

Universidade Federal de Campina Grande
Centro de Engenharia Elétrica e Informática
Coordenação de Pós-Graduação em Ciência da Computação

Eficácia no uso de aprendizagem de máquina para
estimação de risco em contratos públicos e empresas

Talita Lôbo de Menezes

Dissertação submetida à Coordenação do Curso de Pós-Graduação em
Ciência da Computação da Universidade Federal de Campina Grande -
Campus I como parte dos requisitos necessários para obtenção do grau
de Mestre em Ciência da Computação.

Área de Concentração: Ciência da Computação
Linha de Pesquisa: Sistemas de Computação

Nazareno Ferreira de Andrade
(Orientador)

Campina Grande, Paraíba, Brasil

©Talita Lôbo de Menezes, 04/07/2019

**"EFICÁCIA NO USO DE APRENDIZAGEM DE MÁQUINA PARA ESTIMAÇÃO DO
RISCO EM CONTRATOS PÚBLICOS E EMPRESAS"**

TALITA LÔBO DE MENEZES

DISSERTAÇÃO APROVADA EM 29/07/2019

**NAZARENO FERREIRA DE ANDRADE, Dr., UFCG
Orientador(a)**

**JOÃO ARTHUR BRUNET MONTEIRO, Dr., UFCG
Examinador(a)**

**ADRIANO ALONSO VELOSO, Dr., UFMG
Examinador(a)**

CAMPINA GRANDE - PB

M543e Menezes, Talita Lôbo de.
Eficácia no uso de aprendizagem de máquina para estimação de risco em contratos públicos e empresas / Talita Lôbo de Menezes. – Campina Grande, 2019.
77 f. : il. color.

Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2019.
"Orientação: Prof. Dr. Nazareno Ferreira de Andrade".
Referências.

1. Aprendizagem de Máquina. 2. Contratos Públicos. 3. Avaliação de Risco. 4. Gastos Públicos. I. Andrade, Nazareno Ferreira de. II. Título.

CDU 004:35(043)

FICHA CATALOGráfICA ELABORADA PELA BIBLIOTECÁRIA SEVERINA SUELI DA SILVA OLIVEIRA CRB-15/225

Resumo

A contratação de empresas pela Administração Pública é um processo frequente e necessário na prestação de serviços para autarquias e órgãos governamentais em diversos países. Por lei, os contratos celebrados devem ser acompanhados em sua execução. Entretanto, devido à grande quantidade de contratos administrativos gerados entre empresas e entes públicos, e à escassez de recursos da Administração, é impraticável fiscalizar todos os documentos firmados e seus andamentos. Uma das estratégias de controle utilizadas pelos órgãos de persecução criminal e cortes de contas é a priorização de casos a serem investigados a partir do seu nível de risco. Na literatura, o problema da estimação de risco utilizando aprendizado de máquina para entidades da Administração é tratado, na maioria dos casos, no contexto Federal. Atualmente na Paraíba esta escolha de casos é feita na prática através de uma série de heurísticas que compõem uma Matriz de Risco. Através da avaliação experimental da eficácia deste modelo prático e de seis modelos de aprendizagem de máquina na estimação do risco de empresas, concluímos que esta abordagem se sobressaiu em relação àquela. No contexto de contratos públicos municipais do estado da Paraíba, mediante a análise experimental do desempenho de três técnicas de balanceamento de classes e seis modelos de aprendizagem de máquina, foi possível verificar que a utilização destes modelos para o grânulo "contrato" da Administração Pública é, de fato, promissora. Para além disso, observou-se também que a construção de modelos com dados de contratos de áreas específicas (no contexto deste estudo, obras e serviços de engenharia) pode torná-los ainda mais eficazes. Com isso, esperamos auxiliar na otimização dos recursos públicos e no traçado de estratégias de controle interno que permitam subsidiar entidades de controle externo.

Abstract

The hiring of companies by the Public Administration is a frequent and necessary process in providing services to municipalities and government agencies in several countries. By law, contracts concluded must be accompanied by their execution. However, due to the large number of administrative contracts generated between companies and public entities, and the scarcity of management resources, it is impracticable to inspect all signed documents and their progress. One of the control strategies used by criminal prosecution and accountability agencies is the prioritization of cases to be investigated based on their level of risk. In the literature, the problem of risk estimation using machine learning for management entities is dealt with, in most cases, in the Federal context. Currently, in Paraíba, this choice of cases is made in practice through a series of heuristics that make up a Risk Matrix. Through experimental evaluation of the effectiveness of this practical model and six machine learning models to the risk assessment of companies, we concluded that this approach had outperformed that one. In the context of public municipal contracts of the Paraíba state, through experimental analysis of the effectiveness of three class balancing techniques and six models of machine learning, it was validated that the use of these models for public contracts is, indeed, promising. Also, we noted that building models with data from specific areas (in this work, engineering works and services) could improve its performance. With this, we aim to assist the use of public resources and the design of internal control strategies that allow subsidizing external control entities.

Agradecimentos

Agradeço primeiramente a minha família, em especial aos meus pais, Jonas e Tereza, pelo apoio não apenas durante minha jornada acadêmica, mas em qualquer coisa que decida fazer. Vocês são incríveis!

Ao meu orientador e amigo, professor Nazareno Andrade, pela paciência, ensinamentos, inspiração, dedicação e generosidade no processo de orientação deste trabalho.

Agradeço ao time do Laboratório Analytics e Laboratório de Sistemas Distribuídos pelo ótimo ambiente de trabalho, pelas discussões construtivas e momentos de descontração.

Agradeço também ao time Empenhados: Gileade, Júlio, Hugo, Victor e Lucas, pela competência, paciência, companhia e dedicação durante nosso projeto.

Aos amigos, que perto ou longe, envolvidos academicamente ou não, estão presentes durante os bons momentos e também durante os desanimadores.

Aos professores João Arthur Monteiro, Leandro Balby e Raquel Lopes, dos quais recebi apoio e esclareci dúvidas durante esta jornada.

Ao Ministério Público da Paraíba, em especial ao Grupo de Atuação Especial Contra o Crime Organizado, representado na pessoa de Octávio Paulo Neto, pela confiança e disponibilização dos dados utilizados neste estudo. Nazareno e eu esperamos que os resultados aqui presentes possam contribuir durante sua incansável jornada.

Aos professores e funcionários da Copin e do Departamento de Sistemas e Computação.

Finalmente, agradeço ao Governo Brasileiro, que através da CAPES, apoiou financeiramente a execução desta pesquisa. Espero que as Talitas que virão tenham a oportunidade de desfrutar de um sistema educacional público, gratuito e de qualidade, assim como eu. E que a educação do país não seja moeda de troca num sistema político falido e desonesto.

A todos, muito obrigada.

Conteúdo

1	Introdução	1
1.1	Motivação	1
1.2	Objetivos e Contribuições	3
1.3	Estrutura do Documento	4
2	Fundamentação teórica	5
2.1	Execução da Despesa Pública	5
2.1.1	Rescisão Contratual	6
2.2	Operacionalização do Risco	7
2.3	SAGRES	9
2.4	Balanceamento de Classes	10
3	Trabalhos relacionados e estado da arte	12
3.1	Estimação de Risco	12
3.1.1	Compras e Contratos	12
3.1.2	Lances e Participantes em Leilão	14
3.1.3	Licitantes	14
3.1.4	Servidores Públicos	15
3.1.5	Unidades Administrativas	16
3.2	Entidades da Administração Pública	16
3.3	Considerações finais	16
4	Materiais e Método	18
4.1	Fontes de Dados	18
4.2	Tratamento, Filtragem e Preparação de Dados	19

4.3	Atributos	20
4.3.1	Risco Associado a Empresas	21
4.3.2	Risco Associado a Contratos	23
4.4	Variável resposta	24
4.5	Métricas para Treinamento e Avaliação dos Modelos	26
4.6	Projeto dos Experimentos	28
4.6.1	Modelos e Parâmetros de Aprendizagem	28
4.6.2	Infraestrutura	29
5	Resultados	31
5.1	Experimentos Sobre Empresas	32
5.1.1	Avaliação da Matriz de Risco	32
5.1.2	Treinamento de modelos com as tipologias da Matriz	34
5.1.3	Modelagem Através de Novos Atributos	35
5.1.4	Considerações Sobre a Modelagem do Risco Associado a Empresas	37
5.2	Experimento Sobre Contratos de Propósito Geral	38
5.2.1	Balanceamento original	38
5.2.2	<i>Undersampling</i>	40
5.2.3	<i>Oversampling</i>	41
5.2.4	SMOTE	43
5.2.5	Considerações Sobre a Modelagem do Risco Associado a Contratos	44
5.3	Experimento Sobre Contratos de Obras e Serviços de Engenharia	45
5.3.1	Balanceamento original	46
5.3.2	<i>Undersampling</i>	47
5.3.3	<i>Oversampling</i>	48
5.3.4	SMOTE	49
5.3.5	Considerações Sobre a Modelagem do Risco Associado a Contratos de Obra e Serviços de Engenharia	51
6	Conclusões	53
6.1	Discussão	53
6.2	Limitações	55

6.3	Trabalhos Futuros	55
A	Fundamentos de Rescisão Contratual da Lei Geral de Licitação	60
B	Mapeamento de Entidades da Administração Pública	64

Lista de Símbolos

AUC - Area Under the Curve

CEAF - Cadastro de Expulsões da Administração Federal

CEIS - Cadastro Nacional de Empresas Inidôneas ou Suspensas

CNEP - Cadastro Nacional de Empresas Punidas

CNCIAI - Cadastro Nacional de Condenados por Ato de Improbidade Administrativa e por Ato que Implique Inelegibilidade

MSE - Erro Quadrático Médio

CGU - Controladoria Geral da União

LGL - Lei Geral de Licitação

MCC - Coeficiente de Correlação de Matthews

MPPB - Ministério Público do Estado da Paraíba

SAGRES - Sistema de Acompanhamento da Gestão dos Recursos da Sociedade

SIASG - Sistema Integrado de Administração de Serviços Gerais

TCE - Tribunal de Contas do Estado

TCE-PB - Tribunal de Contas do Estado da Paraíba

TCU - Tribunal de Contas da União

Lista de Figuras

4.1	Distribuição do sumário das tipologias por classe	22
4.2	Distribuição do valor dos contratos por classe	23
4.3	Distribuição do valor dos contratos relativos a obras por classe	25
4.4	Total de rescisões longo dos anos	25
4.5	Quantidade de rescisões por área	26
5.1	Intervalo de Confiança - Top 10 observações da Matriz de Risco	33
5.2	Intervalo de Confiança - <i>Tree Bag</i>	35
5.3	Intervalo de Confiança - <i>Tree Bag</i>	36
5.4	Intervalo de confiança da diferença entre o desempenho da <i>Tree Bag</i> e do Top 10 da Matriz de Risco	37
5.5	Intervalo de Confiança - <i>Tree Bag</i>	40
5.6	Intervalo de Confiança (<i>Undersampling</i>) - Floresta Aleatória	41
5.7	Intervalo de Confiança (<i>Oversampling</i>) - <i>Random Forest</i>	42
5.8	Intervalo de Confiança (SMOTE) - <i>Tree Bag</i>	43
5.9	Intervalo de confiança da diferença entre o desempenho da Floresta Aleatória e da <i>Tree Bag</i>	45
5.10	Intervalo de Confiança (Dados sem tratamento) - <i>Tree Bag</i>	47
5.11	Intervalo de Confiança (<i>Undersampling</i>) - Floresta Aleatória	48
5.12	Intervalo de Confiança (<i>Oversampling</i>) - <i>Tree Bag</i>	49
5.13	Intervalo de Confiança (SMOTE) - <i>Tree Bag</i>	50
5.14	Intervalo de confiança da diferença entre o desempenho da <i>Floresta Aleató-</i> <i>ria</i> e da <i>Tree Bag</i>	51
B.1	Resumo do processo de contratação pública	64

B.2	Mapeamento da entidade Contrato	65
B.3	Mapeamento da entidade Licitante(s)	65
B.4	Mapeamento da entidade Jurisdicionado (Unidade Administrativa)	66
B.5	Mapeamento do relacionamento entre as entidades Jurisdicionado e Licitante - Conluio	66
B.6	Mapeamento da entidade Servidor Efetivo	67

Lista de Tabelas

5.1	Avaliação da Matriz de Risco	33
5.2	Avaliação dos modelos treinados com as tipologias da Matriz de Risco . . .	35
5.3	Avaliação do conjunto de modelos utilizando como insumo atributos propostos para as empresas descritas na Matriz de Risco	36
5.4	Avaliação dos modelos utilizando como entrada dados de contratos sem balanceamento	39
5.5	Avaliação dos modelos utilizando como entrada dados de contratos com balanceamento através do <i>Undersampling</i>	40
5.6	Avaliação dos modelos utilizando como entrada dados de contratos com balanceamento através do <i>Oversampling</i>	42
5.7	Avaliação dos modelos utilizando como entrada dados de contratos com balanceamento através do SMOTE	43
5.8	Avaliação dos modelos utilizando como entrada dados de contratos de obras sem balanceamento	46
5.9	Avaliação dos modelos utilizando como entrada dados de contratos de obras com balanceamento através do <i>Undersampling</i>	48
5.10	Avaliação dos modelos utilizando como entrada dados de contratos de obras com balanceamento através do <i>Oversampling</i>	49
5.11	Avaliação dos modelos utilizando como entrada dados de contratos de obras com balanceamento através do SMOTE	50

Capítulo 1

Introdução

A grande quantidade de dados e documentos produzidos pela Administração Pública Federal, Estadual e Municipal tem se mostrado como um entrave na promoção do controle e auditoria. Por vezes, distanciando entidades e órgãos governamentais da consecução de seus objetivos devido a escassez de recursos para o manutenção do seu pleno funcionamento.

Na tentativa de aperfeiçoar a governança e gestão públicas, órgãos como o Ministério Público, Tribunal de Contas e Controladoria Geral da União e Estados têm desenvolvido métodos para lidar com a explosão no volume de dados a tratar e para um melhor aproveitamento dos recursos disponíveis. Os efeitos colaterais de iniciativas como esta estão relacionados à geração de economias decorrentes da implementação de estratégias contra risco de descumprimento ou desvirtuação de princípios ou recursos. [14, p. 29].

A seguir são expostas a motivação por trás deste trabalho, seus objetivos e contribuições para o campo da estimação de risco no contexto da Administração Pública (ou Administração), bem como a descrição da estrutura deste documento.

1.1 Motivação

O contrato administrativo é um instrumento através do qual a Administração Pública se vale para aquisição de bens e contratação de serviços [13]. Em geral, esse instrumento é regulado pelas suas próprias condições, pelos preceitos do Direito Público e pela Lei de Licitação [5, Art. 54]. Porém, na prática, não é incomum em que a ineficiência na gestão ou outros incidentes contratuais terminam por findar o vínculo firmado antes da sua conclusão, impos-

sibilitando que o propósito da contratação seja alcançado através da rescisão do contrato.

Se por um lado é de interesse público o manutenção do contrato e finalização da prestação do serviço nos conformes da lei, também é importante que a fiscalização destes contratos ocorra sem onerar em excesso sua execução. Assim, a própria gestão responsável, os órgãos de persecução criminal e cortes de contas têm um papel importante na escolta, investigação e auditoria destes contratos, respectivamente.

No contexto da Administração Pública Federal, a quantidade de contratos celebrados por ano é da ordem de dezenas de milhares, e nestes dígitos sequer estão inclusos os contratos celebrados pelas administrações Estaduais e Municipais. A Paraíba é um estado brasileiro com 223 municípios, e somente nestes municípios, em 2016, foram firmados mais de 18 mil contratos que somados chegam a quase de 3 bilhões de reais¹.

Investigar ou auditar este montante de contratos requer um alto custo tanto de recursos humanos quanto recursos financeiros, tornando impraticável que todo e qualquer um destes documentos seja averiguado em sua totalidade. Como implicação desta limitação, os órgãos encarregados precisam fazer uso de estratégias de seleção e priorização em suas investidas.

Uma vez que todas as atividades de uma organização pública envolvem riscos inerentes ao seu próprio funcionamento e natureza, uma prática comum para controle interno é o gerenciamento de riscos associados às entidades. Assim, a partir da identificação e avaliação do risco, é possível antecipá-lo, monitorar seu acontecimento ou mitigá-lo [14, p. 23].

Proativamente conduzir investigações com o objetivo de recuperar recursos que não tiveram seus fins de direito tem, na maioria das vezes, um custo mais caro do que o próprio recurso [14, p. 34]. Isso faz com que a criação de planos de prevenção e mitigação de riscos sejam de grande valor, tanto para as instituições fiscalizadoras e cofres públicos, quanto para o povo - que é o beneficiário final dos contratos firmados pela máquina pública.

A identificação e priorização de casos a serem apurados em minúcia tem se mostrado um plano de grande valor no projeto e consecução das ações de controle promovidas pelos órgãos responsáveis, como é o caso da previsão do risco de corrupção de servidores públicos [9], previsão do risco de corrupção associado à unidades de gestão públicas [8; 3], estimação do risco associado aos contratos celebrados pela administração pública [28; 25; 26], avaliação do risco do fracionamento de licitações [7], dentre outros.

¹Fonte: SAGRES.

Entretanto, antever situações de risco que impactem o funcionamento e objetivos governamentais não é uma tarefa trivial. Há diversos desafios inerentes à estimação de riscos no contexto deste trabalho, como: entendimento das características e peculiaridades do processo de contratação pública; obtenção, tratamento e cruzamento de dados relacionados relevantes (visto que, em geral, parte dos conjuntos de dados utilizados na área ainda não se encontram disponíveis publicamente); definição e desenvolvimento de métricas que possibilitem a medição dos padrões de interesse; no caso de abordagens utilizando métodos supervisionados de aprendizagem: busca ou construção de bases de dados cujas classes estejam devidamente rotuladas; e mapeamento de entidades cujo risco associado possa ser mensurado.

As técnicas envoltas pela ciência de dados têm sido ferramentas cada vez mais utilizadas para a obtenção, tratamento e extração de informações dos grandes volumes de dados gerados pela máquina pública em diversos contextos [12; 16; 17; 21; 22]. A aplicação destas técnicas no ambiente da Administração Pública tem sido estudada com o objetivo de permitir às unidades de persecução criminal e auditoria praticar o controle de forma mais ativa e eficaz.

Contudo, a partir da literatura pública e do nosso conhecimento até então, avaliamos que há uma escassez de avaliações experimentais que afirmam rigorosamente a eficácia dos métodos empregados na prática na avaliação do risco de entes públicos, ou que comparem modelos, características de dados e parâmetros de aprendizado utilizando aprendizado de máquina. Também há uma lacuna no que diz respeito à estimação de risco a nível Municipal para empresas e contratos, e na aplicação de modelos de estimação de risco em áreas específicas - por exemplo: construção, medicamentos, alimentação escolar, entre outros.

É válido destacar que o presente trabalho foi desenvolvido durante um projeto firmado entre o Laboratório Analytics da Universidade Federal de Campina Grande e o Ministério Público do Estado da Paraíba, cujo objetivo era utilizar técnicas de ciências de dados no auxílio ao combate à corrupção. Durante diversas etapas contamos com contribuições essenciais de Júlio Silva, Gileade Kelvin, Hugo Silva, José Lucas e Victor Andrade.

1.2 Objetivos e Contribuições

Com o objetivo de avançar o estado da arte no uso de aprendizagem de máquina para priorização de contratos e empresas a serem acompanhados ou investigados pelos órgãos com-

petentes, foram desenvolvidas as seguintes atividades no contexto dos municípios do Estado da Paraíba:

- Avaliação experimental da prática (no MPPB) para estimação do risco associado a empresas. Comparação entre esta estratégia e modelos de aprendizado de máquina.
- Avaliação experimental sobre seis modelos para classificação de empresas e contratos em relação ao seu risco.
- Avaliação experimental de três técnicas utilizadas na literatura para mitigar o problema do desbalanceamento severo entre classes no conjunto de treinamento, para estimação do risco associado a contratos.
- Aplicação e avaliação de modelos de aprendizagem de máquina em uma área de competência específica, neste caso, nos contratos relativos a obras e serviços de engenharia.

A partir das atividades descritas, será possível mensurar a eficácia das técnicas atualmente utilizadas e propor melhorias aos órgãos competentes e unidades gestoras municipais. Assim, auxiliando a Administração Pública paraibana no refinamento das estratégias de controle, e também na melhor utilização dos recursos financeiros e humanos disponíveis.

1.3 Estrutura do Documento

O restante deste documento está estruturado da seguinte maneira:

No Capítulo 2 são apresentados os fundamentos teóricos necessários para a compreensão desse trabalho e o contexto no qual está inserido. No Capítulo 3 são descritos e comentados os trabalhos relacionados a estimação de risco em entidades da Administração Pública. Os materiais e métodos utilizados no planejamento e desenvolvimento deste estudo são descritos no Capítulo 4. No Capítulo 5 são expostos os resultados obtidos e discussões geradas. Por fim, no Capítulo 6 apresentamos as conclusões deste trabalho de dissertação, limitações e perspectivas para atividades futuras.

Capítulo 2

Fundamentação teórica

Este capítulo apresenta os fundamentos teóricos utilizados no trabalho desenvolvido. A Seção 2.1 expõe os conceitos e objetivos das licitações e contratos públicos. O tópico 2.2 apresenta a materialização do conceito de risco e como o interpretamos. A seção 2.3 apresenta o Sistema de Acompanhamento da Gestão dos Recursos da Sociedade. Por fim, o item 2.4 discorre sobre técnicas utilizadas para o balanceamento artificial entre classes de conjuntos de dados.

2.1 Execução da Despesa Pública

Datada de 21 de junho de 1993, a Lei Geral de Licitação (LGL) - Lei Nº 8.666 - regimenta o artigo 37 da Constituição Federal Brasileira, e:

"Estabelece normas gerais sobre licitações e contratos administrativos pertinentes a obras, serviços, inclusive de publicidade, compras, alienações e locações no âmbito dos Poderes da União, dos Estados, do Distrito Federal e dos Municípios".

O procedimento administrativo da Licitação é o processo prévio a celebração do contrato administrativo, e tem por objetivo selecionar a proposta mais vantajosa, promover o desenvolvimento nacional e garantir um tratamento igualitário perante a lei (isonomia) para com as pessoas físicas ou jurídicas participantes (chamados licitantes) do referido processo [1].

À parte da administração contratante estão designadas as seguintes tarefas: publicação e divulgação do edital ou convite convocatório aos interessados, recebimento e homologação das propostas, classificação dos licitantes, adjudicação¹ e outros atos intermediários necessários. Por outro lado, os participantes deste procedimento estão responsáveis pela ciência do documento do edital, apresentação de propostas, desistência do certame, apresentação de recursos, impugnações e, por fim, prestação do serviço ou fornecimento do bem constante como objeto da licitação [23]. Além de princípios, a LGL detalha também as modalidades de licitação praticadas, as fases da licitação, os tipos de licitação, tipos de contrato, aspectos da qualificação das empresas concorrentes e também antevê sanções administrativas ou penais para toda e qualquer prática que obstrua as diretrizes estabelecidas na lei [20].

Realizada a reserva orçamentária necessária referente ao valor da contratação (a este passo se dá o nome de empenho) e homologada a licitação, a Administração Pública convocará o licitante vencedor para assinatura do termo de contrato ou equivalente [5, Art. 64]. Com a finalidade de formação de um vínculo, o documento do contrato deve conter com clareza as vontades e obrigações de ambos os órgãos expressas em cláusulas [1]. Segundo a LGL:

"Parágrafo único: Para os fins desta Lei, considera-se contrato todo e qualquer ajuste entre órgãos ou entidades da Administração Pública e particulares, em que haja um acordo de vontades para a formação de vínculo e a estipulação de obrigações recíprocas, seja qual for a denominação utilizada".

Em suas atribuições, a Administração tem o poder e o dever de fiscalizar a execução do contrato. Gerir e acompanhar esta ferramenta ao longo do seu andamento objetiva comprovar o cumprimento dos resultados acordados, bem como prestar apoio em questões trabalhistas e processuais [5, Art. 58, 67].

2.1.1 Rescisão Contratual

O término do contrato administrativo pode acontecer devido a conclusão da prestação do serviço contratado ou fornecimento do bem, rescisão ou anulação. O não cumprimento das

¹Ato judicial que dá a alguém a posse e propriedade de determinados bens, ou mesmo a atribuição da coisa adjudicada ao adjudicante [19].

cláusulas contratuais, especificações ou prazos estabelecidos, lentidão, atraso, irregularidades ou até mesmo a falência da empresa contratada são alguns dos motivos que podem justificar a rescisão contratual. As razões em sua totalidade podem ser observadas na Seção V da LGL: Da Inexecução e da Rescisão dos Contratos ou no Apêndice A deste documento.

Neste trabalho é do nosso interesse estimar as chances de que motivos relativos à rescisão sejam razão para que um, ou pelo menos um, contrato administrativo não se conclua. Para concretizar esta ideia, é necessário tanto determinar características relativas a contratos ou empresas que possibilitem a materialização do risco de rescisão, como a criação de uma base de dados que contenha casos de contratos de alto e baixo risco de rescisão.

2.2 Operacionalização do Risco

Medir o risco associado a algo ou alguém pode ser definido como estimar o grau de incerteza de que um determinado evento ocorra ou não. Por exemplo, quando alguém solicita um empréstimo a uma instituição bancária, esta realiza uma avaliação que resultará se o solicitante receberá ou não o valor a ser emprestado. Esta avaliação pode ser chamada de análise de risco ou análise de crédito, e pode dispor ao seu critério de informações tais como: o histórico do quão bom ou mau pagador o solicitante é, informações sobre seus bens, recebimentos e outras peculiaridades.

No exemplo acima é intuitivo entender o porquê das características utilizadas auxiliarem na associação entre alguém cujo histórico seja de mau pagador a um futuro problema ou prejuízo ao banco. Por conseguinte, a existência do risco, tal como sua intensidade, estão condicionadas a fatores específicos do domínio em questão. Isto é, caso desejemos aplicar a técnica em outros cenários, é necessário observar e investigar quais características deste cenário influenciam para o aumento ou diminuição do risco.

Diversos setores têm feito uso da estimação de risco como forma de evitar perdas e prejuízos [2; 18]. Neste trabalho, o desafio enfrentado é estimar o risco associado a contratos administrativos ou empresas em contratação com a Administração. No Capítulo 4 definiremos as características a partir das quais modelamos o comportamento de risco, adiante elucidamos o que caracteriza um contrato ou empresa de alto risco.

Na prática, os órgãos de persecução criminal e cortes de contas têm diferentes formas de

tornar palpável a abstração do risco para suas entidades. O processo de avaliação do risco de que uma empresa seja fantasma² ou que um contrato tenha um alto risco de corrupção ou fraude, por exemplo, tomará rumo de acordo com o conhecimento prévio e experiência dos executores desta atividade.

No contexto municipal e estadual, o Ministério Público da Paraíba (MPPB), a partir de uma iniciativa do TCE de Pernambuco e do Tribunal de Contas da União (TCU), faz uso de Tipologias de Risco para ordenar empresas no que diz respeito a sua natureza de ameaça. Estas tipologias funcionam tal qual as características sobre o solicitante do empréstimo no exemplo supracitado; elas auxiliam na tipificação do risco da ocorrência de um evento de interesse. A partir da sumarização dos valores de variáveis criadas para medir as Tipologias, é criado um ranking para nortear os esforços de investigação da instituição. A este conjunto de tipologias e sumário nos referiremos como Matriz de Risco. A estratégia descrita é, até onde temos conhecimento, o estado da prática no que diz respeito a estimação de risco para empresas no Estado da Paraíba.

O conceito de risco neste estudo está associado à rescisão do contrato administrativo, por quaisquer motivos (aqui também estão inclusos sustação e impedimento). A partir de uma base de dados sobre contratos dissolvidos entre os anos de 2014 e 2018, rotulamos duas outras bases: uma sobre contratos públicos celebrados entre 2014 e 2017, outra sobre empresas em contrato com a Administração Pública - ambas no contexto do Estado da Paraíba. Assim, fica definido:

- Um contrato de alto risco é todo e qualquer contrato que tenha sido rescindido entre os anos de 2014 e 2018.
- Uma empresa de alto risco é toda e qualquer empresa que tenha, relacionada a si, pelo menos um contrato rescindido entre os anos de 2014 e 2018.

Dessa maneira, o conjunto de dados final é composto por contratos ou empresas representantes de ambas as classes: contratos de alto e de baixo risco. Sobre a categoria de baixo risco, é interessante mencionar que neste caso não significa estar isento de qualquer suspeita de periculosidade, apenas que não foram direcionados esforços de investigação e nem, até então, constam informações sobre atitudes sancionáveis.

²<https://www.investopedia.com/terms/s/shellcorporation.asp>

Levando em consideração a previsão elaborada pelos algoritmos de classificação, definimos que:

- O risco a ser mensurado para cada contrato é o risco da sua rescisão a partir do momento de sua assinatura. Isto é, prevemos o risco de rescisão do contrato a partir do dia em que este foi celebrado.
- Analogamente, o risco associado a uma empresa é dado pela chance de que pelo menos um contrato firmado por ela seja rescindido a partir de uma data estipulada para a sua avaliação.

Outras fontes de dados podem ser utilizadas como formas de operacionalizar a abstração do risco, através de rotulagem das entidades a serem classificadas. Além da utilizada neste estudo, temos conhecimento do Cadastro Nacional de Empresas Inidôneas ou Suspensas (CEIS), Cadastro Nacional de Empresas Punidas (CNEP), Cadastro Nacional de Condenados por Ato de Improbidade Administrativa e por Ato que Implique Inelegibilidade (CN-CIAI), Cadastro de Expulsões da Administração Federal (CEAF) e outros. Entretanto, a depender do contexto alvo do estudo, é importante verificar se a inclusão de informações nestes cadastros não é subnotificada. O CEIS, por exemplo, é comumente utilizado em trabalhos envolvendo risco na esfera Federal, entretanto é uma fonte de dados incompleta em se tratando do contexto Municipal.

2.3 SAGRES

No Estado da Paraíba, o SAGRES³ (Sistema de Acompanhamento da Gestão dos Recursos da Sociedade) é o sistema responsável por armazenar e disponibilizar dados sobre o processo de despesa pública: licitações, contratos, receitas, despesas, dados sobre pessoas físicas e jurídicas, dentre outros. O sistema foi desenvolvido e implementado no Estado da Paraíba pelo Tribunal de Contas do Estado (TCE) no ano de 2002 como uma forma de atender aos requisitos da Lei de Responsabilidade Fiscal, e é alimentado automaticamente pelos órgãos sob a jurisdição do TCE [27].

³<http://tce.pb.gov.br/sagres-online>

O banco de dados que alimenta o SAGRES conta com diversos módulos. Destes, utilizaremos o módulo referente aos poderes públicos dos municípios paraibanos, aqui chamado de SAGRES Municipal. Este módulo conta com cerca de 190 tabelas, responsáveis por representarem entidades da Administração ou seus relacionamentos. A ideia e arquitetura do SAGRES foram replicadas por diversos estados brasileiros, representando uma prática de transparência pública notável.

No Capítulo 4 serão exibidos mais detalhes de como os dados disponíveis no SAGRES vieram a compor as características utilizadas no treinamento dos modelos avaliados.

2.4 Balanceamento de Classes

A identificação de comportamentos atípicos em cenários financeiros utilizando aprendizagem de máquina tem sido uma tarefa amplamente estudada nos últimos anos, por exemplo: fraudes em leilões [2; 12], serviços bancários e seguros [21], fraudes ou corrupções em contratos públicos [26] ou privados [18].

Uma característica em comum entre as bases de dados utilizadas no treinamento destes modelos é o desbalanceamento entre as classes. A exemplo das fraudes em cartão de crédito, transações fraudulentas representam apenas 2% do total. Lidar com conjuntos de dados onde o desbalanceamento entre classes seja tão severo, impacta diretamente nos resultados dos modelos de classificação, prejudicando sua capacidade preditiva [2].

A utilização de técnicas de balanceamento em outros cenários tem demonstrado que algoritmos de classificação obtêm melhor performance quando operam com bases de dados cujas classes possuem tamanhos equivalentes [2]. Tais técnicas podem funcionar de maneira a proporcionar um aumento no número de observações da classe minoritária (*Oversampling*), uma diminuição do número de casos da classe majoritária (*Undersampling*) ou a combinação de ambas as estratégias.

Para atingir o objetivo do *Oversampling*, observações da classe minoritária foram replicadas aleatoriamente até que o número de instâncias das classes seja o mesmo. Na contramão, para balancear os dados reduzindo o tamanho da classe majoritária, a estratégia utilizada foi simplesmente remover aleatoriamente observações da classe de maior tamanho até que ambas as classes tivessem o mesmo tamanho [15].

Interposto entre o *Oversampling* e o *Undersampling*, o método SMOTE (*Synthetic Minority Over-sampling Technique*) foi utilizado para gerar observações da classe menor, como para reduzir a classe majoritária. SMOTE é uma das abordagens mais conhecidas neste campo, e utiliza a distância entre uma observação e seus K (parâmetro definido pelo usuário) vizinhos para interpolar novos pontos gerados sinteticamente [10].

Capítulo 3

Trabalhos relacionados e estado da arte

Neste capítulo será descrita uma visão geral dos trabalhos prévios que abordaram problemas relacionados a estimação de risco de contratos públicos e outras entidades concernentes à Administração Pública. Os trabalhos relacionados foram divididos por tema: Compras, Contratos e Licitações; Lances e Participantes em Leilão; Licitantes; Servidores Públicos e Unidades Administrativas.

Na seção 3.2 é apresentado um resumo sobre as entidades da Administração onde foram concentrados esforços, bem como quais dados foram utilizados (quando disponível).

3.1 Estimação de Risco

A combinação de mineração de dados e aprendizagem de máquina tem se tornado um conjunto de ferramentas popular na detecção e previsão de fraude ou comportamentos anômalos no âmbito financeiro [21; 29]. Administrações de órgãos públicos e privados têm empregado esforços nesta área, incrementando a quantidade de trabalhos envolvendo a utilização de dados históricos e comportamentais de entes com a finalidade de gerar indicadores de risco, irregularidade e outras práticas sancionáveis.

3.1.1 Compras e Contratos

Sales et al. [26] utilizaram dados de contratações e compras governamentais provenientes do Sistema Integrado de Administração de Serviços Gerais – SIASG foram utilizados para

medir o risco associado a contratos públicos pela Controladoria Geral da União (CGU). A rotulação dos dados para treinamento e teste foi feita através de informações contidas no Cadastro Nacional de Empresas Inidôneas ou Suspensas (CEIS). *Naïve Bayes*, *Tree Augmented Naïve Bayes*, *Tabu Search* e *Hill-Climbing* foram os modelos de classificação avaliados, sendo o primeiro escolhido como modelo mais eficiente.

Dados sobre compras de TI no Governo Federal foram usados como entrada do algoritmo *Deep Auto-Encoder* na identificação de compras anômalas [16]. Dentre os 12 modelos avaliados, o de melhor desempenho era composto por 15 camadas escondidas, 5 iterações, função de ativação *Tahn*, e seu Erro Quadrático Médio (MSE) ficou em torno de 0,0013221. O modelo de melhor desempenho foi aplicado nos dados de teste e avaliado o erro na reconstrução de cada registro. A estratégia de validação utilizada foi selecionar os registros com maior erro (MSE) na sua reconstrução e verificar suas características, a fim de priorizá-los ou não em futuras investigações.

No contexto internacional, Grace et al. [18] avaliaram a eficácia dos modelos *Random Forest*, *Logistic Regression*, *AdaBoost*, *Support Vector Classification (SVC)*, *Gradient Boosting Machine (GBM)* e *K-nearest neighbors (KNN)* para previsão de risco de fraude, corrupção e conluio em contratos internacionais. Este estudo foi desenvolvido em parceria com o Banco Mundial. Foram usados características relativas a empresas e contratos, e dados sobre investigações em contratos passados como rótulo. Uma característica interessante desta base de investigações é que foram documentadas tanto investigações que levaram a algum indício de irregularidade, como também investigações onde nada foi encontrado - tem-se, de fato, uma base com ambas as classes. Utilizando a precisão como métrica de avaliação, o classificador construído com *Gradient Boosting* atingiu o melhor resultado.

É comum na literatura que trabalhos cujo objetivo seja estimar o risco de contratos administrativos utilizem dados para rotular as categorias referentes a empresas. No nosso entendimento fazemos distinção entre estas entidades, visto que a cardinalidade associada ao par empresa-contrato é 1:N, ou seja, uma mesma empresa pode estar associada a um ou mais contratos. Dessa forma, estimando o risco de um contrato apontamos para o menor grânulo do processo de contratação cujo risco pode ser mensurado, ao tempo que ao estimar o risco de uma empresa será necessária uma varredura em todos os contratos associados a esta.

3.1.2 Lances e Participantes em Leilão

O pregão eletrônico¹ opera como um leilão inverso e é uma das diversas modalidades da qual a Administração dispõe para licitar a compra de bens ou serviços. A detecção de figurantes (participantes cujo o objetivo é influenciar o resultado do pregão, prejudicando o comprador e/ou vendedor) ou de lances possivelmente fraudulentos é tema abordado na literatura. Esta tarefa consiste na estimação da chance de que um participante do leilão ou um lance sejam irregulares.

Utilizando dados do portal de compras do Governo Federal, o ComprasNet² (atual SI-ASG), foi proposto um conjunto de indicadores relacionados a situações suspeitas para fundamentar a classificação de um lance como possivelmente irregular, partindo de um figurante. Rebouças et. al.[24] propôs e buscou na literatura referente a leilões ordinários, indicadores que pudessem ser portados ao contexto dos pregões eletrônicos. Após atribuir pesos a estes indicadores, gerou um conjunto de dados contendo figurantes e lances artificiais; assim, a partir do cálculo do score dos lances utilizando as características propostas, foi possível mensurar o risco desses e verificar sua autenticidade.

Ainda que proveniente de uma área de estudo distinta, as contribuições trazidas por Anowar et al. [2] podem ser aplicadas ao contexto da Administração. Este estudo avaliou a utilização de técnicas para balanceamento entre classes na identificação de lances possivelmente fraudulentos em leilões do eBay³. A partir de dados sobre o leilão de um videogame, foi empregado agrupamento hierárquico para rotulação dessa base em grupos significativos, e igualados o número de observações de cada classe através de cinco técnicas distintas de balanceamento: SMOTE, SmoteENN, Smote-TomekLink, NearMiss, ClusterCentroids. Por fim, o conjunto de dados gerado foi utilizado como entrada e teste no modelo SVM. Com base na métrica F1, SmoteENN demonstrou melhores resultados.

3.1.3 Licitantes

Empresas que se submetem ao processo de contratação pública (licitantes) também são entidades abordadas no tema da estimação de risco. Apoiados por informações das contratações

¹<https://www.comprasgovernamentais.gov.br/index.php/pregaoeletronico-faq>

²<https://www.comprasgovernamentais.gov.br/>

³<https://www.ebay.com/>

governamentais, *Logistic Regression* e *Decision Tree* foram métodos utilizados para classificação destas empresas em relação ao seu possível grau de irregularidade [25]. Os dados sobre os licitantes foram rotulados a partir do CEIS. Neste trabalho é importante ressaltar que as classes da base de dados final foram balanceadas a partir de uma amostragem aleatória da classe de maior tamanho. Sua contribuição diz respeito a avaliação de características relevantes na distinção entre fornecedores possivelmente inadimplentes ou não, e destaca atributos relativos ao tamanho / porte da empresa como bons discriminantes.

A primeira aplicação de *Deep Neural Networks* no contexto da auditoria governamental [28] comparou o desempenho desta técnica com *Tradicional Neural Networks*, *Logistic Regression* e *Function Discriminant Analysis*. Foram utilizadas características de empresas que firmaram contratos acima de um milhão de reais com a Administração Pública federal, provenientes de sete bases de dados, e dados sobre as penalidades sofridas por essas empresas em caso de atitudes sancionáveis. As classes no conjunto de dados de treino foram balanceadas. O modelo utilizando *Deep Neural Networks* demonstrou melhores resultados, entretanto um entrave nesta técnica é a interpretabilidade de seus resultados. Até onde temos conhecimento, este é o trabalho mais recente abordando o problema do risco em entidades governamentais no Brasil.

3.1.4 Servidores Públicos

Com o objetivo de prever o risco de corrupção de servidores públicos federais, foram utilizados dados referentes a filiação partidária de servidores públicos [9]. Este trabalho proporciona uma comparação entre os modelos de classificação utilizados e um modelo construído na prática por especialistas da área. Acreditando na ligação moral entre as esferas políticas e públicas, os autores deste artigo fazem uso dos modelos *Bayesian Networks*, *Support Vector Machines (SVM)*, *Random Forest* e *Artificial Neural Networks with backpropagation* para classificação. Para rotular seus dados de treinamento, foram usados registros disponíveis no sistema de administração de pessoal do governo, e filtrados os casos em que servidores foram expulsos devido a corrupção. *Random Forest* foi o modelo que demonstrou melhores resultados - ainda que sua precisão tenha sido semelhante ao modelo utilizado pelos especialistas (86%), a revocação obteve um valor aproximadamente duas vezes mais alto (32%).

3.1.5 Unidades Administrativas

Num intento de otimizar estratégias de controle, a CGU utilizou modelos probabilísticos (*Naïve Bayes*, *Tree Augmented Naïve Bayes*, e *Attribute Weighted Naïve Bayes*) para estimar o risco de corrupção em unidades administrativas federais [8]. Utilizando dados sobre as unidades administrativas e sobre servidores expulsos devido a corrupção foi possível criar uma base de treinamento composta por unidades corruptas ou não corruptas. O modelo cuja performance se destacou foi o *Naïve Bayes*, atingindo 0,76 em sua métrica de interesse, AUC (*Area Under the Curve*).

Já no Tribunal de Contas da União (TCU) foi testada uma estratégia cujo objetivo era estimar o risco combinado entre unidades governamentais e empresas privadas na execução de contratos [3] (corrupção e conluio). *Naïve Bayes* foi utilizado numa abordagem não-supervisionada, e a validação da solução proposta foi feita através do conhecimento e opinião de especialistas na área.

3.2 Entidades da Administração Pública

Compreender o processo de despesa pública não é tarefa simples. Por essa razão, em adição a esta revisão de literatura produzimos diagramas com o objetivo de mapear para quais entes da Administração já foram empregados esforços de estudo, bem como de onde partiram os dados para realização dos respectivos trabalhos. Este mapeamento conceitual está descrito no Apêndice B deste documento.

3.3 Considerações finais

Os trabalhos citados representam grande contribuição para o estado da arte tanto na área computacional, como no controle social. Neste último, auxiliando as entidades fiscalizadoras a migrarem para uma postura mais proativa, antevendo casos de corrupção, fraude, rescisões e outras atitudes que possam levar a prejuízos futuros.

Como observado, no panorama do Brasil, todos os trabalhos apresentados tratam da estimativa do do risco - em algum ente - na esfera federal. Até onde temos conhecimento, não há evidências de uma avaliação experimental da prática na estimativa do risco em contratos

e empresas em contratação com a administração municipal do Estado da Paraíba, bem como não há registro da utilização de aprendizagem de máquina neste contexto.

Neste trabalho desejamos abordar os seguintes aspectos: i) Avaliação experimental do método utilizado na prática pelo MPPB para estimação de risco de empresas; ii) Proposição de características e de um modelo para estimação de risco para os contratos celebrados no âmbito dos municípios do Estado da Paraíba; iii) Avaliação experimental na modelagem do risco de contratos de uma área de competência específica, neste caso: obras públicas.

É importante mencionar que trabalhos acima que tangem a área pública brasileira [3; 8; 9; 16; 24; 25; 26; 28], contaram com a participação direta de, pelo menos, um órgão ou membro de órgão governamental na sua concepção. Este estudo contou com a generosa colaboração do Ministério Público da Paraíba e Tribunal de Contas do Estado da Paraíba, todavia foi desenvolvido e orientado por membros da sociedade civil.

Capítulo 4

Materiais e Método

Neste capítulo são detalhados os dados e métodos usados em nossos experimentos.

4.1 Fontes de Dados

Neste trabalho utilizamos diferentes conjuntos de dados sobre as empresas prestadoras de serviço à Administração Pública e sobre os contratos firmados entre estas duas entidades. As fontes de dados foram obtidas em parceria com o MPPB e o TCE-PB, e a seguir são listadas juntamente a suas descrições:

- **Matriz de Risco** - Representa o estado da prática atualmente em ao menos dois órgãos cuja realidade conhecemos. Ou seja, esses são dados relacionados à forma como o MPPB e outros órgãos atualmente mensuram o risco associado a empresas. Esta ferramenta foi desenvolvida com o intuito de orientar os esforços de investigação em fraudes e outras atitudes sancionáveis, e reúne atributos estratégicos sobre mais de 29.000 pessoas jurídicas em contrato com os municípios (e suas respectivas unidades administrativas) da Paraíba. O código fonte para geração da Matriz foi produzido em SQL (Linguagem de Consulta Estruturada, em tradução livre), codificando em uma linguagem de programação características de empresas que os órgãos reconhecem como suspeitas. Isto é, os atributos contidos nesse banco se baseiam no conhecimento prévio de servidores sobre o processo de contratação, composição societária, características cadastrais e de contratação de funcionários, entre outros dados.

- **SAGRES** - Como mostrado na Seção 2.3, o SAGRES contém dados sobre a atuação das empresas, características de licitações, empenhos e contratos relacionados, dados cadastrais de servidores e fornecedores, pagamentos e outros documentos decorrentes do processo de despesa da máquina pública estadual e municipal na PB. Somente o módulo do SAGRES municipal beira os 73 GB, divididos em 294 tabelas.
- **Cadastro de Empresas** - A Receita Federal do Brasil é responsável por manter um cadastro de todas as empresas em operação em todo o Brasil. Neste cadastro estão contidos dados como: data de abertura, porte da empresa e situação do cadastro. Neste estudo filtramos apenas dados sobre empresas em contratação com a Administração paraibana.
- **Obras** - Ainda que o SAGRES também contenha informações sobre as obras públicas do estado, o TCE-Paraíba dispõe de um banco de dados especializado sobre o tema. Esta base reúne informações sobre o georreferenciamento e status da obra, valores repassados, andamento do contrato e outros detalhes documentados e armazenados no banco de dados por especialistas. Uma visualização desses dados está disponível no sistema Painel de Obras¹.
- **Tramita** - Este sistema² do TCE-Paraíba permite acompanhar a situação de tramitação dos processos nos municípios do estado. A base à qual tivemos acesso é referente às rescisões contratuais entre os anos de 2014 a 2018.

A partir de filtros, transformações e cruzamentos nas bases acima citadas foram construídos os conjuntos de dados para análise ou treinamento dos modelos de classificação. Nas seções a seguir são esclarecidos os detalhes sobre como os arquivos de dados foram gerados e quais suas finalidades.

4.2 Tratamento, Filtragem e Preparação de Dados

Na maioria das vezes a inserção de dados em sistemas como o SAGRES é manual e feita pelas próprias unidades administrativas de cada município, estando sujeita a erros e incon-

¹<http://paineldeobras.tce.pb.gov.br/>

²<https://tramita.tce.pb.gov.br/tramita/pages/main.jsf>

sistências. Devido à diversidade de origens e formatos de dados em que foram recebidos os dados, foi necessário converter, tratar e filtrá-los para que pudessem ser utilizados neste estudo.

Originalmente, o SAGRES é mantido em um banco de dados Oracle³, e a cópia à qual tivemos acesso dispunha deste mesmo formato. Foi necessário realizar a migração dos dados para um sistema gerenciador de banco de dados não proprietário que pudesse ser utilizado por nós. O eleito foi o MariaDB⁴ 5.5.58.

Também em relação ao SAGRES foram: criados ou corrigidos índices de tabelas inconsistentes; removidos contratos, licitações ou empenhos cujo código de identificação fossem inválidos, bem como removidas empresas cujo CNPJ fosse inválido.

Neste trabalho estudamos o risco associado a duas entidades da Administração Pública: empresas e contratos. Para uma melhor compreensão, nas Seções 4.3.1 e 4.3.2 detalhamos separadamente, por finalidade, como foram gerados os conjuntos de dados que permitiram a estimação e avaliação do risco das referidas entidades. Na Seção 4.3 é brevemente exposto o processo de criação dos atributos de treinamento.

4.3 Atributos

Como visto no Capítulo 3, alguns trabalhos descrevem os atributos propostos ou utilizados como entrada em seus modelos, enquanto outros apenas descrevem o contexto no qual estão inseridos a fim de manter sigilo sobre como é mensurado o risco das entidades alvo. Visto que este é o primeiro trabalho do qual temos conhecimento no contexto da estimação de risco utilizando aprendizagem de máquina para entidades municipais na Paraíba, não detalharemos os atributos utilizados um por um, a fim de resguardar informações estratégicas dos órgãos de controle.

A primeira referência de características que pudessem ser úteis como indicadores de risco para empresas e contratos veio da literatura, visto que há trabalhos [25; 26] que descrevem os atributos utilizados e avaliam sua importância na modelagem do problema. Em seguida, a partir da replicação do trabalho desenvolvido por Sun et al. [28] na estimação de risco

³<https://www.oracle.com/br/database/>

⁴<https://mariadb.org/>

de contratos na Esfera Federal, foi possível verificar a importância de atributos e bases de dados do mais recente trabalho na área do qual temos notícia, e elencar quais características eram relevantes e existentes em nosso contexto. Com base no conhecimento dos auditores e promotores com os quais tivemos contato e da nossa própria experiência ao longo do estudo, também foram propostos atributos.

Fundamentados nesta revisão e criação de variáveis preditoras do risco, o próximo passo foi verificar quais eram possíveis de serem extraídas das fontes de dados disponíveis.

Os atributos gerados relativos a empresas são 23 e compreendem seu comportamento licitatório, atuação geográfica, aspectos cadastrais e composição societária. No caso dos atributos relativos a contratos, que são 30, além das características anteriormente citadas que remetem às empresas prestadoras dos contratos, foram adicionadas características inerentes aos próprios contratos, como o tipo da licitação que gerou aquele contrato, valor total do contrato, entre outros. Por fim, no caso específico dos contratos de obras e serviços de engenharia, foram adicionados três características próprias desta área de atuação e removida uma, totalizando 32 atributos.

4.3.1 Risco Associado a Empresas

Para estimar e avaliar o risco associados a empresas através de algoritmos de classificação de abordagem supervisionada, foi necessário utilizar bases de dados contendo informações sobre empresas cuja atuação foi correta, e empresas que tiveram registro de problemas durante a execução de seu contrato.

Matriz de Risco

Como visto na Seção 1.2, nosso primeiro objetivo é avaliar a eficácia do método de estimação de risco utilizado na prática pelo MPPB e outros órgãos (Matriz de Risco).

O arquivo da Matriz de Risco foi gerado em 17 de junho de 2017 e é composto por pouco mais de 29.000 observações, com 62 variáveis, dentre as quais 57 são tipologias de risco. Cada observação descreve a atuação de uma empresa nos seus aspectos históricos, cadastrais e políticos.

Um ranking do risco associado às empresas pode ser gerado a partir da ordenação da

variável que representa o sumário (soma ponderada) de todas as tipologias numéricas (Total Tipologias). Para os fins desta avaliação, selecionamos 15.000 empresas deste ranking, do maior para o menor valor de risco atrelado.

Após este filtro, os dados resultantes foram combinados com a base de rescisões contratuais, para que fosse possível rotular as empresas em relação ao seu risco real. Por fim, temos um dado composto de 14.730 empresas de baixo risco (sem rescisões segundo o Tramita) e 270 de alto risco (com rescisões no Tramita). Este primeiro arquivo gerado será empregado na avaliação da Matriz de Risco (Seção 5.1.1), e também como insumo para treinamento dos modelos de classificação (Seção 5.1.2).

Na Figura 4.1, onde cada ponto representa uma empresa, é possível observar a distribuição do sumário de tipologias em relação às duas classes: alto e baixo risco. A diferença da quantidade de observações entre as classes mostra que o desbalanceamento é uma característica evidente deste conjunto de dados.

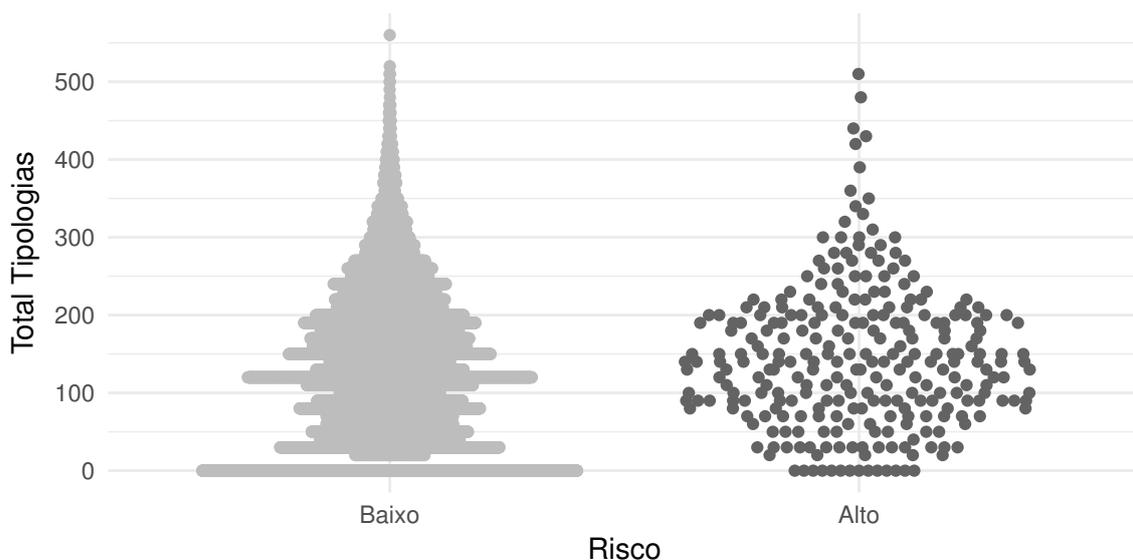


Figura 4.1: Distribuição do sumário das tipologias por classe

Recaracterizando as Empresas da Matriz de Risco

Após a avaliação da Matriz de Risco, também era do nosso interesse verificar a eficácia das tipologias utilizadas como indicadores de risco. Com esse intuito, geramos um novo arquivo contendo o conjunto de atributos descrito na Seção 4.3 para as mesmas empresas presentes

na Matriz, composto por 7.485 observações (7.227 com baixo e 258 com alto risco), após remoção de observações incompletas, e 23 atributos. Com esse segundo conjunto de dados é possível avaliar a eficácia de modelos de aprendizagem que usam como insumo as mesmas variáveis da matriz de risco e ao mesmo tempo compara esta eficácia com a de modelos informados por outras variáveis que propomos (Seção 5.1.3).

4.3.2 Risco Associado a Contratos

Contratos de Propósito Geral

Como visto, nosso segundo objetivo é estimar o risco associado a contratos administrativos. Para isto, foi realizada uma varredura na base do SAGRES e filtrados os contratos municipais iniciados entre os anos de 2014 e 2017. Em seguida, para cada contrato foi criado um conjunto de atributos proposto por nós e descrito na Seção 4.3. O resultado é um arquivo contendo 12.567 observações (12.428 da classe de baixo risco, 139 de alto risco) e 30 atributos, onde cada observação representa um contrato distinto, que foi rotulado a partir da base de contratos rescindidos.

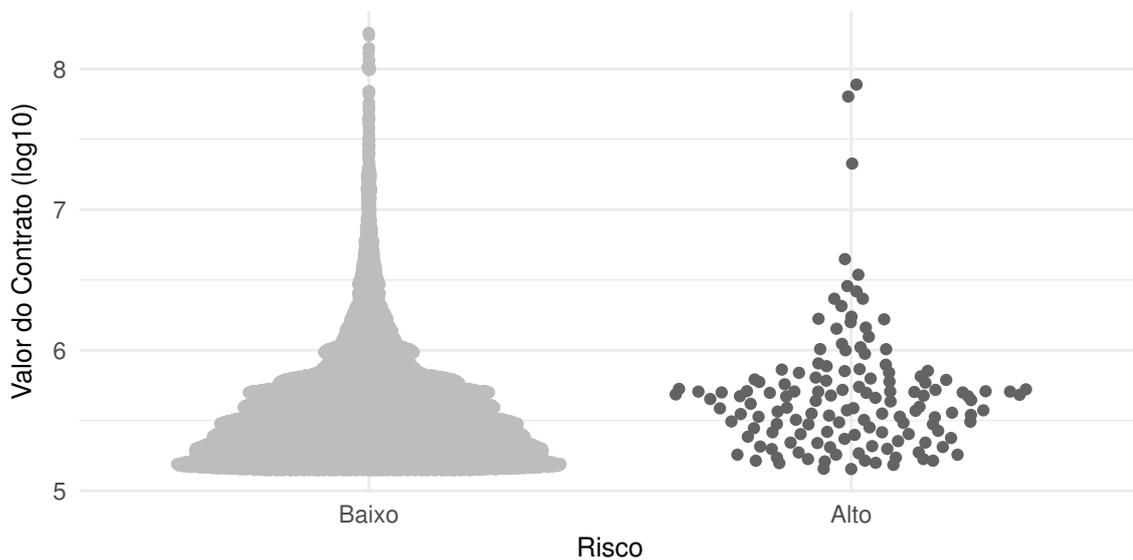


Figura 4.2: Distribuição do valor dos contratos por classe

Com base em uma análise de correlação e variância dos atributos numéricos, foram removidos aqueles cuja informação fosse redundante (correlação igual ou superior a 0.9) e também cuja variância fosse próxima a zero (pouca relevância na modelagem). Devido ao

viés nas suas distribuições, estas características também passaram pela transformação de Yeo Johnson, que busca aproximar o formato das distribuições ao formato normal⁵, o que pode melhorar o desempenho de alguns modelos de classificação. Por fim, o arquivo resultante é composto de 12.567 observações, 20 atributos e será utilizado como insumo dos modelos de classificação (Seção 5.2).

Contratos de Obras e Serviços de Engenharia

Com o objetivo de verificar o desempenho da classificação de contratos em uma área de competência específica, selecionamos contratos da área de obras e serviços de engenharia. Foram filtrados contratos cujo valor fosse igual ou superior a 140.000 reais, visto que o limite legal⁶ vigente até o ano de 2017 para que empresas fossem contratadas via modalidade convite era de 150.000 reais⁷. Essa filtragem se dá pelo fato de que, por vezes, no intento de fraudar o processo licitatório, governantes fracionam licitações de obras em parcelas menores, para que se encaixem nos requisitos da modalidade convite. Dessa forma, em lugar da ampla concorrência entre os interessados em contratar, empresas em conluio com a gestão são favorecidas, ferindo o princípio de igualdade da licitação.

A rotulagem destes contratos foi feita utilizando a base de rescisões e também informações sobre o estado de cada obra - obras abandonadas, encerradas judicialmente ou inacabadas também foram consideradas de alto risco. Após filtragens, o quarto arquivo produzido contém 505 observações (297 contratos de baixo risco, 208 de alto risco) e 32 atributos.

Neste contexto, também aplicamos as análises de correlação, variância e distribuição realizadas na seção anterior, resultando um arquivo com 505 observações e 23 atributos, que também será utilizado como insumo do conjunto de modelos de classificação (Seção 5.3).

4.4 Variável resposta

Como descrito na Seção 2.2, o conceito de risco discutido neste estudo está ligado à rescisão do contrato administrativo. As rescisões disponíveis na base de dados utilizada podem ter

⁵<https://www.rdocumentation.org/packages/VGAM/versions/1.1-1/topics/yeo.johnson>

⁶Valores limites atualizados em 19 de junho de 2018 via Decreto Nº 9.412/2018.

⁷<http://www.planejamento.gov.br/noticias/decreto-atualiza-valores-para-licitacoes-e-contratos>

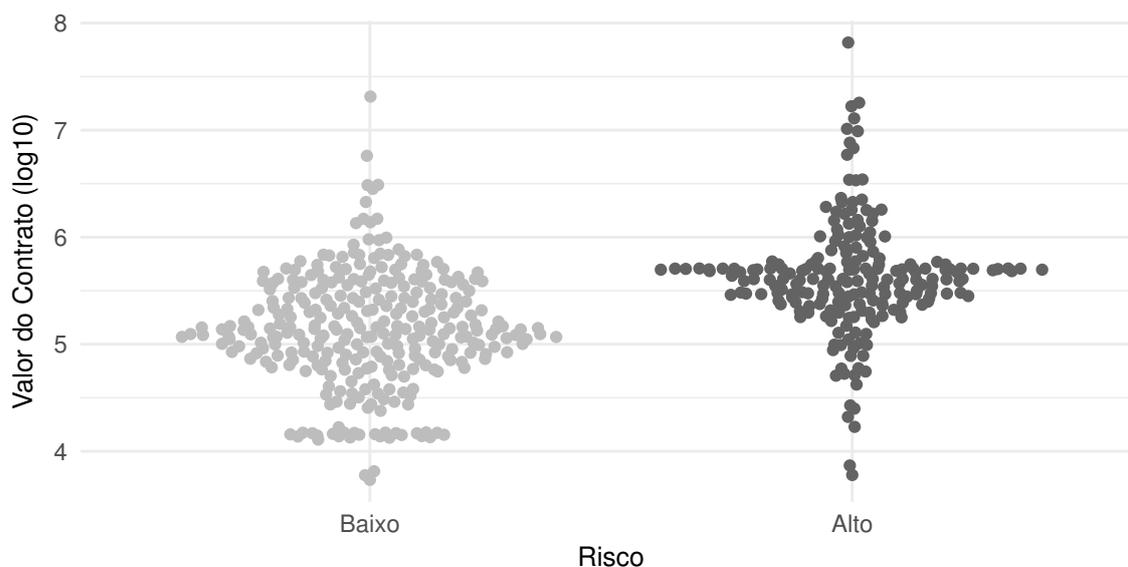


Figura 4.3: Distribuição do valor dos contratos relativos a obras por classe

ocorrido por vontade de ambas as partes: do jurisdicionado contratante ou do contratado. É possível que a rescisão do contrato também se dê através do seu impedimento ou sustação.

Na Figura 4.4, observa-se que a quantidade de rescisões contratuais cresce ao longo dos anos até 2017, o último ano completamente representado nos dados.

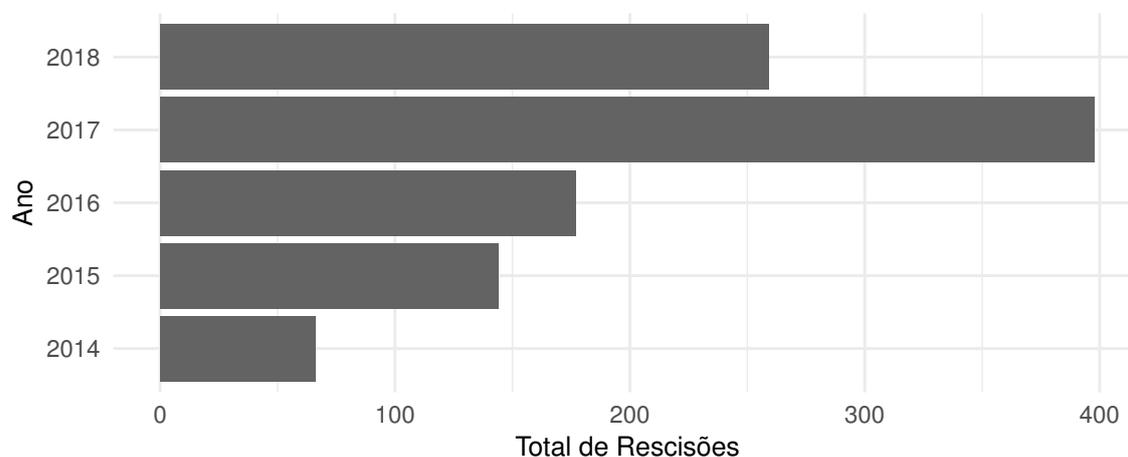


Figura 4.4: Total de rescisões longo dos anos

Cada rescisão é cadastrada noTramita com uma justificativa. As justificativas em nosso conjunto de dados incluem: desistência do licitante, falta de condições da empresa contratada de efetuar os serviços determinados, atraso no fornecimento, descumprimento das cláusulas contratuais, óbito do contratado, impedimento de contratar com a união, estados

ou municípios; vontade de ambas as partes e outras.

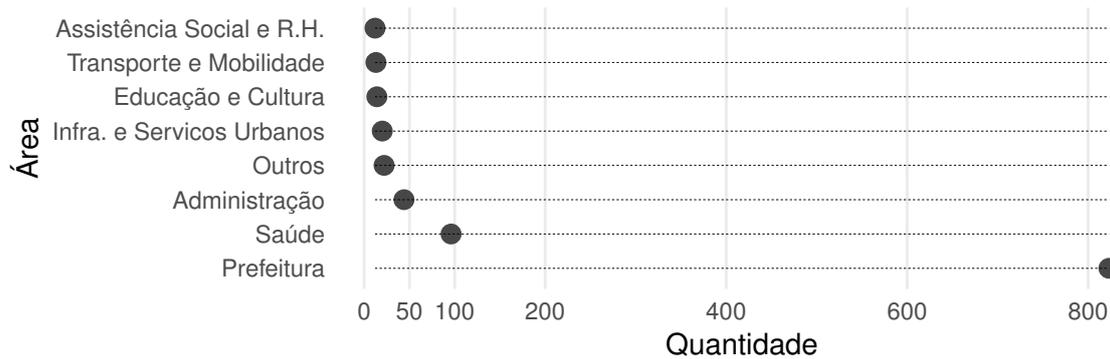


Figura 4.5: Quantidade de rescisões por área

De maneira manual categorizamos por área os jurisdicionados presentes na base de rescisões. Exemplos de jurisdicionados são: Fundos Municipais, Prefeituras Municipais, Secretarias, Consórcios, Câmaras Municipais, Departamentos e outros. Devido à grande quantidade de rescisões nas Prefeituras Municipais, optamos por separá-las do restante da Administração, como observado na Figura 4.5.

4.5 Métricas para Treinamento e Avaliação dos Modelos

Na literatura, diversos trabalhos que abordam a detecção de fraudes ou eventos que acontecem com significativa menor proporção ressaltam a importância da utilização de métricas apropriadas para a avaliação de modelos nesse cenário.

Em cenários de severo desbalanceamento, as métricas mais comumente utilizadas para avaliação de modelos como precisão, revocação, acurácia e F1 podem levar a conclusões problemáticas.

Para ilustrar, lembramos que considerando as quatro classes da matriz de confusão, Verdadeiros Positivos (VP), Verdadeiros Negativos (VN), Falsos Positivos (FP) e Falsos Negativos (FN), temos:

$$Precisao = \frac{VP}{VP + FP} \quad (4.1)$$

$$Revocacao = \frac{VP}{VP + FN} \quad (4.2)$$

$$Acuracia = \frac{VP + VN}{VP + VN + FP + FN} \quad (4.3)$$

Por exemplo, considerando um cenário hipotético de 10 empresas, sete de baixo risco e três de alto risco:

Caso o classificador considere que todas as empresas possuam um alto risco associado a si (gerando Falsos Positivos), sua revocação será máxima (1), enquanto a precisão e acurácia serão iguais a 0,3. Este cenário é desvantajoso visto que sete empresas seriam alvos desnecessários de investigação - desperdiçando recursos humanos e financeiros do Estado.

Caso o classificador considere todas as empresas possuam um baixo risco associado a si (gerando Falsos Negativos), sua revocação será 0. Entretanto é gerado um reflexo positivo na acurácia (0,7), dando uma falsa impressão de eficiência. A desvantagem desta cenário é clara, pois como nenhuma empresa é recomendada para investigação, as que venham a possuir atitudes sancionáveis passariam impunes.

Com visto, fica claro que modelos cujas classificações produzam significativas quantidades tanto de Falsos Positivos como de Falsos Negativos são custosas. A medida F1, definida como a média harmônica entre a precisão e a revocação, inclui tanto FN com FP em seu cálculo e foi utilizada com parâmetro de treinamento na escolha dos hiperparâmetros dos modelos.

$$F1 = \frac{2 * VP}{2 * VP + FP + FN} \quad (4.4)$$

Para avaliação do desempenho dos modelos no conjunto de dados de teste entendemos que era necessário uma medida cujo desempenho não fosse afetado pelo severo desbalanceamento entre as classes. O Coeficiente de Correlação de Matthews (MCC) é citado na literatura como uma métrica consistente e adequada neste caso [4; 11].

$$MCC = \frac{VP * VN - FP * FN}{\sqrt{(VP + FP)(VP + FN)(VN + FP)(TN + FN)}} \quad (4.5)$$

Ainda que em alguns tratamentos tentemos mitigar os efeitos do desbalanceamento, através de técnicas como *Oversampling* e *Undersampling*, é preciso atentar para o uso de métricas de avaliação robustas ao desequilíbrio entre classes. O MCC pode assumir valores num

intervalo de $[-1$ a $+1]$, o que assemelha-se a uma correlação entre o conjunto de observações previsto pelo classificador e o conjunto de observações utilizado como rótulo.

4.6 Projeto dos Experimentos

Com o objetivo de facilitar a compreensão das decisões metodológicas tomadas, nos referiremos ao treinamento e avaliação de modelos com dados de empresas e contratos administrativos como Experimentos Sobre Empresas e Experimentos Sobre Contratos. Em ambos os casos enfrentamos um problema de classificação, visto que desejamos classificar empresas ou contratos quanto ao seu risco associado.

Os arquivos de dados experimentais foram divididos em duas partes: uma para treinamento dos modelos e outra para avaliação destes modelos em um conjunto ainda não visto - aos quais nos referiremos como conjuntos de treino e teste. Os conjuntos de treino e teste foram compostos de amostras aleatórias de 80% e 20% dos dados originais, respectivamente.

Nos Experimentos Sobre Contratos foram aplicados três tratamentos no conjunto de dados de treino, com o objetivo de comparar a efetividade das técnicas de balanceamento de classes: *Undersampling* e *Oversampling* aleatório e SMOTE. Os conjuntos de testes não sofreram nenhuma alteração visto que deveriam refletir o cenário real de desbalanceamento de classes.

4.6.1 Modelos e Parâmetros de Aprendizagem

Utilizamos seis algoritmos de aprendizagem supervisionada: Árvore, Floresta Aleatória, *Tree Bag*, Rede Neural, Regressão Logística e *Stochastic Gradient Boosting* (SGB). A seguir serão detalhados os parâmetros de aprendizagem de cada modelo, bem como suas especificações.

O treinamento de todos os modelos foi realizado com a biblioteca *caret*, implementada na linguagem R. A biblioteca implementa uma interface uniforme para a busca de quais valores cada parâmetro dos modelos deve assumir. Para cada combinação de parâmetros o método de treinamento realiza uma validação cruzada com 5 *folds*; para cada um dos *folds* é computada a métrica de performance definida, F1 (vide Seção 4.5). Por fim, a média do F1 é computada através de cada um dos *folds*, e a combinação de parâmetros que tenha a melhor

média da métrica avaliativa é considerada como os melhores parâmetros para o modelo.

A arquitetura de Rede Neural que utilizamos dispõe de uma camada de neurônios escondida. Os hiperparâmetros de aprendizado a serem definidos são *size* - que ajusta a quantidade de neurônios existentes na camada escondida - e *decay* - um parâmetro de regularização que define o fator de decaimento dos pesos de cada neurônio. O primeiro assumiu os valores 1, 3 e 5, enquanto o segundo 0; 0,1; 0,0001. A biblioteca que implementa este algoritmo é *nnet*⁸.

Para os modelos baseados em árvores, o m_{try} - 2, 14 e 26 - é o parâmetro de ajuste para a Floresta Aleatória e determina a quantidade de variáveis disponíveis para ramificação das árvores em cada nó. No caso do SGB a profundidade assumiu os valores 1, 2 e 3; enquanto o número de árvores (*ntrees*) ficou entre 50, 100 e 150. Para a Árvore e *Tree Bag* não há necessidade do ajuste de nenhum parâmetro de aprendizado. As bibliotecas que implementam esses modelos são: *rpart*⁹, *randomForest*¹⁰ e *gbm*¹¹.

4.6.2 Infraestrutura

Durante a fase de pré-processamento de dados a ferramenta utilizada foi a linguagem e ambiente de computação R (versão 3.5.1). Ainda na fase de pré-processamento foram utilizadas as bibliotecas *RPostgres*¹², *DBI*¹³, *Config*¹⁴, *RMySQL*¹⁵, *Tidyverse*¹⁶, *Here*¹⁷ e *Tibble*¹⁸. Durante a avaliação e modelagem foram utilizadas as bibliotecas *Caret*¹⁹, *Tidyverse*, *MCCR*²⁰ e *Unbalanced*²¹. Para geração das visualizações foram utilizados as bibliotecas *Tidyverse*, *gg-*

⁸<https://www.rdocumentation.org/packages/nnet/versions/7.3-12>

⁹<https://www.rdocumentation.org/packages/rpart/versions/4.1-15>

¹⁰<https://www.rdocumentation.org/packages/randomForest/versions/4.6-14>

¹¹<https://www.rdocumentation.org/packages/gbm/versions/2.1.5>

¹²<https://www.rdocumentation.org/packages/RPostgres/versions/1.1.1>

¹³<https://www.rdocumentation.org/packages/DBI/versions/0.5-1>

¹⁴<https://www.rdocumentation.org/packages/config/versions/0.3>

¹⁵<https://www.rdocumentation.org/packages/RMySQL/versions/0.10.17>

¹⁶<https://www.rdocumentation.org/packages/tidyverse/versions/1.2.1>

¹⁷<https://www.rdocumentation.org/packages/here/versions/0.1>

¹⁸<https://www.rdocumentation.org/packages/tibble/versions/1.4.2>

¹⁹<https://www.rdocumentation.org/packages/caret/versions/6.0-84>

²⁰<https://www.rdocumentation.org/packages/mccr/versions/0.4.4>

²¹<https://www.rdocumentation.org/packages/unbalanced/versions/2.0>

plot2²², GridExtra²³ e Ggbeeswarm²⁴. Para cálculo e geração dos intervalos de confiança utilizando *bootstrap* foram utilizadas as bibliotecas Broom²⁵ e Boot²⁶.

²²<https://www.rdocumentation.org/packages/ggplot2/versions/3.2.0>

²³<https://www.rdocumentation.org/packages/gridExtra/versions/2.3>

²⁴<https://www.rdocumentation.org/packages/ggbeeswarm/versions/0.5.3>

²⁵<https://www.rdocumentation.org/packages/broom/versions/0.5.2>

²⁶<https://www.rdocumentation.org/packages/boot/versions/1.3-22>

Capítulo 5

Resultados

Este capítulo apresenta resultados relacionados a nossos experimentos em dois contextos: empresas e contratos. A diferenciação entre estas entidades, neste trabalho, se dá por acharmos que prever o risco do contrato é uma maneira menos onerosa de prever riscos. Visto que, a partir da previsão do risco da empresa, seria necessário revisar todos os contratos desta para verificar possíveis suspeitas.

O primeiro contexto compreende três experimentos: o primeiro avalia a eficácia da Matriz de Risco em relação a sua capacidade de detectar empresas cujo risco associado seja alto. O segundo avalia o desempenho de seis modelos de classificação utilizando como entrada as tipologias de risco da própria Matriz. Por fim, o terceiro, utilizando as mesmas empresas presentes na Matriz de Risco caracterizadas por outros atributos, avalia a eficácia de seis modelos na diferenciação do risco de empresas nos municípios do Estado da Paraíba.

No segundo contexto, avaliamos o desempenho de seis modelos de classificação, e de três técnicas para o balanceamento artificial de classes em dois domínios: contratos de propósito geral e contratos específicos de obras e serviços de engenharia.

O desempenho dos modelos em geral foi mensurado, como descrito na Seção 4.5, através do MCC. A partir disso, os dois melhores modelos em cada ensaio foram comparados por meio do intervalo de confiança da diferença entre os valores de MCC.

Os intervalos de confiança foram calculados utilizando *bootstraps* com os seguintes parâmetros: 4 mil reamostras e nível de confiança de 95%. O método utilizado para computar o intervalo a partir dos bootstraps foi o *BCa* [6], exceto nos casos onde este não convergiu (Seções 5.1.1, 5.1.4 e 5.2.5); nestes casos foi utilizado o método *basic*. Analogamente, o

intervalo de confiança da diferença foi calculado com base na diferença entre o MCC dos modelos avaliados em cada caso.

Também é importante registrar que, durante o treinamento de diversos modelos, devido a pequena quantidade de contratos rotulados como alto risco, alguns conjuntos na reamostragem dos dados realizadas pela validação cruzada não continham nenhuma observação da classe minoritária de contratos, resultando em valores não existentes para a métrica de interesse do treinamento. Assim, foi possível considerar apenas as amostras onde havia uma métrica avaliativa válida.

5.1 Experimentos Sobre Empresas

Neste primeiro experimento iremos estimar e avaliar o risco associado às empresas em contratação com a Administração Pública.

5.1.1 Avaliação da Matriz de Risco

Como visto, a Matriz de Risco é o método criado e utilizado pelo TCE de Pernambuco e TCU, e utilizado atualmente pelo MPPB para mensuração do risco associado a empresas.

Por se tratar de um ranking, optamos por avaliar a eficácia da Matriz em relação a empresas possivelmente suspeitas nas primeiras 10, 25, 50, 100, 300, 500 e 1.000 observações. Durante esta avaliação é importante lembrar a natureza de severo desbalanceamento entre as classes aqui presentes: das 15.000 empresas listadas, apenas 270 (menos de 2%) foram rotuladas como alto risco.

Na Tabela 5.1 foram listadas as métricas avaliativas calculadas para os sete conjuntos de empresas selecionados.

Se interpretarmos a Matriz de Risco como um classificador, a partir da avaliação do MCC - que em todos os casos é próximo ou igual a zero - é possível dizer que o desempenho da Matriz de Risco para identificar empresas que terão contratos rescindidos é próximo ao de um palpite aleatório. O valor da Área Sob a Curva (AUC), que varia de 0 a 1, também traz informações sobre a capacidade dos modelos em diferenciar as classes. Neste caso, os valores próximos a 0,5 sugerem dificuldade na separação entre empresas de baixo e alto

Modelo	F1	AUC	Precisão	Revoc.	MCC	VP	VN	FP	FN
Top 10	0,014	0,503	0,200	0,007	0,035	2	14722	8	268
Top 25	0,014	0,503	0,080	0,007	0,019	2	14707	23	268
Top 50	0,025	0,506	0,080	0,015	0,027	4	14684	46	266
Top 100	0,027	0,506	0,050	0,019	0,020	5	14635	95	265
Top 300	0,028	0,505	0,027	0,030	0,009	8	14438	292	262
Top 500	0,026	0,502	0,020	0,037	0,003	10	14240	490	260
Top 1.000	0,028	0,500	0,018	0,067	0,000	18	13748	982	252

Tabela 5.1: Avaliação da Matriz de Risco

risco; uma interpretação consoante à proporcionada pelo MCC.

Considerando os dados como uma amostra e estimando o valor do MCC para a população de empresas postas a classificação, estimamos que o MCC na população estaria no intervalo exibido na Figura 5.1. O fato de que o intervalo engloba valores negativos, evidencia que é possível que a classificação desse recorte de dados para a população tenha um desempenho inferior a um palpite aleatório.

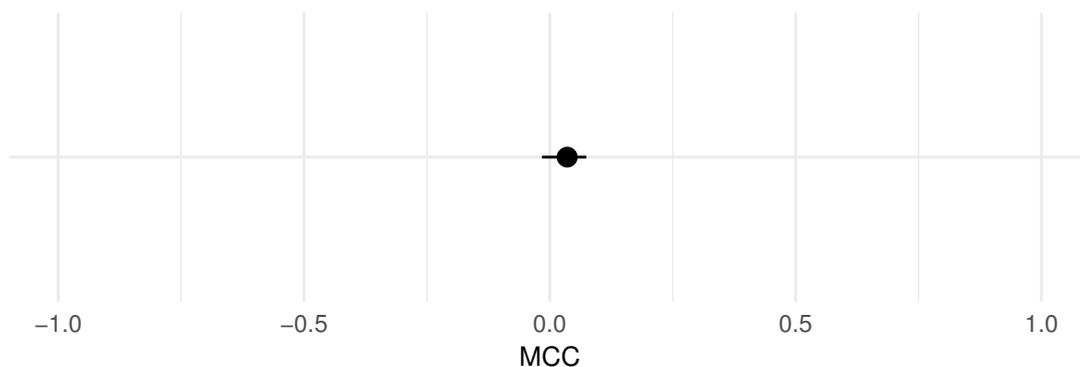


Figura 5.1: Intervalo de Confiança - Top 10 observações da Matriz de Risco

Este intervalo de confiança do MCC para as Top 10 primeiras empresas varia de -0,016 até 0,074 para a amostra, diferença total de 0,090.

O Top 10 observações foi o recorte de dados que obteve o maior valor para o MCC. Ainda assim, das dez empresas identificadas como alto risco, apenas duas (20%) eram suspeitas justificadas. No Top 100 a proporção de acerto fica ainda menor: das 100 que seriam

investigadas, apenas cinco deveriam ter sido (5%). A medida que o conjunto de empresas selecionadas cresce, a quantidade de Falsos Positivos positivos é consideravelmente superior à quantidade de Verdadeiros Positivos.

A revocação reflete a habilidade do classificador de identificar o maior número possível de observações alvo. Ao passo que aumentamos o recorte de dados, a revocação tem uma variância de apenas 0,0004 (sd 0,02); o número de Verdadeiros Positivos cresce lentamente enquanto o de Falsos Negativos decai na mesma intensidade. A partir desta avaliação, é possível perceber com clareza a desvantagem de utilizar a revocação isoladamente.

Por outro lado, à medida que o recorte das empresas cresce, a quantidade de Falsos Positivos aumenta de maneira considerável, o que para os órgãos de controle externo é refletido em maiores custos operacionais, financeiros e de pessoal.

Ressaltamos que esta avaliação realizada sobre a Matriz de Risco utiliza como base para rotulagem de risco os contratos rescindidos entre os anos de 2014 a 2018. É possível que, após investigação, a população de empresas cujo risco associado seja alto venha a crescer.

5.1.2 Treinamento de modelos com as tipologias da Matriz

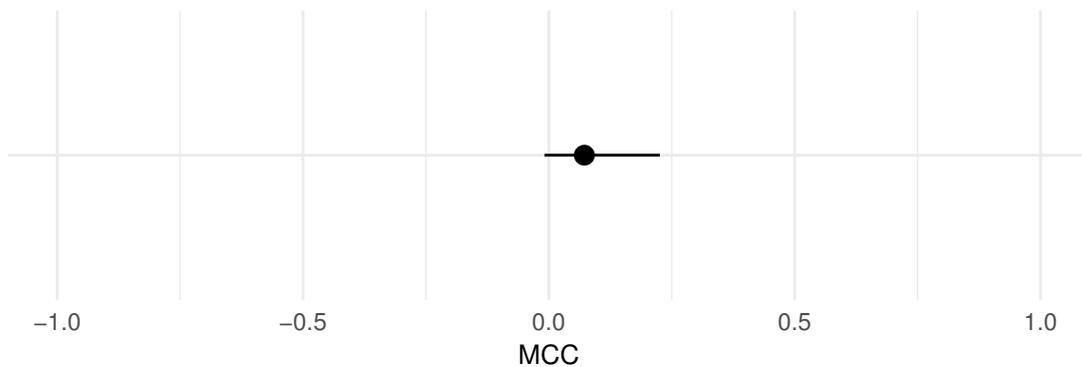
Com o propósito de verificar a eficácia das tipologias de risco da Matriz na diferenciação entre empresas de baixo e alto risco, utilizamos essas tipologias como variáveis preditoras para o conjunto de modelos de aprendizagem de máquina utilizado neste estudo.

Após a partição aleatória do arquivo da Matriz em arquivos de treino e teste, ambos ficaram compostos das seguintes proporções, respectivamente: 10.251 observações de baixo risco e 205 de alto risco, e 2562 observações de baixo risco e 51 de alto risco.

A classificação de empresas quanto ao seu risco associado, neste ensaio, apresenta pouca ou nenhuma melhora em relação ao ranqueamento de risco proposto na Matriz avaliada anteriormente em relação à eficácia. A *Tree Bag*, modelo que apresentou melhor valor para o MCC, classificou corretamente apenas duas das 51 empresas de alto risco, porém errou ao classificar 10 empresas cujo risco era de fato baixo. Dos seis modelos avaliados, três tiveram seus MCC iguais a zero e um levado a valores negativos. No segundo caso, é possível dizer que não só o modelo não conseguiu diferenciar ambas as classes, como está indo na direção errada da classificação - a Floresta Aleatória não possui nenhum Verdadeiro Positivo, apenas Falsos Positivos e Negativos.

Modelo	F1	AUC	Precisão	Revoc.	MCC	VP	VN	FP	FN
Reg. Logística	0,026	0,505	0,038	0,020	0,014	1	2537	25	50
Rede Neural	0,000	0,500	0,000	0,000	0,000	0	2562	0	51
SGB	0,000	0,500	0,000	0,000	0,000	0	2562	0	51
<i>Tree Bag</i>	0,063	0,518	0,167	0,039	0,072	2	2552	10	49
Floresta Aleatória	0,000	0,500	0,000	0,000	-0,003	0	2561	1	51
Árvore	0,000	0,500	0,000	0,000	0,000	0	2562	0	51

Tabela 5.2: Avaliação dos modelos treinados com as tipologias da Matriz de Risco

Figura 5.2: Intervalo de Confiança - *Tree Bag*

A partir do intervalo de confiança apresentado na Figura 5.2, vê-se que o MCC para a população de empresas deste ensaio está compreendido entre -0,008 e 0,225. A faixa de valores plausíveis para o MCC na população de empresas inclui portanto valores muito próximos a zero e um efeito positivo relativamente pequeno (0,225).

5.1.3 Modelagem Através de Novos Atributos

No intuito de verificar a capacidade dos indicadores propostos na Seção 4.3 em distinguir empresas em relação ao seu nível de risco, geramos os referidos indicadores associados às mesmas empresas que constam na Matriz de Risco anteriormente avaliada, e modelamos seu comportamento em relação ao risco.

Alguns dos modelos demonstram melhora na classificação das empresas presentes na Matriz de Risco. Em comparação com o resultado dos modelos na Tabela 5.2, há elevação na

Modelo	F1	AUC	Precisão	Revoc.	MCC	VP	VN	FP	FN
Reg. Logística	0,109	0,529	0,750	0,059	0,204	3	1444	1	48
Rede Neural	0,000	0,500	0,000	0,000	0,000	0	1445	0	51
SGB	0,000	0,500	0,000	0,000	0,000	0	1445	0	51
<i>Tree Bag</i>	0,194	0,557	0,545	0,118	0,243	6	1440	5	45
Floresta Aleatória	0,070	0,518	0,333	0,039	0,105	2	1441	4	49
Árvore	0,140	0,539	0,667	0,078	0,221	4	1443	2	47

Tabela 5.3: Avaliação do conjunto de modelos utilizando como insumo atributos propostos para as empresas descritas na Matriz de Risco

quantidade de Verdadeiros Positivos detectados e também diminuição no número de Falsos Positivos.

Tree Bag foi o modelo que demonstrou melhor desempenho na separação das classes desse conjunto de dados. A partir da amostra, calculamos o valor do MCC para a população de empresas em geral. Na Figura 5.3, observa-se que o MCC para os dados da população pode assumir valores entre 0,095 e 0,404. Essa faixa de valores plausíveis inclui os efeitos irrisórios e pequenos do experimento imediatamente anterior, e admitem uma eficácia considerável (0,404) como também plausível.

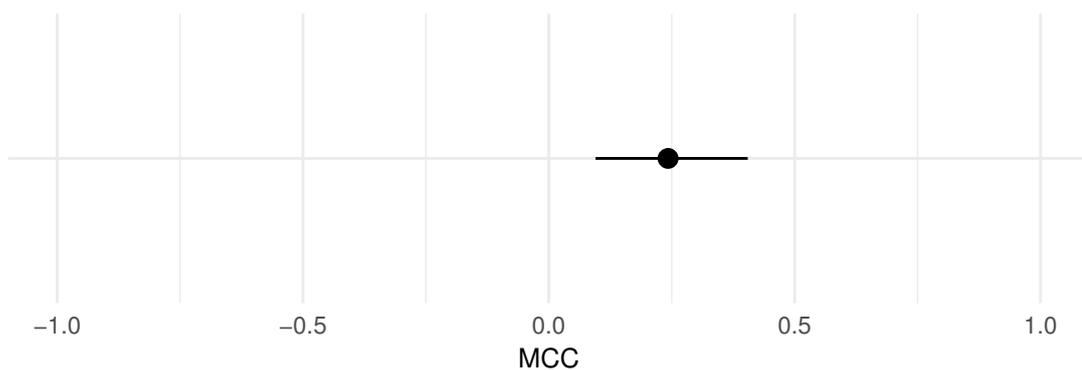


Figura 5.3: Intervalo de Confiança - *Tree Bag*

Ainda que as características propostas como descritores para as empresas tenham aprimorado o desempenho dos modelos, o número de Falsos Negativos é expressivo. Dado a natureza de severo desbalanceamento entre as classes da variável alvo, é esperado que os al-

goritmos não consigam obter informações da classe minoritária suficientes para diferenciá-la da sua oposta.

5.1.4 Considerações Sobre a Modelagem do Risco Associado a Empresas

Entre os ensaios das Seções 5.1.2 e 5.1.3, o melhor modelo utilizando aprendizagem de máquina proposto possui $MCC = 0,243$ (*Tree Bag*), ao tempo que na avaliação dos subconjuntos do ranking da Matriz de Risco, o Top 10 foi o recorte de dados que resultou o maior valor para o MCC (0,035).

Das 11 empresas classificadas pela *Tree Bag* como pertencentes à classe de alto risco, seis de fato tiveram pelo menos um contrato rescindido entre os anos de 2014 e 2018. Enquanto nas 10 primeiras empresas selecionadas na Matriz de Risco, apenas 2 seriam relevantes em futuras investigações.

Nas Figuras 5.1 e 5.3 é possível observar o intervalo no qual estimamos que o MCC produzido pelas duas abordagens está contido, com uma confiança de 95%.

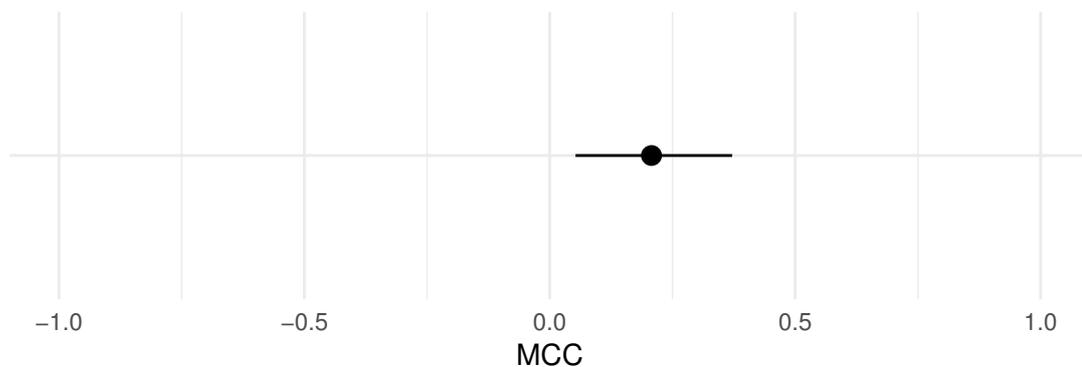


Figura 5.4: Intervalo de confiança da diferença entre o desempenho da *Tree Bag* e do Top 10 da Matriz de Risco

A estimação do intervalo de confiança da diferença entre os MCCs para ambos os modelos indica não apenas a direção da diferença (se o desempenho da *Tree Bag* é melhor ou pior que o Top 10 da Matriz), mas também a magnitude da diferença (se, na prática, temos uma diferença relevante).

Podemos observar que há evidência de uma diferença positiva entre os dois estimadores

de risco comparados. Entretanto, as barras de erro do intervalo de confiança na Figura 5.4 indicam que há margem para que essa diferença assuma valores grandes ou pequenos. Quanto menor o valor da diferença, mais próximo o desempenho da *Tree Bag* seria da classificação das empresas através do ranking Matriz de Risco. Ao passo que a diferença aumenta, a *Tree Bag* mostraria melhor capacidade em distinguir empresas de ambas as classes de risco em relação ao método utilizado pelo MPPB.

Na prática, podemos afirmar que a utilização da *Tree Bag* como modelo de seleção de empresas a serem acompanhadas mais proximamente traria benefícios, porém que não é possível com nossos dados precisar se esse benefício seria substancial ou marginal.

5.2 Experimento Sobre Contratos de Propósito Geral

Nos experimentos seguintes, utilizando dados de contratos administrativos caracterizados pelos atributos descritos na Seção 4.3, e comparamos a eficácia dos modelos tendo como insumo dados sem nenhum tratamento e dados que passaram por tratamentos a fim de terem suas classes balanceadas. Há de se notar que, por se tratar de um contexto onde a norma seja o cumprimento dos contratos, as classes de alto e baixo risco estão agressivamente desbalanceadas.

Originalmente dispomos de 12.428 contratos rotulados como baixo risco e 139 contratos pertencentes a classe de alto risco. Após a divisão dos dados entre treino e teste, os arquivos gerados seguem a seguinte distribuição entre classes: treino - 9943 baixo risco e 112 alto risco, teste - 2.485 baixo risco e 27 alto risco.

5.2.1 Balanceamento original

Nos dados utilizados como insumo para treinamento dos modelos, a classe minoritária representa cerca de 1% dos dados totais. Diante disso, é possível que na amostra disponível não haja informações suficientes sobre a classe de menor tamanho para que os modelos consigam diferenciá-la da sua oposta.

O valor negativo do MCC indica que a classificação das observações do conjunto de teste está indo na contramão da informação de referência utilizada (contratos rescindidos). Neste caso, apesar de negativo, o MCC é muito próximo de zero, o que aproxima a classificação

Modelo	F1	AUC	Precisão	Revoc.	MCC	VP	VN	FP	FN
Reg. Logística	-	-	-	-	-	-	-	-	-
Rede Neural	-	-	-	-	-	-	-	-	-
SGB	0,000	0,500	0,000	0,000	-0,002	0	2484	1	27
<i>Tree Bag</i>	0,167	0,554	0,333	0,111	0,188	3	2479	6	24
Floresta Aleatória	0,111	0,536	0,222	0,074	0,123	2	2478	7	25
Árvore	0,000	0,499	0,000	0,000	-0,004	0	2481	4	27

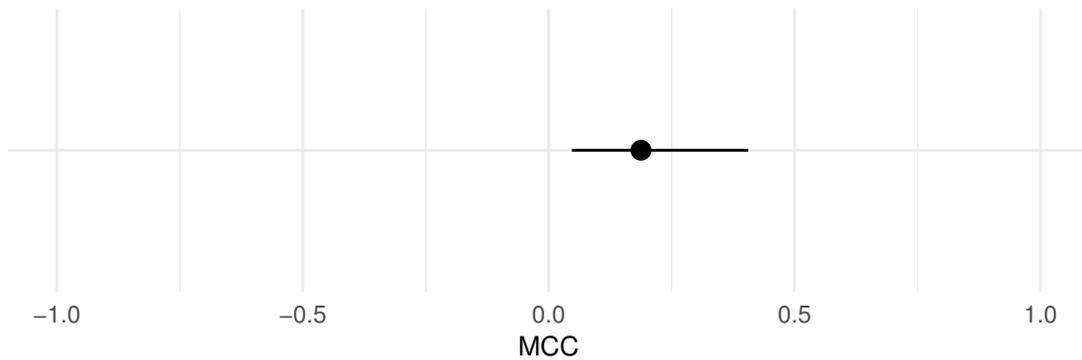
Tabela 5.4: Avaliação dos modelos utilizando como entrada dados de contratos sem balanceamento

de um palpite aleatório. A *Árvore* e a *Tree Bag* não possuem nenhum contrato classificado corretamente estimado como de alto risco (Verdadeiro Positivo), enquanto classificam erroneamente contratos de baixo risco como alto. Neste ensaio, tanto a Regressão Logística quanto a Rede Neural não convergiram em seu treinamento.

Utilizando dados de contratos administrativos em sua distribuição de classes original, o melhor desempenho na classificação do conjunto de testes foi da *Tree Bag*, três dos 27 contratos rescindidos foram classificados corretamente.

Considerando o conjunto de contratos utilizado nestes experimentos como uma amostra dos contratos administrativos a serem classificados, é possível calcular o intervalo de confiança no qual está contido o MCC da população (Figura 5.5). Entre 0,047 e 0,405, o intervalo de confiança do classificador evidencia, na prática, que seu desempenho para a população de contratos pode estar bastante próximo a um palpite aleatório, como também pode, no melhor caso, ser mediano (quando mais próximo de 0,5).

Como observado, o desbalanceamento entre classes pode afetar negativamente o aprendizado e desempenho dos algoritmos de classificação [2]. As Seções 5.2.2, 5.2.3 e 5.2.4 têm como objetivo mitigar artificialmente o problema do desbalanceamento entre classes tratando o conjunto de treino utilizado neste ensaio com técnicas frequentemente utilizadas na literatura.

Figura 5.5: Intervalo de Confiança - *Tree Bag*

5.2.2 Undersampling

Aplicado o *Undersampling* aos dados de treinamento, a distribuição entre as classes resultou em 112 observações para ambas. Enquanto o teste se manteve com a proporção original.

Modelo	F1	AUC	Precisão	Revoc.	MCC	VP	VN	FP	FN
Reg. Logística	0,025	0,544	0,013	0,593	0,018	16	1229	1256	11
Rede Neural	0,000	0,500	0,000	0,000	0,000	0	2485	0	27
SGB	0,026	0,557	0,014	0,556	0,024	15	1389	1096	12
<i>Tree Bag</i>	0,033	0,617	0,017	0,630	0,049	17	1503	982	10
Floresta Aleatória	0,037	0,660	0,019	0,741	0,067	20	1437	1048	7
Árvore	0,024	0,542	0,012	0,815	0,019	22	668	1817	5

Tabela 5.5: Avaliação dos modelos utilizando como entrada dados de contratos com balanceamento através do *Undersampling*

O modelo de pior desempenho foi a Rede Neural, que classificou todos os contratos como pertencentes a classe de baixo risco. Talvez pela pouca quantidade de dados resultante do *Undersampling* no conjunto de treinamento, a Rede não consiga aprender a distinguir as classes presentes. Classificar todas as observações como de uma mesma classe, nos dados de teste, é uma estratégia do algoritmo para reduzir o erro na classificação.

Com base na comparação entre os valores do MCC, o melhor desempenho foi atribuído a Floresta Aleatória. A Árvore conseguiu classificar dois contratos de alto risco corretamente além da Floresta, em contrapartida gerou 769 Falsos Positivos a mais que esta.

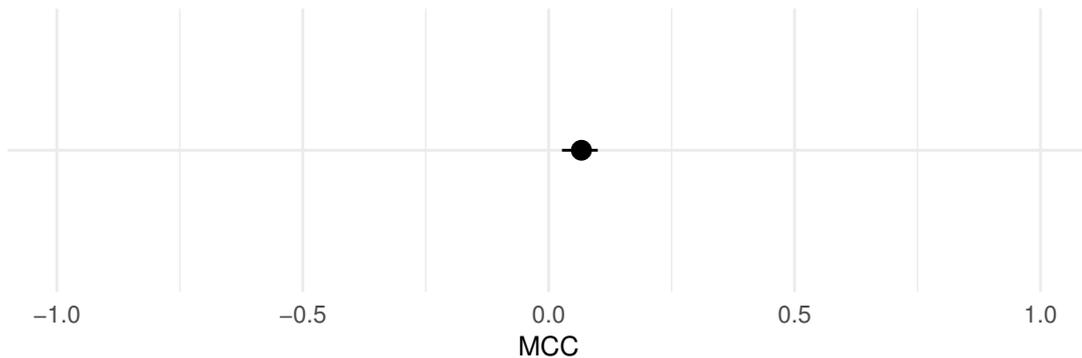


Figura 5.6: Intervalo de Confiança (*Undersampling*) - Floresta Aleatória

Como efeito do tratamento aplicado para balanceamento houve piora no desempenho dos modelos em relação aos resultados utilizando dados sem nenhum balanceamento (Tabela 5.4), gerando inclusive um número maior de Falsos Positivos. Os modelos erram amplamente em ambas as direções: classificam incorretamente contratos de alto e baixo risco como baixo e alto, respectivamente. Sob esta perspectiva, é possível supor que devido ao número reduzido de observações do conjunto de treino é provável que não hajam informações suficientes sobre ambas as classes para discriminá-las, comprometendo a capacidade de generalização dos classificadores.

Consideramos que neste ensaio nenhum dos seis modelos resultou em um desempenho satisfatório. Os valores do MCC todos abaixo de 0,1 indicam uma classificação análoga a um palpite aleatório.

Na Figura 5.5 é possível observar o intervalo no qual está localizado o MCC da população de contratos, utilizando como modelo de classificação a Floresta Aleatória.

5.2.3 *Oversampling*

Após o tratamento dos dados originais através do *Oversampling*, o número de contratos das classes igualou-se no valor de 9.943 observações. O conjunto de teste seguiu com o número de observações original.

De acordo com o MCC o modelo que melhor classificou novas observações foi a Floresta Aleatória, não só reconhecendo quatro (Verdadeiros Positivos) dos 27 contratos de alto risco, como também classificando a menor quantidade de Falsos Positivos.

Assim como na Seção 5.2.2, os classificadores também erram em ambas as direções. A

Modelo	F1	AUC	Precisão	Revoc.	MCC	VP	VN	FP	FN
Reg. Logística	0,041	0,657	0,021	0,630	0,069	17	1699	786	10
Rede Neural	0,022	0,507	0,011	0,889	0,004	24	310	2175	3
SGB	0,053	0,617	0,029	0,370	0,070	10	2148	337	17
<i>Tree Bag</i>	0,150	0,554	0,231	0,111	0,154	3	2475	10	24
Floresta Aleatória	0,195	0,572	0,286	0,148	0,200	4	2475	10	23
Árvore	0,041	0,586	0,022	0,333	0,048	9	2085	400	18

Tabela 5.6: Avaliação dos modelos utilizando como entrada dados de contratos com balanceamento através do *Oversampling*

Rede Neural obteve o pior desempenho dentre os seis modelos, desta vez com a classificação enviesada para a classe de alto risco: acertou 24 dos 27 contratos de alto risco, entretanto 2.175 contratos de baixo risco foram selecionados injustamente. Relativo ao grande número de Falsos Positivos, a Árvore, SGB e Regressão Logística tiveram comportamentos semelhantes à Rede Neural. Caso considerássemos apenas o valor da AUC para comparar os modelos, a Regressão Logística estaria à frente dos demais, entretanto este modelo teve o segundo maior valor para Falsos Positivos.

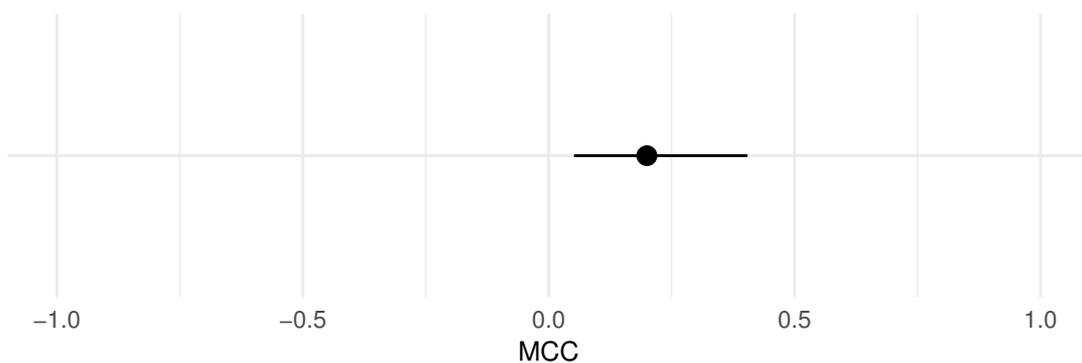


Figura 5.7: Intervalo de Confiança (*Oversampling*) - *Random Forest*

Utilizando o modelo da Floresta Aleatória, o valor estimado do MCC para a população de contratos a serem classificados quanto ao seu risco característico está entre 0,051 e 0,404 (Figura 5.7).

5.2.4 SMOTE

Modelo	F1	AUC	Precisão	Revoc.	MCC	VP	VN	FP	FN
Reg. Logística	0,057	0,610	0,031	0,333	0,071	9	2203	282	18
Redes Neurais	0,040	0,557	0,022	0,222	0,038	6	2218	267	21
SGB	0,054	0,585	0,030	0,259	0,060	7	2261	224	20
<i>Tree Bag</i>	0,092	0,692	0,051	0,481	0,130	13	2241	244	14
Floresta Aleatória	0,085	0,645	0,048	0,370	0,108	10	2286	199	17
Árvore	0,031	0,539	0,017	0,222	0,023	6	2129	356	21

Tabela 5.7: Avaliação dos modelos utilizando como entrada dados de contratos com balanceamento através do SMOTE

Através do tratamento dos dados de treino com o SMOTE, o balanceamento entre as classes resultou nos seguintes números de contratos de baixo e alto risco, nessa ordem: 448 e 336. Novamente, o conjunto utilizado para teste manteve suas proporções originais a fim de validar o modelo em um dado similar ao cenário real.

A *Árvore* foi o modelo com o maior número de contratos classificados erroneamente. Errou 21 dos 27 contratos de alto risco e 356 dos 2.485 de baixo risco.

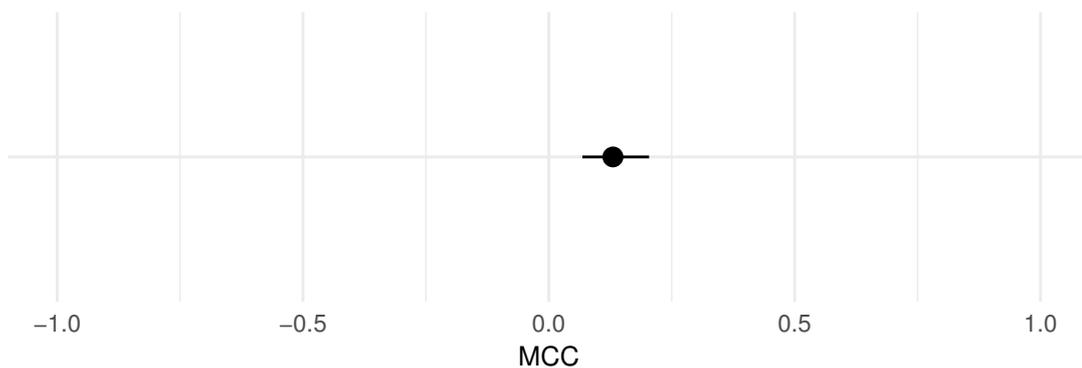


Figura 5.8: Intervalo de Confiança (SMOTE) - *Tree Bag*

Neste ensaio a *Tree Bag* exibiu melhor desempenho em relação aos demais classificadores, categorizando justamente 13 dos 27 contratos que, de fato, foram rescindidos. Na Figura 5.8, é possível observar o intervalo de confiança (de 0,068 a 0,203) para a *Tree Bag*

a partir do MCC da população de contratos a serem classificados. De modo geral, o desempenho dos modelos foi comprometido pela grande quantidade de Falsos Positivos, o que demonstra a tendência dos modelos em relação a classe de alto risco.

5.2.5 Considerações Sobre a Modelagem do Risco Associado a Contratos

Nestes ensaios foram comparadas as eficácias entre modelos de aprendizagem de máquina e também entre técnicas de balanceamento de classes. A partir do conjunto de modelo treinado utilizando como insumo dados provenientes dos diferentes tratamentos de balanceamento, foram escolhidos os melhores modelos de cada tratamento: Dados sem nenhum tipo de tratamento - *Tree Bag*, *Undersampling* - Floresta Aleatória, *Oversampling* - Floresta Aleatória, e SMOTE - *Tree Bag*.

A replicação de amostras artificialmente através do *Oversampling* a fim de igualar o tamanho das classes minoritária e majoritária forneceu o maior MCC. Por outro lado, reduzir o número de observações da classe majoritária e utilizar dados sem nenhum tratamento foram estratégias pouco promissoras no contexto da classificação de contratos públicos municipais.

Neste experimento é importante destacar o Coeficiente de Correlação de Matthews (MCC) como uma métrica avaliativa robusta, que permitiu comparar modelos treinados com conjuntos de dados de cardinalidades severamente distintas. No cenário da rescisão contratual a nível municipal, o grande desbalanceamento entre as classes se apresenta como um desafio. Não utilizar técnicas de amostragem a fim de incrementar classes, ou métricas apropriadas para avaliação e comparação entre modelos pode levar a conclusões equivocadas.

Em todos os ensaios pelo menos um modelo baseado em árvore obteve desempenho relativamente superior aos demais. Para fins de interpretabilidade dos resultados, isto é marcante, visto que, caso necessário, é possível explicar os motivos da classificação realizada.

Considerando os dois modelos de melhor desempenho entre os quatro tratamentos de dados: Floresta Aleatória (*Oversampling*) e *Tree Bag* (Dados sem balanceamento entre as classes), é possível verificar a diferença na eficácia dos dois modelos quanto a sua aplicação utilizando dados relativos a população de contratos administrativos municipais.

As barras de erro do intervalo de confiança exibido na Figura 5.9 variam de -0,232 até

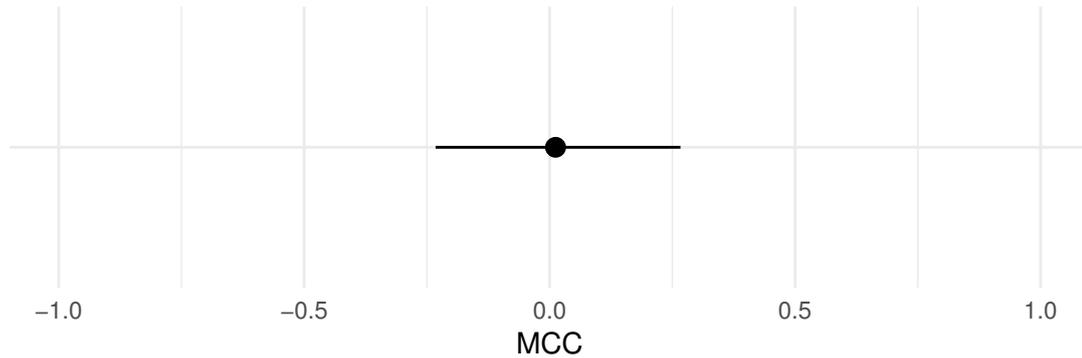


Figura 5.9: Intervalo de confiança da diferença entre o desempenho da Floresta Aleatória e da *Tree Bag*

0,266, podendo assumir valores tanto negativos como positivos. Isto evidencia que é possível que a Floresta Aleatória tenha um desempenho substancialmente melhor, ou que o desempenho da *Tree Bag* se sobressaia consideravelmente, ou que ambas possuam um desempenho bastante próximo (caso o MCC seja zero).

O treinamento e teste dos modelos acima avaliados foram realizados a partir de uma partição do arquivo total de contratos em arquivos de treino e teste. É relevante exibir que também foi realizado um experimento seccionando os dados de contratos desde uma perspectiva temporal: treinamos os modelos com contratos iniciados de 30/05/2003 até 30/12/2015, e aplicamos os modelos em dados de contratos iniciados de 02/01/2016 até 19/04/2017. Nestes moldes, o ensaio realizado teve o objetivo de verificar o desempenho dos modelos em uma abordagem utilizando janelas temporais para o particionamento de dados. Entretanto, a eficácia dos modelos neste caso foi significativamente inferior aos utilizando o particionamento convencional de amostragem aleatória.

5.3 Experimento Sobre Contratos de Obras e Serviços de Engenharia

Após a modelagem dos contratos administrativos dos municípios paraibanos em relação ao seu risco associado, também era propósito deste estudo verificar a eficácia da modelagem de contratos de uma área específica de atuação das empresas contratadas. Neste caso, utilizamos contratos referentes a obras e serviços de engenharia, rotulados e caracterizados pelos

atributos descritos na Seção 4.3.

Contamos com 505 contratos de obras: 297 pertencentes a classe de baixo risco e 208 da classe de alto risco. Em comparação ao descrito na Seção 5.2 dispomos de um volume menor de dados, porém o cenário de desbalanceamento neste caso se mostra menos crítico. Após a divisão dos dados em conjuntos de treino e teste, temos 238 e 59 contratos de baixo risco, 167 e 41 contratos de alto risco no conjunto de treinamento e teste, respectivamente.

A seguir experimentaremos a modelagem dos dados em sua distribuição entre classes original, bem como tratados utilizando técnicas de balanceamento artificial.

5.3.1 Balanceamento original

Modelo	F1	AUC	Precisão	Revoc.	MCC	VP	VN	FP	FN
Reg. Logística	0,564	0,641	0,595	0,537	0,288	22	44	15	19
Rede Neural	0,000	0,500	0,000	0,000	0,000	0	59	0	41
SGB	0,650	0,707	0,667	0,634	0,417	26	46	13	15
<i>Tree Bag</i>	0,706	0,747	0,682	0,732	0,490	30	45	14	11
Floresta Aleatória	0,675	0,732	0,722	0,634	0,476	26	49	10	15
Árvore	0,642	0,698	0,650	0,634	0,398	26	45	14	15

Tabela 5.8: Avaliação dos modelos utilizando como entrada dados de contratos de obras sem balanceamento

Nesta primeira experimentação, dos dados utilizados para treinar os modelos de classificação, 41% destes eram da classe de alto risco. De maneira geral o conjunto de modelos teve um desempenho superior em comparação ao ensaio utilizando contratos administrativos de propósito geral (descrito na Seção 5.2.1).

A Rede Neural foi o modelo com pior desempenho, classificando todos os contratos como contratos de baixo risco. É possível que a pequena quantidade de dados tenha comprometido o aprendizado deste modelo, e ao tentar minimizar o erro de classificação todas as observações foram rotuladas como pertencentes ao grupo de baixo risco.

O modelo mais eficaz na distinção entre as classes foi a *Tree Bag*, que classificou corretamente 30 dos 41 contratos tidos como de alto risco, o maior valor de Verdadeiros Positivos

dentre todos os modelos nesse ensaio. A Floresta Aleatória ficou atrás do melhor modelo devido sua dificuldade em identificar contratos da classe de alto risco, classificando quatro observações a mais que a *Tree Bag* como Falsos Negativos.

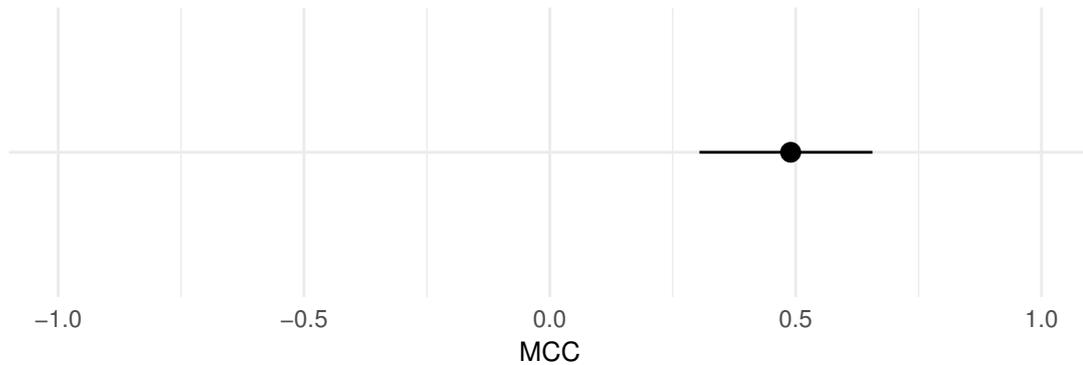


Figura 5.10: Intervalo de Confiança (Dados sem tratamento) - *Tree Bag*

Presumindo o conjunto de contratos utilizado como uma amostra de todos os contratos de obras a serem rotulados pelo melhor modelo aqui encontrado, podemos calcular o intervalo de confiança onde estaria contido o MCC para o modelo aplicado nesta população de contratos. Na Figura 5.10 é exibido o intervalo de confiança, que varia de 0,304 até 0,656.

5.3.2 *Undersampling*

Após aplicar *Undersampling* no conjunto de dados de treinamento, o arquivo de dados ficou composto de 167 contratos em cada uma das classes.

Observando o MCC, o modelo Floresta Aleatória apresentou o melhor desempenho neste ensaio, classificando corretamente 34 dos 41 contratos realmente rescindidos. Este também foi o modelo com o menor número de Falsos Positivos, ou seja, que menos errou na identificação de contratos de baixo risco. Apesar da *Tree Bag* ter identificado um contrato de alto risco a mais que a Floresta Aleatória, aquela produziu sete contratos Falsos Positivos além desta.

O intervalo de confiança do MCC para a Floresta Aleatória na população de contratos de obras pode ser observado na Figura 5.11, e varia de 0,408 a 0,725.

Os piores resultados na classificação foram obtidos pela Rede Neural. É possível que o desempenho insatisfatório deste modelo decorra da quantidade insuficiente de dados para seu treinamento efetivo.

Modelo	F1	AUC	Precisão	Revoc.	MCC	VP	VN	FP	FN
Reg. Logística	0,674	0,709	0,608	0,756	0,410	31	39	20	10
Rede Neural	0,574	0,591	0,483	0,707	0,183	29	28	31	12
SGB	0,667	0,705	0,612	0,732	0,403	30	40	19	11
<i>Tree Bag</i>	0,674	0,718	0,644	0,707	0,431	29	43	16	12
Floresta Aleatória	0,764	0,796	0,708	0,829	0,583	34	45	14	7
Árvore	0,722	0,749	0,625	0,854	0,493	35	38	21	6

Tabela 5.9: Avaliação dos modelos utilizando como entrada dados de contratos de obras com balanceamento através do *Undersampling*

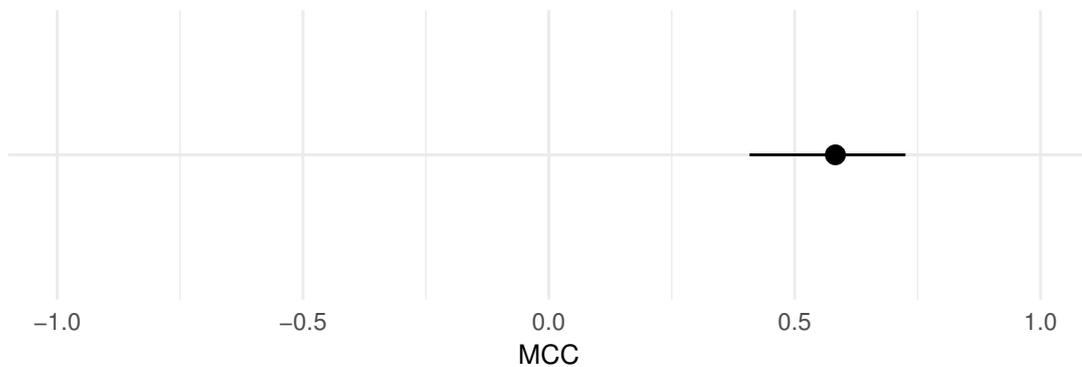


Figura 5.11: Intervalo de Confiança (*Undersampling*) - Floresta Aleatória

5.3.3 *Oversampling*

Após tratadas a classe minoritária dos dados originais através do *Oversampling*, o conjunto de treino restou com 238 observações em cada classe de risco.

Com suporte no MCC, a *Tree Bag* foi o modelo que obteve melhor desempenho, classificando corretamente 28 dos 41 contratos de alto risco. Apesar de não ter sido o modelo com maior número de Verdadeiros Positivos, demonstrou um balanço na classificação de contratos de ambas as classes. O desempenho da Floresta Aleatória ficou pouco abaixo da *Tree Bag*, apesar do maior número de Verdadeiros Positivos associado, teve mais erros na classificação de contratos de baixo risco.

Considerando o conjunto de contratos de obras como uma amostra, na Figura 5.12 podemos observar o intervalo de confiança para o MCC da *Tree Bag* na população de todos os

Modelo	F1	AUC	Precisão	Revoc.	MCC	VP	VN	FP	FN
Reg. Logística	0,659	0,701	0,617	0,707	0,396	29	41	18	12
Rede Neural	0,674	0,709	0,608	0,756	0,410	31	39	20	10
SGB	0,690	0,735	0,674	0,707	0,467	29	45	14	12
<i>Tree Bag</i>	0,700	0,748	0,718	0,683	0,501	28	48	11	13
Floresta Aleatória	0,706	0,747	0,682	0,732	0,490	30	45	14	11
Árvore	0,650	0,707	0,667	0,634	0,417	26	46	13	15

Tabela 5.10: Avaliação dos modelos utilizando como entrada dados de contratos de obras com balanceamento através do *Oversampling*

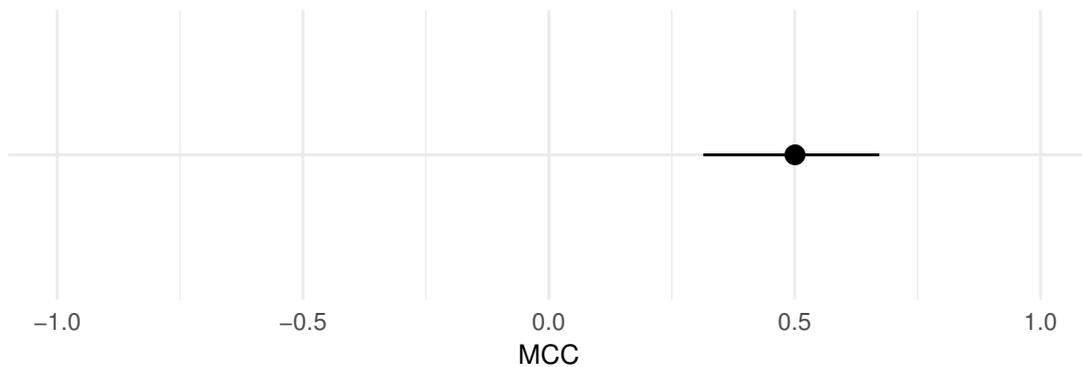


Figura 5.12: Intervalo de Confiança (*Oversampling*) - *Tree Bag*

contratos de obras a serem classificados. Este intervalo varia entre 0,314 e 0,672.

Surpreendentemente, a Regressão Logística foi o modelo com mais dificuldade em diferenciar classes neste caso. Obteve o segundo menor valor para Verdadeiros Positivos, e segundo maior valor de Falsos Positivos. Em geral, no contexto de estimação de risco, este algoritmo obtém desempenhos notáveis [18; 25; 28].

5.3.4 SMOTE

Com o SMOTE, o balanço entre as classes resultante foi de 334 contratos tanto na classe de baixo como de alto risco. Após o balanceamento dos dados, o conjunto de modelos foi submetido ao treinamento.

A *Tree Bag* novamente teve seu desempenho em destaque, acertando o nível de risco de

Modelo	F1	AUC	Precisão	Revoc.	MCC	VP	VN	FP	FN
Reg. Logística	0,659	0,701	0,617	0,707	0,396	29	41	18	12
Rede Neural	0,667	0,710	0,630	0,707	0,414	29	42	17	12
SGB	0,698	0,739	0,667	0,732	0,472	30	44	15	11
<i>Tree Bag</i>	0,745	0,774	0,660	0,854	0,541	35	41	18	6
Floresta Aleatória	0,705	0,742	0,660	0,756	0,478	31	43	16	10
Árvore	0,593	0,657	0,600	0,585	0,315	24	43	16	17

Tabela 5.11: Avaliação dos modelos utilizando como entrada dados de contratos de obras com balanceamento através do SMOTE

35 dos 41 contratos rescindidos no conjunto de teste. Entretanto, juntamente com a Regressão Logística, atingiu o maior número de Falsos Positivos. O menor número de Verdadeiros Positivos e maior de Falsos Negativos foram obtidos pela Árvore, cujo MCC alcançou o menor valor deste ensaio.

O intervalo de confiança, de 0,363 a 0,681, desse modelo pode ser observado na Figura 5.13, para a população de contratos de obras.

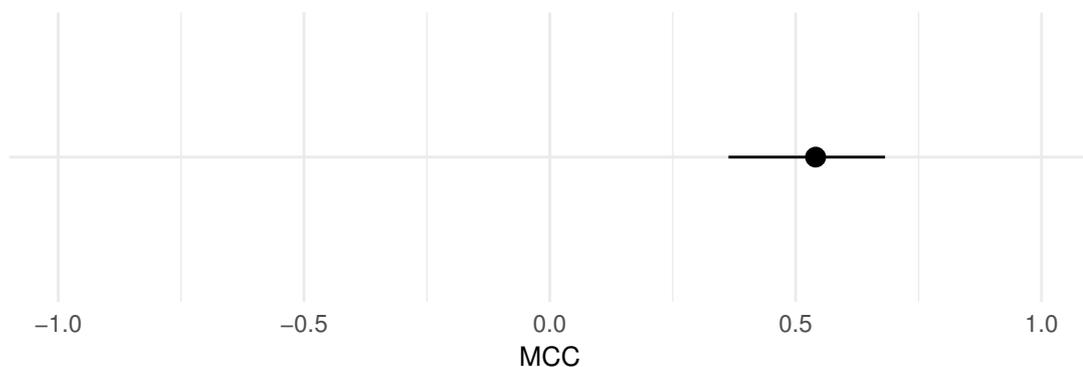


Figura 5.13: Intervalo de Confiança (SMOTE) - *Tree Bag*

5.3.5 Considerações Sobre a Modelagem do Risco Associado a Contratos de Obra e Serviços de Engenharia

No cenário da modelagem do risco associado a contratos administrativos relativos a obras e serviços de engenharia, os modelos que melhor diferenciaram as classes de alto e baixo risco foram, para cada tratamento: Dados sem nenhum tipo de tratamento - *Tree Bag*, *Undersampling* - Floresta Aleatória, *Oversampling* - *Tree Bag*, e SMOTE - *Tree Bag*.

A utilização de dados de uma área específica de atuação das empresas produziu uma classificação mais eficaz por parte dos modelos, que alcançaram, em geral, valores maiores para o MCC. A remoção aleatória de observações da classe de maior tamanho (*Undersampling*) foi o tratamento com maior MCC. Não é possível inferir se não utilizar nenhuma técnica de balanceamento entre classes, neste caso, afetou de alguma forma o desempenho dos modelos, visto que as classes dos dados originais tinham um tamanho aproximado.

Novamente, em todos os ensaios os modelos de maior destaque foram modelos baseados em Árvores de decisão, simplificando a compreensão das decisões fornecidas pelos classificadores.

A partir do intervalo de confiança exibido na Figura 5.14, para a diferença entre o desempenho da Floresta Aleatória (*Undersampling*) e da *Tree Bag* (SMOTE), podemos verificar com 95% de confiança o sentido e grandeza da diferença entre a aplicação dos modelos na população de contratos de obras.

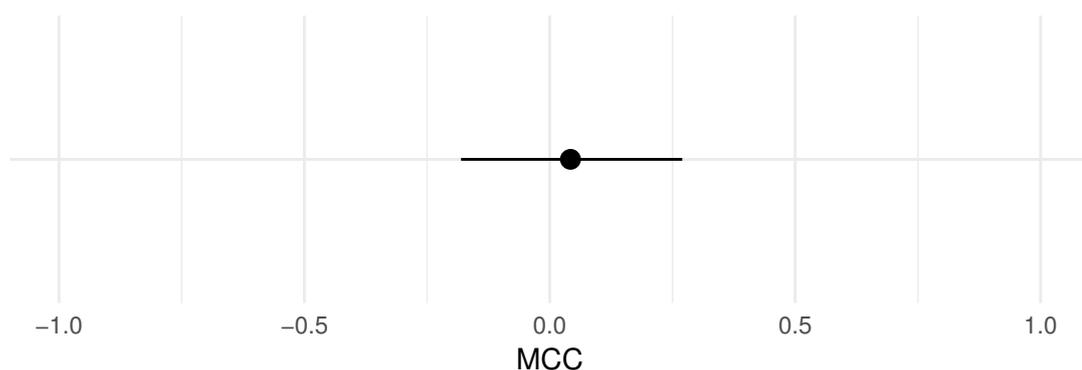


Figura 5.14: Intervalo de confiança da diferença entre o desempenho da *Floresta Aleatória* e da *Tree Bag*

Variando de -0,180 a 0,269, o intervalo de confiança da diferença entre os dois classifica-

dores evidencia que pode existir uma diferença entre ambos: em caso de diferença positiva, a Floresta Aleatória teria um desempenho superior a *Tree Bag*; em caso de diferença negativa, a *Tree Bag* teria um desempenho superior a Floresta Aleatória; e também é possível que não exista diferença entre o desempenho de ambos os modelos utilizando como insumo a população de contratos de obras.

Capítulo 6

Conclusões

6.1 Discussão

O objetivo deste trabalho foi avaliar o cenário atual e propor melhorias na estimação do risco associado a contratos e empresas em contratação com a Administração Pública dos municípios do Estado da Paraíba, visto que não temos conhecimento de abordagens utilizando aprendizagem de máquina neste nível governamental. Desejamos também que este trabalho possa ser útil a gestores de outros municípios e estados.

Buscamos modelos que não apenas classifiquem o máximo possível de observações positivas, mas com forte capacidade de discernimento entre as classes de risco. Se por um lado é necessário otimizar os recursos públicos traçando estratégias de priorização, por outro é fundamental utilizar estes recursos de forma justificada e temperada.

No contexto da estimação do risco associado a empresas foi avaliado o desempenho da Matriz de Risco, método utilizado na prática pelos órgãos de persecução criminal na Paraíba. Em seguida, seis modelos de aprendizagem foram avaliados a partir do seu treinamento com dados de empresas descritas por dois conjuntos de atributos distintos: as próprias características disponíveis na Matriz de Risco, e características propostas por nós.

A partir da comparação do modelo de melhor desempenho (*Tree Bag*) com o a classificação mais eficaz da Matriz de Risco, inferimos que o modelo de aprendizagem de máquina proposto, neste enquadramento, possui uma capacidade superior em diferenciar empresas de baixo risco de empresas de alto risco, e que o conjunto de atributos propostos se revelou como melhor descritor do comportamento de risco para empresas, proporcionando aos mo-

delos classificações mais acertivas. Ainda que, na prática, a diferença entre os dois estimadores de risco pode vir a ser de baixa a moderada, mostramos que há espaço para melhorias na estimação do risco associado a empresas.

No caso da estimação do risco associado a contratos administrativos, foram avaliados seis modelos em quatro cenários de balanceamento entre classes distintos, tanto para contratos em geral, como para contratos relativos a obras e serviços de engenharia. Em ambos os casos utilizando os atributos descritos na Seção 4.3.

Para os contratos de ampla abrangência, a partir do conjunto de modelos treinado com os resultados de cada técnica de balanceamento entre classes, selecionamos o modelo de melhor desempenho para cada técnica. A partir do intervalo de confiança da diferença entre o MCC dos dois modelos de melhor desempenho (Floresta Aleatória utilizando *Oversampling* e *Tree Bag* sem nenhum tratamento nos dados), podemos dizer que um ou outro modelo pode ter um desempenho superior, ou que não há evidências de diferença entre o desempenho dos modelos. Dentre as estratégias para mitigar o desbalanceamento entre classes, o *Oversampling* e o *Undersampling* forneceram o balanço entre as classes mais e menos eficazes, respectivamente.

Por fim, avaliamos o desempenho dos modelos de classificação e do conjunto de técnicas de balanceamento de classes utilizando como insumo dados dos contratos relacionados a obras e serviços de engenharia. Neste caso, tal como no experimento anterior, a comparação entre os modelos de melhor desempenho forneceu evidências de que é possível que tanto a Floresta Aleatória como a *Tree Bag* tenham desempenhos superiores, ou que não haja uma diferença relevante entre o desempenho dos modelos comparados. Ainda assim, a utilização de contratos de uma área específica, proporcionou modelos com uma maior capacidade de separação entre as classes, com base no valor do MCC.

É interessante notar que em todos os ensaios, pelo menos um modelo baseado em Árvores de Decisão teve um desempenho relevante. Visto que lidamos com informações sensíveis, é preciso que as razões que ocasionaram classificações estejam disponíveis e possam ser interpretadas por servidores capacitados, ou até mesmo pela população.

De modo geral, a modelagem das características de empresas utilizando modelos de aprendizagem de máquina resultou num melhor desempenho na classificação destas entidades em relação ao seu risco, em comparação com métodos que cremos que são hoje utiliza-

dos. A modelagem de dados de contratos administrativos também sucedeu-se em resultados promissores, especialmente em se tratando de contratos de uma área específica.

O apoio e atenção dos órgãos de controle para a criação e continuidade de fontes de dados que amparem este tipo de estudo é de grande relevância. Dados e informações como os utilizados e produzidos neste trabalho, são formas de instrumentalizar o controle e lapidar as técnicas e métodos já existentes.

6.2 Limitações

Entendemos que os resultados trazidos neste estudo são um ponto de partida para a estimação do risco para empresas e contratos administrativos, e que o objetivo da estimação não é identificar todo o conjunto de empresas ou contratos cujo risco associado seja alto.

A construção dos atributos utilizados como preditores dos modelos vem, desde a própria literatura, em sua maioria, baseada na experiência de servidores públicos que se debruçaram sobre esse tema de estudo. Logo, é possível que este olhar sobre as causas do risco tenha um viés inerente ao contexto no qual estes profissionais estão inseridos.

O dado utilizado como rótulo da classe de alto risco representa os contratos rescindidos sabidos pela Administração Pública. É possível, e provável, que existam processos de contratação envolvendo atitudes sancionáveis, não auditados ou investigados, que não constem na base utilizada.

6.3 Trabalhos Futuros

No mesmo contexto, outras pesquisas podem ser desenvolvidas a partir do presente estudo, como por exemplo: modelagem de contratos de outras áreas específicas de contratações (medicamentos, merenda escolar, dentre outras), análises sobre quais atributos são mais relevantes em determinados tipos de comportamentos de risco, verificação do viés e justiça dos atributos utilizados na caracterização das entidades classificadas, e comparação de outras técnicas de balanceamento entre classes.

Bibliografia

- [1] Victor Aguiar Jardim de Amorim. Licitações e contratos administrativos: teoria e jurisprudência. *Coordenação de Edições Técnicas*, 2017.
- [2] Samira e Mouhoub-Malek Anowar, Farzana e Sadaoui. Auction fraud classification based on clustering and sampling techniques. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 366–371. IEEE, 2018.
- [3] Pierre e Mazer-Emmanuel e Cobbe Paulo Balaniuk, Remis e Bessiere. Risk based government audit planning using naïve bayes classifiers. In *Advances in Knowledge-Based and Intelligent Information and Engineering Systems*, 2012.
- [4] Fethi e El-Anbari Mohammed Boughorbel, Sabri e Jarray. Optimal classifier for imbalanced data using matthews correlation coefficient metric. *PloS one*, 12(6):e0177678, 2017.
- [5] Brasil. Lei nº 8.666, de 21 de junho de 1993. regulamenta o art. 37, inciso XXI, da Constituição Federal, institui normas para licitações e contratos da administração pública e dá outras providências. *Diário Oficial União*, 1993.
- [6] James Carpenter and John Bithell. Bootstrap confidence intervals: when, which, what? a practical guide for medical statisticians. *Statistics in medicine*, 19(9):1141–1164, 2000.
- [7] Leonardo J e Da Rocha Henrique A e Mendes Gilson L Carvalho, Rommel N e Sales. Using bayesian networks to identify and prevent split purchases in brazil. In *Proceedings of the Eleventh UAI Conference on Bayesian Modeling Applications Workshop-Volume 1218*, pages 70–78. CEUR-WS. org, 2014.

- [8] Rommel Novaes Carvalho, Ricardo Silva e Carvalho. Bayesian models to assess risk of corruption of federal management units. In *BMA@ UAI*, pages 28–35, 2016.
- [9] Rommel Novaes e Ladeira Marcelo e Monteiro Fernando Mendes e de Oliveira Mendes Gilson Libório Carvalho, Ricardo Silva e Carvalho. Using political party affiliation data to measure civil servants’ risk of corruption. In *2014 Brazilian Conference on Intelligent Systems*, pages 166–171. IEEE, 2014.
- [10] Kevin W e Hall Lawrence O e Kegelmeyer W Philip Chawla, Nitesh V e Bowyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.
- [11] Davide Chicco. Ten quick tips for machine learning in computational biology. *BioData mining*, 10(1):35, 2017.
- [12] Maurizio Conti, Pier Luigi e Naldi. Detection of anomalous bids in procurement auctions. *Decision Support Systems*, 46(1):420–428, 2008.
- [13] Brasil Tribunal de Contas da União. Licitações e contratos: orientações e jurisprudência do tcu. *Secretaria Especial de Editoração e Publicações*, 2010.
- [14] Tribunal de Contas da União. Referencial de combate a fraude e corrupção: Aplicável a órgãos e entidades da administração pública. *Brasília: Secretaria de Métodos e Suporte ao Controle Externo (Semec)*, 2017.
- [15] Olivier e Waterschoot Serge e Bontempi Gianluca Dal Pozzolo, Andrea e Caelen. Racing for unbalanced methods selection. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 24–31. Springer, 2013.
- [16] Rommel N e Carvalho Ricardo S e Ramos Guilherme N Domingos, Silvio L e Carvalho. Identifying it purchases anomalies in the brazilian government procurement system using deep learning. In *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 722–727. IEEE, 2016.
- [17] Rommel Novaes e Carvalho Ricardo Silva Gomes, Thiago Alencar e Carvalho. Identifying anomalies in parliamentary expenditures of brazilian chamber of deputies with

- deep autoencoders. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 940–943. IEEE, 2017.
- [18] Ankit e Redmiles Elissa e Ghani Rayid Grace, Emily e Rai. Detecting fraud, corruption, and collusion in international development contracts: The design of a proof-of-concept automated system. In *2016 IEEE International Conference on Big Data (Big Data)*, pages 1444–1453. IEEE, 2016.
- [19] Antônio e de Salles Villar Mauro e de Melo Franco Francisco Manuel Houaiss, Antônio e Houaiss. *Dicionário Houaiss da Língua Portuguesa*. Temas e debates, 2003.
- [20] Toshio Mukai. *O Novo Estatuto Jurídico Das Licitações E Contratos Públicos: Comentários À Lei 8,666/93: com as Alterações Promovidas Pela Lei 8.883/94*. Editora Revista dos Tribunais, 1994.
- [21] Yong e Wong Yiu Hing e Chen Yijun e Sun Xin Ngai, Eric WT e Hu. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision support systems*, 50(3):559–569, 2011.
- [22] Marcelo e Carvalho Rommel N e Marzagao Thiago Paula, Ebberth L e Ladeira. Deep learning anomaly detection as support fraud investigation in brazilian exports and anti-money laundering. In *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 954–960. IEEE, 2016.
- [23] Maria Sylvia Zanella Di Pietro. Direito administrativo. In *Direito administrativo*, volume 27. Editora Atlas, 2014.
- [24] Hércules Antonio e Ferneda Edilson e Balaniuk Remis Rebouças, Rômulo Roseo e do Prado. Detecção de figurantes em pregões eletrônicos do governo federal brasileiro. *Informação & Tecnologia*, 2(2):5–21, 2017.
- [25] Leonardo J Sales. Risk prevention of public procurement in the brazilian government using credit scoring. Technical report, Observatório de Economia e Gestão de Fraude, 2013.
- [26] Rommel Novaes Sales, Leonardo Jorge e Carvalho. Measuring the risk of public contracts using bayesian classifiers. In *BMA@ UAI*, pages 7–13, 2016.

-
- [27] Ed Wilson Fernandes de Santana. *A utilização da informação contábil como um instrumento de controle social dos municípios paraibanos com mais de 50.000 habitantes*. PhD thesis, Dissertação (Mestrado em Contabilidade)-UnB, UFPB, UFRN e UFPE), Brasília, 2008.
- [28] Leonardo J Sun, Ting e Sales. Predicting public procurement irregularity: An application of neural networks. *Journal of Emerging Technologies in Accounting*, 15(1):141–154, 2018.
- [29] Maumita West, Jarrod e Bhattacharya. Intelligent financial fraud detection: a comprehensive review. *Computers & security*, 57:47–66, 2016.

Apêndice A

Fundamentos de Rescisão Contratual da Lei Geral de Licitação

Seção V

Da Inexecução e da Rescisão dos Contratos

Art. 77. A inexecução total ou parcial do contrato enseja a sua rescisão, com as conseqüências contratuais e as previstas em lei ou regulamento.

Art. 78. Constituem motivo para rescisão do contrato:

- I - o não cumprimento de cláusulas contratuais, especificações, projetos ou prazos;
- II - o cumprimento irregular de cláusulas contratuais, especificações, projetos e prazos;
- III - a lentidão do seu cumprimento, levando a Administração a comprovar a impossibilidade da conclusão da obra, do serviço ou do fornecimento, nos prazos estipulados;
- IV - o atraso injustificado no início da obra, serviço ou fornecimento;
- V - a paralisação da obra, do serviço ou do fornecimento, sem justa causa e prévia comunicação à Administração;
- VI - a subcontratação total ou parcial do seu objeto, a associação do contratado com outrem, a cessão ou transferência, total ou parcial, bem como a fusão, cisão ou incorporação, não admitidas no edital e no contrato;
- VII - o desatendimento das determinações regulares da autoridade designada para acompanhar e fiscalizar a sua execução, assim como as de seus superiores;
- VIII - o cometimento reiterado de faltas na sua execução, anotadas na forma do § 1º do art. 67 desta Lei;

IX - a decretação de falência ou a instauração de insolvência civil;

X - a dissolução da sociedade ou o falecimento do contratado;

XI - a alteração social ou a modificação da finalidade ou da estrutura da empresa, que prejudique a execução do contrato;

XII - razões de interesse público, de alta relevância e amplo conhecimento, justificadas e determinadas pela máxima autoridade da esfera administrativa a que está subordinado o contratante e exaradas no processo administrativo a que se refere o contrato;

XIII - a supressão, por parte da Administração, de obras, serviços ou compras, acarretando modificação do valor inicial do contrato além do limite permitido no § 1º do art. 65 desta Lei;

XIV - a suspensão de sua execução, por ordem escrita da Administração, por prazo superior a 120 (cento e vinte) dias, salvo em caso de calamidade pública, grave perturbação da ordem interna ou guerra, ou ainda por repetidas suspensões que totalizem o mesmo prazo, independentemente do pagamento obrigatório de indenizações pelas sucessivas e contratualmente imprevistas desmobilizações e mobilizações e outras previstas, assegurado ao contratado, nesses casos, o direito de optar pela suspensão do cumprimento das obrigações assumidas até que seja normalizada a situação;

XV - o atraso superior a 90 (noventa) dias dos pagamentos devidos pela Administração decorrentes de obras, serviços ou fornecimento, ou parcelas destes, já recebidos ou executados, salvo em caso de calamidade pública, grave perturbação da ordem interna ou guerra, assegurado ao contratado o direito de optar pela suspensão do cumprimento de suas obrigações até que seja normalizada a situação;

XVI - a não liberação, por parte da Administração, de área, local ou objeto para execução de obra, serviço ou fornecimento, nos prazos contratuais, bem como das fontes de materiais naturais especificadas no projeto;

XVII - a ocorrência de caso fortuito ou de força maior, regularmente comprovada, impeditiva da execução do contrato.

XVIII – descumprimento do disposto no inciso V do art. 27, sem prejuízo das sanções penais cabíveis. (Incluído pela Lei nº 9.854, de 1999)

Parágrafo único. Os casos de rescisão contratual serão formalmente motivados nos autos do processo, assegurado o contraditório e a ampla defesa.

Art. 79. A rescisão do contrato poderá ser:

I - determinada por ato unilateral e escrito da Administração, nos casos enumerados nos incisos I a XII e XVII do artigo anterior;

II - amigável, por acordo entre as partes, reduzida a termo no processo da licitação, desde que haja conveniência para a Administração;

III - judicial, nos termos da legislação;

IV - (VETADO)

IV - (Vetado). (Redação dada pela Lei nº 8.883, de 1994)

§ 1º A rescisão administrativa ou amigável deverá ser precedida de autorização escrita e fundamentada da autoridade competente.

§ 2º Quando a rescisão ocorrer com base nos incisos XII a XVII do artigo anterior, sem que haja culpa do contratado, será este ressarcido dos prejuízos regularmente comprovados que houver sofrido, tendo ainda direito a:

I - devolução de garantia;

II - pagamentos devidos pela execução do contrato até a data da rescisão;

III - pagamento do custo da desmobilização.

§ 3º (VETADO)

§ 3º (Vetado). (Redação dada pela Lei nº 8.883, de 1994)

§ 4º (VETADO)

§ 4º (Vetado). (Redação dada pela Lei nº 8.883, de 1994)

§ 5º Ocorrendo impedimento, paralisação ou sustação do contrato, o cronograma de execução será prorrogado automaticamente por igual tempo.

Art. 80. A rescisão de que trata o inciso I do artigo anterior acarreta as seguintes consequências, sem prejuízo das sanções previstas nesta Lei:

I - assunção imediata do objeto do contrato, no estado e local em que se encontrar, por ato próprio da Administração;

II - ocupação e utilização do local, instalações, equipamentos, material e pessoal empregados na execução do contrato, necessários à sua continuidade, na forma do inciso V do art. 58 desta Lei;

III - execução da garantia contratual, para ressarcimento da Administração, e dos valores das multas e indenizações a ela devidos;

IV - retenção dos créditos decorrentes do contrato até o limite dos prejuízos causados à Administração.

§ 1º A aplicação das medidas previstas nos incisos I e II deste artigo fica a critério da Administração, que poderá dar continuidade à obra ou ao serviço por execução direta ou indireta.

§ 2º É permitido à Administração, no caso de concordata do contratado, manter o contrato, podendo assumir o controle de determinadas atividades de serviços essenciais.

§ 3º Na hipótese do inciso II deste artigo, o ato deverá ser precedido de autorização expressa do Ministro de Estado competente, ou Secretário Estadual ou Municipal, conforme o caso.

§ 4º A rescisão de que trata o inciso IV do artigo anterior permite à Administração, a seu critério, aplicar a medida prevista no inciso I deste artigo.

Apêndice B

Mapeamento de Entidades da Administração Pública

Com objetivo de complementar a revisão de literatura do Capítulo 3, a Figura B.1 resume o processo de contratação de uma ou mais empresas pela Administração Pública.

As figuras seguintes descrevem, a partir da literatura revisada, quais entidades foram alvo de estudo sobre estimação de risco (colorido em laranja) e de quais entidades adviram os dados utilizados para possibilitar esta estimação (colorido em azul). Para a Figura B.4, não ficou claro para nós se foram utilizados dados referentes a entidade colorida em verde.

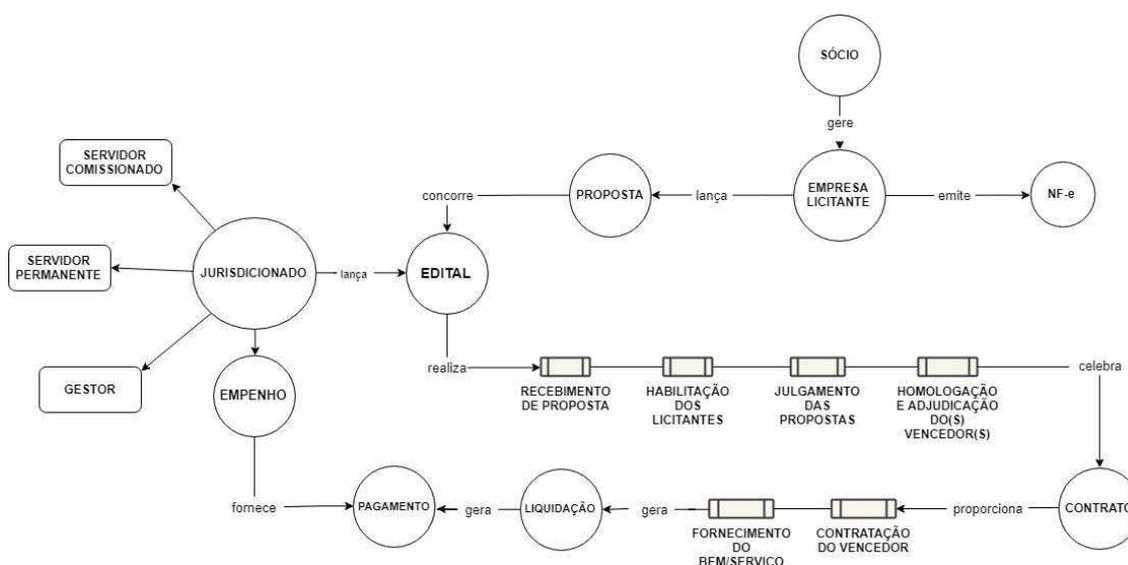


Figura B.1: Resumo do processo de contratação pública

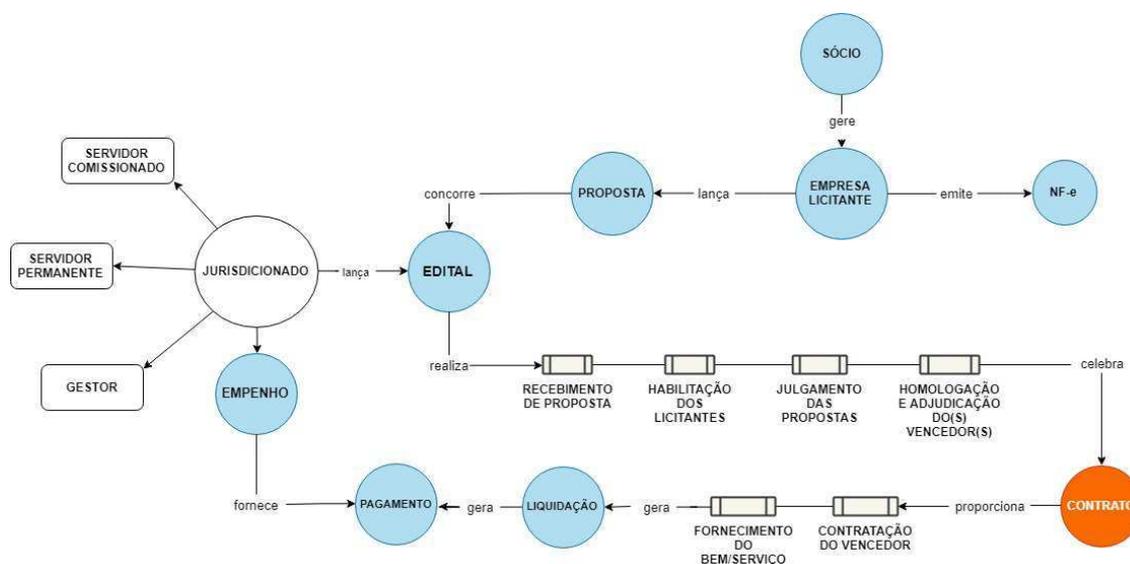


Figura B.2: Mapeamento da entidade Contrato

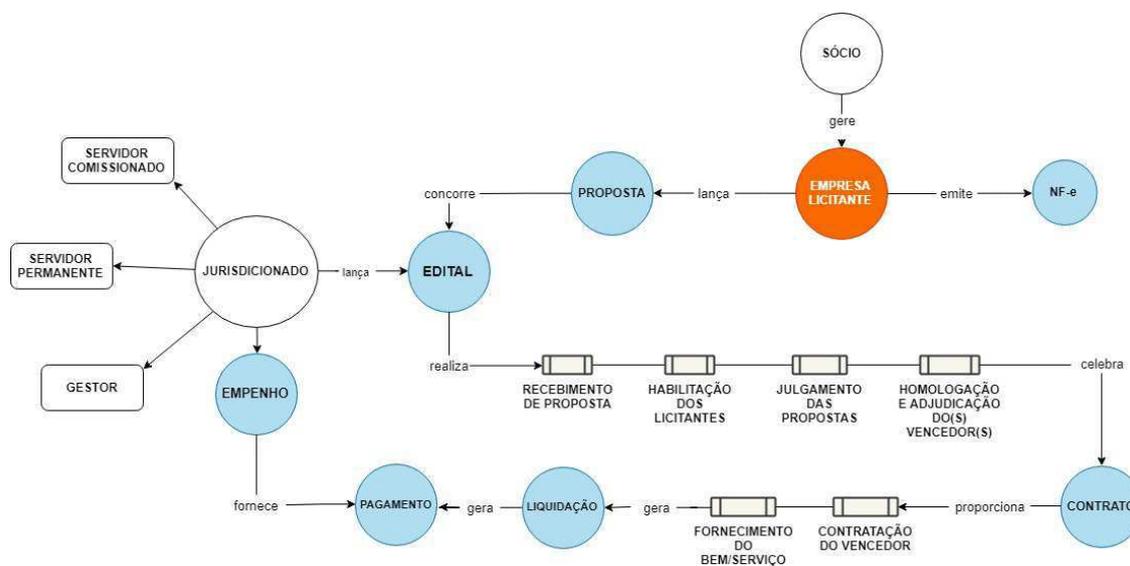


Figura B.3: Mapeamento da entidade Licitante(s)

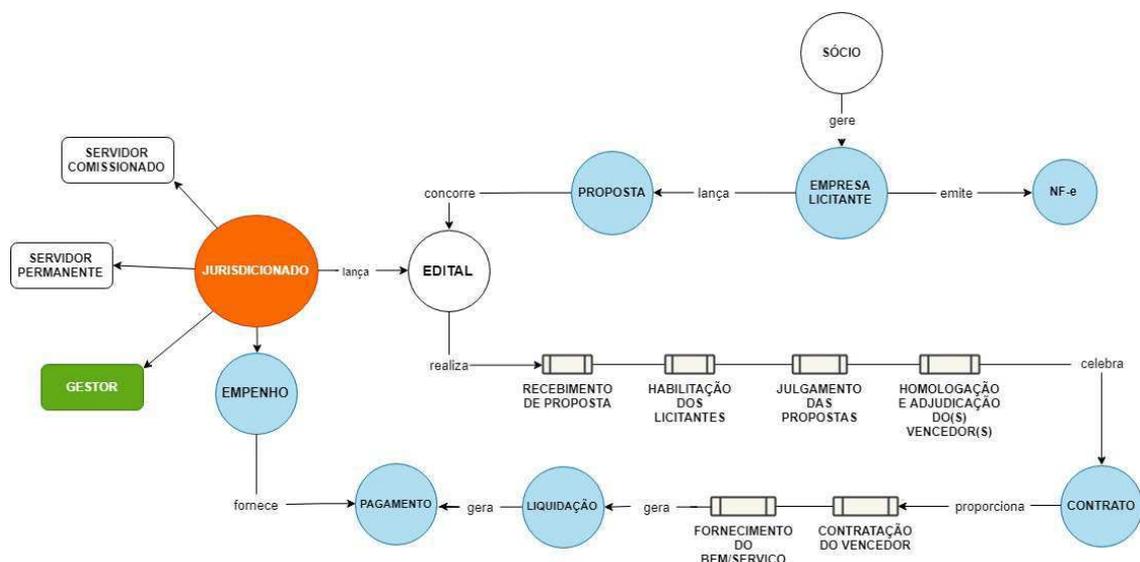


Figura B.4: Mapeamento da entidade Jurisdicionado (Unidade Administrativa)

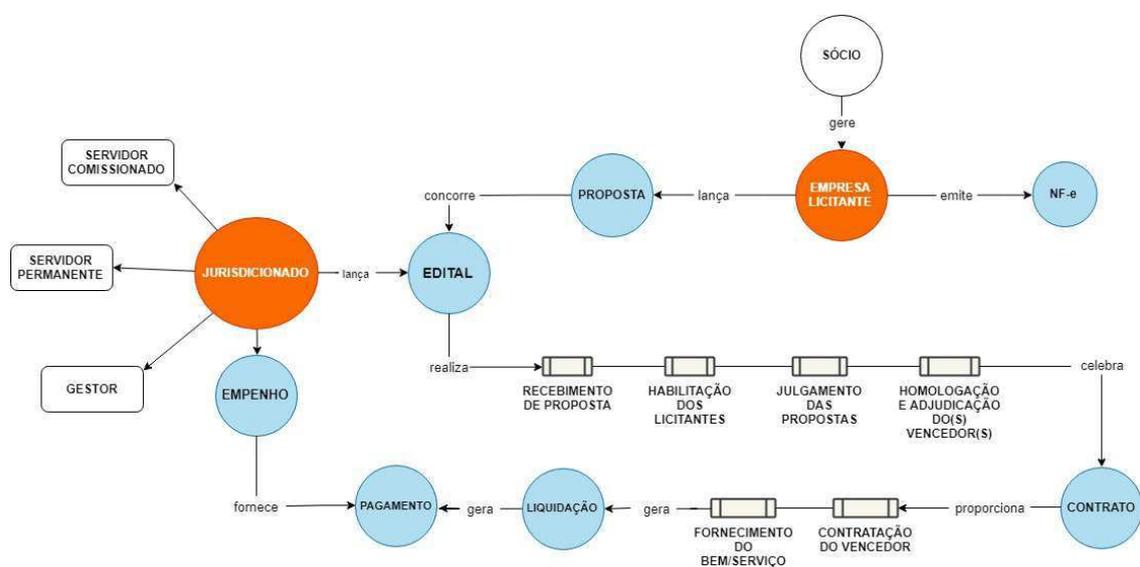


Figura B.5: Mapeamento do relacionamento entre as entidades Jurisdicionado e Licitante - Conluio

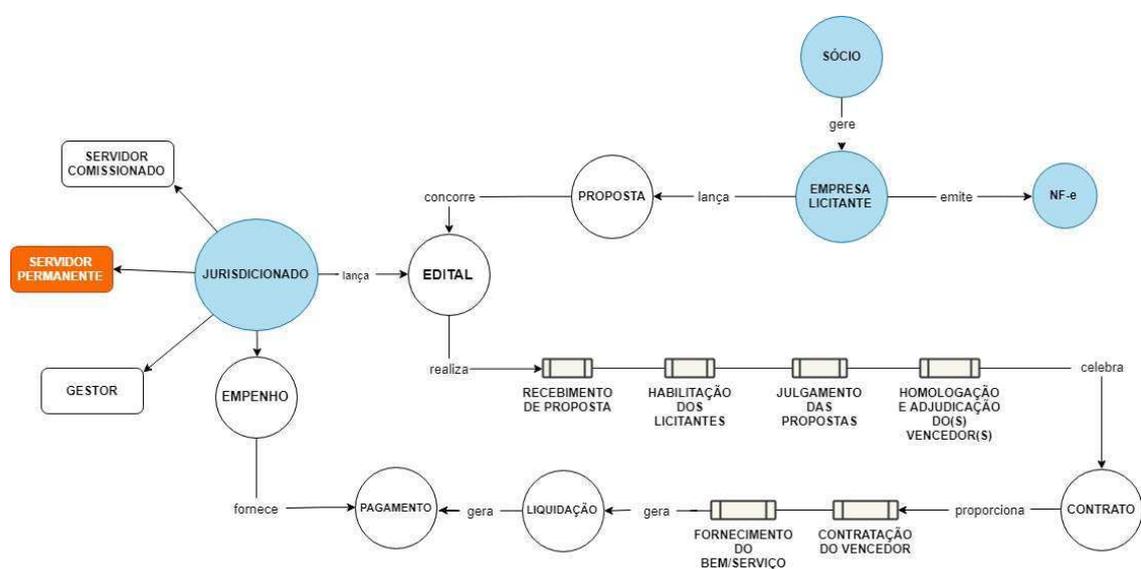


Figura B.6: Mapeamento da entidade Servidor Efetivo