

# Anode Temperature Classification of Liquid Metal in an Electric Arc Furnace using K-Nearest Neighbors

Tiago P. Ferreira<sup>1</sup>, Karla Figueiredo<sup>2</sup>, Marley M. B. R. Vellasco<sup>1</sup>

<sup>1</sup>Departamento de Engenharia Elétrica – Pontifícia Universidade Católica (PUC-Rio)  
Rio de Janeiro – RJ – Brasil

<sup>2</sup>Departamento de Ciências da Computação - UERJ – Universidade do Estado do Rio de Janeiro (UERJ) – Rio de Janeiro – RJ – Brasil

{tiagopf,marley}@ele.puc-rio.br, {karlafigueiredo}@ime.uerj.br

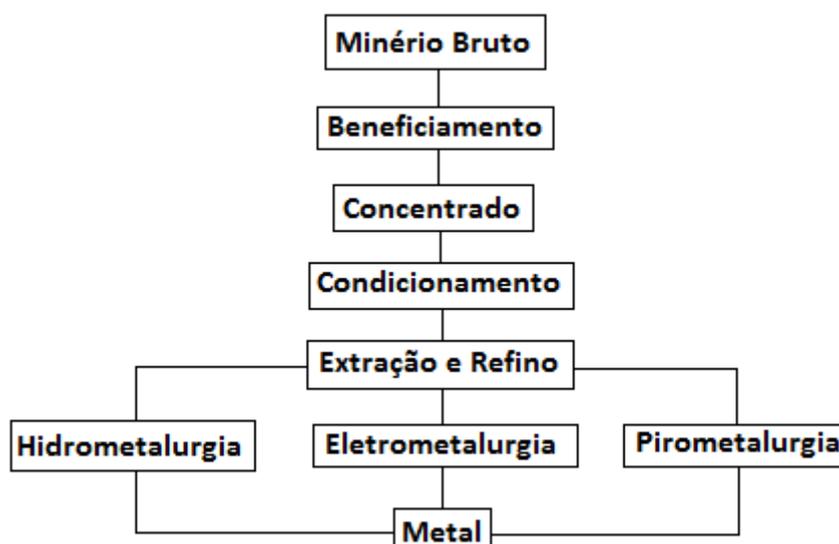
**Abstract.** *Metallurgical and steel industries adopt arc electric furnaces in their processes for the refining of the chemical composition of liquid metals. Therefore, it is necessary to control the anode temperature of these reactors, so that it does not get into excessively high or low ranges, which leads to the loss of the metallic bath or the early wear of the furnace refractory. Given a pilot database of attributes that involve this process, Data Mining and Machine Learning techniques were applied to preprocess the data and create models that solve the anode temperature classification problem at low, normal and high temperatures. The Machine Learning model used is the KNN (k-Nearest Neighbors) algorithm which, based on distance between test data point and training data points, classifies the output variable by majority vote. Three different approaches to creating KNN models are used and the results of each are presented and discussed.*

**Resumo.** *As indústrias metalúrgica e siderúrgica adotam, em seus processos, fornos elétricos a arco para o refino da composição química de metais líquidos. Para tanto, faz-se necessário controlar a temperatura de anodo destes reatores, de modo que a mesma não entre em faixas excessivamente elevadas ou baixas, o que leva a perda do banho metálico ou ao desgaste precoce do refratário dos fornos. Assim, esse trabalho apresenta técnicas de Mineração de Dados e Machine Learning que solucionam o problema de classificação da temperatura do anodo em temperaturas baixas, normais e altas. Os modelos de Aprendizado de Máquina se valeram do algoritmo KNN (k-Nearest Neighbors). Três diferentes abordagens são utilizadas para criação dos modelos KNN e os resultados de cada uma delas são apresentados e discutidos.*

## 1. Introdução

Diversas indústrias do ramo de metalurgia e siderurgia utilizam, em seus processos, equipamentos e subprocessos que visam refinar a composição química dos metais. Em alguns casos, tais equipamentos apresentam comportamentos de processos puramente físicos como, por exemplo, separadores magnéticos (que visam separar tipos de partículas de determinado metal por meio de um agitador eletromagnético, por exemplo), ou ainda processos de fundição [Da Costa e Silva 2013] onde, por meio de variações de temperatura, transformações físicas no metal podem ser efetuadas sem alterar suas propriedades e composições químicas.

Além disso, os processos químicos também são largamente empregados com o objetivo de alterar ou refinar a composição química da liga ou estrutura metálica (matéria-prima). Estes processos requerem, assim como os processos físicos, controles apurados de grande quantidade de variáveis. Entretanto, processos físicos normalmente requerem monitoramento e controle de variáveis físicas (como velocidade, fusão, nível, temperatura) e de equipamentos de processo, enquanto que processos químicos necessitam, muitas vezes, de monitoramento e controle de variáveis químicas (como composição química, concentração química, PH), físicas e de equipamentos de processo. A Figura 1 esquematiza os processos, de maneira geral, que caracterizam a indústria metalúrgica.



**Figura 1: Fluxograma geral de processos metalúrgicos. Fonte: [Mourão 2019], adaptado.**

A Figura 1 registra etapas concernentes a processos metalúrgicos que envolvem tanto metais leves (Alumínio, Manganês, Lítio, etc.), quanto nobres (Ouro, Prata, Platina, etc.), pesados (Cobre, Níquel, Zinco, etc.), secundários (Arsênio, Cobalto, Bismuto, etc.) e raros (Ítrio, Tungstênio, Nióbio, Índio, etc.) [Mourão 2019].

Este artigo aborda um subprocesso de refino de metais tipicamente ferrosos (siderurgia), pesados e leves (e, em alguns casos, nobres e raros): o forno elétrico a arco. Ressalva-se que versões similares à abordagem aqui descrita são adotadas na indústria siderúrgica, indústria com processos majoritariamente em bateladas [ABM 1978].

A temperatura do anodo, presente na parte inferior do reator, é controlada por meio da potência entregue através do arco. Se o reator, em seu funcionamento, operar em temperaturas muito altas, o anodo pode ser danificado, bem como a composição química do banho metálico pode ser afetada perdendo-se o produto presente no forno. Em reatores de fusão catalítica de alumina, por exemplo, o grau de superaquecimento da liga metálica reduzida acima de 2000°C se torna um problema [O'Brien & Ford 1998]. Portanto, o monitoramento e controle de temperatura visando evitar altas temperaturas devem ser feito de maneira minuciosa. Além disso, deve ser dada atenção também aos casos de temperaturas baixas de anodo, pois ainda que não seja um caso crítico tanto quanto altas temperaturas, valores excessivamente baixos de temperatura podem solidificar os metais presentes no banho ou, em casos de vazamento de metal e escória,

até mesmo dificultar o processamento da corrida seguinte (isto é, a próxima sequência de metais a serem inseridos no forno). Bergman & Kjellberg (2001) detalham o equacionamento matemático da potência demandada para fornos elétricos a arco similares baseados tanto em corrente contínua quanto em corrente alternada.

Ressalta-se ainda a importância da medição da temperatura do anodo. Sem o valor correto desta temperatura, torna-se complexo estabelecer o controle da mesma, já que é fundamental conhecer em qual faixa a temperatura do anodo se encontra (baixa, normal ou alta) para saber qual a ação a ser executada no reator no sentido de restabelecer a temperatura para valores dentro das faixas desejadas. Além disso, sabe-se que a *performance* global da fusão do conjunto metálico de fornos elétricos a arco é monitorada em estágios específicos do forno e é baseada, dentre outros fatores, nas medições de crescimento desta temperatura [Rogers & Cramer 1989].

O objetivo deste trabalho é encontrar um modelo, utilizando técnicas de Mineração de Dados e Aprendizado de Máquina, que classifique as temperaturas de anodo do forno elétrico baseado nas informações medidas por sensores e/ou calculadas pelo sistema do forno. Uma correta classificação da faixa da temperatura de anodo permitirá evitar que se percam corridas de banhos metálicos, reduzindo-se as perdas; bem como evitará desgastes prematuros de refratário do forno, evitando-se, portanto, manutenções corretivas antes do final da vida útil do equipamento.

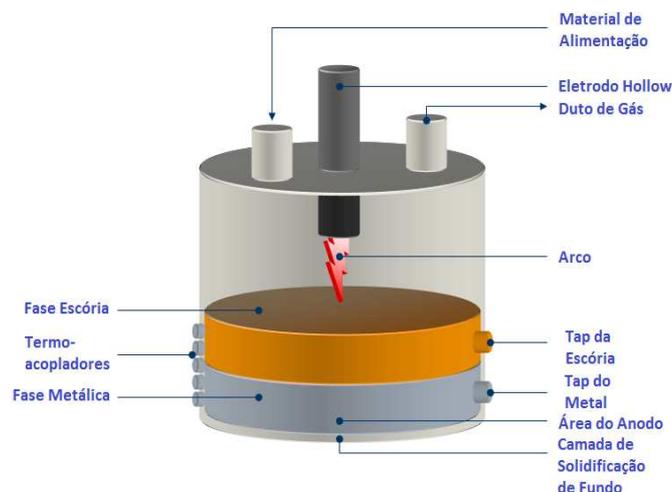
O restante do trabalho inclui uma seção com uma sucinta descrição dos fornos elétricos a arco. A seção III indica as técnicas de Mineração de Dados utilizadas para classificação da temperatura. A seção IV, por fim, pontua as conclusões deste trabalho e futuros trabalhos.

## 2. Forno Elétrico a Arco

A Figura 2 ilustra o processo do forno elétrico a arco (FEA). Trata-se, conforme se pode verificar na figura, de um forno/reator a altas temperaturas. Este reator é usado para separar metal de minério. As entradas do reator incluem o minério metálico a ser purificado (por meio de altas temperaturas provenientes do eletrodo), um agente oxidante, bem como agentes fluxantes como, por exemplo, o calcário. O metal líquido é produzido pela adição de redutores, calcário e minério em paralelo a presença de um arco elétrico. A camada de metal encontra-se em contato com o anodo [Jones 2011].

O arco formado pela passagem de uma alta corrente entre o anodo condutor (carbono) e a sobrecarga do reator através de um não condutor (ar) faz com que ele se ionize e, no processo, libere energia térmica. O arco é transferido do catodo (normalmente, um eletrodo de grafite) para o anodo por meio de um caminho de condução [O'Brien & Ford 1998]. Impurezas no metal são oxidadas, sendo transformadas em óxidos que, pelas suas naturezas, são menos densos que o metal e flutuam sobre o metal reduzido (formando a escória). O forno, portanto, é aquecido de maneira a derreter o metal, separando-o do minério, com escória e gases como coprodutos. O metal e a escória são vazados pelos seus respectivos bocais periodicamente, resfriando o reator.

O método de *Machine Learning* utilizado para efetuar este modelo de classificação é o KNN (*k-Nearest Neighbors*). Este método, assim como as etapas de pré-processamento de dados e a análise dos resultados encontrados baseados nos dados obtidos de uma aplicação piloto, será explicitado nas próximas seções.



**Figura 2: O processo de um forno elétrico a arco, presente nas indústrias siderúrgica e metalúrgica.**

Fernandez et al. (2008) implementaram uma estimativa da temperatura do *tap* de um FEA usando Redes Neuro-Fuzzy. Todavia, esta aplicação visava prever uma temperatura diferente daquela verificada neste trabalho. Em Blachnik et al. (2010), um modelo de previsão de temperatura de aço líquido em um FEA é realizado usando-se técnicas de Máquinas de Vetores de Suporte (SVM) e KNN. Entretanto, o autor se vale de um conjunto de variáveis e medições distintas daquelas expostas neste trabalho, bem como realiza uma inferência de temperatura do aço, e não uma classificação. Chen et al. (2018) fez uma modelagem de consumo de energia de um FEA usando uma técnica de *Deep Learning*, sendo que as temperaturas do forno foram consideradas entradas do processo e a saída era a potência entregue ao forno.

### 3. Etapas de Mineração de Dados

#### 3.1. Pré-Processamento

Uma base de dados contendo valores de atributos para um típico processo de forno elétrico a arco foi obtida. O seguinte conjunto de variáveis (atributos) foi identificado:

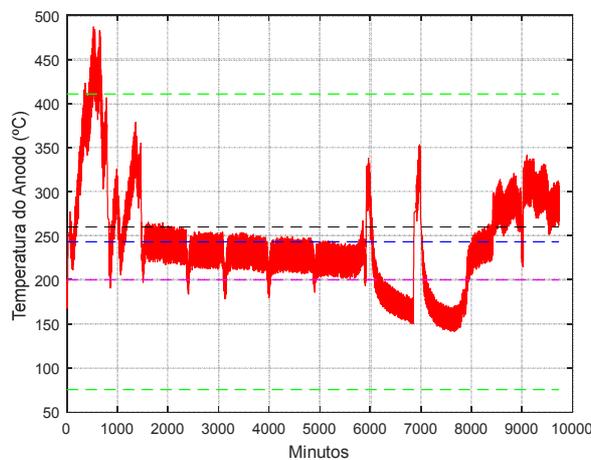
- **CrO/Cr<sub>2</sub>O<sub>3</sub>**: Taxa entre composição química de CrO e Cr<sub>2</sub>O<sub>3</sub> do refratário do forno.
- **CaO\_Flux**: Taxa de agente fluxante (Calcário) adicionado, que auxilia na separação Escória/Metal. Nas indústrias metalúrgicas são exigidos calcários puros para uso, em geral, como fluxantes [Sampaio & Almeida 2016].
- **MW**: Medida de entrada da potência elétrica, de um transformador elétrico.
- **Ore\_Feed**: A taxa de alimentação de minério metálico no reator.
- **Reductant**: Taxa de adição de redutor. Usada para reduzir os óxidos de metal, purificando, portanto, o metal.
- **CO**: Fluxo de Gás CO (oxidante).
- **H<sub>2</sub>**: Fluxo de Gás H<sub>2</sub>, coproduto das reações de redução.

- **CO<sub>2</sub>**: Fluxo de Gás CO<sub>2</sub>. Trata-se de um oxidante, bem como um agente formador de espuma no banho.
- **Bottom Freeze Layer**: Medida da camada de solidificação de fundo do reator. Protege o forno de metais aquecidos.
- **Metal\_Depth**: Profundidade da camada de metal.
- **Feed Power Product**: Produto entre a taxa de alimentação e a potência por tonelada de alimentação.
- **Slag\_Depth**: Profundidade da camada de escória. Afeta a temperatura, variando a condutividade.
- **Temp\_anode**: Medida de temperatura (que se deseja classificar) do anodo.

A temperatura do anodo foi classificada conforme especificação de processo da seguinte maneira:

- Temperatura < 200°C → Temperatura Baixa.
- 200°C ≤ Temperatura < 260°C → Temperatura Normal.
- Temperatura ≥ 260°C → Temperatura Alta.

Foram obtidos, da massa de dados, 9728 registros contendo as variáveis acima explicitadas, amostradas periodicamente de 1 em 1 minuto. A Figura 3 registra a evolução da temperatura do anodo ao longo do tempo.



**Figura 3: Dados da temperatura do anodo ao longo do tempo. A linha roxa representa o limiar para temperaturas baixas (200°C), enquanto que a linha preta representa o limiar para temperaturas altas (260°C).**

Vale ressaltar que, apesar de algumas variáveis da base de dados terem apresentado valores que são discrepantes e que se aproximam de 0, estes valores não podem ser considerados *outliers* e, portanto, serem expurgados da base de análise. Isto ocorre pois o modelo deve justamente captar e entender este momento do sistema; isto é, não basta que o modelo obtido por Aprendizado de Máquina registre com exatidão somente os dados de temperatura normal. Faz-se necessário modelar e classificar, com precisão, dados de temperaturas baixas e altas. Por este mesmo motivo (e por explicar a variabilidade de dados que ocasionam altas temperaturas), dados que aparentam ser *outliers* devido a seus altos valores também não foram expurgados da base de dados.

Por outro lado, detectou-se que as variáveis CO e CO<sub>2</sub> apresentam valores negativos devido a algumas falhas de medições dos sensores destes gases. Estes valores foram corrigidos, sendo lançado o valor 0 aos mesmos (valor correto em todos os casos).

A Figura 4 registra a correlação entre todas as variáveis do processo. Nesta figura fica evidente que diversas variáveis são bem correlacionadas entre si, sugerindo a possibilidade de que uma redução ou transformação de variáveis poderia ser benéfica ao modelo de classificação, podendo alcançar resultados similares mesmo em caso de redução de dimensionalidade dos dados de entrada.

	CrO/ Cr2O3	CaO flux	MW	Ore feed	Reduc tant	CO	H2	CO2	Bottom freeze layer	metal depth	Feed power product	slag_ depth	Temp_ level
CrO/ Cr2O3	1	-0,212	-0,203	-0,265	-0,221	-0,1874	-0,174	0,031	-0,3231	-0,2044	-0,28	0,2176	-0,1482
CaO flux	-0,2117	1	0,7966	0,9292	0,9438	0,7221	0,7218	0,67	0,1591	0,1528	0,9437	-0,5705	0,5135
MW	-0,203	0,7966	1	0,8458	0,8438	0,6415	0,6383	0,5822	0,1518	0,1544	0,8421	-0,517	0,5286
Ore feed	-0,2645	0,9292	0,8458	1	0,9844	0,7797	0,7743	0,6989	0,1997	0,1956	0,9889	-0,5962	0,5264
Reductant	-0,2206	0,9438	0,8438	0,9844	1	0,7743	0,7732	0,7102	0,1797	0,1708	0,9826	-0,6056	0,5456
CO	-0,1874	0,7221	0,6415	0,7797	0,7743	1	0,9967	0,887	0,223	0,2337	0,7738	-0,6896	0,6125
H2	-0,1736	0,7218	0,6383	0,7743	0,7732	0,9967	1	0,8941	0,2147	0,2252	0,7683	-0,6942	0,6208
CO2	0,031	0,67	0,5822	0,6989	0,7102	0,887	0,8941	1	0,0006	0,0517	0,6828	-0,5626	0,5308
Bottom freeze	-0,3231	0,1591	0,1518	0,1997	0,1797	0,223	0,2147	0,0006	1	0,7427	0,2077	-0,4223	0,3378
metal depth	-0,2044	0,1528	0,1544	0,1956	0,1708	0,2337	0,2252	0,0517	0,7427	1	0,1999	-0,4809	0,4175
Feed power product	-0,28	0,9437	0,8421	0,9889	0,9826	0,7738	0,7683	0,6828	0,2077	0,1999	1	-0,6035	0,5253
slag_ depth	0,2176	-0,571	-0,517	-0,596	-0,606	-0,6896	-0,694	-0,5626	-0,4223	-0,4809	-0,6035	1	-0,9139
Temp_ level	-0,1482	0,5135	0,5286	0,5264	0,5456	0,6125	0,6208	0,5308	0,3378	0,4175	0,5253	-0,9139	1

**Figura 4: Matriz de correlação entre dados de variáveis oriundas do processo estudado.**

Destaca-se, na Figura 4, que a variável Temperatura do Anodo já se encontra apresentada como “Classificação da Temperatura” (Temp\_Level), absorvendo os valores: 1 → Temperatura Baixa; 2 → Temperatura Normal; 3 → Temperatura Alta.

A última coluna da Figura 4 mostra a correlação entre as entradas e a temperatura classificada (saída). Em verde, foi marcada a variável de maior correlação (“Slag\_Depth”).

Visando a construção do modelo por meio do método KNN, aplicou-se o método de normalização denominado MIN-MAX a todos os atributos.

### 3.2. Abordagens para Construção do Modelo

Visando a construção do modelo via *Machine Learning* e tendo como base o pré-processamento realizado, optou-se por avaliar duas abordagens:

1. Abordagem 1: **Transformar Variáveis**

- a. O critério estabelecido envolve: (a) Aplicar o método Análise de Componentes Principais (ACP) nas variáveis originais do problema; (b) Avaliar a quantidade de componentes a serem utilizadas como sendo as variáveis de entrada do modelo; (c) Aplicar o método KNN ao novo conjunto de variáveis (componentes escolhidas) para classificar a temperatura do anodo; (d) Analisar resultados.
2. Abordagem 2: **Estabelecer critério de seleção de variáveis para o conjunto inicial de variáveis.**

O critério estabelecido envolve: (a) Eliminar todas as variáveis que tenham módulo de correlação abaixo de 30%; (b) Gerar modelo KNN de classificação da temperatura do anodo; (c) Analisar resultados.

No método de Análise de Componentes Principais (ACP) [Jackson 1991; Jolliffe 2002] as primeiras componentes registram a maior parte da variabilidade dos dados, enquanto, em geral, as últimas variáveis registram baixa variabilidade dos dados (modelando, em alguns casos, o ruído do processo).

Dada a matriz de correlação dos atributos deste problema, entende-se que a aplicação do método ACP aos dados de entrada reduziria a dimensionalidade sem perder a variabilidade significativa do processo.

### 3.3. Método KNN (k-Nearest Neighbors)

O método KNN [Parsian 2015] foi utilizado para solucionar o problema de classificação de Temperatura do Anodo. Para isto, os registros da base dados foram divididos em 85% para treinamento (8269 registros) e 15% para teste (1459 registros).

Neste algoritmo, as instâncias de treinamento são usadas para identificar qual a classe que mais se parece com a nova instância que se deseja identificar. A função distância euclidiana é calculada entre o ponto da massa de dados de teste e todos os pontos da base de dados de treinamento.

A função de classificação leva em consideração as  $K$  menores distâncias entre o ponto a se avaliar e a base de dados de treinamento. Define-se, portanto, a classe do novo ponto de dados por meio da votação majoritária. Ou seja, para cada um destes  $K$  pontos *dataset* de treinamento, verificam-se as suas respectivas classificações. Aquela classe que apresentar maior quantidade de votos destes  $K$  pontos vence a disputa e classifica o novo ponto.

As próximas duas seções apresentam os resultados para as Abordagens 1 e 2.

### 3.4. Análise de Resultados – Abordagem 1

Após a aplicação da técnica de ACP na base de dados das variáveis independentes, obteve-se a matriz de autovalores, que explicam a variabilidade acumulada desses dados na proporção mostrada na Tabela 1. Dados os valores registrados na Tabela 1, optou-se por gerar modelos KNN que utilizavam, como entrada, 2, 3, 4 e 5 componentes principais (ou *scores*), respectivamente, uma vez que, com 2 *scores*, 97,96% da variabilidade dos dados de entrada já estariam embutidas; enquanto que, para 5 *scores*, 99,99% da variabilidade dos dados de entrada já teria sido explicada. Destaca-se, portanto, a queda na quantidade de atributos de entrada do modelo (de 12 entradas para modelos com 2, 3, 4 ou 5 componentes principais).

**Tabela 1: Variabilidade acumulada explicada pelas componentes principais aplicadas aos dados originais de entrada.**

	Autovalor 1	Autovalor 2	Autovalor 3	Autovalor 4	Autovalor 5	Autovalor 6
Variabilidade acumulada explicada	80,73%	97,96%	99,67%	99,94%	99,99%	100,00%

O método KNN foi aplicado com valores de K pertencentes a um progressão aritmética com razão 3, e valores inicial e final iguais 1 a 40; isto é,  $K = 1, 4, 7, 10$  e assim sucessivamente até 40. Este critério foi aplicado já que, para valores distintos de K dentro deste intervalo, poder-se-ia haver empate na votação majoritária de decisão da classe, impedindo uma avaliação acurada da classe de determinado padrão.

A Figura 5 registra a matriz de confusão contendo os resultados obtidos para  $K = 1$  e 2 *scores*. Nesta figura, deve-se considerar a discretização 1, 2 e 3 equivalente a Temperatura Baixa, Normal e Alta, respectivamente. Além disso, os valores de *Precision* e *Recall* foram adicionados para facilitar a análise dos resultados.

		Matriz de Confusão			Real		
		1	2	3	Precision		
Predito	1	246	9	0	96,47%		
	2	13	773	31	94,61%		
	3	7	34	346	89,41%		
Recall		92,48%	94,73%	91,78%			
Acurácia		93,56%					

**2 SCORES**

**Figura 5: Matriz de Confusão para modelo KNN = 1, após aplicação do ACP, e mantendo-se 2 componentes principais.**

Nota-se, por meio da Figura 5, o resultado para  $K = 1$  e duas componentes principais mantidas. Em amarelo foram marcados os valores considerados críticos, isto é, os  $FalseNegative_3$  (que indicam que a temperatura prevista foi classificada como baixa ou normal, sendo que, na prática, ela era alta). Em verde foram marcados os valores considerados moderadamente críticos, isto é, os  $FalseNegative_1$  (que indicam que a temperatura prevista foi classificada como alta ou normal, sendo que, na prática, ela era baixa).

A Tabela 2 registra todos os resultados encontrados, comparando os resultados dos modelos. O modelo que apresentou melhores resultados para esta abordagem encontra-se destacado na Tabela 2, e apresentou  $K = 22$  e 3 *scores*. A matriz de confusão deste modelo está representada na Figura 6.

**Tabela 2: Comparação dos resultados obtidos para os modelos KNN, após aplicação de ACP, e mantendo-se de 2 a 5 componentes principais.**

KNN	Acurácia (2 Scores)	Acurácia (3 Scores)	Acurácia (4 Scores)	Acurácia (5 Scores)
<b>1</b>	93,56%	93,21%	93,15%	93,49%
<b>4</b>	95,00%	94,31%	94,17%	94,04%
<b>7</b>	95,20%	95,07%	94,93%	95,00%

10	95,61%	95,07%	95,13%	95,07%
13	95,54%	95,07%	95,20%	95,00%
16	95,68%	95,48%	95,48%	95,48%
19	95,48%	95,68%	95,48%	95,61%
22	95,68%	95,89%	95,54%	95,54%
25	95,61%	95,34%	95,54%	95,54%
28	95,41%	95,48%	95,34%	95,34%
31	95,41%	95,41%	95,34%	95,34%
34	95,34%	95,34%	95,07%	95,07%
37	95,41%	95,34%	95,13%	95,13%
40	95,48%	95,34%	95,41%	95,47%

		Matriz de Confusão			Real		
		1	2	3	Precision		
Predito	1	249	6	0	97,65%		
	2	13	788	16	96,45%		
	3	7	18	362	93,54%		
Recall		92,57%	97,04%	95,77%			
Acurácia		95,89%					

3 SCORES

Figura 6: Matriz de Confusão para melhor modelo obtido para a abordagem 1. K = 22, mantendo-se 3 componentes principais.

### 3.5. Análise de Resultados – Abordagem 2

O primeiro passo desta abordagem envolvia eliminar todas as variáveis de entrada que tivessem módulo de correlação com a saída menor que 30%. De todas as variáveis de entrada do modelo, apenas a variável “CrO/Cr<sub>2</sub>O<sub>3</sub>” possuía módulo de correlação menor que 30%. Além disso, trata-se de uma variável oriunda de análise laboratorial e, portanto, sua atualização era menos recorrente do que outras variáveis de processo. Pelos motivos acima expostos, tal variável foi retirada do modelo.

Após a devida normalização das variáveis, foram gerados modelos KNN de classificação da temperatura do anodo com as variáveis restantes. O resultado encontrado para K = 1 encontra-se registrado na Figura 7.

Nota-se, por meio da Figura 7, a eficácia do modelo obtido, dado que não há registros de *FalseNegative*<sub>3</sub> e *FalseNegative*<sub>1</sub> para esta base de dados de teste. A Tabela 3 registra resultados obtidos para todos os modelos em função de K.

Tabela 3: Comparação dos resultados obtidos para os modelos KNN, excluindo a variável “CrO/Cr<sub>2</sub>O<sub>3</sub>” dos atributos de entrada.

KNN	1	4	7	10	13	16	19
Acurácia	99,86%	96,85%	96,57%	96,57%	96,44%	96,44%	96,23%
KNN	22	25	28	31	34	37	40
Acurácia	96,44%	96,44%	96,37%	96,37%	96,23%	96,09%	96,09%

O modelo que apresentou melhores resultados para esta abordagem encontra-se destacado na Tabela 3, e apresentou  $K = 1$ . A matriz de confusão deste modelo está representada na Figura 7.

		Matriz de Confusão		Real			Precision
				1	2	3	
Predito	1	255	2	0	99,22%		
	2	0	815	0	100,00%		
	3	0	0	387	100,00%		
Recall		100,00%	99,76%	100,00%			
Acurácia		99,86%					

$K = 1$

**Figura 7: Matriz de Confusão para modelo KNN com  $K=1$ , excluindo a variável “CrO/Cr<sub>2</sub>O<sub>3</sub>” dos atributos de entrada.**

Em virtude dos resultados obtidos para esta abordagem serem ligeiramente superiores àqueles envolvendo a Abordagem 1, uma nova “Abordagem 3” foi realizada.

- **Abordagem 3: Manter apenas as variáveis de entrada melhor correlacionadas e gerar modelos com dimensões reduzidas em relação à Abordagem 2.**

O critério estabelecido envolve: a) Manter apenas as seis variáveis de entrada melhor correlacionadas com a saída; b) Gerar modelo KNN de classificação da temperatura do anodo; e c) Analisar resultados.

### 3.6. Análise de Resultados – Abordagem 3

As variáveis mantidas para a confecção do modelo, conforme se pode verificar na matriz de correlação (Figura 4) foram: slag\_depth, H<sub>2</sub>, CO, Reductant, CO<sub>2</sub> e MW.

Após a devida normalização das variáveis, foram gerados modelos KNN de classificação da temperatura do anodo com as variáveis restantes. Os resultados encontrados para  $K = 1$  encontram-se registrados na Figura 8.

		Matriz de Confusão		Real			Precision
				1	2	3	
Predito	1	254	1	0	99,61%		
	2	3	814	0	99,63%		
	3	0	0	387	100,00%		
Recall		98,83%	99,88%	100,00%			
Acurácia		99,73%					

$K = 1$

**Figura 8: Matriz de Confusão para modelo KNN com  $K=1$ , mantendo-se as 6 variáveis melhor correlacionadas como entradas do modelo.**

Nota-se, por meio da Figura 8, a similaridade com o resultado obtido na Abordagem 2. Essa análise é confirmada por meio da Tabela 4, que registra os resultados obtidos para todos os modelos em função de  $K$ .

Ressalta-se, ainda, que os resultados obtidos na “Abordagem 3” são bastante similares àqueles verificados com a “Abordagem 2”. As diferenças se devem ao fato de

que a eliminação de determinados atributos de entrada do modelo resultou em pequena perda de variabilidade relevante do processo, o que resultou em pequena queda de desempenho para a “Abordagem 2”. No entanto, os resultados da “Abordagem 3” são também consistentes, uma vez que o percentual de erro (neste caso, igual ao MAPE) para  $K = 1$  mantém-se inferior a 0,5%, e para  $K = 40$  mantém-se inferior a 5%.

**Tabela 4: Comparação dos resultados obtidos para os modelos KNN, mantendo-se as 6 variáveis melhor correlacionadas com a saída.**

KNN	1	4	7	10	13	16	19
Acurácia (Com Normalização)	99,73%	96,57%	96,85%	96,98%	96,78%	96,64%	96,50%
KNN	22	25	28	31	34	37	40
Acurácia (Com Normalização)	96,57%	96,30%	96,23%	96,16%	96,02%	96,02%	95,96%

#### 4. Conclusões

Neste artigo abordou-se a construção de modelos baseados em técnicas de Mineração de Dados e *Machine Learning* no intuito de solucionar o problema de classificação da temperatura do anodo de um FEA, tipicamente utilizado em indústrias dos setores siderúrgicos e metalúrgicos. Para tanto, após a etapa de pré-processamento de dados, três abordagens para a solução do problema foram adotadas. Na primeira, aplicou-se a técnica de ACP sobre os atributos de entrada do modelo. Em seguida, foram gerados modelos baseados no algoritmo KNN que foram bastante eficazes para o problema, além de ter baixa complexidade computacional. Na segunda abordagem, foram gerados modelos KNN excluindo apenas a variável atrelada à análise laboratorial. Por fim, na terceira abordagem, o modelo KNN foi aplicado a um grupo de dados que continha, como variáveis de entrada, apenas as seis variáveis independentes do grupo original melhor correlacionadas com a variável dependente.

A “Abordagem 2” apresentou os melhores resultados, sendo que o modelo KNN com  $K = 1$  apresentou acurácia superior a 99,5% dos dados de teste. Todavia, ressalta-se que o modelo KNN com variáveis reduzidas, contemplado pela “Abordagem 3”, também apresentou resultados próximos àqueles da “Abordagem 2”, sendo que o modelo KNN com  $K = 1$  apresentou acurácia de classificação superior a 99,5%. Por fim, os modelos KNN obtidos após a aplicação de ACP nas variáveis originais do modelo apresentaram bons resultados (ainda que piores que as duas abordagens anteriores), sendo o melhor modelo aquele obtido para  $K = 22$ , e mantendo as três principais componentes oriundas do modelo ACP.

Ressalta-se, portanto, a relevância da aplicação do algoritmo KNN bem como de técnicas de pré-processamento de dados em problemas de importante aplicação nos setores metalúrgico e siderúrgico. A correta classificação e/ou inferência da temperatura do forno resulta em ações adequadas para manutenção do reator dentro das faixas toleradas de temperatura, evitando-se perdas de corridas metálicas com alto valor agregado e desgaste precoce de refratário do forno (o que anteciparia uma manutenção corretiva e reduziria a vida útil do forno elétrico). Vale destacar também a característica pioneira deste trabalho, uma vez que não foram identificados registros, na literatura avaliada, de aplicações do método KNN (juntamente com a técnica ACP) para efetuar a classificação de temperatura de anodo em FEA.

Como trabalhos futuros, sugere-se a realização de inferências e previsões *multi-step* da temperatura do anodo valendo-se do mesmo conjunto de variáveis e dados coletados. A aplicação de técnicas envolvendo Lógica *Fuzzy* em conjunto com o uso de Redes Neurais Profundas [Haykin 1998; Goodfellow et al. 2016] que possam classificar e prever a temperatura de anodo mostra-se relevante para comparar os resultados aqui obtidos e avaliar a aplicabilidade da técnica KNN para este problema.

## Bibliografia

- ABM, (1978) “Fusão em Fornos Elétricos - Princípios de Funcionamentos de Fornos Elétricos: Forno a Arco Direto, a Arco Indireto, a Resistência; Forno a Indução e Forno Especiais Processo Duplex e Triplex”. Associação Brasileira de Metais - Fundição, 10ª edição.
- Bergman, K., Kjellberg, B. (2001) “Arc Furnace Technology Applied to Smelting Applications”. INFACON IX – 9ª Inter. Ferroalloys Cong., Quebec, Canadá.
- Blachnik M., Maćzka K., Wiczorek T. (2010) “A Model for Temperature Prediction of Melted Steel in the Electric Arc Furnace (EAF)”. Artificial Intelligence and Soft Computing. ICAISC 2010. vol 6114. Springer, Berlin, Heidelberg.
- Chen, C., Liu, Ying, Kumar, M., Qin, J. (2018) “Energy Consumption Modelling Using Deep Learning Technique — A Case Study for EAF”. Elsevier. 51º CIRP Conferência em Sistemas de Manufatura. p. 1063 - 1068.
- Da Costa e Silva, A. (2013) Refino dos Aços - Capítulo 7: Forno Elétrico a Arco.
- Fernandéz, J. M. M., Cabal, V.A., Montequin, V.R., Balsera, J.V. (2008) “Online estimation of electric arc furnace tap temperature by using fuzzy neural networks”. Elsevier. Engineering Applications of Artificial Intelligence. p. 1001 – 1012.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016) Deep learning. MIT press.
- Jackson, J. (1991) A User’s Guide to Principal Components, Wiley-Interscience Publication, John Wiley & Sons, Inc.
- Haykin, S. (1998) Neural Networks: A Comprehensive Foundation, Prentice Hall, 2ª Edição.
- Jolliffe, I. (2002) Principal Component Analysis, Springer.
- Jones, R. (2011) “Smelting Applications of DC Arc Furnaces. Fray Inter. Symposium on Metals and Materials Processing in a Clean Environment”. Cancun, México.
- Mourão, M. (2019) “PMT 2420 – Metalurgia Geral”, USP, São Paulo.
- O’Brien, N. and Ford, M. (2018) “Practical Aspects of DC Furnace Design”. INFACON VIII, Pequin, China.
- Parsian, M. (2015) Data Algorithms – Recipes for Scaling Up with Hadoop and Spark, O’Reilly Media.
- Rogers, C. and Cramer, C. (1989) “Advanced Process Control for Electric Arc Furnaces”. Carnegie Mellon Research Institute, Pittsburg, Pennsylvania, EUA.
- Sampaio, J. and Almeida, S. (2005) “Calcário e dolomito In: Rochas & minerais industriais: usos e especificação”, RJ: CETEM/MCT, Parte II. Cap. 15. p.327-350.