

UNIVERSIDADE FEDERAL DE JUIZ DE FORA  
FACULDADE DE ENGENHARIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM COMPUTACIONAL

**RODRIGO BARBOSA DE SANTIS**

**PREVISÃO DE FALTA DE MATERIAIS NO CONTEXTO DE GESTÃO  
INTELIGENTE DE INVENTÁRIO: UMA APLICAÇÃO DE APRENDIZADO  
DESBALANCEADO**

JUIZ DE FORA

2018

**RODRIGO BARBOSA DE SANTIS**

**PREVISÃO DE FALTA DE MATERIAIS NO CONTEXTO DE GESTÃO  
INTELIGENTE DE INVENTÁRIO: UMA APLICAÇÃO DE APRENDIZADO  
DESBALANCEADO**

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora, na área de concentração em Sistemas Computacionais Aplicados, como requisito parcial para obtenção do título de Mestre em Modelagem Computacional.

Orientador: D. Sc., Leonardo Goliatt da Fonseca

Coorientador: D. Sc., Eduardo Pestana de Aguiar

JUIZ DE FORA

2018

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF  
com os dados fornecidos pelo(a) autor(a)

Santis, Rodrigo Barbosa de.

PREVISÃO DE FALTA DE MATERIAIS NO CONTEXTO DE GESTÃO  
INTELIGENTE DE INVENTÁRIO: UMA APLICAÇÃO DE APRENDIZADO  
DESBALANCEADO / RODRIGO BARBOSA DE SANTIS. – 2018.

72 f. : il.

Orientador: D. Sc., Leonardo Goliatt da Fonseca

Coorientador: D. Sc., Eduardo Pestana de Aguiar

Dissertação de (Mestrado) – UNIVERSIDADE FEDERAL DE JUIZ  
DE FORA, FACULDADE DE ENGENHARIA. PROGRAMA DE PÓS-  
GRADUAÇÃO EM MODELAGEM COMPUTACIONAL, 2018.

1. Palavra-chave. 2. Palavra-chave. 3. Palavra-chave. I. Sobrenome, Nome  
do orientador, orient. II. Título.

**RODRIGO BARBOSA DE SANTIS**

**PREVISÃO DE FALTA DE MATERIAIS NO CONTEXTO DE GESTÃO  
INTELIGENTE DE INVENTÁRIO: UMA APLICAÇÃO DE APRENDIZADO  
DESBALANCEADO**

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora, na área de concentração em Sistemas Computacionais Aplicados, como requisito parcial para obtenção do título de Mestre em Modelagem Computacional.

Aprovada em:

**BANCA EXAMINADORA**

---

D. Sc., Leonardo Goliatt da Fonseca  
Universidade Federal de Juiz de Fora

---

D. Sc., Eduardo Pestana de Aguiar  
Universidade Federal de Juiz de Fora

---

D. Sc., Heder Soares Bernardino  
Universidade Federal de Juiz de Fora

---

D. Sc., Ana Laura Cruz Fabiano  
OM Partners

## **AGRADECIMENTOS**

Aos orientadores Leonardo Goliatt da Fonseca e Eduardo Pestana de Aguiar, pelo incentivo, companheirismo e confiança para desenvolvimento desta proposta de estudo.

Aos colegas de mestrado, pela amizade e momentos compartilhados durante as disciplinas, trabalhos, projetos e congressos.

À CAPES, à Universidade Federal de Juiz de Fora e à MRS Logística, pela apoio e fomento da pesquisa desenvolvida.

À minha família, por todo o amor e carinho em todos os momentos que precisei de apoio.

Em especial, à minha tia Regina, por me ensinar a importância da formação continuada no processo de desenvolvimento humano.

“Estratégia é definir-se à parte da competição. Não é uma questão de ser melhor no que você faz, é uma questão de fazer diferente.”

Michael Porter

## RESUMO

Falta de materiais é um problema comum na cadeia de suprimentos, impactando o nível de serviço e eficiência de um sistema de inventário. A identificação de materiais com grande riscos de falta antes da ocorrência do evento pode apresentar uma enorme oportunidade de melhoria no desempenho geral de uma empresa. No entanto, a complexidade deste tipo de problema é alta, devido ao desbalanceamento das classes de itens faltantes e não faltantes no inventário, que podem chegar a razões de 1 para 100. No presente trabalho, algoritmos de classificação são investigados para proposição de um modelo preditivo para preencher esta lacuna na literatura. Algumas métricas específicas como a área abaixo das curvas de Característica Operacionais do Receptor e de Precisão-Abrangência, bem como técnicas de amostragem e comitês de aprendizado são aplicados nesta tarefa. O modelo proposto foi testado em dois estudos de caso reais, nos quais verificou-se que adoção da ferramenta pode contribuir com o aumento do nível de serviço em uma cadeia de suprimentos.

Palavras-chave: Gestão da cadeia de suprimentos. Planejamento e controle de inventário. Aprendizado desbalanceado. Métodos de amostragem. Comitês de classificadores.

## **ABSTRACT**

Material backorder (or stockout) is a common supply chain problem, impacting the inventory system service level and effectiveness. Identifying materials with the highest chances of shortage prior its occurrence can present a high opportunity to improve the overall company's performance. However, the complexity of this sort of problem is high, due to class imbalance between missing items and not missing ones in inventory, which can achieve proportions of 1 to 100. In this work, machine learning classifiers are investigated in order to fulfill this gap in literature. Specific metrics such as area under the Receiver Operator Characteristic and precision-recall curves, sampling techniques and ensemble learning are employed to this particular task. The proposed model was tested in two real case-studies, in which it was verified that the use of the tool may contribute with the improvement of the service level in the supply chain.

**Keywords:** Supply chain management. Inventory planning and control. Imbalanced learning. Sampling methods. Ensembles of classifiers.

## LISTA DE ILUSTRAÇÕES

- Figura 1 – Ciclo de novas tecnologias para PCS (GARTNER, 2015). Aplicações de aprendizado de máquina encontra-se na primeira fase de gatilho de inovação, sendo uma das grandes promessas de desenvolvimento na área para os próximos 5-10 anos. . . . . 16
- Figura 2 – De acordo com a classificação de aplicações de aprendizado de máquinas na GCS proposta por (WANG et al., 2016). O presente trabalho aborda uma aplicação de logística e inventário, cuja natureza é preditiva, e possui foco operacional no nível de processo. . . . . 17
- Figura 3 – Fatores de aumento de complexidade comuns a problemas de classificação em conjuntos desbalanceados são a) sobreposição de classes, no qual há ocorrências da classe positiva e negativa no mesmo espaço de entrada; e b) conjuntos distintos, onde as ocorrências da classe positiva espalham-se em pequenos grupos disjuntos. . . . . 21
- Figura 4 – As curvas PA, que apresentam a curva de compromisso existente entre as métricas precisão e abrangência, são utilizadas na avaliação de diferentes classificadores. No exemplo, o classificador A apresenta melhores níveis operacionais do que o classificador B, que por sua vez apresenta melhores níveis do que o classificador C. . . . . 23
- Figura 5 – As curvas COR, que apresentam a curva de compromisso existente entre as taxas de  $T_p$  e  $F_p$ , são utilizadas na avaliação de diferentes classificadores. No exemplo, o classificador A apresenta melhores níveis operacionais do que o classificador B, que por sua vez apresenta melhores níveis do que um classificador aleatório ( $ACOR = 0,5000$ ). . . . . 24
- Figura 6 – Representação gráfica de amostra de tamanho  $N = 5.000$  sorteada aleatoriamente e sem reposição, em duas dimensões a partir das duas componentes principais – método ACP. Verifica-se a ocorrência de sobreposição de classes e formação de conjuntos disjuntos nos espaços amostrais. Embora o conjunto de dados da MRS apresente menor razão de desbalanceamento, os conjuntos de ocorrência positiva são mais dispersos comparados ao conjunto de dados do Kaggle. . . . . 28
- Figura 7 – Estrutura de uma árvore de decisão (CART). No método, regras são definidas recursivamente em cada nó usando um dos atributos ( $x, z, \dots$ ) e um valor limite ( $c1, c2, \dots$ ), a fim de separar o conjunto de dados em dois novos grupos que seja o mais distinto o possível um do outro, usando um critério de separação conhecido como o ganho de informação ou entropia. . . . . 32

Figura 8 – Ilustração dos métodos de amostragem RUS e SMOTE. Enquanto o método RUS seleciona aleatoriamente uma amostra da classe negativa de tamanho igual ou definido da classe negativa, o método SMOTE gera novas instâncias a partir da similaridade entre as ocorrências da classe positiva existentes. . . . .	35
Figura 9 – Fluxograma do desenvolvimento da proposta do presente trabalho: 1) Definição de atributos de entrada, 2) Pré-processamento a partir de técnicas de normalização; imputação de dados faltantes e codificação; 3) Seleção de modelos a partir do ajuste fino de parâmetros; e 4) Avaliação de resultados a partir das curvas COR, PA e métricas de área sob a curva. . . . .	41
Figura 10 – Divisão do conjunto de treino em $K = 5$ dobras para validação cruzada. Cada um dos modelos é treinado usando 4 das dobras e sua métrica calculada em uma das dobras. O processo é realizado iterativamente, e a métrica de avaliação é dada pela média obtida nos conjuntos de validação. O conjunto de teste não é alterado, sendo este igual a 15% das amostras totais disponíveis no conjunto de dados. . . . .	44
Figura 11 – Histograma de atributos após a aplicação das técnicas de pré-processamento - Kaggle. A função de probabilidade é representada em preto enquanto os histogramas são visualizados como retângulos azuis. . . . .	47
Figura 12 – Curvas COR para os diferentes métodos em uma instância do conjunto de dados Kaggle. O classificador que apresentou maior capacidade de discriminação entre classes positivas e negativas foi o UB, enquanto o menor resultado foi obtido por RF. . . . .	48
Figura 13 – Curvas PA para uma instância do problema do conjunto de dados Kaggle. O classificador RF apresentou dominância em relação aos demais, enquanto o RUS apresentou o pior resultado. . . . .	49
Figura 14 – Representação gráfica em barras de erro dos resultados das métricas ACOR e ACPA na aplicação Kaggle. Os modelos baseados em comitê (RF, GB e UB) dominaram os demais, atingido melhores níveis em ambas as métricas definidas. . . . .	50
Figura 15 – Análise de importância de atributos de cada método na aplicação Kaggle. . . . .	54
Figura 16 – Histograma de atributos após a aplicação das técnicas de pré-processamento - MRS. A função de probabilidade é representada em preto enquanto os histogramas são visualizados como retângulos azuis. . . . .	56
Figura 17 – Curvas COR para os diferentes métodos em uma instância do conjunto de dados MRS. O classificador que apresentou maior capacidade de discriminação entre classes positivas e negativas foi o UB, enquanto o menor resultado foi obtido por GB. . . . .	57

Figura 18 – Curvas PA para uma instância do conjunto de dados MRS. O método RUS apresentou melhor curva para abrangência $< 0,25$ , enquanto o método LGST apresentou melhores resultados para valores maiores. . . . .	58
Figura 19 – Representação gráfica em barras de erro dos resultados das métricas ACOR e ACPA na aplicação MRS. O método LGST parece obter melhor resultados que os demais, enquanto dois grupos muito próximos são formados: o primeiro grupo com os comitês GB, RF e UB, e o segundo com os métodos CART, RUS e SMT. . . . .	59
Figura 20 – Análise de importância de atributos de cada método na aplicação MRS. . . .	62

## LISTA DE TABELAS

Tabela 1 – Matriz de confusão para um problema de classificação binário. Classes e previsões positivas e negativas são utilizadas para cálculo de métricas específicas: $V_p, V_n, F_p, F_n$ .	22
Tabela 2 – Síntese dos conjuntos de dados. Nas colunas são mostrados o número de atributos, amostras, ocorrências positivas e negativas, razão de desbalanceamento e nível de serviço são dados para cada um dos conjuntos utilizados no presente trabalho.	28
Tabela 3 – Parâmetros e amplitudes de busca adotados para os classificadores. A coluna 1 apresenta a sigla do método, a coluna 2 o nome do parâmetro, e a coluna 3 o conjunto de valores adotados para o parâmetro.	33
Tabela 4 – Parâmetros adotados para as técnicas de amostragem. A coluna 1 apresenta a sigla do método, a coluna 2 o nome do parâmetro, e a coluna 3 os valores adotados para o parâmetro.	35
Tabela 5 – Parâmetros adotados para os comitês. A coluna 1 apresenta a sigla do método, a coluna 2 o nome do parâmetro, e a coluna 3 os valores adotados para o parâmetro.	40
Tabela 6 – Lista de atributos utilizados em cada uma das aplicações, e técnicas de pré-processamento aplicadas a cada.	42
Tabela 7 – Resumo dos métodos adotados no presente trabalho, com o nome, sigla, referência e descrição sucinta.	44
Tabela 8 – Média e desvio-padrão obtidos no conjunto de teste na aplicação Kaggle em $N=30$ instâncias definidas para o problema.	49
Tabela 9 – Moda dos melhores parâmetros obtidos a partir do processo de ajuste de modelos na aplicação Kaggle.	50
Tabela 10 – Teste ANOVA das métricas para os resultados dos modelos na aplicação Kaggle. Um $p$ -valor menor do que 0,05 indica a rejeição da hipótese nula que afirma que a média das métricas de todos os métodos são iguais.	51
Tabela 11 – Teste de Tukey da métrica ACOR para os resultados dos modelos na aplicação Kaggle. Um $p$ -valor menor do que 0,05 indica a rejeição da hipótese nula de que as métricas possuem média igual.	51
Tabela 12 – Teste de Tukey da métrica ACPA para os resultados dos modelos na aplicação Kaggle. Um $p$ -valor menor do que 0,05 indica a rejeição da hipótese nula de que os dados não possuem distribuição normal.	52
Tabela 13 – Média desvio-padrão obtidos no conjunto de teste na aplicação MRS em $N=30$ instâncias definidas para o problema.	58
Tabela 14 – Moda dos melhores parâmetros obtidos a partir do processo de ajuste de modelos na aplicação MRS.	58

Tabela 15 – Teste ANOVA das métricas para os resultados dos modelos na aplicação MRS. Um $p$ -valor menor do que 0,05 indica a rejeição da hipótese nula que afirma que a média das métricas de todos os métodos são iguais. . . . .	59
Tabela 16 – Teste de Tukey da métrica ACOR para os resultados dos modelos na aplicação MRS. Um $p$ -valor menor do que 0,05 indica a rejeição da hipótese nula de que os dados não possuem distribuição normal. . . . .	60
Tabela 17 – Teste de Tukey da métrica ACPA para os resultados dos modelos na aplicação MRS. Um $p$ -valor menor do que 0,05 indica a rejeição da hipótese nula de que os dados não possuem distribuição normal. . . . .	61
Tabela 18 – Estatísticas descritivas do conjunto de dados Kaggle após pré-processamento.	70
Tabela 19 – Estatísticas descritivas do conjunto de dados MRS após pré-processamento.	70
Tabela 20 – Importância dos atributos dos modelos obtidos no conjunto de dados Kaggle.	71
Tabela 21 – Importância dos atributos dos modelos obtidos no conjunto de dados MRS. .	72

## LISTA DE ABREVIATURAS E SIGLAS

AC	Acurácia
ACOR	Área sob a curva COR
ACP	Análise de Componentes Principais
ACPA	Área sob a curva PA
CART	Árvore de classificação e regressão
COR	Características Operacionais do Receptor
FN	Falso Negativo
FP	Falso Positivo
GB	Gradient Boosting
GCS	Gestão da Cadeia de Suprimentos
GINI	Ganho de Informação
IVE	Índice de Variância Explicada
LGST	Regressão Logística
MVS	Máquina de Vetor Suporte
PA	Precisão-Abrangência
PCS	Planejamento da Cadeia de Suprimentos
RF	Floresta Randômica
RUS	Random Undersampling
SMT	Synthetic Minority Over-sampling Technique
UB	Under Bagging
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO . . . . .</b>	<b>15</b>
1.1	CONSIDERAÇÕES INICIAIS . . . . .	15
1.2	JUSTIFICATIVA . . . . .	15
1.3	ESCOPO DO TRABALHO: GESTÃO DE INVENTÁRIO . . . . .	18
1.4	OBJETIVOS . . . . .	19
1.5	ESTRUTURA DO TRABALHO . . . . .	20
<b>2</b>	<b>APRENDIZADO DESBALANCEADO . . . . .</b>	<b>21</b>
2.1	O PROBLEMA DO DESBALANCEAMENTO DE CLASSES . . . . .	21
2.2	MÉTRICAS DE AVALIAÇÃO . . . . .	22
2.3	MÉTODOS DE AMOSTRAGEM . . . . .	24
2.4	COMITÊS DE APRENDIZADO . . . . .	25
<b>3</b>	<b>MATERIAIS E MÉTODOS . . . . .</b>	<b>27</b>
3.1	CONJUNTOS DE DADOS . . . . .	27
3.2	PRÉ-PROCESSAMENTO . . . . .	29
3.2.1	Normalização . . . . .	29
3.2.2	Imputação de Dados Faltantes . . . . .	29
3.2.3	Codificação de Variáveis Categóricas . . . . .	30
3.3	SISTEMAS DE CLASSIFICADORES DE APRENDIZAGEM . . . . .	30
3.3.1	Regressão Logística – LGST . . . . .	30
3.3.2	Árvore de Classificação e Regressão – CART . . . . .	31
3.4	TÉCNICAS DE AMOSTRAGEM . . . . .	33
3.4.1	Sub-Amostragem Aleatória – RUS . . . . .	33
3.4.2	SMOTE – SMT . . . . .	34
3.5	COMITÊS DE APRENDIZADO . . . . .	34
3.5.1	Floresta Randômica – RF . . . . .	36
3.5.2	Gradient Tree Boosting – GB . . . . .	37
3.5.3	Under Bagging – UB . . . . .	39
3.6	SOFTWARE E HARDWARE . . . . .	39
<b>4</b>	<b>DESENVOLVIMENTO DA PROPOSTA . . . . .</b>	<b>41</b>
4.1	DEFINIÇÃO DE ATRIBUTOS . . . . .	41
4.2	PRÉ-PROCESSAMENTO . . . . .	43
4.3	SELEÇÃO DE MODELOS . . . . .	43
4.4	AVALIAÇÃO DE RESULTADOS . . . . .	45

<b>5</b>	<b>RESULTADOS E DISCUSSÃO</b> . . . . .	<b>46</b>
5.1	ESTUDO DE CASO I: KAGGLE . . . . .	46
5.2	ESTUDO DE CASO II: MRS LOGÍSTICA . . . . .	55
<b>6</b>	<b>CONCLUSÃO</b> . . . . .	<b>63</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>65</b>

# 1 INTRODUÇÃO

## 1.1 CONSIDERAÇÕES INICIAIS

A inteligência computacional tem causado profundas mudanças na indústria, ao passo que as barreiras de sua implementação – tais como custo computacional, poder de processamento e acessibilidade a plataformas abertas – vêm diminuindo cada vez mais. O aprendizado de máquina é a área de estudo que trata dos aspectos de projeto e desenvolvimento de modelos preditivos para apoio à tomada de decisão, provendo *insights* essenciais às empresas que buscam entender melhor e reagir às mudanças em sua operação.

Um tema há muito discutido na gestão de cadeia de suprimentos é o planejamento do inventário, uma atividade essencial para qualquer empresa que busca determinar a decisão de quando e quanto pedir, dados os diferentes mecanismos de controle adotados (TSOU, 2008). A maioria das abordagens propostas até então formulam a atividade como um problema de otimização mono ou multi-objetivo, nos quais custos de armazenagem e pedido deve ser mínimo, enquanto o nível de serviço é levado ao melhor possível. Uma abordagem diferente para gestão de inventário de forma mais eficiente – e complementar aos modelos desenvolvidos na literatura – é a identificação dos materiais com risco de falta antes do evento ocorrer (em inglês referenciado como *backorder* ou *stockout*), garantindo ao negócio tempo hábil para tomada de ações corretivas.

Uma complicação a este tipo de aplicação de aprendizado supervisionado surge, dado que em um sistema de inventário regular o número de materiais faltantes é significativamente inferior ao número de materiais ativos. Esta situação caracteriza o problema de desbalanceamento de classes (HE; MA, 2013; LÓPEZ et al., 2013; HE; GARCIA, 2009) e é comum a diversas outras das áreas de aplicações como telecomunicações (HOENS et al., 2012), internet (ZHU et al., 2003), finanças (PHUA et al., 2004; CHAN et al., 1999), biologia (RAO et al., 2006), entre outras (KUBAT et al., 1998); e requer a adoção de técnicas apropriadas para auxiliar a construção do modelo preditivo desejado.

O seguinte trabalho propõe a aplicação de modelos de aprendizado supervisionado para previsão de ocorrência de falta de materiais na gestão de inventário, baseando-se na combinação de métodos de amostragem e comitês de classificadores, e apresenta os resultados obtidos em dois estudos de casos em conjuntos de dados reais.

## 1.2 JUSTIFICATIVA

De acordo com o ciclo de novas tecnologias para o processo de Planejamento de Cadeia de Suprimentos (PCS), o aprendizado de máquina e desempenha um papel importante dentre as tecnologias emergentes nesta indústria, contando com um longo período previsto de desenvolvimento até atingir seu pico – de 5 a 10 anos (GARTNER, 2015).

A Figura 1, extraída do último relatório disponibilizado pela Gartner (GARTNER, 2015) e aqui não traduzida a fim de não alterar o conteúdo original do texto, apresenta o nível atual de desenvolvimento das tecnologias de PCS, examinados em 5 possíveis estágios. São eles: 1) gatilho de inovação, 2) pico de expectativas inflacionadas, 3) calha de desilusão, 4) declive da iluminação e 5) planalto de produtividade. A classificação adotada pela instituição é baseada no modelo econométrico de teoria de bolhas, proposto por (SMITH et al., 1988), inicialmente desenvolvido para aplicação no mercado de ações. Verifica-se que as aplicações de aprendizado de máquina e se encontra na fase de gatilho de inovação com previsão para atingir o pico nos próximos 5-10 anos, o que atrai o interesse de organizações e projetos neste tema.

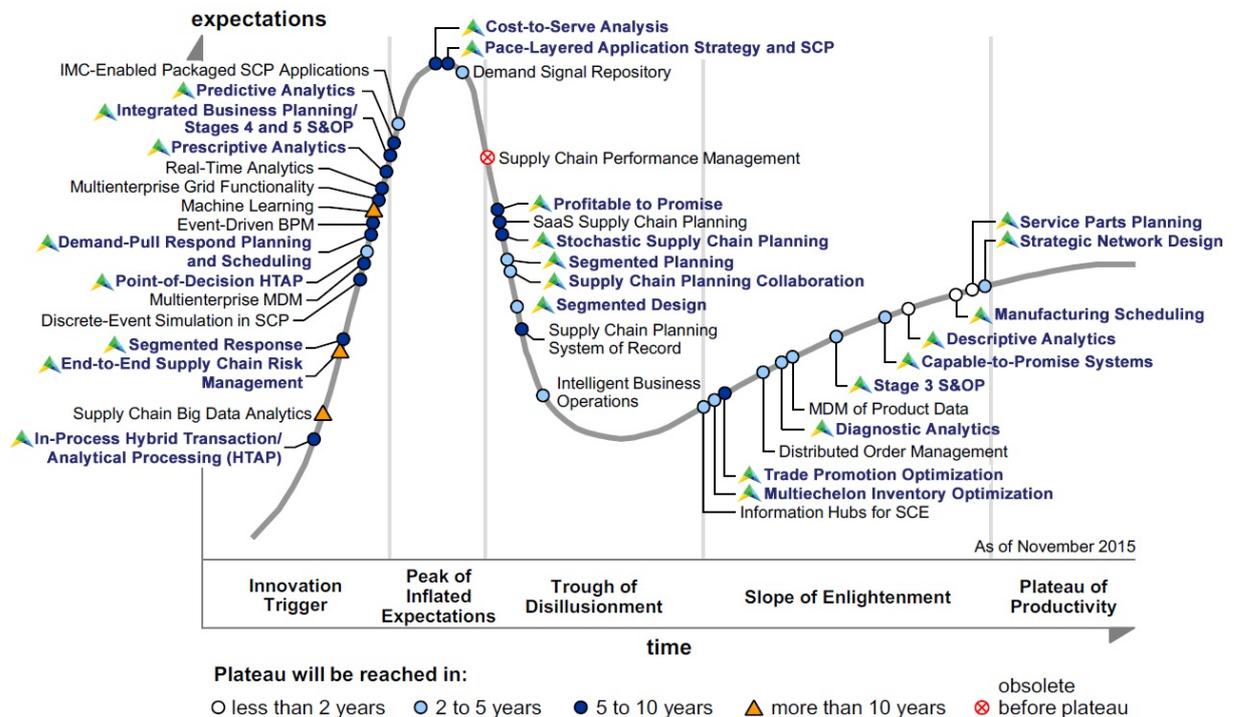


Figura 1 – Ciclo de novas tecnologias para PCS (GARTNER, 2015). Aplicações de aprendizado de máquina encontra-se na primeira fase de gatilho de inovação, sendo uma das grandes promessas de desenvolvimento na área para os próximos 5-10 anos.

Para pesquisadores e executivos, isto indica um enorme campo de trabalho para projetos e pesquisas em busca do desenvolvimento de modelos preditivos e ferramentas de análise que contribuam com a gestão mais eficiente das cadeias de suprimentos, que se tornam cada dia mais complexas com a globalização e mudança dos requisitos dos clientes. Destaca-se no entanto a necessidade que estes profissionais detenham o domínio de ambas habilidades analíticas e de entendimento sobre o negócio no contexto da gestão da cadeia de suprimento (GCS), a fim de formular as perguntas certas e serem capazes de analisar criticamente os resultados obtidos pelos métodos computacionais aplicados (WALLER; FAWCETT, 2013).

Quanto aos tipos de aplicações de aprendizado de máquina no contexto de GCS, é pro-

posta a classificação dos modelos de análise 1) quanto à sua natureza: descritiva, preditiva ou prescritiva; 2) quanto ao foco da operação: estratégico ou operacional; e 3) quanto ao nível de maturidade, avaliado a partir de quatro níveis: funcional, baseado em processo, colaborativo, ágil ou sustentável (WANG et al., 2016). Nas aplicações com foco estratégico destacam-se aquelas relacionadas ao processo de compras estratégicas, projeto de rede de cadeia de suprimentos e projeto e desenvolvimento de produto; enquanto nas aplicações de foco operacional destacam-se o processo de planejamento de demanda, compras, produção, logística e inventário, sendo este último parte do objeto de estudo deste trabalho. A aplicação de predição de demanda proposta é classificada de acordo com o modelo proposto por (WANG et al., 2016) conforme ilustrado na Figura 2.

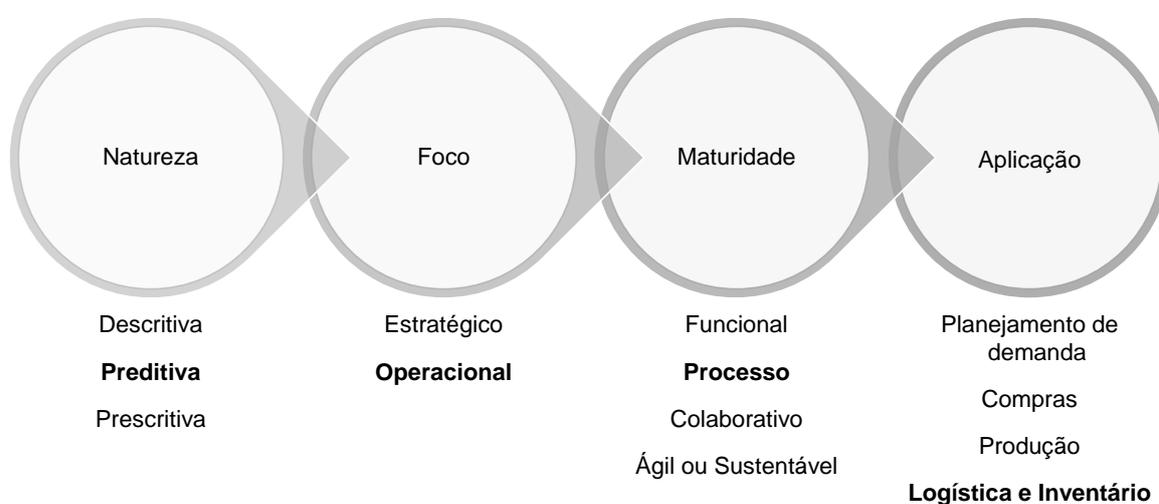


Figura 2 – De acordo com a classificação de aplicações de aprendizado de máquinas na GCS proposta por (WANG et al., 2016). O presente trabalho aborda uma aplicação de logística e inventário, cuja natureza é preditiva, e possui foco operacional no nível de processo.

Um exemplo que destaca como a aplicação de tais técnicas são capazes de promover melhorias benéficas na GCS pode ser identificado a partir da técnica de sensoriamento e modelamento da demanda, que se originou a partir do planejamento colaborativo da previsão de demanda e abastecimento de estoque e tem contribuído com resultados significativos na indústria de bens de consumo, através de previsões mais precisas capazes de responder a eventos do mundo-real como mudanças de mercado, mudanças climáticas, desastres naturais, comportamentos de consumo, etc (NGUYEN et al., 2017; FOLINAS; RABI, 2012).

A técnica de sensoriamento de demanda tem como base a utilização de mais informações internas à organização e sua cadeia do que apenas o histórico de vendas, também referidos como dados a jusante – por exemplo informação de produtos: histórico de vendas, categorias, promoções (BELVEDERE; GOODWIN, 2017; DONSELAAR et al., 2016; ALI et al., 2009), lançamentos: entrada/saída de linhas de produtos, plano de desaceleração (GARCIA, 2005); inventário: nível de estoque por inventário e vendas por canais (HARTZEL; WOOD, 2017).

A previsão de falta de materiais possui alta identidade com as aplicações de sensoria-mento e modelagem de demanda pois faz uso da mesma base de dados de entrada, porém sendo formulada como um problema de classificação em previsão de séries temporais – visto que a saída do modelo é discreta (item em falta *vs* item regular), e não contínua como a quantidade prevista para determinado produto num período futuro, objeto de saída de modelos de previsão de demanda.

### 1.3 ESCOPO DO TRABALHO: GESTÃO DE INVENTÁRIO

Os problemas de controle de inventário consistem na determinação da frequência de quantidade dos pedidos de forma a minimizar os custos totais ou maximizar os lucros da empresa. Muitos trabalhos foram desenvolvidos nesta área nas últimas décadas como por exemplo o modelo de lote econômico ou de quantidade de produção que buscam atingir o lucro máximo ou custo mínimo de operação (TALEIZADEH et al., 2011; PASANDIDEH et al., 2010; SAHA et al., 2010). Este tipo de abordagem é criticada dado que muitos sistemas de controle de inventário precisam operar com custo mínimo enquanto mantido um nível de serviço esperado pelos clientes, sendo estes dois objetivos conflitantes e incomensuráveis (TSOU, 2008).

O problema pode ser formulado de diferentes formas, dependendo do tipo da complexidade da aplicação e da cadeia de suprimentos: quanto ao nível de camadas escalonadas de planejamento – único ou multi-escalão; quanto ao número de funções objetivo – único ou multi-objetivo (TSOU, 2008); quanto ao número de produtos – produto único ou múltiplos produtos; quanto ao número de períodos no horizonte de planejamento – único ou múltiplos períodos (MOUSAVI et al., 2014); em tempo contínuo ou discreto (TÜRK et al., 2017); e quanto a natureza das variáveis – determinísticas ou estocásticas (GUTJAHR; PICHLER, 2016).

As funções objetivo são normalmente construídas a partir dos custos de armazenagem, pedido, depreciação do estoque, nível de serviço (numero de pedidos e/ou quantidade de itens do pedido não atendidos), risco de suprimento, dentre outros. Para problemas de natureza multi-objetivos, ferramentas de tomada de decisões multi-critérios podem ser adotadas de forma a decidir dentre o conjunto de soluções ótimas, qual melhor atende as expectativas dos decisores (TSOU, 2008).

Técnicas de otimização podem ser aplicadas como programação linear simples e programação linear mista inteira, ou otimização meta-heurística, como algoritmo evolutivo (TÜRK et al., 2017), otimização por enxame de partículas (MOUSAVI et al., 2014; TSOU, 2008), dentre outras. A escolha da melhor técnica de otimização é definida a partir do equilíbrio da formulação e complexidade do problema, e o tempo de resposta necessário para a resolução deste em sua aplicação. Em casos mais complexos, podem ser aplicadas regras simples, ou heurísticas, ou até combinações destas regras mais simples que podem ser desenvolvidas a partir de métodos de aprendizado automatizado de heurísticas, também conhecidos como hiper-heurísticas (BURKE et al., 2013).

A principal limitação destes modelos, são que os planos ótimos (ou quase-ótimo), por serem construídos em um dado momento específico de tempo a partir de dados registrados e cenários estáticos, podem estar sujeitos à reprogramação quase ao mesmo tempo em que são gerados, visto que os sistemas de inventários complexos que muitas empresas dispõem atualmente são dinâmicos. Outro problema é advindo da grande dificuldade de representação das variáveis que compõem o sistema, pois simples variáveis como a taxa de produção de um item, tempo de entrega, estão em constante mudança e atreladas a eventos desprezados na formulação do problema.

O modelo atual busca complementar o processo de planejamento ativo de inventário, possibilitando a identificação e resposta da companhia às situações de faltas, previstas a partir da construção de um sistema de aprendizado online, que leva em consideração fatores externos àqueles considerados no plano de abastecimento. Algumas premissas gerais do modelo proposto são destacadas a seguir:

1. A saída do modelo representa um risco associado a falta de cada componente, entre 0 e 1, e não um valor estimado de venda/consumo em um determinado período, como geralmente visto em modelos de previsão de demanda;
2. Previsão deve ser aplicada apenas em itens de requisições independentes, que são aqueles cuja demanda são externas à cadeia de suprimentos, tais como as de clientes finais. Isto porque qualquer alteração realizada a nível interno da cadeia precisa estar alinhada com as atividades de planejamento dos demais agentes que a compõem, de forma a evitar o acúmulo de estoque provocado pelo efeito chicote (LEE et al., 1997);
3. O período de horizonte de falta é tal que a empresa seja capaz de responder às situações de alto risco de falta, devendo ser este definido levando em consideração o tempo médio de entrega/produção e o tempo de respostas necessário para tomada de ações corretivas.

#### 1.4 OBJETIVOS

Embora a ocorrência de falta de materiais seja um fenômeno comum e de interesse da área de GCS, não foi encontrado na literatura nenhum modelo de predição baseado em aprendizado de máquina. Desta forma, o presente trabalho busca promover a discussão do projeto de modelos desta natureza, a partir da seguinte indagativa: Como (e quais) dados históricos podem ser utilizados na predição de falta de materiais e quais algoritmos de aprendizados devem ser utilizados, de forma a promover uma gestão de inventário mais eficiente?

Os objetivos secundários do trabalho são apresentados a seguir:

- Quais métricas de classificação devem ser utilizadas?
- Quais dos algoritmos (LR, CART, RUS, SMT, GB ou RF) apresentam melhor acurácia?

- Como o modelo preditivo pode ser integrado com o processo de planejamento operacional na prática?
- Quais dos atributos (inventário, histórico de vendas, previsão de vendas, classificação do material) apresentam maior contribuição para prever o problema de falta de material?

As principais conclusões obtidas com o estudo são:

- As métricas propostas ACOR e ACPA e a avaliação das curvas características apresentaram desempenho satisfatório para avaliação dos modelos;
- Não foi possível estabelecer um único algoritmo que obtivesse melhor resultado que os demais em ambas as métricas e/ou aplicação, no entanto verificou-se que a aplicação de comitês obtiveram resultados superiores quando comparados à aplicação dos classificadores ou técnicas de amostragem isolados;
- Os resultados obtidos nas aplicações de caso obtiveram resultados favoráveis a implementação do sistema de detecção de materiais faltantes, no entanto existe vasto campo para melhoria dos resultados dos modelos e avaliação deste em sistemas de inventários de outras empresas e segmentos;
- Os atributos definidos tiveram uma capacidade satisfatória em descrever o fenômeno previsto, com destaque aos relacionados ao nível de estoque, previsão, histórico de consumo e desempenho do fornecedor, sendo que os algoritmos mais complexos obtiveram maior êxito em considerar demais atributos de entrada.

## 1.5 ESTRUTURA DO TRABALHO

Este documento é organizado em 5 capítulos: seguindo esta Introdução, o Capítulo 2 provê uma revisão sobre o problema de desbalanceamento de classes em algoritmos de aprendizados, as métricas de avaliações específicas adotadas neste domínio e o estado da arte das técnicas aplicadas. O Capítulo 3 apresenta os materiais e métodos utilizados neste estudo. O Capítulo 4 apresenta os resultados e discussões provenientes do estudo em duas aplicações, comparando o desempenho dos diferentes modelos e atributos selecionados. Finalmente, o Capítulo 5 conclui o trabalho e inclui algumas recomendações para trabalhos futuros.

## 2 APRENDIZADO DESBALANCEADO

### 2.1 O PROBLEMA DO DESBALANCEAMENTO DE CLASSES

Em aprendizado supervisionado, um conjunto de dados é dito desbalanceado quando o número de instâncias de uma dada classe de interesse é raro quando comparado a outra (ou outras, no caso de problemas multi-classes). Este é um problema de interesse à pesquisa dado o diverso número de problemas de classificação com esta natureza no mundo real, tais como detecção de poluição, gestão de risco, detecção de fraudes e diagnóstico médico (LÓPEZ et al., 2013; GALAR et al., 2012; HAIXIANG et al., 2017).

Em tais aplicações, a razão de desbalanceamento entre a classe minoritária e majoritária pode atingir grandezas na ordem de 1:100, 1:1.000, e 1:10.000. A maior parte dos classificadores de aprendizagem, cujo treinamento é realizado através da minimização da taxa de erro, geralmente apresentam resultados insuficientes, por negligenciar as classes minoritárias – que são interpretadas pelo classificador como ruído. Diversos fatores podem contribuir com o aumento da complexidade do problema de desbalanceamento, como presença de pequenos grupos disjuntos de amostras minoritárias, sobreposição de classes dificultando a indução de um modelo genérico e insuficiência de exemplos da classe minoritária, conforme ilustrado pela Figura 3.

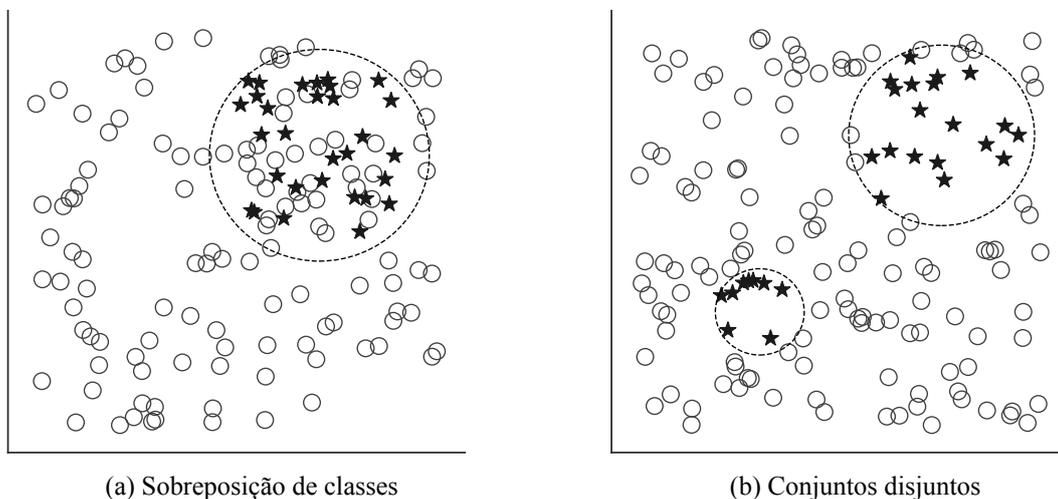


Figura 3 – Fatores de aumento de complexidade comuns a problemas de classificação em conjuntos desbalanceados são a) sobreposição de classes, no qual há ocorrências da classe positiva e negativa no mesmo espaço de entrada; e b) conjuntos distintos, onde as ocorrências da classe positiva espalham-se em pequenos grupos disjuntos.

Algumas abordagens internas são propostas de forma a prover modificações a algoritmos de classificação existentes, favorecendo o aprendizado das classes positivas; as abordagens externas são aplicadas no nível do conjunto de dados para ajustar a distribuição das classes antes da aplicação dos classificadores; enquanto o aprendizado sensível ao custo ocupa uma posição intermediária entre as abordagens internas e externas, dado que aplica transformações no nível

externo (dados), estabelecendo penalizações aos erros de classificações de cada instância, e no nível interno (algoritmos), por meio de incorporação dos custos durante o processo de treinamento (GALAR et al., 2012).

Para avaliar sistemas de aprendizado preditivos desenvolvidos em aplicações de aprendizado de classes desbalanceadas, métricas específicas de avaliação são necessárias e exploradas na próxima sessão.

## 2.2 MÉTRICAS DE AVALIAÇÃO

A escolha das métricas de avaliação apropriadas é um fator determinante para o projeto e desenvolvimento de um modelo preditivo (GALAR et al., 2012). Em um problema de classificação binário, a matriz de confusão – exibida na Tabela 1 – registra a quantidade das amostras correta e incorretamente classificadas de cada classe.

Tabela 1 – Matriz de confusão para um problema de classificação binário. Classes e previsões positivas e negativas são utilizadas para cálculo de métricas específicas:  $V_p$ ,  $V_n$ ,  $F_p$ ,  $F_n$ .

	Previsão positiva	Previsão negativa
Classe positiva	Verdadeiro positivo – $V_p$	Falso negativo – $F_n$
Classe negativa	Falso positivo – $F_p$	Verdadeiro negativo – $V_n$

A acurácia (Eq. 2.1) é tida como a medida padrão para problemas de classificação, e em um problema de classificação binário é obtida a partir da soma das ocorrências de verdadeiros positivos  $V_p$  e verdadeiros negativos  $V_n$  e a soma total de ocorrências geral, que também inclui as ocorrências de falsos positivos  $F_p$  e falsos negativos  $F_n$ .

$$Ac = \frac{V_p + V_n}{V_p + F_n + F_p + V_n} \quad (2.1)$$

No contexto de classes desbalanceadas, a acurácia não é uma métrica ideal, pois é incapaz de distinguir o número de exemplos classificados corretamente para diferentes classes isoladamente (HE; MA, 2013). Um exemplo típico é o caso de um estimador que classifica todas as amostras como negativas: embora a acurácia do modelo seja alta, este é incapaz de identificar separadamente a ocorrência de classe positivas – que é considerada pelo classificador como ruído.

Diversas métricas específicas são propostas de forma a considerar a distribuição das classes: a precisão ( $\frac{V_p}{V_p + F_p}$ ) expressa a acurácia de um estimador quando prevendo apenas as ocorrências da classe positiva, enquanto a sensibilidade ( $\frac{V_p}{V_p + F_n}$ ), também conhecida como taxa de verdadeiro positivo ou abrangência, indica a capacidade de identificar todas as ocorrências de amostras positivas existentes.

As curvas de precisão-abrangência (PA), exemplificadas na Figura 4, representam o conflito existente entre as métricas de precisão e sensibilidade e são comumente utilizadas em clas-

sificações binárias para compreender o comportamento da saída de um classificador, quando alterados os limites da função de ativação do modelo.

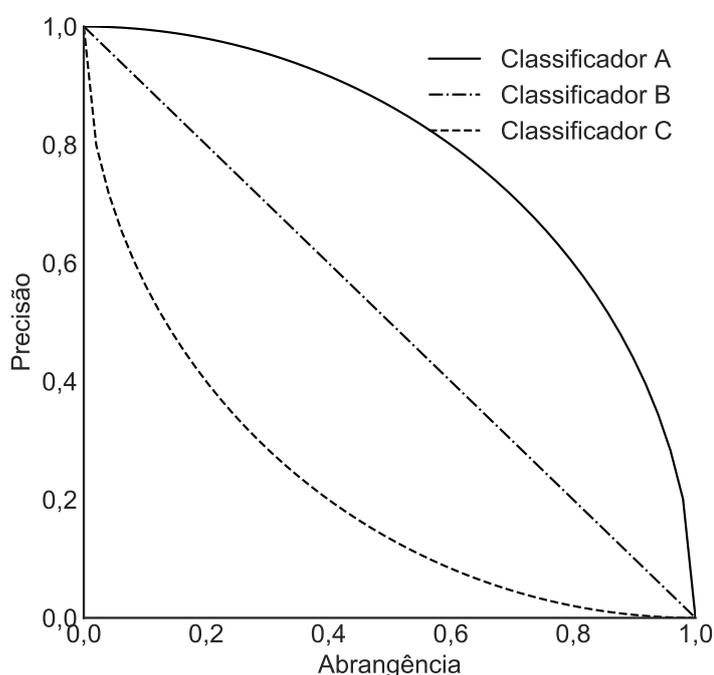


Figura 4 – As curvas PA, que apresentam a curva de compromisso existente entre as métricas precisão e abrangência, são utilizadas na avaliação de diferentes classificadores. No exemplo, o classificador A apresenta melhores níveis operacionais do que o classificador B, que por sua vez apresenta melhores níveis do que o classificador C.

Outra métrica de interesse obtida da análise da matriz de confusão é a taxa de falso positivos ou decaimento ( $\frac{F_p}{F_p+V_n}$ ), que representa o número de falsos positivos dividido pelo número total de amostras negativas. O decaimento é importante para a construção do gráfico das Características Operacionais do Receptor (COR), representado na Figura 5. Este gráfico permite a visualização do conflito existente entre precisão e decaimento, evidenciando que um dado classificador qualquer é incapaz de aumentar o número de verdadeiros positivos sem aumentar também o número de falsos positivos.

A área sob a curva COR (ACOR) corresponde à probabilidade de um sistema diferenciar um estímulo que seja apenas ruído de um que represente um sinal mais ruído. A métrica foi inicialmente desenvolvida para avaliar sistemas discriminadores de sinais de rádio e acabou se tornando uma referência padrão para problemas desbalanceados. No contexto de previsão de falta de materiais, a métrica está relacionada à capacidade dos modelos em distinguir materiais faltantes de materiais não-faltantes. Atualmente a área sob a curva COR é a mais adotada na avaliação de classificadores em problemas desbalanceados (HAIXIANG et al., 2017) e oferece uma medida única para avaliação do classificador mais adequado para um problema específico.

De forma análoga, a área abaixo da curva PA (ACPA) também é utilizada como métrica de avaliação, complementar à ACOR, pois um modelo que apresenta melhor desempenho em

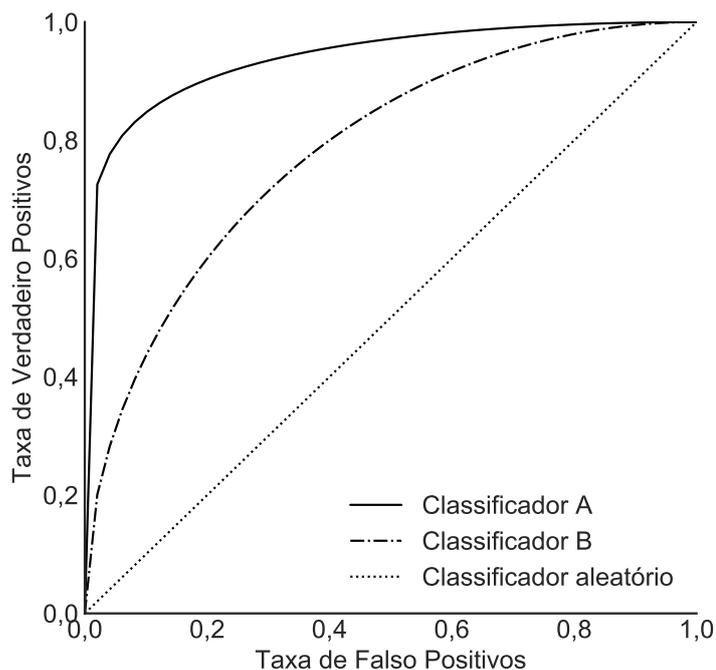


Figura 5 – As curvas COR, que apresentam a curva de compromisso existente entre as taxas de  $T_p$  e  $F_p$ , são utilizadas na avaliação de diferentes classificadores. No exemplo, o classificador A apresenta melhores níveis operacionais do que o classificador B, que por sua vez apresenta melhores níveis do que um classificador aleatório ( $ACOR = 0,5000$ ).

uma das métricas, não necessariamente apresenta um resultado melhor na outra. Experimentos práticos realizados por (LANDGREBE et al., 2006) mostraram que a ACPA é mais sensível a variações na taxa de desbalanceamento entre classes do que a ACOR. Por outro lado, a métrica ACPA é considerada mais informativa na avaliação de classificadores binários do que a adoção apenas da métrica ACOR (SAITO; REHMSMEIER, 2015).

### 2.3 MÉTODOS DE AMOSTRAGEM

Em sua essência, a ideia de transformar um conjunto de classes desproporcionais em um conjunto balanceado pode parecer uma solução razoável para o problema de desbalanceamento. Essa noção genérica levou ao desenvolvimento de diversas técnicas, conhecidas como métodos de amostragem, que podem ser agrupados em dois grandes grupos (GALAR et al., 2012): sub-amostragem, nas quais instâncias da classe majoritária são eliminadas para ajuste do equilíbrio; ou super-amostragem, onde instâncias da classe minoritária são replicadas de forma a igualar-se à classe majoritária. A aplicação destes métodos é justificada pelo já comprovado ganho no desempenho geral dos classificadores em conjunto de dados balanceados quando comparado à sua aplicação direta em conjuntos desbalanceados (LÓPEZ et al., 2013). Uma vantagem destes métodos é que como estes modificam o conjunto de dados, estes podem ser combinados com qualquer algoritmo de classificação (HE; GARCIA, 2009).

Diversas abordagens de amostragem já foram aplicadas até então são: a replicação (ou

eliminação) aleatória de instâncias de classes (LUSTIG et al., 2007); a sub-amostragem informada, que busca superar a deficiência da perda de informação aplicando regras específicas para determinar quais instâncias da classe majoritária serão abandonadas; a amostragem sintética, que busca a criação artificial de dados baseando-se na proximidade existente entre os exemplos da classe minoritária (CHAWLA et al., 2002); técnicas de limpeza de dados, aplicadas para remover a sobreposição de classes antes da aplicação de estimadores (KOTSIANTIS et al., 2006); e técnicas baseada em agrupamento, que produzem instâncias sintéticas para cada classe do problema baseado na média de cada agrupamento (HAN et al., 2005).

A combinação de estratégias de amostragem com comitês de aprendizado vêm sido amplamente discutida pela literatura, justificada pelos melhores resultados quando comparada a aplicação das técnicas isoladas (GALAR et al., 2012).

## 2.4 COMITÊS DE APRENDIZADO

Os comitês de aprendizado, também conhecidos como sistemas de múltiplos classificadores, são modelos de aprendizado supervisionado desenvolvidos a partir da combinação de classificadores individuais. Esta técnica é amplamente aplicada, dada sua capacidade de gerar resultados melhores do que o produzido por um classificador isolado (LÓPEZ et al., 2013). Em aprendizado desbalanceado, estes sistemas são classificados como comitês sensíveis ao custo ou comitês de pré-processamento (GALAR et al., 2012), de acordo com o método base empregado em sua construção.

Outra categorização dos modelos é dada de acordo com seu processo de construção. Comitês iterativos são aqueles nos quais os classificadores são treinados sequencialmente, de modo que cada novo classificador busca especializar seu aprendizado naquelas amostras às quais seus predecessores obtiveram piores resultados. Os algoritmos de *boosting* e alguns algoritmos evolutivos são os métodos mais aplicados neste grupo. Os comitês paralelos, nos quais os classificadores podem ser treinados de forma independente, apresentam desempenho comparável aos modelos iterativos e são mais recomendados para resolução de problemas práticos por apresentar ganhos no tempo de execução. Os métodos normalmente empregados neste grupo incluem *bagging*, comitês baseados em seleção de atributos e comitês baseados em amostragem (HAI-XIANG et al., 2017).

O desempenho de um comitê está diretamente associada à acurácia individual de cada classificador e à diversidade existente entre estes (WANG; YAO, 2009), sendo diversidade a unidade de medida que avalia a habilidade do grupo de classificadores em tomar diferentes decisões frente a um mesmo problema. Um equilíbrio entre acurácia e diversidade pode levar ao melhor desempenho de um comitê (WANG; YAO, 2013), no entanto, ainda é desconhecida a relação de compromisso entre as duas variáveis.

A aplicação de um único modelo específico para resolução de todos os tipos de proble-

mas de aprendizado desbalanceado é tida como insuficiente, dada as inúmeras variações possíveis nas taxas de desbalanceamento, número de atributos e classes (YIJING et al., 2016). A aprendizagem do modelo possui um alto grau de incerteza na construção do comitê, que pode ser afetado pela forma em que as amostras são selecionadas para o treino, as escolhas dos classificadores base e regras. Deste modo, (REN et al., 2017) propõem a utilização de aprendizado adaptativo para determinação automática das melhores taxas de amostragem, enquanto para o caso de aprendizado sensível ao custo, (KRAWCZYK et al., 2014) propõem o aprendizado adaptativo para determinação da matriz de custo das instâncias de classe.

### 3 MATERIAIS E MÉTODOS

#### 3.1 CONJUNTOS DE DADOS

Neste trabalho, são considerados o conjunto de dados disponível no Kaggle "*Can You Predict Product Backorders?*"<sup>1</sup> e um conjunto de dados construído a partir de dados fornecidas pela operadora ferroviária MRS Logística<sup>2</sup>.

O conjunto de dados do Kaggle é proposto por uma empresa que preferiu não identificar sua razão social ou setor de atuação. Por se tratar de um conjunto de dados para teste e não uma competição oficial, não há definição formal de métricas de avaliação e, conseqüentemente, não há uma lista de classificação de predições/modelos, como visto em outras competições na plataforma. No entanto, a proposta de prever os materiais faltantes no inventário atraiu a atenção de diversos profissionais e pesquisadores da área, com a criação de dezenas de *kernels* públicos para análise exploratória dos dados e aplicação de métodos preditivos. Em menos de 9 meses, os três *kernels* públicos mais visitados receberam juntos mais de 8.600 visitas. O maior problema enfrentado dentre os pesquisadores no entanto é a falta de definição de métricas de avaliação para o problema em aberto.

O conjunto de dados da MRS é construído a partir da proposta anterior, sendo adaptado à operadora logística a partir das informações disponíveis nos bancos de dados da organização. É do interesse da empresa o aumento da disponibilidade de materiais no inventário, em especial aqueles relacionados a manutenções críticas, dado que um dos objetivos estratégicos da empresa é o aumento da eficiência geral de equipamentos (NAKAJIMA, 1988), sendo a falta de materiais uma das principais causas de indisponibilidade de fatores produtivos com altos custos de oportunidade, como locomotivas e outros equipamentos chaves.

A Tabela 2 sintetiza as propriedades dos conjunto de dados: o número de atributos, o número de ocorrências da classe positiva, o número de ocorrências de classes positivas, o número total de registros, a razão de desbalanceamento e o nível de serviço atual do sistema. O nível de serviço do sistema de inventário é apresentado e o objetivo da proposta é auxiliar na identificação de componentes com maior risco de falta antes do evento, de forma que ações de curto prazo possam ser realizadas para mitigar riscos e aumentar o desempenho geral do sistema. Os conjuntos de dados contém registros de níveis de estoque e quantidades em movimento dado um momento específico, e a predição é realizada para as próximas 8 semanas.

A Figura 6 apresenta uma representação 2D de uma dada amostra de tamanho  $N = 5.000$  de cada um dos conjuntos de dados, obtida através do gráfico de dispersão da decomposição dos atributos nos dois componentes mais representativos, utilizando o método de análise de

<sup>1</sup> Kaggle é uma plataforma para competições de modelos preditivos e analíticos a partir de conjuntos de dados enviados por empresas e usuários. <<https://www.kaggle.com/>>

<sup>2</sup> MRS Logística S.A. é uma concessionária responsável pela operação da malha ferroviária federal da região sudeste. <<https://www.mrs.com.br/>>

Tabela 2 – Síntese dos conjuntos de dados. Nas colunas são mostrados o número de atributos, amostras, ocorrências positivas e negativas, razão de desbalanceamento e nível de serviço são dados para cada um dos conjuntos utilizados no presente trabalho.

Conjunto de dados	# Atributos	# Positivos	# Negativos	# Total	Desb.	Nív. Serviço
Kaggle	22	13.981	1.915.954	1.929.936	1:137	99,27%
MRS	17	321	16.320	16.641	1:50	98,03%

componentes principais (ACP), proposto por (JOLLIFFE, 1986). As amostras foram sorteadas de forma aleatória sem reposição e para fins representativos. O índice de variância explicada (IVE) é a medida que representa o quanto as novas componentes, geradas a partir da combinação linear dos vetores de entrada, explicam os atributos originais.

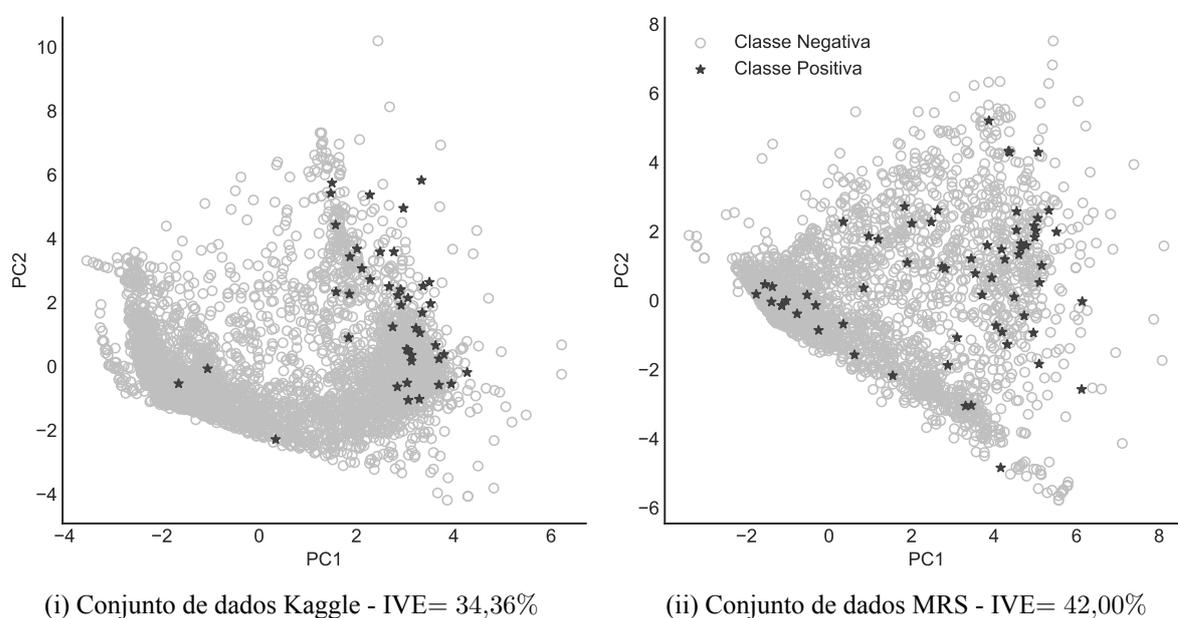


Figura 6 – Representação gráfica de amostra de tamanho  $N = 5.000$  sorteada aleatoriamente e sem reposição, em duas dimensões a partir das duas componentes principais – método ACP. Verifica-se a ocorrência de sobreposição de classes e formação de conjuntos disjuntos nos espaços amostrais. Embora o conjunto de dados da MRS apresente menor razão de desbalanceamento, os conjuntos de ocorrência positiva são mais dispersos comparados ao conjunto de dados do Kaggle.

A ocorrência de classes positivas (minoritária) é bem menos comum do que a classe negativa (majoritária), o que caracteriza a complexidade de desbalanceamento destes conjuntos. O primeiro conjunto de dados apresenta menor ocorrência relativa de classe positivas, porém estas estão mais concentradas entre si, com alguns pequenos casos isolados em meio aos itens normais. No segundo estudo de caso, existem mais ocorrências relativas de casos positivos, no entanto, estas são mais dispersas por todo o espaço bi-dimensional construído.

## 3.2 PRÉ-PROCESSAMENTO

### 3.2.1 Normalização

Normalização é o processo de pré-processamento no qual cada amostra com pelo menos um componente diferente de zero é redimensionada independentemente das demais amostras, de modo que sua norma seja igual a um. A transformação das entradas para normas unitárias é uma operação muito comum nas aplicações de classificação de texto, clusterização de instâncias ou processamento de imagens (PEDREGOSA et al., 2011).

Os tipos de normalização mais comum são realizados a partir das normas L1, L2 e Max, descritas pelas equações

$$L1 : z = \|x\|_1 = \sum_{i=1}^n |x_i| \quad (3.1)$$

$$L2 : z = \|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2} \quad (3.2)$$

$$\text{Max} : z = \max x_i \quad (3.3)$$

em que  $z$  é o valor da norma, e  $x_i$  o valor original da variável  $i$ , para  $i = 1, 2, \dots, n$ .

A normalização pode ser realizada por atributo (coluna) ou por amostra (linha), de acordo com a escolha do especialista. A vantagem da utilização de atributos normalizados como entrada em algoritmos de aprendizado supervisionado é que a técnica proporciona resultados mais estáveis (quanto a convergência dos modelos) e/ou superiores comparados com a aplicação direta dos atributos (DIETTERICH, 2000).

Para identificar a necessidade e o melhor tipo de normalização, são necessários testes junto aos modelos preditivos adotados a fim de verificar se a normalização contribui com o aumento da eficiência das métricas adotadas, e se a aplicação da estratégia é coerente com a aplicação desenhada.

### 3.2.2 Imputação de Dados Faltantes

Em mineração de dados em bases de dados reais, é extremamente comum a ocorrência de dados faltantes, que são campos não preenchidos ou não encontrados na obtenção de uma amostra. Isso pode ocorrer por diversos fatores, internos ou externos, como custo de obtenção, falha em equipamentos, esquecimento, dentre outros. A fim de não prejudicar o treinamento dos modelos, algumas estratégias podem ser aplicadas para correção destas irregularidades, desde a exclusão de amostras do conjunto de dados analisado, até métodos mais complexos que utilizam classificadores para prever e completar os atributos faltantes a partir de atributos existentes, como o exemplo dos algoritmos KNNimpute e SVDimpute (TROYANSKAYA et al., 2001).

Uma estratégia muito aplicada é a inferência dos dados faltantes a partir da porção conhecida dos dados, inserindo nos campos faltantes o valor de uma medida de tendência central da série, como média ou mediana (PEDREGOSA et al., 2011). A principal vantagem da aplicação desta é evitar a perda de dados que possam ser valiosos (embora incompletos), com um baixo custo computacional de pré-processamento.

### 3.2.3 Codificação de Variáveis Categóricas

Atributos categóricos precisam ser convertidos em numéricos para que possam ser utilizados nos algoritmos de aprendizado. Os tipos de transformações mais comuns são a codificação binária para atributos que possuem apenas duas classes (ex: sim/não); codificação de rótulos, onde possíveis classes são enumeradas e associadas com um identificador numérico, de forma em que esta ordem exprime uma relação de proximidade entre as classes; e por fim na codificação uma-de- $K$  (ou *dummy coding*), em que as  $m$  classes possíveis são transformadas em  $m$  atributos binários, com apenas um desses atributos ativo, de forma que cada classe é independente das demais.

## 3.3 SISTEMAS DE CLASSIFICADORES DE APRENDIZAGEM

Sistemas de classificadores de aprendizagem são algoritmos únicos capazes de gerar modelos e realizar tarefas complexas como classificação de eventos (JI et al., 2013), imagens (CHAN et al., 2015), áudio (LEE et al., 2009) dentre outras. Estes modelos são desenvolvidos através de um processo de aprendizado supervisionado, onde parâmetros são ajustados – ou regras são construídas, dependendo da natureza do algoritmo – enquanto a taxa de erro ou outra métrica específica é minimizada ou maximizada (PRINCIPE et al., 2000).

Alguns destes algoritmos apresentam modificações internas para manuseio de conjunto de dados desbalanceados, tais como aqueles aplicados neste estudo. Outro benefício dos seguintes modelos é que eles são capazes de prover um valor de saída entre 0 e 1, que pode ser interpretado como a probabilidade de falta. Assim, um valor de corte pode ser adotado, tornando o projeto dos modelos mais flexível.

### 3.3.1 Regressão Logística – LGST

A regressão logística é uma das abordagens mais simples e populares para classificação em problemas binominais. A diferença entre a regressão logística e a regressão linear está nas escolhas do modelo paramétrico e suas premissas. O método apresenta sensibilidade aumentada no aprendizado desbalanceado, pois seu treinamento é baseado na maximização do logaritmo da probabilidade binomial. Dado sua sofisticação e simplicidade, o método é aplicado para prover uma solução base para o problema apresentado.

Considera-se uma coleção de  $p$  variáveis independentes denotada pelo vetor  $X'$ . Seja a probabilidade condicional representada por  $P(Y = 1|x) = \pi(x)$ . O modelo de regressão logística múltipla é dado pela equação (HOSMER JR et al., 2013)

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (3.4)$$

no qual o modelo de regressão logística é

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}} \quad (3.5)$$

em que  $\pi(x)$  é a transformação logística,  $g(x)$  o valor previsto do método, e  $\beta_j$  o coeficiente ajustado para cada variável de entrada  $j = 1, 2, \dots, p$ .

A inclusão direta no método de variáveis categóricas é inapropriada, sendo necessária a codificação destas em variáveis numéricas representativas, conforme explicado na Subseção 3.2.3. Os parâmetros de configuração possíveis para o método são o tipo de penalização utilizado na normalização (L1 ou L2), e o coeficiente de regularização, responsável pela flexibilização do coeficiente atribuído a cada uma das variáveis.

Assumindo uma amostra de  $n$  observações independentes  $(x_i, y_i)$ ,  $i = 1, 2, \dots, n$ , o ajuste do método requer estimação do vetor  $\beta' = (\beta_0, \beta_1, \dots, \beta_p)$ . As equações de probabilidades resultantes são expressas por

$$\sum_{i=1}^n [y_i = \pi(x_i)] = 0 \quad (3.6)$$

e

$$\sum_{i=1}^n x_{ij} [y_i = \pi(x_i)] = 0 \quad (3.7)$$

para  $j = 1, 2, \dots, p$ . Seja  $\hat{\beta}$  a solução destas equações, os valores do modelo de regressão logística são  $\hat{\pi}(x_i)$ , o valor da expressão na Equação 3.5 é obtida usando  $\hat{\beta}$  e  $x_i$ .

### 3.3.2 Árvore de Classificação e Regressão – CART

A árvore de classificação e regressão (BREIMAN et al., 1993) é um modelo preditivo baseado na construção de árvores binárias, na qual regras são associadas a cada ramo criado no processo de treinamento supervisionado. A adoção de critérios de separação resulta na insensibilidade do método às variações nas distribuições de classes na amostra (DRUMMOND; HOLTE, 2000). Embora o CART seja considerado um algoritmo fraco, o modelo apresenta alguns benefícios interessantes tais como legibilidade, robustez a *outliers* e capacidade de seleção de atributos.

O método trabalha a partir do particionamento recursivo dos dados, em um processo gradual que divide os dados sequencialmente em grupos menores. Uma característica importante deste método é que neste processo, os estágios anteriores não são revisitados após o conhecimento dos estágios posteriores.

A função básica de particionamento é dada por

$$f(\mathbf{X}) = \sum_{j=1}^p \sum_{m=1}^{M_j} \beta_{jm} h_{jm}(\mathbf{X}) \quad (3.8)$$

em que cada um dos  $p$  preditores possui seu próprio conjunto de transformações, e todas as transformações de todos os preditores, cada qual com seu próprio peso  $\beta_{jm}$ , são combinadas de forma linear.

A Figura 7 mostra uma representação visual do funcionamento do método, no qual regras de particionamento são definidas em cada um dos nós, de modo a gerar terminações para classificação das amostras. Os atributos de cada observação (neste caso,  $x$  e  $z$  são os atributos avaliados) e comparados a um valor referência ( $c1$  e  $c2$ ), caso a condição seja verdadeira, a observação pode ser alocada em uma terminação ou ser avaliada por outro nó.

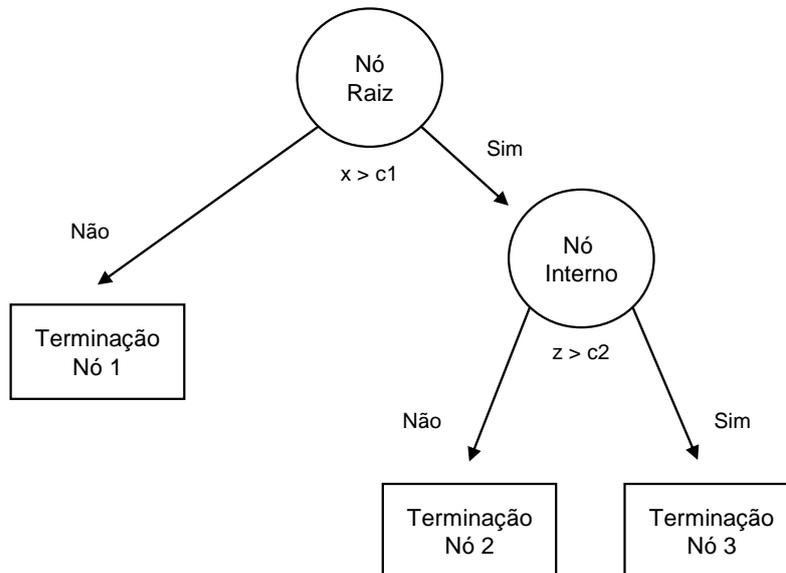


Figura 7 – Estrutura de uma árvore de decisão (CART). No método, regras são definidas recursivamente em cada nó usando um dos atributos ( $x, z, \dots$ ) e um valor limite ( $c1, c2, \dots$ ), a fim de separar o conjunto de dados em dois novos grupos que seja o mais distinto o possível um do outro, usando um critério de separação conhecido como o ganho de informação ou entropia.

Cada subconjunto da árvore é criado de forma que os valores dos atributos de cada grupo seja o mais similar possível, e para isso é adotado um critério de separação. São dois os critérios de separação mais aplicados, Gini ou Entropia ( $H$ )

$$Gini(E) = 1 - \sum_{j=1}^c p_j^2 \quad (3.9)$$

ou

$$H(E) = - \sum_{j=1}^c p_j \log p_j \quad (3.10)$$

em que  $p_j$  é a saída probabilística do modelo que descreve o risco de falta do material.

A entropia pode apresentar melhores resultados quando comparada ao ganho de informação em aplicações desbalanceadas, no entanto requer maior custo computacional para o cálculo do logaritmo da probabilidade. A profundidade máxima da árvore e o número mínimo de amostras por folha são alguns parâmetros normalmente adotados a fim de evitar-se sobre-ajuste no treino. Os parâmetros e amplitude de busca adotados para os classificadores são exibidos na Tabela 3.

Tabela 3 – Parâmetros e amplitudes de busca adotados para os classificadores. A coluna 1 apresenta a sigla do método, a coluna 2 o nome do parâmetro, e a coluna 3 o conjunto de valores adotados para o parâmetro.

Método	Parâmetros	Amplitude
LGST	Penalização	L1, L2
	Regularização - $C$	1, 10, 100, 1000
CART	Critério	Gini, Entropia
	Profundidade máx.	3, 4, 5, 7, 10, 15, 20, 30, 50
	Mín. amostra folha	5

### 3.4 TÉCNICAS DE AMOSTRAGEM

A amostragem é utilizada para balancear as classes antes da aplicação do estimador, considerado que o desempenho geral de um algoritmo de classificação é melhor em conjunto de dados balanceados. Dado que a amostragem é uma abordagem externa e não possui capacidade de classificação por si só, um classificador deve ser aplicado em conjunto aos métodos apresentados.

#### 3.4.1 Sub-Amostragem Aleatória – RUS

No método de sub-amostragem aleatória, um novo conjunto é criado a partir da combinação do conjunto de todos os exemplos da classe minoritária e uma amostra (com ou sem reposição) da classe majoritária de tamanho igual ou pré-definido (HE; GARCIA, 2009). Um problema particular que pode ocorrer a partir da aplicação deste método é a perda de alguma informação relevante sobre a classe majoritária, prejudicando o desempenho do modelo quando aplicado a todo o conjunto de dados (LUSTIG et al., 2007). A proporção entre as classes pode também ser ajustada, ao invés do padrão 1:1, embora no atual estudo esta tenha sido mantido constante.

### 3.4.2 SMOTE – SMT

A técnica de super-amostragem SMOTE (Synthetic Minority Over-sampling Technique) (CHAWLA et al., 2002) cria dados artificiais baseando-se na similaridade dos atributos existente no espaço das amostras minoritárias, a partir dos  $K$ -vizinhos mais próximos no espaço Euclidiano. Para gerar uma amostra sintética, um vizinho aleatório é selecionado e multiplicado pelo correspondente vetor de diferença e um valor aleatório entre 0 e 1.

A diferença  $\delta$  entre o valor de dois atributos é definida pela distância geométrica em um espaço de valores finitos (COST; SALZBERG, 1993), calculada por

$$\delta(\mathbf{V}_1, \mathbf{V}_2) = \sum_{i=1}^n \left| \frac{\mathbf{C}_{1i}}{\mathbf{C}_1} - \frac{\mathbf{C}_{2i}}{\mathbf{C}_2} \right|^k \quad (3.11)$$

em que  $\mathbf{V}_1$  e  $\mathbf{V}_2$  são os vetores correspondentes de valor,  $\mathbf{C}_j$  é o vetor com o número total de ocorrências dos valores dos atributos  $V_j$ , e  $\mathbf{C}_{ji}$  o número de ocorrências do valor  $V_j$  para a classe  $i$ , para  $j = 1, 2$ , e  $k$  é uma constante de valor usualmente igual a 1. A distância entre dois vetores de atributos é dada por

$$\Delta(\mathbf{X}, \mathbf{Y}) = w_x w_y \sum_{i=1}^N \delta(x_i, y_i)^r \quad (3.12)$$

em que  $w_x$  e  $w_y$  são os pesos relativos e  $r$  a constante de valor. Quando  $r = 1$ , este refere-se à distância Manhattan e, quando  $r = 2$ , à distância Euclidiana (COST; SALZBERG, 1993).

Algumas variações do método também muito populares são o Borderline-SMOTE (HAN et al., 2005), no qual apenas a minoria de exemplos próximos a linha de fronteira é super amostrada, e o SVM-SMOTE (TANG et al., 2009), que de forma análoga gera as novas instâncias minoritárias próximas às linhas de fronteira estabelecidas, com o auxílio de um classificador máquina de vetor suporte (MVS).

A Figura 8 apresenta de forma ilustrativa a aplicação dos dois métodos de amostragem a um conjunto de dados exemplo. No método RUS, o conjunto de classes negativa é reduzido a partir da sub-amostragem de suas classes, enquanto no método SMOTE, novas ocorrências da instância positiva são criadas entre as classes já existentes.

O número de vizinhos mais próximos  $K$  a serem considerados na seleção por amostragem é um dos parâmetros do modelo, sendo 5 o número adotado no presente estudo. A proporção adotada é de 1:1. Os demais parâmetros e amplitude de busca adotados para os métodos de amostragem são exibidos na Tabela 4.

## 3.5 COMITÊS DE APRENDIZADO

Comitês são utilizados para combinar a predição de diversos estimadores bases construídos com um dado algoritmo de aprendizado, a fim de melhorar a capacidade de generalização dos

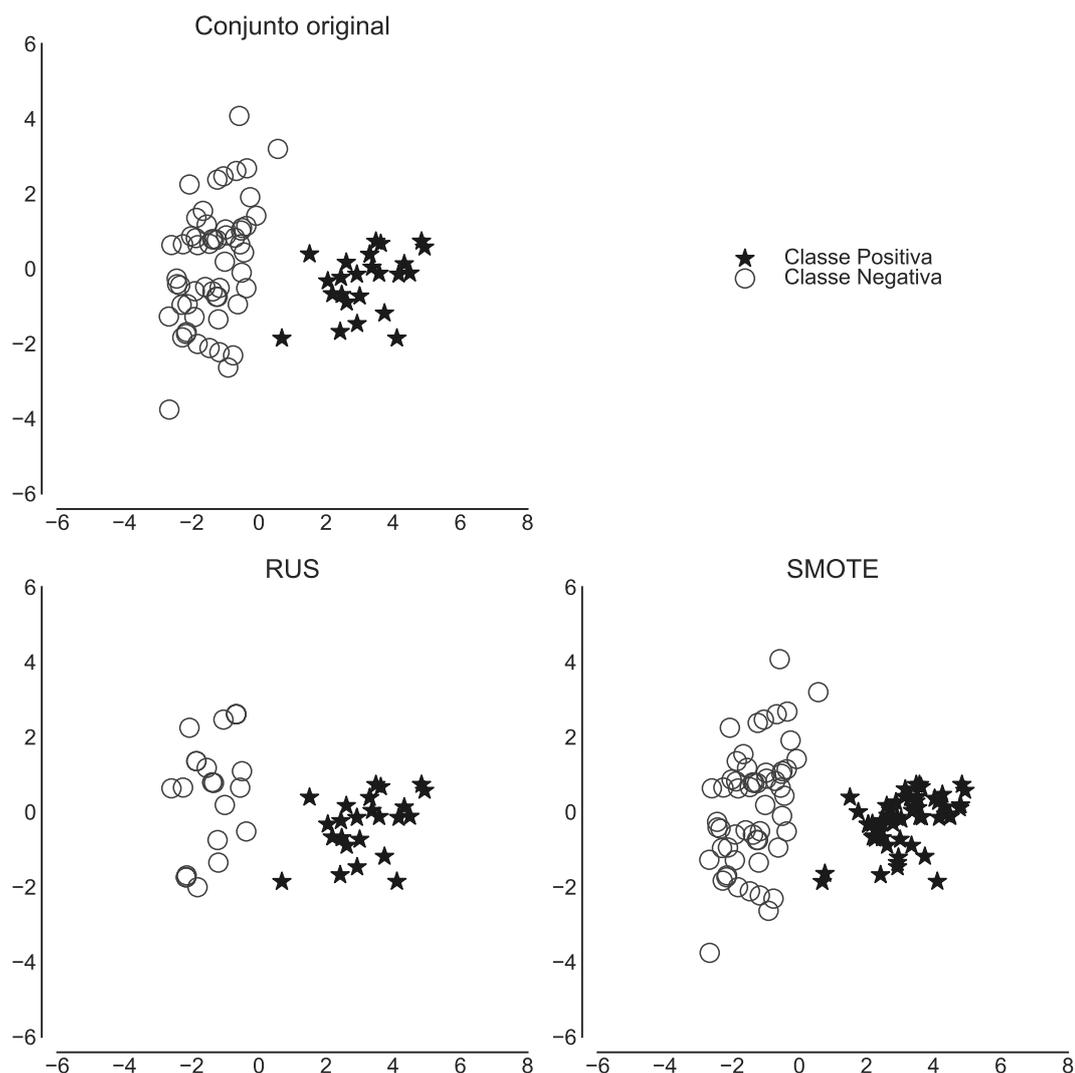


Figura 8 – Ilustração dos métodos de amostragem RUS e SMOTE. Enquanto o método RUS seleciona aleatoriamente uma amostra da classe negativa de tamanho igual ou definido da classe negativa, o método SMOTE gera novas instâncias a partir da similaridade entre as ocorrências da classe positiva existentes.

Tabela 4 – Parâmetros adotados para as técnicas de amostragem. A coluna 1 apresenta a sigla do método, a coluna 2 o nome do parâmetro, e a coluna 3 os valores adotados para o parâmetro.

Método	Parâmetros	Valor
RUS	Proporção	1:1
SMT	Distância	Euclidiana
	Proporção	1:1
	Número de vizinhos - $K$	5

modelos quando comparado à de um único estimador (GALAR et al., 2012). Duas abordagens básicas podem ser exploradas: no *bagging*, estimadores são treinados de forma independente e a média de sua predição é atribuída ao comitê, enquanto a técnica de *boosting* constrói estimadores sequencialmente buscando melhores resultados no domínio no qual o estimador anterior

não tenha sido satisfatório. A motivação para utilização destes métodos, especialmente combinado com a técnica de amostragem, é a incorporação de diversos estimadores fracos em comitês robustos com alta capacidade de distinção das classes majoritária e minoritária.

### 3.5.1 Floresta Randômica – RF

Floresta randômica é um comitê construído utilizando uma amostra (técnica conhecida também como *bootstrap*), onde os lotes de treino são sorteados com reposição (LIAW et al., 2002). Durante a construção da árvore, a divisão selecionada é a melhor dentre uma sub-amostra aleatória de atributos, levando a uma aleatoriedade que tende a gerar resultados melhores da floresta comparado a uma única árvore não-randômica. O aumento do viés é compensado pela diminuição da variação média, obtida pela combinação da probabilidade da predição dos classificadores base.

Dado um comitê de classificadores  $h_1(x), h_2(x), \dots, h_K(x)$  e com o conjunto de treino amostrado aleatoriamente das distribuições dos vetores aleatórios  $x$  e  $y$ , a função marginal é definida por (BREIMAN, 2001)

$$mg(x,y) = av_k I(h_k(x) = y) - \max_{j \neq y} av_k I(h_k(x) = j) \quad (3.13)$$

em que  $I(\cdot)$  é a função indicadora. A função marginal mede a extensão na qual a média do número de votos em  $(x,y)$  para a classe correta excede a média de votos para qualquer outra classe, sendo que quanto maior a margem, maior a confiança na classificação. O erro de generalização é dado por

$$PE^* = P_{x,y}(mg(x,y) < 0) \quad (3.14)$$

onde  $x,y$  indica a probabilidade sobre o espaço  $x,y$ . O pseudocódigo com o processo de execu-

ção do método é dado pelo Algoritmo 1.

**Entrada:** Conjunto de treino  $D$  com  $N$  amostras, profundidade de cada árvore  $K$ , taxa de subamostragem  $\eta$ , número de atributos  $d$  e número de estimadores base  $M$

**Saída:** Aproximação  $\hat{f}(x)$  para  $f(x)$

**para**  $m=1, \dots, M$  **faça**

1. Selecionar aleatoriamente  $\eta \times N$  amostras do conjunto (etapa de sub-amostragem)
2. Ajustar a árvore de classificação com profundidade  $k$  utilizando o novo espaço sub-amostrado. Em cada nó são selecionados  $d$  atributos sem reposição.
3. Podar as árvores para minimizar o erro de generalização

**fim**

**retorna**  $L_E = \arg \max_{\gamma} \sum_{t=1}^T 1(y = h(x))$

**Algoritmo 1:** Random Forest (RF), em que  $M$  é o número de árvores de classificação,  $d$  o número de atributos,  $K$  é a profundidade de cada árvore e  $0 < \eta \leq 1$  é a fração de amostras de treinamento (sem reposição).

### 3.5.2 Gradient Tree Boosting – GB

Dado um conjunto de treino  $D = \{(x_i, y_i)\}_{i=1}^N$  com  $N$  amostras, busca-se encontrar a função  $f(x)$  que minimize o valor esperado da função de perda (FRIEDMAN, 2001)

$$L(y, \hat{f}(x)) = \sum_i^N [y_i - \hat{f}(x_i)]^2. \quad (3.15)$$

O método GB aproxima  $f$  por uma expansão aditiva do tipo

$$\hat{f} = \sum_{m=1}^M \beta_m h(x, a_m) \quad (3.16)$$

em que as funções  $h(x, a_m)$  é um  $K$ -nó da árvore de classificação e os parâmetros  $\beta, a$  são ajustados concomitantemente ao conjunto de treino em um processo sequencial. Em cada iteração  $m$ , a árvore de classificação particiona o espaço de variáveis em regiões distintas  $R_{km}^K$  na  $m$ -ésima iteração. A constante  $\gamma_{jm}$  é associado para cada uma região e a regra preditiva descrita por  $x \in R_{jm} \Rightarrow f(x) = \alpha_{jm}$ . Usando a notação indicadora, a saída de  $h$  para o atributo de entrada  $x$  pode ser escrito como

$$h(x, R_{km}^K) = \sum_{k=1}^K \gamma_{km} I(x \in R_{km}), \quad I(\cdot) = \begin{cases} 1 & \text{se } x \in R_{km} \\ 0 & \text{se } x \notin R_{km} \end{cases} \quad (3.17)$$

com parâmetros  $R_{km}, \gamma_{km}$ ,  $k = 1, 2, \dots, J$ ,  $m = 1, \dots, M$ , em que  $\gamma_{km}$  é o valor previsto para a região  $R_{km}$ .

Como a Eq. 3.17 prediz um valor constante para cada região  $R_{km}$ , a solução se reduz a

$$\gamma_{km} = \arg \min_{\gamma} \sum_{x_i \in R_{km}} L(y_i, f_{m-1}(x_i) + \gamma), \quad \gamma \text{ constante.} \quad (3.18)$$

A aproximação atual de  $f_{m-1}(x)$  é atualizada a partir da seguinte regra

$$\hat{f}_m(x) = \hat{f}_{m-1}(x) + \lambda \sum_{k=1}^K \gamma_{km} I(x \in R_{km}) \quad (3.19)$$

em que o parâmetro  $0 < \lambda \leq 1$  é chamado de taxa de aprendizagem. O pseudocódigo com o processo de execução do método é dado pelo Algoritmo 2.

**Entrada:** Conjunto de treino  $D$  com  $N$  amostras em que  $y_i = f(x_i)$ , uma função de perda diferenciável  $L(y, f(x))$ , profundidade de cada árvore  $K$ , e taxa de subamostragem  $\eta$ , taxa de aprendizado  $\lambda$  e número de iterações  $M$

**Saída:** Aproximação  $\hat{f}(x)$  para  $f(x)$

Inicializar  $\hat{f}_0(x) = \arg \min_{\gamma} \sum_{i=1}^N L(y_i, \gamma)$

**para**  $m=1, \dots, M$  **faça**

1. Computar os pseudo-resíduos como o gradiente negativo da resposta

$$r_{im} = - \left[ \frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right]_{f(x)=\hat{f}_{m-1}(x)}, \text{ para } i = 1, 2, \dots, N$$

2. Selecionar aleatoriamente  $\eta \times N$  amostras do conjunto (etapa de sub-amostragem)

3. Ajustar a árvore de classificação com profundidade  $K$  para os pseudo-resíduos  $r_{im}$  usando as  $\eta \times N$  amostras. Cada árvore é ajustada utilizando apenas a sub-amostra aleatória

4. Computar os parâmetros ótimos  $\gamma_{1m}, \dots, \gamma_{Km}$  como

$$\gamma_{km} = \arg \min_{\gamma} \sum_{x_i \in R_{km}} L(y_i, \hat{f}_{m-1}(x_i) + \gamma), \text{ para } k = 1, 2, \dots, K \text{ em que } R_{km} \text{ é o conjunto de } x \text{ que define o nó } k$$

5. Atualizar  $\hat{f}_m(x) = \hat{f}_{m-1}(x) + \lambda \sum_{k=1}^K \gamma_{km} I(x \in R_{km})$ .

**fim**

**retorna**  $\hat{f}(x) = \hat{f}_M(x)$

**Algoritmo 2:** Gradient Boosting (GB), em que  $M$  é o número de iterações,  $L(y, f(x))$  é a função de perda,  $N$  é o número de amostras no conjunto de treino,  $K$  é a profundidade de cada árvore. O parâmetro  $0 < \lambda \leq 1$  controla a taxa de aprendizado do procedimento de boosting e atua como parâmetro de regularização, enquanto  $0 < \eta \leq 1$  é a fração de amostras de treinamento (sem reposição) usada para computar os parâmetros ótimos do modelo na iteração  $m$ .

### 3.5.3 Under Bagging – UB

Também conhecido por *easy ensemble*, quando aplicado usando o classificador AdaBoost como estimador base (LIU et al., 2009), o *under bagging* é um comitê especial que aplica a técnica de sub-amostragem aleatória a um comitê construído em paralelo (BARANDELA et al., 2003). Cada estimador base é treinado utilizando um lote contendo todos os exemplos da classe minoritária e uma sub-amostra de igual tamanho da classe majoritária, embora possa haver redução também da classe minoritária a fim de construir-se comitês mais diversos.

O modelo provém maior capacidade de generalização, visto que uma maior parcela da classe majoritária é considerada pelo sistema comparado a uma única árvore construída nas mesmas circunstâncias. No entanto, é provável que algumas instâncias negativas importantes não sejam levadas em consideração pelo comitê. O pseudocódigo com o processo de execução do método é dado pelo Algoritmo 3.

**Entrada:** Conjunto de treino  $D$  com  $N$  amostras, número de estimadores base  $M$ , algoritmos base de classificação  $L_t$  ( $t=1, \dots, T$ )

**Saída:** Função de classificação do comitê

**para**  $m=1, \dots, M$  **faça**

1. Selecionar aleatoriamente uma amostra  $D_t$  a partir de uma subamostra de  $D$ , obtida a partir do método RUS
2. Ajustar árvore de classificação com profundidade  $K$  da o conjunto de treino  $D_t$  de forma que  $h_t = L_t(D_t)$

**fim**

**retorna**  $H(x) = \arg \max_{\gamma} \sum_{t=1}^T 1(y = h(x))$

**Algoritmo 3:** Under Bagging (UB), em que  $D$  é o conjunto de treino,  $D_t$  a amostra balanceada de treino,  $M$  o número de estimadores base,  $L_t$  a função dos classificadores base e  $1(a)$  é igual a 1 se  $a$  é verdadeiro e 0, caso contrário.

Para todos os comitês, é definido um valor padrão de 10 classificadores-base, cabendo a ressalva de que os resultados deste tipo de modelo tende a ser melhor, quanto maior seja o número de estimadores – e consequentemente, maior o tempo de treinamento. Os demais parâmetros e amplitude de busca adotados para os comitês de aprendizagem são exibidos na Tabela 5, na qual destacam-se em negrito aqueles que apresentaram melhor desempenho nos estudos de caso.

## 3.6 SOFTWARE E HARDWARE

As rotinas foram implementadas na linguagem de programação Python 3.6.1, utilizando as bibliotecas *Scikit-learn* 0.19.0 (PEDREGOSA et al., 2011) e *Imbalanced-learn* 0.2.1 (LEMAITRE et al., 2017) de aprendizado de máquina. A implementação do modelo Under Bagging é

Tabela 5 – Parâmetros adotados para os comitês. A coluna 1 apresenta a sigla do método, a coluna 2 o nome do parâmetro, e a coluna 3 os valores adotados para o parâmetro.

Método	Parâmetros	Valor
RF	Critério	Entropia
	Máx. de atributos – $d$	$\sqrt{m}$
	Mín. amostra folha	5
	Número de estimadores – $M$	10
	Taxa de subamostragem – $\eta$	1,0
GB	Critério	Friedman MSE
	Função de perda	Desvio
	Mín. amostra folha	5
	Número de estimadores – $M$	10
	Taxa de aprendizado – $\lambda$	0,1
	Taxa de subamostragem – $\eta$	1,0
UB	Critério	Entropia
	Mín. amostra folha	5
	Número de estimadores – $M$	10
	Proporção	1:1

fornecida por G. Louppe e T. Fawcett disponível no repositório do Silicon Valley Data Science<sup>3</sup>. Todos os códigos e dados estão disponíveis online no repositório do GitHub<sup>4</sup>.

As especificações de hardware utilizadas para executar a implementação do algoritmo e gerar os modelos de aprendizado são dadas a seguir: CPU AMD Opteron Processor 6272 (64 núcleos de 2.1GHz e cachê de memória de 2MB), RAM de 250 GB e sistema operacional Linux Ubuntu 14.04.4 LTS. Uma vez construído, cada modelo otimizado realiza a predição rapidamente, permitindo a análise e teste de parâmetros no ciclos de projeto.

<sup>3</sup> <https://svds.com/>

<sup>4</sup> [https://github.com/rodrigasantisl/backorder\\_prediction/](https://github.com/rodrigasantisl/backorder_prediction/)

## 4 DESENVOLVIMENTO DA PROPOSTA

No presente estudo, aplica-se o fluxograma representado na Figura 9, no qual os dois conjuntos de dados são avaliados a partir de quatro processo principais. São eles: 1) Definição de atributos; 2) Pré-processamento; 3) Seleção de modelos; e 4) Avaliação de resultados.

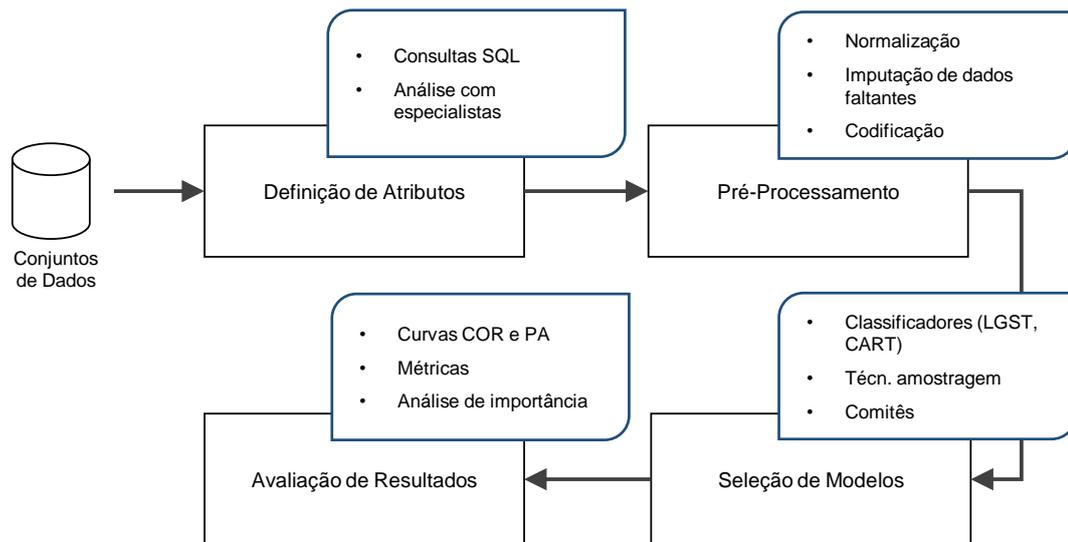


Figura 9 – Fluxograma do desenvolvimento da proposta do presente trabalho: 1) Definição de atributos de entrada, 2) Pré-processamento a partir de técnicas de normalização; imputação de dados faltantes e codificação; 3) Seleção de modelos a partir do ajuste fino de parâmetros; e 4) Avaliação de resultados a partir das curvas COR, PA e métricas de área sob a curva.

### 4.1 DEFINIÇÃO DE ATRIBUTOS

A etapa de definição de atributos consiste na obtenção de características (entradas) que sejam capazes de descrever a variável a qual se deseja prever (saída). Esta etapa é fundamental para o sucesso de qualquer modelo preditivo, visto que o desempenho de todo modelo, do mais simples ao mais complexo, é diretamente relacionado à qualidade dos atributos de entrada. Além disso, no projeto de um modelo preditivo, é necessária a aplicação de atributos que permitam a legibilidade das soluções obtidas. O maior número de atributos utilizados também interfere no tempo de treinamento dos modelos, sendo que um número muito alto de características de entrada pode prejudicar tanto o desempenho computacional como a acurácia do modelo desenvolvido.

No estudo de caso que utiliza o conjunto de dados do Kaggle, os atributos foram definidos pela própria empresa e seus especialistas internos, divulgando-se apenas o nome dos

atributos e a descrição destes. No segundo estudo de caso MRS, o problema da falta de materiais foi discutido com especialistas da empresa, que elencaram quais atributos além dos já aplicados poderiam descrever o problema de falta de materiais, a partir de suas experiências prévias com este. Um conjunto de atributos relacionados ao nível de estoque, vendas e previsão de consumo, informações de cadastro do item, foram definidos para dar início a construção do conjunto de dados de treino. Posteriormente, diversas consultas foram realizadas no banco de dados da empresa utilizando SQL (*Structured Query Language*). Os Dados são por fim tratados utilizando editores de planilhas eletrônicas e scripts em Python, criando um arquivo de texto com os atributos e amostras do conjunto.

Uma lista com todos os atributos aplicados neste estudo é exibida na Tabela 6. Verifica-se que alguns atributos não foram possíveis ser obtidos no conjunto de dados MRS, tais como desempenho de fornecedor e algumas classificações de risco, pois a empresa não apresenta tais informações em seus processos internos de controle. Neste caso, foram utilizados alguns outros atributos relacionados, como o preço registrado, atraso de pedidos em trânsito e a criticidade do item – definida pelos executivos da empresa no processo interno de cadastro de materiais.

Tabela 6 – Lista de atributos utilizados em cada uma das aplicações, e técnicas de pré-processamento aplicadas a cada.

Atributos	Descrição	Aplicação		Pré-processamento		
		MRS	Kaggle	Norm.	Imput.	Codif.
$x_1$	Nível de inventário atual do material	○	○	○		
$x_2$	Tempo de entrega registrado	○	○		○	
$x_3$	Quantidade em trânsito	○	○	○	○	○
$x_4$	Atraso de pedidos em trânsito	○		○	○	○
$x_5$	Previsão de consumo para os próximos 3 meses	○	○	○		
$x_6$	Previsão de consumo para os próximos 6 meses	○	○	○		
$x_7$	Previsão de consumo para os próximos 9 meses		○	○		
$x_8$	Previsão de consumo para os próximos 12 meses	○		○		
$x_9$	Unidades vendidas no último mês	○	○	○		
$x_{10}$	Unidades vendidas nos últimos 3 meses	○	○	○		
$x_{11}$	Unidades vendidas nos últimos 6 meses	○	○	○		
$x_{12}$	Unidades vendidas nos últimos 9 meses		○	○		
$x_{13}$	Unidades vendidas nos últimos 12 meses	○		○		
$x_{14}$	Quantidade em estoque mínima recomendada	○	○	○		
$x_{15}$	Preço registrado	○				
$x_{16}$	Desempenho do fornecedor nos últimos 6 meses		○		○	
$x_{17}$	Desempenho do fornecedor nos últimos 12 meses		○		○	
$x_{18}$	Criticidade do item no estoque	○				○
$x_{19}$	Criticidade do item para manutenção	○				○
$x_{20}$	Criticidade do item para compra	○				○
$x_{21}$	Potencial de problemas		○			○
$x_{22}$	Peças em atraso		○	○		
$x_{23}$	Pedidos não atendidos em aberto		○	○		
$x_{24}$	Risco de convés		○			○
$x_{25}$	Restrição operacional		○			○
$x_{26}$	Processo de aprovação parcial de produção		○			○
$x_{27}$	Material suprido por compra automática		○			○
$x_{28}$	Estoque material por revisão contínua		○			○
$y$	Material em falta nos próximos 2 meses	○	○			

## 4.2 PRÉ-PROCESSAMENTO

Nesta etapa, são realizadas análises exploratórias dos conjuntos de dados a respeito dos atributos levantados. Diversos testes foram realizados junto a um método preditivo, de forma a compreender dentre as possíveis transformações dos atributos de entradas, quais contribuem com o aumento do desempenho dos modelos em geral.

O método escolhido para testar os resultados foi o LGST, devido o baixo custo computacional do método e sua capacidade de considerar todos os atributos na construção – desta forma qualquer transformação realizada nos atributos de entrada é refletida no desempenho do método, mesmo que a importância destes seja relativamente menor. Outros modelos como as árvores de decisão, que realizam a seleção de atributos, tornam-se mais insensíveis à transformação de atributos menos importantes quando comparado aos de maior contribuição na variação do critério de avaliação de separação.

Para atributos relacionados com a quantidade do item tais como o nível de inventário, quantidade em trânsito, previsão e histórico de vendas e consumo, quantidade em estoque mínima recomendada e peças em atraso são aplicados a normalização L2. Para o atributo de desempenho do fornecedor realiza-se a imputação de dados faltantes a partir da média do grupo. Para as variáveis categóricas de cadastro do material, são utilizadas as transformações binárias e uma-de- $K$ , com exceção do nível de criticidade de compras, manutenção e estoque, que por apresentar relação de proximidade (Ex: A para muito, B médio e C pouco crítico), aplica-se a codificação de rótulo. A Tabela 6 apresenta um resumo de quais técnicas foram aplicadas em cada um dos atributos disponíveis.

## 4.3 SELEÇÃO DE MODELOS

Os métodos base utilizados neste trabalho, descritos no Capítulo 3, são sintetizados na Tabela 7, que apresenta a abreviação adotada, o nome, referência, e a descrição sucinta destes. São sete os métodos aqui apresentados, e estes foram escolhidos de forma a se obter um conjunto diverso, dentre os mais populares de cada tipo de abordagem.

Os conjuntos de treino e teste foram separados utilizando uma proporção de 0,85:0,15, gerando aleatoriamente 30 diferentes instâncias para cada problema, de forma a preservar a razão entre as classes. Adotou-se também um esquema de validação cruzada em  $K$ -dobras (KOHAVI et al., 1995), com  $K = 5$ , a fim de evitar-se sobre-ajuste no treinamento, conforme exemplificado na Figura 10.

Todas as combinações de parâmetros dentro das amplitudes de busca definidas são enumeradas e os modelos afinados através do procedimento de busca exaustiva (CHEUNG; NG, 1995), utilizando a métrica ACOR para avaliação, por se tratar de uma métrica mais informativa e estável para problemas desbalanceados. O treinamento de cada combinação de parâmetros possível foi realizado paralelamente, de forma a aumentar a produtividade do fluxo de trabalho,

Tabela 7 – Resumo dos métodos adotados no presente trabalho, com o nome, sigla, referência e descrição sucinta.

Abrev.	Método	Referência	Descrição sucinta
LGST	Regressão Logística	(HOSMER JR et al., 2013)	Método de regressão estatística aplicando função logística.
CART	Árvore de Decisão	(BREIMAN et al., 1993)	Algoritmo de aprendizado clássico baseado em árvores de classificação e regressão.
RUS	Sub-Amostragem Aleatória	(HE; GARCIA, 2009)	CART aplicado a conjunto de dados previamente transformado usando RUS.
SMT	SMOTE	(CHAWLA et al., 2002)	CART aplicado a conjunto de dados previamente transformado usando SMOTE.
RF	Floresta Randômica	(BREIMAN, 2001)	Comitê <i>bagging</i> aplicado à lotes de treinos e atributos sub-amostrados.
GB	Gradient Boosting	(FRIEDMAN, 1999)	Comitê <i>boosting</i> treinado a partir de uma função de perda diferenciável.
UB	Under Bagging	(BARANDELA et al., 2003)	Comitê <i>bagging</i> em que cada classificador é aplicado em um conjunto previamente transformado usando RUS.

registrando-se o tempo de treino de cada iteração.

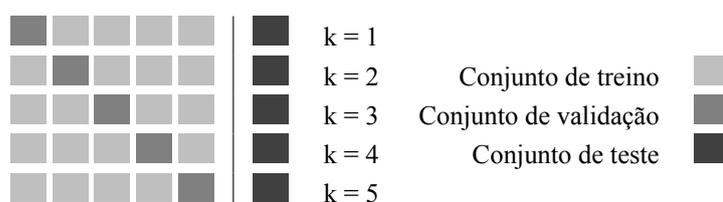


Figura 10 – Divisão do conjunto de treino em  $K = 5$  dobras para validação cruzada. Cada um dos modelos é treinado usando 4 das dobras e sua métrica calculada em uma das dobras. O processo é realizado iterativamente, e a métrica de avaliação é dada pela média obtida nos conjuntos de validação. O conjunto de teste não é alterado, sendo este igual a 15% das amostras totais disponíveis no conjunto de dados.

Os estimadores, resultados e configurações de parâmetros são registrados a cada iteração em uma tabela, salva ao final do treinamento de cada modelo. Diversos testes foram realizados para determinação das amplitudes de busca dos parâmetros de forma experimental. As instâncias de cada modelo treinado foram salvas em arquivos para possibilitar a persistência dos modelos e a avaliação dos resultados.

#### 4.4 AVALIAÇÃO DE RESULTADOS

Os resultados são avaliados a partir do desempenho obtido pelos modelos nos conjuntos de teste de cada instância do problema, sendo esta até então desconhecida pelo classificador. As métricas utilizadas são o ACOR e o ACPA, descritas na Seção 2.2.

As curvas COR e PA são também informativas e analisadas no projeto do sistema, trazendo informações do comportamento dos diversos modelos a partir da variação do valor limite da função de decisão. Isto porque nem sempre os modelos com melhores valores para área sob a curva irão dominar os demais modelos em todos os pontos do gráfico.

Por fim, a importância de cada atributo é obtida da diminuição da impureza (neste estudo, aplica-se o critério de entropia) a partir de cada nó associado ao atributo (BREIMAN et al., 1993). No caso de comitês, aplica-se a média da importância dentre todos classificadores-base (BREIMAN, 2001; FRIEDMAN, 2001).

## 5 RESULTADOS E DISCUSSÃO

### 5.1 ESTUDO DE CASO I: KAGGLE

O primeiro estudo de caso apresentado foi proposto a partir do conjunto de dados obtido na plataforma de aprendizado de máquina Kaggle. O conjunto de dados contém uma grande variedade de itens, informações como tempo de entrega, previsão e vendas passadas, conforme descrito anteriormente. A função de probabilidade e histogramas dos atributos do conjunto de dados após a etapa de pré-processamento é exibida na Figura 11, cujas estatísticas descritivas encontram-se disponíveis na Tabela 18 no Anexo A - Estatísticas Descritivas.

O atributo  $x_1$ , relacionado ao nível atual de estoque, apresenta um acúmulo de distribuição próximo ao valor 1, o que indica capacidade de atendimento do inventário às demandas previstas e históricos de consumo para o dado item. Os itens com valores próximos ou inferiores a 0 indicam situações de alto risco de falta ou quantidade requisitada maior do que disponível no inventário.

O atributo  $x_2$  está relacionado com o tempo de entrega do item, que se distribui entre 0 e 52 dias, sendo que a maior parte possui tempo de entrega inferior ou igual a 8 dias – 3º quartil igual a 8. Dado que o horizonte de falta de materiais do modelo é de 2 meses (60 dias), isto indica uma boa capacidade de reação da empresa no caso de identificação de falta.

O atributo  $x_3$  representa uma distribuição próxima a 0, indicando que o tamanho do lote de entrega é menos representativo do que a quantidade total estocada. Enquanto o atributo  $x_{14}$  representa a quantidade mínima do estoque, com distribuição relativamente superior aos lotes em trânsito.

Os atributos  $x_5, x_6, x_7$  são relacionados à previsão de consumo para os próximos meses, da mesma forma que os atributos  $x_9, x_{10}, x_{11}, x_{12}$  relacionam-se com o histórico de vendas dos produtos. A distribuição destes atributos em sua maioria encontra-se próximo a 0, o que indica que muitos itens não possuem previsão de consumo para os próximos meses e/ou não possuem consumo recorrente.

Os atributos  $x_{16}$  e  $x_{17}$  de desempenho do fornecedor nos últimos 6 e 12 meses apresentam perfis similares, com distribuição assimétrica à direita. Os demais atributos são binários e relacionados à classificação do item, com concentração maior em uma das classes.

Após a aplicação dos métodos de classificação para uma instância do problema, é possível analisar o comportamento dos métodos preditivos de classificação, a partir da análise das curvas COR de cada modelo ajustado usando os parâmetros de profundidade máxima, conforme ilustrado na Figura 12. Nesta, a área sob a curva (ACOR) de cada método é exibida na legenda, ao lado da sigla do mesmo. O resultado de todos os métodos foram superiores a 0,90, sendo que a pela definição da métrica um resultado entre 0,50 – 0,60 é considerado falho, entre 0,60 – 0,70 é insuficiente, entre 0,70 – 0,80 é tido como um bom teste, e entre 0,90 – 1,00 é considerado exce-

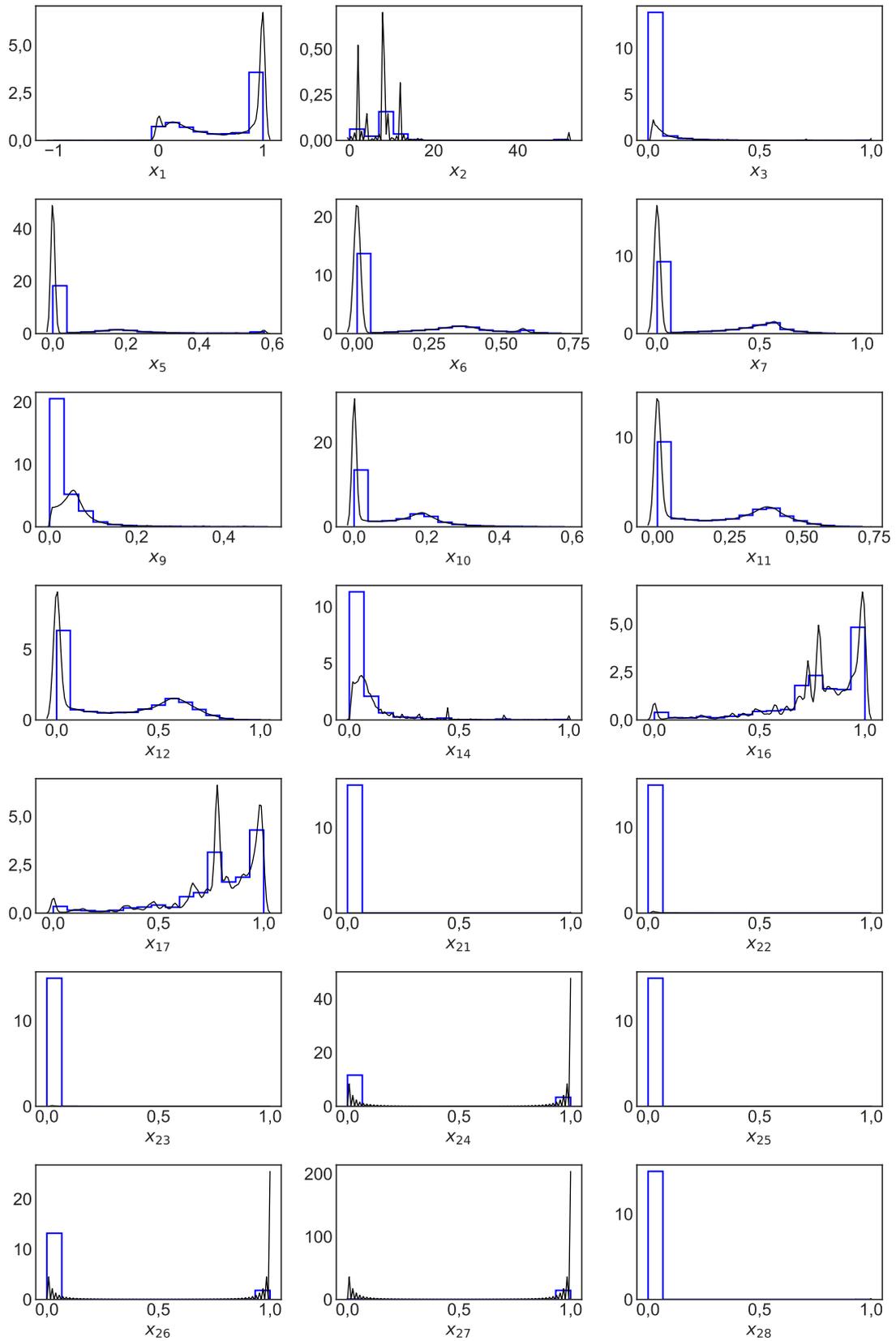


Figura 11 – Histograma de atributos após a aplicação das técnicas de pré-processamento - Kaggle. A função de probabilidade é representada em preto enquanto os histogramas são visualizados como retângulos azuis.

lente (HANLEY; MCNEIL, 1983). O desempenho base de 0,50 para um classificador aleatório e, portanto, sem valor expressivo é também exibido na Figura 12.

O método UB obteve o melhor resultado, indicando uma alta capacidade de distinção entre classes positivas e negativa (0,9518), seguido do GB (0,9494), RUS (0,9338), SMT (0,9316), CART (0,9289), LGST (0,9227) e finalmente, RF (0,9131). No método RF é perceptível uma brusca queda na taxa de  $V_p$  com o aumento da taxa de  $F_p$ , quando este atinge valores superiores a 0,20.

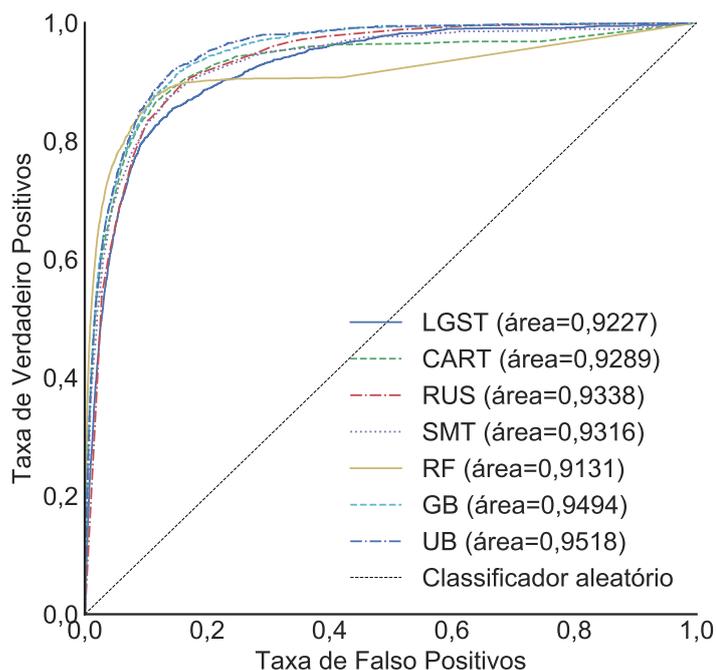


Figura 12 – Curvas COR para os diferentes métodos em uma instância do conjunto de dados Kaggle. O classificador que apresentou maior capacidade de discriminação entre classes positivas e negativas foi o UB, enquanto o menor resultado foi obtido por RF.

Fazendo a análise das curvas PA na Figura 13, é notável a dominância da RF sob os demais métodos por quase todo o limite da função de probabilidade adotada no sistema. O RUS apresentou a curva com menor área, seguida pelo LGST e SMT. Os demais modelos GB, UB e CART apresentaram curvas com perfis bem próximos entre si, com precisão variando entre 0,40 para abrangências baixas e decaimento constante com o aumento da abrangência do sistema.

A partir da análise da melhor curva (RF) é possível verificar possíveis níveis de operação do sistema de detecção: para um sistema com abrangência de 0,10 é possível obter-se uma precisão de 0,65; para abrangência de 0,20, uma precisão de 0,50; para abrangência de 0,40, uma precisão de 0,35; e assim em diante. A escolha do nível de operação do sistema deve ser adotado pela organização de acordo com os custos de falta de materiais e de ação corretiva sobre os itens com risco de falta.

Trinta execuções do problema são geradas e calculadas as métricas ACOR e ACPA obtida em cada um dos conjuntos de teste. Desta forma é possível obter-se o valor médio e o

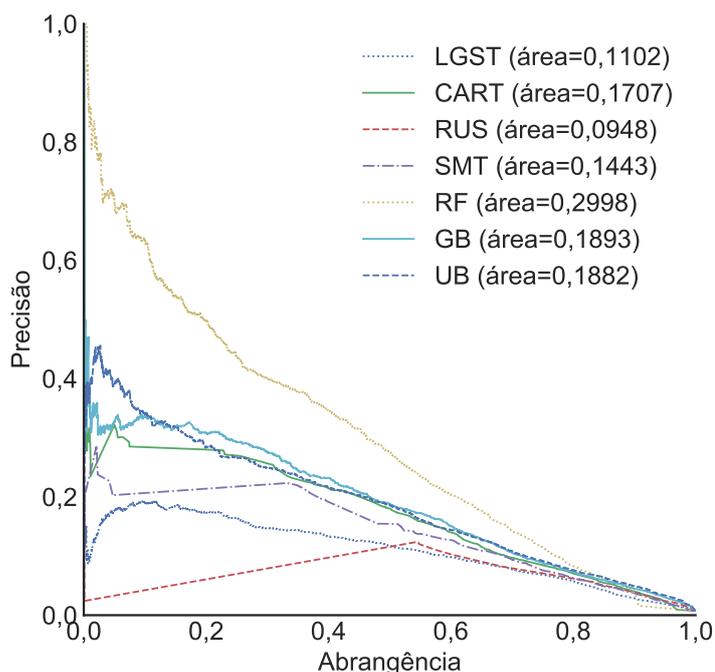


Figura 13 – Curvas PA para uma instância do problema do conjunto de dados Kaggle. O classificador RF apresentou dominância em relação aos demais, enquanto o RUS apresentou o pior resultado.

desvio padrão de cada um dos modelos ajustados, conforme exibido na Tabela 8. Nesta aplicação, o modelo UB apresentou melhor resultado para a métrica ACOR, e valor intermediário para ACPA, enquanto a RF apresentou o melhor resultado para a métrica ACPA e o pior valor para ACOR.

Tabela 8 – Média e desvio-padrão obtidos no conjunto de teste na aplicação Kaggle em N=30 instâncias definidas para o problema.

Método	ACOR	ACPA
LGST	0,9199 ± 0,0023	0,1096 ± 0,0029
CART	0,9262 ± 0,0031	0,1633 ± 0,0049
RUS	0,9341 ± 0,0017	0,1069 ± 0,0059
SMT	0,9305 ± 0,0532	0,1454 ± 0,0079
RF	0,9099 ± 0,0049	0,2981 ± 0,0085
GB	0,9456 ± 0,0017	0,1852 ± 0,0052
UB	0,9497 ± 0,0017	0,1750 ± 0,0061

O procedimento de busca exaustiva é realizado para ajuste dos modelos, e a moda dos parâmetros que obtiveram melhor desempenho no conjunto de validação do procedimento de seleção de modelos é exibida na Tabela 9. Para o modelo LGST, a penalização que adota a normalização L1 e a regularização com valor igual a 1000 obtiveram melhores resultados. No modelo CART, o critério de entropia obteve melhor desempenho como critério de separação do que o ganho de informação. A profundidade máxima, que determina o número de níveis de profundidade e logo, o número de regras e complexidade do modelo, obteve valor igual a 10 na maioria dos modelos, com exceção do modelo RUS, que obteve valor igual a 7, e o RF, que

apresentou modelos com profundidade de até 50 níveis.

Tabela 9 – Moda dos melhores parâmetros obtidos a partir do processo de ajuste de modelos na aplicação Kaggle.

Método	Melhores parâmetros
LGST	Penalização: L1, Regularização ( $C$ ): 1000
CART	Critério: Entropia, Profundidade máx.: 10
RUS	Profundidade máx.: 7
SMT	Profundidade máx.: 10
RF	Profundidade máx.: 50
GB	Profundidade máx.: 10
UB	Profundidade máx.: 10

Na Figura 14 são apresentados os mesmos resultados das métricas da Tabela 8 de forma gráfica – média como ponto central, desvio padrão como linhas nas direções dos eixos das abscissas e coordenadas. Desta forma fica clara a visualização do *trade-off* existente entre as métricas: nem sempre o ACPA aumenta com o aumento do ACOR, e vice-versa. De forma genérica, as técnicas de comitês (RF, GB e UB) dominaram em sua maioria as demais técnicas, obtendo melhores resultados em ambas as métricas. Dependendo da aplicação, deverá ser avaliado qual métrica é mais desejável (curvas PA ou curvas COR) para então definir-se qual dos modelos escolher.

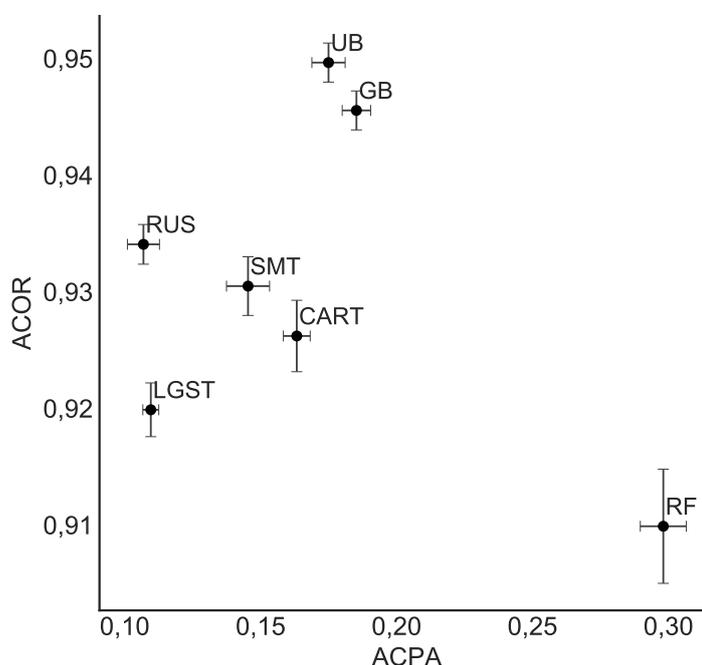


Figura 14 – Representação gráfica em barras de erro dos resultados das métricas ACOR e ACPA na aplicação Kaggle. Os modelos baseados em comitê (RF, GB e UB) dominaram os demais, atingindo melhores níveis em ambas as métricas definidas.

Para realizar a comparação do desempenho dos métodos de acordo com as métricas definidas, são realizados testes estatísticos. O primeiro teste realizado tem como hipótese nula

$H_0$  que as diferentes amostras possuem média  $\mu$  iguais. Para as métricas é utilizado o teste de comparação de médias ANOVA (HEIMAN, 2002). Em ambas as métricas, a hipótese nula é rejeitada conforme exibido na Tabela 10, assim não pode se afirmar que os resultados obtidos pelos diferentes métodos são iguais.

Tabela 10 – Teste ANOVA das métricas para os resultados dos modelos na aplicação Kaggle. Um  $p$ -valor menor do que 0,05 indica a rejeição da hipótese nula que afirma que a média das métricas de todos os métodos são iguais.

Métrica	$F$ -valor	$p$ -valor	Rejeita $H_0$ ?
ACOR	759,8306	0,0000	Sim
ACPA	3270,9309	0,0000	Sim

O teste de Tukey (TUKEY, 1949) é aplicado a seguir para realizar a comparação pareada entre as médias das métricas dos métodos, de forma complementar ao teste ANOVA. Os resultados obtidos para as métricas ACOR e ACPA são exibidos nas Tabelas 11 e 12, respectivamente. A partir da análise dos resultados, é possível verificar que a hipótese nula é rejeitada em todas as comparações dois-a-dois, com exceção da comparação entre os métodos LGST e RUS para a métrica ACPA. Isto indica que indica que o método  $UB > GB > RUS > SMT > CART > LGST > RF$ , na perspectiva da métrica ACOR; e  $RF > GB > UB > CART > SMT > LGST/RUS$ , na perspectiva da métrica ACPA.

Tabela 11 – Teste de Tukey da métrica ACOR para os resultados dos modelos na aplicação Kaggle. Um  $p$ -valor menor do que 0,05 indica a rejeição da hipótese nula de que as métricas possuem média igual.

$G_1$	$G_2$	$\mu_{G_1} - \mu_{G_2}$	Inferior	Superior	Rejeita $H_0$ ?
CART	GB	0,0193	0,0172	0,0215	Sim
CART	LGST	-0,0063	-0,0085	-0,0042	Sim
CART	RF	-0,0163	-0,0184	-0,0142	Sim
CART	RUS	0,0079	0,0057	0,0100	Sim
CART	SMT	0,0043	0,0022	0,0064	Sim
CART	UB	0,0234	0,0213	0,0256	Sim
GB	LGST	-0,0257	-0,0278	-0,0235	Sim
GB	RF	-0,0357	-0,0378	-0,0335	Sim
GB	RUS	-0,0115	-0,0136	-0,0094	Sim
GB	SMT	-0,0151	-0,0172	-0,0129	Sim
GB	UB	0,0041	0,0020	0,0062	Sim
LGST	RF	-0,0100	-0,0121	-0,0079	Sim
LGST	RUS	0,0142	0,0121	0,0163	Sim
LGST	SMT	0,0106	0,0085	0,0127	Sim
LGST	UB	0,0298	0,0276	0,0319	Sim
RF	RUS	0,0242	0,0221	0,0263	Sim
RF	SMT	0,0206	0,0185	0,0227	Sim
RF	UB	0,0398	0,0376	0,0419	Sim
RUS	SMT	-0,0036	-0,0057	-0,0015	Sim
RUS	UB	0,0156	0,0135	0,0177	Sim
SMT	UB	0,0192	0,0170	0,0213	Sim

Tabela 12 – Teste de Tukey da métrica ACPA para os resultados dos modelos na aplicação Kaggle. Um  $p$ -valor menor do que 0,05 indica a rejeição da hipótese nula de que os dados não possuem distribuição normal.

$G_1$	$G_2$	$\mu_{G_1} - \mu_{G_2}$	Inferior	Superior	Rejeita $H_0$ ?
CART	GB	0,0219	0,0172	0,0267	Sim
CART	LGST	-0,0537	-0,0584	-0,0489	Sim
CART	RF	0,1348	0,1301	0,1396	Sim
CART	RUS	-0,0564	-0,0611	-0,0517	Sim
CART	SMT	-0,0179	-0,0227	-0,0132	Sim
CART	UB	0,0117	0,0069	0,0164	Sim
GB	LGST	-0,0756	-0,0804	-0,0709	Sim
GB	RF	0,1129	0,1081	0,1176	Sim
GB	RUS	-0,0783	-0,0831	-0,0736	Sim
GB	SMT	-0,0399	-0,0446	-0,0351	Sim
GB	UB	-0,0103	-0,0150	-0,0055	Sim
LGST	RF	0,1885	0,1838	0,1933	Sim
LGST	RUS	-0,0027	-0,0075	0,0020	Não
LGST	SMT	0,0358	0,0310	0,0405	Sim
LGST	UB	0,0654	0,0606	0,0701	Sim
RF	RUS	-0,1912	-0,1960	-0,1865	Sim
RF	SMT	-0,1527	-0,1575	-0,1480	Sim
RF	UB	-0,1231	-0,1279	-0,1184	Sim
RUS	SMT	0,0385	0,0337	0,0432	Sim
RUS	UB	0,0681	0,0633	0,0728	Sim
SMT	UB	0,0296	0,0249	0,0343	Sim

Por fim, realiza-se a análise de importância dos atributos, calculada a partir da diminuição do ganho de informação das regras dos atributos para cada um dos melhores modelos selecionados nas instâncias do problema, para todos os métodos com exceção do LGST. A Figura 15 apresenta o valor médio da importância de cada atributo representado pelas barras azuis e o intervalo de confiança considerando um coeficiente de 0,95 representado pela linha preta. As informações podem ser também verificadas na Tabela 20 disponível no Anexo B - Importância dos Atributos.

Em todos os métodos, o atributo  $x_1$  que representa o nível de inventário, apresentou a maior importância, chegando ser responsável sozinha por até 0,7514 de toda a importância no modelos RUS. Verifica-se também que os métodos CART, RUS, SMT e UB, apresentam um perfil semelhante, com maior concentração de importância no atributo  $x_1$  em detrimento aos demais. Os métodos RF e GB, por outro lado, foram capazes de considerar outros atributos presentes, apresentando perfis mais distribuídos por todos os outros atributos. Isto é devido à aleatoriedade gerada pelo processo de seleção de amostras e atributos no método RF, e a capacidade do GB em especializar seu treinamento em outros exemplos no qual os modelos-base anterior obtiveram menor desempenho.

Um grupo de atributos que obteve uma parcela representativa de importância foram aqueles relacionados a previsão de consumo para os próximos meses – atributos  $x_5$ ,  $x_6$  e  $x_7$  – que juntos contribuíram com 0,2510 da importância dos modelos RF e 0,1635 da importância

dos modelos GB. De forma análoga, aqueles relacionados ao histórico de vendas – atributos  $x_9$ ,  $x_{10}$ ,  $x_{11}$  e  $x_{12}$  – também apresentaram um peso significativo de 0,2492 nos modelos RF e 0,3010 nos modelos GB. De forma geral, a previsão e histórico de consumo de períodos mais próximos apresentaram importâncias maiores em períodos a curto prazo do que de aquelas de horizontes maiores.

O desempenho do fornecedor – atributos  $x_{16}$  e  $x_{17}$  – apresentaram importância de 0,1015 e 0,1130, para os modelos RF e GB, respectivamente. Isto indica que quanto melhor o desempenho do fornecedor, maior a probabilidade deste honrar com a entrega do item ou realizar um suprimento de emergência em tempo hábil. Finalmente, alguns outros atributos com alguma contribuição relativa foram: o tempo de entrega  $x_2$ , com média de 0,0237 em todos os modelos; a quantidade em trânsito  $x_3$ , com média de 0,0720 em todos os modelos; o estoque mínimo  $x_{14}$ , com média de 0,0240 em todos os modelos.

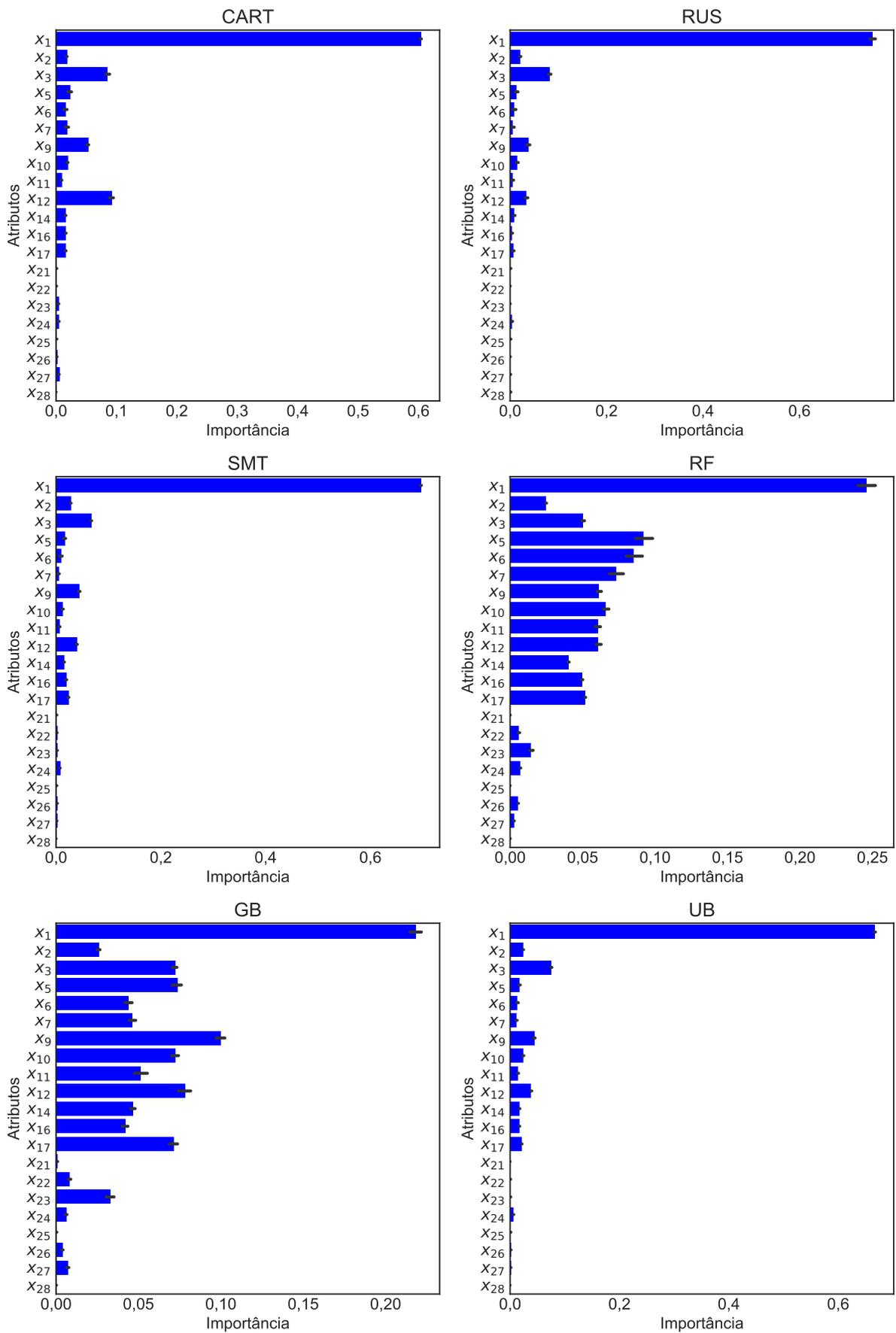


Figura 15 – Análise de importância de atributos de cada método na aplicação Kaggle.

## 5.2 ESTUDO DE CASO II: MRS LOGÍSTICA

O segundo estudo de caso apresentado, foi desenvolvido em parceria a uma operadora ferroviária nacional, que possui cerca de 20 centros de armazenagem responsáveis pelo atendimento das operações de manutenção e suporte de toda sua malha ferroviária, que cobre os estados de Minas Gerais, Rio de Janeiro e São Paulo. Os atributos selecionados e processados podem ser visualizados na Figura 16, e as estatísticas descritivas dos mesmos estão disponíveis na Tabela 19 do Anexo A - Estatísticas Descritivas.

De maneira análoga à aplicação anterior, verifica-se que o atributo  $x_1$  relacionado ao nível atual de estoque apresenta acúmulo de distribuição próximo a 1, o que indica a capacidade de atendimento à demandas futuras, enquanto os itens com estoque próximo a 0 apresentam alto risco de falta.

O atributo  $x_2$ , relacionado ao tempo de entrega do item, se distribui entre 10 e 320 dias, sendo o tempo médio de entrega de 63,55 dias. Comparado à aplicação anterior, isto indica uma menor capacidade de resposta da empresa à situações de risco de suprimentos. O atributo  $x_3$ , relacionado a quantidade em trânsito apresenta um número relativamente baixo das entregas, em relação ao total estocado na companhia; enquanto o atributo  $x_4$ , que indica o atraso médio dos pedidos em trânsito, mostra que a maioria dos pedidos não atrasam, sendo o 1º, 2º e 3º quartis iguais a 0.

Os atributos  $x_5$ ,  $x_6$  e  $x_8$ , relacionados à previsão de consumo para os próximos meses, apresenta quantidades relativas inferiores ao nível de estoque, assim como os atributos  $x_9$ ,  $x_{10}$ ,  $x_{11}$  e  $x_{13}$  de consumo nos últimos meses. Verifica-se que a mediana de todas estas variáveis são iguais a 0, o que mostra que mais de 50% do estoque não possui previsão para os próximos meses ou consumo nos últimos meses.

O atributo  $x_{14}$  relacionado ao estoque mínimo do item no inventário apresenta perfil bem parecido a aplicação anterior, com distribuição concentrada próxima a 0. O atributo  $x_{15}$  descreve a distribuição do preço unitário dos itens, enquanto as variáveis  $x_{18}$ ,  $x_{19}$  e  $x_{20}$ , representam a classificação do item quanto a criticidade para estoque, manutenção e compra, respectivamente.

A partir da análise dos atributos deste conjunto de dados em relação ao da aplicação anterior, é possível perceber que o sistema de inventário da organização dispõem de maiores níveis de estoque, justificado pelo maior tempo de entrega/fornecimento, pouca fluidez do consumo e, conseqüentemente, menos previsibilidade da demanda. Todos estes fatores contribuem com o aumento da complexidade desta segunda aplicação, junto ao menor número de amostras da classe minoritária neste conjunto de dados.

Os métodos são aplicados para uma instância do problema e construídas a curvas COR dos modelos ajustados, conforme representado na Figura 17. A área sob a curva é apresentada entre parênteses ao lado da sigla do método, pelo o método UB obteve melhor resultado (0,8512), seguido de LGST (0,8472), RF (0,8406), CART (0,7964), RUS (0,7917), SMT (0,7826) e GB

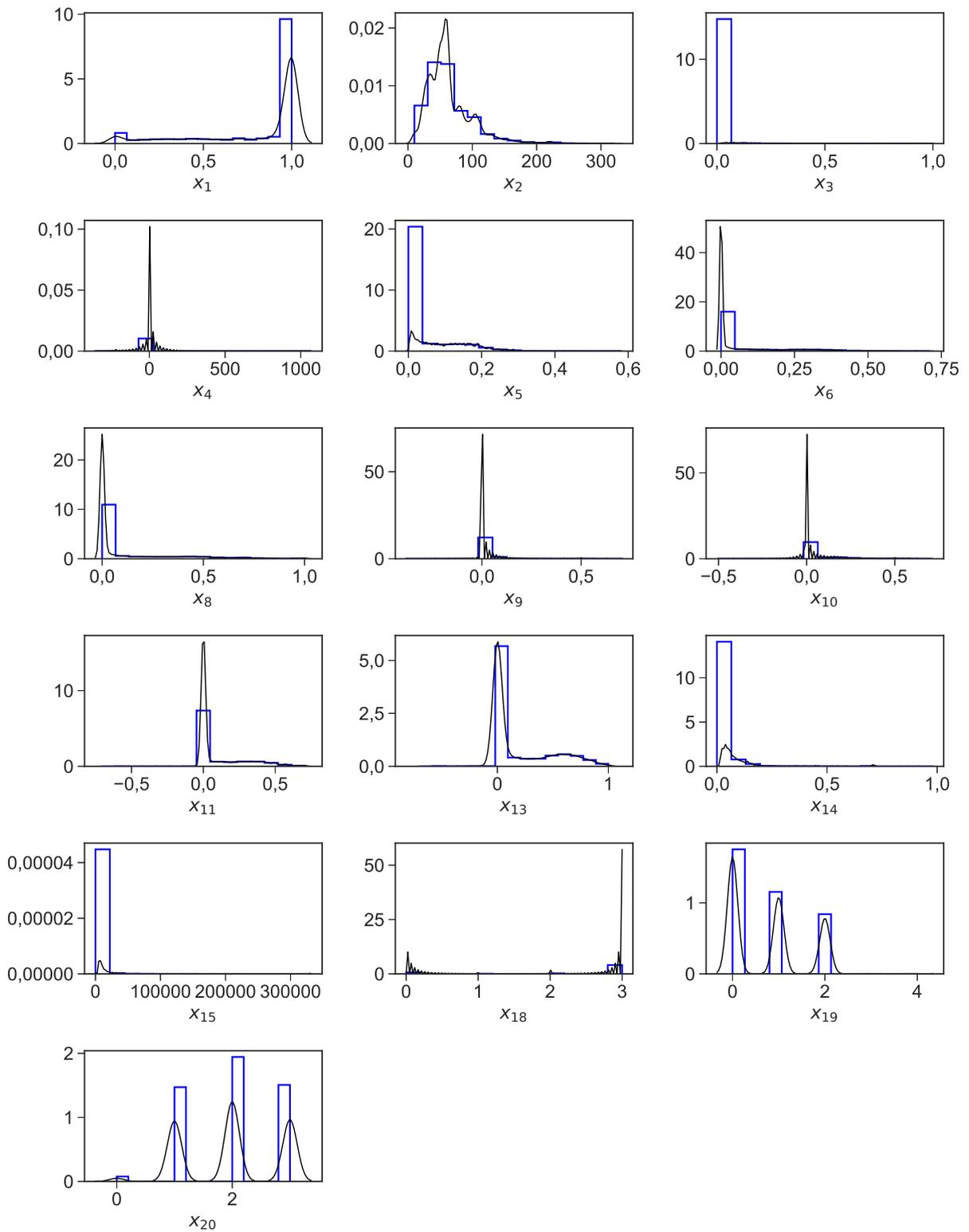


Figura 16 – Histograma de atributos após a aplicação das técnicas de pré-processamento - MRS. A função de probabilidade é representada em preto enquanto os histogramas são visualizados como retângulos azuis.

(0,7283). A maioria dos métodos apresentaram uma diminuição na taxa de  $T_p$  para  $F_p$  maiores

do que 0,30.

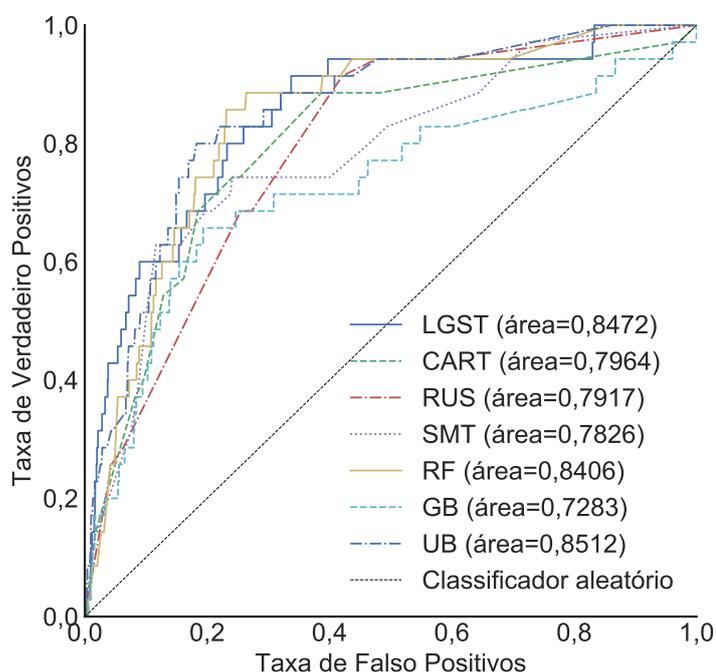


Figura 17 – Curvas COR para os diferentes métodos em uma instância do conjunto de dados MRS. O classificador que apresentou maior capacidade de discriminação entre classes positivas e negativas foi o UB, enquanto o menor resultado foi obtido por GB.

As curvas PA exibidas na Figura 18 demonstram o *trade-off* enfrentado para a seleção do valor limítrofe da função de decisão. Os métodos de amostragem apresentaram melhores resultados dentre os intervalos de abrangência entre 0,00 – 0,25; para valores superiores a 0,25, LGST apresenta resultados operacionais mais interessantes, e a partir de 0,60, todos os métodos apresentam curvas bem próximas, sendo difícil distinguir um método preferível.

A partir da análise das melhores curvas (SMT e LGST), é possível enumerar possíveis níveis de operação do sistema de detecção: para um sistema com abrangência de 0,10, é possível obter-se precisão de 0,70; para abrangência de 0,20, precisão de 0,30; para abrangência de 0,40, precisão de 0,15; e assim em diante. Novamente, a escolha do nível de operação do sistema deve ser adotada de acordo com os interesse e custos de ação da organização.

Após execução dos métodos em 30 instâncias diferentes e cômputo das métricas ACOR e ACPA nos conjuntos de teste, são calculadas as médias e desvio padrão das métricas conforme exibido na Tabela 13. Não é possível afirmar qual modelo obteve melhor desempenho, visto que o desvio padrão de cada métrica foi substancial quando comparado à distância entre as métricas.

A moda do conjunto de parâmetros dos melhores modelos obtidos através do procedimento de busca exaustiva são exibidos na Tabela 14. Para o modelo de LGST a penalização L1 e o coeficiente de regularização igual a 1000 obtiveram melhores resultados, enquanto no modelo CART, o critério de separação entropia obteve melhores resultados, assim como na aplicação anterior. A profundidade máxima das árvores de decisão base dos modelos foram menores do

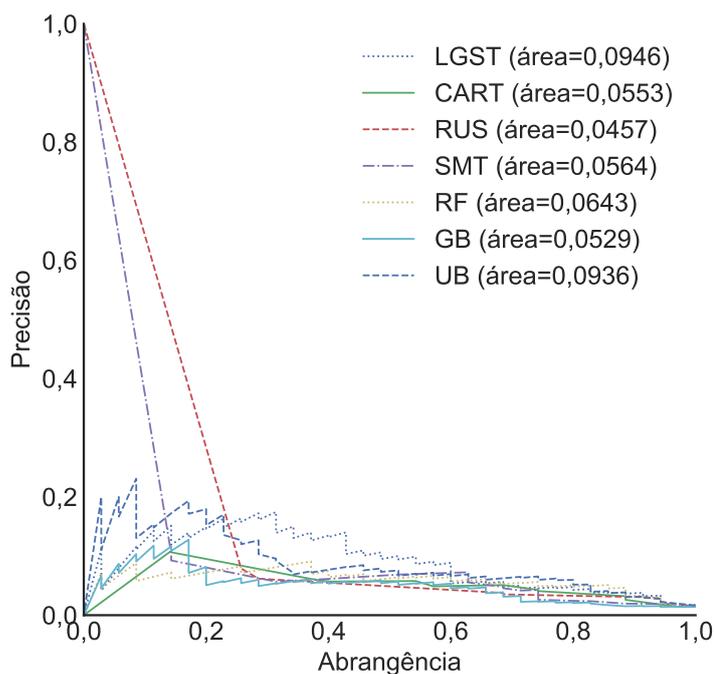


Figura 18 – Curvas PA para uma instância do conjunto de dados MRS. O método RUS apresentou melhor curva para abrangência  $< 0,25$ , enquanto o método LGST apresentou melhores resultados para valores maiores.

Tabela 13 – Média desvio-padrão obtidos no conjunto de teste na aplicação MRS em  $N=30$  instâncias definidas para o problema.

Método	ACOR	ACPA
LGST	$0,8364 \pm 0,0342$	$0,0967 \pm 0,0224$
CART	$0,7794 \pm 0,0463$	$0,0595 \pm 0,0121$
RUS	$0,7840 \pm 0,0372$	$0,0445 \pm 0,0095$
SMT	$0,7893 \pm 0,0532$	$0,0533 \pm 0,0126$
RF	$0,8287 \pm 0,0451$	$0,0819 \pm 0,0198$
GB	$0,8310 \pm 0,0313$	$0,0825 \pm 0,0189$
UB	$0,8321 \pm 0,0294$	$0,0835 \pm 0,0260$

que a aplicação anterior, entre 3 e 5, o que mostra que os modelos gerados na aplicação Kaggle tiveram maior complexidade quando compara a esta aplicação.

Tabela 14 – Moda dos melhores parâmetros obtidos a partir do processo de ajuste de modelos na aplicação MRS.

Método	Melhores parâmetros
LGST	Penalização: L1, Regularização ( $C$ ): 1000
CART	Critério: Entropia, Profundidade máx.: 4
RUS	Profundidade máx.: 3
SMT	Profundidade máx.: 4
RF	Profundidade máx.: 5
GB	Profundidade máx.: 3
UB	Profundidade máx.: 4

As mesmas informações da Tabela 13 são apresentadas de forma gráfica na Figura 5.2,

no qual a média de cada métrica é dada pelo ponto no gráfico, e o intervalo de confiança de um desvio-padrão pelas linhas associadas. Nesta forma de visualização, é possível observar que o modelo LGST domina as demais soluções em ambas as métricas. É possível perceber também a formação de 3 diferentes grupos, de acordo com a similaridade dos resultados: o primeiro grupo com os modelos baseado em amostragem e o CART, o segundo grupo com os modelos baseado em comitês, e o terceiro grupo com o modelo LGST. No entanto, dado o considerável desvio padrão do conjunto de soluções, não é possível afirmar com exatidão se um dado modelo é melhor ou pior do que outro.

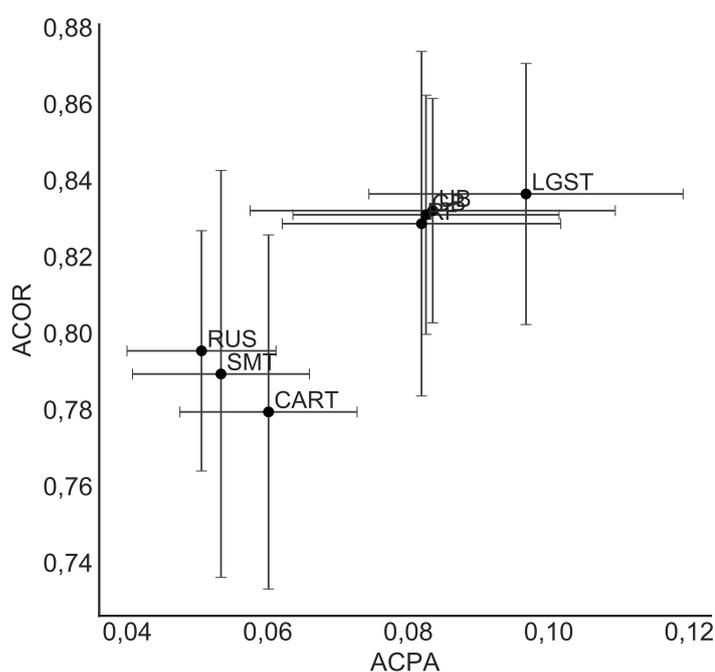


Figura 19 – Representação gráfica em barras de erro dos resultados das métricas ACOR e ACPA na aplicação MRS. O método LGST parece obter melhores resultados que os demais, enquanto dois grupos muito próximos são formados: o primeiro grupo com os comitês GB, RF e UB, e o segundo com os métodos CART, RUS e SMT.

Assim, primeiramente aplica-se o teste ANOVA para verificar se a média das métricas obtidas por todos os métodos são iguais. Como a hipótese nula é rejeitada conforme mostrado na Tabela 15, não é possível afirmar que os resultados obtidos pelos diferentes métodos são iguais.

Tabela 15 – Teste ANOVA das métricas para os resultados dos modelos na aplicação MRS. Um  $p$ -valor menor do que 0,05 indica a rejeição da hipótese nula que afirma que a média das métricas de todos os métodos são iguais.

Métrica	$F$ -valor	$p$ -valor	Rejeita $H_0$ ?
ACOR	11,0645	0,0000	Sim
ACPA	28,1979	0,0000	Sim

O teste de Tukey é aplicado a seguir a fim de realizar a comparação pareada entre os métodos, adicionalmente ao primeiro teste. Os resultados obtidos para as métricas ACOR e

ACPA são exibidos na Tabela 16 e 17, respectivamente. É possível concluir a partir da análise dos resultados que os métodos CART, RUS e SMT apresentam métricas de igual valor estatístico, enquanto GB, LGST, RF e UB apresentam igual valor. A única afirmação que pode ser feita é de que este segundo grupo possui melhor desempenho no dado problema do que o primeiro grupo.

Tabela 16 – Teste de Tukey da métrica ACOR para os resultados dos modelos na aplicação MRS. Um  $p$ -valor menor do que 0,05 indica a rejeição da hipótese nula de que os dados não possuem distribuição normal.

$G_1$	$G_2$	$\mu_{G_1} - \mu_{G_2}$	Inferior	Superior	Rejeita $H_0$ ?
CART	GB	0,0516	0,0211	0,0821	Sim
CART	LGST	0,0570	0,0265	0,0875	Sim
CART	RF	0,0492	0,0188	0,0797	Sim
CART	RUS	0,0160	-0,0145	0,0465	Não
CART	SMT	0,0099	-0,0206	0,0404	Não
CART	UB	0,0527	0,0222	0,0832	Sim
GB	LGST	0,0055	-0,0250	0,0359	Não
GB	RF	-0,0023	-0,0328	0,0282	Não
GB	RUS	-0,0356	-0,0661	-0,0051	Sim
GB	SMT	-0,0417	-0,0721	-0,0112	Sim
GB	UB	0,0011	-0,0294	0,0316	Não
LGST	RF	-0,0078	-0,0383	0,0227	Não
LGST	RUS	-0,0411	-0,0715	-0,0106	Sim
LGST	SMT	-0,0471	-0,0776	-0,0166	Sim
LGST	UB	-0,0044	-0,0349	0,0261	Não
RF	RUS	-0,0333	-0,0638	-0,0028	Sim
RF	SMT	-0,0393	-0,0698	-0,0088	Sim
RF	UB	0,0034	-0,0271	0,0339	Não
RUS	SMT	-0,0061	-0,0365	0,0244	Não
RUS	UB	0,0367	0,0062	0,0672	Sim
SMT	UB	0,0427	0,0123	0,0732	Sim

A Figura 20 apresenta a média e intervalo de confiança de 0,95 da importância dos atributos nos modelos ajustados às instâncias do problema. Os valores podem ser consultados também na Tabela 21 disponível no Anexo B - Importância dos Atributos.

Com exceção do SMT, em todos os modelos o nível de inventário  $x_1$  obteve maior importância, de forma similar à aplicação anterior, chegando a ser responsável por até 0,5649 no modelo CART. O modelo SMT favoreceu no entanto os atributos de classificação do item –  $x_{18}$ ,  $x_{19}$  e  $x_{20}$  – que obtiveram 0,2195 da importância, os atributos relacionados ao histórico de vendas –  $x_9$ ,  $x_{10}$ ,  $x_{11}$  e  $x_{13}$  – com 0,4568 e a previsão de vendas –  $x_5$ ,  $x_6$  e  $x_8$  – com 0,2727 da importância relativa.

Considerando o grupo formado pelos métodos GB, RF e UB, os atributos relacionados ao histórico de vendas –  $x_9$ ,  $x_{10}$ ,  $x_{11}$  e  $x_{13}$  – obtiveram importância igual a 0,2269, enquanto a previsão de vendas –  $x_5$ ,  $x_6$  e  $x_8$  – obteve o valor de 0,3344 da importância total. Assim como a aplicação anterior, estes grupos de atributos se mostram essenciais para o desenvolvimento do sistema de detecção de falta. A importância média do atributo de tempo de entrega  $x_2$  foi de

Tabela 17 – Teste de Tukey da métrica ACPA para os resultados dos modelos na aplicação MRS. Um  $p$ -valor menor do que 0,05 indica a rejeição da hipótese nula de que os dados não possuem distribuição normal.

$G_1$	$G_2$	$\mu_{G_1} - \mu_{G_2}$	Inferior	Superior	Rejeita $H_0$ ?
CART	GB	0,0224	0,0083	0,0365	Sim
CART	LGST	0,0367	0,0225	0,0508	Sim
CART	RF	0,0218	0,0077	0,0359	Sim
CART	RUS	-0,0095	-0,0237	0,0046	Não
CART	SMT	-0,0068	-0,0209	0,0073	Não
CART	UB	0,0234	0,0093	0,0375	Sim
GB	LGST	0,0142	0,0001	0,0284	Sim
GB	RF	-0,0006	-0,0147	0,0135	Não
GB	RUS	-0,0320	-0,0461	-0,0178	Sim
GB	SMT	-0,0292	-0,0433	-0,0151	Sim
GB	UB	0,0010	-0,0132	0,0151	Não
LGST	RF	-0,0149	-0,0290	-0,0008	Sim
LGST	RUS	-0,0462	-0,0603	-0,0321	Sim
LGST	SMT	-0,0434	-0,0575	-0,0293	Sim
LGST	UB	-0,0133	-0,0274	0,0008	Não
RF	RUS	-0,0313	-0,0454	-0,0172	Sim
RF	SMT	-0,0285	-0,0427	-0,0144	Sim
RF	UB	0,0016	-0,0125	0,0157	Não
RUS	SMT	0,0028	-0,0113	0,0169	Não
RUS	UB	0,0329	0,0188	0,0470	Sim
SMT	UB	0,0301	0,0160	0,0443	Sim

0,0244; da quantidade em trânsito  $x_3$  de 0,0123; do atraso médio dos pedidos  $x_4$  de 0,0187; do preço registrado  $x_{15}$  de 0,0347; e da classificação do item –  $x_{18}$ ,  $x_{19}$  e  $x_{20}$  – igual a 0,0619.

De forma análoga à aplicação anterior, os métodos RF e GB foram capazes de considerar outros atributos além daqueles considerados nos demais métodos, pelos motivos já explicados. A previsão e histórico de consumos de períodos mais próximos apresentaram importância maiores do que dos horizontes maiores, assim como na aplicação anterior, o que reforça a maior importância dos meses mais recentes como atributos de entrada.

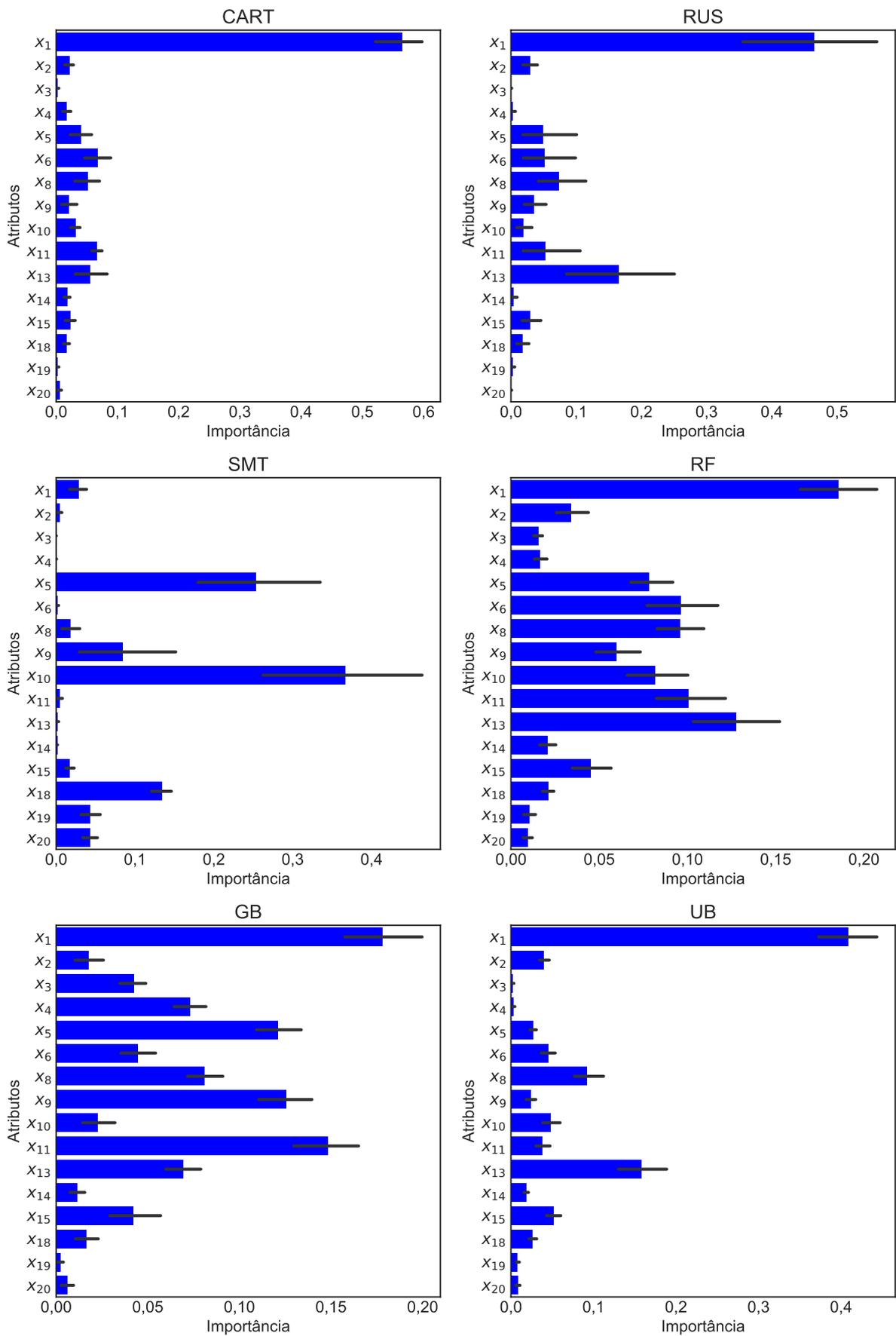


Figura 20 – Análise de importância de atributos de cada método na aplicação MRS.

## 6 CONCLUSÃO

O presente trabalho apresentou os resultados da aplicação de algoritmos de classificações na construção e projeto de um sistema preditivo para controle de inventário, em extensão aos modelos de planejamento de inventário já há muito discutidos na literatura. Em ambos os estudos de caso aqui apresentados, foram obtidos resultados positivos à aplicação de tal mecanismo de controle, capaz de indicar materiais com alto risco de falta no estoque em um período de tempo definido. Tal identificação possibilita a promoção de ações para mitigar tais problemas em tempo hábil.

Os materiais que faltam (classe positiva) são raros quando comparados aos que não (classe negativa), fazendo necessário assim a adoção de métodos e métricas particulares para o projeto, desenvolvimento e avaliação dos modelos neste problema de classes desbalanceadas. As curvas COR e PA apresentam um importante papel na avaliação dos modelos, que podem obter diferentes níveis operacionais de acordo com o limite adotado para a função de decisão.

A respeito dos atributos utilizados, destaca-se em ambos estudos o nível de estoque do item, as quantidades de consumo previstos e histórico, e outras informações transacionais como quantidade em trânsito e em atraso. As informações relacionadas ao fornecedor do material desempenham uma contribuição importante, sendo sua indisponibilidade no segundo estudo de caso pode ser um dos fatores que contribuíram para um resultado menos satisfatório.

As informações gerais do cadastro do itens como criticidade, risco e tempo de entrega tiveram uma contribuição considerável. Outro atributo que poderiam ser adicionado é o erro das previsões (ex.: desvio padrão absoluto), de forma a complementar os atributos de valor previsto que, conforme discutido, possui uma contribuição alta no desempenho dos modelos; assim como outras métricas estatística relacionadas à variância no consumo nos meses passados.

Quanto aos modelos adotados, não foi possível eleger um único modelo que obtivesse melhor resultado em todas as aplicações, ou mesmo em ambas as métricas. No entanto, é possível concluir que a adoção de comitês obtiveram resultados melhores do que a aplicação de classificadores isolados ou combinados com as técnicas de amostragem, em parte por estes serem capazes de atribuir importância a mais atributos do que os modelos mais simples.

Em ambas aplicações, os resultados obtidos pelos melhores modelos foram positivos à adoção da ferramenta para resposta a eventos no inventário, no entanto, aplicações em outros sistemas de inventário de diferentes naturezas e indústrias. Outra aplicação de igual complexidade que pode ser conferida a partir do desdobramento da proposta, é a previsão de pedidos – para casos de inventários em que a maioria do itens são inativos, ou seja, não apresentam demanda estável.

Existe amplo espaço para incorporação de outros atributos não abordados nos estudos de casos apresentados, como frequência de compras, outros indicadores estatísticos de alta ordem

relacionados ao histórico de transações, entre outros. Por fim, o horizonte adotado nas aplicações foi de 8 semanas, havendo espaço para análise da sensibilidade dos resultados obtidos com a variação destes horizontes.

Demais sugestões para trabalhos futuros incluem explorar outros modelos baseados em comitês e técnicas de amostragem não abordadas neste estudo, além de algoritmos de diferente natureza como as máquinas de vetor suporte e redes neurais, para verificar possíveis melhorias de desempenho. A análise de sensibilidade pode ser utilizada em combinação a estes outros atributos, para avaliação da importância dos atributos e comparação com aqueles aqui apresentados. Outro campo vasto não adotado neste trabalho é o aprendizado sensível ao custo, no qual custos de classificação errada são incorporados, possibilitando a análise das curvas de custos no projeto dos modelos.

## REFERÊNCIAS

- ALI, Ö. G.; SAYIN, S.; WOENSEL, T. V.; FRANSOO, J. Sku demand forecasting in the presence of promotions. *Expert Systems with Applications*, Elsevier, v. 36, n. 10, p. 12340–12348, 2009.
- BARANDELA, R.; VALDOVINOS, R. M.; SÁNCHEZ, J. S. New applications of ensembles of classifiers. *Pattern Analysis & Applications*, Springer, v. 6, n. 3, p. 245–256, 2003.
- BELVEDERE, V.; GOODWIN, P. The influence of product involvement and emotion on short-term product demand forecasting. *International Journal of Forecasting*, Elsevier, v. 33, n. 3, p. 652–661, 2017.
- BREIMAN, L. Random forests. *Machine Learning*, Springer, v. 45, n. 1, p. 5–32, 2001.
- BREIMAN, L.; FRIEDMAN, J.; OLSHEN, R.; STONE, C. Classification and regression trees, wadsworth international group, belmont, ca, 1984. *Case Description Feature Subset Correct Missed FA Misclass*, v. 1, p. 1–3, 1993.
- BURKE, E. K.; GENDREAU, M.; HYDE, M.; KENDALL, G.; OCHOA, G.; ÖZCAN, E.; QU, R. Hyper-heuristics: A survey of the state of the art. *Journal of the Operational Research Society*, Springer, v. 64, n. 12, p. 1695–1724, 2013.
- CHAN, P. K.; FAN, W.; PRODRONIDIS, A. L.; STOLFO, S. J. Distributed data mining in credit card fraud detection. *IEEE Intelligent Systems and Their Applications*, IEEE, v. 14, n. 6, p. 67–74, 1999.
- CHAN, T.-H.; JIA, K.; GAO, S.; LU, J.; ZENG, Z.; MA, Y. Pcanet: A simple deep learning baseline for image classification? *IEEE Transactions on Image Processing*, IEEE, v. 24, n. 12, p. 5017–5032, 2015.
- CHAWLA, N. V.; BOWYER, K. W.; HALL, L. O.; KEGELMEYER, W. P. Smote: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, v. 16, p. 321–357, 2002.
- CHEUNG, B. K.; NG, A. C. An efficient and reliable algorithm for nonsmooth nonlinear optimization. *Neural, Parallel & Scientific Computations*, Dynamic Publishers, Inc., v. 3, n. 1, p. 115–128, 1995.
- COST, S.; SALZBERG, S. A weighted nearest neighbor algorithm for learning with symbolic features. *Machine Learning*, Springer, v. 10, n. 1, p. 57–78, 1993.
- DIETTERICH, T. G. Ensemble methods in machine learning. In: SPRINGER. *International workshop on multiple classifier systems*. [S.l.], 2000. p. 1–15.
- DONSELAAR, K. V.; PETERS, J.; JONG, A. D.; BROEKMEULEN, R. Analysis and forecasting of demand during promotions for perishable items. *International Journal of Production Economics*, Elsevier, v. 172, p. 65–75, 2016.
- DRUMMOND, C.; HOLTE, R. C. Exploiting the cost (in) sensitivity of decision tree splitting criteria. In: *International Conference on Machine Learning*. [S.l.: s.n.], 2000. v. 1, n. 1.

- FOLINAS, D.; RABI, S. Estimating benefits of demand sensing for consumer goods organisations. *Journal of Database Marketing & Customer Strategy Management*, Springer, v. 19, n. 4, p. 245–261, 2012.
- FRIEDMAN, J. H. Stochastic gradient boosting. *Computational Statistics & Data Analysis*, Elsevier, v. 38, n. 4, p. 367–378, 1999.
- FRIEDMAN, J. H. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, JSTOR, p. 1189–1232, 2001.
- GALAR, M.; FERNANDEZ, A.; BARRENECHEA, E.; BUSTINCE, H.; HERRERA, F. A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, v. 42, n. 4, p. 463–484, 2012.
- GARCIA, R. Uses of agent-based modeling in innovation/new product development research. *Journal of Product Innovation Management*, Wiley Online Library, v. 22, n. 5, p. 380–398, 2005.
- GARTNER. *Hype Cycle for Supply Chain Planning, 2015*. Gartner, Inc., 2015. Disponível em: <<https://www.gartner.com/doc/3172030/hype-cycle-supply-chain-planning>>.
- GUTJAHR, W. J.; PICHLER, A. Stochastic multi-objective optimization: a survey on non-scalarizing methods. *Annals of Operations Research*, Springer, v. 236, n. 2, p. 475–499, 2016.
- HAIXIANG, G.; YIJING, L.; SHANG, J.; MINGYUN, G.; YUANYUE, H.; BING, G. Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications*, v. 73, p. 220–239, 2017.
- HAN, H.; WANG, W.-Y.; MAO, B.-H. Borderline-smote: a new over-sampling method in imbalanced data sets learning. *Advances in Intelligent Computing*, Springer, p. 878–887, 2005.
- HANLEY, J. A.; MCNEIL, B. J. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, v. 148, n. 3, p. 839–843, 1983.
- HARTZEL, K. S.; WOOD, C. A. Factors that affect the improvement of demand forecast accuracy through point-of-sale reporting. *European Journal of Operational Research*, Elsevier, v. 260, n. 1, p. 171–182, 2017.
- HE, H.; GARCIA, E. A. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, v. 21, n. 9, p. 1263–1284, 2009.
- HE, H.; MA, Y. *Imbalanced learning: foundations, algorithms, and applications*. [S.l.]: John Wiley & Sons, 2013.
- HEIMAN, G. *Research Methods in Statistics*. [S.l.]: None, 2002. 394–397 p.
- HOENS, T. R.; POLIKAR, R.; CHAWLA, N. V. Learning from streaming data with concept drift and imbalance: an overview. *Progress in Artificial Intelligence*, Springer, v. 1, n. 1, p. 89–101, 2012.
- HOSMER JR, D. W.; LEMESHOW, S.; STURDIVANT, R. X. *Applied logistic regression*. [S.l.]: John Wiley & Sons, 2013. v. 398.

JI, S.; XU, W.; YANG, M.; YU, K. 3d convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 35, n. 1, p. 221–231, 2013.

JOLLIFFE, I. T. Principal component analysis and factor analysis. In: *Principal component analysis*. [S.l.]: Springer, 1986. p. 115–128.

KOHAVI, R. et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: STANFORD, CA. *International Joint Conference on Artificial Intelligence*. [S.l.], 1995. v. 14, n. 2, p. 1137–1145.

KOTSIANTIS, S.; KANELLOPOULOS, D.; PINTELAS, P. et al. Handling imbalanced datasets: A review. *GESTS International Transactions on Computer Science and Engineering*, v. 30, n. 1, p. 25–36, 2006.

KRAWCZYK, B.; WOŹNIAK, M.; SCHAEFER, G. Cost-sensitive decision tree ensembles for effective imbalanced classification. *Applied Soft Computing*, v. 14, p. 554–562, 2014.

KUBAT, M.; HOLTE, R. C.; MATWIN, S. Machine learning for the detection of oil spills in satellite radar images. *Machine learning*, Springer, v. 30, n. 2-3, p. 195–215, 1998.

LANDGREBE, T. C.; PACLIK, P.; DUIN, R. P. Precision-recall operating characteristic (p-roc) curves in imprecise environments. In: IEEE. *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*. [S.l.], 2006. v. 4, p. 123–127.

LEE, H.; PHAM, P.; LARGMAN, Y.; NG, A. Y. Unsupervised feature learning for audio classification using convolutional deep belief networks. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2009. p. 1096–1104.

LEE, H. L.; PADMANABHAN, V.; WHANG, S. Information distortion in a supply chain: The bullwhip effect. *Management science*, INFORMS, v. 43, n. 4, p. 546–558, 1997.

LEMAITRE, G.; NOGUEIRA, F.; ARIDAS, C. K. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research*, v. 18, n. 17, p. 1–5, 2017.

LIAW, A.; WIENER, M. et al. Classification and regression by randomforest. *R News*, v. 2, n. 3, p. 18–22, 2002.

LIU, X.-Y.; WU, J.; ZHOU, Z.-H. Exploratory undersampling for class-imbalance learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, IEEE, v. 39, n. 2, p. 539–550, 2009.

LÓPEZ, V.; FERNÁNDEZ, A.; GARCÍA, S.; PALADE, V.; HERRERA, F. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information Sciences*, v. 250, p. 113–141, 2013.

LUSTIG, M.; DONOHO, D.; PAULY, J. M. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine*, Wiley Online Library, v. 58, n. 6, p. 1182–1195, 2007.

MOUSAVI, S. M.; HAJIPOUR, V.; NIAKI, S. T. A.; AALIKAR, N. A multi-product multi-period inventory control problem under inflation and discount: a parameter-tuned particle swarm optimization algorithm. *The International Journal of Advanced Manufacturing Technology*, Springer, v. 70, n. 9-12, p. 1739–1756, 2014.

- NAKAJIMA, S. Introduction to tpm: Total productive maintenance. *Productivity Press, Inc.*, 1988, p. 129, 1988.
- NGUYEN, T.; LI, Z.; SPIEGLER, V.; IEROMONACHOU, P.; LIN, Y. Big data analytics in supply chain management: A state-of-the-art literature review. *Computers & Operations Research*, Elsevier, 2017.
- PASANDIDEH, S. H. R.; NIAKI, S. T. A.; NIA, A. R. An investigation of vendor-managed inventory application in supply chain: the eoq model with shortage. *The International Journal of Advanced Manufacturing Technology*, Springer, v. 49, n. 1-4, p. 329–339, 2010.
- PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V. et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, v. 12, n. Oct, p. 2825–2830, 2011.
- PHUA, C.; ALAHAKOON, D.; LEE, V. Minority report in fraud detection: classification of skewed data. *Acm sigkdd explorations newsletter*, ACM, v. 6, n. 1, p. 50–59, 2004.
- PRINCIPE, J. C.; EULIANO, N. R.; LEFEBVRE, W. C. *Neural and adaptive systems: fundamentals through simulations*. [S.l.]: Wiley New York, 2000. v. 672.
- RAO, R. B.; KRISHNAN, S.; NICULESCU, R. S. Data mining for improved cardiac care. *ACM SIGKDD Explorations Newsletter*, ACM, v. 8, n. 1, p. 3–10, 2006.
- REN, F.; CAO, P.; LI, W.; ZHAO, D.; ZAIANE, O. Ensemble based adaptive over-sampling method for imbalanced data learning in computer aided detection of microaneurysm. *Computerized Medical Imaging and Graphics*, v. 55, p. 54–67, 2017.
- SAHA, A.; ROY, A.; KAR, S.; MAITI, M. Inventory models for breakable items with stock dependent demand and imprecise constraints. *Mathematical and Computer Modelling*, Elsevier, v. 52, n. 9, p. 1771–1782, 2010.
- SAITO, T.; REHMSMEIER, M. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PloS One*, Public Library of Science, v. 10, n. 3, p. e0118432, 2015.
- SMITH, V. L.; SUCHANEK, G. L.; WILLIAMS, A. W. Bubbles, crashes, and endogenous expectations in experimental spot asset markets. *Econometrica: Journal of the Econometric Society*, Wiley, p. 1119–1151, 1988.
- TALEIZADEH, A. A.; SADJADI, S. J.; NIAKI, S. T. A. Multiproduct epq model with single machine, backordering and immediate rework process. *European Journal of Industrial Engineering*, Inderscience Publishers Ltd, v. 5, n. 4, p. 388–411, 2011.
- TANG, Y.; ZHANG, Y.-Q.; CHAWLA, N. V.; KRASSER, S. Svms modeling for highly imbalanced classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, IEEE, v. 39, n. 1, p. 281–288, 2009.
- TROYANSKAYA, O.; CANTOR, M.; SHERLOCK, G.; BROWN, P.; HASTIE, T.; TIBSHIRANI, R.; BOTSTEIN, D.; ALTMAN, R. B. Missing value estimation methods for dna microarrays. *Bioinformatics*, Oxford University Press, v. 17, n. 6, p. 520–525, 2001.

TSOU, C.-S. Multi-objective inventory planning using mopso and topsis. *Expert Systems with Applications*, v. 35, n. 1, p. 136–142, 2008.

TUKEY, J. W. Comparing individual means in the analysis of variance. *Biometrics*, JSTOR, p. 99–114, 1949.

TÜRK, S.; ÖZCAN, E.; JOHN, R. Multi-objective optimisation in inventory planning with supplier selection. *Expert Systems with Applications*, Elsevier, v. 78, p. 51–63, 2017.

WALLER, M. A.; FAWCETT, S. E. Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. *Journal of Business Logistics*, Wiley Online Library, v. 34, n. 2, p. 77–84, 2013.

WANG, G.; GUNASEKARAN, A.; NGAI, E. W.; PAPADOPOULOS, T. Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics*, Elsevier, v. 176, p. 98–110, 2016.

WANG, S.; YAO, X. Diversity analysis on imbalanced data sets by using ensemble models. In: IEEE. *Computational Intelligence and Data Mining, 2009. CIDM'09. IEEE Symposium on*. [S.l.], 2009. p. 324–331.

WANG, S.; YAO, X. Using class imbalance learning for software defect prediction. *IEEE Transactions on Reliability*, v. 62, n. 2, p. 434–443, 2013.

YIJING, L.; HAIXIANG, G.; XIAO, L.; YANAN, L.; JINLING, L. Adapted ensemble classification algorithm based on multiple classifier system and feature selection for classifying multi-class imbalanced data. *Knowledge Based Systems*, v. 94, p. 88–104, 2016.

ZHU, T.; GREINER, R.; HÄUBL, G. Learning a model of a web user's interests. In: SPRINGER. *User Modeling*. [S.l.], 2003. v. 2702, p. 65–75.

## ANEXO A - ESTATÍSTICAS DESCRITIVAS

As Tabelas 18 e 19 a seguir apresentam as estatísticas descritivas básicas (média, desvio-padrão, mínimo, mediana, máximo, 1º e 3º quartis) dos dois conjuntos de dados após a etapa de pré-processamento, para cada uma dos atributos  $x$  disponíveis.

Tabela 18 – Estatísticas descritivas do conjunto de dados Kaggle após pré-processamento.

Atributo	Média - $\mu$	Desvio - $\sigma$	Mínimo	1º Quartil	Mediana	3º Quartil	Máximo
$x_1$	0,6330	0,3837	-1,0000	0,2311	0,8165	0,9988	1,0000
$x_2$	7,8859	6,8397	0,0000	4,0000	8,0000	8,0000	52,0000
$x_3$	0,0208	0,0860	0,0000	0,0000	0,0000	0,0000	1,0000
$x_5$	0,0723	0,1352	0,0000	0,0000	0,0000	0,1231	0,5774
$x_6$	0,1275	0,1896	0,0000	0,0000	0,0000	0,2914	0,7071
$x_7$	0,1840	0,2519	0,0000	0,0000	0,0000	0,4472	1,0000
$x_9$	0,0270	0,0473	0,0000	0,0000	0,0000	0,0471	0,5000
$x_{10}$	0,0868	0,1037	0,0000	0,0000	0,0261	0,1735	0,5774
$x_{11}$	0,1773	0,1852	0,0000	0,0000	0,1056	0,3587	0,7071
$x_{12}$	0,2706	0,2724	0,0000	0,0000	0,1838	0,5434	1,0000
$x_{14}$	0,0542	0,1156	0,0000	0,0000	0,0000	0,0657	1,0000
$x_{16}$	0,7820	0,2280	0,0000	0,7300	0,8200	0,9600	1,0000
$x_{17}$	0,7769	0,2222	0,0000	0,7200	0,8100	0,9500	1,0000
$x_{21}$	0,0005	0,0226	0,0000	0,0000	0,0000	0,0000	1,0000
$x_{22}$	0,0013	0,0182	0,0000	0,0000	0,0000	0,0000	1,0000
$x_{23}$	0,0004	0,0070	0,0000	0,0000	0,0000	0,0000	1,0000
$x_{24}$	0,2256	0,4180	0,0000	0,0000	0,0000	0,0000	1,0000
$x_{25}$	0,0002	0,0123	0,0000	0,0000	0,0000	0,0000	1,0000
$x_{26}$	0,1205	0,3255	0,0000	0,0000	0,0000	0,0000	1,0000
$x_{27}$	0,9634	0,1877	0,0000	1,0000	1,0000	1,0000	1,0000
$x_{28}$	0,0004	0,0208	0,0000	0,0000	0,0000	0,0000	1,0000

Tabela 19 – Estatísticas descritivas do conjunto de dados MRS após pré-processamento.

Atributo	Média - $\mu$	Desvio - $\sigma$	Mínimo	1º Quartil	Mediana	3º Quartil	Máximo
$x_1$	0,8005	0,3180	0,0000	0,6707	1,0000	1,0000	1,0000
$x_2$	63,553	35,002	10,000	41,000	60,000	79,000	320,00
$x_3$	0,0043	0,0418	0,0000	0,0000	0,0000	0,0000	1,0000
$x_4$	-2,9050	28,379	-357,00	0,0000	0,0000	0,0000	1.070,0
$x_5$	0,0310	0,0643	0,0000	0,0000	0,0000	0,0178	0,5774
$x_6$	0,0621	0,1255	0,0000	0,0000	0,0000	0,0413	0,7071
$x_8$	0,1070	0,2068	0,0000	0,0000	0,0000	0,0988	1,0000
$x_9$	0,0178	0,0580	-0,3824	0,0000	0,0000	0,0000	0,7071
$x_{10}$	0,0506	0,1106	-0,5000	0,0000	0,0000	0,0269	0,7071
$x_{11}$	0,0907	0,1665	-0,7071	0,0000	0,0000	0,1309	0,7071
$x_{13}$	0,1817	0,2799	-0,6980	0,0000	0,0000	0,3665	1,0000
$x_{14}$	0,0168	0,0605	0,0000	0,0000	0,0000	0,0000	0,9806
$x_{15}$	1.317,1	7.109,5	0,0100	12,021	76,820	474,688	331.919
$x_{18}$	2,5346	1,0320	0,0000	3,0000	3,0000	3,0000	3,0000
$x_{19}$	0,7572	0,7965	0,0000	0,0000	1,0000	1,0000	4,0000
$x_{20}$	1,9762	0,8106	0,0000	1,0000	2,0000	3,0000	3,0000

## ANEXO B - IMPORTÂNCIA DOS ATRIBUTOS

As Tabelas 20 e 21 a seguir apresentam a importância dos atributos calculada a partir do decaimento do critério de separação nas regras criadas pelas árvores de decisão base dos métodos aplicados aos conjuntos de dados.

Tabela 20 – Importância dos atributos dos modelos obtidos no conjunto de dados Kaggle.

Atributos	CART	GB	RF	RUS	SMT	UB
$x_1$	0,6038 ± 0,0024	0,2183 ± 0,0101	0,2458 ± 0,0179	0,7514 ± 0,0160	0,6962 ± 0,0031	0,6672 ± 0,0025
$x_2$	0,0180 ± 0,0010	0,0258 ± 0,0024	0,0248 ± 0,0006	0,0211 ± 0,0030	0,0282 ± 0,0012	0,0243 ± 0,0010
$x_3$	0,0849 ± 0,0096	0,0720 ± 0,0041	0,0504 ± 0,0025	0,0820 ± 0,0073	0,0671 ± 0,0017	0,0755 ± 0,0023
$x_5$	0,0226 ± 0,0093	0,0734 ± 0,0079	0,0923 ± 0,0171	0,0133 ± 0,0086	0,0160 ± 0,0060	0,0174 ± 0,0033
$x_6$	0,0155 ± 0,0082	0,0439 ± 0,0061	0,0854 ± 0,0157	0,0085 ± 0,0093	0,0090 ± 0,0066	0,0137 ± 0,0033
$x_7$	0,0176 ± 0,0081	0,0462 ± 0,0058	0,0733 ± 0,0147	0,0057 ± 0,0072	0,0050 ± 0,0025	0,0120 ± 0,0023
$x_9$	0,0525 ± 0,0036	0,0998 ± 0,0074	0,0615 ± 0,0043	0,0378 ± 0,0091	0,0443 ± 0,0044	0,0444 ± 0,0032
$x_{10}$	0,0189 ± 0,0021	0,0721 ± 0,0061	0,0662 ± 0,0053	0,0147 ± 0,0056	0,0122 ± 0,0034	0,0242 ± 0,0025
$x_{11}$	0,0091 ± 0,0014	0,0511 ± 0,0110	0,0605 ± 0,0052	0,0046 ± 0,0061	0,0069 ± 0,0016	0,0148 ± 0,0018
$x_{12}$	0,0921 ± 0,0097	0,0780 ± 0,0108	0,0610 ± 0,0049	0,0342 ± 0,0083	0,0400 ± 0,0016	0,0386 ± 0,0024
$x_{14}$	0,0158 ± 0,0013	0,0465 ± 0,0039	0,0403 ± 0,0010	0,0088 ± 0,0033	0,0154 ± 0,0014	0,0174 ± 0,0011
$x_{16}$	0,0161 ± 0,0014	0,0418 ± 0,0050	0,0497 ± 0,0012	0,0043 ± 0,0019	0,0196 ± 0,0019	0,0172 ± 0,0008
$x_{17}$	0,0159 ± 0,0016	0,0712 ± 0,0077	0,0518 ± 0,0008	0,0072 ± 0,0027	0,0237 ± 0,0019	0,0214 ± 0,0008
$x_{21}$	0 ± 0	0,0006 ± 0,0007	9,6710 ± 3,6007	0 ± 0	0 ± 0	1,1445 ± 6,2689
$x_{22}$	0,0007 ± 0,0003	0,0082 ± 0,0021	0,0060 ± 0,0014	0,0001 ± 0,0002	0,0018 ± 0,0005	0,0007 ± 0,0002
$x_{23}$	0,0043 ± 0,0006	0,0327 ± 0,0067	0,0145 ± 0,0032	0,0001 ± 0,0003	0,0023 ± 0,0006	0,0011 ± 0,0001
$x_{24}$	0,0047 ± 0,0008	0,0060 ± 0,0015	0,0071 ± 0,0004	0,0043 ± 0,0020	0,0075 ± 0,0009	0,0064 ± 0,0006
$x_{25}$	0 ± 0	0,0003 ± 0,0004	1,3383 ± 1,3049	0 ± 0	0 ± 0	0 ± 0
$x_{26}$	0,0015 ± 0,0004	0,0038 ± 0,0011	0,0056 ± 0,0001	0,0002 ± 0,0004	0,0024 ± 0,0005	0,0016 ± 0,0003
$x_{27}$	0,0050 ± 0,0005	0,0071 ± 0,0017	0,0028 ± 0,0002	0,0007 ± 0,0010	0,0016 ± 0,0004	0,0013 ± 0,0004
$x_{28}$	7,6506 ± 6,6121	0,0001 ± 0,0002	8,3788 ± 3,0572	0 ± 0	3,0929 ± 6,4450	7,6755 ± 1,9295

Tabela 21 – Importância dos atributos dos modelos obtidos no conjunto de dados MRS.

Atributos	CART	GB	RF	RUS	SMT	UB
$x_1$	0,5649 ± 0,1129	0,1778 ± 0,0617	0,1857 ± 0,0645	0,4654 ± 0,2865	0,0279 ± 0,0303	0,4083 ± 0,1069
$x_2$	0,0210 ± 0,0209	0,0177 ± 0,0225	0,0342 ± 0,0255	0,0292 ± 0,0327	0,0043 ± 0,0076	0,0401 ± 0,0166
$x_3$	0,0017 ± 0,0065	0,0421 ± 0,0188	0,0155 ± 0,0072	0 ± 0	9,9329 ± 0,0004	0,0022 ± 0,0032
$x_4$	0,0172 ± 0,0184	0,0729 ± 0,0249	0,0166 ± 0,0103	0,0026 ± 0,0102	0,0002 ± 0,0009	0,0032 ± 0,0042
$x_5$	0,0400 ± 0,0507	0,1208 ± 0,0347	0,0782 ± 0,0350	0,0491 ± 0,1184	0,2533 ± 0,2218	0,0269 ± 0,0113
$x_6$	0,0674 ± 0,0627	0,0443 ± 0,0287	0,0964 ± 0,0587	0,0521 ± 0,1187	0,0013 ± 0,0048	0,0450 ± 0,0241
$x_8$	0,0511 ± 0,0553	0,0807 ± 0,0287	0,0958 ± 0,0379	0,0737 ± 0,1050	0,0181 ± 0,0347	0,0925 ± 0,0492
$x_9$	0,0202 ± 0,0360	0,1256 ± 0,0393	0,0596 ± 0,0363	0,0357 ± 0,0492	0,0842 ± 0,1873	0,0238 ± 0,0176
$x_{10}$	0,0312 ± 0,0228	0,0225 ± 0,0255	0,0817 ± 0,0493	0,0193 ± 0,0326	0,3663 ± 0,2874	0,0484 ± 0,0298
$x_{11}$	0,0655 ± 0,0239	0,1481 ± 0,0481	0,1008 ± 0,0546	0,0532 ± 0,1237	0,0044 ± 0,0091	0,0376 ± 0,0244
$x_{13}$	0,0550 ± 0,0810	0,0691 ± 0,0289	0,1276 ± 0,0684	0,1650 ± 0,2627	0,0019 ± 0,0035	0,1585 ± 0,0855
$x_{14}$	0,0174 ± 0,0140	0,0113 ± 0,0116	0,0207 ± 0,0130	0,0042 ± 0,0132	0,0009 ± 0,0020	0,0185 ± 0,0078
$x_{15}$	0,0227 ± 0,0256	0,0420 ± 0,0405	0,0453 ± 0,0311	0,0300 ± 0,0416	0,0169 ± 0,0154	0,0515 ± 0,0244
$x_{18}$	0,0166 ± 0,0136	0,0162 ± 0,0172	0,0213 ± 0,0090	0,0176 ± 0,0266	0,1338 ± 0,0343	0,0264 ± 0,0135
$x_{19}$	0,0021 ± 0,0066	0,0022 ± 0,0042	0,0104 ± 0,0093	0,0024 ± 0,0081	0,0430 ± 0,0344	0,0081 ± 0,0053
$x_{20}$	0,0052 ± 0,0099	0,0059 ± 0,0094	0,0095 ± 0,0074	0 ± 0	0,0427 ± 0,0270	0,0082 ± 0,0076