadrados são imparciais, mas suas variações são grandes, e podem, portanto, estar longe do verdadeiro valor. Assim, a RCP reduz o erro padrão, ou seja, esperam-se estimativas mais confiáveis (HAIR *et al.*, 2009). Em seguida, para que fosse possível uma comparação de resultados, fez-se a regressão múltipla, que analisa o relacionamento entre a variável dependente (projeção ou não projeção) e as variáveis de processo independentes. Por fim, o modelo matemático foi testado.

Assim, a sequência das análises do primeiro grupo pode ser resumida como demonstrado na Figura 4.2.

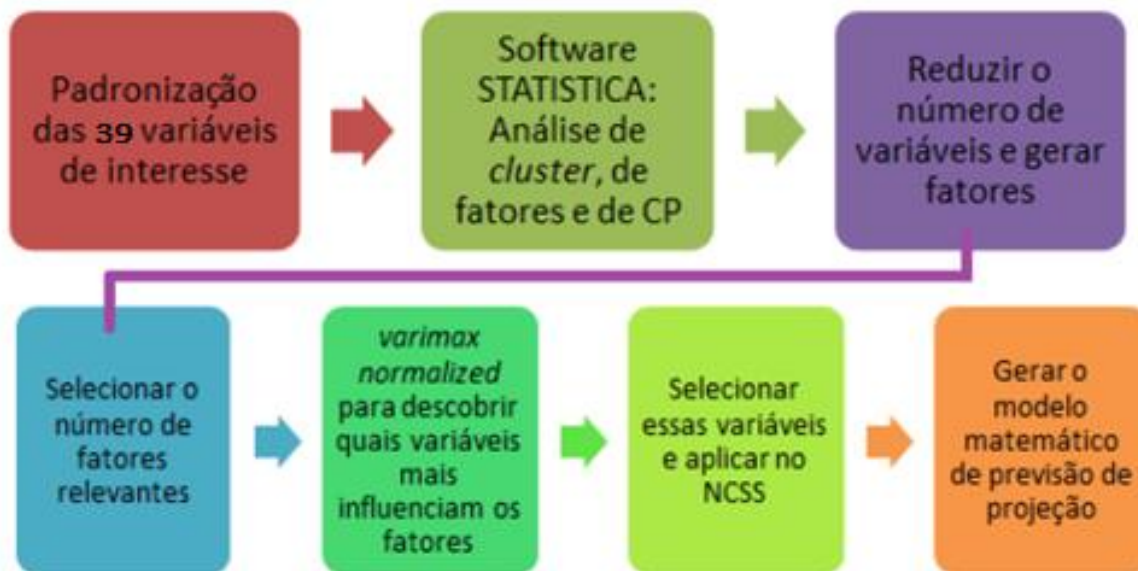


Figura 4.2 – Esquema com a demonstração do passo a passo das análises do primeiro grupo.

4.2 Análises do Segundo Grupo

O maior desafio encontrado para o início da análise de projeção, a partir dos resultados anteriores, foi com relação à veracidade da informação da variável projeção, posto que ao contrário de todas as outras, o registro desta não é automático, sendo que é reportada de acordo com a interpretação do operador. Assim, existe, por exemplo, a possibilidade de se encontrar corridas com projeção, mas sem registro, o que pode influenciar na interpretação final da análise.

Para a mitigação das incertezas, dois tipos de dados foram analisados em conjunto: a planilha com os dados com a variável projeção e a outra planilha com os dados do uso do botão projeção.

De forma geral, há quatro possibilidades de respostas:

1. O operador informou que houve projeção e pressionou o botão;
2. O operador não informou que houve projeção e não pressionou o botão;
3. O operador não informou que houve projeção, mas pressionou o botão;
4. O operador informou que houve projeção e não apertou o botão.

Assim, ao se considerar apenas os dados com as possibilidades 1 e 2, e eliminar as de número 3 e 4, as chances de se trabalhar com dados mais precisos são maiores, uma vez que a 1 foi considerada como corridas com projeção e a 2 como corridas sem projeção. Outros testes utilizando as opções 1, 2, 3 e 4; 1, 2, e 4 foram realizados, para efeito de comparação dos erros.

Após a compatibilização das informações das duas planilhas (uma com informações de todas as corridas e suas variáveis; outra com dados do botão de controle) e com a identificação das corridas de interesse, o próximo passo foi realizar uma filtragem das 160 variáveis iniciais, com o objetivo de eliminar dados em repetição. Com a planilha em formato de matriz, com 44 as variáveis selecionadas (vide Anexo II), também foi feita a padronização dos dados.

4.2.1 Determinação das variáveis de influência da projeção

Para a determinação das variáveis que mais estão relacionadas com a variável projeção, realizou-se, inicialmente, a análise de componentes principais com o objetivo de se obter o número ideal de variáveis que deveriam seguir para as próximas etapas de análises, sem perdas de informações, e eliminação daquelas de menor significância para o processo. Para isso, utilizou-se o software *Statistica*.

A análise de trilha foi realizada no software *Genes*, um programa livre também utilizado para estatística experimental. A análise foi iniciada ao se fornecer ao software *Gene* o nome do arquivo que contém os dados, além de alguns parâmetros como o número de variáveis e seus respectivos nomes. Após essa etapa, segue-se um diagnóstico de multicolinearidade.

A necessidade da verificação da multicolinearidade vem a partir da indicação de existência de forte correlação entre duas (ou mais) variáveis independentes e da existência de relação linear entre uma variável explicativa e as demais. Assim, se as variáveis forem muito correlacionadas, as interpretações das análises podem ser pouco confiáveis ou até mesmo errôneas.

Com a utilização do software *Rbio*, foi possível fazer o diagnóstico de multicolinearidade. Neste processo foi verificado o valor de VIF (Fator de Inflação de Variância). Se o valor informado após a análise fosse maior que 10 significa que há multicolinearidade. Então, caso o diagnóstico de multicolinearidade for estabelecido, é preciso analisar o número de condição (NC), que é a razão entre o maior e o menor autovalor da matriz. Sendo NC inferior a 100, a multicolinearidade é considerada moderada e não estabelece problema para a análise; se NC estiver entre 100 a 1000, a multicolinearidade é forte, e se NC for maior que 1000, a mesma é considerada severa.

Verificada a multicolinearidade, a análise de trilha escolhida foi de acordo com o resultado de haver ou não a colinearidade no software *Genes*. A partir deste ponto se fez a análise de trilha

com o software citado, que permitiu estabelecer os efeitos diretos de cada variável na variável principal projeção, e a partir dos resultados da ACP, escolheu-se exatamente o número de variáveis que o *Statistica* indicou anteriormente, seguindo a ordem de quem tem maior influência sobre a projeção, como indicado no *Genes*.

Uma vez feito isso, foi possível realizar mais uma breve análise visual utilizando a técnica de rede de correlações no *Rbio* que mostrou as variáveis que mais se relacionam uma com a outra, sendo possível identificar quais as que mais se conectam com a variável projeção, como uma forma de confirmação em relação à análise de trilha.

4.2.2 Montagem do modelo matemático a partir de uma rede neural

Para desenvolver a rede neural e montar o modelo para predição de corridas com projeção foi utilizada a linguagem *Python*. Essa linguagem teve sua primeira publicação em 1991 e possui características que possibilitam escrever os mesmos requisitos em menos linhas de código que o necessário em outras linguagens de programação, como por exemplo, o Java e o C#. Atualmente, o *Python* é muito utilizado na construção de soluções web, assim como em aplicações que lidam com processamento de texto, *machine learning*, redes neurais e recomendação de conteúdo (YAMAZACK, 2018). As RNA também foram escolhidas para gerar um modelo matemático por apresentarem, segundo a literatura, resultados com erros muito menores se comparados com outras análises, como regressões.

Os resultados obtidos a partir da análise de trilha e confirmadas pelas análises de redes de correlações foram transportados para o software *Spyder* para continuação das análises. Após a criação do programa específico em *Python* para predição de projeção, a rede neural foi treinada a partir das variáveis que mais estavam envolvidas com a projeção. Por fim, após o treinamento, um modelo de predição de projeção foi desenvolvido, com a indicação de que corridas com possibilidades de projeção indicaram o número 1, enquanto corridas sem possibilidades de projeção indicaram o número 0.

A técnica utilizada para validação foi a *k-fold*, onde *k* foi determinado, através de tentativas, o que apresentou um melhor resultado, juntamente com o valor de número de camadas escondidas. Foram utilizadas 5956 corridas para treino, destas, 216 eram corridas com resposta 1, ou seja, que tiveram projeção. Para validar o modelo, foram separadas outras 425 corridas, e destas, 25 resultaram em projeção, porém, as respostas foram omitidas, para que o próprio modelo da RN respondesse, e o resultado pudesse ser comparado com o real.

Assim, o procedimento dos melhores testes pode ser resumido como representado no esquema da Figura 4.3, a seguir.



Figura 4.3 – Esquema com a demonstração com o passo a passo das análises do segundo grupo.

5 RESULTADOS E DISCUSSÃO

5.1 Análises de *Cluster*, Fatores e CP

Foram primeiramente selecionadas 39 variáveis. A partir do dendograma produzido pelo software *Statistica*, decidiu-se por um intercepto próximo de 1,6, em destaque na Figura 5.1, de forma a identificar um grupo com número aceitável de variáveis para análise. Cada número da horizontal no diagrama identifica uma variável contida nos dados. Todas as 7 contidas no agrupamento número 3, identificado no retângulo roxo na Figura 5.1, foram utilizadas na etapa seguinte da análise pois as variáveis de projeção estão melhor conectadas. Assim, com base no uso de análise fatorial para a redução do número de variáveis, obteve-se a Tabela 5.1.

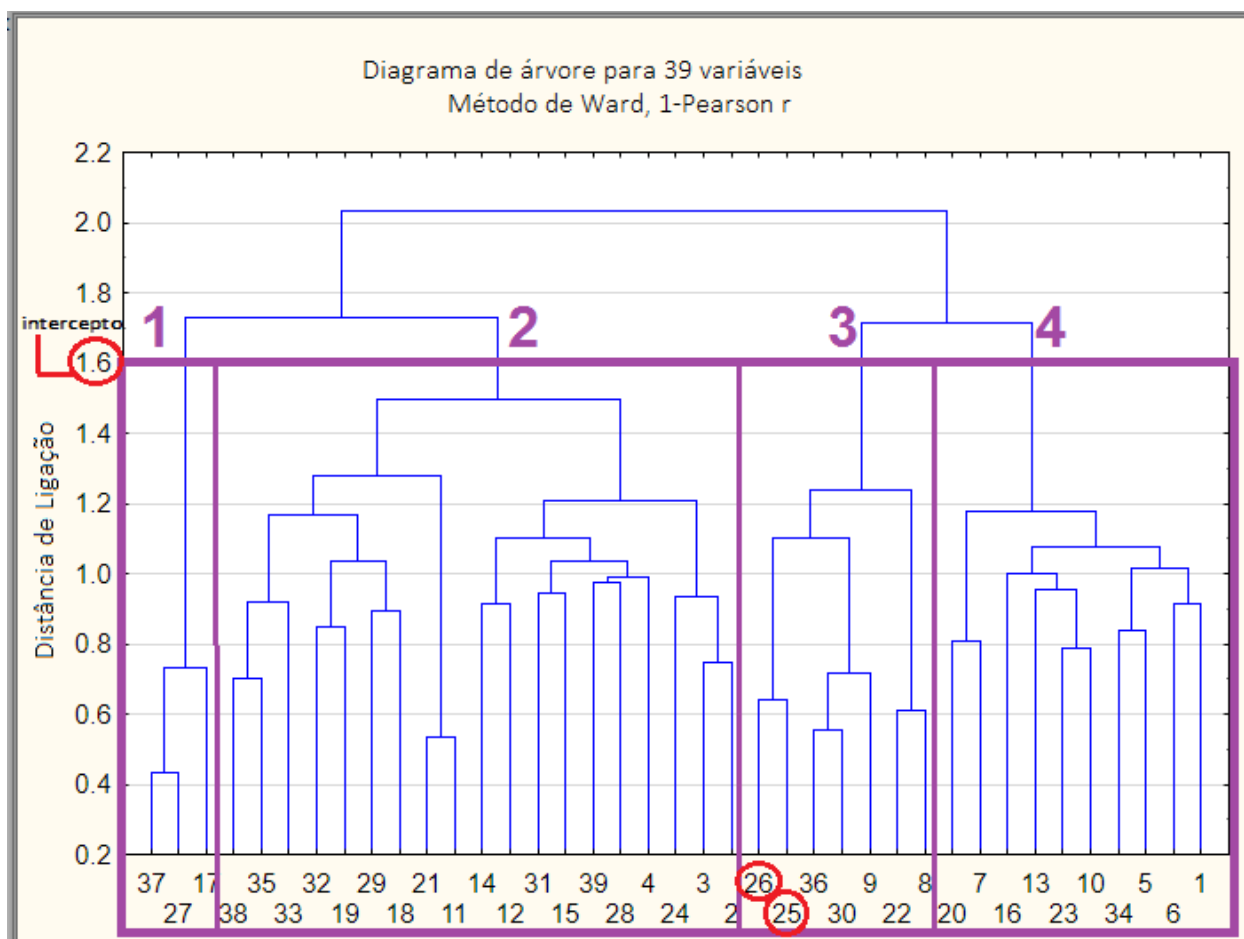


Figura 5.1 – Dendograma formado a partir da Análise de *Cluster*.

Tabela 5.1 – Resultado gerado no *Statistica* com os números de autovalores, variância total, autovalor cumulativo e porcentagem cumulativa de variância para 7 CP.

Número	Autovalor	Variância Total (%)	Autovalor Cumulativo	Cumulativo (%)
1	2,194318	31,34740	2,194318	31,3474
2	1,302686	18,60981	3,497004	49,9572
3	1,029008	14,70012	4,526013	64,6573
4	0,753853	10,76932	5,279865	75,4266
5	0,635194	9,07421	5,915060	84,5009
6	0,579698	8,28141	6,494758	92,7823
7	0,505242	7,21774	7,000000	100,0000

Para as 7 variáveis, o software indicou 7 fatores, porém, nem todas foram consideradas, seguindo-se o objetivo de reduzir o número de variáveis. Para análises estatísticas, considera-se, geralmente, até o fator que apresente porcentagem cumulativa maior que 70, logo, neste caso, até o fator número 4, como destacado de vermelho.

De acordo com a análise das CP, o mesmo conteúdo da Tabela 5.1 foi obtido. Assim, foi possível confirmar o número de CP (fatores) a serem analisados como influências de projeção. A etapa seguinte consistiu no emprego da ferramenta *Varimax Normalized*, a qual verificou quais variáveis originais mais contribuíram para a formação dos fatores. As variáveis selecionadas foram aquelas que possuíam valores significativos (maiores que 0,7) e que mais contribuem para a formação de cada fator, conforme a Tabela 5.2. Neste caso, cada fator foi explicado por apenas uma variável.

Tabela 5.2 – Tabela de contribuição das variáveis originais em cada fator.

Variável	Fator 1	Fator 2	Fator 3	Fator 4
8	0.037918	0.971831	0.056711	0.022496
9	0.044692	0.022098	0.121799	0.972485
22	0.009853	0.193379	0.087873	0.015571
25	0.086824	0.117793	0.215251	0.177252
26	0.177202	0.023608	0.043500	0.099435
30	0.087407	0.058543	0.959315	0.126233
36	0.975468	0.037477	0.083696	0.044286

As variáveis de maior contribuição, para cada fator, são apresentadas na Tabela 5.3.

Tabela 5.3 – Variáveis que mais influenciam nas projeções de acordo com análises prévias

Variável	Fator	Nome
36	1	Si no gusa líquido (%)
8	2	Peso de minério adicionado no início do sopro (t)
30	3	Tempo de sopro de oxigênio (min)
9	4	Peso de cal calcítica (t)

O uso da Análise de *Cluster*, AF e ACP não indica a relação causa efeito das variáveis na projeção, o que precisa ser avaliado pelo operador do processo. Entretanto, estas técnicas permitem agrupar as variáveis correlacionadas à projeção, e reduzir o número de variáveis existentes para análise.

5.1.1 Regressão Linear no NCSS

Os resultados da regressão linear geraram um modelo matemático para a previsão de corridas com projeção. O primeiro teste feito foi a partir da regressão por componentes principais (RCP). Esta análise gera dados e, automaticamente, os compara com o uso da técnica de mínimos quadrados de regressão linear múltipla (RMQ). As variáveis respostas da Tabela 5.3 foram utilizadas de forma a gerar um modelo matemático pelo software NCSS. Foram obtidos os dados da Tabela 5.4. A primeira coluna e a última coluna identificam os fatores e os nomes das variáveis selecionadas. A terceira coluna identifica o coeficiente de cada variável por RCP e a quinta coluna mostra o coeficiente de cada variável por RMQ. Percebe-se, então, que a linha de regressão linear múltipla está melhor ajustada uma vez que apresenta um valor de r^2 superior ao r^2 da regressão por componentes principais. Tomando a regressão linear múltipla, portanto, como a mais adequada dentre as duas, tem-se o seguinte modelo matemático (5.1) gerado para previsão de corridas com projeções:

$$M_{pi} = a + \sum B_i X_i \quad (5.1)$$

Onde M_{pi} é a variável dependente (com valor de 0 para não projeção ou 1 para projeção), a é o intercepto, B_i são os coeficientes da regressão e X_i são as variáveis com os valores encontrados em destaque de cada fator na Tabela 5.2.

Tabela 5.4 – Coeficientes gerados por RCP, Regressão por Mínimos Quadrados (RMQ), via NCSS.

Variável	Fator	B_i (RCP)	Erro (RCP)	B_i (RMQ)	Erro (RMQ)	X_i
		-0,3078		-0,2970		Intercepto
C34	1	0,7547	0,04	0,8609	0,09	Si no gusa líquido (%)
C8	2	0,0862	0,03	0,0862	0,03	Peso de minério adicionado (t)
C28	3	0,8814	0,09	0,7840	0,12	Tempo de sopro de O ₂ (min)
C9	4	0,6648	0,06	0,6378	0,06	Peso de cal calcítica (t)
	r^2	0,5959		0,6570		

Para testar a equação obtida pelo modelo da equação (5.1), utilizaram-se os dados das próprias 7735 corridas. Foram separadas as corridas com projeção e corridas sem projeção. Os resultados para M_{pi} são mostrados na Tabela 5.5.

Tabela 5.5 – Principais valores encontrados para a variável projeção (M_{pi}).

	Corridas com projeção	Corridas sem projeção
Máximo; Média; Mínimo	0,97; 0,24; 0,16	0,86; 0,12; -0,25

O ideal é encontrar o valor 1 para corridas com projeção e 0 para corridas sem projeção, o que não foi alcançado. A imprecisão pode estar diretamente ligada à subjetividade do valor real de M_{pi} , pois existem corridas em que houve projeção e o valor da variável foi anotado como 0, ou seja, dessa forma ficou claro que para resultados mais exatos realmente os dados precisam ser refinados. Vale considerar, ainda, que a reta de regressão se ajusta em 65,70%, ou seja, também há os valores que estão fora deste ajuste.

5.2 Análises Finais

Dessa vez, ampliou-se o leque de variáveis e foram selecionadas 44 que podem estar ligadas diretamente com a ocorrência de projeção. Fez-se a ACP para obter um número próximo ao ideal, sem perdas de informação, para identificação de quantas variáveis seguiriam para análise de acordo com a técnica seguinte. Obtiveram-se os resultados mostrados na Tabela 5.6.

Tabela 5.6 – Resultado gerado no *Statistica* com os números de autovalores, variância total, autovalor cumulativo e porcentagem cumulativa de variância para 19 CP.

Número	Autovalor	Variância Total (%)	Autovalor Cumulativo	Cumulativo (%)
1	4,101394	9,321350	4,10139	9,3213
2	3,441826	7,822333	7,54322	17,1437
3	2,625222	5,966414	10,16844	23,1101
4	2,340119	5,318452	12,50856	28,4285
5	1,795515	4,080716	14,30408	32,5093
6	1,624863	3,692871	15,92894	36,2021
7	1,552256	3,527855	17,48120	39,7300
8	1,433120	3,257092	18,91432	42,9871
9	1,393211	3,166389	20,30753	46,1535
10	1,294330	2,941659	21,60186	49,0951
11	1,224133	2,782119	22,82599	51,8772
12	1,171149	2,661703	23,99714	54,5390
13	1,133366	2,575831	25,13050	57,1148
14	1,089539	2,476224	26,22004	59,5910
15	1,048932	2,383937	27,26898	61,9749
16	1,027263	2,334689	28,29624	64,3096
17	0,967764	2,199465	29,26400	66,5091
18	0,952493	2,164757	30,21650	68,6739
19	0,914109	2,077521	31,13061	70,7514
20	0,903253	2,052847	32,03386	72,8042
21	0,848763	1,929007	32,88262	74,7332
22	0,836334	1,900760	33,71896	76,6340
23	0,813395	1,848624	34,53235	78,4826
24	0,788698	1,792495	35,32105	80,2751
25	0,776089	1,763839	36,09714	82,0389
26	0,742004	1,686373	36,83914	83,7253

27	0,690803	1,570007	37,52994	85,2953
28	0,676942	1,538505	38,20689	86,8338
29	0,655109	1,488883	38,86199	88,3227
30	0,628618	1,428677	39,49061	89,7514
31	0,603662	1,371960	40,09428	91,1234
32	0,561424	1,275963	40,65570	92,3993
33	0,531938	1,208949	41,18764	93,6083
34	0,460483	1,046551	41,64812	94,6548
35	0,423965	0,963557	42,07208	95,6184
36	0,392746	0,892604	42,46483	96,5110
37	0,338013	0,768212	42,80284	97,2792
38	0,293628	0,667336	43,09647	97,9465
39	0,290061	0,659230	43,38653	98,6058
40	0,248912	0,565708	43,63544	99,1715
41	0,196325	0,446192	43,83177	99,6177
42	0,093286	0,212013	43,92505	99,8297
43	0,073102	0,166141	43,99816	99,9958
44	0,001844	0,004190	44,00000	100,0000

Deste modo, para as 44 variáveis, o software gerou 44 componentes principais (CP), com 44 variáveis mais explicativas para cada uma; contudo, nem todas foram consideradas, pois o objetivo foi reduzir o número de variáveis. Para análises estatísticas, como citado anteriormente, consideram-se, geralmente, as CP que possuam porcentagem cumulativa de variância maior que 70%. Logo, foram selecionadas as 19 primeiras, destacadas de vermelho.

5.2.1 Análise de Trilha no Software *Genes*

Antes da análise de trilha ser feita de fato, foi realizada o diagnóstico de multicolinearidade com a utilização do software *Rbio*. Logo, o programa informou que o valor de VIF, era menor que 10, ou seja, não existia, ou a multicolinearidade era fraca.

Então, para a análise de trilha sem colinearidade, todas as 44 variáveis foram consideradas, e logo, a partir do resultado do software *Genes*, foram selecionadas 19 variáveis, como indicado pela ACP, que mais influenciam a variável principal projeção, como demonstrado na Tabela 5.7.

Tabela 5.7 – Variáveis em ordem de maior influência nas projeções

	Número da Variável	Nome
1	20	Peso de briquete misto adicionado durante o sopro (t)
2	34	Análise do Si no gusa carregado (%)
3	15	Peso de dunito adicionado durante o sopro (t)
4	3	Peso de sucata de gusa moldado (t)
5	41	Teor de P analisado na amostra de final de sopro (%)
6	25	Vida do convertedor (número de corridas)
7	42	Teor de S analisado na amostra de final de sopro (%)
8	29	Peso total de gusa sólido carregado (t)
9	1	Peso de sucata de gusa lingotado (t)
10	26	Tempo de sopro de O ₂ da corrida (min)
11	35	Análise real do Mn no gusa líquido carregado na corrida (%)
12	14	Peso de escória adicionado durante o sopro de O ₂ (t)
13	40	Teor de Mn analisado na amostra de final de sopro (%)
14	8	Peso de minério adicionado durante o sopro de O ₂ (t)
15	43	Teor de O ₂ dissolvido no aço líquido medido no final do sopro (ppm)
16	16	Peso de sucata F1 adicionado durante o sopro de O ₂ (t)
17	19	Peso de sinter adicionado durante o sopro de O ₂ (t)
18	22	Peso minério adicionado no início do sopro (t)
19	23	Volume total de O ₂ no sopro (Nm ³)

Deste modo, a partir das análises multivariadas de dados foi possível identificar 19 variáveis que mais apresentam influência nas projeções. É possível perceber que três das quatro variáveis encontradas nas análises do primeiro grupo estão presentes nas respostas das variáveis encontradas no segundo grupo. Esses resultados são consequências de análises estatísticas, mas que de certa forma, conectam-se as explicações metalúrgicas, uma vez que são as variáveis comumente citadas na literatura como possíveis causas de projeção.

É possível destacar, por exemplo, o aparecimento da variável 20 (peso de briquete misto adicionado durante o sopro). Percebeu-se que em todas as vezes que se adicionou esse tipo de briquete (mistura de lama de aciaria com carepa de laminação) durante o sopro, houve projeção. O efeito das adições de refrigerantes, que é o caso do briquete, do minério e do sínter, é endotérmico, o que favorece a diminuição da temperatura, e, portanto, o aumento da viscosidade da emulsão, provocando o seu crescimento, o que facilita a ocorrência de projeção.

Em relação ao silício, argumenta-se que sua presença no gusa líquido está relacionada a projeção de acordo com sua quantidade. Quanto maior o percentual deste elemento, maior o volume de escória formada e de emulsão.

O excesso de adição de fundentes como dunito pode contribuir para o acréscimo da viscosidade e para inibir a liberação dos gases das reações, o que também auxilia no crescimento da escória.

Sobre as sucatas, vale a observação sobre a sua composição. Quanto maior o nível de impurezas, maior a geração de óxidos formados na escória e maior é a propensão à projeção devido ao aumento do volume da escória.

O teor de fósforo analisado na amostra final de sopro influencia de acordo com a necessidade de mais agentes desfosforantes, para aumentar a eficiência da eliminação deste elemento. A maior adição de cal, por exemplo, pode implicar em acréscimo da viscosidade aparente devido à presença de sólidos dispersos durante o período de dissolução e/ou maior volume de escória, que inibem a liberação dos gases das reações, o que também auxilia no maior Índice de Formação de Espuma. Assim, quanto menor o teor final de P na amostra final, maior teria sido o efeito da desfosforação, o que, provavelmente, gera um aumento da formação da emulsão gás-metal-escória durante o sopro, com maiores riscos a projeção. Os teores de enxofre, manganês e oxigênio também estão envolvidos nessa conjuntura.

A vida do convertedor está diretamente ligada à quantidade de massa refratária contida em seu interior. Quanto mais novo o convertedor, menor o volume interno do BOF, reflexo do menor desgaste do revestimento refratário. Dessa forma, quanto menor o tempo de vida de um convertedor, maiores são as chances de projeções.

O peso total de gusa sólido adicionado está ligado ao volume de material acrescentado, que deve estar de acordo com a capacidade ideal do BOF.

O tempo de sopro de O_2 da corrida e o volume total de O_2 no sopro estão relacionados ao momento em que o operador detecta a possibilidade de projeção, já que a vazão de oxigênio é reduzida (seja manualmente ou automaticamente); o que implica em tempos de sopro mais longos para as mesmas necessidades de oxidação; o que denota que o padrão de sopro original era inadequado.

O peso de escória adicionado é um parâmetro importante a ser levado em consideração. A adição dessa escória contribui para a formação da escória inicial, porém pode acarretar, em seguida, em um volume excessivo de escória, e gerar a projeção.

5.2.2 Análise de Rede de Correlações no Software *RBio*

Como a análise de rede de correlação é basicamente visual, é possível verificar, a partir da Figura 5.2 que a variável projeção está mais bem conectada principalmente com as variáveis 20, 34, 15, (identificadas na Tabela 5.7), pois estão mais próximas e as linhas são mais espessas, indicando maior correlação, o que confirma a análise de trilha.

Além disso, é viável constatar que a variável 25 (vida do convertedor) possui correlação negativa em relação à projeção, uma vez que a linha que faz a ligação entre ambas é vermelha. Nos casos em que as linhas são verdes, a correlação entre as variáveis são positivas.

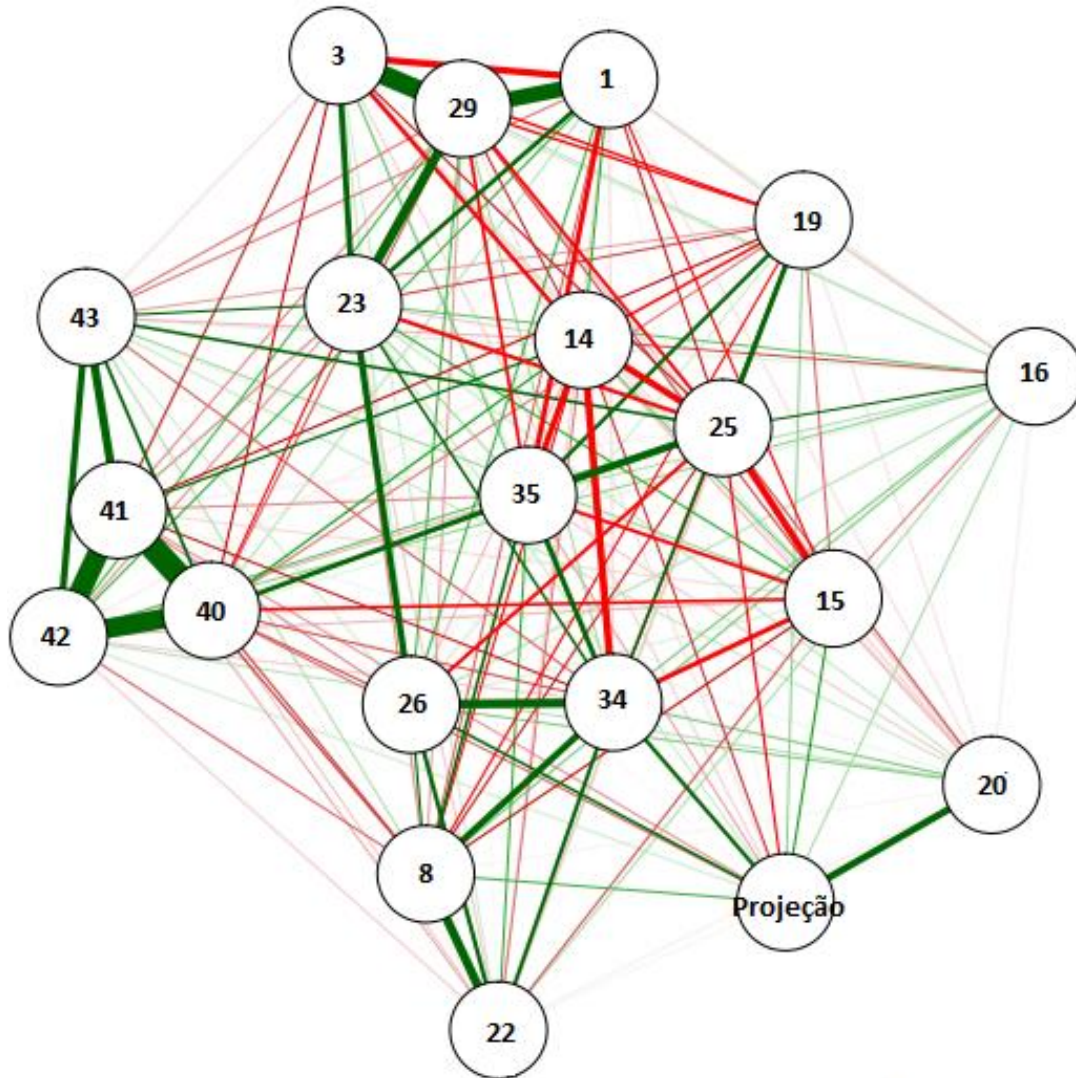


Figura 5.2 – Rede de correlações gerada no Rbio das 19 variáveis com a variável projeção. As linhas mais espessas significam maior grau de correlação entre as variáveis. As linhas verdes afirmam correlação positiva, enquanto as vermelhas, negativas.

5.4 Redes Neurais Artificiais em Python no Software *Spyder*

Selecionando, portanto, as 19 variáveis mais bem correlacionadas à variável projeção, realizaram-se alguns treinos da RNA. Dessa forma, foram obtidos os resultados conforme a Tabela 5.8, utilizando as possibilidades de respostas de 1 a 4 citadas no item 4.2.

Tabela 5.8 – Comparação dos erros dos diferentes testes da RN.

Opção de resposta (1, 2, 3, 4)	Erro a partir da RN
1, 2, 3, 4	27 %
1, 2, 3	12 %
1, 2 (com normalização)	4 %
1, 2 (sem normalização)	20 %
1, 2 (sem análise multivariada prévia, com 44 variáveis)	22 %

De acordo com a Tabela 5.8, é possível verificar que a análise com o melhor resultado obteve um erro de 4% e foi realizada com menor número de corridas, pois incluía apenas com as respostas 1 e 2 (1 – o operador informou que houve projeção e pressionou o botão; 2 – o operador não informou que houve projeção e não pressionou o botão), e com os dados normalizados. Esse resultado indica que realmente os dados requerem a utilização de um filtro prévio para que se eliminem dados (corridas de aço) menos precisos. Além disso, comprova que a normalização é importante e que beneficia o resultado final, pois minimiza as dificuldades oriundas do uso de unidades e dispersões distintas entre as variáveis, uma vez que todos os valores são transformados em números entre 0 e 1. A análise multivariada também se mostrou fortemente necessária, pois elimina as variáveis menos correlacionadas com a variável de interesse, o que permite treinos de RN mais apurados, e consequentemente, com menores erros.

Depois que o erro de 4,27% foi encontrado ao fim da 20000^o iteração, como demonstrado na Tabela 5.9, para que houvesse um teste mais preciso, foi aplicada a técnica *k-fold*, com 6 neurônios na camada escondida, o valor de k igual a 8, e 20000 iterações. Dessa forma, o software foi capaz de mostrar, com teste de precisão, os resultados mostrados na Tabela 5.10.

Tabela 5.9 – Valores dos erros das últimas 5 iterações da RN.

Número da Iteração	Erro
19996	5,03%
19997	4,44%
19998	4,67%
19999	4,65%
20000	4,27%

Tabela 5.10 – Resultado do teste de precisão informado pela RN com $k = 8$.

Precisão	Valor em %
0,95704698	95,70%
0,95838926	95,84%
0,94630872	94,63%
0,95033557	95,03%
0,94489247	94,49%
0,94892473	94,89%
0,94354839	94,35%
0,95833333	95,83%
Média	Desvio Padrão
0,950972431	$\pm 0,006147282$

Com média de precisão de 0,951 e desvio padrão de $\pm 0,00615$, tem-se o valor de confiabilidade de 0,96. Portanto, o valor inicial encontrado de 4,27% de erro, antes mesmo da validação, está dentro da faixa da última análise, assim como a validação do modelo com as outras 425 corridas, que apresentou, após as análises, 4,47% de erro nas respostas.

O valor do erro de 4,27% implica que a confiabilidade do modelo é de que a cada 100 corridas de aço, o modelo é capaz de prever corretamente se haverá ou não projeção em cerca de 96 corridas, errando apenas em 4. Estes números são ótimos para dados industriais, ou seja, todas essas análises são plausíveis de aplicação nessa área, comprovadamente. A minimização dos impactos ambientais, bem como dos riscos de acidentes, das perdas metálicas e da produtividade.

Quando se trata de treinos de redes neurais, entretanto, o resultado poderia ser melhor, uma vez que RN é uma técnica que comumente gera modelos com erros próximos a 0%. Além disso, quando se analisa a porcentagem de acertos realizados separadamente apenas nas corridas com projeção, isto é, àquelas corridas que a resposta era 1, e a RN de fato mostrou 1, o erro aumenta para 28%. Essa diferença está vinculada aos dados iniciais utilizados, que mesmo com o tratamento, ainda apresentam muitas incertezas. Com a utilização de dados coletados com mais acurácia e exatamente para este fim, certamente os resultados de confiabilidade seriam mais próximos a 100%.

6 CONCLUSÕES

As causas da anomalia projeção em um BOF foram identificadas pelo uso estatístico de análises multivariadas, regressão, e um modelo de predição por meio de redes neurais em *Python* foi construído. Os resultados podem ser resumidos como a seguir:

- Os resultados das análises do primeiro grupo geraram uma regressão com um modelo matemático para a previsão de corridas com projeção. O primeiro teste feito foi a partir da Análise de *Cluster*, Análise de Fatores e conseguiram ser confirmadas pela ACP. A regressão linear múltipla demonstrou estar melhor ajustada uma vez que apresenta um valor de r^2 de 0,6570, valor melhor se comparado com a regressão por componentes principais.
- Ao se validar o modelo matemático por regressão encontrado, percebeu-se que o modelo não alcançou bons resultados, ou seja, os dados precisavam ser mais bem filtrados.
- No segundo grupo, após um refino maior de dados, com a análise de componentes principais de 44 variáveis, foi possível determinar o número de 19 variáveis que mais se conectam com as causas da projeção. A análise de trilha identificou quais eram essas variáveis em ordem de maior efeito em porcentagem, e assim, a análise de rede de correlações confirmou:
 - Peso de briquete misto adicionado durante o sopro;
 - Análise do Si no gusa carregado (%);
 - Peso de dunito adicionado durante o sopro;
 - Peso de sucata de gusa moldado (t);
 - Teor de P analisado na amostra de final de sopro (%);
 - Vida do convertedor (número de corridas);
 - Teor de S analisado na amostra de final de sopro (%);
 - Peso total de gusa sólido carregado (t);
 - Peso de sucata de gusa lingotado (t);
 - Tempo de sopro de O₂ da corrida (min);
 - Análise real do Mn no gusa líquido carregado na corrida (%);
 - Peso de escória adicionado durante o sopro de O₂ (t);
 - Teor de Mn analisado na amostra de final de sopro (%);

- Peso de minério adicionado durante o sopro de O_2 (t);
 - Teor de O_2 dissolvido no aço líquido medido no final do sopro (ppm);
 - Peso de sucata F1 adicionado durante o sopro de O_2 (t);
 - Peso de sínter adicionado durante o sopro de O_2 (t);
 - Peso minério no início do sopro (t);
 - Volume total de O_2 no sopro (Nm^3);
- A RN concluiu um modelo matemático de predição de projeção com confiabilidade de 96%, o que indica um resultado importante para dados industriais, de modo que com a aplicação eficiente deste modelo, é possível haver a minimização dos impactos ambientais, dos riscos de acidentes, das perdas metálicas e financeiras.
 - As análises multivariadas possuem significados estatísticos, enquanto as RNA formam modelos matemáticos capazes de aprender pelo exemplo. Ainda assim, as variáveis respostas dessas análises também têm sentido metalúrgico, uma vez que estão de acordo com as causas de projeção citadas na literatura.
 - Os erros apresentados pela RN, embora pequenos, ainda são significativamente maiores que 0%; o que provavelmente se deve a incertezas remanescentes nos dados iniciais, mesmo após filtragem destes. Com a utilização de dados coletados exatamente para este fim, espera-se alcançar uma confiabilidade maior.

7 SUGESTÕES PARA TRABALHOS FUTUROS

- Estudar informações de outros convertedores e dados com maior acurácia especificamente para aplicação em análise multivariada de dados e redes neurais e comparar os resultados;
- Analisar dados de várias corridas para cálculos do Índice de Formação de Espuma, verificar a altura máxima ideal da emulsão durante o sopro para que não haja projeção, e fazer a simulação dos resultados para verificação;
- Fazer a aplicação do modelo de RN de predição de projeção na indústria e verificar as melhorias obtidas quanto à minimização dos impactos ambientais, dos riscos de acidentes, e das perdas de produtividade.

8 REFERÊNCIAS BIBLIOGRÁFICAS

ALBANI, C. B.; FURTADO, H. S.; DE OLIVEIRA, J. R. Estudos das causas de projeção de escória no convertedor LD, São Paulo, 2017. 289-298.

BIRK, W. *et al.* **Foam Level Control in a Water Model of the LD Converter Process.** 10th IFAC Symposium on Automation in Mining, Mineral and Metal Processing. Tóquio: IFAC. 2001. p. 273-278.

BRÄMMING, M. **Tese. Avoiding Slopping in Top-Blown BOS Vessels.** Lulea: Luleå University of Technology, 2010.

BRYANT, F. B.; YARNOLD, P. R. Principal Components Analysis and Exploratory and Confirmatory Factor Analysis. **American Psychological Association**, Washington DC, 2019.

CICUTTI, C. *et al.* Analysis of Slag Foaming During the Operation of an Industrial Converter. **Latin American Applied Research**, Buenos Aires, p. 237/240, 2001. ISSN 32:237-240.

COMPARINI, I. P.; WECHSLER, S. M.; MACHADO, W. D. L. Indicadores emocionais no desenho da figura humana: investigando evidências de validade. **Psicologia: Teoria e Prática**, São Paulo, v. 19, p. 256-269, Dezembro 2017. ISSN 1980-6906.

CORLETA SCHREIBER, J. N. *et al.* **Técnicas de validação de dados para sistemas inteligentes: uma abordagem do software Sdbayes.** Universidade Federal de Santa Catarina e Universidade Nacional de Mar del Plata. Mar del Plata. 2017. (ISBN: 978-85-68618-03-5).

CORRAR, L. J.; PAULO, E.; FILHO, J. M. D. **Análise multivariada para os cursos de administração, ciências contábeis e economia.** São Paulo: Atlas, 2012.

DATA SCIENCE ACADEMY. As 10 Principais Arquiteturas de Redes Neurais. **Deep Learning Book**, 2019. Disponível em: <<http://deeplearningbook.com.br/as-10-principais-arquiteturas-de-redes-neurais/>>. Acesso em: 27 dezembro 2019.

DAVID, S. ; SILVA, M. B.; MACHADO, M. L. P. Artificial Neural Networks Model to Predict the Sulfur Percentage at the Hot Metal in the Blast Furnace, Vitória, 2014.

DEO, B. et al. Control of Slag Formation, Foaming, Slopping, and Chaos in BOF. **Indian Institute of Metals - Springer**, p. 543–554, Dezembro 2013. ISSN DOI 10.1007/s12666-013-0306-2.

DUNTEMAN, G. H. **Principal Components Analysis**. 69. ed. Califórnia: Sage, 1987.

EVESTEDT, M. *et al.* Slopping Warning System for the LD Converter Process - An Extended Evaluation Study. **Lulea University of Technology**, Lulea, 2007.

FERNEDA, E. Redes neurais e sua aplicação em sistemas de recuperação de informação. **Ciência da Informação**, Brasília, v. 35, n. 1, p. 25-30, Abril 2006.

FIORIN, D. V. *et al.* Aplicações de redes neurais e previsões de disponibilidade de recursos energéticos solares. **Revista Brasileira de Ensino de Física**, São Paulo, v. 33, março 2011.

GONÇALVES, H. F. P.; CARNEIRO, C. A. R.; ARAÚJO, T. C. D. S. Estudo microestrutural da escória de convertedor LD e sua relação com o processo de desfosforação. **Tecnologia em Metalurgia, Materiais e Mineração**, São Paulo, Janeiro 2014. 41-46.

HAIR JR., J. F. *et al.* **Análise multivariada de dados**. 6. ed. Porto Alegre: Bookman, 2009.

ILUMINEO. Como funciona a análise de Cluster. **Ilumeo**, 2019. Disponível em: <<https://www.ilumeo.com.br/todos-posts/2019/04/03/como-funciona-a-analise-de-cluster>>. Acesso em: 08 janeiro 2020.

INSTITUTO AÇO BRASIL. Estatística Preliminar. **Instituto Aço Brasil**, 2019. Disponível em: <<http://www.acobrasil.org.br/site2015/estatisticas.asp>>. Acesso em: 05 Fevereiro 2019.

INSTITUTO AÇO BRASIL. Dados do setor. **Instituto Aço Brasil**, 2018. Disponível em: <<http://www.acobrasil.org.br/site2015/dados.asp>>. Acesso em: 05 Fevereiro 2019.

ITO, K.; FRUEHAN, R. Study on the Foaming of CaO-SiO₂-FeO Slags: Part I. Foaming Parameters and Experimental Results. **Metallurgical and Materials Transactions - Springer**, Pittsburgh, v. 20B, p. 509-514, Agosto 1989.

ITO, K.; FRUEHAN, R. J. Study on the foaming of CaO-SiO₂-FeO slags: Part II. Dimensional analysis and foaming in iron and steelmaking processes. **Metallurgical Transactions B**, v. 20, p. 515-521, 1989.

JUNG, S.-M.; FRUEHAN, R. J. Foaming Characteristics of BOF Slags. **ISIJ International**, Pittsburgo, v. 40, n. 2, p. 348–355, 2000.

JÚNIOR, M. A. V. **Modelo híbrido associando cálculos termodinâmicos e redes neurais artificiais para a predição da temperatura do aço desde o fim de sopro até a metalurgia secundária**. Ouro Preto: REDEMAT/UFOP, 2018.

KIT-LEARN DEVELOPERS BSD LICENSE. Cross-validation: evaluating estimator performance. **Scikit-Learn**, 2019. Disponível em: <https://scikit-learn.org/stable/modules/cross_validation.html>. Acesso em: 27 dezembro 2019.

KOBAYASHI, S. *et al.* Prediction and Control of Slag Slopping in BOF Using Microwave Gauge. **IFAC Automation in Mining, Mineral and Metal Processing**, Heisinki, p. 297-301, 1983.

KOSTÚR, K.; FUTÓ, J. The Prediction of Metal Slopping in LD Converter on Base an Acoustic Signal. **Journal Metalurgija**, 30 setembro 2006. 97-101.

LAHIRI, A. K.; PAL, S. **Foaminess of slag: cause and control**. VII International Conference on Molten Slags Fluxes and Salts. Cidade do Cabo: The South African Institute of Mining and Metallurgy. 2004. p. 473-477.

LIM, E.; KIM, H.; SON, B. Slopping Prediction Using Analysis of Behavior in BOF. **IFAC Automation in Mining, Mineral and Metal Processing**, Tokyo, p. 29-33, 2001.

MILETIC, I. *et al.* Experiences in Applying Data-Driven Modelling Technology to Steelmaking Processes. **The Canadian Journal of Chemical Engineering**, Hamilton, v. 86, p. 937–946, Outubro 2008. ISSN DOI 10.1002/cjce.20090.

MONTEIRO, G. Estudo de Regressão. **Rpubs**, 2018. Disponível em: <<https://rpubs.com/gabrieloqm/regressaogodqm1> >. Acesso em: 09 janeiro 2020.

NI, P.; HAGLUND, T.; ERSSON, M. Study on Slopping Prevention in the BOF Steelmaking Process. **Steel Research International**, p. 1-7, Outubro 2016. ISSN DOI: 10.1002/srin.201600399.

OLIVEIRA, B.; RIBEIRO, C. Análise de Cluster. **Oper**, 2019. Disponível em: <<https://operdata.com.br/blog/analise-de-cluster/>>. Acesso em: 08 janeiro 2020.

PORTAL ACTION. Análise de Regressão. **Portal Action**, 2019. Disponível em: <<http://www.portalaction.com.br/analise-de-regressao/regressao-linear-multipla>>. Acesso em: 09 janeiro 2020.

RAJESHA, N.; KHARE, M. R.; PABI, S. K. Feed Forward Neural Network for Prediction of end Blow Oxygen in LD Converter Steel Making. **Materials Research**, Kharagpur, v. 13, p. 15-19, Janeiro 2010.

RBIO, S. Rede de Correlacoes. In: ROSADO, R. **Rbio: A Tool For Biometric And Statistical Analysis Using The R Platform**. [S.l.]: [s.n.], v. 17, 2016. p. 187-190.

REMUS, R. *et al.* **Best Available Techniques (BAT) Reference Document for Iron**. 1. ed. Seville: Joint Research Centre, 2013.

RIZZO, E. M. D. S. **Introdução aos processos de refino primário dos aços nos convertedores a oxigênio**. São Paulo: ABM, 2006.

ROCHA, J. C.; MATOS, F. D.; FREI, F. Utilização de redes neurais artificiais para a determinação do número de refeições diárias de um restaurante universitário. **Revista de Nutrição**, Campinas, p. 735-742, outubro 2011.

SALLA, V. P. *et al.* Análise de trilha em caracteres de frutos de jabuticabeira. **Pesquisa Agropecuária Brasileira**, Brasília, Março 2015. 218-223.

SHAKIROV, M. *et al.* **Prediction and Prevention of Slopping in the BOF**. ISSTech 2003 Conference Proceedings. Indianápolis: Iron & Steel Society. 2003. p. 655-666.

SILVA, A. L. V. D. C. E. **Termodinâmica Aplicada a Aciaria**. Rio de Janeiro: UFF, 1998.

SMITH, L. I. **A tutorial on Principal Components Analysis**. Dunedin: University of Otago, 2002.

SOUZA, T. V. D. **Aspectos estatísticos da análise de trilha (Path Analysis) aplicada em experimentos agrícolas**. Lavras: UFLA, 2013.

SUMMA, R. F.; MACRINI, L. Os determinantes da inflação brasileira recente: estimações utilizando redes neurais. **Nova economia**, Belo Horizonte, v. 24, n. 2, p. 279-295, Agosto 2014. ISSN ISSN 1980-5381.

TEODORO, P. E. *et al.* Interrelations between agronomic and technological fiber traits in upland cotton. **Acta Scientiarum**, Maringá, 2018.

TURKDOGAN, E. T. **Fundamentals of Steelmaking**. Londres: The Institute of Materials, 1996.

TURKDOGAN, E. T.; FRUEHAN, R. J. Fundamentals of Iron and Steelmaking. In: _____ **The Making, Shaping, and Treating of Steel**. 10. ed. Pittsburgo: Aist Steel Foundation, 1998. Cap. 2, p. 123.

VIALI, L. **Série: Estatística Multivariada**. Porto Alegre, p. 3-4. 2005.

VIRTUOUS TECNOLOGIA DA INFORMAÇÃO. Células nervosas. **Só Biologia**, 2008. Disponível em: <<https://www.sobiologia.com.br/conteudos/FisiologiaAnimal/nervoso2.php>>. Acesso em: 02 Abril 2019.

WALKER, D. I.; KEMENY, F. L.; JONES, J. A. T. **Vessel Slopping Detection**. AISTech 2005 Proceedings - Volume I. Pittsburgo: AIST. 2005. p. 711-720.

WRIGHT, S. Correlation and causation. **Journal of Agricultural**, Washington, D.C, 3 Janeiro 1921. 557-585.

YAMAZACK, W. Guia Completo de Python. **DevMedia**, 2018. Disponível em: <<https://www.devmedia.com.br/guia/python/37024>>. Acesso em: 26 Março 2019.

ZUBEN, F. J. V. Redes neurais multicamadas. In: _____ **Redes neurais aplicadas ao controle de máquina de indução**. Campinas: Unicamp, 1993. Cap. 2, p. 7-30.

ZUSHI, K.; MATSUZOE, N. Metabolic profile of organoleptic and health-promoting qualities in two tomato cultivars subjected to salt stress and their interactions using correlation network analysis. **Scientia Horticulturae**, Miyazaki, 07 Janeiro 2015.

9 ANEXO I – RESUMO DE ARTIGOS

Artigo publicado nos anais do Iron & Steel Technology Conference – AisTech 2019, Pitsburgo, EUA, páginas 943-949, DOI 10.1000.377.099.

USE OF MULTIVARIATE ANALYSIS DATA IN THE STUDY OF SLAG SLOPPING OF BOF

Abstract:

Slopping is a phenomenon that occurs in many steelmaking plants with BOF converters. During the oxygen blowing period the formation of the emulsion occurs which often increases in volume and if excessive the emulsion is expelled from the vessel. This occurrence generates some problems for the steelmaking process, such as financial, metallic material loss and environmental hazards. The purpose of this work is to identify the variables that can most influence the occurrence of slopping with the use of statistic tools. In addition, multiple regression analysis was also performed to build a mathematical model capable of forecasting slopping.

Keywords: BOF Converter; Big Data; Slopping; Multivariate Analysis.

Artigo publicado nos anais do 50º Seminário de Aciaria, Fundição e Metalurgia de Não-Ferrosos, São Paulo, parte integrante da ABM Week 2019, páginas 314-321, DOI 10.5151/2594-5300-33530.

IDENTIFICAÇÃO DAS VARIÁVEIS QUE INFLUENCIAM A PROJEÇÃO DE ESCÓRIA NO BOF POR MEIO DE ANÁLISE MULTIVARIADA DE DADOS

Resumo:

A projeção de escória do BOF (*Basic Oxygen Furnace*) ocorre a partir da formação da emulsão, a qual provoca expulsão de escória se o volume da mesma aumenta em demasia. Esse fenômeno gera algumas preocupações para uma aciaria, como diminuição da eficácia do refino, perdas financeiras, perda de material, riscos de acidentes e principalmente problemas ambientais. Arquivos de *Big Data* da ArcelorMittal Tubarão foram estudados para que fosse possível a identificação das causas principais das projeções a partir de estudos estatísticos de análises multivariadas, técnicas ainda pouco usuais para estudos na área de aciaria. Foram utilizados os softwares Statistica, Genes, e Rbio com novas propostas de utilização de ferramentas desta área, como a análise de componentes principais e a análise de trilha e análise de rede de correlações. Foi possível verificar, por exemplo, que o peso de briquete misto adicionado durante o sopro e o teor de silício no gusa carregado estão entre as principais causas de projeção.

Palavras-chave: Projeção, BOF; Big Data; Análise Multivariada.

Artigo publicado nos anais do 2019 Sustainable Industrial Processing Summit and Exhibition, Chipre.

HIERARCHING BOF SLOPPING VARIABLES FROM BIG DATA ANALYSIS

Abstract:

Slopping is a phenomenon which is observed when the volume the emulsion inside of BOF (Basic Oxygen Furnace) is excessive and a fraction of slag and metal is then expelled. This phenomenon concerns steelmakers since it leads to material losses, health hazards, reduction of refining efficiency and, mainly, environmental issues. Big Data files from a steelmaking shop in Brazil have been analysed in order to identify the central causes of the slopping. Statistical techniques including Multivariate Analyzes have been employed in order to identify the main variables that affect slopping. Statistica, Genes and Rbio software were used to do so; also principal components, path analysis, and correlation network were chosen as tools. It was possible to identify and hierarchize the variables most affecting slopping in good agreement with literature. The resulting variables can be used to generate a model to anticipate slopping.

Keywords: **Erro! Indicador não definido.** Slopping; Emulsion; BOF; Multivariate Analyzes.

Artigo aceito pela revista Tecnologia em Metalurgia, Materiais e Mineração, na área de Engenharia de Processos.

ESTUDO DE PROJEÇÕES DE UM BOF UTILIZANDO ANÁLISE MULTIVARIADA DE DADOS

Resumo:

Aprojeção em um BOF é uma anomalia que representa desperdício de matéria prima, risco de acidentes, prejuízo financeiro e, principalmente, traz consequências ambientais. Devido à relevância deste fato foram realizados acompanhamentos *in loco* na ArcelorMittal Tubarão (AMT), e arquivos de *Big Data* advindos da operação de sopro de oxigênio dos convertedores foram analisados com o intuito de investigar as principais variáveis associadas ao fenômeno de projeção. Realizou-se o tratamento descritivo, por meio das técnicas de Análise de *Cluster*, Análise de Fatores, e de Componentes Principais com a finalidade de elaborar uma equação de regressão linear multivariada que pudesse modelar e prever as corridas com projeções. O tratamento estatístico foi feito com a utilização do software *Statistica* e do *NCSS Statistical Software*. A análise conclui que as variáveis teor de silício no gusa líquido, peso de minério adicionado, peso de cal calcítica e tempo de sopro de oxigênio estão entre as principais causas de projeção.

Palavras-chave: Projeção de escória; BOF; Análise Multivariada; Big Data.

10 ANEXO II – VARIÁVEIS DO PRIMEIRO E DO SEGUNDO GRUPO DE ANÁLISES

As variáveis de análise do primeiro grupo são:

1. Peso de sucata de gusa lingotado (t).
2. Peso de sucata de aço tipo C (t).
3. Peso de sucata de gusa moldado (t).
4. Peso sucata de gusa desclassificado(t).
5. Peso sucata de aço tipo A (t).
6. Peso sucata de gusa tipo b.
7. Peso sucata de aço comprada tipo a (t).
8. Peso de minério adicionado durante o sopro de O₂ (t).
9. Peso de cal calcítica adicionado durante o sopro de O₂ (t).
10. Peso de cal dolomítica adicionado durante o sopro de O₂ (t).
11. Peso de dolomita crua adicionado durante o sopro de O₂ (t).
12. Peso de fluorita crua adicionado durante o sopro de O₂ (t).
13. Peso de SIC crua adicionado durante o sopro de O₂ (t).
14. Peso de escória adicionado durante o sopro de O₂ (t).
15. Peso de dunito adicionado durante o sopro de O₂ (t).
16. Peso de sucata F1 adicionado durante o sopro de O₂ (t).
17. Peso de briquete adicionado durante o sopro de O₂ (t).
18. Peso de briquete refrigerante adicionado durante o sopro de O₂ (t).
19. Peso de sínter adicionado durante o sopro de O₂ (t).
20. Peso de briquete misto adicionado durante o sopro de O₂ (t).
21. Peso dolomita crua para proteção do refratário (t).
22. Peso minério no início do sopro (t).
23. Volume total de O₂ no sopro (Nm³).
24. Vida do bico da lança (número de corridas).
25. Código de intensidade de projeção (0 = sem; 1 = média; 2 = forte; 3 = muito forte).
26. Código de intensidade de projeção (0 = sem; 1 = média; 2 = forte; 3 = muito forte).

27. Vida do convertedor (número de corridas/campanha).
28. Tempo de carregamento de gusa e sucata da corrida (min).
29. Tempo entre o final descarregamento e o início do sopro da corrida (min).
30. Tempo de sopro de O₂ da corrida (min).
31. Vida do furo de corrida (furo de vazamento – número de corridas).
32. Peso de gusa líquido carregado (t).
33. Temperatura do gusa líquido carregado, medido na liberação do KR (°C).
34. Peso de aço líquido estimado (t/batelada).
35. Análise real do C no gusa líquido carregado na corrida (%).
36. Análise real do Si no gusa líquido carregado na corrida (%).
37. Análise real do Mn no gusa líquido carregado na corrida (%).
38. Análise real do P no gusa líquido carregado na corrida (%).
39. Análise real do S no gusa líquido carregado na corrida (%).

As variáveis de análise do segundo grupo são:

1. Peso de sucata de gusa lingotado (t).
2. Peso de sucata de aço tipo C (t).
3. Peso de sucata de gusa moldado (t).
4. Peso sucata de gusa desclassificado (t).
5. Peso sucata de aço tipo A (t).
6. Peso sucata de gusa tipo b.
7. Peso sucata de aço comprada tipo a (t).
8. Peso de minério adicionado durante o sopro de O₂ (t).
9. Peso de cal calcítica adicionado durante o sopro de O₂ (t).
10. Peso de cal dolomítica adicionado durante o sopro de O₂ (t).
11. Peso de dolomita crua adicionado durante o sopro de O₂ (t).
12. Peso de fluorita crua adicionado durante o sopro de O₂ (t).
13. Peso de SiC crua adicionado durante o sopro de O₂ (t).
14. Peso de escória adicionado durante o sopro de O₂ (t).

15. Peso de dunito adicionado durante o sopro de O₂ (t).
16. Peso de sucata F1 adicionado durante o sopro de O₂ (t).
17. Peso de briquete adicionado durante o sopro de O₂ (t).
18. Peso de briquete refrigerante adicionado durante o sopro de O₂ (t).
19. Peso de sinter adicionado durante o sopro de O₂ (t).
20. Peso de briquete misto adicionado durante o sopro de O₂ (t).
21. Peso previsto de cales a ser adicionado antes de iniciar o sopro (t).
22. Peso minério no início do sopro (t).
23. Volume total de O₂ no sopro (Nm³).
24. Vida do bico da lança (numero de corridas).
25. Vida do convertedor (numero de corridas/campanha).
26. Tempo de sopro de O₂ da corrida (min).
27. Peso de gusa líquido carregado (t).
28. Temperatura do gusa líquido carregado, medido na liberação do KR (°C).
29. Peso total de gusa sólido carregado (t).
30. Temperatura real de final de sopro (°C).
31. Basicidade real da escória amostrada no final do sopro.
32. Percentagem de C estimado pelo sensor de oxigênio no final do sopro (%).
33. Análise real do C no gusa líquido carregado na corrida (%).
34. Análise real do Si no gusa líquido carregado na corrida (%).
35. Análise real do Mn no gusa líquido carregado na corrida (%).
36. Análise real do P no gusa líquido carregado na corrida (%).
37. Análise real do S no gusa líquido carregado na corrida (%).
38. Teor de C analisado na amostra de final de sopro (%).
39. Teor de Si analisado na amostra de final de sopro (%).
40. Teor de Mn analisado na amostra de final de sopro (%).
41. Teor de P analisado na amostra de final de sopro (%).
42. Teor de S analisado na amostra de final de sopro (%).
43. Teor de oxigênio dissolvido no aço líquido medido no final do sopro (ppm).
44. Projeção.