



Universidade Federal do Piauí
Centro de Ciências da Natureza
Programa de Pós-Graduação em Ciência da Computação

Modelo para Classificação de Fornecedores da Administração Pública Baseado em Aprendizagem de Máquina Supervisionada

Joselito Mendes de Sousa Junior

Teresina-PI, Fevereiro de 2019

Joselito Mendes de Sousa Junior

**Modelo para Classificação de Fornecedores da
Administração Pública Baseado em Aprendizagem de
Máquina Supervisionada**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da UFPI (área de concentração: Sistemas de Computação), como parte dos requisitos necessários para a obtenção do Título de Mestre em Ciência da Computação.

Universidade Federal do Piauí – UFPI

Centro de Ciências da Natureza

Programa de Pós-Graduação em Ciência da Computação

Orientador: Prof. Dr. Vinícius Ponte Machado

Coorientador: Prof. Dr. Marcus Vinícius de Sousa Lemos

Teresina-PI

Fevereiro de 2019

Joselito Mendes de Sousa Junior

Modelo para Classificação de Fornecedores da Administração Pública Baseado em Aprendizagem de Máquina Supervisionada/ Joselito Mendes de Sousa Junior. – Teresina-PI, Fevereiro de 2019-

59 p. : il.; 30 cm.

Orientador: Prof. Dr. Vinícius Ponte Machado

Coorientador: Prof. Dr. Marcus Vinícius de Sousa Lemos

Dissertação (Mestrado) – Universidade Federal do Piauí – UFPI

Centro de Ciências da Natureza

Programa de Pós-Graduação em Ciência da Computação, Fevereiro de 2019.

1. Aprendizagem de Máquina.

2. Aprendizagem Supervisionado.

3. Classificação.

4. Sistema de Avaliação de Fornecedores.

I. Vinícius Ponte Machado. II. Marcus Vinícius de Sousa Lemos III. Universidade Federal do Piauí. IV. Programa de Pós-Graduação em Ciência da Computação. V. Departamento de Computação

CDU 02:141:005.7

Joselito Mendes de Sousa Junior

Modelo para Classificação de Fornecedores da Administração Pública Baseado em Aprendizagem de Máquina Supervisionada

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da UFPI (área de concentração: Sistemas de Computação), como parte dos requisitos necessários para a obtenção do Título de Mestre em Ciência da Computação.

Teresina-PI, 01 de Fevereiro de 2019:

Prof. Dr. Vinícius Ponte Machado
Orientador

**Prof. Dr. Marcus Vinícius de Sousa
Lemos**
Co-Orientador

**Prof. Dr. Pedro de Alcântara dos
Santos Neto**
Membro Interno

Prof. Dr. Raimundo Santos Moura
Membro Interno

Prof. Dr. Raimir Holanda Filho
Membro Externo

Teresina-PI
Fevereiro de 2019

*À minha família e à Lanara,
pelo incentivo e apoio em todas as minhas escolhas e decisões.*

Agradecimentos

Agradeço a Deus.

Agradeço à minha mãe, Euniceles, pela criação, paciência e sacrifícios que sempre fez por mim. Sem suas brigas e exigências diárias, muito provavelmente eu nem estaria aqui.

Agradeço ao meu pai, Joselito, por nunca ter deixado faltar nada em casa. Seus esforços diários na minha criação foram bem aproveitados.

Ao meu irmão de sangue, Guilherme, a quem espero ser sempre um exemplo. Obrigado pela atenção e por inflar o meu ego sempre que preciso.

Ao meu irmão de vida e grande amigo, Iago Breno, por sempre estar ao meu lado. É uma honra tê-lo como amigo, compadre, como revisor, como parceiro. Eu sei que sempre poderei contar contigo, mesmo com as dificuldades e chatices envolvendo a minha pessoa.

Ao meu outro irmão de vida, Roney Lira, pelo apoio nas horas difíceis, pelo impulso profissional, por me mandar estudar quando eu te atrapalhava nas horas de almoço e, principalmente, por ser um grande amigo.

Agradeço ao meu orientador, Vinícius Ponte Machado, por todos os conselhos, pela paciência e ajuda nesse período. E principalmente, por sempre acreditar em mim e em nossos projetos rabiscados em um papel. Sua orientação foi incrível e eu garanto que aprendi muito com os seus ensinamentos.

Ao meu coorientador, Marcus Vinícius, pela paciência e disponibilidade.

À minha namorada, Lanara. Você foi o elo mais forte na reta final dessa dissertação e merece muito crédito por sempre ter me dado amor, atenção, carinho e a certeza de que todos os dias vou dormir tranquilo lembrando de um sorriso teu. Posso ter chegado até aqui antes de te conhecer, mas depois que te conheci, eu descobri que posso muito mais. Obrigado por tudo.

Aos meus primos, em especial ao Washington Junior, Thales Rodrigues e Tiago Rodrigues, com quem dividi minha infância e até hoje me aturam ou surtam.

Ao meu padrinho, Lino Junior, e madrinha, Luciane, por sempre serem presentes na minha vida. Sou muito grato e orgulhoso de vocês.

Aos grandes professores da minha graduação e mestrado, Vinícius, Ricardo Lira, Rodrigo Veras, Laurindo, Ivan, Raimundo, Pedro, André Soares e André Macedo. E também aos que não chegaram a ser meus professores diretamente, Kelson e Érico. Todos vocês são fantásticos! Nunca mudem! E muito obrigado por tudo o que me ensinaram, seja

na vida acadêmica, no futebol, nos jogos ou nas figurinhas da copa.

Aos professores Leonardo Carvalho (Física), Braz (História) e Petrysson (Matemática). Nunca esquecerei seus ensinamentos.

Aos meu treinador e amigo, Yago Vasconcelos, por me dar força e acreditar no meu potencial. Hang!

Aos amigos que fiz ao longo da minha vida acadêmica, em especial ao Selles Gustavo e Sidiney Araújo. Espero sempre contar com vocês.

Ao meu afilhado, Bernardo, por ter o sorriso mais gostoso do mundo. Que eu possa ajudar o seu Brasil a ser um pouquinho melhor que o meu.

Aos companheiros de pesquisa, Aline, Imperes, Sidiney e Tarcísio pelas contribuições no trabalho, pela amizade e pelas festinhas.

Ao seu Zé, pela educação e por manter tudo sempre limpo.

Aos meus familiares e amigos que de alguma forma já fizeram parte da minha vida.

Ao Tribunal de Contas do Estado do Piauí, em especial ao Inaldo, por acreditar na ideia e pelo apoio para realização deste trabalho de pesquisa.

Ao Eddie Vedder, pela música de qualidade.

E, finalmente, a todos que, por um momento, fizeram parte da minha vida e contribuíram de maneira positiva para ajudarem a me tornar quem eu sou hoje.

“Love all, trust a few, do wrong to none.”
(William Shakespeare)

*“O otimista é um tolo. O pessimista, um chato.
Bom mesmo é ser um realista esperançoso.”*
(Ariano Suassuna)

“Acredite, você pode vencer.”
(Autor Desconhecido)

Resumo

Contratos públicos podem ser entendidos como ajustes realizados entre a Administração Pública e particulares, para a consecução de objetivos de interesse público, com regras e condições estabelecidas pela própria Administração (PIETRO, 1999). Em outras palavras, contratos públicos são realizados através de um modelo de compras públicas. No Brasil, respondem por mais de 19% do Produto Interno Bruto (PIB). O modelo de compras públicas passou, nas últimas décadas, por um processo de automação, com a criação do Portal de Compras Governamentais e a implantação de pregões eletrônicos. O crescimento do PIB nesse período foi acompanhado por uma elevação nos gastos públicos que reflete o aumento do número de contratos firmados e licitações ou compras diretas realizadas. Considerando o desafio das instituições de controle governamental brasileiras de garantir eficiência e regularidade desses processos, propõe-se neste trabalho um modelo computacional que utilize aprendizagem de máquina para a classificação de fornecedores públicos. Para o modelo proposto, baseado no problema de classificação binária, deverá aprender as características dos fornecedores considerados de baixo risco (bons) e dos fornecedores considerados de alto risco (ruins) e, com isso, classificar os novos fornecedores inseridos na base de dados dos fornecedores do TCE-PI. A base de dados utilizada foi fornecida pelo Tribunal de Contas do Estado do Piauí e é composta pela união de dados presentes em outras bases, como na Receita Federal, no Tribunal Superior Eleitoral e no Portal da Transparência. A abordagem utilizada nesse trabalho baseia-se em selecionar e preparar os dados presentes na base de dados dos fornecedores para, em seguida, realizar-se uma etapa de testes e uma etapa de desenvolvimento. Na etapa de testes são realizados vários experimentos com a ferramenta *WEKA* para fazer uma análise do melhor algoritmo a ser utilizado no problema de classificação. Nessa etapa, definiu-se como a melhor solução utilizar o algoritmo J48, que apresenta taxa de classificação superior a 82%, podendo, em alguns casos, chegar a 94%. Na etapa de desenvolvimento, foi efetivada a implementação do algoritmo J48 e do Sistema de Avaliação de Fornecedores (SAF) como produto final da pesquisa aplicada ao problema de classificação dos fornecedores.

Palavras-chaves: Aprendizagem de Máquina. Aprendizado Supervisionado. Classificação. Sistema de Avaliação de Fornecedores.

Abstract

Public contracts can be understood as an agreement between the Public and Private Administration, for the achievement of public interest objectives, with the rules and conditions established by the Administration (PIETRO, 1999). In other words, public contracts are made by a model of public acquisition. In Brazil, public contracts account for more than 19% of Gross National Product (GNP). The model of public acquisition have been passed, in the last decades, the public acquisition model has undergone an automation process, with the creation of the Government Acquisition Portal and the implantation of electronic trading sessions. The GNP's growth in this period was accompanied by a rise in public spending, which reflects in the increase in the number of contracts signed and biddings or direct purchases made. Considering the challenge of Brazilian government control institutions in ensuring efficiency and regularity ins these processes, it is proposed in this study a computational model that uses machine learning to public suppliers classification. To the proposed model, this model, based on the binary classification problem, should learn the characteristics of suppliers considered to be low-risk (good) and suppliers considered to be high-risk (bad) and thus be able to classify the new suppliers included in the suppliers database of TCE-PI. The database used was provided by the Court of Auditors of the State of Piauí and is composed by the union of data present in other bases, such as the Federal Revenue, The Superior Electoral Court and the Transparency Portal. The approach used in this work is based on the selection and preparation of the data present in the suppliers to, then, perform a test and development steps. In the test step several experiments are performed with the WEKA tool to make an analysis of the best algorithm to be used in the classification problem. In this step, the best solution was the implementation of the J48 algorithm, which has a classification rate of more than 82%, and in some cases may reach 94%. In the development stage, the implementation of the J48 algorithm was effected and the Supplier Evaluation System (SAF) was implemented as the final product of the research applied to the problem classification of suppliers.

Keywords: Classification. Machine Learning. Supervised Learning. Suppliers Evaluation System.

Lista de ilustrações

Figura 1 – Modelo de uma rede <i>Perceptron</i> de múltiplas camadas (adaptado de Silva, Spatti e Flauzino (2010))	10
Figura 2 – Árvore de classificação gerada pelo J48	12
Figura 3 – Matriz de confusão	13
Figura 4 – Composição da base de dados dos fornecedores	21
Figura 5 – Modelo proposto	28
Figura 6 – Gráfico comparativo entre os testes com as bases de dados balanceadas	38
Figura 7 – Gráfico comparativo entre os testes com as bases de dados balanceadas e seleção de atributos	43
Figura 8 – Fluxograma das atividades referentes à utilização do SAF	46
Figura 9 – Tela inicial do SAF	47
Figura 10 – Tela de treinamento do SAF	47
Figura 11 – Tela de resultados do treinamento do SAF	48
Figura 12 – Fragmento da árvore de treinamento do SAF	48
Figura 13 – Gráfico representativo da curva de acerto referente aos verdadeiros negativos para os testes realizados	51
Figura 14 – Gráfico comparativo das métricas precisão, <i>recall</i> e <i>f-measure</i>	51

Lista de tabelas

Tabela 1 – Comparativo dos trabalhos relacionados utilizando os algoritmos <i>naive bayes</i> (NB), J48 e <i>random forest</i> (RF)	19
Tabela 2 – Relação entre testes com as bases de dados e a métrica de avaliação dos algoritmos	26
Tabela 3 – Resultados dos testes no <i>WEKA</i> com a base de dados desbalanceada .	32
Tabela 4 – Matrizes de confusão para os testes com a base de dados desbalanceada	33
Tabela 5 – Resultados dos testes no <i>WEKA</i> com a base de dados balanceada BB1	33
Tabela 6 – Matrizes de confusão para os testes com a base de dados balanceada BB1	34
Tabela 7 – Resultados dos testes no <i>WEKA</i> com a base de dados balanceada BB2	35
Tabela 8 – Matrizes de confusão para os testes com a base de dados balanceada BB2	35
Tabela 9 – Resultados dos testes no <i>WEKA</i> com a base de dados balanceada BB3	36
Tabela 10 – Matrizes de confusão para os testes com a base de dados balanceada BB3	36
Tabela 11 – Resultados dos testes no <i>WEKA</i> com a base de dados balanceada BB4	37
Tabela 12 – Matrizes de confusão para os testes com a base de dados balanceada BB4	37
Tabela 13 – Resultados dos testes no <i>WEKA</i> com a base de dados balanceada BB1	39
Tabela 14 – Matrizes de confusão para os testes com a base de dados balanceada BB1	39
Tabela 15 – Resultados dos testes no <i>WEKA</i> com a base de dados balanceada BB2	40
Tabela 16 – Matrizes de confusão para os testes com a base de dados balanceada BB2	40
Tabela 17 – Resultados dos testes no <i>WEKA</i> com a base de dados balanceada BB3	41
Tabela 18 – Matrizes de confusão para os testes com a base de dados balanceada BB3	41
Tabela 19 – Resultados dos testes no <i>WEKA</i> com a base de dados balanceada BB4	42
Tabela 20 – Matrizes de confusão para os testes com a base de dados balanceada BB4	43
Tabela 21 – Resultados dos testes no <i>WEKA</i> com as quatro bases de dados balanceadas, seleção de atributos e dados normalizados	44
Tabela 22 – Matrizes de confusão para os testes com as quatro bases de dados balanceadas, seleção de atributos e dados normalizados	45

Lista de abreviaturas e siglas

CC	Cartão de Crédito
CNPJ	Cadastro Nacional de Pessoa Jurídica
IA	Inteligência Artificial
IBGE	Instituto Brasileiro de Geografia e Estatística
FP	Falso Positivo
FN	Falso Negativo
NB	Naive Bayes
PIB	Produto Interno Bruto
RAIS	Relação Anual de Informações Sociais
RL	Regressão Logística
SAF	Sistema de Avaliação de Fornecedores
SIAFEM	Sistema Integrado de Administração Financeira para Estados e Municípios
SIAFE	Sistema Integrado de Administração Financeira do Estado do Piauí
SIASG	Sistema de Administração de Serviços Gerais
SAGRES	Sistema de Acompanhamento da Gestão dos Recursos da Sociedade
TCE-PI	Tribunal de Contas do Estado do Piauí
TSE	Tribunal Superior Eleitoral
UAG	Unidades Administrativas Governamentais
VP	Verdadeiro Positivo
WEKA	<i>Waikato Environment for Knowledge Analysis</i>

Sumário

1	INTRODUÇÃO	1
1.1	Introdução	1
1.2	Objetivo	3
1.3	Motivação	4
1.4	Metodologia	5
1.5	Estrutura Organizacional	5
2	REFERENCIAL TEÓRICO	7
2.1	Aprendizagem de Máquina	7
2.1.1	Aprendizagem Supervisionada	8
2.1.1.1	Redes Neurais Artificiais	8
2.1.1.2	Algoritmo J48	11
2.1.2	Aprendizagem Não-Supervisionada	12
2.1.3	Métricas de Avaliação dos Algoritmos	13
2.2	Tribunal de Contas do Estado e a Prestação de Contas	14
2.3	Trabalhos Relacionados	16
3	MATERIAIS E MÉTODOS	21
3.1	Base de Dados	21
3.2	Preparação da Base de Dados	24
3.3	Detalhes dos Testes com o <i>WEKA</i>	25
3.4	Modelo Proposto	27
4	RESULTADOS E DISCUSSÃO	31
4.1	Análise dos Resultados	31
4.1.1	Testes com a Base de Dados Desbalanceada	32
4.1.2	Testes com as Bases de Dados Balanceadas	33
4.1.2.1	Base de Dados Balanceada BB1	33
4.1.2.2	Base de Dados Balanceada BB2	34
4.1.2.3	Base de Dados Balanceada BB3	35
4.1.2.4	Base de Dados Balanceada BB4	36
4.1.3	Testes com as Bases de Dados Balanceadas e Seleção de Atributos	38
4.1.3.1	Base de Dados Balanceada BB1 e Seleção de Atributos	38
4.1.3.2	Base de Dados Balanceada BB2 e Seleção de Atributos	39
4.1.3.3	Base de Dados Balanceada BB3 e Seleção de Atributos	41
4.1.3.4	Base de Dados Balanceada BB4 e Seleção de Atributos	42

4.1.4	Testes com as Bases de Dados Balanceadas, Seleção de Atributos e Normalização de Dados	44
4.2	Sistema de Avaliação de Fornecedores (SAF)	45
4.2.1	Sistema	46
4.3	Discussão dos Resultados	49
5	CONCLUSÕES	53
5.1	Conclusões	53
	REFERÊNCIAS	57

1 Introdução

Neste capítulo é discutida a importância do tema abordado. Além disso, apresentam-se os objetivos, a metodologia e a estrutura utilizada na elaboração do trabalho.

1.1 Introdução

Contratos públicos são entendidos como ajustes realizados entre a Administração Pública e particulares para a consecução de objetivos de interesse público. As regras e condições desses contratos são estabelecidas pela própria Administração, que pode ser qualquer órgão ou entidade pública das esferas federal, municipal ou estadual (PIETRO, 1999). A essas regras e condições para a consecução de contratos públicos denomina-se modelo de compras públicas.

Os gastos do governo, oriundos de contratos públicos e aquisições diretas de bens e serviços, vêm respondendo por aproximadamente 19% do Produto Interno Bruto (PIB) brasileiro nos últimos anos. Segundo o Instituto Brasileiro de Geografia e Estatística (IBGE), dados publicados no Relatório de Contas Nacionais referente ao primeiro semestre de 2018 quantificam em R\$ 19.9 bilhões o montante de despesas de consumo do governo nesse ano. As licitações e contratos públicos são o meio institucional pelo qual esse consumo se materializa, tendo relevante papel na busca pela eficiência e eficácia dos gastos públicos.

No Brasil, o modelo de compras públicas passou, nas últimas décadas, por um processo de automação, destacando-se a criação do Portal de Compras Governamentais e a implantação dos pregões eletrônicos a partir de 2005 (BRAGA et al., 2008). O crescimento do PIB brasileiro entre 2000 e 2010 foi acompanhado pela elevação do patamar dos gastos públicos, e isso reflete no aumento do número de contratos firmados e licitações ou compras diretas realizadas.

Segundo o Sistema de Administração de Serviços Gerais (SIASG), que possui como finalidade integrar os órgãos da Administração Pública Federal direta, autárquica e fundacional, somente nos três primeiros trimestres de 2018, foram realizadas pelo governo 99.046 processos de compras e assinados 14.897 contratos de prestação de serviços ou entrega continuada de bens, representando mais de R\$ 21 bilhões aos cofres públicos¹. Comparando esse valor com 2008, têm-se um valor 39% maior. Com isso, para os órgãos de controle do governo, o desafio gerado por essa grande quantidade de licitações realizadas e contratos assinados é atuar de maneira eficiente na identificação de fraudes e corrupção.

Toma-se como exemplo o Tribunal de Contas do Estado do Piauí (TCE-PI) — am-

¹ <http://paineldecompras.planejamento.gov.br>

biente onde a pesquisa foi desenvolvida. O TCE-PI é responsável pela análise e julgamento das prestações de contas de natureza contábil, financeira, orçamentária, além de inspeções operacionais e patrimoniais das unidades administrativas dos Poderes Legislativo, Executivo e Judiciário do Estado do Piauí. O TCE-PI conta, hoje, com mais de 500 servidores, não sendo necessariamente todos eles auditores. Considerando a possibilidade de auditar todos os processos de compras dos três primeiros trimestres de 2018 — 1.172 processos de compras² no Estado do Piauí — tem-se um número inviável de processos por cada auditor.

Assim, fica clara a necessidade de otimizar a escolha do que se vai efetivamente auditar, uma vez que o censo completo é impossível e antieconômico. Surge, portanto, a necessidade de racionalizar os recursos gastos com auditorias. Atuar preventivamente, evitando o surgimento de problemas — como fraudes — ou na economia de recursos destinados à auditoria também é algo importante, visto que a maior parte das falhas operacionais geram danos irrecuperáveis, tais como a paralisação de um benefício (obras ou serviços) público ou a necessidade de refazê-lo do início. Segundo [Balaniuk et al. \(2012\)](#), a aplicação da Inteligência Artificial pode auxiliar na resolução de problemas, como na verificação de risco de corrupção das unidades administrativas governamentais ou na quantificação de risco das empresas na execução de contratos.

Outra questão que pode ser resolvida com o uso de Inteligência Artificial refere-se ao problema dos fornecedores. Quando uma licitação é vencida por uma empresa e um contrato é assinado, tem-se uma relação entre o Estado e a empresa (fornecedor). Dentro dessa relação podem ocorrer alguns problemas como quebras contratuais, benefícios paralisados ou baixa qualidade no serviço prestado. Esses problemas que são causados, na maioria das vezes, pela empresa prestadora de serviço são bastante prejudiciais à economia do Estado, uma vez que caracterizam atividades ilícitas — como desvio de verbas que poderiam ser investidas em desenvolvimento, educação, infraestrutura e saúde.

Quebra contratual é o nome dado à situação em que o Estado e uma ou mais empresas assinam um contrato e, posteriormente, um dos envolvidos na assinatura do contrato anuncia o rompimento contratual por algum motivo. A paralisação de um benefício acontece quando o mesmo é iniciado e não é finalizado. Essa situação é ocasionada por vários motivos, como a ausência ou carência de um macroplanejamento, falta de aparelhamento das equipes e ausência de contratos mais equilibrados (com regras melhor definidas). A maior parte dos problemas referentes à paralisação de serviços diz respeito ao saneamento básico e à infraestrutura³. Com isso, tem-se que os problemas referentes aos contratos entre o Estado e uma empresa caracterizam um desperdício do dinheiro público.

Os gastos com auditoria têm ficado cada vez menores perto da necessidade de evitar a grande quantidade de problemas contratuais ao longo dos anos. Com isso, o presente

² <http://paineldecompras.planejamento.gov.br>

³ <http://www.sienge.com.br>

trabalho busca desenvolver um modelo de classificador que auxilie na seleção de objetos a serem auditados. Ou seja, um estudo será realizado através de uma análise referente aos fornecedores estaduais, identificando as características comuns a eles e classificando os fornecedores em alto risco e/ou baixo risco. Essa classificação é dada de acordo com o relatório de prestação de contas feito pelo TCE-PI. Um fornecedor é considerado de alto risco quando foi citado em algum relatório de contas, ou seja, possui envolvimento com algum tipo de fraude ou quebra contratual. Para fornecedores de baixo risco não existem citações em relatórios de auditoria. De maneira geral, fornecedores de alto risco são considerados fornecedores não confiáveis e fornecedores de baixo risco são considerados fornecedores confiáveis.

Assim, espera-se desenvolver um estudo que servirá de base para a implementação de um sistema capaz de aprender o comportamento de fornecedores já classificados e, com isso, avaliar fornecedores que ainda não fazem parte da base dos fornecedores do TCE-PI. Essa avaliação e classificação de novos fornecedores surge como saída para direcionar os gastos com auditoria para possíveis atividades fraudulentas. Dessa forma, o estudo servirá também para evitar que empresas com tendências fraudulentas sejam contratadas sem uma fiscalização mais intensa antes da assinatura do contrato. Em outras palavras, com o estudo buscam-se maneiras para auxiliar o auditor, guiando-o a observar contratos pontuais e racionalizar os recursos gastos em auditorias.

1.2 Objetivo

Esta Dissertação possui como objetivo geral elaborar um modelo de classificação de dados que, através do conhecimento adquirido referente à base de dados dos fornecedores estaduais, avalie e classifique fornecedores desconhecidos (com predição não conhecida). Em outras palavras, os novos fornecedores inseridos na base de dados dos fornecedores do TCE-PI, serão avaliados e classificados quanto ao risco contratual, podendo ser classificados como fornecedores de alto risco ou fornecedores de baixo risco.

Ao alcançar o objetivo geral, espera-se que a taxa de acerto para a classificação dos fornecedores do Piauí seja satisfatória. Para a classificação ser considerada satisfatória, a taxa de acerto deverá ser superior a 75% e a análise da matriz de confusão para a classificação deve obter mais de 90% de classificação correta para os verdadeiros negativos (fornecedores de alto risco).

De acordo com o especialista, uma taxa de classificação em torno de 75% é satisfatória, levando em consideração a redução no escopo do problema. A análise da matriz de confusão permite entender quantos fornecedores de alto ou baixo risco foram classificados de maneira correta. Nesse caso, o objetivo é classificar o máximo de fornecedores de alto risco que são, de fato, fornecedores de alto risco. Com isso, a taxa de acerto referente aos

verdadeiros negativos da matriz de confusão (fornecedores de alto risco) deve ser superior a 90% para ser considerada viável e satisfatória.

Para atingir o objetivo geral da Dissertação, alguns objetivos específicos são relacionados, como:

- Testes com a base de dados e suas variações, derivadas da preparação de dados (balanceamento, normalização e seleção de atributos);
- Escolha do melhor algoritmo para o problema;
- Desenvolvimento do sistema responsável por integrar todo o conhecimento adquirido durante a pesquisa e capaz de classificar os fornecedores com índice de acerto, referente à classificação de fornecedores de alto risco, satisfatório.

1.3 Motivação

A classificação de dados tem sido considerada como um dos tópicos mais relevantes dentre aqueles existentes na área de aprendizagem de máquina e mineração de dados (ALPAYDIN; BACH, 2014). Assim, embora o desenvolvimento e aprimoramento de algoritmos que enfrentam esse problema tenham sido foco de muitos pesquisadores, poucos trabalhos lidam explicitamente com a análise de dados aplicado ao controle de contas.

A utilização de Inteligência Artificial aplicada ao controle de contas permite catalogar e avaliar uma grande quantidade de dados de maneira mais simples. Com a aplicação do estudo de aprendizagem de máquina, mineração de dados e classificação de dados à base dos fornecedores do Estado do Piauí, espera-se elaborar um modelo de classificação de dados que servirá como base para um sistema de avaliação e classificação de fornecedores.

O sistema de avaliação e classificação de fornecedores deverá prever o comportamento de fornecedores do Estado do Piauí e, com isso, auxiliar os auditores (especialistas) nos procedimentos referentes às auditorias de licitações. Além de auxiliar na economia dos recursos empregados com as auditorias, espera-se que, com o estudo, seja possível evitar a assinatura de contratos que são considerados fraudulentos — através da análise dos fornecedores de alto risco. Em outras palavras, o estudo presente nessa Dissertação deverá ocasionar uma queda nos problemas envolvendo contratos e licitações e, com isso, espera-se que a economia referente aos gastos com esse tipo de problema possa gerar desenvolvimento para o Estado do Piauí através de investimentos em outras atividades.

1.4 Metodologia

A metodologia utilizada no presente trabalho consta de algumas etapas que envolvem desde o estudo do problema ao desenvolvimento do sistema final, produto do conhecimento adquirido com a pesquisa.

A primeira etapa é caracterizada pelo estudo dos problemas referentes ao controle de contas, do problema da classificação, dos trabalhos relacionados e dos algoritmos candidatos à resolução do problema. Nessa etapa entendeu-se como funciona o procedimento de prestação de contas e os custos envolvidos no processo de auditoria. Além disso, foi possível observar como acontece a avaliação de riscos através dos estudo de autores que utilizaram aprendizagem de máquina na resolução do problema.

O estudo do problema de classificação permitiu a conexão com a ferramenta de mineração de dados denominada *WEKA*. Essa ferramenta possui alguns módulos, dentre eles o *AUTOWEKA*. Nesse módulo é possível ter como entrada uma base de dados e, a partir de uma série de configurações definidas, obter uma lista com os melhores algoritmos e os valores de seus parâmetros para a resolução de um problema de classificação.

A utilização da ferramenta de mineração *WEKA* sugeriu a utilização de testes envolvendo a linguagem de programação *JAVA*, devido a integração facilitada entre a linguagem e a biblioteca. Com isso, foram realizados vários experimentos utilizando a base de dados dos fornecedores. Os experimentos foram executados utilizando procedimentos de preparação de dados, onde realizou-se balanceamento, normalização e seleção de atributos para a base de dados.

Finalmente, para o desenvolvimento do sistema, a utilização da ferramenta *WEKA* foi relacionada à linguagem de programação *PYTHON*. Essa relação foi definida devido ao excelente desempenho da linguagem na resolução de problemas de aprendizagem de máquina e à estabilidade no desenvolvimento de sistemas computacionais.

1.5 Estrutura Organizacional

Além deste capítulo, este trabalho está organizado tendo como referencial teórico o Capítulo 2, onde são apresentados os assuntos e técnicas utilizadas na abordagem proposta: Aprendizagem de Máquina, Controle de Contas do TCE-PI e Trabalhos Relacionados. No Capítulo 3, são apresentadas a abordagem utilizada na resolução do problema. O Capítulo 4 mostra os resultados obtidos da aplicação do método utilizado na base do TCE-PI e trata da implementação e funcionamento do sistema de avaliação desenvolvido como produto desta Dissertação. Finalmente, o Capítulo 6 discute alguns problemas encontrados e apresenta sugestões de trabalhos futuros.

2 Referencial Teórico

Neste capítulo é apresentado o embasamento teórico para auxiliar a compreensão do presente trabalho. Na primeira seção apresenta-se uma introdução à área de Aprendizagem de Máquina, mais especificamente em seus dois principais paradigmas de aprendizagem: supervisionada e não-supervisionada, com foco na técnica utilizada no modelo proposto. Em seguida, é dada uma explanação sobre o controle de contas realizado no Tribunal de Contas do Estado do Piauí. Finalmente, são apresentados os trabalhos relacionados.

2.1 Aprendizagem de Máquina

Segundo [Mitchell \(1997\)](#), Aprendizagem de Máquina é um ramo da Inteligência Artificial que lida com o estudo de métodos computacionais que permitem programas de computadores obter uma melhoria em seu desempenho, de forma autônoma, em uma determinada tarefa por meio de experiências. Ou seja, o uso de aprendizagem de máquina lida com o problema de modo que a própria máquina irá encontrar, após um processo de aprendizagem, uma hipótese que melhor o define. Para isso, a aprendizagem de máquina baseia-se em ideias de um conjunto diversificado de disciplinas incluindo Inteligência Artificial, Probabilidade, Estatística, Psicologia, Neurobiologia e Filosofia.

Podemos dividir a área de aprendizagem de máquina basicamente em dois paradigmas de aprendizagem: supervisionada e não-supervisionada. Em ambos os casos existe uma busca por um modelo capaz de generalizar dados desconhecidos, sendo diferenciados basicamente pela existência de um rótulo (resposta) presente nos dados utilizados na aprendizagem supervisionada. Na primeira, busca-se a criação de um modelo preciso em relação à predição de valores para novos dados enquanto que na segunda o objetivo é encontrar características que podem resumir os dados.

Existem ainda outras abordagens da aprendizagem de máquina, tais como aprendizagem semi-supervisionado e aprendizagem por reforço (ambas não discutidas no escopo deste trabalho). Na aprendizagem semi-supervisionado, existe uma tentativa de aprimorar um classificador criado a partir de dados rotulados com o uso de amostras não-rotuladas ([BARBER, 2012](#)). A aprendizagem por reforço é um método de aprendizado de máquina não-supervisionado cujo objetivo é a construção de algoritmos que realizam o aprendizado a partir da interação de um agente com um ambiente, e baseia-se nos conceitos matemáticos de programação dinâmica ([BELLMAN, 2003](#)) ([COSTA, 2017](#)).

2.1.1 Aprendizagem Supervisionada

Os métodos de aprendizagem supervisionada são capazes de modelar padrões existentes nos dados. Um elemento existente em uma base de dados qualquer pode ser definido por um conjunto de pares de valores que contêm uma característica (atributo) e seu respectivo valor. Esses valores servirão de entrada para um modelo, sendo utilizados diretamente em seu processo de aprendizagem. Nesse caso, para cada conjunto de valores de entrada existe uma saída que deverá ser apresentada ao processo.

Pode-se fazer uma relação da aprendizagem supervisionada com um supervisor que indica a saída que deverá ser apresentada para determinados valores de entrada. Dessa forma, esse processo pode ser modelado, entre outras coisas, como um problema de classificação onde cada amostra, contendo um conjunto de características como entrada e um conjunto de respostas como saída, é utilizada durante as etapas de treinamento e teste.

Os métodos utilizados nas etapas de treinamento e teste podem variar e devem ser ajustados conforme a técnica aplicada. Nesse contexto, a etapa de treinamento consiste em utilizar parte dos dados para a elaboração de um modelo classificador. Essa parte dos dados é conhecida como conjunto de treinamento. Os valores de entrada são atribuídos ao modelo e seu resultado é então comparado à resposta esperada. O modelo produzido é reajustado sempre que houver erro ou até que se atinja uma margem de erro aceitável. Quando esta condição for finalmente satisfeita - ou que se atinja uma quantidade máxima de iterações - segue-se para a etapa de testes na qual a parte restante dos dados (denominada conjunto de teste) será utilizada para medir a acurácia do modelo produzido.

[Barber \(2012\)](#) apresenta como definição para a aprendizagem supervisionada um conjunto de dados $D = (x_n, y_n), n = 1, \dots, N$ contendo N elementos, no qual o objetivo é criar um modelo capaz de aprender a relação existente entre os valores de entrada x e os valores de saída y , de modo que quando uma nova entrada x^* é fornecida, a predição de sua saída y^* seja precisa.

2.1.1.1 Redes Neurais Artificiais

Segundo ([SILVA; SPATTI; FLAUZINO, 2010](#)), as Redes Neurais Artificiais (RNAs) são modelos computacionais que possuem capacidade de aquisição e manutenção de conhecimento baseado em informações, representados por um conjunto de unidades de processamento.

Com inspiração no sistema nervoso de seres vivos, as RNAs possuem diversas características interessantes que as tornam candidatas a serem aplicadas no presente trabalho: adaptação por experiência, tolerância a falhas, organização dos dados, habilidade de generalização e, especialmente, capacidade de aprendizado.

Historicamente, o primeiro neurônio artificial foi elaborado por [McCulloch e Pitts](#)

(1990). Os autores propuseram o primeiro modelo matemático inspirado em um neurônio biológico resultando na concepção do que seria o primeiro neurônio artificial. Em seguida, uma rede neural artificial em sua forma mais simples (denominada *Perceptron*) foi apresentada por Rosenblatt (1958).

Inicialmente, para o *Perceptron*, existe uma etapa de treinamento, onde é fornecida uma camada de dados de entrada na rede. Após a entrada de dados, a rede é submetida a um processamento onde são calculadas as saídas com base em uma função de ativação. O valor da saída é comparado ao valor esperado e, se o erro não estiver dentro de um limite aceitável, definido previamente, aplica-se um método de treinamento que irá ajustar os pesos da rede. Esse processo, iterativo, de reajuste se repete até que o erro não exista ou esteja dentro de um limite aceitável — ou ainda que o número máximo de iterações seja extrapolado.

Após o processo de treinamento ocorre a etapa de testes que tem como objetivo avaliar a performance do modelo gerado. Nesta etapa são apresentados dados desconhecidos pela rede — isto é, dados não apresentados durante a etapa de treinamento — que passam pelas mesmas operações da etapa anterior. Comparando o resultado apresentado pela rede com o resultado desejado estima-se o quão preciso é o modelo gerado.

O modo de funcionamento das redes neurais atraiu diversos pesquisadores desencadeando um série de trabalhos e pesquisas, conduzindo ao surgimento de diversos novos modelos de redes neurais como redes recorrentes de Hopfield (HOPFIELD, 1982), redes auto-organizáveis de Kohonen (KOHONEN, 1982), entre outras apresentadas em (HAYKIN, 2001).

Dentre os modelos de redes neurais artificiais propostos existe a rede *Perceptron* de múltiplas camadas (PMC), ou em inglês *Multilayer Perceptron* (MLP). Conforme Silva, Spatti e Flauzino (2010), as redes PMC se caracterizam pela presença de pelo menos uma camada intermediária (ou camada oculta) de neurônios, localizada entre a camada de entrada e a camada neural de saída da rede. Assim como o *Perceptron*, as redes PMCs possuem uma topologia do tipo *feedforward*, isto é, as saídas dos neurônios de uma dada camada servem como entrada exclusivamente para neurônios de camadas adiantes. A Figura 1 ilustra uma rede *Perceptron* de múltiplas camadas.

Assim como na rede *Perceptron*, os sinais de entrada de uma rede PMC (x_1, x_2, \dots, x_n) possuem pesos associados a cada neurônio da camada seguinte (primeira camada oculta), representados por $w_{j,i}^{(1)}$, onde j representa o neurônio e i a entrada. O valor de entrada I de cada neurônio j pertencente à camada 1 é calculado pelo somatório dos produtos dos valores de entrada da camada anterior.

Em seguida, o valor de entrada de um neurônio é aplicado a uma função de ativação. Geralmente, a função mais utilizada é a função tangente hiperbólica, embora qualquer

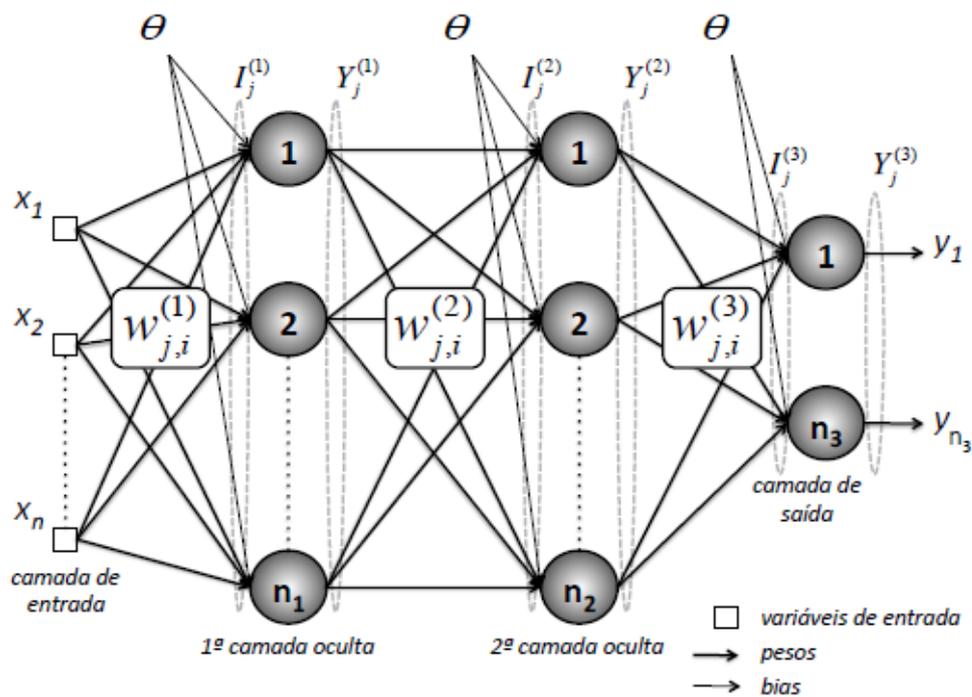


Figura 1 – Modelo de uma rede *Perceptron* de múltiplas camadas (adaptado de [Silva, Spatti e Flauzino \(2010\)](#))

outra função totalmente diferenciável possa ser utilizada como por exemplo as funções logística, linear ou gaussiana. Em um neurônio j , para um dado valor de entrada $I_j^{(1)}$ a função de ativação apresenta, então, um valor de saída Y_j^1 .

Estes valores de saída multiplicados por seus respectivos pesos $w_{j,i}^2$ e somados com as unidades de *bias* servirão como entrada para os neurônios da camada seguinte dando origem às entradas I_j^2 da segunda camada oculta. O processo se repete de forma análoga para a segunda camada e, finalmente, a camada de saída apresenta em cada neurônio os valores finais da rede.

De um modo geral, em uma rede PMC qualquer podem existir L camadas ocultas de modo que o processo de propagação dos sinais de entrada ocorre de maneira análoga ao explicado anteriormente. Dessa forma, cada uma das L camadas possuem pesos, representados por $w_{j,i}^l$ com $1 \leq l \leq L$, que combinados com os valores de entrada e as unidades de *bias* geram os valores de entrada I_j^l e, finalmente, os valores de saída Y_j^l .

Cada saída da rede é então comparada com sua respectiva saída desejada. A existência de um erro inaceitável requer a aplicação de algum método de treinamento que terá como objetivo reajustar os pesos da rede. Os métodos mais comumente utilizados com o algoritmo de *backpropagation* é o método do gradiente descendente embora outros métodos, como por exemplo ([LEVENBERG, 1944](#)), possam ser utilizados. Finalmente, após o processo de treinamento ocorre a etapa de teste da rede.

2.1.1.2 Algoritmo J48

O algoritmo C4.5 foi desenvolvido por (QUINLAN, 1993) e representa uma árvore de decisão. Como essas árvores podem ser utilizadas em classificação de dados, elas são conhecidas como classificadores estatísticos.

A seleção da melhor partição dos nós e o critério de parada são baseados na entropia de Shannon, como é usual em parte da família de indução de árvores de classificação (MITCHELL, 1997). A entropia é normalmente utilizada para medir a pureza ou impureza de um determinado conjunto, e é dada pela Equação 2.1. Para a Equação 2.1, $p(x)$ é a proporção de exemplos em relação a todo o conjunto e $\log_2 p(x)$ é o número no qual tem-se que elevar a 2 para chegar ao valor em questão.

$$\text{entropia} = \sum_{i=1}^n p(x_i) \cdot \log_2 p(x_i) \quad (2.1)$$

Para o presente trabalho, utiliza-se a implementação de uma variação do algoritmo C4.5, implementada em *JAVA* e utilizada no *WEKA*: o algoritmo J48. O J48 é um algoritmo de aprendizagem supervisionada definido como um indutor *top-down* de árvores de classificação que reimplementa o algoritmo C4.5, proposto por (QUINLAN, 1993), no *WEKA* (WITTEN et al., 1999). A utilização do J48 é justificada pelo desempenho explicado na seção de resultados e por contribuir de maneira positiva ao:

- lidar tanto com atributos categóricos como também atributos contínuos. Para lidar com atributos contínuos, o algoritmo J48 define um limiar e então divide os exemplos de forma binária: aqueles cujo valor do atributo são maiores que o limiar e aqueles cujo valor do atributo são menores ou iguais ao limiar;
- utilizar a medida chamada *gain ratio* para selecionar o atributo que melhor divide os exemplos. Essa medida se mostrou superior ao ganho de informação, gerando árvores mais precisas e menos complexas. A medida *gain ratio* é definida como o ganho de informação relativo (ponderado) como critério de avaliação e é dada pela Equação 2.2;

$$\text{gain_ratio_}(nó) = \frac{\text{ganho_de_informação}}{\text{entropia_}(nó)} \quad (2.2)$$

- apresentar um método de pós-poda das árvores geradas, tornando-as menos complexas. O algoritmo J48 faz uma busca na árvore, de baixo para cima, e transforma em nós folha aqueles ramos que não apresentam nenhum ganho significativo (QUINLAN, 1993).

A Figura 2 mostra um exemplo de árvore gerada com a utilização do algoritmo J48 aplicada a um pequeno fragmento da base de dados utilizada no trabalho. Como nó raiz

tem-se o atributo *ST_EPPME* que apresenta valores binários (0 ou 1). Caso o fornecedor tenha valor do atributo igual a 1, é analisado o valor do atributo *CNAE_Qtd_Total*, que possui valor contínuo. Nesse caso, foi definido um limiar e os valores do atributo foram divididos em menores que 4 e maiores ou igual a 4. Caso o fornecedor possua valor de *CNAE_Qtd_Total* menor que 4, o atributo a ser analisado será *Idade_Empresa_2017*. Novamente, por se tratar de um atributo com valor contínuo, foi definido um limiar e os valores do atributo foram divididos em menores que 8 e maiores ou igual a 8. Caso o fornecedor possua valor de *Idade_Empresa_2017* maior ou igual a 8, é considerado um fornecedor de baixo risco (bom).

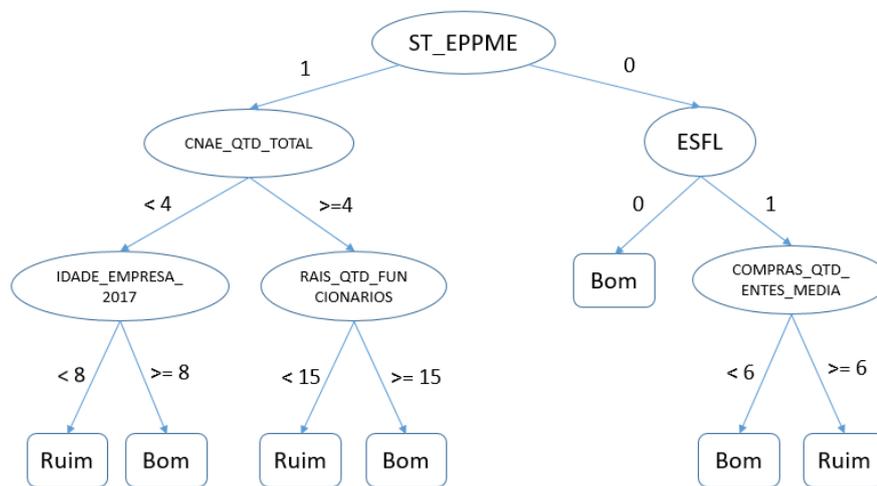


Figura 2 – Árvore de classificação gerada pelo J48

2.1.2 Aprendizagem Não-Supervisionada

Os métodos de aprendizagem não-supervisionada buscam detectar padrões existentes nos dados a fim de representá-los de forma resumida. Diferente da aprendizagem supervisionada, não existe aqui uma resposta para cada conjunto de valores de entrada do problema. Assim, o problema consiste em agrupar diversas amostras em classes distintas de modo que exista uma similaridade intragrupo, ou seja, os elementos de um mesmo grupo devem ser o mais semelhantes possível, e uma dissimilaridade intergrupo, ou seja, elementos de grupos diferentes tendem a ser diferentes entre si.

Formalmente, Barber (2012) define aprendizagem não-supervisionada da seguinte maneira: dado um conjunto de dados $D = (x_n), n = 1, \dots, N$ contendo N elementos, o objetivo é encontrar uma descrição plausível e compacta dos dados. Portanto, aprendizagem não-supervisionada pode ser vista em sua essência como um sinônimo para agrupamento, onde pretende-se atribuir classes aos elementos de uma base de dados.

2.1.3 Métricas de Avaliação dos Algoritmos

Durante a etapa de testes e análise dos resultados são utilizadas algumas métricas de avaliação pela busca do melhor classificador. As métricas escolhidas são bastante utilizadas na literatura e justificam-se pelo objetivo principal do trabalho de encontrar o classificador que melhor possa prever o curso dos fornecedores de alto risco, ou seja, um classificador capaz de acertar uma maior quantidade de verdadeiros negativos (em uma análise de matriz de confusão).

		Classe Prevista	
		Positivo	Negativo
Classe Real	Positivo	VP	FN
	Negativo	FP	VN

Figura 3 – Matriz de confusão

A Figura 3 representa a matriz de confusão na sua forma geral. No eixo das abscissas, encontram-se os valores da Classe Real, enquanto no eixo das ordenadas, encontram-se os valores da Classe Prevista. Quando os valores das classes real e prevista são positivos, ou seja, quando um elemento que possui um valor real *positivo* for classificado como *positivo*, ele é dito como Verdadeiro Positivo (VP). Quando um elemento possui valor real *positivo* e for classificado como *negativo*, ele é dito Falso Negativo (FN). Para um elemento com valor real *negativo* e classificado como *positivo*, sua classificação é Falso Positivo (FP). Por fim, para um elemento com valores real e previsto *negativo*, sua classificação é Verdadeiro Negativo (VN). As métricas escolhidas para a avaliação dos algoritmos são baseadas na ideia de matriz de confusão e estão representadas a seguir:

- *Precision* - Tem como objetivo identificar quantas amostras foram classificadas corretamente, ou seja, nenhum exemplo negativo é incluído. Em outras palavras, é uma medida do quão exato é a classificação para as amostras positivas (GHAMRAWI; MCCALLUM, 2005). Por conta disso, a métrica é conhecida como valor de predição positiva e é definida como o número de verdadeiros positivos (VP) dividido pela soma de verdadeiros positivos (VP) mais falsos positivos (FP), conforme Equação 2.3:

$$Precision = \frac{VP}{VP + FP} \quad (2.3)$$

De maneira geral, pode-se dizer que um baixo valor de *precision*, pode indicar um grande número de falsos positivos na classificação geral.

- *Recall* ou Cobertura - Também conhecido como taxa de sensibilidade ou taxa de verdadeiro positivo, é a taxa com que classificam-se como positivos todos os

elementos que são positivos, ou seja, nenhum exemplo positivo é deixado de fora (GODBOLE; SARAWAGI, 2004). É definida como o número de verdadeiros positivos (VP) dividido pela soma de verdadeiros positivos (VP) mais falsos negativos (FN), conforme Equação 2.4:

$$Recall = \frac{VP}{VP + FN} \quad (2.4)$$

O *recall* pode ser compreendido como uma medida do quão completo é o classificador. De maneira geral, um valor baixo do *recall* indica um alto valor de falsos negativos.

- *fMeasure* - É definido como a média harmônica entre *precision* e *recall* buscando otimizar as duas métricas, uma vez que a média harmônica é mais influenciada pela pior medida. Em outras palavras, a métrica busca um modelo que faça um balanço geral entre as duas métricas anteriores (POWERS, 2011) (Equação 2.5):

$$fMeasure = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (2.5)$$

- *Recall* ou Cobertura da Classe Negativa- Semelhante ao *recall*, é a taxa com que classificam-se como negativos todos os elementos que são negativos, ou seja, nenhum exemplo negativo é deixado de fora. É definida como o número de verdadeiros negativos (VN) dividido pela soma de verdadeiros negativos (VN) mais falsos positivos (FP), conforme Equação 2.6:

$$Recall_Negativo = \frac{VN}{VN + FP} \quad (2.6)$$

2.2 Tribunal de Contas do Estado e a Prestação de Contas

O Tribunal de Contas do Estado do Piauí, em sua trajetória de mais de um século, sempre manteve seu papel de resguardar a lei e o interesse coletivo na aplicação do dinheiro público. Dentre os Tribunais de Contas, foi o primeiro a ser implantado no Brasil, sendo criado depois apenas do Tribunal de Contas da União. Os Tribunais de Contas tem a função técnica de analisar e julgar os gastos públicos e possuem autonomia em relação aos outros poderes.

Ele foi criado através da primeira Constituição republicana estadual em 1899 e, durante a Ditadura Vargas, as suas funções e atribuições passaram a ser exercidas pela Secretaria de Fazenda. O TCE-PI foi restituído em 1946. Com a Constituição de 1988, abriu-se caminho para que o TCE-PI melhorasse a fiscalização dos negócios públicos. Além disso, a Constituição também estabeleceu concurso público para a escolha de auditores fiscais, assessores jurídicos e procuradores do Ministério Público de Contas.

Com a Lei de Responsabilidade Fiscal, o TCE-PI passou a exercer também um papel didático-pedagógico em relação à aplicação dos recursos públicos, em nível estadual e municipal. Sendo assim, ele continua desempenhando, com mais eficácia, sua função técnica de análise e julgamento das prestações de contas de natureza contábil, financeira, orçamentária, além de inspeções operacionais e patrimoniais das unidades administrativas dos Poderes Legislativo, Executivo e Judiciário.

Para auxiliar na execução das suas atividades o TCE-PI baseia-se em algumas instruções normativas, sendo as principais:

- Instrução normativa 06/2017 - Dispõe sobre os Sistemas Licitações, Contratos e Obras Web, especificando a forma e o prazo para o envio de informações relativas a licitações, adesões a sistemas de registros de preços, procedimentos administrativos de dispensa ou inexigibilidade e dos respectivos contratos administrativos ou outros instrumentos hábeis assemelhados, inclusive se relativos a obras e serviços de engenharia, componentes de prestação de contas da administração pública direta ou indireta do TCE-PI¹;
- Instrução normativa 07/2017 - Dispõe sobre a forma e prazo de prestação de contas ao Tribunal de Contas pelos órgãos e entidades dos Poderes Legislativo, Executivo e Judiciário, Ministério Público e Defensoria Pública do Estado do Piauí, e dá outras providências²;
- Instrução normativa 09/2017 - Dispõe sobre a forma e o prazo para o envio da prestação de contas da administração pública municipal direta e indireta ao TCE-PI, e dá outras providências³.

Existem alguns sistemas computacionais vinculados às instruções normativas, sendo os principais o SAGRES, SIAFEM (SIAFE), Documentação Web e Licitações Web.

O SAGRES é um sistema de prestação de contas eletrônica para os municípios e seus órgãos, que não substitui os programas de contabilidade tradicionais, mas que possibilita ao gestor informações gerenciais, além da prestação de contas simplificada ao tribunal de Contas, tendo em vista a progressiva redução de envio de documentação. Uma das características do SAGRES é a possibilidade do gestor que está em fim de mandato poder levar consigo uma cópia de todas as informações em meio eletrônico. Ele ainda é subdividido em SAGRES Contábil, responsável pela contabilidade da prefeitura e câmara, onde é realizada a contabilidade pura do negócio, e SAGRES Folha, responsável pela folha de pagamento do órgão, do pessoal, dos novos concursados e pela quantidade paga pelas prefeitura.

¹ <https://www.tce.pi.gov.br/instrucao-normativa-no-062017/>

² <https://www.tce.pi.gov.br/instrucao-normativa-no-07-2017-de-14-de-dezembro-de-2017/>

³ <https://www.tce.pi.gov.br/instrucao-normativa-no-09-2017-de-14-de-dezembro-de-2017/>

O SIAFEM, hoje conhecido por SIAFE, é um sistema que assemelha-se ao SAGRES Contábil, adequando-se às prestações de contas estaduais. É um sistema pertencente à Secretaria de Fazenda, mas é utilizado dentro do TCE-PI como ferramenta de prestação de contas.

O Documentação Web é um sistema que é utilizado para prestações de contas municipais e estaduais. Ele é responsável por receber informações não estruturadas, como PDFs da lei e notas fiscais. Por fim, o Licitações Web também é um sistema utilizado para prestações de contas municipais e estaduais, sendo responsável por auxiliar os órgãos na prestação de contas relativas aos processos licitatórios.

2.3 Trabalhos Relacionados

A aplicação dos conceitos de aprendizagem de máquina e a utilização de ferramentas estatísticas podem ser observados em diversos estudos nas mais diversas áreas. Especificamente, quanto à criação de modelos preditivos de risco, as aplicações mais comuns são na área de crédito bancário (citados a seguir). Nesta seção, serão abordadas aplicações que usam técnicas semelhantes nessa área, bem como outros trabalhos que usam técnicas diferentes, mas que tem como objetivo a construção de modelos preditivos.

Segundo [Caouette et al. \(2009\)](#), para a avaliação de risco de crédito bancário, a aplicação clássica da chamada pontuação de crédito (*Credit Scoring*) é a seleção de solicitantes ao crédito. O objetivo de [Caouette et al. \(2009\)](#) é mostrar soluções de tecnologia de ponta de risco de crédito juntamente com a perspectiva dos principais especialistas no campo, que as implantaram com sucesso.

Além da pontuação de crédito, existem, na literatura, exemplos de aplicações desses modelos em outras etapas do ciclo de vida do crédito. Ele pode ser aplicado na prospecção de clientes, quando recebe o nome de *Response Scoring*, durante a execução do contrato (*Behavioral Scoring*) ou mesmo após já consumada a inadimplência (*Collection Scoring*), para identificar devedores mais propensos a quitar a dívida ([SEMEDO, 2010](#)). Para [Semedo \(2010\)](#), seu estudo foi dirigido através de testes com a aplicação de regressão logística e redes neurais artificiais, onde apesar dos testes estatísticos indicarem que os modelos apresentam desempenhos semelhantes, as redes neurais representaram uma promissora técnica para a análise e concessão de crédito em Cabo Verde.

As visões relativas ao *Credit Scoring* ampliam a utilidade da ferramenta que, além de prever problemas, passa também a ser utilizada para identificar oportunidades para a organização. Segundo [Thomas, Edelman e Crook \(2002\)](#), modelos matemáticos podem ajudar credores a tomar decisões inteligentes de risco de crédito. Em seus estudos, a abordagem utilizada faz referência a dois tipos básicos de decisões e oferece modelos matemáticos sólidos para auxiliar no processo de tomada de decisão. A primeira delas

é conceder crédito a um novo candidato (*Credit Scoring*), e a segunda abordagem trata sobre como ajustar as restrições de crédito a um cliente atual (*Behavioral Score*).

Como exemplo de aplicação tradicional do *Credit Scoring*, [Vasconcellos \(2002\)](#) propõe uma metodologia de análise da concessão de crédito a pessoas físicas a partir do estudo matemático e estatístico, baseado na implementação do método de regressão logística. Em seu trabalho, foi utilizado como objeto de estudo, os hábitos de pagamentos e variáveis cadastrais, financeiras, patrimoniais e de relacionamento com a instituição credora dos clientes analisados.

Já [Araujo e Carmona \(2009\)](#), estudando uma instituição de microcrédito, aborda a possibilidade de utilização do *Credit Scoring* na modalidade *Behavioral Scoring* e compara os desempenhos de duas técnicas: Regressão Logística e Análise Discriminante. Os resultados do estudo demonstraram que os modelos *Credit Scoring* obtêm desempenho satisfatório quando utilizados na análise de risco de crédito na instituição de microcrédito, alcançando um percentual de classificação correta dos clientes de cerca de 80%, além de fornecer subsídios à instituição, auxiliando na prevenção e redução da inadimplência e na diminuição dos seus custos operacionais.

[Devacena \(2015\)](#), estudando um conjunto de dados de crédito alemão, realizou uma pesquisa examinando diferentes classificadores baseados em árvores de decisão (*random forest*, *REP Tree* e *J48*) para a previsão de crédito. De acordo com os dados utilizados em seu trabalho, os melhores resultados encontrados foram com a utilização do *random forest* e do *J48*. As métricas avaliadas foram a acurácia, o erro médio absoluto e o tempo de construção do modelo.

Na esfera governamental, [Carvalho \(2015\)](#) propõe um modelo de quantificação do risco de corrupção de servidores públicos, utilizando Regressão Logística e um conjunto de variáveis para atingir bons resultados preditivos. [Carvalho e Carvalho \(2016\)](#) aplica um modelo de Regressão Bayesiana na verificação do risco de corrupção de unidades administrativas governamentais no Brasil. Em seus estudos, os dados foram processados usando técnicas como análise de correlação e variância por classe. Além disso, foram comparados dois diferentes métodos de discretização: o Coeficiente de Contingência do Atributo Classe e o Princípio da Descrição do Tamanho Mínimo.

Existem outros trabalhos que estudam modelos econômicos relacionando-os com licitações e/ou contratos. Esses trabalhos têm abordado aspectos específicos dessas áreas, como formação de cartéis ou concorrência. No contexto dos contratos públicos no Brasil, alguns estudos já foram realizados com o objetivo de implementar modelos de predição de riscos envolvendo os fornecedores do governo. [Balaniuk et al. \(2012\)](#) utiliza o algoritmo *Naive Bayes* para uma abordagem não supervisionada que quantifica o risco conjunto de empresas e unidades governamentais na execução dos contratos.

Panigrahi et al. (2009) propõe uma abordagem para a detecção de fraudes em cartões de crédito, combinando evidências do comportamento atual e passado. Seu sistema de detecção de fraudes consiste de quatro componentes: o filtro baseado em regras, o somador *Dempster-Shafer*, o banco de dados de histórico de transações e o aprendizado Bayesiano. Esse aprendizado é utilizado para reforçar ou enfraquecer a crença nas transações de acordo com a sua similaridade com o histórico de transações fraudulentas. A fusão desses quatro componentes mostrou um impacto positivo elevado ao ser comparado com outros métodos.

Ngai et al. (2011) propõe um esquema de classificação utilizando a aplicação de técnicas de mineração de dados para a detecção de fraudes financeiras. Em seu trabalho foram analisados e classificados quatro categorias de fraudes financeiras (bancária, de seguros, de valores e outras fraudes financeiras relacionadas) e utilizadas seis classes de técnicas de mineração de dados (classificação, regressão, agrupamento, previsão, detecção de *outliers* e visualização). Ngai et al. (2011) fez uma revisão sistemática e constatou que as técnicas de minerações de dados têm sido aplicadas mais extensivamente na detecção de fraudes de seguros, embora os outros tipos de fraudes tenham atraído muita atenção nos últimos anos.

Librelotto e Mozzaquatro (2013) apresentou um estudo sobre as técnicas de mineração de dados aplicadas na identificação e classificação de indicadores de saúde, a fim de gerar um perfil de usuário. Em seu trabalho foi proposto um modelo que utiliza os algoritmos J48 e Apriori para comparar dados e medir a eficiência de ambos na classificação de informações em banco de dados.

Amin e Habib (2015) apresentou uma tese com o objetivo de mostrar a comparação de diferentes algoritmos de classificação utilizando o *WEKA* e descobrir qual algoritmo é o mais adequado para trabalhar com dados hematológicos. Em seu modelo, médicos e pacientes podem prever os dados hematológicos e o diagnóstico é dado por um aplicativo móvel. O algoritmo com melhores resultados foi o J48 com uma precisão de 97.16% na classificação dos dados e o algoritmo com menor erro médio foi o *naive bayes* com erro de 29.71%.

O presente trabalho tem como objetivo inicial realizar um estudo acerca do problema de classificação dos fornecedores da administração pública do Estado do Piauí. Esse estudo será responsável por desenvolver uma metodologia de classificação dos dados baseada em algoritmos de aprendizagem supervisionada e, com isso, desenvolver um sistema de avaliação e classificação de fornecedores em baixo risco (bons) e alto risco (ruins).

A Tabela 1 mostra um comparativo entre os trabalhos relacionados que utilizam aprendizagem de máquina. A proposta a ser utilizada assemelha-se com o proposto por (BALANIUK et al., 2012), divergindo quanto da utilização do algoritmo de aprendizagem de máquina, sendo o proposto, um algoritmo de aprendizagem supervisionada

baseado em árvore de decisão. Nota-se que grande parte da literatura utiliza algoritmos probabilísticos para verificação de riscos dentro da esfera Governamental. Em outros trabalhos, como em [Amin e Habib \(2015\)](#) e [Librelotto e Mozzaquatro \(2013\)](#), utilizou-se o J48 para classificação de dados voltada para a área da saúde e os resultados obtidos foram satisfatórios. [Ngai et al. \(2011\)](#) não foi relacionado na Tabela 1 por se tratar de uma revisão sistemática.

Tabela 1 – Comparativo dos trabalhos relacionados utilizando os algoritmos *naive bayes* (NB), J48 e *random forest* (RF)

	Verificação de Risco					Área da Saúde		Algoritmo de IA		
	UAG ⁴	CC ⁵	Contrato	Forn. ⁶	AC ⁷	CP ⁸	DH ⁹	NB	J48	RF
Panigrahi, 2009		X						X		
Balaniuk, 2012	X		X	X				X		
Librelotto, 2013						X			X	
Amin, 2015							X	X	X	
Devacena, 2015					X				X	X
Carvalho, 2016	X							X		
Presente Trabalho	X		X	X					X	

A utilização de árvore de decisão é justificada por alguns fatores, dentre eles a disposição dos dados presentes na base de dados (dados discretos e contínuos), a facilidade em interpretar os dados gerados através da visualização da árvore e o desempenho elevado com os testes realizados. A utilização do modelo proposto tem como objetivo analisar os fornecedores já conhecidos como fraudulentos e, através da semelhança com os novos fornecedores, evitar fraudes e auxiliar na economia dos recursos gastos com os procedimentos de auditorias. Ainda nesse estudo, a base de dados foi submetida a vários testes em sua forma original e subdividida em várias bases balanceadas.

A seguir temos materiais e métodos, onde é explicada toda a base de dados e justificada a escolha do melhor algoritmo para o problema de classificação proposto.

⁴ Unidade Administrativa Governamental

⁵ Cartão de Crédito

⁶ Fornecedor

⁷ Análise de Crédito

⁸ Criação de Perfil

⁹ Dados Hematológicos

3 Materiais e Métodos

Neste capítulo é discutida a base de dados utilizada no trabalho, bem como os detalhes de testes. Além disso, também é mostrado o modelo proposto no presente trabalho, composto por 6 etapas de desenvolvimento e suas respectivas explicações.

3.1 Base de Dados

A base de dados utilizada no presente trabalho (Figura 4) refere-se à identificação dos fornecedores presentes nos contratos firmados através de licitações no Estado do Piauí. Ela foi criada através da integração de outras bases de dados, como a base de CNPJ, presente na Receita Federal, a base Relação Anual de Informações Sociais (RAIS), presente no Ministério do Trabalho e na Previdência Social, além de dados provenientes do Tribunal Superior Eleitoral (TSE) e do Portal da Transparência do Governo Federal.

O contexto de sua aplicação é na identificação de fornecedores ajudando na classificação dos mesmos em alto ou baixo risco, de acordo com relatório de prestação de contas. A classificação em alto risco dar-se quando o fornecedor já foi citado em algum relatório de auditoria por envolvimento com fraude ou quebra contratual. Para os fornecedores de baixo risco não existem citações em relatórios de auditoria por motivos de problemas contratuais, ou seja, são fornecedores com credibilidade para fins contratuais.

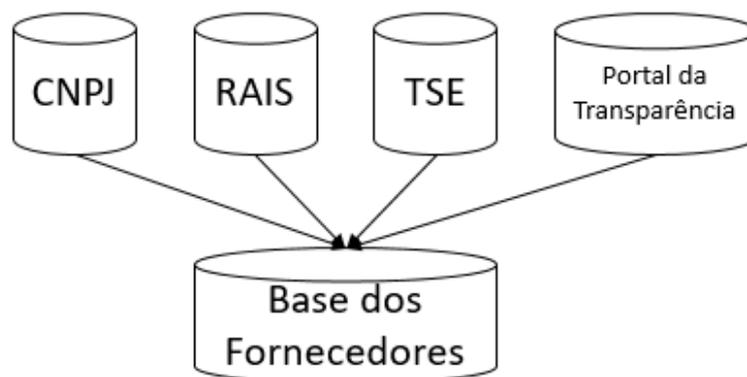


Figura 4 – Composição da base de dados dos fornecedores

A base de dados dos fornecedores é formada por 29.755 instâncias, cada uma representada por 46 atributos. No entanto, alguns dos atributos não foram povoados na base, reduzindo a quantidade de atributos para 22, sendo 21 numéricos e 1 não-numérico (Razão Social). Os atributos numéricos são listados abaixo e cada um de seus significados é fornecido por uma tabela de indicadores do TCE-PI:

1. **ID:** Identificador único da instância;
2. **CNPJ:** Número do CNPJ cadastrado ao fornecedor;
3. **ST_EPPME:** Variável do tipo *string* com valor binário que indica se o fornecedor é de pequeno porte ou microempresa (1) ou não (0);
4. **TSE_Valor_Doado_Total:** Valor total doado pela empresa nas campanhas de 2010, 2012, 2014 e 2016;
5. **Servidor_Qtd_Socio_Servidor:** Quantidade de sócios que também são servidores dos municípios e também do estado do Piauí;
6. **Servidor_Qtd_Funcionarios_Servidor:** Quantidade de funcionários que também são servidores dos municípios e também do estado do Piauí;
7. **RFB_Qtd_Socios_Media:** Quantidade média de sócios da empresa entre 2010 e 2017;
8. **RFB_Qtd_Socios_2017:** Quantidade de sócios da empresa em 2017;
9. **Pagto_2016:** Valor recebido pela empresa em 2016;
10. **Idade_Empresa_2017:** Idade da empresa em 2017;
11. **Rais_Qtd_Funcionarios_2017:** Quantidade de funcionários em 2017;
12. **ESFL:** Variável binária que indica se o fornecedor é Entidade Sem Fins Lucrativos (1) ou não (0);
13. **Convenios_Valor_Conveniente:** Valor recebido pela empresa na condição de conveniente em convênios;
14. **Compras_Soma_Valor_Homologado_Total:** Valor total em licitações vencidas pela empresa (entre 2010 e 2017);
15. **Compras_Soma_Valor_Homologado_2017:** Valor total em licitações vencidas pela empresa em 2017;
16. **Compras_Qtd_Entes_Media:** Quantidade de Entes (estado e municípios) distintos onde se localizam unidades governamentais para as quais a empresa vendeu (média entre 2010 e 2017);
17. **Compras_Qtd_Entes_2017:** Quantidade de Entes (estado e municípios) distintos onde se localizam unidades governamentais para as quais a empresa vendeu em 2017;

18. **Compras_Qtd_Orgaos_Media:** Quantidade de órgãos públicos para os quais a empresa vendeu, segregando-se Prefeituras e Câmaras (média entre 2010 e 2017);
19. **Compras_Qtd_Orgaos_2017:** Quantidade de órgãos públicos para os quais a empresa vendeu no ano de 2017, segregando-se Prefeituras e Câmaras;
20. **CNAE_Qtd_Total:** Quantidade de CNAEs (Classificação Nacional de Atividades Econômicas) secundários cadastrados pela empresa;
21. **Label:** Variável binária dependente do modelos. Identifica se um fornecedor é de alto risco (1) ou baixo risco (0), de acordo com o relatório de prestação de contas.

Inicialmente, alguns desses atributos merecem destaque. Por exemplo, o atributo *ST_EPPME* é responsável por dizer se um fornecedor é microempresa ou pequeno porte (recebendo valor 1) ou não (recebendo valor 0). Neste caso, quando o atributo recebe valor 1, indica que o seu faturamento não deve ultrapassar R\$ 360 mil para microempresas ou R\$ 3.6 milhões para empresas de pequeno porte. Fornecedores que possuem microempresas ou empresas de pequeno porte são fiscalizadas com menos rigor do que grandes empresas. Baseado nisso, espera-se uma atenção maior sobre esse tipo de empresa, pois elas são mais propensas a irregularidades.

Outro atributo que merece destaque é o *CNAE_Qtd_Total* que quantifica o número de CNAEs secundários cadastrados pela empresa. Em outras palavras, esse atributo diz em quantas atividades diferentes a empresa trabalha. Por exemplo, uma empresa que trabalhe no ramo de construção, transporte, alimentação e fornecimento de medicamentos, possui valor do atributo igual a 4. Normalmente, uma grande quantidade de CNAEs para uma empresa pode caracterizar envolvimento com fraudes.

Finalmente, o atributo classe (*Label*) que determina se um fornecedor é de alto risco (valor 1) ou é de baixo risco (valor 0) é fornecido pelo Tribunal de Contas do Estado do Piauí de acordo com a análise de relatórios de prestação de contas realizados anteriormente. Neste caso, para valor do atributo classe igual a 1, existem fraudes relacionadas a contratos envolvendo o fornecedor, por isso o mesmo é classificado como alto risco. Para o atributo classe com valor igual a 0, tem-se fornecedores confiáveis, justificando a classificação como baixo risco.

Em um segundo momento, a base foi acrescida de alguns atributos que não haviam sido povoados anteriormente. Os atributos acrescentados são todos numéricos e são observados a seguir:

1. **Pagto_Total:** Valor total recebido pela empresa entre 2010 e 2017;
2. **Rais_Pagto_Por_Funcionarios_2017:** Valor recebido do governo por funcionário na empresa (média entre 2010 e 2017);

3. **Rais_Qtd_Funcionarios_Media:** Quantidade média de sócios com ocupação de nível baixo entre 2010 e 2017;
4. **N_UG:** Número da unidade gestora;
5. **N_Modalidade:** Número da modalidade de contrato que a empresa participou;
6. **N_Microreg:** Número da microrregião a que a empresa pertence.

Outra alteração na base foi a exclusão do atributo *TSE_Valor_Doado_Total*, uma vez que a partir do ano de 2018, nenhuma empresa poderá fazer doações em campanhas eleitorais. Por fim, a base atualizada dos fornecedores é composta por 27 atributos.

3.2 Preparação da Base de Dados

A base de dados utilizada no trabalho é composta por 46 atributos e contém 29.755 instâncias, sendo *Label* o atributo classe. Para o contexto de aprendizagem de máquina, o atributo classe refere-se ao atributo que vai determinar se um fornecedor pertence ao grupo dos fornecedores de alto risco ou fornecedores de baixo risco. Inicialmente, 24 dos 46 atributos da base de dados eram não povoados, tornando-se necessário a sua exclusão junto a base, por se tratarem de atributos que não influenciariam de qualquer forma no processo de classificação. Após a remoção dos atributos não povoados, a base foi convertida para o formato *.arff* (utilizado na leitura do *WEKA*). Além disso, o atributo *ST_EPPME* é convertido de *string* para binário, pois o valor da variável é apenas 0 ou 1.

Uma importante etapa da preparação de dados é o balanceamento da base. O balanceamento acontece com o objetivo de equilibrar as duas classes em quantidade de instâncias. Para a base dos fornecedores (29.755 instâncias) tem-se 23.682 instâncias pertencentes à classe majoritária dos fornecedores de baixo risco e 6.073 instâncias pertencentes à classe minoritária dos fornecedores de alto risco. A relação existente na base é de aproximadamente 80% das instâncias pertencentes a uma classe e 20% pertencentes a outra classe, caracterizando um desbalanceamento. Com isso, a base original dos fornecedores (desbalanceada) foi subdividida em duas bases menores: a base dos fornecedores de baixo risco (*bons*, com valor do atributo classe igual a 0), contendo 23.682 instâncias e base dos fornecedores de alto risco (*ruins*, com valor do atributo classe igual a 1), contendo 6.073 instâncias.

Após a subdivisão em duas bases menores, formaram-se quatro bases de dados aleatórias através da união da base de dados dos fornecedores de alto risco com cada uma das quatro bases de dados dos fornecedores de baixo risco (BB1, BB2, BB3 e BB4), cada uma com 12.146 instâncias. A base BB1 foi formada através da união de 6.073 fornecedores de alto risco com 6.073 fornecedores de baixo risco. A base BB2 foi formada através da

união dos mesmos 6.073 fornecedores de alto risco com outros 6.073 fornecedores de baixo risco. A mesma regra vale para as bases BB3 e BB4, sendo a base BB4 formada pelos 6.073 fornecedores de alto risco e os 5.463 fornecedores de baixo risco restantes. Os 610 fornecedores de baixo risco que faltaram para completar o balanceamento exato das classes foram fornecidos pelo terceiro fragmento dos 6.073 fornecedores de baixo risco.

Foi utilizado um filtro de ganho de informação para a base desbalanceada e para as quatro bases balanceadas. O filtro de ganho de informação seleciona cada um dos atributos pertencentes à base de dados e quantifica a importância desse atributo em relação ao atributo classe escolhido. Com isso, o filtro de ganho de informação seleciona, ranqueia e avalia cada um dos atributos pertencentes às bases de dados em relação ao atributo *Label*. Para os *rankings* formados, notou-se uma diferença muito grande de importância entre os dois primeiros colocados e os demais atributos. Assim, os atributos mais relevantes para o processo de classificação são: *Idade_Empresa_2017* e *Compras_Soma_Valor_Homologado_Total*.

Além da análise com o filtro de ganho de informação, foram selecionados alguns atributos, através de uma análise feita por especialistas (auditores de controle externo junto ao TCE-PI). Essa análise foi realizada com base nos conhecimentos do problema e de acordo com o grau de importância de cada atributo dentro do processo de auditoria. Assim, os atributos selecionados foram: *Compras_Qtd_Entes_Media*, *Pagto_2016*, *ESFL*, *Rais_Qtd_Funcionarios_2017*, *ST_EPPME* e *CNAE_Qtd_Total*.

Com isso, para cada uma das bases de dados foi utilizada a seleção dos atributos obtida através da utilização do filtro do ganho de informação mais os atributos selecionados pelos especialistas mais o atributo classe, totalizando 9 atributos. Por fim, os dados presentes nas bases de dados são submetidos à normalização, visto que grande parte dos dados encontra-se em escala de milhares ou milhões, enquanto alguns deles possuem valores binários ou decimais.

3.3 Detalhes dos Testes com o WEKA

Os experimentos apresentados neste capítulo foram realizados através de testes com a ferramenta *WEKA*. A utilização dessa ferramenta permitiu a extração de informações que possibilitaram a escolha do algoritmo J48 na linguagem de programação JAVA com suporte da ferramenta ECLIPSE, que possibilita a utilização de diversos algoritmos de aprendizagem supervisionada e não-supervisionada. Além disso, a linguagem JAVA possui boa integração com a biblioteca do *WEKA*, utilizada na implementação do algoritmo de classificação.

A escolha do algoritmo J48 como favorito à implementação, deu-se por diversos fatores, dentre eles os testes utilizando a ferramenta *WEKA*. O *WEKA* conta com um módulo chamado *AUTOWEKA*, onde através dele é possível escolher uma base de dados e

definir uma série de parâmetros de execução que serão utilizados em sucessivos testes na busca de um melhor algoritmo e os seus valores de parâmetros para resolver um problema de classificação.

Um dos parâmetros definidos no *AUTOWEKA* é a métrica de busca que é responsável por, durante os sucessivos testes, encontrar o melhor algoritmo que otimize o valor da métrica escolhida. Essa métrica pode ser a *precision*, o *recall*, o *f-measure*, o *recall* da classe negativa, dentre outros. Um exemplo de uso do *AUTOWEKA* pode ser observado no presente trabalho, onde ao selecionar a base de dados, deseja-se encontrar o melhor algoritmo de classificação através da manipulação das métricas *precision*, *recall*, *f-measure* e *recall* da classe negativa.

Os testes realizados com as bases de dados e a ferramenta *WEKA* são testes de avaliação dos algoritmos, pois ao final deles, será escolhido o algoritmo mais adequado ao problema de classificação e à base de dados dos fornecedores. Esses testes foram divididos em quatro etapas e são apresentados na Tabela 2:

1. Inicialmente foram realizados testes com a base desbalanceada;
2. A segunda etapa é definida por testes com as bases balanceadas BB1, BB2, BB3 e BB4;
3. Na terceira etapa são realizados testes com as bases balanceadas BB1, BB2, BB3 e BB4 utilizando seleção de atributos definidos pelo filtro de ganho de informação e pela análise dos especialistas.
4. Por fim, são realizados testes com as bases balanceadas BB1, BB2, BB3 e BB4 utilizando seleção de atributo e aplicando normalização nos dados.

Tabela 2 – Relação entre testes com as bases de dados e a métrica de avaliação dos algoritmos

Base de Dados	Métrica de Avaliação			
	Precision	Recall	F-Measure	<i>Recall</i> da Classe Negativa
Base Desbalanceada	T1	T2	T3	-
Base Balanceada BB1	T4 / T20	T5 / T21	T6 / T22	T7 / T23 / T36
Base Balanceada BB2	T8 / T24	T9 / T25	T10 / T26	T11 / T27 / T37
Base Balanceada BB3	T12 / T28	T13 / T29	T14 / T30	T15 / T31 / T38
Base Balanceada BB4	T16 / T32	T17 / T33	T18 / T34	T19 / T 35 / T39

Para a Tabela 2 tem-se cada um dos 39 testes realizados. Para cada um dos testes é feita uma relação entre cada uma das bases de dados utilizadas e a métrica de avaliação dos algoritmos. Os testes T1, T2 e T3 referem-se aos testes realizados com a base de dados

desbalanceada buscando encontrar o melhor algoritmo baseado na otimização da métrica *precision*, *recall* e *f-measure*, respectivamente.

Ainda na Tabela 2 os testes de T4 até T19 referem-se aos testes realizados com as bases de dados balanceadas (BB1, BB2, BB3 e BB4) e buscam encontrar o melhor algoritmo baseado na otimização das métricas *precision*, *recall*, *f-measure* e *recall* da classe negativa. Para esses testes, foi incluída a análise da métrica *recall* da classe negativa pela necessidade de aumentar a sensibilidade da classe negativa, ou seja, aumentar a frequência em que o classificador encontra os exemplos de uma classe. Nesse caso, o classificador tenta priorizar os acertos da classe referente aos fornecedores de alto risco (ruins).

Os testes de T20 até T35 referem-se aos testes realizados com as bases de dados balanceadas (BB1, BB2, BB3 e BB4) e seleção de atributos por filtro de ganho de informação e análise de especialistas. Eles buscam encontrar o melhor algoritmo baseado na otimização das métricas *precision*, *recall*, *f-measure* e *recall* da classe negativa. Finalmente, os testes T36, T37, T38 e T39 referem-se aos testes com as bases de dados balanceadas BB1, BB2, BB3 e BB4, respectivamente, seleção de atributo e normalização de dados. Esses testes levaram em consideração apenas a métrica *recall* da classe negativa, pois o objetivo geral do trabalho é acertar os fornecedores considerados de alto risco (ruins).

3.4 Modelo Proposto

A metodologia adotada no presente trabalho fundamentou-se em cinco etapas:

1. Estudo do problema, do processo de análise e controle de contas, dos trabalhos relacionados e dos algoritmos candidatos a serem implementados;
2. Preparação dos dados;
3. Testes com as bases de dados e identificação dos principais algoritmos para o processo de classificação;
4. Avaliação dos resultados;
5. Implementação do sistema.

As etapas dois e três são mostradas através de um fluxograma, que representa a metodologia proposta (Figura 5).

A primeira etapa caracterizou-se pelo estudo do problema de racionalizar os recursos de auditorias junto ao TCE-PI e pelo entendimento no que diz respeito ao processo de análise e prestação de contas. Além disso, incluiu-se uma análise acerca dos trabalhos já existentes na área e o estudo de algoritmos candidatos à resolução do problema de

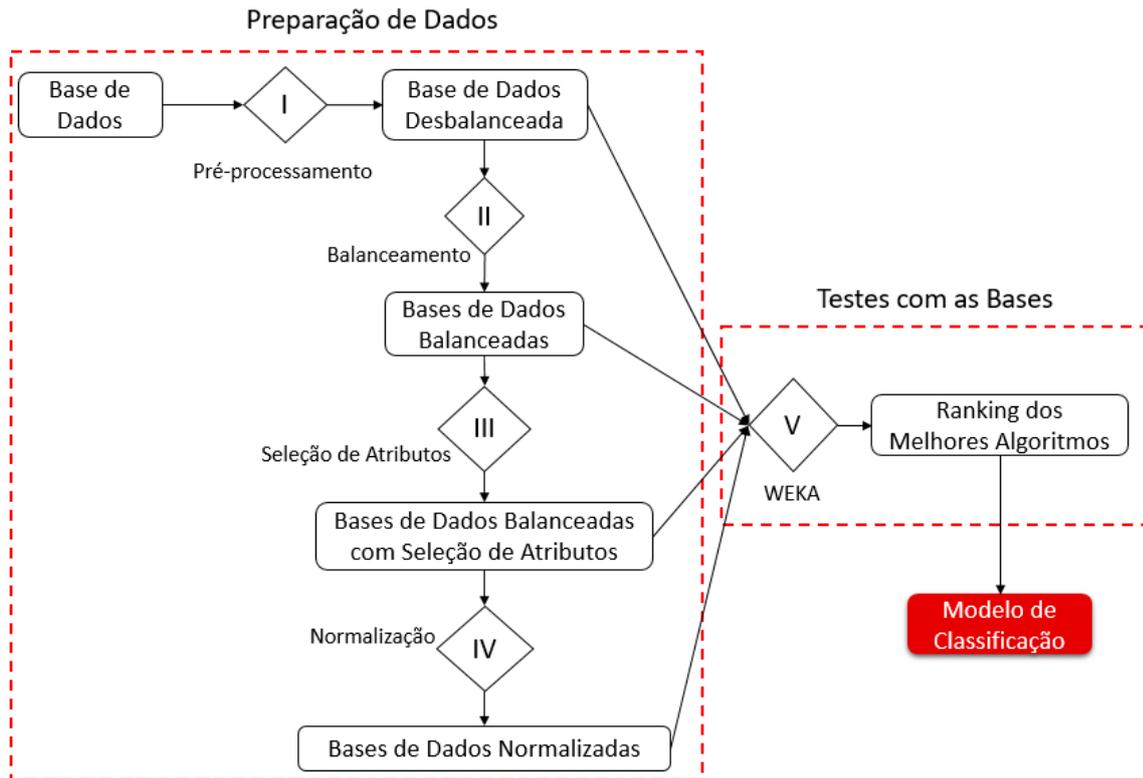


Figura 5 – Modelo proposto

classificação (Redes Neurais Artificiais, Árvores de Decisão, *Naive Bayes*). A escolha desses algoritmos como candidatos ao processo de classificação deu-se por serem amplamente utilizados na literatura e na aplicação da inteligência artificial dentro da resolução de problemas. O estudo dos algoritmos candidatos serviu para identificar o melhor uso dos parâmetros a serem utilizados por cada um deles.

De acordo com a análise da Figura 5, a segunda etapa caracterizou-se como a etapa de pré-processamento ou preparação de dados (I), onde foram feitos ajustes à(s) base(s). Nessa etapa, a base de dados foi disponibilizada pelo Tribunal de Contas do Estado do Piauí. Inicialmente, a base formada por aproximadamente 30 mil instâncias e 46 atributos foi submetida a procedimentos de exclusão de atributos e conversão de tipos. Alguns atributos foram excluídos por possuírem valores *NULL* e outros por se relacionarem à situações de mudança dentro da esfera governamental. Um exemplo de atributo relacionado à mudanças é o *TSE_Vvalor_Doador_Total* que, a partir de 2018, deixou de ser utilizado, uma vez que empresas não poderão mais fazer doações a campanhas eleitorais. As conversões de tipos referem-se ao atributo *ST_EPPME*, onde seu tipo inicial era *string* e passou a ser binário, e à base de dados dos fornecedores que foi convertida para o formato de leitura *.arff* — formato de execução para a ferramenta *WEKA*.

Ainda nessa etapa foram realizados os procedimentos de balanceamento (II) gerando quatro novas bases de dados balanceadas. O balanceamento da base de dados dos fornecedores deu-se com o objetivo de equilibrar a quantidade de instâncias pertencentes a

cada uma das classes. Para a base de dados original, contendo aproximadamente 30 mil instâncias, tem-se a relação 80% dos dados pertencentes à classe majoritária — fornecedores de baixo risco — e 20% dos dados pertencentes à classe minoritária — fornecedores de alto risco. Dessa forma, o balanceamento permitiu a subdivisão da base de dados original em quatro bases de dados menores e balanceadas: BB1, BB2, BB3 e BB4. Para cada uma das bases de dados balanceadas, tem-se um total de 6073 instâncias classificadas como fornecedores de baixo risco e 6073 instâncias classificadas como fornecedores de alto risco, mantendo a relação 50% dos dados para cada uma das classes predefinidas.

Além do balanceamento foi realizada a seleção de atributos (III), onde foram selecionados atributos com base no filtro de ganho de informação e na análise dos especialistas do problema. O filtro de ganho de informação utiliza um cálculo para quantificar a importância de um atributo em relação ao atributo classe *Label*. Essa importância é calculada para todos os atributos e, ao final, é gerado um *ranking* que classifica cada um dos atributos em relação ao atributo classe. Para esse *ranking* são selecionados os dois atributos mais relevantes, uma vez que a partir do terceiro atributo melhor colocado, os valores de importância decaem bastante em relação aos dois primeiros colocados. Também foi realizada uma análise preditiva com especialistas do problema. Essa análise foi realizada de acordo com o conhecimento do problema e com o grau de importância de cada atributo dentro do processo de auditoria por parte dos auditores. A análise dos especialistas permitiu selecionar oito atributos, sendo dois deles já selecionados com o filtro de ganho de informação. Com isso, ao final da seleção de atributos determinou-se a utilização de nove atributos para a classificação da base de dados dos fornecedores, sendo oito escolhidos pelo filtro de ganho de informação e análise de especialistas mais o atributo classe *Label*.

Finalmente, após a seleção de atributos, aconteceu a normalização de dados (IV). A importância da fase de normalização dar-se pela disparidade de escala dos dados relacionados nas bases de dados. Alguns atributos, como *CNAE_Qtd_Total* possuem amplitude de valores entre 0 e 10. Para outros atributos, como *Pagto_2016* a amplitude de valores está entre 0 e 4.130.267.649,08. Como a classificação busca encontrar características em comum que tornem possível a comparação e relação de duas instâncias, existe a necessidade de manter os dados estruturados, preferencialmente utilizando a mesma escala. Com isso, as bases com balanceamento dos dados e seleção de atributos foram submetidas à normalização dos dados, onde os dados com escalas em binário, decimal, milhares e milhão, foram colocados em uma única escala.

Ainda de acordo com a Figura 5, a terceira etapa caracterizou-se como a etapa de testes utilizando a ferramenta de mineração de dados. Nessa etapa a base desbalanceada, as bases balanceadas, as bases balanceadas com seleção de atributos e as bases balanceadas com os dados normalizados foram submetidos a testes com a ferramenta *WEKA* (V) alternando as métricas de avaliação dos algoritmos (*precision*, *recall*, *fMeasure* e *recall*

da classe negativa). Os testes realizados com auxílio da ferramenta *WEKA* permitiram emitir um relatório com as informações detalhadas desses testes, facilitando a análise comparativa dos algoritmos apontados como candidatos à implementação.

A quarta etapa (apresentada no capítulo 4) refere-se à avaliação dos resultados, onde são discutidos os resultados de cada um dos testes para cada uma das bases e suas variações. Finalmente, a última etapa caracterizou-se pela implementação do sistema de avaliação de fornecedores (SAF) utilizando o algoritmo J48. Nessa etapa, utilizou-se a linguagem de programação *PYTHON* integrada com a biblioteca do *WEKA*. A justificativa da utilização do algoritmo J48 deu-se por apresentar melhores resultados de classificação (discutidos no capítulo 4) para o problema proposto. Além disso, a frequência com a qual o algoritmo J48 é destacada como melhores soluções para o problema, contaram como a escolha do mesmo.

4 Resultados e Discussão

Neste capítulo, são discutidos os resultados obtidos conforme os testes e a implementação do modelo discutido no Capítulo 3. Os resultados foram obtidos a partir da base de dados dos fornecedores do TCE-PI com modificações de balanceamento e seleção de atributos. Além disso, são discutidos os detalhes de implementação, as telas e o funcionamento do sistema, produto da pesquisa.

4.1 Análise dos Resultados

A análise dos resultados aconteceu com base nos testes realizados com auxílio da ferramenta de mineração de dados *WEKA*. Essa ferramenta possui um módulo chamado *AUTOWEKA* que permite executar múltiplos testes com a base alterando-se parâmetros como tempo de execução, métrica de avaliação do algoritmo, quantidade de memória RAM utilizada nos testes e quantidade de configurações (algoritmos + parâmetros e seus valores) esperadas como solução para o processo de classificação.

Para todos os testes realizados, foram utilizados os seguintes parâmetros no módulo do *AUTOWEKA*:

- tempo de execução fixo em 1440min (24 horas). A não utilização de um tempo de execução maior, deu-se por a maioria dos testes alcançarem a melhor solução antes das 18 horas de execução. Com isso, justifica-se o tempo escolhido;
- utilização de 2 Gigabytes de memória RAM dedicada ao processo de execução dos testes. A quantidade de memória RAM utilizada é limitada à capacidade da ferramenta. Em alguns casos foi utilizada uma quantidade maior de memória RAM. Para eles, os testes travaram e a execução não foi concluída;
- busca das 10 melhores configurações (algoritmos + parâmetros e seus valores) candidatas.

As Tabelas 3, 5, 7, 9, 11, 13, 15, 17, 19 e 21 apresentam os valores obtidos para a quantidade de instâncias corretamente classificadas, a quantidade de instâncias incorretamente classificadas e o melhor algoritmo encontrado para trabalhar com a base em relação às métricas de avaliação.

As Tabelas 4, 6, 8, 10, 12, 14, 16, 18, 20 e 22 apresentam as matrizes de confusão relacionadas a cada um dos testes realizados. Para elas é dada a quantidade de elementos

classificados como verdadeiros positivos, falsos positivos, falsos negativos e verdadeiros negativos. As tabelas e os resultados são discutidos nas subseções a seguir.

4.1.1 Testes com a Base de Dados Desbalanceada

Para os testes realizados com a base de dados desbalanceada, foram utilizadas como variações nas métricas de avaliação do algoritmo: *precision*, *recall* e *fMeasure*. Os resultados são apresentados na Tabela 3.

Tabela 3 – Resultados dos testes no *WEKA* com a base de dados desbalanceada

Teste Realizado	Instâncias Classificadas Corretamente	Instâncias Classificadas Incorretamente	Melhor Algoritmo
T1	23704 (79.67%)	6049 (20.33%)	<i>Logistic Model Tree</i> (Árvore de Decisão)
T2	23680 (79.59%)	6073 (20.41%)	<i>Multilayer Peceptron</i>
T3	25811 (86.75%)	3942 (13.25%)	<i>Random Forest</i> (Árvore de Decisão)

O teste T1, tendo como entrada a base desbalanceada, sugere a utilização de um algoritmo baseado em árvore de decisão (*logistic model tree*). Um modelo que implemente esse algoritmo é capaz de acertar 23.704 instâncias em um universo de 29.753, e errar apenas 6.049. O resultado obtido com o teste T2 assemelhasse em taxa de acerto ao teste T1. Para T2 a sugestão de algoritmo é o *multilayer perceptron*, com 23.680 instâncias classificadas corretamente e 6.073 classificadas incorretamente. Para o teste T3 a sugestão de algoritmo também é baseada em árvore de decisão (*random forest*), onde é possível acertar 25.811 instâncias e errar 3.942 instâncias.

Inicialmente tem-se boas taxas de acerto para um possível classificador. No entanto ao analisar a matriz de confusão para cada um dos três testes, observa-se que as instâncias corretamente classificadas referem-se basicamente aos fornecedores de baixo risco (positivos) e a maior parte das instâncias classificadas incorretamente referem-se aos fornecedores de alto risco (negativos).

Além disso, a análise das matrizes de confusão presentes na Tabela 4 permite entender, com exemplos reais, o significado dos termos verdadeiro positivo (dado pela intersecção da linha *true* com a coluna *true*), falso positivo (dado pela intersecção da linha *true* com a coluna *negative*), falso negativo (dado pela intersecção da linha *negative* com a coluna *true*) e verdadeiro negativo (dado pela intersecção da linha *negative* com a coluna *negative*).

Como o problema de classificação em questão tem por objetivo classificar corretamente o máximo de fornecedores de alto risco (verdadeiros negativos na matriz de confusão), os resultados apresentados na Tabela 3 não são satisfatórios.

Tabela 4 – Matrizes de confusão para os testes com a base de dados desbalanceada

	T1		T2		T3	
	<i>True</i>	<i>Negative</i>	<i>True</i>	<i>Negative</i>	<i>True</i>	<i>Negative</i>
<i>True</i>	23680	0	23680	0	23675	5
<i>Negative</i>	6049	24	6073	0	3937	2136

4.1.2 Testes com as Bases de Dados Balanceadas

A base desbalanceada original consiste em 29.753 instâncias, sendo 23.680 pertencentes à classe dos fornecedores de baixo risco e 6.073 pertencentes à classe dos fornecedores de alto risco. Após a análise dos testes realizados na base desbalanceada, surge como solução para melhorar os resultados do classificador, um processo de balanceamento da base.

A maneira encontrada para realizar o balanceamento foi criar quatro bases de dados aleatórias distintas (BB1, BB2, BB3 e BB4), mantendo a proporção de 6.073 fornecedores de alto risco para 6.073 fornecedores de baixo risco. Assim, foram realizados testes com as quatro bases de dados balanceadas, mantendo as mesmas características da base de dados original. Para os testes realizados com as bases de dados balanceadas, foram utilizadas como variações nas métricas de avaliação do algoritmo: *precision*, *recall*, *fMeasure* e *recall* da classe negativa. Os resultados são apresentados nas subseções a seguir.

4.1.2.1 Base de Dados Balanceada BB1

Os testes realizados com a base de dados balanceada BB1 são T4, T5, T6 e T7. A análise da Tabela 5 permite observar que o teste T6 apresenta os melhores resultados obtidos com classificação correta de 9.469 instâncias (aproximadamente 78%) e classificação incorreta de 2.677 instâncias. O teste T4 apresenta os piores resultados com classificação correta de 6.112 instâncias e classificação incorreta de 6.034 instâncias. Para T6 a sugestão de melhor algoritmo para o problema de classificação é baseada em árvore de decisão (*logistic model tree*) e para T4 a sugestão de melhor algoritmo é o *naive bayes*.

Tabela 5 – Resultados dos testes no *WEKA* com a base de dados balanceada BB1

Teste Realizado	Instâncias Classificadas Corretamente	Instâncias Classificadas Incorretamente	Melhor Algoritmo
T4	6112 (50.32%)	6034 (49.68%)	<i>Naive Bayes</i>
T5	8670 (71.38%)	3476 (28.12%)	<i>Multilayer Peceptron</i>
T6	9469 (77.96%)	2677 (22.04%)	<i>Logistic Model Tree</i> (Árvore de Decisão)
T7	9215 (75.87%)	2931 (24.13%)	J48 (Árvore de Decisão)

Ainda analisando a Tabela com os resultados dos testes T4, T5, T6 e T7 pode-se observar que, para os testes com a base de dados balanceada BB1, existe uma predominância em utilizar algoritmos baseados em árvore de decisão (2 dos 4 casos de testes) para o problema de classificação dos fornecedores do Estado do Piauí.

Comparando os resultados obtidos através dos testes da base de dados balanceada BB1 com os testes da base de dados desbalanceada, nota-se que a taxa de acerto geral foi reduzida. A análise das matrizes de confusão para cada um dos quatro testes (Tabela 6) mostra que essa redução aconteceu em quantidade, no entanto houve uma melhoria na qualidade dos dados buscados. É possível perceber que com exceção do teste T4, todas os outros obtiveram uma taxa de acerto superior a 83% em relação aos verdadeiros negativos.

Tabela 6 – Matrizes de confusão para os testes com a base de dados balanceada BB1

	T4		T5		T6		T7	
	<i>True</i>	<i>Neg.</i>	<i>True</i>	<i>Neg.</i>	<i>True</i>	<i>Neg.</i>	<i>True</i>	<i>Neg.</i>
<i>True</i>	6066	7	3481	2592	4411	1662	4094	1979
<i>Neg.</i>	6062	46	884	5189	1015	5058	952	5121

Ainda analisando a Tabela com as matrizes de confusão para os testes com a base de dados balanceada BB1, observa-se que o teste T6 também apresenta boa taxa de acerto referente aos verdadeiros negativos (fornecedores de alto risco que são, de fato, classificados como alto risco ou ruins), com 5.058 acertos para 6.073 instâncias (83.3% de acerto dos verdadeiros negativos).

4.1.2.2 Base de Dados Balanceada BB2

Os testes realizados com a base de dados balanceada BB2 são T8, T9, T10 e T11. A análise da Tabela 7 referente a esses testes permite observar que o teste T11 apresenta os melhores resultados obtidos com classificação correta de 9.344 instâncias (aproximadamente 77%) e classificação incorreta de 2.802 instâncias. O teste T8 apresenta os piores resultados com classificação correta de 6.113 instâncias e classificação incorreta de 6.033 instâncias. Para T11 a sugestão de melhor algoritmo para o problema de classificação é baseada em árvore de decisão (J48) e para T8 a sugestão de melhor algoritmo é o *naive bayes*.

Ainda analisando a Tabela com os resultados dos testes T8, T9, T10 e T11 pode-se observar que, para os testes com a base de dados balanceada BB2, existe uma predominância em utilizar algoritmos baseados em árvore de decisão (3 dos 4 casos de testes) para o problema de classificação dos fornecedores do Estado do Piauí.

Comparando os resultados obtidos através dos testes da base de dados balanceada BB2 com os testes da base de dados desbalanceada, nota-se que a taxa de acerto geral também foi reduzida. A análise das matrizes de confusão para cada um dos quatro testes

Tabela 7 – Resultados dos testes no *WEKA* com a base de dados balanceada BB2

Teste Realizado	Instâncias Classificadas Corretamente	Instâncias Classificadas Incorretamente	Melhor Algoritmo
T8	6113 (50.32%)	6033 (49.68%)	<i>Naive Bayes</i>
T9	9129 (75.17%)	3017 (24.83%)	J48 (Árvore de Decisão)
T10	8746 (71.01%)	3400 (27.97%)	J48 (Árvore de Decisão)
T11	9344 (76.99%)	2802 (23.01%)	J48 (Árvore de Decisão)

(Tabela 8) mostra que essa redução aconteceu em quantidade, no entanto houve uma melhoria na qualidade dos dados buscados. É possível perceber que com exceção do teste T8, todas os outros obtiveram uma taxa de acerto superior a 83% em relação aos verdadeiros negativos.

Tabela 8 – Matrizes de confusão para os testes com a base de dados balanceada BB2

	T8		T9		T10		T11	
	<i>True</i>	<i>Neg.</i>	<i>True</i>	<i>Neg.</i>	<i>True</i>	<i>Neg.</i>	<i>True</i>	<i>Neg.</i>
<i>True</i>	6066	7	4086	1986	3382	2690	4120	1953
<i>Neg.</i>	6061	47	1031	5043	709	5365	849	5224

Ainda analisando a Tabela com as matrizes de confusão para os testes com a base de dados balanceada BB2, observa-se que o teste T11 também apresenta boa taxa de acerto referente aos verdadeiros negativos (fornecedores de alto risco que são, de fato, classificados como alto risco ou ruins), com 5.224 acertos para 6.073 instâncias (86% de acerto dos verdadeiros negativos).

4.1.2.3 Base de Dados Balanceada BB3

Os testes realizados com a base de dados balanceada BB3 são T12, T13, T14 e T15. A análise da Tabela 9 referente a esses testes permite observar que o teste T15 apresenta os melhores resultados obtidos com classificação correta de 9.407 instâncias (aproximadamente 77%) e classificação incorreta de 2.739 instâncias. O teste T12 apresenta os piores resultados com classificação correta de 6.437 instâncias e classificação incorreta de 5.709 instâncias. Para T15 a sugestão de melhor algoritmo para o problema de classificação é baseada em árvore de decisão (J48) e para T12 a sugestão de melhor algoritmo é o *naive bayes*.

Ainda analisando a Tabela com os resultados dos testes T12, T13, T14 e T15 pode-se observar que, embora o melhor resultado seja baseado na utilização de árvore de decisão, para os testes com a base de dados balanceada BB3, existe uma predominância

Tabela 9 – Resultados dos testes no *WEKA* com a base de dados balanceada BB3

Teste Realizado	Instâncias Classificadas Corretamente	Instâncias Classificadas Incorretamente	Melhor Algoritmo
T12	6437 (53.99%)	5709 (47.01%)	<i>Naive Bayes</i>
T13	8718 (71.78%)	3428 (28.22%)	<i>Multilayer Perceptron</i>
T14	9198 (75.73%)	2948 (24.27%)	<i>Multilayer Perceptron</i>
T15	9407 (77.45%)	2739 (22.55%)	J48 (Árvore de Decisão)

em utilizar algoritmos baseados em redes neurais (2 dos 4 casos de testes) para o problema de classificação dos fornecedores do Estado do Piauí.

Comparando os resultados obtidos através dos testes da base de dados balanceada BB3 com os testes da base de dados desbalanceada, nota-se que a taxa de acerto geral foi reduzida. A análise das matrizes de confusão para cada um dos quatro testes (Tabela 10) mostra que essa redução aconteceu em quantidade, no entanto houve uma melhoria na qualidade dos dados buscados. É possível perceber que com exceção do teste T12, todas os outros obtiveram uma taxa de acerto superior a 82% em relação aos verdadeiros negativos.

Tabela 10 – Matrizes de confusão para os testes com a base de dados balanceada BB3

	T12		T13		T14		T15	
	<i>True</i>	<i>Neg.</i>	<i>True</i>	<i>Neg.</i>	<i>True</i>	<i>Neg.</i>	<i>True</i>	<i>Neg.</i>
<i>True</i>	6066	7	3540	2533	4164	1909	3940	2133
<i>Neg.</i>	5702	371	895	5178	1039	5034	606	5467

Ainda analisando a Tabela com as matrizes de confusão para os testes com a base de dados balanceada BB3, observa-se que o teste T15 também apresenta boa taxa de acerto referente aos verdadeiros negativos (fornecedores de alto risco que são, de fato, classificados como alto risco ou ruins), com 5.467 acertos para 6.073 instâncias (90% de acerto dos verdadeiros negativos).

4.1.2.4 Base de Dados Balanceada BB4

Os testes realizados com a base de dados balanceada BB4 são T16, T17, T18 e T19. A análise da Tabela 11 referente a esses testes permite observar que o teste T19 apresenta os melhores resultados obtidos com classificação correta de 9.376 instâncias (aproximadamente 77%) e classificação incorreta de 2.770 instâncias. O teste T16 apresenta os piores resultados com classificação correta de 6.433 instâncias e classificação incorreta de 5.713 instâncias. Para T19 a sugestão de melhor algoritmo para o problema de classificação é baseada em árvore de decisão (J48) e para T16 a sugestão de melhor algoritmo é o *naive bayes*.

Tabela 11 – Resultados dos testes no *WEKA* com a base de dados balanceada BB4

Teste Realizado	Instâncias Classificadas Corretamente	Instâncias Classificadas Incorretamente	Melhor Algoritmo
T16	6433 (53.99%)	5713 (47.01%)	<i>Naive Bayes</i>
T17	8831 (72.71%)	3315 (27.29%)	J48 (Árvore de Decisão)
T18	8783 (72.11%)	3363 (27.69%)	J48 (Árvore de Decisão)
T19	9376 (77.19%)	2770 (22.81%)	J48 (Árvore de Decisão)

Ainda analisando a Tabela com os resultados dos testes T16, T17, T18 e T19 pode-se observar que, para os testes com a base de dados balanceada BB4, existe uma predominância em utilizar algoritmos baseados em árvore de decisão (3 dos 4 casos de testes) para o problema de classificação dos fornecedores do Estado do Piauí.

Comparando os resultados obtidos através dos testes da base de dados balanceada BB4 com os testes da base de dados desbalanceada, nota-se que a taxa de acerto geral novamente foi reduzida. A análise das matrizes de confusão para cada um dos quatro testes (Tabela 12) mostra que essa redução aconteceu em quantidade, no entanto houve uma melhoria na qualidade dos dados buscados. É possível perceber que com exceção do teste T16, todos os outros obtiveram uma taxa de acerto superior a 86% em relação aos verdadeiros negativos.

Tabela 12 – Matrizes de confusão para os testes com a base de dados balanceada BB4

	T16		T17		T18		T19	
	<i>True</i>	<i>Neg.</i>	<i>True</i>	<i>Neg.</i>	<i>True</i>	<i>Neg.</i>	<i>True</i>	<i>Neg.</i>
<i>True</i>	6062	7	3568	2505	3230	2843	3718	2315
<i>Neg.</i>	5706	371	810	5263	520	5553	415	5658

Ainda analisando a Tabela com as matrizes de confusão para os testes com a base de dados balanceada BB4, observa-se que o teste T19 também apresenta boa taxa de acerto referente aos verdadeiros negativos (fornecedores de alto risco que são, de fato, classificados como alto risco ou ruins), com 5.658 acertos para 6.073 instâncias (93% de acerto dos verdadeiros negativos).

A Figura 6 representa uma comparação dos resultados obtidos através dos testes realizados com as bases de dados balanceadas BB1, BB2, BB3 e BB4. Nela, observa-se que os testes T4, T8, T12 e T16 não apresentam bons resultados, visto que a quantidade de instâncias classificadas de maneira correta e incorreta são bem semelhantes. Além disso, também é possível verificar a qualidade dos testes T6, T11, T15 e T19 (discutidos na análise de cada base de dados balanceada individualmente).

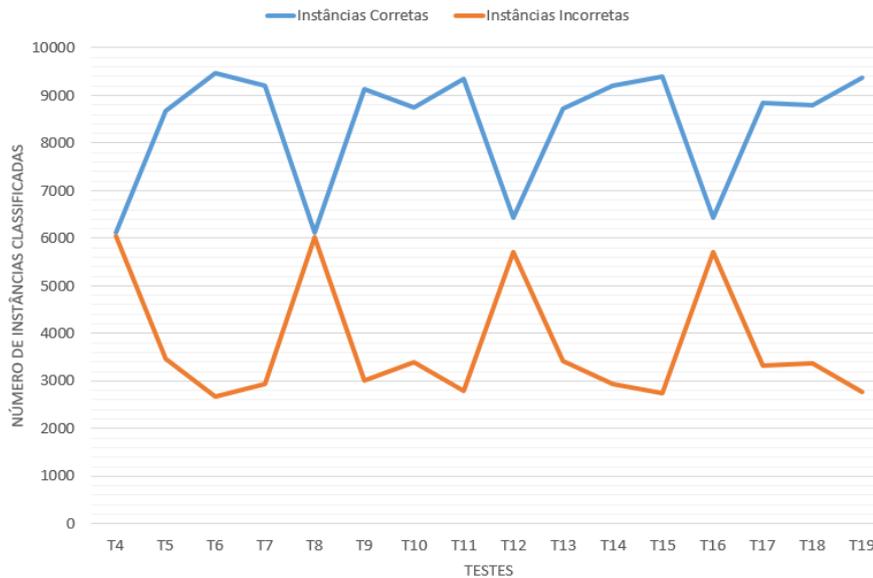


Figura 6 – Gráfico comparativo entre os testes com as bases de dados balanceadas

4.1.3 Testes com as Bases de Dados Balanceadas e Seleção de Atributos

Foi realizada uma terceira etapa de testes através da análise das bases balanceadas (BB1, BB2, BB3 e BB4) com filtros de seleção de atributos. Ao rodar um filtro para ranquear os atributos, tem-se uma diferença muito grande de importância entre os dois primeiros colocados e os demais atributos. Sendo assim, os dois atributos que obrigatoriamente devem fazer parte do classificador são: *Idade_Empresa_2017* e *Compras_Soma_Valor_Homologado_Total*.

Por se tratar de uma quantidade pequena de atributos para a avaliação junto ao classificador, decidiu-se por considerar uma análise feita por especialistas (auditores de controle externo junto ao TCE-PI). Com isso, também foram considerados os seguintes atributos, em ordem decrescente de ranqueamento: *Compras_Qtd_Entes_Media*, *Pagto_2016*, *ESFL*, *ST_EPPME*, *Rais_Qtd_Funcionarios_2017* e *CNAE_Qtd_Total*. Além dos 8 atributos considerados, utilizou-se também o atributo classe *Label*.

Para os testes realizados com as bases de dados balanceadas e seleção de atributos, foram utilizadas como variações nas métricas de avaliação do algoritmo: *precision*, *recall*, *fMeasure* e *recall* da classe negativa. Os resultados são apresentados nas subseções a seguir.

4.1.3.1 Base de Dados Balanceada BB1 e Seleção de Atributos

Os testes realizados com a base de dados balanceada BB1 são T20, T21, T22 e T23. A análise da Tabela 13 referente a esses testes permite observar que o teste T23 apresenta os melhores resultados obtidos com classificação correta de 11.086 instâncias (aproximadamente 91%) e classificação incorreta de 1.060 instâncias. O teste T21 apresenta os piores resultados com classificação correta de 10.657 instâncias e classificação incorreta de

1.489 instâncias. Para T23 a sugestão de melhor algoritmo para o problema de classificação é baseada em árvore de decisão (J48) e para T21 a sugestão de melhor algoritmo também é baseada em árvore de decisão (J48).

Tabela 13 – Resultados dos testes no *WEKA* com a base de dados balanceada BB1

Teste Realizado	Instâncias Classificadas Corretamente	Instâncias Classificadas Incorretamente	Melhor Algoritmo
T20	10903 (89.77%)	1243 (10.23%)	J48 (Árvore de Decisão)
T21	10657 (87.74%)	1489 (12.26%)	J48 (Árvore de Decisão)
T22	10788 (88.82%)	1358 (11.18%)	J48 (Árvore de Decisão)
T23	11086 (91.27%)	1060 (9.73%)	J48 (Árvore de Decisão)

Ainda analisando a Tabela com os resultados dos testes T20, T21, T22 e T23 pode-se observar que, em todos os testes com a base de dados balanceada BB1 e seleção de atributos, a sugestão de melhor algoritmo para o problema de classificação de fornecedores do Estado do Piauí é baseada em árvores de decisão, sendo específica ao uso do J48.

Comparando os resultados obtidos através dos testes da base de dados balanceada BB1 com e sem seleção de atributos, nota-se que a taxa de acerto geral aumentou. A análise das matrizes de confusão para cada um dos quatro testes (Tabela 14) mostra que todos os quatro obtiveram uma taxa de acerto superior a 88% em relação aos verdadeiros negativos.

Tabela 14 – Matrizes de confusão para os testes com a base de dados balanceada BB1

	T20		T21		T22		T23	
	<i>True</i>	<i>Negative</i>	<i>True</i>	<i>Negative</i>	<i>True</i>	<i>Negative</i>	<i>True</i>	<i>Negative</i>
<i>True</i>	5477	595	5280	792	5297	643	5549	524
<i>Negative</i>	648	5426	697	5377	715	5491	536	5537

Ainda analisando a Tabela com as matrizes de confusão para os testes com a base de dados balanceada BB1 e seleção de atributos, observa-se que o teste T23 também apresenta boa taxa de acerto referente aos verdadeiros negativos (fornecedores de alto risco que são, de fato, classificados como alto risco ou ruins), com 5.537 acertos para 6.073 instâncias (91.2% de acerto dos verdadeiros negativos).

4.1.3.2 Base de Dados Balanceada BB2 e Seleção de Atributos

Os testes realizados com a base de dados balanceada BB2 são T24, T25, T26 e T27. A análise da Tabela 15 referente a esses testes permite observar que o teste T25

apresenta os melhores resultados obtidos com classificação correta de 11.459 instâncias (aproximadamente 94%) e classificação incorreta de 687 instâncias. O teste T24 apresenta os piores resultados com classificação correta de 11.023 instâncias e classificação incorreta de 1.123 instâncias. Para T25 a sugestão de melhor algoritmo para o problema de classificação é baseada em árvore de decisão (J48) e para T24 a sugestão de melhor algoritmo também é baseada em árvore de decisão (J48).

Tabela 15 – Resultados dos testes no *WEKA* com a base de dados balanceada BB2

Teste Realizado	Instâncias Classificadas Corretamente	Instâncias Classificadas Incorretamente	Melhor Algoritmo
T24	11023 (90.75%)	1123 (9.25%)	J48 (Árvore de Decisão)
T25	11459 (94.34%)	687 (5.66%)	J48 (Árvore de Decisão)
T26	11082 (91.24%)	1064 (8.76%)	<i>Logistic Model Tree</i> (Árvore de Decisão)
T27	11354 (93.48%)	792 (6.52%)	J48 (Árvore de Decisão)

Ainda analisando a Tabela com os resultados dos testes T24, T25, T26 e T27 pode-se observar que, para os testes com a base de dados balanceada BB2 e seleção de atributos, existe uma predominância em utilizar algoritmos baseados em árvore de decisão (4 dos 4 casos de testes) para o problema de classificação dos fornecedores do Estado do Piauí.

Comparando os resultados obtidos através dos testes da base de dados balanceada BB2 com e sem seleção de atributos, nota-se que a taxa de acerto geral aumentou. A análise das matrizes de confusão para cada um dos quatro testes (Tabela 16) mostra que todos os quatro obtiveram uma taxa de acerto superior a 86% em relação aos verdadeiros negativos.

Tabela 16 – Matrizes de confusão para os testes com a base de dados balanceada BB2

	T24		T25		T26		T27	
	<i>True</i>	<i>Negative</i>	<i>True</i>	<i>Negative</i>	<i>True</i>	<i>Negative</i>	<i>True</i>	<i>Negative</i>
<i>True</i>	5755	416	5876	196	5707	365	5935	138
<i>Negative</i>	707	5268	491	5583	699	5375	654	5419

Ainda analisando a Tabela com as matrizes de confusão para os testes com a base de dados balanceada BB2 e seleção de atributos, observa-se que o teste T25 também apresenta boa taxa de acerto referente aos verdadeiros negativos (fornecedores de alto risco que são, de fato, classificados como alto risco ou ruins), com 5.583 acertos para 6.073 instâncias (91.9% de acerto dos verdadeiros negativos).

4.1.3.3 Base de Dados Balanceada BB3 e Seleção de Atributos

Os testes realizados com a base de dados balanceada BB3 são T28, T29, T30 e T31. A análise da Tabela 17 referente a esses testes permite observar que o teste T31 apresenta os melhores resultados obtidos com classificação correta de 10.761 instâncias (aproximadamente 88%) e classificação incorreta de 1.385 instâncias. O teste T28 apresenta os piores resultados com classificação correta de 10.144 instâncias e classificação incorreta de 2.002 instâncias. Para T31 a sugestão de melhor algoritmo para o problema de classificação é baseada em árvore de decisão (J48) e para T28 a sugestão de melhor algoritmo também é baseada em árvore de decisão (J48).

Tabela 17 – Resultados dos testes no *WEKA* com a base de dados balanceada BB3

Teste Realizado	Instâncias Classificadas Corretamente	Instâncias Classificadas Incorretamente	Melhor Algoritmo
T28	10144 (83.52%)	2002 (16.48%)	J48 (Árvore de Decisão)
T29	10713 (88.20%)	1433 (11.80%)	<i>Logistic Model Tree</i> (Árvore de Decisão)
T30	10509 (86.52%)	1637 (13.48%)	J48 (Árvore de Decisão)
T31	10761 (88.60%)	1385 (11.40%)	J48 (Árvore de Decisão)

Ainda analisando a Tabela com os resultados dos testes T28, T29, T30 e T31 pode-se observar que, para os testes com a base de dados balanceada BB3 e seleção de atributos, existe uma predominância em utilizar algoritmos baseados em árvore de decisão (4 dos 4 casos de testes) para o problema de classificação dos fornecedores do Estado do Piauí.

Comparando os resultados obtidos através dos testes da base de dados balanceada BB3 com e sem seleção de atributos, nota-se que a taxa de acerto geral aumentou. A análise das matrizes de confusão para cada um dos quatro testes (Tabela 18) mostra que todos os quatro obtiveram uma taxa de acerto superior a 85% em relação aos verdadeiros negativos.

Tabela 18 – Matrizes de confusão para os testes com a base de dados balanceada BB3

	T28		T29		T30		T31	
	<i>True</i>	<i>Negative</i>	<i>True</i>	<i>Negative</i>	<i>True</i>	<i>Negative</i>	<i>True</i>	<i>Negative</i>
<i>True</i>	4937	1164	5112	817	5040	991	5122	951
<i>Negative</i>	838	5207	616	5601	646	5469	434	5639

Ainda analisando a Tabela com as matrizes de confusão para os testes com a base de dados balanceada BB3 e seleção de atributos, observa-se que o teste T31 também

apresenta boa taxa de acerto referente aos verdadeiros negativos (fornecedores de alto risco que são, de fato, classificados como alto risco ou ruins), com 5.639 acertos para 6.073 instâncias (92.8% de acerto dos verdadeiros negativos).

4.1.3.4 Base de Dados Balanceada BB4 e Seleção de Atributos

Os testes realizados com a base de dados balanceada BB4 são T32, T33, T34 e T35. A análise da Tabela 19 referente a esses testes permite observar que o teste T33 apresenta os melhores resultados obtidos com classificação correta de 10.873 instâncias (aproximadamente 89%) e classificação incorreta de 1.273 instâncias. O teste T34 apresenta os piores resultados — dentre todos os testes com seleção de atributos — com classificação correta de 10.007 instâncias e classificação incorreta de 2.139 instâncias. Para T33 a sugestão de melhor algoritmo para o problema de classificação é baseada em árvore de decisão (J48) e para T34 a sugestão de melhor algoritmo também é baseada em árvore de decisão (J48).

Tabela 19 – Resultados dos testes no *WEKA* com a base de dados balanceada BB4

Teste Realizado	Instâncias Classificadas Corretamente	Instâncias Classificadas Incorretamente	Melhor Algoritmo
T32	10437 (85.99%)	1703 (14.01%)	J48 (Árvore de Decisão)
T33	10873 (89.52%)	1273 (10.48%)	J48 (Árvore de Decisão)
T34	10007 (82.39%)	2139 (17.61%)	J48 (Árvore de Decisão)
T35	10389 (85.54%)	1757 (14.46%)	<i>Logistic Model Tree</i> (Árvore de Decisão)

Ainda analisando a Tabela com os resultados dos testes T32, T33, T34 e T35 pode-se observar que, para os testes com a base de dados balanceada BB4 e seleção de atributos, existe uma predominância em utilizar algoritmos baseados em árvore de decisão (4 dos 4 casos de testes) para o problema de classificação dos fornecedores do Estado do Piauí.

Comparando os resultados obtidos através dos testes da base de dados balanceada BB4 com e sem seleção de atributos, nota-se que a taxa de acerto geral aumentou. A análise das matrizes de confusão para cada um dos quatro testes (Tabela 20) mostra que todos os quatro obtiveram uma taxa de acerto superior a 73% em relação aos verdadeiros negativos.

Ainda analisando a Tabela com as matrizes de confusão para os testes com a base de dados balanceada BB4 e seleção de atributos, observa-se que o teste T33 também

Tabela 20 – Matrizes de confusão para os testes com a base de dados balanceada BB4

	T32		T33		T34		T35	
	<i>True</i>	<i>Negative</i>	<i>True</i>	<i>Negative</i>	<i>True</i>	<i>Negative</i>	<i>True</i>	<i>Negative</i>
<i>True</i>	5339	644	5624	708	5548	524	5574	499
<i>Negative</i>	1059	5098	565	5252	1615	4459	1258	4815

apresenta boa taxa de acerto referente aos verdadeiros negativos (fornecedores de alto risco que são, de fato, classificados como alto risco ou ruins), com 5.252 acertos para 6.073 instâncias (86.4% de acerto dos verdadeiros negativos). Já para o teste T34 a taxa de acerto referente aos verdadeiros negativos é de 73.4% (com 4.459 acertos para 6.073 instâncias).

A Figura 7 representa uma comparação dos resultados obtidos através dos testes realizados com as bases de dados balanceadas BB1, BB2, BB3 e BB4 e seleção de atributos. Nela, observa-se que o teste com melhor resultado (relativo a taxa de acerto) é o teste T25 que classificou poucas instâncias a mais que o teste T37. Além disso, também é possível observar que as piores taxas de acerto com os testes T34 e T28, respectivamente.

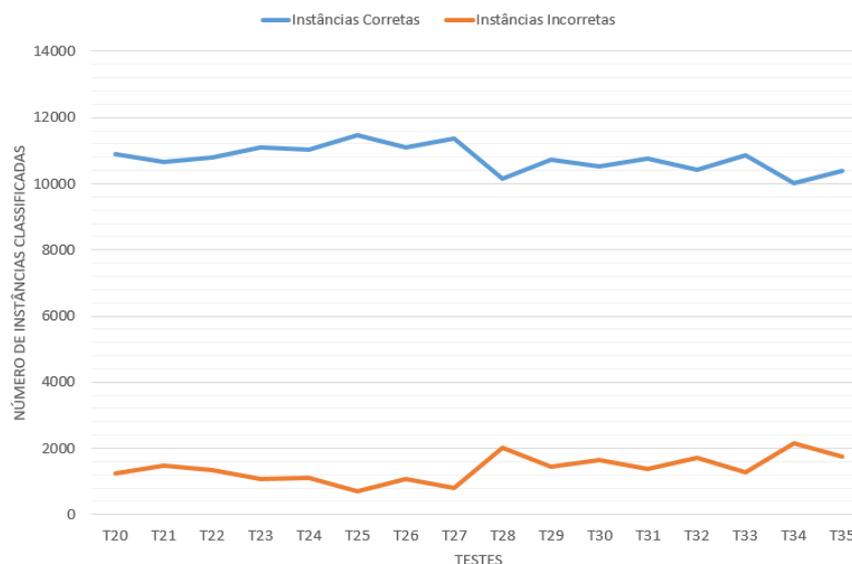


Figura 7 – Gráfico comparativo entre os testes com as bases de dados balanceadas e seleção de atributos

Para os testes com as bases de dados balanceadas e seleção de atributos observou-se uma melhoria na classificação dos dados em termos quantitativos e através da análise das matrizes de confusão. Essa melhoria deu-se principalmente pela redução da generalização do modelo. Com isso, tem-se um modelo mais específico. Em outras palavras, quando o algoritmo tenta analisar uma base de dados que inicialmente tinha 27 atributos e agora faz a mesma análise para a base de dados com apenas 9 atributos (sendo eles com maior ganho de informação em relação ao atributo classe), tem-se uma análise mais representativa e compacta.

4.1.4 Testes com as Bases de Dados Balanceadas, Seleção de Atributos e Normalização de Dados

Finalmente, foi realizada uma quarta etapa de testes através da análise das bases de dados balanceadas (BB1, BB2, BB3 e BB4) com filtros de seleção de atributos e os dados normalizados. A normalização é importante, visto que a maioria dos valores dos atributos estavam dispostos em escalas diferentes. Essa característica é observada ao comparar os atributos *CNAE_Qtd_Total* e *Pagto_2016*, sendo a variação dos valores dos atributos de 0 a 10 e de 0 a 4.130.267.649,08, respectivamente.

Para os testes realizados com as bases de dados balanceadas e seleção de atributos, foi utilizada como métrica de avaliação do algoritmo apenas o *recall* da classe negativa. Os resultados são apresentados na Tabela 21.

Tabela 21 – Resultados dos testes no *WEKA* com as quatro bases de dados balanceadas, seleção de atributos e dados normalizados

Teste Realizado	Instâncias Classificadas Corretamente	Instâncias Classificadas Incorretamente	Melhor Algoritmo
T36	9349 (76.97%)	2797 (23.03%)	J48 (Árvore de Decisão)
T37	9662 (79.55%)	2484 (20.45%)	J48 (Árvore de Decisão)
T38	9215 (75.87%)	2931 (24.13%)	J48 (Árvore de Decisão)
T39	9194 (75.70%)	2952 (24.30%)	J48 (Árvore de Decisão)

Os testes realizados com a base de dados balanceada BB1, BB2, BB3 e BB4 são T36, T37, T38 e T39, respectivamente. A análise da Tabela 21 referente a esses testes permite observar que o teste T37 apresenta os melhores resultados obtidos com classificação correta de 9.662 instâncias (aproximadamente 79.5%) e classificação incorreta de 2.484 instâncias. O teste T39 apresenta os piores resultados com classificação correta de 9.194 instâncias e classificação incorreta de 2.952 instâncias.

Ainda analisando a Tabela com os resultados dos testes T36, T37, T38 e T39 pode-se observar que, para todos os testes com as base de dados balanceadas BB1, BB2, BB3 e BB4, seleção de atributos e normalização de dados, a sugestão para resolução do problema de classificação dos fornecedores do Estado do Piauí é baseada na utilização do J48.

Comparando os resultados obtidos através dos testes das bases de dados balanceadas BB1, BB2, BB3 e BB4 com e sem normalização de dados, nota-se que a taxa de acerto geral diminuiu. No entanto, a análise das matrizes de confusão para cada um dos quatro

testes (Tabela 22) mostra que todos os quatro obtiveram uma taxa de acerto superior a 87.5% em relação aos verdadeiros negativos.

Tabela 22 – Matrizes de confusão para os testes com as quatro bases de dados balanceadas, seleção de atributos e dados normalizados

	T36		T37		T38		T39	
	<i>True</i>	<i>Negative</i>	<i>True</i>	<i>Negative</i>	<i>True</i>	<i>Negative</i>	<i>True</i>	<i>Negative</i>
<i>True</i>	3880	2193	3838	2235	3898	2175	3812	2261
<i>Negative</i>	604	5469	249	5824	756	5317	691	5382

Ainda analisando a Tabela com as matrizes de confusão para os testes com as bases de dados balanceadas, seleção de atributos e normalização de dados, observa-se que o teste T37 também apresenta boa taxa de acerto referente aos verdadeiros negativos (fornecedores de alto risco que são, de fato, classificados como alto risco ou ruins), com 5.824 acertos para 6.073 instâncias (95.9% de acerto dos verdadeiros negativos). Já para o teste T39 a taxa de acerto referente aos verdadeiros negativos é de 88.6% (com 5.382 acertos para 6.073 instâncias).

4.2 Sistema de Avaliação de Fornecedores (SAF)

Os detalhes apresentados nesta seção referem-se à implementação do sistema. Na etapa de testes, utilizou-se da linguagem JAVA e da biblioteca do *WEKA* para realizar a implementação do J48 e fazer alguns testes além dos envolvidos com o *AUTOWEKA*.

Para o desenvolvimento do sistema, optou-se por realizar a implementação utilizando a linguagem de programação *PYTHON*. A escolha da linguagem, deu-se pela vasta difusão da mesma no universo acadêmico e profissional, além do excelente desempenho ao trabalhar com funções e desenvolvimento web. Além da escolha da linguagem *PYTHON*, também utilizou-se a biblioteca do *WEKA*, responsável por toda a parte referente ao algoritmo J48 e seu desempenho referente à classificação dos dados. Utilizaram-se ainda outras dependências para o desenvolvimento do sistema, como:

- *Django - framework* para desenvolvimento web;
- *pygraphviz* - interface usada para criar, ler, escrever e desenhar grafos;
- *pymssql* - interface usada para estabelecer conexão ao banco de dados *SQL*;
- *SQLAlchemy* - biblioteca de mapeamento objeto-relacional *SQL*;
- *uwsgi* - software que tem como objetivo desenvolver uma pilha completa para a construção de serviços de hospedagem.

Inicialmente, a utilização do *WEKA* conjunta com o *PYTHON* representou um desafio, devido as divergências de versões para ambos e para as dependências. Como solução para o problema, a versão do *PYTHON* utilizada foi a 3.4.4 (64-bits). Ela é a versão mais recente que possui implementação de todas as dependências necessárias para o desenvolvimento do sistema.

4.2.1 Sistema

O presente trabalho tem como objetivo, além de buscar uma solução para o problema de classificação dos fornecedores do Estado do Piauí, desenvolver um sistema, produto da pesquisa, que auxilie no trabalho de análise por parte dos auditores do TCE-PI. O sistema desenvolvido é chamado de Sistema de Avaliação de Fornecedores (SAF). O fluxograma da Figura 8 mostra a ordem das tarefas desde o acesso ao sistema à análise dos resultados e/ou classificação de novos fornecedores. A Figura 9 mostra a tela inicial do sistema e auxilia na análise e compreensão do fluxograma.

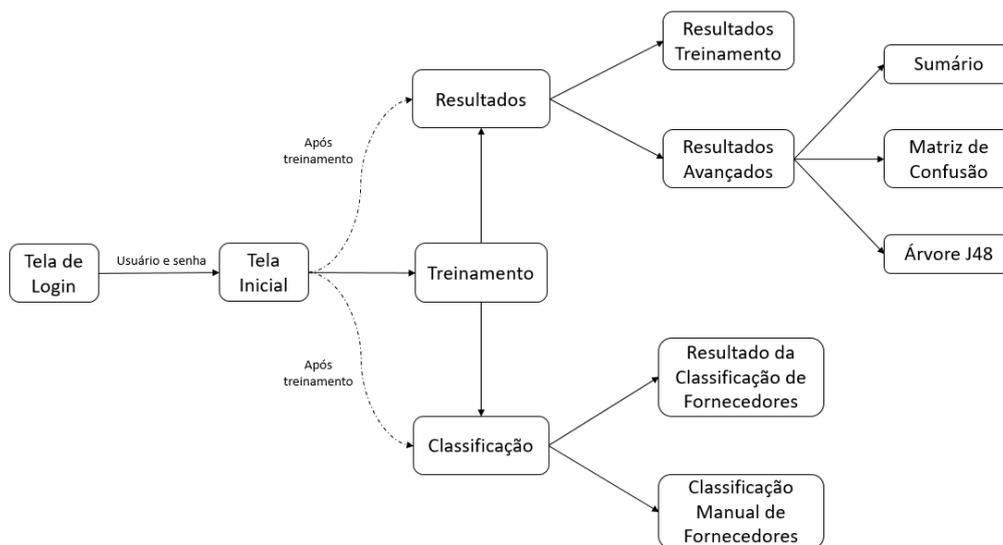


Figura 8 – Fluxograma das atividades referentes à utilização do SAF

Ao realizar acesso ao sistema, através do usuário e senha, a tela inicial é carregada (Figura 9). Nela existem quatro componentes principais:

1. Cabeçalho do Sistema - Contém o nome e a sigla;
2. Menu do Sistema - Contém o acesso ao Treinamento, aos Resultados e à classificação;
3. Corpo do Sistema - Contém as informações exibidas através do processamento de dados, como tabelas, árvores e formulários;
4. Rodapé do Sistema - Contém acesso as informações dos responsáveis pela pesquisa e desenvolvimento do sistema: Tribunal de Contas do Estado do Piauí, Laboratório de Inteligência Artificial e Universidade Federal do Piauí.



Figura 9 – Tela inicial do SAF

Seguindo o fluxograma (Figura 8), na tela inicial é possível acessar os menus de treinamento, resultados e classificação. No entanto, para acessar os resultados e a classificação, é necessário que seja realizado o treinamento dos dados. O treinamento é a primeira tarefa que deve ser realizada após a implantação do sistema no servidor. É através do treinamento que é possível gerar resultados e classificar novas instâncias. A (Figura 10) mostra a página de treinamento do sistema, onde são carregados dois arquivos de entrada, um para treinamento e outro para teste e exibição dos resultados.



Figura 10 – Tela de treinamento do SAF

Após o usuário realizar o treinamento dos dados, o sistema carrega automaticamente a página com os resultados do treinamento (Figura 11). Nela é possível observar o *ID* do fornecedor, os valores para os atributos, a chance do fornecedor ser de alto risco ou de baixo risco e se o classificador acertou a análise (caso tenha acertado, a cor em destaque é o verde, e caso tenha errado, a cor em destaque é o vermelho).

SAF
Sistema de Avaliação de Fornecedores

Home Treinamento Resultados Treinamento Resultados Avançados Classificação de Fornecedores Classificar Sair

ID	ST EPPME	Pagto 2016	Idade Empresa 2017	Rais Qtd Funcionarios 2017	ESFL	Compras Soma Valor Homologado Total	Compras Qtd Entes Media	Cnae Qtd Total	Acertou	Chance de Ser Alto Risco	Chance de Ser Baixo Risco
4260	1,0	3000,0	13,0	3,0	0,0	0,0	0,0	5,0	SIM	70.37%	29.63%
4259	0,0	30710,0	13,0	734,0	0,0	810679,0	5,0	4,0	NÃO	29.63%	70.37%
4258	0,0	0,0	13,0	313,0	0,0	16000,0	2,0	5,0	SIM	75.22%	24.78%
4257	1,0	0,0	13,0	16,0	0,0	0,0	0,0	1,0	SIM	75.22%	24.78%
4256	1,0	0,0	13,0	8,0	0,0	0,0	0,0	0,0	SIM	75.22%	24.78%

Figura 11 – Tela de resultados do treinamento do SAF

Além dos resultados do treinamento, existe um menu de resultados avançados que contém o sumário (resumo) do treinamento (com taxa de acerto, erro quadrático médio e número total de instâncias), a representação da matriz de confusão e a árvore de treinamento para o J48. A Figura 12 mostra um fragmento da árvore. Nela é possível perceber o nó raiz sendo o atributo *Idade_Empresa_2017*. O limiar definido para esse atributo foi o 13, sendo as empresas divididas em empresas com idade menor ou igual a 13 e empresas com idade maior que 13. Caso seja selecionada uma empresa aleatória com idade igual a 18, analisa-se o atributo *ESFL*. Ao analisar o valor do atributo e verificar que o mesmo possui valor 0, novamente analisa-se o atributo *Idade_Empresa_2017*. A análise desse atributo para um limiar igual a 23, como observado na árvore, permite dizer que a classificação da empresa seria com o valor 1, ou seja, uma empresa de baixo risco.

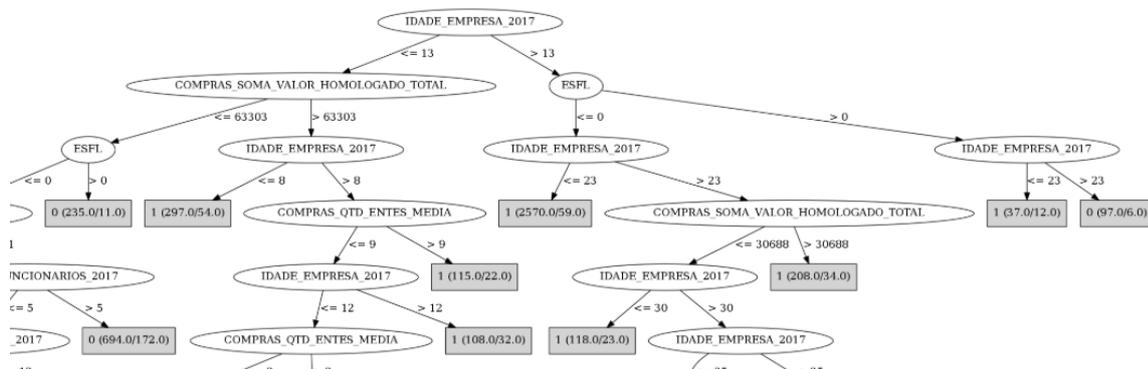


Figura 12 – Fragmento da árvore de treinamento do SAF

Existem ainda duas opções de menu: a classificação manual dos fornecedores e a exibição dos resultados gerados dessa classificação. A classificação manual é dada pelo menu **Classificar**. Ao selecionar esse menu, será carregado um formulário para a inserção dos valores dos atributos:

1. CNPJ;
2. Nome ou Razão Social;
3. ST_EPPME;
4. Pagto_2016;
5. Idade_Empresa_2017;
6. Rais_Qtd_Funcionarios_2017;
7. ESFL;
8. Compras_Soma_Valor_Homologado_Total;
9. Compras_Qtd_Entes_Media;
10. CNAE_Qtd_Total.

Os resultados da classificação manual podem ser observados ao acessar o menu **Classificação de Fornecedores**. Nele é exibida uma tabela com todas as informações preenchidas através do formulário, além da probabilidade da empresa ser de baixo risco ou de alto risco e a respectiva predição. Essa probabilidade é dada de acordo com a semelhança com cada uma das classes determinadas. Caso uma empresa seja de baixo risco, sua predição será representada pela cor verde, e caso uma empresa seja de alto risco, sua predição será representada pela cor vermelha.

Ainda no menu de **Classificação de Fornecedores**, percebe-se o CNPJ de cada empresa em azul. Caso o auditor não concorde com a classificação dada pelo algoritmo, é possível clicar no CNPJ da empresa e fazer a alteração na predição dada pelo classificador. Neste caso, o dado corrigido será submetido novamente ao treinamento e irá gerar um novo conhecimento. Em outras palavras, a utilização de aprendizagem de máquina trará novas formas de conhecimento para o sistema e para o auditor. Além disso, é possível fazer uma busca por CNPJ ou nome de uma determinada empresa que seja de interesse do auditor.

4.3 Discussão dos Resultados

Após a análise dos resultados, observa-se uma tendência na utilização de algoritmos baseados em árvore de decisão para o problema de classificação proposto (31 dos 39 testes utilizam árvore de decisão). Dentre esses algoritmos de árvores de decisão, 25 referem-se ao J48, 5 referem-se ao *Logistic Model Tree* e 1 refere-se ao *Random Forest*.

Para os primeiros testes, realizados com base de dados desbalanceada, os algoritmos sugeridos foram redes neurais e árvores de decisão (*Logistic Model Tree* e *Random Forest*).

No entanto, esses resultados não são satisfatórios, pois as matrizes de confusão referentes aos mesmos inviabilizam encontrar uma solução viável (taxa de acerto referente aos verdadeiros negativos superior a 80%) para o problema de classificação de fornecedores do Estado do Piauí.

Em seguida, foram feitos testes com as quatro bases de dados balanceadas (BB1, BB2, BB3 e BB4) e com os resultados obtidos, pode-se observar que os algoritmos encontrados para solucionar o problema continuaram sendo Redes Neurais e Árvores de Decisão (J48 e *Logistic Model Tree*), além do aparecimento do *Naive Bayes* (algoritmo probabilístico). Os resultados encontrados com esses algoritmos foram satisfatórios se comparados ao primeiro teste (de acordo com análise das matrizes de confusão), no entanto, ainda tornava-se necessário fazer alguns ajustes buscando resultados melhores.

Em uma terceira etapa, foram realizados testes com as bases balanceadas (BB1, BB2, BB3 e BB4) aplicando seleção de atributos a elas. Os testes com seleção de atributos mostraram uma superioridade do algoritmo J48 em relação aos outros. Isso é observado por conta de alguns fatores, tais como a presença de atributos com valores binários (*ST_EPPME* e *ESFL*) e a presença de vários atributos com valores contínuos (*Idade_Empresa_2017*, *CNAE_Qtd_Total*, *Rais_Qtd_Funcionarios_2017*, *Compras_Qtd_Entes_Media*, *Compras_Soma_Valor_Homologado_Total* e *Pagto_2016*). Para os atributos com valores contínuos, o algoritmo define um limiar e então divide a amplitude dos valores de forma binária. Em outras palavras, o algoritmo divide os valores de um atributo em dois grupos de valores: os que são maiores que o limiar e os que são menores ou igual ao limiar.

Finalmente, acontece a etapa dos testes com normalização dos valores dos atributos. A normalização é importante devido à disparidade dos valores dos atributos em termos de escala. Como discutido anteriormente, ao comparar os atributos *CNAE_Qtd_Total* com amplitude de 0 a 10 e *Pagto_2016* com amplitude de 0 a 4.130.267.649,08, existe a necessidade de manter os dados estruturados, justificando o processo de normalização. Os resultados obtidos através dos testes com as quatro bases balanceadas e seus dados normalizados, apresentaram uma queda sutil na taxa de classificação geral. No entanto, ao analisar as matrizes de confusão para os testes, observou-se um bom rendimento quanto da taxa de acerto referente aos verdadeiros negativos — fornecedores de alto risco que são, de fato, classificados como alto risco ou ruins.

A Figura 13 representa uma comparação referente a todos os testes realizados (T1 a T39) quanto à taxa de acerto dos valores verdadeiros negativos das matrizes de confusão. Nela observa-se que com a preparação de dados, o balanceamento, a seleção de atributos e a normalização dos dados, houve um incremento no acerto dos verdadeiros negativos.

Em algumas situações, como é o caso dos testes com normalização de dados, a taxa de acerto geral do algoritmo apresentou uma queda sutil em relação aos testes com seleção de atributos. No entanto, para o problema de classificação de fornecedores, é muito mais

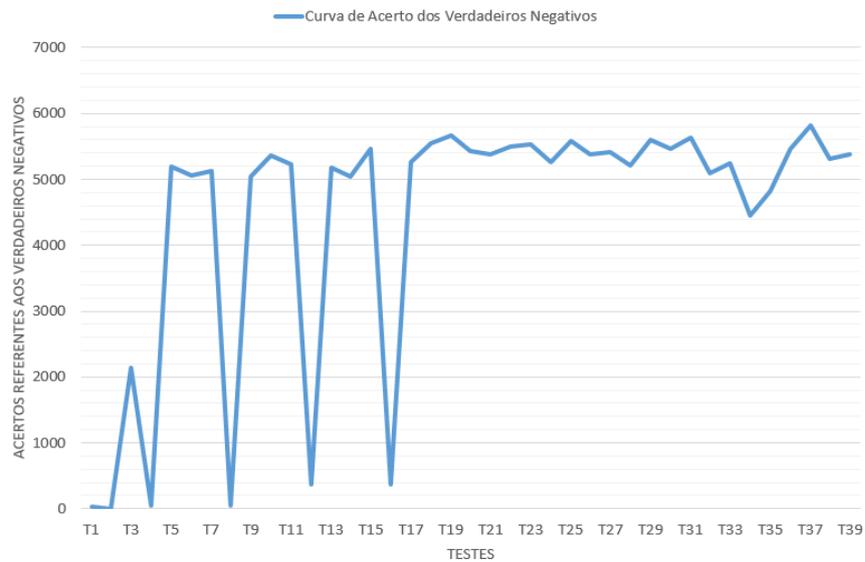


Figura 13 – Gráfico representativo da curva de acerto referente aos verdadeiros negativos para os testes realizados

importante avaliar e classificar de maneira correta os fornecedores de alto risco ou ruins (verdadeiros negativos da matriz). Com isso, os resultados obtidos com a normalização de dados, foram considerados satisfatórios.

Outra análise a ser considerada é a verificação da evolução das métricas utilizadas para avaliar os algoritmos. A Figura 14 mostra a evolução na taxa de acerto (0 a 100%) das métricas *precision*, *recall* e *f-measure* ao longo dos testes realizados.

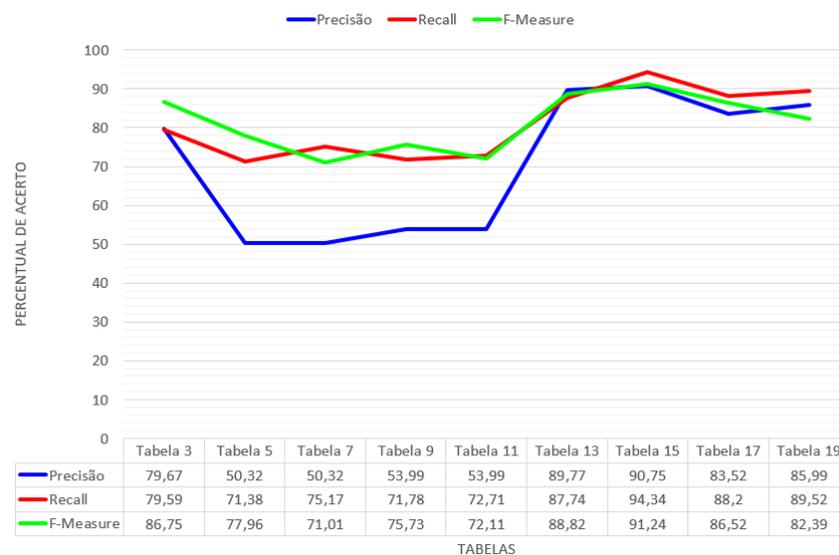


Figura 14 – Gráfico comparativo das métricas precisão, *recall* e *f-measure*

Para a Tabela 3 os testes apresentados foram realizados com a base desbalanceada. Os valores das métricas *precision*, *recall* e *f-measure* são de aproximadamente 80%, mas a análise das matrizes de confusão mostram a falta de qualidade dos dados classificados. As

Tabelas 5, 7, 9 e 11 apresentam os testes realizados com as bases de dados balanceadas. Para esses testes, os valores da métrica *precision* são de aproximadamente 50% e as métricas *recall* e *f-measure* variam de 70% a 77%, mas a análise das matrizes de confusão mostram melhoras na qualidade dos dados classificados.

Ainda na Figura 14, para as Tabelas 13, 15, 17 e 19 são apresentados os resultados dos testes para as bases de dados balanceadas com seleção de atributos. Para esses testes, os valores da métrica *precision*, *recall* e *f-measure* variam de 82% a 94% e a análise das matrizes de confusão mostram melhoras significativas na qualidade dos dados classificados.

Finalmente, a análise das quatro etapas de testes permitiram concluir e preparar o estudo para a implementação de um sistema com base no algoritmo J48 com normalização dos dados e seleção de atributos. A escolha do algoritmo deu-se pela presença na maioria dos testes e pela qualidade dos resultados. Considerando T37 com taxa de acerto aproximada de 80% e análise de matriz de confusão apontando 95.8% de acerto na classe desejada (verdadeiros negativos), tem-se que o trabalho do auditor ficaria reduzido a um escopo de 80% da quantidade inicial de dados, sendo 95.8% desses dados confiáveis para a classe desejada. Em outras palavras, essa taxa de acerto, considerando a matriz de confusão associada a ela, é bastante satisfatória.

5 Conclusões

Neste capítulo são discutidas as conclusões que foram atingidas através do desenvolvimento do estudo e do sistema. Adicionalmente, são listados possíveis trabalhos futuros relacionados a melhorias para o sistema.

5.1 Conclusões

Modelos preditivos utilizando aprendizagem de máquina são largamente utilizados no âmbito empresarial privado, especialmente por instituições que buscam prevenir fraudes ou medir o risco de transações financeiras. No escopo governamental ainda são poucas as experiências consolidadas de uso efetivo de tais técnicas. No Brasil, especificamente, há uma visível subutilização dos dados relativos a contratações públicas.

Este trabalho buscou suprir essa lacuna a partir da utilização massiva dos dados fornecidos pelo Tribunal de Contas do Estado do Piauí, com o complemento de diversas bases que se relacionam indiretamente com o processo de aquisições governamentais, através de uma pesquisa e posterior desenvolvimento de um sistema classificador que seja capaz de avaliar os fornecedores do Estado do Piauí em fornecedores de baixo risco ou fornecedores de alto risco.

A classificação de dados pode ser uma tarefa complexa para um sistema computacional. Vários fatores devem ser ponderados no momento da classificação, como verificar a disposição dos dados, a escolha do algoritmo utilizado, se a base encontra-se balanceada ou desbalanceada, se os dados encontram-se normalizados ou não, dentre outros. O objetivo deste trabalho consiste em elaborar um modelo de classificação de fornecedores do Estado que, através do conhecimento adquirido na etapa de treinamento, seja capaz de prever o comportamento de futuros prestadores de serviços que irão ser adicionados à base.

Para atingir os objetivos da pesquisa, utilizou-se uma abordagem dividida em etapa de testes e etapa de desenvolvimento do sistema. Na etapa de testes percebeu-se o problema da classificação, bem como houve uma preparação de dados e foram realizados vários experimentos. Para a preparação de dados, cabe destacar as etapas de balanceamento e normalização. Esses procedimentos são responsáveis por tornar os dados estruturados e melhorar o desempenho do algoritmo escolhido para o processo de classificação. Além disso, é na preparação de dados que foi feita a seleção dos atributos utilizados no processo. Os experimentos realizados nessa etapa foram divididos segundo as análises feitas na preparação de dados e totalizaram 39 testes.

Para os testes realizados, alguns algoritmos foram apontados inicialmente como can-

didatos na resolução do problema. A maioria deles são baseados em árvore de decisão, além das redes neurais e do *naive bayes*. Nos testes iniciais — com base de dados desbalanceada — os resultados com a classificação apresentaram taxa de acerto de aproximadamente 80%, mas as matrizes de confusão associadas aos testes não foram satisfatórias. Em seguida, os testes realizados com as bases de dados balanceadas obtiveram taxa de acerto média de 70%, com exceção dos testes com a métrica *precision* (com taxa de acerto aproximada de 50%). Para esses testes, a matriz de confusão mostrou-se bem melhor do que a apresentada nos testes iniciais.

Ainda foram realizados os testes com seleção de atributos, onde a taxa de acerto ficou entre 82% e 94%. Nesses testes as métricas avaliadas foram: *precision*, *recall*, *f-measure* e *recall* da classe negativa. As matrizes de confusão referentes aos testes foram bastante satisfatórias, com elevada taxa de acerto para os verdadeiros negativos. Finalmente, foram realizados testes com normalização de dados, onde a taxa de acerto geral foi reduzida. Embora a taxa de acerto tenha sido reduzida (taxa média de 80%), as matrizes de confusão associadas a cada uma das bases de dados testadas obtiveram índices satisfatórios, com o melhor caso sendo de 95.8% referente aos verdadeiros negativos (fornecedores de alto risco classificados, de fato, como fornecedores de alto risco).

Ao final dos testes, percebeu-se uma predominância em utilizar árvores de decisão para o problema de classificação binária. Para 31 dos 39 casos de testes foram utilizadas árvores de decisão, sendo 35 desses casos, caracterizados pela utilização do J48. Os bons resultados obtidos aconteceram principalmente pelo formato contínuo dos dados, onde foi possível utilizar um limiar de divisão para cada atributo. A utilização desse limiar também aconteceu após a normalização de dados, onde os mesmos passaram a ser discretos. Finalmente, foi selecionado o algoritmo J48 para resolver o problema de classificação binária dos fornecedores do Estado do Piauí e realizou-se a implementação do sistema.

Na etapa de desenvolvimento do sistema, foram utilizados os conhecimentos adquiridos na etapa de teste. O produto gerado nessa etapa é um sistema que avalia e classifica fornecedores com base em atributos pré-selecionados, após uma etapa de treinamento de dados.

Os resultados obtidos através desse estudo apontaram para uma taxa de acerto variável de acordo com a preparação de dados realizada. Para a versão final, utilizada no sistema, a abordagem consta de balanceamento de dados, seleção de atributos e normalização de dados. A taxa de acerto observou-se aproximada de 80% quanto à classificação geral dos dados. No entanto, observando a matriz de confusão para esse experimento, percebe-se uma elevada taxa de acerto quanto ao objetivo da pesquisa. A taxa referente aos verdadeiros negativos (fornecedores de alto risco que são, de fato, classificados como alto risco) é superior a 95%.

A utilização do sistema, inicialmente, é pretendida pelo Tribunal de Contas do

Estado do Piauí, pois a análise dos dados foi feita de acordo com os fornecedores do Estado do Piauí. No entanto, o sistema é possível de ser utilizado em outros Tribunais de Contas Estaduais. Para esse fim, algumas alterações devem ser realizadas, como ajustes nos atributos pertencentes às bases de dados de entrada e ajustes na conexão com o banco de dados pertencente ao Tribunal de Contas em questão. Os ajustes de pré-processamento de dados podem ser os mesmos utilizados na pesquisa e os resultados vão depender do formato dos dados utilizados nos testes.

Finalmente, algumas melhorias e alternativas ainda podem ser desenvolvidas:

- aplicar a verificação de outros valores do limiar para os atributos, conforme a necessidade exigida em cada caso;
- analisar os resultados obtidos com outras técnicas de aprendizagem de máquina;
- utilizar outras técnicas de seleção de atributos, como a análise dos componentes principais (PCA).

Referências

- ALPAYDIN, E.; BACH, F. *Introduction to Machine Learning*. 2nd. ed. [S.l.]: MIT Press, 2014. Citado na página 4.
- AMIN, N.; HABIB, A. Comparison of different classification techniques using weka for hematological data. *American Journal of Engineering Research*, Stanford, USA, v. 4, n. 3, 2015. Citado 2 vezes nas páginas 18 e 19.
- ARAUJO, E. A.; CARMONA, C. U. de M. Desenvolvimento de modelos credit scoring com abordagem de regressão logística para a gestão da inadimplência de uma instituição de microcrédito. *Contabilidade Vista Revista*, v. 18, n. 3, p. 107–131, 2009. Citado na página 17.
- BALANIUK, R. et al. Risk based government audit planning using naive bayes classifiers. In: *Knowledge-Based and Intelligent Information and Engineering Systems*. Spain: s.n., 2012. Citado 3 vezes nas páginas 2, 17 e 18.
- BARBER, D. *Bayesian Reasoning and Machine Learning*. New York, NY, USA: Cambridge University Press, 2012. Citado 3 vezes nas páginas 7, 8 e 12.
- BELLMAN, R. E. *Dynamic Programming*. New York, NY, USA: Dover Publications, Inc., 2003. Citado na página 7.
- BRAGA, L. V. et al. *O papel do governo eletrônico no fortalecimento da governança do setor público*. 1. ed. Brasília: Revista do Serviço Público, 2008. Citado na página 1.
- CAOQUETTE, J. B. et al. *Gestão do risco de crédito: o grande desafio dos mercados financeiros globais*. 2. ed. Rio de Janeiro: Qualitymark, 2009. Citado na página 16.
- CARVALHO, R. S. *Modelos Preditivos para Avaliação de Risco de Corrupção de Servidores Públicos Federais*. Dissertação (Mestrado) — Universidade de Brasília, 2015. Citado na página 17.
- CARVALHO, R. S.; CARVALHO, R. N. Bayesian models to assess risk of corruption of federal management units. In: *13th Bayesian Modelling Applications Workshop*. [S.l.]: s.n., 2016. Citado na página 17.
- COSTA, M. L. d. *Uma Abordagem Utilizando Aprendizagem Por Reforço Hierárquica e Computação Paralela para Problema dos K-Servos*. Dissertação (Doutorado) — Universidade Federal do Rio Grande do Norte, 2017. Citado na página 7.
- DEVACENA, L. Comparative analysis of random forest, rep tree and j48 classifiers for credit risk prediction. *International Journal of Computer Applications*, New York, NY, USA, n. 3, 2015. Citado na página 17.
- GHAMRAWI, N.; MCCALLUM, A. Collective multi-label classification. In: *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*. New York, NY, USA: ACM, 2005. p. 195–200. Citado na página 13.

- GODBOLE, S.; SARAWAGI, S. Discriminative methods for multi-labeled classification. *Advances in Knowledge Discovery and Data Mining*, p. 22–30, 2004. Citado na página 14.
- HAYKIN, S. *Redes neurais: princípios e prática*. 2. ed. [S.l.]: Bookman, 2001. Citado na página 9.
- HOPFIELD, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, v. 79, p. 2554–2558, 1982. Citado na página 9.
- KOHONEN, T. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, v. 43, p. 59–69, 1982. Citado na página 9.
- LEVENBERG, K. A method for the solution of certain non-linear problems in least squares. *Quarterly Journal of Applied Mathematics*, II, n. 2, p. 164–168, 1944. Citado na página 10.
- LIBRELOTTO, S. R.; MOZZAQUATRO, P. M. Análise dos algoritmos de mineração j48 e apriori aplicados na detecção de indicadores da qualidade de vida e saúde. *Revista Interdisciplinar de Ensino, Pesquisa e Extensão*, Cruz Alta, Rio Grande do Sul, Brasil, v. 1, n. 1, 2013. Citado 2 vezes nas páginas 18 e 19.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. 1943. *Bulletin of mathematical biology*, v. 52, p. 99–115; discussion 73, 1990. Citado na página 9.
- MITCHELL, T. M. *Machine Learning*. 1. ed. New York, NY, USA: McGraw-Hill, Inc., 1997. Citado 2 vezes nas páginas 7 e 11.
- NGAI, E. W. T. et al. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, v. 50, n. 3, p. 559–569, 2011. Citado 2 vezes nas páginas 18 e 19.
- PANIGRAHI, S. et al. Credit card fraud detection: A fusion approach using dempster-shafer theory and bayesian learning. *Inf. Fusion*, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, v. 10, n. 4, 2009. Citado na página 18.
- PIETRO, M. S. Z. di. *Direito Administrativo*. 1. ed. São Paulo: Atlas, 1999. Citado 3 vezes nas páginas 11, 13 e 1.
- POWERS, D. M. W. Evaluation: From precision, recall and f-measure to roc. *Journal of Machine Learning Technologies*, p. 37–63, 2011. Citado na página 14.
- QUINLAN, J. R. *C4.5: Programs for Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993. Citado na página 11.
- ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, p. 65–386, 1958. Citado na página 9.
- SEMEDO, D. *Credit Scoring: Aplicação da Regressão Logística vs Redes Neurais Artificiais na Avaliação do Risco de Crédito no Mercado Cabo-Verdiano*. Dissertação (Mestrado) — Instituto Superior de Estatística e Gestão da Informação, Universidade Nova de Lisboa, 2010. Citado na página 16.

SILVA, I. N. d.; SPATTI, D. H.; FLAUZINO, R. A. *Redes neurais artificiais para engenharia e ciências aplicadas*. [S.l.]: Artliber Editora, 2010. Citado 4 vezes nas páginas 15, 8, 9 e 10.

THOMAS, L. C.; EDELMAN, D. B.; CROOK, J. N. *Credit Scoring and Its Applications*. Philadelphia, PA, USA: Siam, 2002. Citado na página 16.

VASCONCELLOS, M. S. de. *Proposta de método para análise de concessões de crédito a pessoas físicas*. Tese (Doutorado) — Universidade de São Paulo, 2002. Citado na página 17.

WITTEN, I. H. et al. *Weka: Practical Machine Learning Tools and Techniques with Java Implementations*. 1999. Citado na página 11.