

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
ESCOLA DE ENGENHARIA  
DEPARTAMENTO DE METALURGIA

HELIOIR ANTÔNIO DA SILVA JÚNIOR

**DESENVOLVIMENTO DE MODELO ESTATÍSTICO PARA PREVISÃO DA  
LINGOTABILIDADE NA PRODUÇÃO DE AÇOS LONGOS ESPECIAIS**

PORTO ALEGRE  
2018

HELIOIR ANTÔNIO DA SILVA JÚNIOR

**DESENVOLVIMENTO DE MODELO ESTATÍSTICO PARA PREVISÃO DA  
LINGOTABILIDADE NA PRODUÇÃO DE AÇOS LONGOS ESPECIAIS**

Trabalho de Conclusão de Curso apresentado à  
Escola de Engenharia da Universidade Federal do  
Rio Grande do Sul como requisito parcial à obtenção  
do grau de Bacharel em Engenharia Metalúrgica.

Orientador: Prof. Dr. Wagner Viana Bielefeldt

Porto Alegre  
2018

Este trabalho é dedicado a minha  
família, aos meus amigos e aos  
meus colegas de trabalho.

## **AGRADECIMENTOS**

Agradeço, primeiramente, a minha família que respaldou toda trajetória acadêmica percorrida.

Agradeço ao Laboratório de Siderurgia da UFRGS por fornecer os recursos intelectuais e apoio necessários para a execução do trabalho.

Agradeço aos colegas de trabalho, em especial ao futuro engenheiro Matheus Funck e ao Eng. Pedro Rehfeld Schoen pela contribuição criativa e técnica que foi necessária para o desenvolvimento do modelo.

Agradeço ao Prof. Dr. Wagner Viana Bielefeldt e ao MSc. Vinicius Cardoso da Rocha pela orientação, conhecimento e oportunidade para colocar em prática este trabalho.

Agradeço também à equipe da aciaria da Gerdau Charqueadas pelos dados fornecidos para o desenvolvimento do trabalho.

## RESUMO

A melhoria da qualidade do aço e do processo siderúrgico têm se tornado parte da rotina dos engenheiros metalúrgicos e empresas siderúrgicas, em um mercado exigente de produtos de qualidade e preços altamente competitivos. O lingotamento contínuo é uma das etapas mais importantes do processo de fabricação do aço e, no mundo, cerca de 85 % das siderúrgicas utilizam deste método para solidificar o aço. Contudo, muitos problemas associados a este processo podem surgir, como inclusões não-metálicas, que são originadas de diferentes formas, podendo provocar severos danos ao produto final ou até mesmo prejudicar a fabricação do aço com fenômenos como o *clogging*. Para resolver este problema, diversas soluções são propostas, como práticas para minimizar as inclusões, popularmente conhecidas como limpeza do aço. No presente trabalho, sugeriu-se o desenvolvimento de um modelo estatístico capaz de, a partir de dados históricos, conseguir prever o êxito na lingotabilidade bem como elencar as principais variáveis que influenciam neste processo. Para isso, utilizou-se o modelo de regressão logística que foi capaz de prever, com acurácia de 85,6 %, se houve problema de lingotabilidade na corrida. Também foi possível determinar a probabilidade de sucesso de cada corrida bem como a influência das variáveis empregadas para desenvolvimento do modelo. Sequenciamento no distribuidor, temperatura *liquidus*, *superheat* e teores de alumínio e enxofre foram os principais parâmetros que impactaram nas chances de sucesso do lingotamento contínuo.

Palavras-chave: Lingotabilidade; inclusões não-metálicas; *clogging*; modelagem estatística preditiva; regressão logística.

## **ABSTRACT**

Improving the steel quality and the steelmaking process has become part of the routine of metallurgical engineers and steel companies in a demanding market for quality products and highly competitive prices. Continuous casting is one of the most important stages in the steelmaking process, in the world, about 85% of steel companies use this method to solidify steel. However, many problems associated with this process may arise, such as non-metallic inclusions, which originate in different ways, might causing severe damage to the final product or even undermine the fabrication of steel with phenomena such as clogging. To solve this problem, several solutions are proposed, such as practices to minimize inclusions, popularly known as steel cleanliness. In the present work, it was suggested the development of a statistical model, based on historical data, able to predict the success in castability as well as to list the main variables that influence in this process. Therefore, a logistic regression model was used, which was able to predict, with an accuracy of 85.6%, if there was a castability problem in the process. Also was possible to determine the probability of success of each heat as well as the influence of the variables used to develop the model. Sequencing in the tundish, liquidus temperature, superheat and aluminum and sulfur contents were the main parameters that impacted on the success of continuous casting.

Keywords: Castability; non-metallic inclusions; clogging; predictive statistical modeling; logistic regression.

## LISTA DE FIGURAS

<b>Figura 1</b> – Representação esquemática do lingotamento contínuo.....	14
<b>Figura 2</b> – Representação esquemática dos fenômenos no distribuidor. ....	15
<b>Figura 3</b> – N° de inclusões em um cubo de aço com volume de 1 cm <sup>3</sup> em função do teor de oxigênio e diâmetro da inclusão. ....	17
<b>Figura 4</b> - Representação esquemática da válvula submersa e as regiões de ocorrência de clogging .....	18
<b>Figura 5</b> – Seção longitudinal de válvula submersa onde a entrada é constituída por MgO-C, o volume constituído por Al <sub>2</sub> O <sub>3</sub> -C e a linha de saída é reforçada com ZrO <sub>2</sub> -C .....	19
<b>Figura 6</b> - Efeito do oxigênio e do enxofre na janela líquida à 1550 °C .....	20
<b>Figura 7</b> - Comparativo entre regressão linear e logística .....	22
<b>Figura 8</b> – Curva ROC com pontos importantes. ....	26
<b>Figura 9</b> – Gráfico do KS para duas classes. ....	29
<b>Figura 10</b> – Exemplo de matriz de confusão binária.....	30
<b>Figura 11</b> – Fluxograma da metodologia aplicada no trabalho. ....	33
<b>Figura 12</b> – Ambiente de desenvolvimento Spyder .....	39
<b>Figura 13</b> – Esquema da divisão da base de dados.....	41
<b>Figura 14</b> – Esquema do treinamento da base via <i>cross-validation</i> . ....	42
<b>Figura 15</b> – Exemplo de equações da reta com e sem <i>fit_intercept</i> . ....	43
<b>Figura 16</b> – Curva ROC referente à base de treino .....	46
<b>Figura 17</b> – Curva ROC referente à base de teste .....	47
<b>Figura 18</b> – Gráfico do KS. ....	48
<b>Figura 19</b> – Valor comparativo do KS faixa a faixa. ....	49
<b>Figura 20</b> – Matriz de confusão .....	50
<b>Figura 21</b> – Classificador de distribuição da regressão logística com gradiente de probabilidades. ....	52
<b>Figura 22</b> – Classificador de distribuição da regressão logística com <i>threshold</i> em 0,5. ....	52

## LISTA DE TABELAS

<b>Tabela 1</b> – Exemplo de tabela de distribuição de duas classes para determinação do KS. ....	28
<b>Tabela 2</b> – Principais resultados da validação cruzada estratificada .....	45
<b>Tabela 3</b> – Tabela de distribuição do KS. ....	48
<b>Tabela 4</b> – Tabela com precisão, revocação e F1-Score.....	51
<b>Tabela 5</b> – Valores dos coeficientes das principais variáveis. ....	53

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>11</b>
<b>2 REVISÃO DA LITERATURA</b> .....	<b>13</b>
2.1 Lingotamento contínuo (LC) .....	13
2.2 Inclusões não-metálicas .....	16
2.3 <i>Clogging</i> .....	17
2.3.1 Válvula submersa .....	18
2.3.2 Tratamento de inclusões com cálcio .....	20
2.4 Modelo estatístico de classificação .....	21
2.5 Validação cruzada estratificada ( <i>stratified cross-validation</i> ) .....	24
2.6 Avaliação do modelo de classificação .....	25
2.6.1 <i>Curva ROC, área abaixo da curva (AUC) e acurácia</i> .....	25
2.6.2 <i>Teste de KS (Kolmogorov–Smirnov test)</i> .....	28
2.6.3 <i>Matriz de confusão binária</i> .....	29
2.6.4 <i>Precisão, revocação e F1-Score</i> .....	30
<b>3 METODOLOGIA</b> .....	<b>33</b>
3.1 Bases de dados .....	34
3.1.1 <i>Base: Parâmetros de processo</i> .....	34
3.1.2 <i>Base: Adição de elementos à escória</i> .....	34
3.1.3 <i>Base: Adição Al e CaSi</i> .....	35
3.1.4 <i>Base: Posição da válvula tampão</i> .....	35
3.1.5 <i>Base: Aços sucateados por problemas no lingotamento</i> .....	35
3.1.6 <i>Base: Provas de escória</i> .....	35
3.1.7 <i>Base: Provas de aço</i> .....	35
3.2 Pré-processamento das bases de dados .....	36
3.3 Plataforma computacional .....	38

3.3.1	<i>Hardware</i> .....	38
3.3.2	<i>Software</i> .....	39
3.4	Processamento das bases de dados .....	40
3.5	Separação das bases de treino e teste .....	41
3.6	Treino, validação e teste do modelo.....	41
3.6.1	<i>Validação cruzada estratificada (stratified cross-validation)</i> .....	42
3.6.2	<i>Treinamento do modelo</i> .....	44
3.6.3	<i>Teste do modelo</i> .....	44
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b> .....	<b>45</b>
4.1	Validação cruzada estratificada ( <i>stratified cross-validation</i> ).....	45
4.2	Treinamento do modelo .....	46
4.3	Teste do modelo .....	47
4.3.1	<i>Curva ROC, AUC e acurácia</i> .....	47
4.3.2	<i>Teste de KS</i> .....	48
4.3.3	<i>Matriz de confusão</i> .....	49
4.3.4	<i>Precisão, revocação e F1-Score</i> .....	50
4.3.5	<i>Classificador de distribuição da regressão logística</i> .....	51
4.3.6	<i>Coeficientes das variáveis da regressão logística</i> .....	53
<b>5</b>	<b>CONCLUSÕES</b> .....	<b>55</b>
<b>6</b>	<b>SUGESTÕES PARA TRABALHOS FUTUROS</b> .....	<b>57</b>
	<b>REFERÊNCIAS</b> .....	<b>58</b>
	<b>APÊNDICE A</b> .....	<b>62</b>
	<b>APÊNDICE B</b> .....	<b>65</b>

## 1 INTRODUÇÃO

A crescente demanda por aços especiais com alta qualidade tem impulsionado a indústria siderúrgica para uma exigência cada vez mais elevada. Um dos fatores que podem influenciar no processo de um aço de boa qualidade é sua “limpeza”, ou seja, produzir um material com o menor número de inclusões possível. Estas, principalmente não-metálicas, que podem ser nocivas tanto para o produto final quanto para o próprio processo de fabricação. Neste podendo muitas vezes levar a reparos ou até mesmo sucateamento total (ZHANG, 2006).

Diversos fatores podem influenciar a formação de inclusões durante o processo de fabricação do aço, podendo ser fatores físicos como temperatura, turbulência e pressão, bem como fatores químicos como composição química e reações com diferentes elementos.

Quando a limpeza do aço, durante seu processo de produção, não é severamente tratada, a tendência de formar-se inclusões e acarretar problemas durante o processo de solidificação, mais especificadamente no lingotamento contínuo, é alta. O acúmulo destas inclusões na válvula submersa, durante o processo de lingotamento contínuo é chamado de *clogging*, que é um dos principais problemas enfrentados visto que inclusões não-metálicas são inerentes ao processo de fabricação do aço.

Visto que para se promover uma boa lingotabilidade é necessária uma boa fluidez do metal, bem como evitar obstrução da válvula durante o fluxo do aço do distribuidor para os veios, autores sugerem diferentes soluções para se trabalhar com aços propensos à formação de *clogging* como controlar teor de FeO + MnO, tempo de vácuo durante a desgaseificação, variação do nível da válvula tampão, teor de Al, nível de basicidade, teor de SiO<sub>2</sub> e Al<sub>2</sub>O<sub>3</sub> na escória (BUBOLTZ, 2011).

A fim de determinar quais as principais variáveis que prejudicam o processo de lingotamento contínuo na fabricação de aços especiais e também de mensurar a probabilidade de o processo apresentar problemas, propõe-se como principal objetivo deste trabalho o desenvolvimento de um modelo estatístico de predição em termos de lingotabilidade a partir de dados coletados pela usina siderúrgica nos últimos três anos. Sugere-se que cada parâmetro durante o processo de fabricação do aço

apresente certa influência para o sucesso do lingotamento contínuo e para conseguir avaliar cada um deles de forma menos subjetiva possível, utiliza-se ferramentas estatísticas que, hoje, já auxiliam diversos setores da sociedade como instituições financeiras, institutos meteorológicos, medicina, corporações varejistas e departamentos de transportes. Por conseguinte, avaliar os parâmetros, bem como cada variável do processo, tornam-se os objetivos específicos do trabalho.

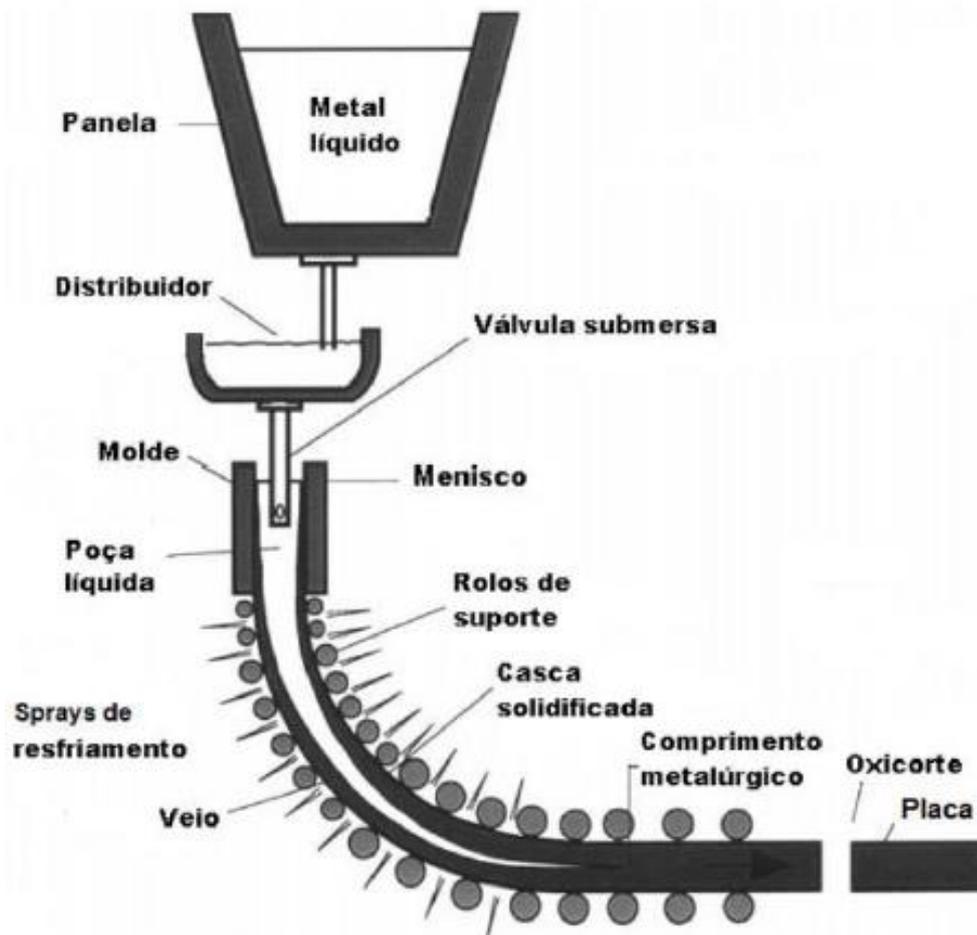
## 2 REVISÃO DA LITERATURA

Para entender os motivos que levam a problemas durante o processo de lingotamento contínuo, faz-se necessário uma revisão. São apresentados esquemas e ilustrações comentadas, a fim de se obter um esclarecimento sobre o assunto.

### 2.1 Lingotamento contínuo (LC)

O lingotamento é o último processo dentro de uma aciaria, em que se trabalha com o metal ainda líquido e tem como objetivo solidificá-lo na forma desejada.

O início do processo é caracterizado pela chegada da panela contendo o aço do refino secundário, em que é vazado para o distribuidor, o aço ainda líquido, através de uma válvula refratária posicionada na parte inferior da panela para um tubo longo que protege o fluxo de aço do contato com a atmosfera. Este tubo leva o aço líquido até o distribuidor, recipiente onde ocorre a homogeneização e limpeza do aço, auxiliando no processo de remoção de inclusões ainda presentes no banho. O distribuidor também é responsável pela alimentação dos moldes, através de uma válvula submersa, com aço líquido para posterior solidificação. A Figura 1 mostra um esquema ilustrando os principais itens que compõe o processo de lingotamento contínuo.

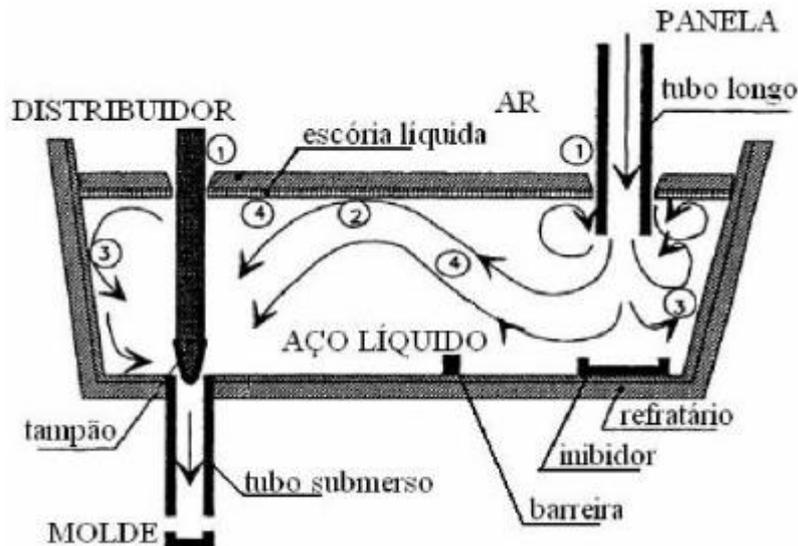


**Figura 1** – Representação esquemática do lingotamento contínuo

Fonte: BIELEFELDT, 2013.

É a partir da saída para os veios em que ocorre a solidificação, pois o aço líquido flui por um molde de cobre resfriado por água em sua superfície externa, promovendo a extração de calor necessária para uma solidificação de fora para dentro, ou seja, solidifica primeiro a camada externa e depois o núcleo. Para evitar o contato do ar atmosférico no momento desta transferência do distribuidor para os veios, uma válvula submersa é responsável para controlar esta passagem. E ela tem papel fundamental no processo de lingotamento contínuo, visto que o fluxo e velocidade precisam permanecer os mais constantes possíveis a fim de que se evitem turbulências e defeitos na solidificação (BANNENBERG, 1995; RIZZO, 2005).

O papel do distribuidor, no que tange a limpeza ou ainda descontaminação do aço, é deveras importante a fim de que se evite tanto problema durante o processo quanto má qualidade no produto final. Pois o objetivo final, no que diz respeito à limpeza, são dois: menores níveis de elementos residuais na composição química do aço e menor nível inclusionário (MAGALHÃES, 2010). Esquemáticamente, os fenômenos que podem ocorrer dentro do distribuidor, são mostrados na Figura 2.



**Figura 2** – Representação esquemática dos fenômenos no distribuidor.

Fonte: ZHANG, THOMAS, 2003.

- (1) Oxidação do banho pelo ar no vazamento da panela e absorção de nitrogênio;
- (2) Oxidação de alumínio pelo FeO, MnO e SiO<sub>2</sub> da escória; incorporação de alumina pelo aço; e absorção das inclusões pela escória devido a flotação.
- (3) Erosão dos refratários;
- (4) Reações de desoxidação e remoção de algumas inclusões.

Nota-se ainda a presença de barreiras e inibidores no fundo do distribuidor, de acordo com o esquema, que auxiliam na obtenção de um escoamento ascendente do aço líquido dentro do distribuidor, havendo maior interação do aço com a escória sobrenadante. Esses fatores promovem uma remoção de inclusões mais facilitada o que ajuda na limpeza no aço (ZHANG, 2003). Já o material refratário do distribuidor, seu revestimento interno, deve ser extremamente estável, visto que a elevadas temperaturas não pode ocorrer reações de redução ou de reoxidação do banho

líquido, as quais podem promover a formação de inclusões não-metálicas (MILLMAN, 2004).

## 2.2 Inclusões não-metálicas

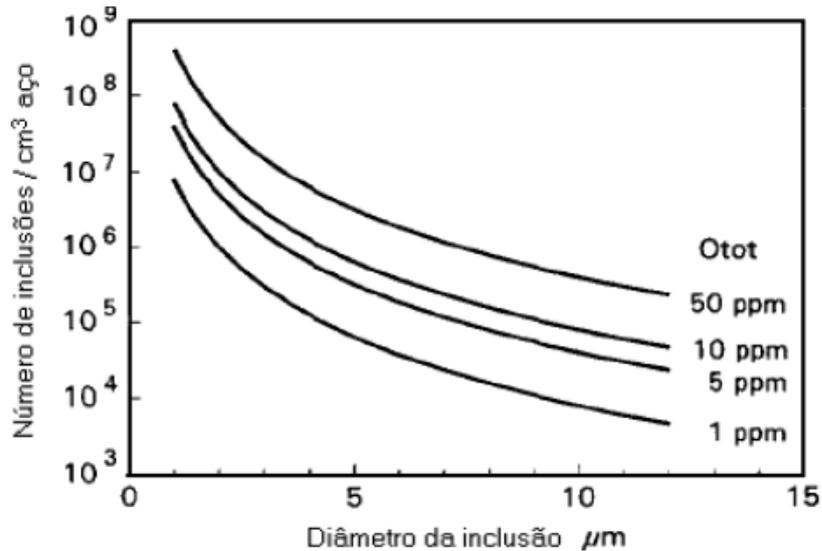
Como foi abordado na subseção anterior, alguns problemas durante a limpeza do aço durante o processo de lingotamento contínuo podem ocorrer, dentre eles, as inclusões não-metálicas. Estudá-las torna-se importante pois elas podem ser responsáveis por produzir defeitos também no produto final, como alterações nas suas propriedades mecânicas, tais como ductilidade, tenacidade e resistência à fadiga (HELLE; HOLAPPA, 1995).

As inclusões não-metálicas são formadas, geralmente, quando o aço ainda está no estado líquido e se diferem de precipitados pois eles surgem durante a solidificação do aço e podemos classificá-las em dois tipos: endógenas e exógenas (MILLMAN, 2004; DEKKERS, 2002).

(1) Inclusões endógenas: resultantes do processo de desoxidação do aço (durante as reações de refino) ou resultantes do processo de solidificação do aço (partículas com origem na precipitação).

(2) Inclusões exógenas: são formadas por fatores externos que, geralmente, são oriundos de reações incidentais (como reoxidação) ou ainda interação mecânica do aço líquido com o refratário ou arraste de escória (KIESSLING, 1978; ZHANG, 2006).

Alguns fatores influenciam diretamente no tamanho e na quantidade das inclusões formadas. O efeito do oxigênio ativo no aço, possui influência logarítmica, como pode ser visto na Figura 3, sobre o diâmetro das inclusões. Este é um dos motivos de se evitar a exposição do aço líquido com o ar atmosférico e, para isso, utiliza-se válvulas e gases inertes (MILLMAN, 2004).



**Figura 3** – N<sup>o</sup> de inclusões em um cubo de aço com volume de 1 cm<sup>3</sup> em função do teor de oxigênio e diâmetro da inclusão.

Fonte: HOLAPPA; HELLE (1995).

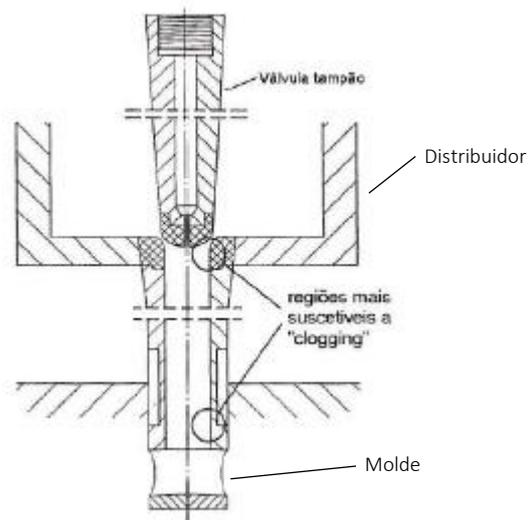
A influência na formação de inclusões, não são somente afetadas pelo oxigênio, mas também por outros fatores, como: composição química da escória, elementos das ferroligas (como Ca, Al, Si, Mg, K), teor de alumina e sílica no aço, teor de FeO, teor de MnO, entre outros óxidos metálicos (OTOTANI, 1981; GHOSH, 2001; MAGALHÃES, 2010).

### 2.3 Clogging

Durante o lingotamento contínuo, quando o aço líquido é transferido do distribuidor para os moldes, é a válvula submersa que realiza o trabalho de evitar o contato com o ar atmosférico. Um tampão, que varia sua posição controla o fluxo de aço líquido sobre os veios. No momento desta transferência, inclusões não-metálicas podem se acumular nesta válvula submersa, perturbando ou até mesmo impedindo completamente (obstruindo) o fluxo do aço. Este fenômeno é chamado de *clogging* (VERMEULEN, 2000; MERMARPOUR, 2011).

### 2.3.1 Válvula submersa

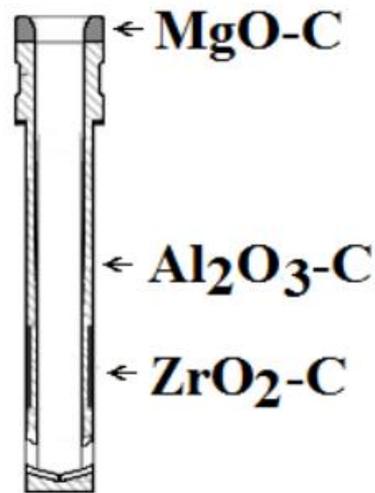
Visto que as inclusões não-metálicas são inerentes ao processo de fabricação do aço, deseja-se o mais alto nível de limpidez possível para evitar fenômenos com o *clogging*, que acabam prejudicando boa parte ou até mesmo toda corrida, acarretando no seu possível sucateamento. Na Figura 4 podem ser observadas as regiões mais sensíveis à formação de *clogging* (JANKE et Al, 2000).



**Figura 4** - Representação esquemática da válvula submersa e as regiões de ocorrência de clogging

**Fonte:** BANNENBERG, 1995.

As válvulas submersas consistem em diferentes materiais de base refratários (RBM), como  $\text{Al}_2\text{O}_3\text{-C}$ ,  $\text{MgO-C}$  e  $\text{ZrO}_2\text{-C}$ , conforme pode ser visto no esquema da Figura 5 (SVENSSON, 2017).



**Figura 5** – Seção longitudinal de válvula submersa onde a entrada é constituída por MgO-C, o volume constituído por Al<sub>2</sub>O<sub>3</sub>-C e a linha de saída é reforçada com ZrO<sub>2</sub>-C

Fonte: SVENSSON, 2017.

Como o processo de lingotamento contínuo ocorre a temperaturas próximas de 1565 °C, a interação entre o aço líquido e o RBM é inevitável. Além disso, a válvula submersa é aquecida antes do vazamento, tanto para evitar o choque térmico tanto para evitar que o aço solidifique diretamente quando entrar em contato. Uma vez que este RBM contém grafita, existe um alto risco que a válvula seja descarbonetada devido ao aumento da atividade de oxigênio próximo à parede interna. É extremamente importante evitar esta descarbonetação, visto que ela aumenta a taxa de entupimento (GOSH, 2000; SUNAYAMA, KAWAHARA, 2005; MERMAPOUR, 2011)

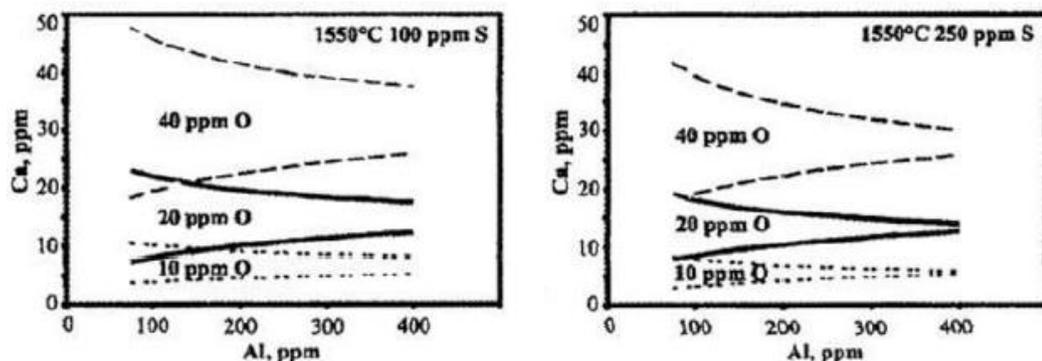
O *clogging* ocorre devido a diferentes fatores, como uma limpeza pobre do aço, baixa temperatura do aço, velocidade do fluxo, pureza e estabilidade do RBM. A transferência gasosa através do RBM será maior na superfície descarbonetada e essa descarbonetação resultará em uma formação de CO, e pela oxidação do alumínio levará a formação de inclusões de Al<sub>2</sub>O<sub>3</sub>. Uma vez que a primeira camada depositada de inclusões não-metálicas tenha sido formada, a superfície da parede da válvula se torna áspera, portanto mais suscetível ao acúmulo de aglomerados (POIRIER, 1995; VERMEULEN, 2002; TUTTLE, 2005).

### 2.3.2 Tratamento de inclusões com cálcio

Transformar inclusões sólidas de óxidos de alumínio, por exemplo – que possam gerar *clogging* – em inclusões líquidas ou inclusões menos nocivas é uma das soluções. Para isso, comumente é utilizado o tratamento com cálcio.

A adição de cálcio é comumente utilizada a fim de formar inclusões de alumina em sulfeto de cálcio, obviamente dependendo das concentrações de alumínio e enxofre no banho. Além de tratar inclusões de alumina, o cálcio é um poderoso desoxidante, podendo diminuir o teor de oxigênio a patamares bem baixos (DEKKERS, 2002; STORY, 2004).

Como o oxigênio influencia diretamente na quantidade do cálcio que será necessário para converter as inclusões de  $Al_2O_3$  em cálcio-aluminatos, é extremamente importante controlar a quantidade de  $O_2$  no banho para atingir uma “janela líquida”, que é a condição que irá gerar inclusões líquidas. A Figura 6 mostra o efeito de três teores de oxigênio para três janelas líquidas de inclusões para as concentrações de 100 e 250 ppm de enxofre numa temperatura constante de 1550 °C. A adição de cálcio em função do teor de Al deve estar entre as linhas do gráfico, chamadas linhas de saturação de CaS (superiores) e linhas de saturação de cálcio-aluminatos sólidos (inferiores). Quanto menor o teor de  $O_2$  mais estreita será a janela, entretanto menos Ca adicionado será necessário (JANKE, 2000; HOLAPPA, 2003; MEDEIROS, 2014).



**Figura 6** - Efeito do oxigênio e do enxofre na janela líquida à 1550 °C

Fonte: HOLAPPA, 2003.

As formas de adição de cálcio no banho, geralmente, são realizadas na forma de ligas como  $\text{CaC}_2$ ,  $\text{CaAl}$  ou ainda  $\text{CaSi}$  que é a mais utilizada (BIELEFELDT, 2005).

## 2.4 Modelo de regressão logística

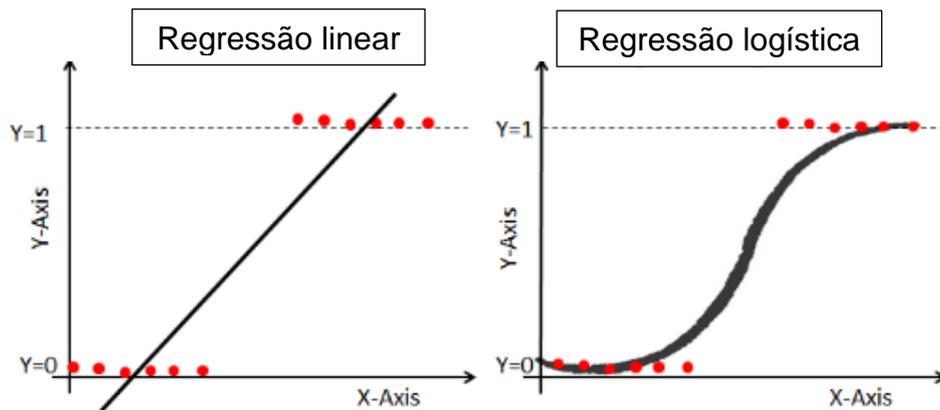
A modelagem para problemas de classificação tem o objetivo de prever uma saída qualitativa, dadas variáveis explicativas. A seguinte subseção aborda sobre algoritmo de regressão logística, modelo aplicado no presente trabalho, e as metodologias para a estimativa da assertividade de modelos serão abordadas na próxima seção.

A regressão logística, uma das ferramentas mais populares para resolver tarefas de classificação, sempre recebeu grande interesse de estatísticos e pesquisadores de *machine learning*, que pode ser definido como uma estratégia computacional para prever automaticamente um resultado específico para problemas genéricos. A regressão logística está dentro do grupo de modelos que se chama de Modelos Generalizados Lineares (GLM) (HOSMER JR et al, 2013; JORDAN, MITCHELL, 2015).

Dados de ocorrência/não-ocorrência de um fenômeno de interesse são amplamente encontrados em várias disciplinas. Esse tipo de variável é conhecido como binário ou dicotômico e o acontecimento deste evento é representado pela variável aleatória  $Y$  e, geralmente, registramos a ocorrência por  $Y=1$  e a não ocorrência por  $Y=0$  (BELL et al.,1994; ALBERINI, 1995; ARANA, LEON, 2005).

A regressão logística, apesar de seu nome, é um modelo linear de classificação e não de regressão. Ela também é conhecida na literatura como regressão *logit*, classificação máxima de entropia (*MaxEnt*) ou de classificador *log-linear*. Neste modelo, as probabilidades que descrevem os resultados possíveis de um único teste são modeladas usando uma função logística. Comparando com a regressão linear, que oferece uma saída contínua e é estimada usando a técnica de mínimos quadrados ordinários (OLS), a regressão logística oferece uma saída constante, utilizando a abordagem de máxima verossimilhança (MLE). Maximizar a função de verossimilhança determina os parâmetros com maior probabilidade de produzir os

dados observados. Do ponto de vista estatístico, o MLE define a média e a variância como parâmetros na determinação dos valores paramétricos específicos para um determinado modelo. Esse conjunto de parâmetros pode ser usado para prever os dados necessários em uma distribuição normal. Na Figura 7 é possível observar um gráfico de comparação entre as duas regressões.



**Figura 7 - Comparativo entre regressão linear e logística**

**Fonte:** DataCamp Community

Há diferenças também na forma de calcular, visto que a equação utilizada para regressão linear é representada pela Equação 1:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (1)$$

Onde:

$\beta$  – Coeficiente da variável

$\beta_0$  – Termo independente (interceptador)

$X_i$  – Variáveis

Já a regressão logística é deduzida a partir da Equação 2, função de Sigmoid:

$$p = \frac{1}{1 + e^{-y}} \quad (2)$$

Onde:

$p$  – Probabilidade de o evento acontecer  
 $y$  – Função linear

Na aplicação do algoritmo do modelo de regressão logística, alguns parâmetros são adicionados como um regularizador/penalizador  $\|w\|^2$  controlado por um  $\lambda$  que auxiliar a controlar sua influência. A vantagem de se ter um penalizador é reduzir os coeficientes irrelevantes e limitar o impacto da colinearidade. A Equação 3 mostra a equação de Sigmoid adaptada e a Equação 4 como é realizado o cálculo da regressão logística pelo algoritmo (e.g., CARRARA et al.; 1991, ATKINSON, MASSARI, 1998; CHUNG, FABBRI, 1999; LEE, MIN, 2001; DAI, LEE, 2002, 2003; OHLMACHER, DAVIS, 2003; CAN et al., 2005; AYALEW, YAMAGISHI, 2005; CHANG et al., 2007A, 2007B; GRECO et al., 2007).

$$p(y_i|x_i) = \frac{1}{1 + e^{-y_i w^2 x_i}} \quad (3)$$

$$\min_w \frac{\lambda}{2} \|w\|^2 + \sum_{x_i} \log(1 + e^{-y_i w^2 x_i}) \quad (4)$$

Onde:

$w$  – Vetor regularizador determinado durante o treino  
 $\lambda$  – Controlador do penalizador  
 $x_i$  – Variável do treino  
 $y_i$  – Variável do teste  
 (YANG, LOOG, 2018).

Para treinar um modelo de classificação, os dados rotulados são essenciais quando uma estrutura de treinamento/teste é adotada, e seu desempenho de classificação depende tanto do tamanho quanto da qualidade dos assuntos de treinamento usados para a aprendizagem. Quando treinamos um classificador de forma ativa, precisamos anotar os dados não rotulados e recrutá-los para o conjunto de treinamento, o que pode ser feito com as informações de um modelo construído nos dados rotulados no estágio atual. Na literatura, observa-se que as pessoas geralmente conseguem aprender um modelo econômico satisfatório com tal procedimento. Existem muitos índices de desempenho de classificação, e o processo de seleção das variáveis deve depender do índice alvo. Um exemplo deste índice alvo,

é o procedimento de aprendizagem ativa que maximiza a área sob a curva ROC (AUC) (COHN et al., 1994; HSU, 2010; SETTLES, 2011, 2012).

## 2.5 Validação cruzada estratificada (*stratified cross-validation*)

Para obtenção de uma estimativa da capacidade preditiva do modelo, pode-se seccionar a base total em três partes que não precisam, necessariamente, ter tamanhos iguais, denomina-se de treino, validação e teste. Utiliza-se a base de treino para construção efetiva do modelo, enquanto a de validação é utilizada para selecionar o nível correto de flexibilidade – para que se possa minimizar o erro ocasionado pela variância e pelo *bias*, que é a tendência de um processo de medição superestimar ou subestimar o valor de um parâmetro. Já a base de teste, permite uma avaliação final da capacidade do modelo (HASTIE et al., 2008).

Esta metodologia apresenta desvantagens (JAMES et al., 2013).

- (1) Subutilização da informação que está contida nos dados. Visto que algoritmos, geralmente, apresentam piora com uma quantidade menor de dados, e portanto, as bases de validação e de teste representam um certo desperdício de informação;
- (2) Pode ser variável, pois o modelo selecionado pode variar com as amostras utilizadas para validação e para treino.

Objetivando-se atenuar essa contrariedade, utiliza-se, comumente, o método *K-fold cross-validation*, que consiste em dividir a base randomicamente em K partes de tamanhos iguais. Sendo uma parte para validação e o restante para ser treinado. Este procedimento é, então, repetido para todas K partes. Há ainda uma variação deste método, que é o *stratified cross-validation*. Parecido com o *k-fold cross-validation*, dividindo a base em K partes, entretanto, os grupos são estratificados para conterem aproximadamente a mesma proporção das classes da base original em cada subparte (KOHAVI, 1995).

O valor do K pode afetar diretamente na elaboração do modelo. Visto que com a segregação em muitos subgrupos, o modelo tende a apresentar menor *bias*, entretanto, pode apresentar uma alta variância pois as bases de treinos podem ser

muito semelhantes. Todavia, geralmente, o erro estimado pelo *cross-validation* ainda é menor que o erro em bases de teste. A vantagem deste método é poder alcançar uma alta performance devido ao nível de flexibilidade adequado ao modelo, isto é, consegue identificar os parâmetros que atenuam o erro do modelo quando aplicado à dados futuros (EFRON, TIBSHIRANI, 1995).

Autores ainda salientam que, técnicas de seleção de variáveis devem ocorrer dentro do *cross-validation* e não externamente (AMBROISE, MCLACHLAN, 2002).

## 2.6 Avaliação do modelo de classificação

Técnicas de classificação têm sido aplicadas em vários campos da ciência. Nos modelos de classificação, os dados de treinamento são usados para construir um modelo preditivo de classificação, prever o rótulo de classe para uma amostra teste. É de extrema importância saber a metodologia certa para medir a assertividade do modelo, a fim de minimizar os erros que os modelos possam deixar passar. A análise de tais métricas e sua significância devem ser interpretadas corretamente para avaliar o algoritmo de aprendizado. A maioria dessas medidas são escalares e algumas delas também gráficas (THARWAT, 2018).

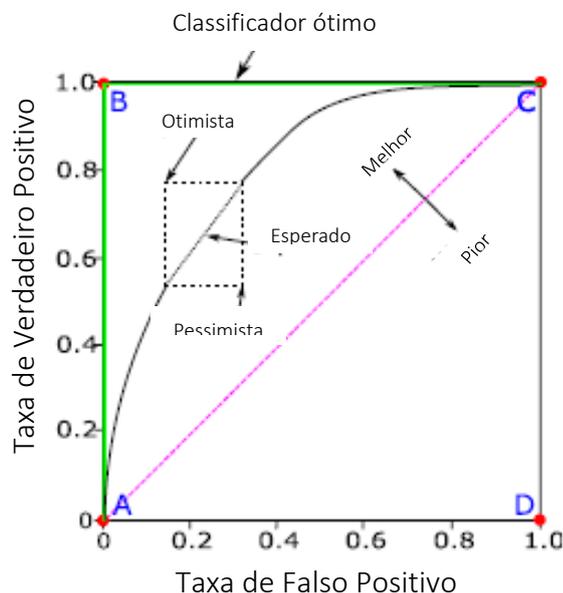
Para entender as medidas de avaliação de classificação que foram utilizadas como avaliação do modelo preditivo deste trabalho, faz-se necessário uma revisão. São apresentadas nas seguintes subseções, um esclarecimento sobre o assunto.

### 2.6.1 Curva ROC, área abaixo da curva (AUC) e acurácia

Além da metodologia de treinamento ser essencial para uma boa performance, utilizar o indicador adequado para medir a capacidade de classificação do modelo é de suma importância. Um indicador comumente utilizado é a área sob a curva ROC, vulgo AUC.

A curva ROC (*receiver operating characteristic curve*) é um gráfico bidimensional em que a taxa de verdadeiros positivos representa o eixo das ordenadas e a taxa de falsos positivos representa o eixo das abscissas. A fins de entendimento do gráfico, há quatro pontos importantes, conforme mostrado na Figura

8. O ponto A, situado na origem, representa a estratégia de considerar todos como negativos, isto é, todas as amostras negativas estão classificadas corretamente. O ponto C, situado em (1,1), representa o contrário, indicando todos os casos como positivos, então todas as classificações positivas estão corretas. O ponto D, situado em (1,0), representa uma classificação onde todos os positivos e negativos estão classificados incorretamente. O ponto B, situado em (0,1), representa o ponto ideal, onde todas as classificações estão corretas. A reta  $\overline{AC}$  representa uma classificação completamente aleatória, abaixo dela teremos uma classificação inaproveitável (PROVOST, FAWCETT, 1997; ZOU, 2002).



**Figura 8** – Curva ROC com pontos importantes.

**Fonte:** THARWAT, 2018.

Comparar diferentes classificadores na curva ROC não é uma tarefa tão fácil, pois não há um valor escalar que represente a performance esperada. Entretanto, a área abaixo da curva ROC (AUC) é uma métrica usada para quantificar esta performance. O valor do AUC é sempre calculado entre 0 e 1 e representa a probabilidade de um par aleatório de casos ser corretamente ordenado em função da sua classe.

A Equação 5 mostra o cálculo do AUC (área sob a curva ROC) (HANLEY, MCNEIL, 1982).

$$AUC = \int_0^1 TVP(x)dx \quad (5)$$

Onde:

$TVP$  – Taxa dos verdadeiros positivos

$x$  – Taxa dos falsos positivos

Autores destacam a importância de ordenar adequadamente os casos para formular estratégias do mundo real e também que o AUC é uma medida mais consistente e discriminante frente a outras, como por exemplo a acurácia. O AUC mostra várias propriedades desejáveis, como sensibilidade nos testes de variância e independência do limite de decisão (dado que a curva incorpora todos os limites e invariabilidade com o desequilíbrio das classes) a qual afeta vigorosamente a acurácia (BRADLEY, 1997; LING et al., 2003).

A acurácia é o indicador mais comumente utilizado para classificar performance de modelo e é definido pela relação do que foi classificado corretamente pelo total de amostras, como mostra a Equação 6:

$$Acurácia = \frac{VP + VN}{VP + VN + FP + FN} \quad (6)$$

Onde:

$VP$  – Verdadeiros positivos

$VN$  – Verdadeiros negativos

$FP$  – Falsos positivos

$FN$  – Falsos negativos

Como fora abordado anteriormente, a acurácia é sensível a dados não balanceados. Por isso que se deve ter cuidado ao utilizá-la como único indicador de performance (GARCIA, 2009).

### 2.6.2 Teste de KS (Kolmogorov–Smirnov test)

O teste de KS, amplamente utilizado para medições estatísticas, é um teste não paramétrico referente a igualdade de distribuições de probabilidades unidimensionais contínuas que podem ser utilizadas a fim de comparação com uma distribuição de probabilidade de referência. Isto é, este teste quantifica a distância entre a função de distribuição empírica que foi prevista e a função de distribuição cumulativa da distribuição de referência (o que realmente aconteceu). Quanto maior for esta distância entre as funções, maior será o grau de separação, por conseguinte maior será a confiabilidade da função de predição (YOUNG, 1977).

A Tabela 1 abaixo mostra um exemplo de distribuição referente a duas classes (1 e 2), onde obteve-se um KS de 52%. Baseada nela, é possível observar o gráfico do KS na Figura 9.

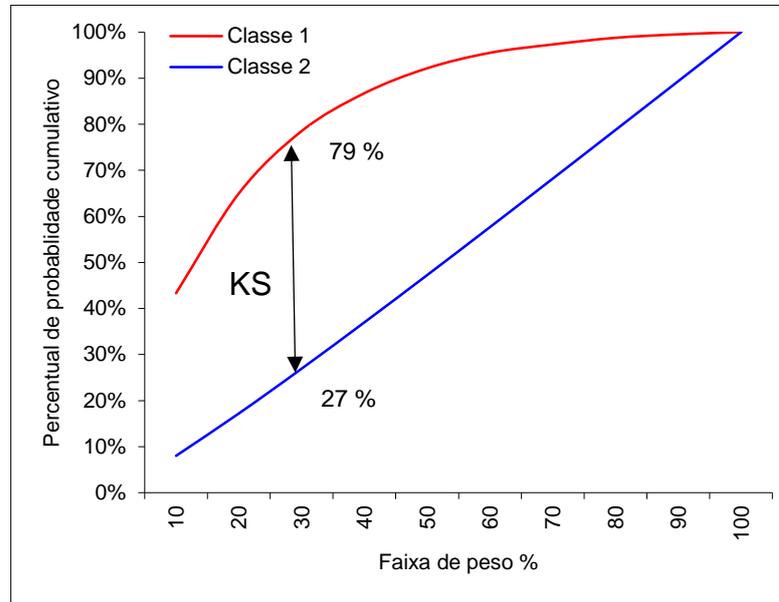
**Tabela 1** – Exemplo de tabela de distribuição de duas classes para determinação do KS.

Pesos %	Valores absolutos		Perc. Acumulado		KS
	Classe 1	Classe 2	Classe 1	Classe 2	Acum.
10	3.728	14.207	43%	8%	35%
20	1.864	16.305	65%	17%	48%
30	1.165	17.269	79%	27%	<b>52%</b>
40	708	17.839	87%	37%	50%
50	466	18.200	92%	47%	45%
60	289	18.472	96%	58%	38%
70	156	18.657	97%	68%	29%
80	124	18.748	99%	79%	20%
90	65	18.815	100%	89%	10%
100	41	18.863	100%	100%	0%
<b>Total</b>	<b>8.606</b>	<b>177.375</b>	<b>100%</b>	<b>100%</b>	<b>52%</b>

**Fonte:** o autor.

A Figura 9 exemplifica a máxima distância de separação entre as funções, isto é, o valor do KS encontrado. O eixo das ordenadas mostra a probabilidade acumulada, já o eixo das abcissas mostra a base de dados distribuídas em 10 faixas de pesos

iguais ordenadas pela classe, em que as faixas mais baixas apresentam maior quantidade de casos na classe 1 e nas maiores faixas apresentam maior quantidade de casos na classe 2. O resultado onde houve uma melhor separação – o KS propriamente dito – foi na terceira faixa.



**Figura 9** – Gráfico do KS para duas classes.

Fonte: o autor.

Como o resultado de confiabilidade depende do tamanho da amostra, KS superiores a 30% com amostras grandes são resultados considerados confiáveis no que diz respeito a predição de modelos econômicos. Este exemplo mostra um KS de 52%, valor considerado bom para uma amostra de mais de 180 mil casos.

### 2.6.3 Matriz de confusão binária

Também conhecida por matriz de erro, a matriz de confusão é uma tabela frequentemente utilizada para descrever a performance de um modelo de classificação. Ela permite identificar facilmente a confusão entre as classes que foram erroneamente ordenadas, pois mostra resumidamente o número de predições corretas e incorretas, divididas pela classe. Esta é vantagem do seu uso, pois nos dá uma visão não apenas dos erros cometidos pelo modelo, mas, mais importante, dos tipos de erros que foram cometidos (XU, 1992; SHAFFI, 2011; THARWAT, 2018).

Para uma matriz de confusão binária, só se tem duas classes, que seriam verdadeiro e falso, isto é, uma matriz de tamanho 2 x 2. A Figura 10 representa uma matriz de confusão em que no quadrante superior esquerdo temos os valores verdadeiros positivos, onde o modelo acertou a classificação positiva. No quadrante inferior direito estão os verdadeiros negativos, em que o modelo acertou a classificação negativa. No quadrante inferior esquerdo estão os falsos negativos, isto é, onde o modelo previu negativo, porém a classe era positiva. Já no quadrante superior direito estão os falsos positivos, em que o modelo previu positivo, entretanto a classe era negativa. O melhor cenário possível para esse teste, é quando os quadrantes falsos positivos e falsos negativos apresentam o menor número de casos.

		Valor Previsto	
		Positivo	Negativo
Valor Verdadeiro	Negativo	Verdadeiros Positivos	Falsos Negativos
	Positivo	Falsos Positivos	Verdadeiros Negativos

**Figura 10** – Exemplo de matriz de confusão binária.

Fonte: LAMBLET, 2018.

#### 2.6.4 Precisão, revocação e F1-Score

Precisão e revocação são amplamente utilizadas para avaliar a performance de classificação. Ambos indicadores estão relacionados pela quantidade de verdadeiros e falsos positivos e negativos.

A precisão, também chamada de valor de predição positiva ou valor de predição negativa, dependendo da classe, representa a proporção de amostras positivas (ou

negativas) que foram corretamente classificadas pelo total de positivos (ou negativos) previstos.

As Equações 7 e 8 mostram as precisões para as duas classes (POWERS, 2011).

$$\text{Precisão positiva (classe bom)} = \frac{VP}{VP + FP} \quad (7)$$

$$\text{Precisão negativa (classe ruim)} = \frac{VN}{VN + FN} \quad (8)$$

Onde:

$VP$  – Verdadeiros positivos

$VN$  – Verdadeiros negativos

$FP$  – Falsos positivos

$FN$  – Falsos negativos

A revocação (*recall* ou *hit rate*), também chamada de sensibilidade (*sensitivity*), é um indicador de classificação que representa amostras positivas (ou negativas) corretamente classificadas pelo total do número de amostras positivas (ou negativas). Ao contrário da precisão, o denominador não é apenas a quantidade de amostras (positivas ou negativas) previstas. As Equações 9 e 10 expressam o cálculo deste indicador para cada classe (SKOLOVA, 2009).

$$\text{Revocação (classe bom)} = \frac{VP}{VP + FN} = \frac{TP}{P} \quad (9)$$

$$\text{Revocação (classe ruim)} = \frac{VN}{VN + FP} = \frac{TN}{N} \quad (10)$$

Onde:

$VP$  – Verdadeiros positivos

$VN$  – Verdadeiros negativos

$FP$  – Falsos positivos

$FN$  – Falsos negativos

$P$  – Positivos

$N$  – Negativos

O F1-Score, também chamado de *F-measure*, representa a média harmônica da precisão com a revocação. O valor pode ir de 0 até 1, onde valores maiores

representam um indicador com melhor performance de classificação. Importante salientar que, assim como para precisão e para revocação, há um F1-Score para cada classe. Uma solução para encontrar a melhor forma de avaliar o modelo de classificação é ponderar para ambas as classes.

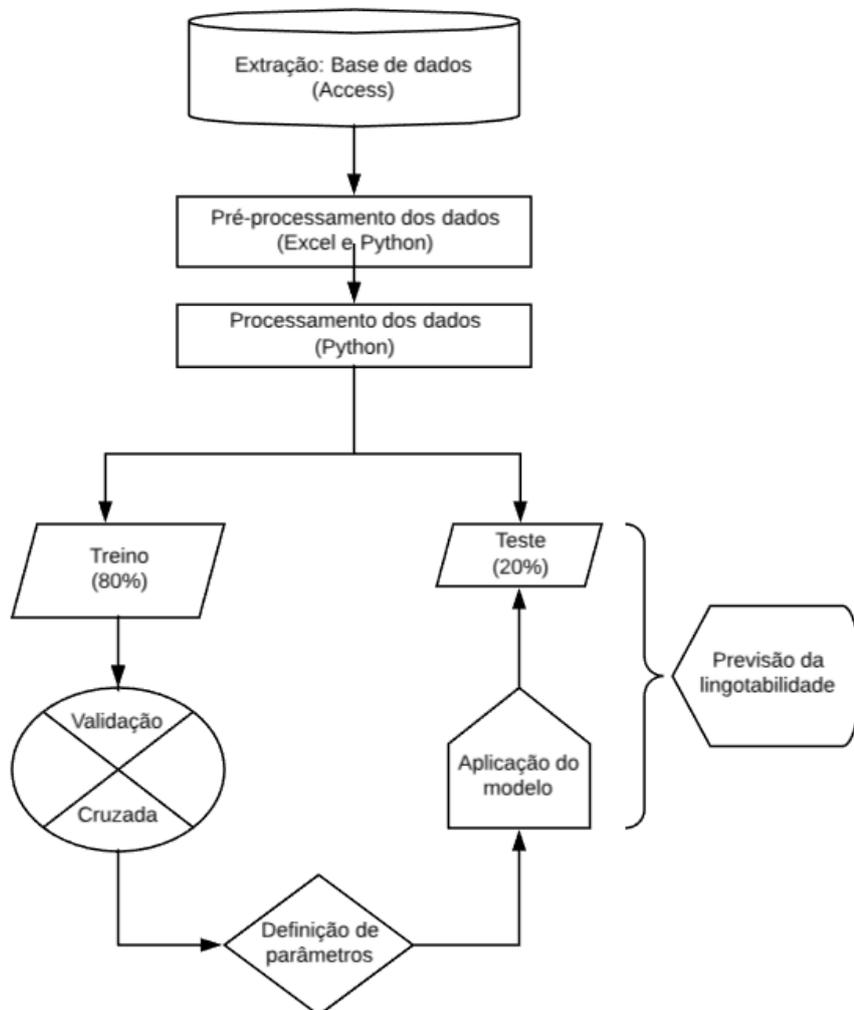
A Equação 11 mostra como é feito o cálculo para o *F-measure* (SKOLOVA, 2009).

$$F1 - Score = 2 * \frac{Revocação * Precisão}{Revocação + Precisão} \quad (11)$$

### 3 METODOLOGIA

Nesta seção será abordada, detalhadamente, a metodologia desde as extrações das bases até o desenvolvimento do modelo estatístico propriamente dito bem como a obtenção dos resultados.

A metodologia deste trabalho envolveu 7 etapas que seguiram uma ordem cronológica. O fluxograma resumido pode ser verificado na Figura 11, o qual ilustra a ordem das etapas realizadas, que serão comentadas na sequência.



**Figura 11** – Fluxograma da metodologia aplicada no trabalho.

### 3.1 Bases de dados

As bases de dados foram coletadas referentes a um histórico dos últimos três anos de corridas realizadas em uma usina de aços longos especiais, extraídas pelo *software* de armazenamento de dados Access, o que totalizou uma amostra de mais de 20 mil corridas, das quais 17.460 foram utilizadas para o desenvolvimento do modelo estatístico de predição. A variável chave para agregar todas as bases foi o nº da corrida. A seguir são apresentadas as bases empregadas neste estudo.

#### 3.1.1 Base: Parâmetros de processo

Representa a base contendo parâmetros de processo como: qualidade do aço, seção do tarugo, produto, tipo de distribuidor, posição, sequência do distribuidor, sequência total do distribuidor, processo de escória (Alta/Baixa alumina, Fluorita, Lafarge ou Wollastonita), oxigênio ativo de vazamento, temperatura objetivada no Forno-Panela, temperatura real no Forno-panela, temperatura objetivada no desgaseificador a vácuo, temperatura real no desgaseificador a vácuo, temperatura objetivada, temperatura na panela com 60 t de carga, temperatura na panela com 50 t de carga, temperatura na panela com 40 t de carga, temperatura na panela com 30 t de carga, temperatura na panela com 20 t de carga, temperatura na panela com 10 t de carga, temperatura na panela com 5 t de carga, temperatura *liquidus* e temperatura de sobre aquecimento (superheat).

Embora alguma das variáveis tenham pouca amplitude de parametrização, a fim de determinar suas influências no processo, todas foram utilizadas.

#### 3.1.2 Base: Adição de elementos à escória

Representa a base contendo elementos e compostos adicionados à escória em diferentes etapas do processo de fabricação do aço, como: alumina, CaF<sub>2</sub> caixa, CaF<sub>2</sub> em pedra, CaF<sub>2</sub> em pó, carbureto no FEA, carbureto, cal, FeSi em pó, MgO, O<sub>2</sub>, SiO<sub>2</sub>, Wolastonita, flor de enxofre, tijolo de enxofre fio de enxofre, além de outros produtos reguladores de acidez da escória.

Embora seja uma base rica em informações para determinar a composição da escória, ela não influenciou positivamente para o modelo e, portanto, não foi utilizada para o desenvolvimento do modelo, como pode ser visto na próxima seção.

### 3.1.3 Base: Adição Al e CaSi

Representa a base contendo a quantidade em metros que foi adicionado de lingote de alumínio ou cálcio-silício durante a corrida.

### 3.1.4 Base: Posição da válvula tampão

Representa a base contendo a posição da válvula tampão a cada cinco minutos de cada uma das três válvulas referentes aos três veios do distribuidor.

### 3.1.5 Base: Aços sucateados por problemas no lingotamento

Representa a base contendo 155 corridas nas quais foram registrados problemas de lingotabilidade e a respectiva quantidade de aço (em quilogramas) que foi sucateada.

### 3.1.6 Base: Provas de escória

Representa a base contendo a composição química da prova que foi coletada em determinada etapa do processo. São, ao todo, 12 etapas em que cada uma é denominada por: E0, E1, E2, E3, E4, E5, E6, E7, E8, E9, E10 e E11.

Foram analisadas e quantificadas as seguintes composições:  $\text{Al}_2\text{O}_3$  (alumina),  $\text{CaF}_2$  (fluoreto de cálcio),  $\text{CaO}$  (óxido de cálcio),  $\text{Cr}_2\text{O}_3$  (óxido de cromo),  $\text{FeO}$  (óxido de ferro),  $\text{FeO} + \text{MnO}$ ,  $\text{K}_2\text{O}$  (óxido de potássio),  $\text{MgO}$  (óxido de magnésio),  $\text{MnO}$  (óxido de manganês),  $\text{Na}_2\text{O}$  (óxido de sódio),  $\text{P}_2\text{O}_5$  (pentóxido de difósforo), S (enxofre),  $\text{SiO}_2$  (sílica),  $\text{TiO}_2$  (dióxido de titânio) e  $\text{V}_2\text{O}_5$  (pentóxido de vanádio). Também é medida a basicidade ( $\text{CaO}/\text{SiO}_2$ ).

### 3.1.7 Base: Provas de aço

Representa a base contendo a composição química da prova que foi coletada em determinada etapa do processo. São, ao todo, 13 etapas em que cada uma é denominada por: P0, P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, P11 e PF.

Foram analisadas e quantificadas as seguintes composições: carbono, silício, manganês, fósforo, enxofre, cromo, níquel, molibdênio, cobre + estanho, alumínio (solubilizado), alumínio (total), arsênio, boro (solubilizado), boro (total), cálcio, cobalto, cobre, hidrogênio, nitrogênio, nióbio, chumbo, antimônio, selênio, estanho, telúrio, titânio, titânio (total), vanádio, tungstênio, zinco e zircônio.

### 3.2 Pré-processamento das bases de dados

Todas as bases que foram extraídas apresentavam diferentes formatos, nomenclaturas e tamanhos. Por este motivo, necessitou-se trabalhá-las a fim de padronizar os dados para imputar no código de programação e posterior desenvolvimento do modelo.

A base de adição de elementos na escória foi agregada por corrida, de forma que se somaram os elementos adicionados durante o processo, independente da etapa em que ele foi inserido.

As bases de provas de escória bem como a base de provas de aço foram selecionadas apenas as etapas mais relevantes do processo. Para escória as etapas E6 e E7, que correspondem às etapas do Forno-panela, e a etapa E11, que corresponde à etapa no desgaseificador. Já para as provas de aço, as etapas P11 (desgaseificador) e PF (lingotamento contínuo). A fim de manter-se a variável chave (corrida) sem duplicatas, cada composição química de cada etapa foi computada como uma coluna.

A base posição da válvula tampão foi agregada de forma que fosse considerado somente a variação da válvula tampão durante o processo de lingotamento, ou seja, a posição final subtraída pela posição inicial.

A base de aço sucateados por problemas no lingotamento foi resumida categoricamente, em que as 155 corridas que apresentaram problemas, foram classificadas como classe “mau”, ou seja, lingotabilidade falha.

Por falta de registros sobre problemas relacionados à lingotabilidade, as demais 17.305 corridas foram classificadas de acordo com a variação da válvula tampão

durante o processo de lingotamento contínuo. Para isso, foi calculada uma variação de corte, ou seja, afirmou-se que a partir deste *delta* foram observados problemas relacionados ao lingotamento, portanto a corrida seria classificada como classe “mau” e as demais que ficaram abaixo do ponto de corte, como classe “bom”.

As Equações 12-19 apresentam os passos propostos para a determinação da classe a partir das variações das válvulas tampões.

a) Cálculo da média de variação das válvulas por corrida:

$$M = \frac{(V1 + V2 + V3)}{3} \quad (12)$$

Onde:

$V_i$  – Variação da válvula tampão em milímetros.

$M$  – Média de variação das válvulas da corrida.

b) Desvio padrão da média das válvulas:

$$\sigma = \sqrt{\frac{\sum(M - \bar{M})^2}{N}} \quad (13)$$

Onde:

$M$  – Média de variação das válvulas da corrida.

$\bar{M}$  – Média da média de variação das válvulas de todas as corridas.

$N$  – Número de corridas.

$\sigma$  – Desvio padrão.

c) Obtenção da média máxima de cada válvula:

$$M(1) = \frac{\sum(V_1)}{N} \quad (14)$$

$$M(2) = \frac{\sum(V_2)}{N} \quad (15)$$

$$M(3) = \frac{\sum(V_3)}{N} \quad (16)$$

Onde:

M(i) – Média de todas as corridas do tampão.

N – Número total de corridas.

d) Máximo valor médio das válvulas:

$$M(2) > M(3) > M(1) \quad (17)$$

$$M_{(max)} = 4.37626$$

Onde:

M<sub>(max)</sub> - Máximo valor médio das válvulas

f) Definição classe:

$$\underline{\text{BOM}}: M < (\sigma + M_{(max)}) \quad (18)$$

$$\underline{\text{MAU}}: M > (\sigma + M_{(max)}) \quad (19)$$

### 3.3 Plataforma computacional

Para o trabalho de pré-processamento, processamento e desenvolvimento do modelo, foram necessários equipamentos e instalações específicas, tanto de *hardware* quanto de *software*.

#### 3.3.1 Hardware

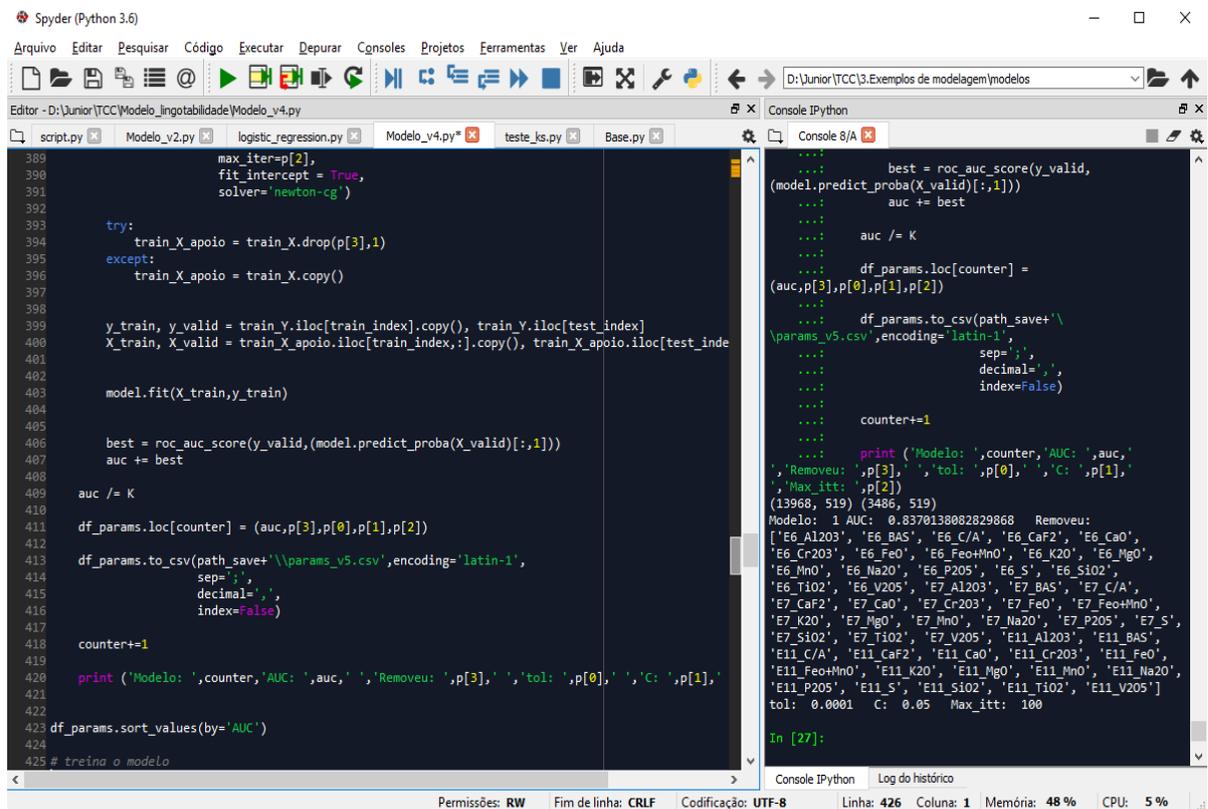
O equipamento utilizado foi um computador pessoal com configurações básicas, sendo elas:

- Processador Intel i7-7500U 2.7 GHz 4 MB Cache
- Memória RAM DDR4 16 GB
- Placa de Vídeo NVIDIA GeForce 940 MX 4 GB
- Sistema Operacional Windows 10

### 3.3.2 Software

Para o pré-processamento das bases de dados foi utilizado o *software* Microsoft Excel, que proporciona uma ampla manipulação com tabelas de até um milhão de linhas.

Para o processamento das variáveis e desenvolvimento do modelo estatístico, foi utilizada a linguagem de programação Python, que possui grande riqueza de pacotes e bibliotecas estatísticas para análise de dados bem como implementação de modelos *machine learning*. A Figura 12 representa o ambiente de desenvolvimento *Spyder*, onde é utilizada a linguagem de programação Python.



```

389         max_iter=p[2],
390         fit_intercept = True,
391         solver='newton-cg')
392
393     try:
394         train_X_apoio = train_X.drop(p[3],1)
395     except:
396         train_X_apoio = train_X.copy()
397
398
399     y_train, y_valid = train_Y.iloc[train_index].copy(), train_Y.iloc[test_index]
400     X_train, X_valid = train_X_apoio.iloc[train_index,:].copy(), train_X_apoio.iloc[test_inde
401
402
403     model.fit(X_train,y_train)
404
405
406     best = roc_auc_score(y_valid,(model.predict_proba(X_valid)[:,:].T))
407     auc += best
408
409     auc /= K
410
411     df_params.loc[counter] = (auc,p[3],p[0],p[1],p[2])
412
413     df_params.to_csv(path_save+'\\params_v5.csv',encoding='latin-1',
414                     sep=',',
415                     decimal=',',
416                     index=False)
417
418     counter+=1
419
420     print ('Modelo: ',counter,'AUC: ',auc,' ', 'Removido: ',p[3],' ', 'tol: ',p[0],' ', 'C: ',p[1],
421
422
423     df_params.sort_values(by='AUC')
424
425 # treina o modelo

```

```

...
...     best = roc_auc_score(y_valid,
... (model.predict_proba(X_valid)[:,:].T))
...     auc += best
...
...     auc /= K
...
...     df_params.loc[counter] =
(auc,p[3],p[0],p[1],p[2])
...
...     df_params.to_csv(path_save+'\\
\\params_v5.csv',encoding='latin-1',
...     sep=',',
...     decimal=',',
...     index=False)
...
...     counter+=1
...
...     print ('Modelo: ',counter,'AUC: ',auc,'
', 'Removido: ',p[3],' ', 'tol: ',p[0],' ', 'C: ',p[1],
(13968, 519) (3486, 519)
Modelo: 1 AUC: 0.8370138082829868 Removido:
['E6_Al2O3', 'E6_BAS', 'E6_C/A', 'E6_CaF2', 'E6_CaO',
'E6_Cr2O3', 'E6_FeO', 'E6_Feo+MnO', 'E6_K2O', 'E6_MgO',
'E6_MnO', 'E6_Na2O', 'E6_P2O5', 'E6_S', 'E6_SiO2',
'E6_TiO2', 'E6_V2O5', 'E7_Al2O3', 'E7_BAS', 'E7_C/A',
'E7_CaF2', 'E7_CaO', 'E7_Cr2O3', 'E7_FeO', 'E7_Feo+MnO',
'E7_K2O', 'E7_MgO', 'E7_MnO', 'E7_Na2O', 'E7_P2O5', 'E7_S',
'E7_SiO2', 'E7_TiO2', 'E7_V2O5', 'E11_Al2O3', 'E11_BAS',
'E11_C/A', 'E11_CaF2', 'E11_CaO', 'E11_Cr2O3', 'E11_FeO',
'E11_Feo+MnO', 'E11_K2O', 'E11_MgO', 'E11_MnO', 'E11_Na2O',
'E11_P2O5', 'E11_S', 'E11_SiO2', 'E11_TiO2', 'E11_V2O5']
tol: 0.0001 C: 0.05 Max_itt: 100
In [27]:

```

**Figura 12** – Ambiente de desenvolvimento *Spyder*.

**Fonte:** o autor

Ao observar a Figura 12, é possível observar parte do *script* que foi escrito para o desenvolvimento do modelo estatístico bem como o *output* de determinada parte do código. A interface é bastante amigável e a linguagem de programação utilizada é exclusivamente Python.

### 3.4 Processamento das bases de dados

Nesta etapa, as variáveis foram tratadas com a finalidade de aumentar a robustez dos algoritmos para a construção do modelo. Variáveis categóricas, como seção do tarugo, produto, distribuidor, posição, processo de escória, qualidade do aço, sequência e sequência total do distribuidor, foram transformadas em variáveis binárias (*dummy variables*), resultando em uma variável para cada categoria. Dessa forma, há um aumento na quantidade de variáveis bem como de melhora na performance do modelo, visto que a introdução de variáveis qualitativas é uma ferramenta que flexibiliza o modelo, sendo capaz de lidar com problemas em estudos empíricos (MISSIOL, 2007).

As variáveis contínuas como: oxigênio ativo no Forno Panela, temperaturas, *superheat*, teores de elementos, tempo de processo e quantidade de elemento adicionado foram normalizadas. Em geral, os algoritmos de *machine learning* se beneficiam da padronização do conjunto de dados, pois se alguns *outliers* estiverem presentes no conjunto, normalizá-los impedirá sua influência negativa sobre o modelo. Na prática, para realizar a normalização, dois objetivos são impostos: desvio padrão da variável ser igual a 1 e a média dos valores da variável ser igual a 0. Para isso, as Equações 18-21 ilustram os cálculos:

a) Normalização dos valores das variáveis contínuas:

$$\mu = \frac{\sum(x_i)}{N} \quad (18)$$

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}} \quad (19)$$

$$X = \frac{x - \mu}{\sigma} \quad (20)$$

$$X_{NORM} = \frac{(X - X_{MIN})}{X_{MAX} - X_{MIN}} \quad (21)$$

Onde:

$\mu$  – Média dos valores da variável contínua.

$\sigma$  – Desvio padrão.

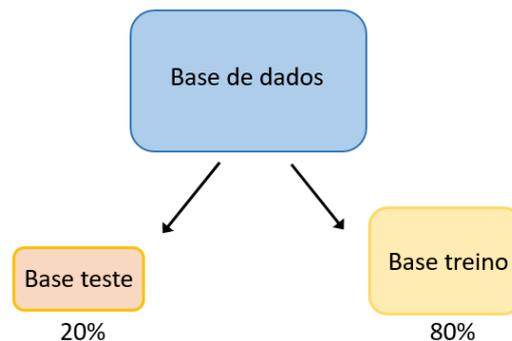
$x_i$  – Valor de cada corrida referente à variável contínua.

$X$  – Número uniformizado.

$X_{NORM}$  – Número normalizado, a partir da diferença dos valores mínimos e máximos da variável contínua.

### 3.5 Separação das bases de treino e teste

Para a fase denominada de treinamento, os dados de entrada e a saída são apresentados conjuntamente. Este treino é realizado até que o modelo aprenda a mapear os dados e a identificar os padrões entre as entradas e as saídas. Quando o modelo, apresentar uma assertividade considerada apta para realizar previsões, ele é validado e então aplicado na base de teste. Para o presente trabalho, a base total das 17.460 corridas foi dividida em 80 % para treino, que representam 13.968 corridas, e 20 % para teste, que representam 3.486 corridas. Na Figura 13, podemos ver um esquema representativo.



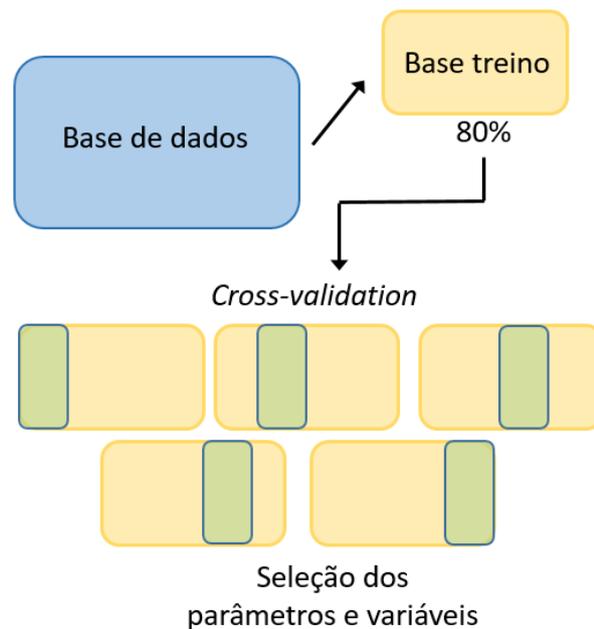
**Figura 13** – Esquema da divisão da base de dados.

### 3.6 Treino, validação e teste do modelo

Na fase de adequação de dados, foram testados diferentes parâmetros para o algoritmo com o objetivo de aumentar a assertividade do modelo. Também foram testadas as remoções de diferentes variáveis.

### 3.6.1 Validação cruzada estratificada (*stratified cross-validation*)

Utilizou-se a validação cruzada estratificada (*stratified cross-validation*). Em que a base foi dividida em cinco partes (quatro partes de treino e uma de teste para validação do modelo gerado), de tamanhos aproximadamente iguais e com semelhante proporção de classes. A Figura 14 ilustra uma representação esquemática do processo de validação cruzada. As comparações entre os modelos gerados com diferentes parâmetros e diferentes variáveis foram realizadas através do indicador AUC (KOHAVI, 1995).

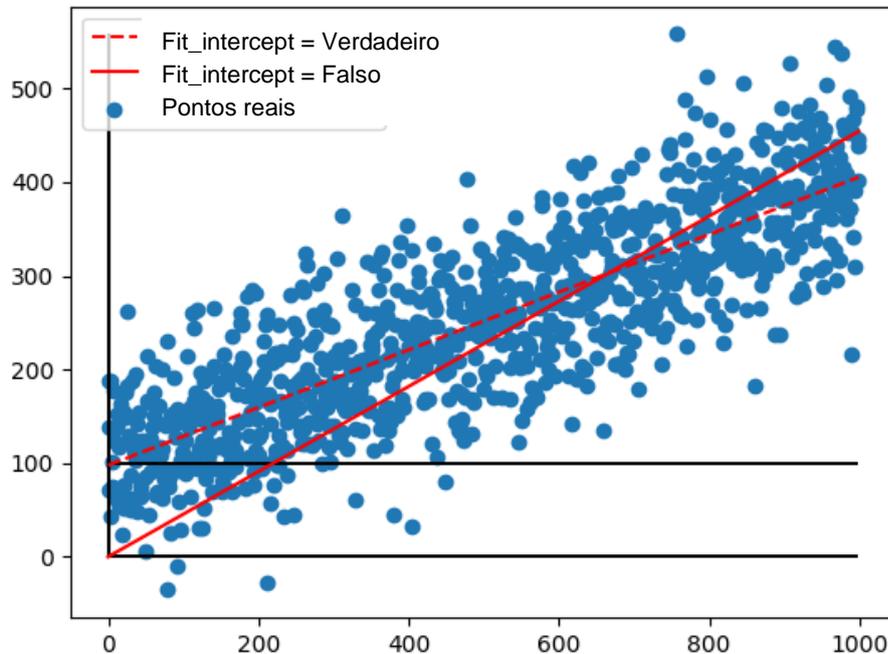


**Figura 14** – Esquema do treinamento da base via *cross-validation*.

Para a regressão logística quatro parâmetros foram testados: *fit\_intercept*, *tol*, *C* e *max\_iter*, descritos a seguir:

- *Tol*: tolerância para os critérios de parada, ou seja, o modelo para de procurar um mínimo (ou máximo) uma vez que esta tolerância seja alcançada. Tolerâncias testadas: [0.0001; 0.001].
- *Fit\_intercept*: decisão da regressão interceptar ou não a origem (0,0). Caso esteja marcado como verdadeiro (*Verdadeiro*), ela interceptará a origem e não haverá coeficiente linear. Caso esteja marcado como falso (*Falso*),

haverá um coeficiente que interceptará o eixo das ordenadas. Este parâmetro é bastante importante, pois influenciará no coeficiente da equação gerada pela regressão logística. Na Figura 15 há um exemplo que demonstra a interceptação.



**Figura 15** – Exemplo de equações da reta com e *sem fit\_intercept*.

- *C*: parâmetro de regularização, onde é aplicada uma penalidade para aumentar a magnitude dos valores dos parâmetros a fim de se reduzir o *overfitting*, isto é, evitar o ajuste de dados a um conjunto limitado que ocasionaria uma tendência que não reflete a realidade dos dados. A penalização geralmente é multiplicada por um coeficiente  $\lambda$ , e quanto maior este  $\lambda$ , menor é a probabilidade de que os parâmetros sejam aumentados em magnitude simplesmente para ajustar pequenas perturbações nos dados. Então o parâmetro *C* é alterado, onde  $C = 1 / \lambda$ . Parâmetros testados: [0.052; 0.053; 0.054].
- *Max\_iter*: máximo número de iterações que o modelo realiza a fim de convergir para um mínimo global. Parâmetros testados: [100; 200]

Foram removidos grupos de variáveis, a fim de testar a influência no modelo, bem como o aumento do indicador AUC. Os grupos de variáveis removidas foram:

tempos por processo, parâmetros de processo, provas de escória, provas de aço, adição de elementos na escória e qualidades do aço.

### 3.6.2 *Treinamento do modelo*

Definidos todos os parâmetros e variáveis a serem utilizadas, treinou-se o modelo com os 80 % do total da base (13.968 corridas). A partir do treino e da validação com resultados satisfatórios, o modelo apresentou-se apto para ser finalizado e ser testado na base de teste.

### 3.6.3 *Teste do modelo*

Com o modelo pronto foi possível testar a base teste, que em nenhum momento foi levada em consideração para desenvolvimento do modelo, a fim de se evitar o *overfitting*.

A partir deste momento, puderam-se extrair os resultados da aplicação do modelo. Acurácia, AUC, coeficientes, peso dos coeficientes, revocação (*recall*), F1-score, precisão (*precision*), curva ROC, teste de KS (Kolmogorov–Smirnov), matriz de confusão e gráfico de distribuição foram os indicadores escolhidos para fazer sua avaliação.

## 4 RESULTADOS E DISCUSSÃO

### 4.1 Validação cruzada estratificada (*stratified cross-validation*)

A Tabela 2 apresenta os resultados resumidos, ordenados do melhor para o pior, obtidos após 96 iterações no *cross-validation* alterando tanto os parâmetros da regressão logística quanto a remoção dos diferentes grupos de variáveis. A Tabela 2 completa encontra-se no apêndice A do presente trabalho.

**Tabela 2** – Principais resultados da validação cruzada estratificada.

Iteração	AUC	Removeu	Tol	C	max_iter	fit_intercept
1	0,8370	Provas escória	0,001	0,053	100	Verdadeiro
2	0,8359	Processo	0,001	0,052	100	Falso
3	0,8359	Qualidade	0,001	0,052	100	Falso
4	0,8354	Provas aço	0,0001	0,053	100	Verdadeiro

Percebe-se que a iteração com maior AUC apresentado, foi com os parâmetros  $C=0.053$ ,  $max\_iter=100$ ,  $fit\_intercept=Verdadeiro$  e  $tol=0.001$ . Já as variáveis retiradas que mais aumentaram o desempenho foram as provas de escória. Com a retirada das variáveis de processo, qualidade e provas de aço, o indicador AUC oscilou menos, o que representa uma baixa sensibilidade com estas respectivas variáveis.

Como o indicador AUC aumentou com a remoção das variáveis de prova de escória, isso mostra que a influência da medição da composição de escória nas etapas E6, E7 e E11 não foram relevantes para o modelo a fim de se determinar a lingotabilidade.

Notou-se ainda a indiferença da variação do parâmetro  $max\_iter$ , que modifica bastante o tempo de processamento do modelo, portanto fixá-lo em 100 torna-se o ponto ideal, visto que é o valor mínimo para esta regressão.

O  $fit\_intercept$  apresentou comportamento ambíguo, visto que na retirada das variáveis de provas de escória e provas de aço, quando usado como verdadeiro, o modelo apresentou um melhor desempenho. Em contrapartida, para a retirada das variáveis de processo ou qualidade, usá-lo como falso apresentou melhor

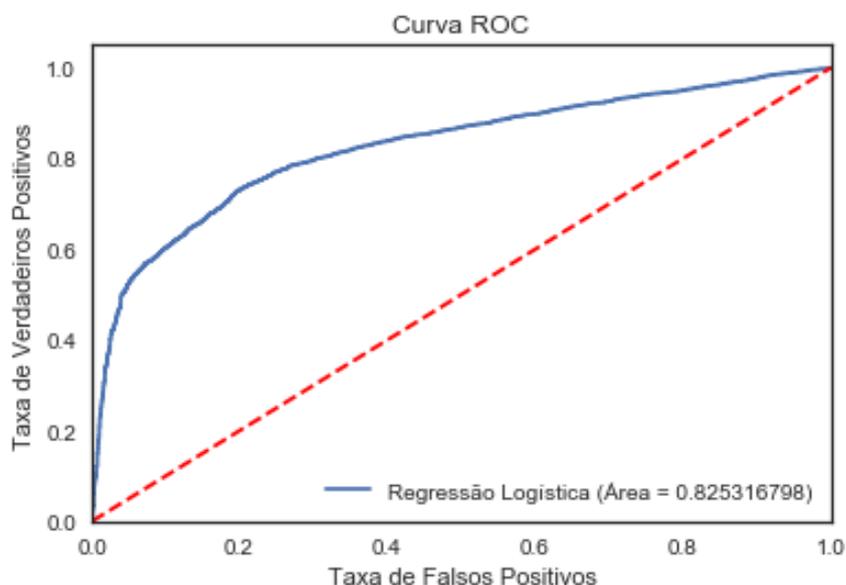
desempenho. O que pode ser explicado por um melhor *fit* graficamente, havendo independência de inclinação da reta.

O parâmetro de tolerância (*tol*) modificou pouco no valor do AUC, visto que sua influência só foi notada a partir do 6º algarismo depois da vírgula na validação com a remoção da variável de provas de escória, conforme podemos observar no apêndice A do trabalho. Visto que sua influência no processamento não é tão perceptível quanto o parâmetro *max\_iter*, pode-se optar pelo resultado que garantiu melhor performance.

O parâmetro C é o mais sensível para o modelo de regressão linear, este foi o motivo pelo qual foram testados valores tão próximos (0,052, 0,053 e 0,054). Percebeu-se que para as retiradas das variáveis de provas de escória e provas de aço, seu valor ótimo foi 0,053, já para remoção das variáveis de qualidade e processo, o melhor valor foi 0,052. O que significa um valor muito baixo de controlador de regularização da penalidade da regressão logística.

#### 4.2 Treinamento do modelo

Como o esperado, o AUC apresentado no treinamento da base de treino com os parâmetros apropriados ficou bastante próximo da validação cruzada estratificada, com valor de 0.8253, conforme pode ser visualizado na Figura 16.



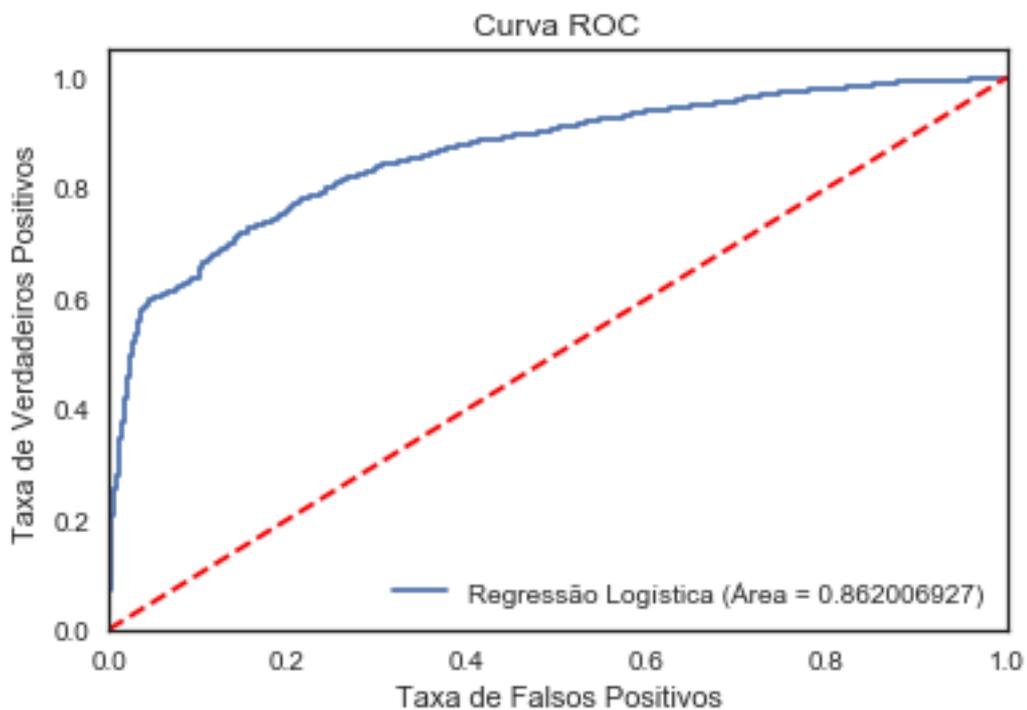
**Figura 16** – Curva ROC referente à base de treino.

Pela Figura 16, valores de AUC próximos à 0.50 sugerem uma distribuição aleatória, ou seja, que o modelo não tem uma capacidade suficiente de separação de classes. A acurácia da base de treino do modelo ficou em 0.825 e o AUC ficou em 0.825. As comparações e discussões serão feitas na próxima subseção.

#### 4.3 Teste do modelo

##### 4.3.1 Curva ROC, AUC e acurácia

Com a aplicação do modelo já treinado na base de teste, o resultado do AUC foi ainda mais satisfatório visto que alcançou um valor superior a 0.86, sendo ainda maior que na validação cruzada estratificada e treino. Na Figura 17 pode se observar a curva ROC bem como o valor da área abaixo da curva (AUC).



**Figura 17** – Curva ROC referente à base de teste.

Isso significa que o modelo apresentou mais de 86% de chance de conseguir distinguir as classes, ou seja, de acertar se houve problema com a lingotabilidade do aço ou não. Como comparação, na área da medicina para modelos de previsão, trabalha-se com AUC de aproximadamente 0.90. A acurácia no teste também teve um resultado satisfatório e superior ao do treino, 0.856 (TRÉPO; SHARMA, 2018)

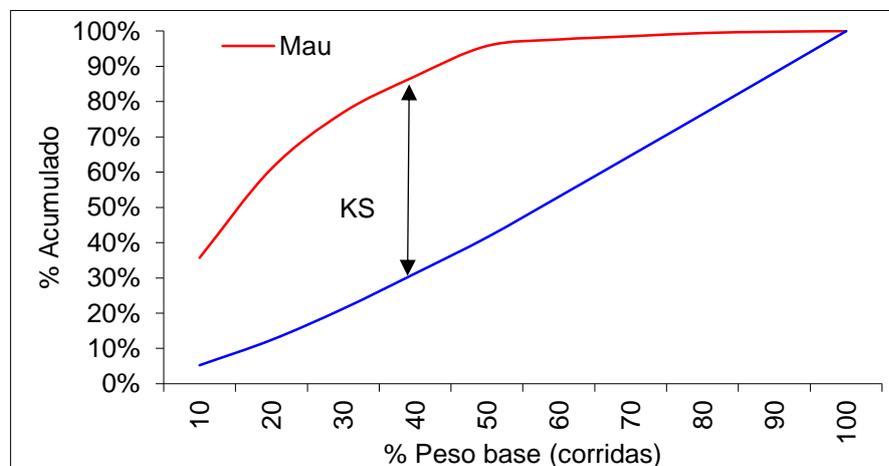
### 4.3.2 Teste de KS

A Tabela 3 apresenta os resultados do teste sobre a igualdade de distribuições de probabilidade contínuas e unidimensionais, o teste de KS.

**Tabela 3** – Tabela de distribuição do KS.

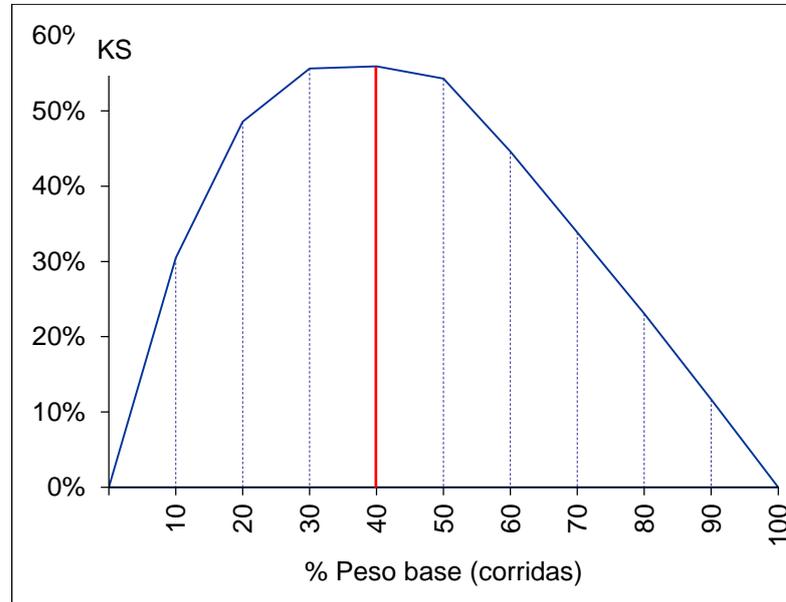
Pesos %	Bom	Mau	Percentual _Mau	Percentual _Bom	Perc_Mau_acumulado	Perc_Bom_acumulado	KS
10	154	195	36%	5%	36%	5%	30,5%
20	211	138	25%	7%	61%	12%	48,6%
30	261	87	16%	9%	77%	21%	55,6%
40	293	56	10%	10%	87%	31%	55,9%
50	301	47	9%	10%	96%	41%	54,3%
60	339	10	2%	12%	98%	53%	44,6%
70	343	5	1%	12%	99%	65%	33,8%
80	344	5	1%	12%	99%	76%	23,1%
90	346	2	0%	12%	100%	88%	11,7%
100	348	1	0%	12%	100%	100%	0,0%

Como o maior valor máximo entre as faixas é o resultado do KS, o resultado foi de 55,92%, que representa uma boa distância entre a função distribuição de referência e a distribuição prevista no modelo. Embora ainda seja uma base de amostra considerada, estatisticamente, pequena (3.486 corridas).



**Figura 18** – Gráfico do KS.

A Figura 18 representa a diferença de distância em cada faixa que foi segregada a base. Percebeu-se que a faixa onde houve uma melhor separação na faixa de 40% de peso, conforme destacado pela Figura 19.

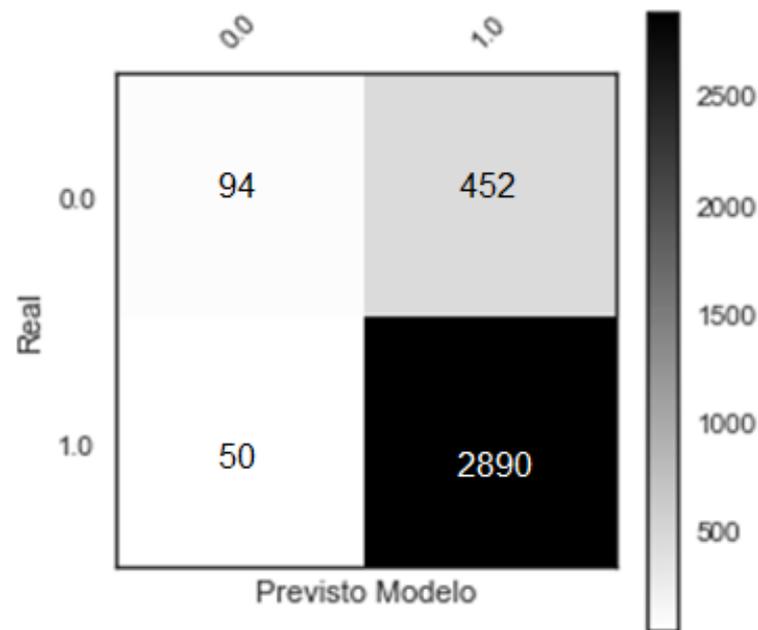


**Figura 19** – Valor comparativo do KS faixa a faixa.

Com a Figura 19 é possível comparar graficamente faixa a faixa o desempenho de separação. Notou-se uma melhor separação entre as faixas de 20 a 50 % em peso.

#### 4.3.3 Matriz de confusão

A Figura 20 apresenta a matriz de confusão, composta por verdadeiros positivos, verdadeiro negativos, falsos positivos e falsos negativos. Quanto menor forem os valores de falsos positivos e negativos, mais assertivo o modelo se apresenta.



**Figura 20** – Matriz de confusão.

Percebeu-se que das 3.486 corridas, o modelo previu erroneamente 492, das quais 452 foram falsos positivos e apenas 50 foram falsos negativos. Este resultado mostrou uma lacuna ao tentar encontrar casos com problemas, pois ao sugerir 452 corridas como boas, mostra que o modelo é pouco sensível a prever classe “mau”. Em contrapartida, as 50 previsões falso negativas, mostram que o modelo erra apenas 1,4 % dos casos ao prever que terá problemas no lingotamento.

Uma possível solução para reduzir a taxa de falsos positivos, seria a mobilidade da *threshold*, ou seja, alterando o limite de decisão seria possível obter um modelo mais conservador. Dessa forma, diminuiriam os erros oriundos de uma previsão falsamente dita como boa.

#### 4.3.4 Precisão, revocação e F1-Score

A Tabela 4 apresenta os valores de precisão, revocação e F1-Score para ambas as classes e também o ponderado.

**Tabela 4** – Tabela com precisão, revocação e F1-Score.

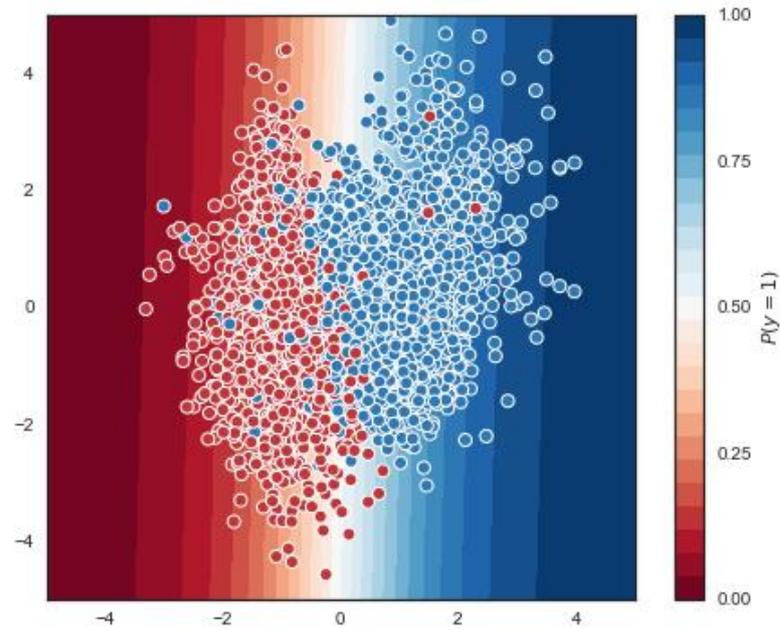
	Precisão	Revocação	F1-Score
MAU	0,65	0,17	0,27
BOM	0,86	0,98	0,92
Ponderado	0,83	0,86	0,82

Através da Tabela 4, percebeu-se que para classe “bom” os resultados foram muito mais satisfatórios do que para classe “mau” e isso é explicado pelo que podemos notar na matriz de confusão, em que apresentou uma quantidade elevada de falsos positivos. Por conseguinte, ao ponderar os resultados, os valores continuaram bastantes aceitáveis, visto que a quantidade de verdadeiros positivos é deveras elevada em relação aos demais classificadores.

#### 4.3.5 Classificador de distribuição da regressão logística

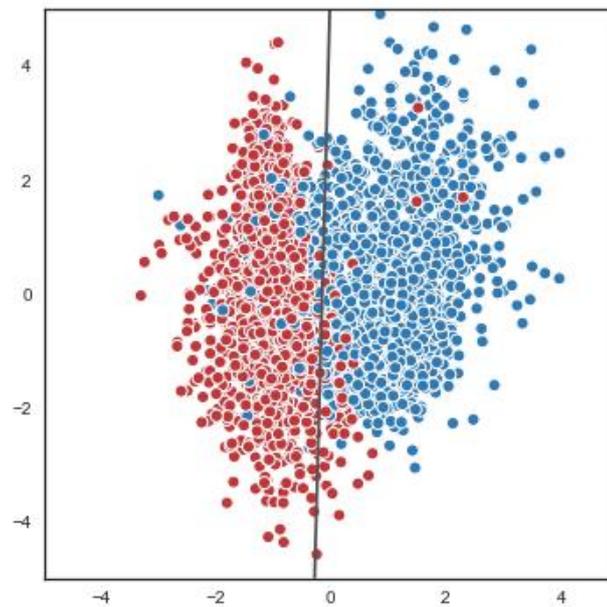
A vantagem do classificador de regressão logística é que, uma vez ajustado, é possível obter probabilidades para qualquer vetor de amostra.

A regressão logística permite classificar novas amostras com base em qualquer limite desejado, de modo que não tenha inerentemente um "limite de decisão". Uma regra de decisão comum a usar é 0.5, de forma que a *threshold* siga a curva da regressão logística, porém ajuste-se próximo a este valor. A Figura 21 ilustra o gráfico de distribuição bem como a probabilidade de cada ponto. Percebe-se que, assim como a matriz de confusão, a revocação e o F1-Score mostraram, a quantidade de falsos positivos bem mais elevada do que falsos negativos, isto é possível observar pela quantidade de pontos vermelhos além da *threshold* ser bem menor do que a quantidade de pontos azuis ser maior aquém da *threshold*.



**Figura 21** – Classificador de distribuição da regressão logística com gradiente de probabilidades.

Imputando a *threshold* no gráfico e retirando o apoio da probabilidade, como é possível observar na Figura 22, consegue-se observar melhor os pontos em que o modelo previu aquém ou além do “limite de decisão” definido.



**Figura 22** – Classificador de distribuição da regressão logística com *threshold* em 0,5.

Pela Figura 22, é observado que a grande concentração de erro ficou bastante próxima da *threshold*, o que se conclui que o modelo previu uma probabilidade muito próxima de 50% e, portanto, o modelo conseguiu ordenar corretamente estas corridas.

#### 4.3.6 Coeficientes das variáveis da regressão logística

Os coeficientes, de cada variável que foi utilizada para o modelo de regressão logística, podem ser visualizados na Tabela que está no apêndice B, abaixo está apresentada a Tabela 5 referente às variáveis que apresentaram maior valor no seu coeficiente. Ao aplicar a equação de Sigmoid, apresentada na revisão do presente trabalho, multiplicando cada valor de variável por seu coeficiente, é possível chegar na probabilidade de a classe ser “bom”, ou seja, não ocorrer problemas no lingotamento.

**Tabela 5** – Valores dos coeficientes das principais variáveis.

Coeficientes	Variável
-0,803	Posição_Inicio
-0,803	1º SEQUENCIA
-0,382	1º SEQUENCIA TOTAL
-0,343	Posição Fim
-0,263	2º SEQUENCIA TOTAL
-0,245	Temperatura superheat
-0,197	PF_Altot
-0,134	PF_S
0,201	7º SEQUENCIA TOTAL
0,206	Temperatura Liquidus
0,277	2º SEQUENCIA
0,286	3º SEQUENCIA
1,147	Posição Meio

As variáveis com coeficientes mais negativos mostram que ao serem verdadeiras (no caso das *dummies*) ou apresentarem valor alto (no caso das contínuas escalares) ao serem multiplicadas pelo seu coeficiente representarão uma queda de probabilidade de sucesso no lingotamento. Em contrapartida, as variáveis com coeficientes mais positivos, aumentarão as chances de sucesso na lingotabilidade.

As variáveis de posição início, meio e fim, bem como as variáveis de ordem de sequenciamento da corrida de determinada qualidade do aço no distribuidor, denominada “Xº SEQUENCIA”, foram as que mais influenciaram no modelo. Caso esteja na posição início, será a 1ª sequência, por consequência será a primeira corrida com aquele distribuidor. Pelo resultado dos coeficientes, notou-se uma piora na probabilidade da lingotabilidade nesta situação, fato explicado pois as primeiras corridas de um sequencial tendem a ser mais sujas, ou seja, apresentam maior probabilidade de conter impurezas provenientes dos detritos remanescentes da montagem do distribuidor. Analogamente, as variáveis de sequências intermediárias e a posição meio, colaboram para uma boa probabilidade de lingotabilidade, pois após certo tempo, a limpeza do aço tende a estabilizar em valores menores (BANNENBERG, 1993; FERNANDES, 2007; GARCIA, 2009).

As variáveis que medem a temperatura como “Temperatura *superheat*” (temperatura de sobre-aquecimento) e temperatura *liquidus*, apresentaram forte influência no modelo. Isso pode ser explicado devido ao tipo de fonte que provoca o *clogging*. O tipo 1 corresponde a formação de óxidos no fundo do bocal da válvula; o tipo 2 corresponde aos óxidos que estão presentes no banho e são transportados para as paredes da válvula submersa; e o tipo 3 – que explica o resultado do modelo – ocorre como consequência do resfriamento e solidificação nas paredes da válvula, pois o baixo grau de *superheat* ou a elevada taxa de extração de calor estão associadas ao aumento local da temperatura *liquidus* (BUBOLTZ, 2011).

Teores altos na prova final de alumínio total e enxofre no aço tiveram forte impacto negativo para a lingotabilidade, o que pode ser explicado pela tendência de formação de inclusões sólidas de alumina e CaS (STORY, 2004; MEDEIROS, 2014).

## 5 CONCLUSÕES

A partir das análises empregadas no presente estudo, permite-se concluir que:

- A aplicação de modelagem estatística para predição de eventos em ambiente siderúrgico mostrou capacidade de ser uma ferramenta robusta para aumento de eficiência de processo.

- A definição de uma *threshold* estratégica torna-se essencial na aplicação da tomada de decisão a partir da probabilidade de êxito no processo, pois assim consegue se determinar o grau de segurança para garantir uma boa corrida. Os resultados do trabalho foram balizados pelo limite de decisão de 50 %.

- Tanto o indicador do AUC (área abaixo da curva ROC), com valor de 0,862, quanto de acurácia, com valor de 0,856, apresentaram resultados bastante satisfatórios – do ponto de vista estatístico – visto que pôde ser comparado com resultados de modelos aplicados na área de medicina, onde se exige alto grau de confiança.

- As variáveis com maior influência para o processo, como o sequenciamento no distribuidor, temperaturas *liquidus* e *superheat*, e teores de alumínio e enxofre, corroboraram com o que já fora abordado na literatura. Isso mostra a coerência do modelo, bem como a confiança nos trabalhos referenciados.

- A aplicabilidade do modelo em ambiente industrial, como ferramenta para auxiliar na parametrização do processo, torna-se uma solução inteligente, visto que sua implementação depende apenas de integração do sistema de *preset* de parâmetros e controle de composição química.



## 6 SUGESTÕES PARA TRABALHOS FUTUROS

Dados resultados, análises e discussões, apresentam-se sugestões após a realização deste trabalho:

- Registrar novas variáveis que possam influenciar no processo, para que possam ser incorporadas ao modelo a fim de aumentar seu desempenho de predição;
- Desenvolver modelos preditivos além de regressão logística, como árvore de decisões ou redes neurais, a fim de se obter melhores resultados;
- Analisar e estudar variáveis que apresentaram coeficientes suficientemente elevados, com o objetivo de determinar a real influência delas no processo de lingotamento contínuo;
- Implementar o modelo em escala industrial e medir sua capacidade de aumento de eficiência. O resultado do modelo mostrou facilidade de implementação, visto que com o *output* dos coeficientes, juntamente com a equação de Sigmoid, é possível determinar a probabilidade de êxito no processo com uma simples planilha de Excel, por exemplo.

## REFERÊNCIAS

- A. ALBERINI. **Testing willingness-to-pay models of discrete-choice contingent valuation survey data.** Land Econ., 71 (1), p. 83-95, 1995.
- ARANA, J.E., LEON, C.J. **Flexible mixture distribution modeling of dichotomous choice contingent valuation with heterogeneity.** J. Environ. Econ. Manage., 50 (1), p. 170-188, 2005.
- ATKINSON, P.M., MASSARI, R. **Generalised linear modelling of susceptibility of landsliding in the Central Apennines, Italy.** Computers & Geosciences, 24, p. 373-385, 1998.
- AYALEW, L., YAMAGISHI, H. **The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan.** Geomorphology, 65, p. 15-31, 2005.
- BANNENBERG, N., HARSTE, K. **Improvements in Steel Cleanliness by Tundish Intertization.** La Revue de Métallurgie, p. 71-76, Janvier 1993.
- BANNENBERG, N.; LACHMUND H. **Reactions between tundish lining and steel and their unfluence on steel cleanness.** Proceedings METEC Congress '94, 2<sup>ND</sup> European Continuous Casting Conference, 6<sup>th</sup> Internacional Rolling Conference, V.1, p. 25-31. Düsseldorf, Germany, 1994.
- BELL, C.D., ROBERTS, R.K., ENGLISH, B.C., W.M. PARK. **A logit analysis of participation in Tennessee's forest stewardship program.** J. Agric. Appl. Econ., 26 (2), p. 463-472, 1994.
- BIELEFELDT, W. V. **Estudo do tratamento de inclusões com cálcio na fabricação do aço SAE 8620.** Dissertação. PPGEM: UFRGS, 2005.
- BIELEFELDT, W. V., VILELA, A. C. F. **Curso básico de siderurgia.** Porto Alegre, 2013.
- BUBOLTZ, F. F. **Análise e verificação de variáveis do processo como causas para o sucateamento por macroinclusões em uma aciaria de aços especiais.** Dissertação: UFRGS. Porto Alegre, 2011.
- BUBOLTZ, C. B. **Desenvolvimento de modelos teóricos para dureza Brinell, utilizando método de regressão linear múltipla e fórmulas de carbono equivalente para aço SAE 52100.** Dissertação: UFRGS. Porto Alegre, 2016.
- CARRARA, A., CARDINALI, M., DETTI, R., GUZZETTI, F., PASQUI, V., REICHENBACH, P. **GIS techniques and statistical models in evaluating landslide hazard.** Earth Surface Processes and Landforms, 16, p. 427-445, 1991.
- CHANG, K.T. **Analysing the relationship between typhoon-triggered landslides and critical rainfall conditions.** Earth Surface Processes and Landforms, 33, p. 1261-1271, 2007B.

CHANG, K.T. **Modeling typhoon- and earthquake-induced landslides in a mountainous watershed using logistic regression.** *Geomorphology*, 89, p. 335-347, 2007A.

CHUNG, C.F., FABRI, A.G. **Probabilistic prediction models for landslide hazard mapping.** *Photogrammetric Engineering and Remote Sensing*, 65, p. 1389-1399, 1999.

COHN, D. A. **Active learning with statistical models.** Proceedings of the 7th International Conference on Neural Information Processing Systems, NIPS'94, MIT Press, Cambridge, MA, USA, p. 705-712, 1994.

DAI, F.C., LEE, C.F. **A spatiotemporal probabilistic modeling of storm-induced shallow landsliding using aerial photographs and logistic regression.** *Earth Surface Processes and Landforms*, 28, p. 527-545, 2003.

DAI, F.C., LEE, C.F. **Landslide risk assessment and management: an overview.** *Engineering Geology*, 64, p. 65-87, 2002.

DEKKERS, R. **Non-metallic Inclusions in Liquid Steel.** Germany, 2002, Doctoral Thesis, Department of Geography and Geology, KU Leuven.

DEKKERS, R et al. **Non-metallic Inclusions in Aluminium Killed Steels.** *Ironmaking and Steelmaking*. V.29, n. 6, 2002.

EFRON, Bradley; TIBSHIRANI, Robert J. **Cross-validation and the bootstrap: Estimating the error rate of a prediction rule.** Division of Biostatistics, Stanford University, 1995.

GARCIA, V. S. JR. **Verificação do efeito de injeção de gás através de válvulas submersas mulltiporos sobre a lingotabilidade e limpeza do aço.** Dissertação. PPGEM: UFRGS, 2009.

GHOSH, A. **Secondary steelmaking – Principles and Applications.** Kanpur: CRC Press, 2001.

GRECO, R. **Logistic regression analysis in the evaluation of mass movements susceptibility: the Aspromonte case study, Calabria, Italy.** *Engineering Geology*, 89, p. 47-66, 2007.

HOLAPPA, L. et al. **Thermodynamic Examination on Inclusion Modification and Precipitation from Calcium Treatment to Solidified Steel.** *Ironmaking and Steelmaking*, V.30, n. 6, p. 111-115, 2003.

HSU, D.J. **Algorithms for Active Learning** (Ph.D. thesis), Columbia University, 2010

JAMES, Gareth; WITTEN, Daniela; HASTIE, Trevor; TIBSHIRANI, Robert. **An introduction to statistical learning.** New York: springer, 2013.

JANKE, D. et al. **Improvement of Castability and Quality of Continuously Cast Steel.** *ISIJ International*, V.40, n. 1, 2000.

JORDAN, M.I., MITCHELL, T.M. **Machine learning: Trends, perspectives, and prospects**. Science, 349, p. 255-260, 2015.

KIESSLING, R., LANGE, N. **Non-metallic inclusions in steel**. London: Metals Society, 1978.

KOHAVI, Ron. **A study of cross-validation and bootstrap for accuracy estimation and model selection**. In: Ijcai. 1995. p. 1137-1145.

LEE, S., MIN, K. **Statistical analysis of landslide susceptibility at Yongin, Korea**. Environmental Geology, 40, p. 1095-1113, 2001.

MAGALHÃES, H. L. G. **Melhoria da limpidez do aço SAE 1045 desoxidado ao alumínio com aplicação na indústria automobilística, utilizando termodinâmica computacional**. Dissertação de mestrado: UFOP. Ouro Preto, 2010.

MEDEIROS, S. M. **Verificação do efeito do teor de alumínio na lingotabilidade, na limpeza inclusionária e no tamanho de grão do aço DIN 38B3**. Dissertação: UFRGS. Porto Alegre, 2014.

MEMARPOUR, A.; BRABIE, V.; JÖNSSON, P. G. **Studies on decarburization of commercial submerged entry nozzles**. Steel Grips, Journal of Steel and Related Materials, Plants and Equipment, V. 9, P. 121-127, 2011.

MILLMAN, S. **Clean Steel – Basic Features and Operation Practices**. IISI Study on Clean Steel, IISI Committee on Technology, 2004.

OHLMACHER, G.C., DAVIS, J.C. **Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA**. Engineering Geology, 69, p. 331-343, 2003.

POWERS, D. M. W. **Evaluation: from precision, recall and f-measure to roc, informedness, markedness & correlation**. Journal of Machine Learning Technologies, V.2, p.37-63. Australia, 2011.

RIZZO, E. **Introdução aos processos siderúrgicos**. Associação Brasileira de Metalurgia, Materiais e Mineração, São Paulo, 2005.

SETTLES B. **Active Learning. Synthesis Lectures on Artificial Intelligence and Machine Learning**. Morgan & Claypool Publishers, San Rafael, 2012.

SETTLES, B. **From theories to queries: Active learning in practice**. Journal of Machine Learning Research, Workshop on Active Learning and Experimental Design, Workshop and Conference Proceedings, V.16, p. 1–18, 2011.

SHAFFI, A. S. **Measures Derived from a 2 x 2 Table for an Accuracy of a Diagnostic Test**. Journal of Biometrics & Biostatistics. Saudi Arabia, 2011.

SHARMA, A., KANDLAKUNTA, H. **Model to Determine Risk of Pancreatic Cancer in Patients with New-Onset Diabetes**. Official Journal of The AGA Institute, Gastroenterology, V.155, p. 730-739, 2018.

- SKOLOVA, M. **A systematic analysis of performance measures for classification tasks.** Information Processing and Management, V.45, p. 427-437. Canada, 2009.
- STORY, S. R. et al. **Inclusion Analysis to Predict Casting Behavior.** Iron & Steel Technology, p. 163-169, Indianapolis, 2004.
- SVENSSON, J. K. S.; MEMARPOUR, A.; EKEROT, S.; BRABIE, V.; JÖNSSON, P. G. **Studies of new coating materials to prevent clogging of submerged entry nozzle (SEN) during continuous casting of Al killed low carbon steels.** Ironmaking & Steelmaking – Processes, Products and Applications. V. 44, n. 2, p. 117-127, 2017.
- THARWAT, Alaa. **Classification assessment methods.** Applied Computing and Informatics, 2018.
- THOMAS, B. G., BAI. **Tundish nozzle clogging – Application of computational models.** 18rd Process Technology Division Conference Proceedings. V. 18, Iron and Steel Society, Warrendale, PA, 2001.
- TRÉPO, E., GOOSSENS, N., FUJIWARA, N. **Combination of Gene Expression Signature and Model for End-Stage Liver Disease Score Predicts Survival of Patients with Severe Alcoholic Hepatitis.** Official Journal of The AGA Institute, Gastroenterology, V.154, p. 965-975, 2018.
- VERMEULEN, Y.; COLETTI, B.; BLANPAIN, B.; WOLLANTS, P.; VLEUGELS, J. **Material Evaluation to Prevent Nozzle Clogging during Continuous Casting of Al Killed Steels.** ISIJ International, V. 42, n. 11, p. 1234-1240, 2002.
- XU, L., 1992. **Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition.** IEEE Transactions on Systems, Man and Cybernetics, V.22, n.3, 1992.
- YANG, Y., LOOG, M. **A benchmark and comparison of active learning for logistic regression.** Pattern Recognition. V.83, p. 401-415, 2018.
- YOUNG, I.T. **Proof without prejudice: use of the Kolmogorov–Smirnov test for the analysis of histograms from flow systems and other sources.** Journal of Histochemistry and Cytochemistry, 25, p. 935, 1977.
- ZHANG, Li-Feng. **Inclusion and bubble in steel - A review.** Journal of iron and Steel Research International, v.13, 2006.
- ZHANG, L., THOMAS, B. G. **State of the Art in Evaluation and Control of Steel Cleanliness.** ISIJ International, V.43, n. 3, p. 271-291, 2003.
- ZHANG, L. **Indirect methods of detecting and evaluating inclusions in steel – A review.** Journal of Iron and Steel Research, V.13, p. 1-8, 2006.

## APÊNDICE A

Iteração	AUC	Removeu	<i>tol</i>	<i>C</i>	<i>max_iter</i>	<i>fit_intercept</i>
1	0,837025598	Provas escória	0,001	0,053	100	Verdadeiro
2	0,837025598	Provas escória	0,001	0,053	200	Verdadeiro
3	0,837025382	Provas escória	0,0001	0,053	100	Verdadeiro
4	0,837025382	Provas escória	0,0001	0,053	200	Verdadeiro
5	0,837021243	Provas escória	0,0001	0,054	100	Verdadeiro
6	0,837021243	Provas escória	0,0001	0,054	200	Verdadeiro
7	0,837021065	Provas escória	0,001	0,054	100	Verdadeiro
8	0,837021065	Provas escória	0,001	0,054	200	Verdadeiro
9	0,837018611	Provas escória	0,0001	0,052	100	Verdadeiro
10	0,837018611	Provas escória	0,0001	0,052	200	Verdadeiro
11	0,837018233	Provas escória	0,001	0,052	100	Verdadeiro
12	0,837018233	Provas escória	0,001	0,052	200	Verdadeiro
13	0,836843011	Provas escória	0,0001	0,054	100	Falso
14	0,836843011	Provas escória	0,0001	0,054	200	Falso
15	0,836842477	Provas escória	0,001	0,054	100	Falso
16	0,836842477	Provas escória	0,001	0,054	200	Falso
17	0,836832126	Provas escória	0,001	0,053	100	Falso
18	0,836832126	Provas escória	0,001	0,053	200	Falso
19	0,836831941	Provas escória	0,0001	0,053	100	Falso
20	0,836831941	Provas escória	0,0001	0,053	200	Falso
21	0,836813142	Provas escória	0,0001	0,052	100	Falso
22	0,836813142	Provas escória	0,0001	0,052	200	Falso
23	0,836812757	Provas escória	0,001	0,052	100	Falso
24	0,836812757	Provas escória	0,001	0,052	200	Falso
25	0,835898798	Processo	0,001	0,052	100	Falso
26	0,835898798	Qualidade	0,001	0,052	100	Falso
27	0,835898798	Processo	0,001	0,052	200	Falso
28	0,835898798	Qualidade	0,001	0,052	200	Falso
29	0,83589862	Processo	0,0001	0,052	100	Falso
30	0,83589862	Qualidade	0,0001	0,052	100	Falso
31	0,83589862	Processo	0,0001	0,052	200	Falso
32	0,83589862	Qualidade	0,0001	0,052	200	Falso
33	0,835892795	Processo	0,0001	0,054	100	Falso
34	0,835892795	Qualidade	0,0001	0,054	100	Falso
35	0,835892795	Processo	0,0001	0,054	200	Falso
36	0,835892795	Qualidade	0,0001	0,054	200	Falso
37	0,835892619	Processo	0,001	0,054	100	Falso
38	0,835892619	Qualidade	0,001	0,054	100	Falso
39	0,835892619	Processo	0,001	0,054	200	Falso
40	0,835892619	Qualidade	0,001	0,054	200	Falso
41	0,835885242	Processo	0,001	0,053	100	Falso

42	0,835885242	Qualidade	0,001	0,053	100	Falso
43	0,835885242	Processo	0,001	0,053	200	Falso
44	0,835885242	Qualidade	0,001	0,053	200	Falso
45	0,835885064	Processo	0,0001	0,053	100	Falso
46	0,835885064	Qualidade	0,0001	0,053	100	Falso
47	0,835885064	Processo	0,0001	0,053	200	Falso
48	0,835885064	Qualidade	0,0001	0,053	200	Falso
49	0,835692543	Processo	0,001	0,052	100	Verdadeiro
50	0,835692543	Qualidade	0,001	0,052	100	Verdadeiro
51	0,835692543	Processo	0,001	0,052	200	Verdadeiro
52	0,835692543	Qualidade	0,001	0,052	200	Verdadeiro
53	0,835692344	Processo	0,0001	0,052	100	Verdadeiro
54	0,835692344	Qualidade	0,0001	0,052	100	Verdadeiro
55	0,835692344	Processo	0,0001	0,052	200	Verdadeiro
56	0,835692344	Qualidade	0,0001	0,052	200	Verdadeiro
57	0,835683417	Processo	0,0001	0,053	100	Verdadeiro
58	0,835683417	Qualidade	0,0001	0,053	100	Verdadeiro
59	0,835683417	Processo	0,0001	0,053	200	Verdadeiro
60	0,835683417	Qualidade	0,0001	0,053	200	Verdadeiro
61	0,835683033	Processo	0,001	0,053	100	Verdadeiro
62	0,835683033	Qualidade	0,001	0,053	100	Verdadeiro
63	0,835683033	Processo	0,001	0,053	200	Verdadeiro
64	0,835683033	Qualidade	0,001	0,053	200	Verdadeiro
65	0,835665652	Processo	0,0001	0,054	100	Verdadeiro
66	0,835665652	Qualidade	0,0001	0,054	100	Verdadeiro
67	0,835665652	Processo	0,0001	0,054	200	Verdadeiro
68	0,835665652	Qualidade	0,0001	0,054	200	Verdadeiro
69	0,835665652	Processo	0,001	0,054	100	Verdadeiro
70	0,835665652	Qualidade	0,001	0,054	100	Verdadeiro
71	0,835665652	Processo	0,001	0,054	200	Verdadeiro
72	0,835665652	Qualidade	0,001	0,054	200	Verdadeiro
73	0,835406498	Provas aço	0,0001	0,053	100	Verdadeiro
74	0,835406498	Provas aço	0,0001	0,053	200	Verdadeiro
75	0,835406299	Provas aço	0,001	0,053	100	Verdadeiro
76	0,835406299	Provas aço	0,001	0,053	200	Verdadeiro
77	0,835396365	Provas aço	0,001	0,052	100	Verdadeiro
78	0,835396365	Provas aço	0,001	0,052	200	Verdadeiro
79	0,835396164	Provas aço	0,0001	0,052	100	Verdadeiro
80	0,835396164	Provas aço	0,0001	0,052	200	Verdadeiro
81	0,835392278	Provas aço	0,0001	0,054	100	Verdadeiro
82	0,835392278	Provas aço	0,0001	0,054	200	Verdadeiro
83	0,835391701	Provas aço	0,001	0,054	100	Verdadeiro
84	0,835391701	Provas aço	0,001	0,054	200	Verdadeiro
85	0,835377716	Provas aço	0,001	0,054	100	Falso

86	0,835377716	Provas aço	0,001	0,054	200	Falso
87	0,835377687	Provas aço	0,0001	0,054	100	Falso
88	0,835377687	Provas aço	0,0001	0,054	200	Falso
89	0,835370068	Provas aço	0,0001	0,053	100	Falso
90	0,835370068	Provas aço	0,0001	0,053	200	Falso
91	0,835370063	Provas aço	0,001	0,053	100	Falso
92	0,835370063	Provas aço	0,001	0,053	200	Falso
93	0,835368961	Provas aço	0,0001	0,052	100	Falso
94	0,835368961	Provas aço	0,0001	0,052	200	Falso
95	0,835368586	Provas aço	0,001	0,052	100	Falso
96	0,835368586	Provas aço	0,001	0,052	200	Falso

## APÊNDICE B

Coef.	Variável	Coef.	Variável	Coef.	Variável
-0,803	Posição Início	-0,002	QUALIDADE_O1	0,016	QUALIDADE_1095C
-0,803	1º SEQUENCIA	-0,002	P11_Zn	0,016	QUALIDADE_5140HE
-0,382	1º SEQUENCIA TOTAL	-0,002	QUALIDADE_4140T	0,017	QUALIDADE_5160HF
-0,343	Posição Fim	-0,001	P11_Si	0,017	PF_Ca
-0,268	QUALIDADE_5120HB	0,000	Distribuidor Ankertun	0,017	QUALIDADE_6135A
-0,263	2º SEQUENCIATOTAL	0,000	QUALIDADE_1010A	0,017	QUALIDADE_1540HM
-0,245	Temperatura superheat	0,000	QUALIDADE_1010D	0,018	QUALIDADE_5140HX
-0,197	PF_Altot	0,000	QUALIDADE_1010E	0,018	QUALIDADE_1035A
-0,164	QUALIDADE_1055HJ	0,000	QUALIDADE_1015D	0,018	QUALIDADE_1018HMAG
-0,134	PF_S	0,000	QUALIDADE_1015I	0,019	QUALIDADE_8620HQ
-0,134	PF_Se	0,000	QUALIDADE_1020T	0,019	QUALIDADE_9840A
-0,133	PF_H2	0,000	QUALIDADE_1020V	0,019	QUALIDADE_4122HC
-0,123	QUALIDADE_8620HE	0,000	QUALIDADE_1025B	0,019	QUALIDADE_1541HC
-0,120	7º SEQUENCIA	0,000	QUALIDADE_1042HB	0,019	QUALIDADE_8620HC
-0,119	QUALIDADE_5120HV	0,000	QUALIDADE_1045HK	0,020	QUALIDADE_4140HAA
-0,116	QUALIDADE_5120A	0,000	QUALIDADE_1048C	0,020	QUALIDADE_4140HD
-0,115	QUALIDADE_8620HD	0,000	QUALIDADE_1053HF	0,020	QUALIDADE_4320HX
-0,115	QUALIDADE_41V20A	0,000	QUALIDADE_1060C	0,021	QUALIDADE_4125HP
-0,109	P11_Co	0,000	QUALIDADE_1080B	0,021	QUALIDADE_5160HD
-0,105	QUALIDADE_1055HL	0,000	QUALIDADE_1095A	0,021	QUALIDADE_15V40A
-0,104	P11_Cu	0,000	QUALIDADE_1108MBI	0,021	QUALIDADE_1040B
-0,103	QUALIDADE_5120HBB	0,000	QUALIDADE_1117MAG	0,022	QUALIDADE_8620HMAG
-0,103	QUALIDADE_4320HAK	0,000	QUALIDADE_1141B	0,022	QUALIDADE_10B20A
-0,103	QUALIDADE_1045HM	0,000	QUALIDADE_1141C	0,022	QUALIDADE_41B30HA
-0,102	QUALIDADE_4140Z	0,000	QUALIDADE_1141O	0,022	QUALIDADE_1006B
-0,097	QUALIDADE_5160REP	0,000	QUALIDADE_1212A	0,022	QUALIDADE_4130HA
-0,096	P11_Bsol	0,000	QUALIDADE_1515E	0,023	QUALIDADE_5130HC
-0,093	QUALIDADE_5120HO	0,000	QUALIDADE_1524HA	0,023	QUALIDADE_15V46A
-0,091	QUALIDADE_1053HA	0,000	QUALIDADE_1536D	0,023	QUALIDADE_9820HB
-0,091	QUALIDADE_1045HE	0,000	QUALIDADE_15B35HC	0,024	QUALIDADE_4125HE
-0,091	QUALIDADE_1045HMAG	0,000	QUALIDADE_15V31A	0,024	QUALIDADE_10B38C
-0,089	QUALIDADE_5160B	0,000	QUALIDADE_15V38P	0,024	QUALIDADE_4140J
-0,087	QUALIDADE_10V40M	0,000	QUALIDADE_15V43HA	0,024	QUALIDADE_1540G
-0,085	QUALIDADE_4125HK	0,000	QUALIDADE_3017H	0,024	Distribuidor Jetmag
-0,085	QUALIDADE_4120HC	0,000	QUALIDADE_3218HA	0,024	QUALIDADE_1045HD
-0,082	P11_P	0,000	QUALIDADE_4027HA	0,024	QUALIDADE_P20
-0,081	QUALIDADE_15V27A	0,000	QUALIDADE_4118B	0,025	QUALIDADE_5134HD
-0,077	Temperatura VD Real	0,000	QUALIDADE_4119HAA	0,025	QUALIDADE_15V44D
-0,077	4º SEQUENCIA TOTAL	0,000	QUALIDADE_4120B	0,025	QUALIDADE_1144A
-0,077	QUALIDADE_15B35HF	0,000	QUALIDADE_4122HB	0,025	PF_Bi
-0,073	QUALIDADE_15V45E	0,000	QUALIDADE_4124HD	0,026	QUALIDADE_1570B

-0,072	PF_AlsoI	0,000	QUALIDADE_4125HC	0,026	QUALIDADE_4140HMAG
-0,071	QUALIDADE_1080E	0,000	QUALIDADE_4125HCR	0,026	P11_Sn
-0,066	QUALIDADE_1045AQ	0,000	QUALIDADE_4125HJ	0,027	QUALIDADE_5120HAU
-0,066	QUALIDADE_1047HA	0,000	QUALIDADE_4130HB	0,027	QUALIDADE_3015H
-0,066	QUALIDADE_1045HAD	0,000	QUALIDADE_4130HC	0,027	QUALIDADE_8640HE
-0,066	QUALIDADE_1018E	0,000	QUALIDADE_4135B	0,028	QUALIDADE_8620HAC
-0,063	PROCESSO ESCORIA nº1	0,000	QUALIDADE_4135F	0,029	13º SEQUENCIA TOTAL
-0,063	QUALIDADE_4140HI	0,000	QUALIDADE_4140HAG	0,029	QUALIDADE_5160HL
-0,063	QUALIDADE_1056HA	0,000	QUALIDADE_4140HAI	0,029	QUALIDADE_15B37HA
-0,063	QUALIDADE_4320HAJ	0,000	QUALIDADE_4140HR	0,030	Produto DEG
-0,062	QUALIDADE_9254HM	0,000	QUALIDADE_4140HZ	0,030	QUALIDADE_1045HG
-0,061	QUALIDADE_5145HE	0,000	QUALIDADE_4150AMAG	0,030	PF_Ni
-0,060	PF_Titot	0,000	QUALIDADE_41V40HA	0,031	QUALIDADE_4140L
-0,059	QUALIDADE_5115HQ	0,000	QUALIDADE_4230HK	0,031	QUALIDADE_4320HC
-0,059	QUALIDADE_51B60A	0,000	QUALIDADE_4320HAT	0,031	QUALIDADE_50B40HA
-0,057	QUALIDADE_4320HA	0,000	QUALIDADE_4320HCR	0,031	QUALIDADE_4140HF
-0,054	QUALIDADE_4130OEG	0,000	QUALIDADE_4340OIL	0,032	8º SEQUENCIA TOTAL
-0,053	PF_W	0,000	QUALIDADE_50095C	0,032	QUALIDADE_15V38HC
-0,053	P11_Zr	0,000	QUALIDADE_50100HD	0,033	PF_As
-0,052	QUALIDADE_4340J	0,000	QUALIDADE_50B34C	0,034	11º SEQUENCIA 11
-0,050	3º SEQUENCIATOTAL	0,000	QUALIDADE_5115HD	0,034	QUALIDADE_1017A
-0,049	QUALIDADE_10B40HA	0,000	QUALIDADE_5115HG	0,035	QUALIDADE_4340HMAG
-0,049	QUALIDADE_61B32HB	0,000	QUALIDADE_5115HK	0,035	QUALIDADE_15B20HJ
-0,048	P11_Mn	0,000	QUALIDADE_5115HX	0,035	QUALIDADE_1053HD
-0,048	QUALIDADE_50B34HD	0,000	QUALIDADE_5119HA	0,036	10º SEQUENCIA
-0,045	QUALIDADE_1040W	0,000	QUALIDADE_5119HAA	0,037	QUALIDADE_51V52HA
-0,044	QUALIDADE_5135HG	0,000	QUALIDADE_5119HB	0,037	QUALIDADE_4140HAC
-0,044	QUALIDADE_15V40K	0,000	QUALIDADE_5120HAJ	0,037	6º SEQUENCIA
-0,043	P11_W	0,000	QUALIDADE_5120HAS	0,037	QUALIDADE_1060E
-0,043	QUALIDADE_1010R	0,000	QUALIDADE_5120HR	0,037	QUALIDADE_1538D
-0,042	QUALIDADE_6150HA	0,000	QUALIDADE_5124HA	0,037	QUALIDADE_51B20HAG
-0,042	QUALIDADE_4320A	0,000	QUALIDADE_5130HA	0,038	P11_Se
-0,042	P11_Pb	0,000	QUALIDADE_5134HC	0,038	QUALIDADE_5115HT
-0,041	QUALIDADE_50095A	0,000	QUALIDADE_5162HA	0,039	QUALIDADE_1050HC
-0,040	Distribuidor Massa seca	0,000	QUALIDADE_5170HB	0,040	QUALIDADE_5115HV
-0,039	QUALIDADE_1552HB	0,000	QUALIDADE_51V31HA	0,040	QUALIDADE_5140I
-0,039	P11_Nb	0,000	QUALIDADE_6150A	0,040	QUALIDADE_15B35HA
-0,038	PROCESSO ESCORIA nº2	0,000	QUALIDADE_8620ABI	0,041	QUALIDADE_6150HF
-0,038	Produto AÇO	0,000	QUALIDADE_8620HAF	0,041	P11_Btot
-0,037	QUALIDADE_5120HAQ	0,000	QUALIDADE_8620HR	0,041	QUALIDADE_1035E
-0,036	QUALIDADE_5115B	0,000	QUALIDADE_8625HM	0,041	QUALIDADE_5120HAI
-0,036	QUALIDADE_51B20HZF	0,000	QUALIDADE_9254HE	0,041	QUALIDADE_5135HI
-0,035	QUALIDADE_3018HH	0,000	QUALIDADE_94V25A	0,042	P11_Ni
-0,035	QUALIDADE_10B38HBI	0,000	QUALIDADE_9818HA	0,043	QUALIDADE_4145A

-0,035	QUALIDADE_4320HAG	0,000	QUALIDADE_9820HA	0,043	8º SEQUENCIA
-0,035	QUALIDADE_1030A	0,000	QUALIDADE_15V38HD	0,043	QUALIDADE_6031A
-0,035	QUALIDADE_5120HU	0,001	QUALIDADE_51105HA	0,046	QUALIDADE_4135HE
-0,034	P11_Bi	0,001	QUALIDADE_50B30HA	0,046	PF_Zn
-0,034	QUALIDADE_15V46F	0,001	QUALIDADE_41V55HA	0,050	P11_N2
-0,033	PF_Cu	0,001	QUALIDADE_4150HB	0,050	9º SEQUENCIA
-0,032	QUALIDADE_1045HAA	0,001	QUALIDADE_15B20E	0,050	QUALIDADE_4130HZ
-0,032	QUALIDADE_5120HAK	0,001	QUALIDADE_5140HC	0,051	QUALIDADE_1540HJ
-0,032	PF_Te	0,002	QUALIDADE_1045HY	0,051	P11_As
-0,032	QUALIDADE_1045V	0,002	P11_Cu+10Sn	0,052	P11_V
-0,032	QUALIDADE_1045HV	0,002	QUALIDADE_4120HCR	0,054	QUALIDADE_1046HA
-0,032	QUALIDADE_4120HN	0,002	PROCESSO ESCORIA nº5	0,054	6º SEQUENCIA TOTAL
-0,031	PF_Ti	0,003	QUALIDADE_4125HQ	0,055	QUALIDADE_5115D
-0,031	QUALIDADE_5145HB	0,003	QUALIDADE_10B20HN	0,058	QUALIDADE_4142HZ
-0,031	PF_N2	0,004	QUALIDADE_50B40HD	0,058	PF_Si
-0,031	QUALIDADE_4140HA	0,004	QUALIDADE_1080A	0,059	QUALIDADE_4320HAE
-0,031	QUALIDADE_4135D	0,004	PF_Cr	0,060	QUALIDADE_5115HE
-0,030	QUALIDADE_1141HNB	0,004	QUALIDADE_4125HA	0,061	QUALIDADE_15V40C
-0,029	QUALIDADE_5140E	0,004	PF_V	0,062	QUALIDADE_5137HC
-0,029	QUALIDADE_15V45D	0,004	QUALIDADE_1050HI	0,063	QUALIDADE_9260A
-0,029	QUALIDADE_5140HD	0,004	QUALIDADE_5170HA	0,066	PF_Mn
-0,029	QUALIDADE_15V46E	0,005	QUALIDADE_51B60HA	0,066	QUALIDADE_4119HA
-0,028	QUALIDADE_1524A	0,005	QUALIDADE_8630HA	0,066	PF_Bsol
-0,028	QUALIDADE_15B35HG	0,005	QUALIDADE_1045WPR	0,068	P11_Cr
-0,027	PF_Sb	0,005	QUALIDADE_4140HAD	0,069	QUALIDADE_5140HG
-0,027	QUALIDADE_5159HA	0,005	QUALIDADE_5120HP	0,069	P11_Te
-0,026	4º SEQUENCIA	0,005	QUALIDADE_5127HA	0,070	QUALIDADE_8620HS
-0,024	QUALIDADE_15B41HD	0,005	QUALIDADE_5160HA	0,071	QUALIDADE_5140F
-0,024	P11_Ca	0,005	QUALIDADE_50B30HB	0,071	QUALIDADE_10B38HA
-0,024	QUALIDADE_4142HI	0,006	QUALIDADE_1548HD	0,073	QUALIDADE_4320HG
-0,023	QUALIDADE_15V43A	0,006	QUALIDADE_5180HB	0,074	QUALIDADE_5115HJ
-0,023	PF_P	0,006	QUALIDADE_6150REP	0,075	QUALIDADE_5120HBA
-0,023	P11_Mo	0,007	QUALIDADE_1045P	0,075	Temperatura Objetiv
-0,023	QUALIDADE_5035HA	0,007	QUALIDADE_4142HG	0,076	5º SEQUENCIA TOTAL
-0,021	QUALIDADE_5115HL	0,007	QUALIDADE_6150HJ	0,079	PF_Sn
-0,020	QUALIDADE_1005A	0,007	QUALIDADE_15V40HE	0,080	QUALIDADE_10V70D
-0,019	12º SEQUENCIA TOTAL	0,007	13º SEQUENCIA	0,080	QUALIDADE_50B34HB
-0,019	QUALIDADE_1140A	0,007	QUALIDADE_51B20HB	0,084	PF_Cu+10Sn
-0,019	QUALIDADE_15B40HD	0,008	QUALIDADE_1045HA	0,084	PROCESSO ESCORIA nº3
-0,019	QUALIDADE_5135HJ	0,008	QUALIDADE_51105HB	0,086	P11_H2
-0,019	QUALIDADE_50B34HA	0,008	QUALIDADE_4124HC	0,092	P11_Altot
-0,018	QUALIDADE_1045HJ	0,008	QUALIDADE_15V24HA	0,092	P11_Titot
-0,017	QUALIDADE_1060A	0,008	Produto ADEG	0,096	QUALIDADE_1515F
-0,017	QUALIDADE_4121HB	0,008	QUALIDADE_1050B	0,096	PF_Pb

-0,016	QUALIDADE_4142HA	0,009	QUALIDADE_1141HE	0,101	QUALIDADE_4120HA
-0,015	QUALIDADE_1018D	0,009	Al Lingote (m)	0,103	QUALIDADE_5120HAE
-0,014	QUALIDADE_5120HAT	0,009	QUALIDADE_1040A	0,104	P11_C
-0,014	O2 ATIVO	0,009	QUALIDADE_4140HK	0,107	PF_Co
-0,014	QUALIDADE_5140HK	0,010	QUALIDADE_1045HB	0,108	QUALIDADE_1020A
-0,013	Seção 240mm	0,010	QUALIDADE_5120K	0,108	QUALIDADE_8620A
-0,013	QUALIDADE_4125HG	0,010	QUALIDADE_1050Q	0,109	10º SEQUENCIA TOTAL
-0,013	QUALIDADE_8640A	0,011	P11_Alsol	0,124	QUALIDADE_5115HB
-0,013	QUALIDADE_1030J	0,011	QUALIDADE_50095D	0,127	QUALIDADE_5160HI
-0,012	QUALIDADE_10V70C	0,011	QUALIDADE_1015B	0,134	QUALIDADE_50100A
-0,011	PF_Nb	0,012	QUALIDADE_5160HAL	0,139	9º SEQUENCIA TOTAL
-0,011	QUALIDADE_15V38K	0,012	QUALIDADE_5120HK	0,139	QUALIDADE_4320HB
-0,010	QUALIDADE_6150HG	0,012	PF_Zr	0,145	QUALIDADE_15V20F
-0,010	P11_Sb	0,013	12º SEQUENCIA	0,153	11º SEQUENCIA TOTAL
-0,010	QUALIDADE_5135HA	0,013	Seção_155mm	0,163	PF_C
-0,009	QUALIDADE_4340HC	0,013	QUALIDADE_1043HA	0,165	PF_Btot
-0,008	QUALIDADE_1540HK	0,014	QUALIDADE_5120HAL	0,167	5º SEQUENCIA
-0,008	CaSi (m)	0,014	QUALIDADE_10V70E	0,167	P11_S
-0,008	QUALIDADE_15B32HC	0,014	QUALIDADE_5120HA	0,173	QUALIDADE_1045D
-0,007	PF_Mo	0,014	QUALIDADE_5120HAM	0,201	7º SEQUENCIA TOTAL
-0,007	QUALIDADE_1015A	0,015	QUALIDADE_13V25HC	0,204	QUALIDADE_1055HP
-0,007	QUALIDADE_4142A	0,015	QUALIDADE_15B20C	0,206	Temperatura Liquidus
-0,007	QUALIDADE_1070A	0,016	PROCESSO ESCORIA nº4	0,277	2º SEQUENCIA
-0,006	QUALIDADE_1026A	0,016	QUALIDADE_1045AN	0,286	3º SEQUENCIA
-0,005	QUALIDADE_4320HAH	0,016	QUALIDADE_4130HI	1,147	Posição Meio
-0,003	QUALIDADE_3115HA	0,016	P11_Ti		
-0,003	QUALIDADE_4140A	0,016	Distribuidor		