



PROJETO DE GRADUAÇÃO

**PREDIÇÃO DE ENCERRAMENTO DE
MICRO E PEQUENAS EMPRESAS DA
INDÚSTRIA DE TRANSFORMAÇÃO**

Por,
Guilherme Monteiro de Castro Becker

BRASÍLIA, 28 DE SETEMBRO DE 2022.

UNIVERSIDADE DE BRASÍLIA

FACULDADE DE TECNOLOGIA DEPARTAMENTO DE
ENGENHARIA DE PRODUÇÃO

UNIVERSIDADE DE BRASÍLIA
Faculdade de Tecnologia
Departamento de Engenharia de Produção

PROJETO DE GRADUAÇÃO

PREDIÇÃO DE ENCERRAMENTO DE MICRO E PEQUENAS EMPRESAS DA INDÚSTRIA DE TRANSFORMAÇÃO

Por,

Guilherme Monteiro de Castro Becker

Relatório submetido como requisito parcial para
obtenção do grau de Engenheiro de Produção

Banca Examinadora

Prof. Dr. Reinaldo Crispiniano Garcia
UnB/ EPR(Orientador)

Prof. Dr. Sérgio Ronaldo Granemann
UnB/ EPR

Brasília-DF, 28 de setembro de 2022.

AGRADECIMENTOS

Agradeço aos meus pais, Ricardo e Andréa, pelo apoio incondicional e por serem tão presentes na minha vida. Por meio de seus exemplos, aprendo diariamente sobre a importância da educação, da ética, da dedicação e da empatia. Obrigado por sempre acreditarem em mim e me mostrarem que é possível buscarmos os nossos sonhos enquanto lutamos por um mundo mais justo e igualitário.

Estendo o agradecimento ao meu irmão, Gustavo, pela parceria e amizade. Saiba que pode contar comigo para o que precisar e tenho certeza que a recíproca é verdadeira. Sinto um orgulho imenso de ser seu irmão.

Agradeço à minha namorada e melhor amiga, Jordana, por todo companheirismo, cuidado e suporte, inclusive durante o Projeto de Graduação, em que seu apoio foi essencial. Você me inspira a ser uma pessoa melhor e sou muito grato por compartilhar a vida com você.

Agradeço também à toda a minha família, materna e paterna, que independentemente de qualquer distância, sei que sempre estarão ao meu lado. Um obrigado especial ao núcleo de Brasília, por dividirem tantos momentos importantes comigo.

Aos amigos que fiz ao longo da graduação, seja nas disciplinas, no Grupo Gestão, ou nos intervalos entre estas atividades, meu muito obrigado. Até mesmo as aulas de sexta-feira à noite ficavam mais leves com vocês. Agradeço em particular ao Felipe e ao Gabriel, que desde o primeiro semestre estiveram juntos comigo nesta jornada.

Em sequência, agradeço ao meu orientador Reinaldo pela oportunidade, pelos ensinamentos, e também por todo o apoio durante a construção do Projeto de Graduação. O seu entusiasmo em cada reunião que tivemos foi contagiante e me motivou a sempre buscar novos desafios.

Por fim, gostaria de agradecer à Universidade Brasília e a todos os professores e funcionários. A experiência que tive nos últimos 5 anos foi transformadora e me proporcionou aprendizados que carregarei durante toda a minha vida. Como diz Paulo Freire, “O educador se eterniza em cada ser que educa”.

RESUMO

Os pequenos negócios desempenham uma função essencial para o desenvolvimento socioeconômico do país. Ao gerarem um volume significativo de empregos e incentivarem as economias locais, estes empreendimentos promovem a redução das desigualdades e melhoram a qualidade de vida da população. Entretanto, os elevados índices de encerramento das micro e pequenas empresas, sobretudo nos primeiros anos de atividade, evidenciam um problema complexo e revelam a necessidade de ações preventivas. Este cenário é ainda mais crítico para o grupo de micro e pequenas empresas da indústria de transformação, uma vez que o setor não vem apresentando bons resultados nos últimos anos. Diante disso, o presente trabalho tem como objetivo prever o encerramento das micro e pequenas empresas da indústria de transformação brasileira, com base nos dados públicos disponibilizados pela Receita Federal do Brasil. Para tanto, foram identificados e analisados os fatores relacionados com o fechamento destas empresas e, a partir dos atributos levantados, aplicou-se o algoritmo de regressão logística. O modelo desenvolvido obteve uma acurácia média de 70,6% e uma área sob a curva ROC média de 70,35%, de forma que os resultados demonstram a sua capacidade de predição.

Palavras-chave: Micro e pequenas empresas; Indústria de transformação; Fatores de encerramento; Regressão logística; Predição de encerramento.

ABSTRACT

The micro and small businesses play an essential role in a country's socio-economic development. By generating a significant amount of jobs and enhancing the local economies, these companies are able to reduce the inequality and to improve the population's life quality. However, the high closure rate of these firms, especially in the first few years, reveal a complex problem that needs to be solved. This situation is even worse for small businesses in the processing industry, due to its recent negative results. Therefore, this study aims to predict the businesses failure for micro and small Brazilian processing industries, based on a public dataset provided by the government. For this purpose, factors related to the closure of these companies were identified and analysed. Subsequently, these characteristics were applied to develop a model based on logistic regression algorithm. The developed model obtained an average accuracy of 70.6% and an average area under the ROC curve of 70.35%, showing its prediction capability.

Keywords: Micro and small businesses; Processing industry; Failure factors; Logistic Regression; Business failure prediction.

LISTA DE FIGURAS

Figura 1 – Exemplo de código CNAE da indústria de transformação	24
Figura 2 – Exemplo de Comprovante de Inscrição e de Situação Cadastral.....	28
Figura 3 – Grupos de fatores relacionados ao encerramento das MPEs	32
Figura 4 – Curva Logística	42
Figura 5 – Gráfico da Área Sob a Curva ROC	44
Figura 6 – Tabelas do Cadastro Nacional de Pessoa Jurídica	46
Figura 7 – Análise da situação cadastral das empresas	48
Figura 8 – Gráfico de caixa com a variável Capital Social	52
Figura 9 – Classificação do tipo de variáveis	55

LISTA DE QUADROS

Quadro 1 – Número de estabelecimentos por subsetor da Indústria de Transformação	21
Quadro 2 – Classificação das empresas com base em seu faturamento	25
Quadro 3 – Tipos de situações cadastrais das empresas.....	29
Quadro 4 – Matriz de Confusão	43
Quadro 5 – Variáveis selecionadas para o modelo de predição de encerramento das MPEITs	53
Quadro 6 – Parâmetros para aplicação do modelo de regressão logística.....	58

LISTA DE TABELAS

Tabela 1 – Comparação do número de empresas ativas e encerradas por sócio	60
Tabela 2 – Taxa de mortalidade de acordo com a faixa etária dos sócios.....	60
Tabela 3 – Número médio de CNAEs secundários, por situação cadastral	63
Tabela 4 – Taxa de mortalidade de acordo com o porte da empresa.....	63
Tabela 5 – Avaliação do Número de Estabelecimentos Ativos e Encerrados no mesmo CEP	66
Tabela 6 – Resultados alcançados com o primeiro teste do modelo	68

LISTA DE GRÁFICOS

Gráfico 1 – Número de trabalhadores formais da indústria de transformação (mil).....	22
Gráfico 2 – Participação da Indústria de Transformação no PIB	22
Gráfico 3 – Percentual de estabelecimentos da Indústria de Transformação de acordo com o porte	26
Gráfico 4 – Taxa de mortalidade das MPEITs de acordo com qualificação dos sócios	61
Gráfico 5 – Capital Social médio por situação cadastral.....	62
Gráfico 6 – Número médio de sócios, por situação cadastral	64
Gráfico 7 – Avaliação da região da localização e da população de seu estado.....	64
Gráfico 8 – Avaliação do número de estabelecimentos ativos e encerrados no mesmo CEP..	65
Gráfico 9 – Avaliação da opção pelo Sistema Simples Nacional.....	67
Gráfico 10 – Comparação do desempenho do modelo com as duas bases de teste	68

LISTA DE SIGLAS

AUC – *Area Under the ROC Curve*

CNAE – Classificação Nacional de Atividades Econômicas

CNPJ – Cadastro Nacional de Pessoas Jurídicas

CONCLA – Comissão Nacional de Classificação

EIRELI – Empresa Individual de Responsabilidade Limitada de Natureza Empresária

EPP – Empresas de Pequeno Porte

IBGE – Instituto Brasileiro de Geografia e Estatística

ME – Microempresas

MEI – Microempreendedores Individuais

MPEITs – Micro e Pequenas Empresas da Indústria de Transformação

MPEs – Micro e Pequenas Empresas

PIB – Produto Interno Bruto

PO – Pesquisa Operacional

PRONAMPE – Programa Nacional de Apoio às Microempresas e Empresas de Pequeno Porte

RFB – Receita Federal do Brasil

ROC – *Receiver Operating Characteristic*

SEBRAE – Serviço Brasileiro de Apoio às Micro e Pequenas Empresas

SEBRAE-SP – Serviço Brasileiro de Apoio às Micro e Pequenas Empresas do Estado de São Paulo

SMOTE – *Synthetic Minority Oversampling Technique*

SUMÁRIO

1	INTRODUÇÃO.....	13
1.1	CONTEXTUALIZAÇÃO	13
1.2	JUSTIFICATIVA	14
1.3	OBJETIVO	16
1.3.1	Objetivo Geral.....	16
1.3.2	Objetivos Específicos	16
1.4	METODOLOGIA.....	16
1.4.1	Classificação da Pesquisa	16
1.4.2	Etapas da Pesquisa	17
1.5	ESTRUTURA DO TRABALHO.....	18
2	REFERENCIAL TEÓRICO	20
2.1	INDÚSTRIA DE TRANSFORMAÇÃO.....	20
2.2	CÓDIGO CNAE.....	23
2.3	MICRO E PEQUENAS EMPRESAS	24
2.4	MICRO E PEQUENAS EMPRESAS DA INDÚSTRIA DE TRANSFORMAÇÃO..	25
2.5	ÍNDICES DE MORTALIDADE DAS MPES	26
2.5.1	Comprovante de Inscrição e de Situação Cadastral.....	27
2.5.1.1	Situação Cadastral.....	28
2.5.1.2	Matriz e Filial.....	29
2.5.1.3	Natureza Jurídica	30
2.6	FATORES RELEVANTES PARA O FECHAMENTO DAS MPES	31
2.6.1	Características do empreendedor	32
2.6.2	Atributos da empresa	33
2.6.3	Condições Externas	35
2.7	PREVISÃO DE ENCERRAMENTO DAS MPES	36
2.7.1	Pré-processamento dos dados.....	37
2.7.1.1	Técnica de Balanceamento dos Dados	39
2.7.2	Engenharia de Atributos	39
2.7.3	Modelos de Classificação.....	40
2.7.3.1	Regressão Logística	41
2.7.4	Avaliação do desempenho do modelo	42

3	DESENVOLVIMENTO DO MODELO.....	45
3.1	COLETA DOS DADOS.....	45
3.2	PREMISSAS ADOTADAS.....	47
3.3	TRATAMENTO INICIAL DOS DADOS.....	50
3.4	SELEÇÃO E CRIAÇÃO DE ATRIBUTOS.....	52
3.5	TRATAMENTO FINAL DOS DADOS.....	55
3.6	APLICAÇÃO E AVALIAÇÃO DO DESEMPENHO DO MODELO.....	57
4	RESULTADOS.....	59
4.1	ANÁLISE DOS ATRIBUTOS DA EMPRESA.....	59
4.1.1	Características do Empreendedor.....	59
4.1.2	Atributos da Empresa.....	61
4.1.3	Fatores Externos.....	65
4.2	AVALIAÇÃO DAS PREDIÇÕES.....	67
5	CONCLUSÃO.....	69
5.1	ATENDIMENTO DOS OBJETIVOS.....	69
5.2	LIMITAÇÕES DO TRABALHO E OPORTUNIDADE PARA FUTURAS PESQUISA	
	70	
	REFERÊNCIAS.....	71
	APÊNDICE A – CÓDIGO UTILIZADO PARA DISPONIBILIZAÇÃO E	
	TRATAMENTO DA BASE DE DADOS DA PESQUISA.....	77
	APÊNDICE B – CÓDIGO UTILIZADO PARA O DESENVOLVIMENTO E PARA A	
	AVALIAÇÃO DO MODELO DE PREDIÇÃO DE ENCERRAMENTO.....	84

1 INTRODUÇÃO

Este capítulo tem como objetivo introduzir o tema abordado no trabalho. Dessa forma, será apresentado o seu contexto, seguido por uma justificativa de sua relevância, dos objetivos gerais e específicos que se pretende alcançar com o estudo, da metodologia de pesquisa adotada e, ao final, haverá uma breve descrição da estrutura do documento.

1.1 CONTEXTUALIZAÇÃO

As micro e pequenas empresas (MPEs), classificadas dessa forma a partir de seu faturamento anual, exercem um papel fundamental na sociedade brasileira, o que pode ser comprovado pelos seus números. Segundo o Serviço Brasileiro de Apoio às Micro e Pequenas Empresas (SEBRAE, 2019), elas são responsáveis por cerca de 27% do Produto Interno Bruto (PIB) do país e pela geração de mais da metade dos empregos formais. Somente entre os anos de 2006 e 2016, estes empreendimentos criaram mais de 5 milhões de postos de trabalho com carteira assinada (SEBRAE; DIEESE, 2018).

Além da comprovada geração de empregos, as MPEs apresentam uma importante característica quanto à sua manutenção. Mesmo em um cenário de enfraquecimento do mercado, os índices de demissão nestas empresas são inferiores aos observados nas de médio ou grande porte. Em 2015, enquanto elas encerraram 211.400 vagas de emprego com carteira assinada, as médias e grandes demitiram 1.315.308 trabalhadores, cerca de 522% a mais (FGV, 2016).

Estas propriedades tornam as MPEs capazes de minimizar as oscilações de oferta e demanda das empresas de grande porte e, com isso, garantir a empregabilidade de muitos trabalhadores. No entanto, ao desempenharem estas funções, ficam sujeitas à baixa lucratividade e à elevada taxa de mortalidade, fatores críticos que impedem que seu impacto seja ainda maior (GOVERNO FEDERAL, 2019).

Dentro do grupo de micro e pequenas empresas, os empreendimentos compartilham muitos de seus desafios. A dificuldade de sobrevivência durante os primeiros anos de atividade e a falta de gestão profissional são fatores amplamente presentes, independente do segmento em que estão inseridas (SEBRAE, 2021). Entretanto, as empresas também apresentam características específicas de acordo com a sua área de atuação.

No Brasil, classifica-se uma empresa como pertencente a determinado setor a partir da atividade principal que exerce, conforme o código de Classificação Nacional de Atividades

Econômicas (CNAE). Entre eles, pode-se destacar o papel e a relevância da Indústria de Transformação.

Este setor, composto por mais de 90% de MPEs, é um dos mais importantes da indústria nacional. Apesar disso, nos últimos anos, a sua participação no PIB e no número de empregos formais vem se reduzindo de maneira significativa. Além disso, sua taxa de mortalidade aumentou, com 27,3% das empresas fechando em até 5 anos (SEBRAE, 2020). Desse modo, esforços para combater as tendências apresentadas são essenciais.

Também é importante ressaltar o papel que as micro e pequenas empresas da indústria de transformação (MPEITs) exercem nas suas regiões. Como grande parte dos postos de trabalho são ocupados por pessoas que vivem próximas ao empreendimento, estas empresas estimulam o desenvolvimento da economia local e, com isso, contribuem para o aumento da qualidade de vida da população (SARAIVA et al., 2020).

Por outro lado, estas características também evidenciam o elevado custo social ocasionado pelo seu fechamento. Segundo Pereira e Souza (2019), uma das consequências diretas do encerramento de pequenos negócios é a necessidade de ampliar os recursos públicos destinados a programas sociais para famílias de baixa ou nenhuma renda. Os autores ainda destacam que outros agentes, direta ou indiretamente, também são prejudicados, como: fornecedores, clientes, distribuidores, bancos e governos. A redução na arrecadação de impostos e no número de empregos indiretos são exemplos destas perdas.

Portanto, a fim de direcionar as políticas públicas e privadas e garantir que as micro e pequenas empresas da indústria de transformação tenham as condições necessárias para manutenção de suas atividades, torna-se imperativo analisar este tema e desenvolver soluções que contribuam positivamente para este cenário.

1.2 JUSTIFICATIVA

Conforme evidenciado na seção anterior, as micro e pequenas empresas da indústria de transformação são importantes agentes para o desenvolvimento socioeconômico do Brasil. Seu forte poder de geração de emprego e de renda faz com que elas tenham um papel fundamental na redução das desigualdades. De acordo com Luíz Barretto e Valdir Oliveira (2022), respectivamente ex-presidente e superintendente do Sebrae, “não há o que falar em distribuição de renda sem o fomento aos pequenos negócios”.

No entanto, sabe-se que os elevados índices de mortalidade ainda são um dos principais desafios enfrentados por estes empreendimentos, sobretudo nos primeiros anos de atividade,

em que as MPEs ainda buscam se consolidar no mercado e precisam suportar dificuldades financeiras. Com relação à indústria de transformação, tendo em vista o enfraquecimento do setor nos últimos anos, este cenário torna-se ainda mais crítico (SEBRAE, 2020).

Diante deste contexto, observa-se os ganhos gerados ao aumentar o número de pequenos negócios e, principalmente, a necessidade de evitar que encerrem suas atividades após um breve período. Há diversas abordagens que podem ser adotadas com esse objetivo, porém deve-se destacar os modelos de previsão de encerramento das empresas. Segundo Pompe e Bilderbeek (2005), essa área de pesquisa tem crescido em diferentes países e vem se tornando cada vez mais relevante para análise e acompanhamento das pequenas empresas. Os benefícios destas previsões podem ser estendidos para diferentes grupos.

No que se refere aos empresários, um modelo de previsão de encerramento pode auxiliar na compreensão dos riscos associados ao seu negócio. Somado a isso, pode ajudar na obtenção de recursos necessários para sua manutenção (BAIDOUN et al., 2018).

Estudos desta natureza podem ser utilizados como base para o desenvolvimento de políticas públicas destinadas às micro e pequenas empresas. De acordo com Marom e Lusier (2014), os governos entendem a relevância dos pequenos negócios e tendem a investir uma verba significativa neles. No entanto, os autores avaliam que estes recursos não vêm sendo alocados de forma eficiente e entendem que modelos de previsão podem tornar estas medidas mais eficazes.

Deve-se destacar ainda a influência positiva que isso poderá gerar para outros grupos de *stakeholders*, incluindo potenciais sócios e investidores, pessoas que desejam abrir o seu negócio futuramente, instituições que fornecem capital para o desenvolvimento dessas empresas, fornecedores e, de maneira geral, todos que, direta ou indiretamente, são impactados por elas (BAIDOUN et al., 2018).

Apesar de sua relevância, o encerramento de micro e pequenas empresas da indústria de transformação ainda é pouco discutido na literatura, principalmente no Brasil, onde há um número reduzido de pesquisas relacionadas a este tema. Tratando-se especificamente dos modelos de previsão de fechamento ou sucesso das MPEITs, o número de estudos é ainda mais escasso, revelando que há necessidade e espaço para o desenvolvimento de novas pesquisas na área.

Diante deste contexto, pode-se resumir o problema de pesquisa a partir da seguinte pergunta: É possível analisar e prever o encerramento das micro e pequenas empresas brasileiras da indústria de transformação?

1.3 OBJETIVO

Em seguida, será apresentado o objetivo geral da pesquisa, bem como os objetivos específicos que deverão ser cumpridos a fim de alcançá-lo.

1.3.1 Objetivo Geral

Desenvolver um modelo de predição de encerramento de micro e pequenas empresas brasileiras da indústria de transformação.

1.3.2 Objetivos Específicos

- Identificar, a partir de revisão da literatura, fatores que impactam na sobrevivência ou no encerramento das MPEITs;
- Reunir dados de MPEITs brasileiras em atividade e encerradas;
- Criar e selecionar variáveis a partir do conjunto de dados da pesquisa e dos fatores identificados;
- Programar o modelo de predição de encerramento das MPEITs;
- Validar o desempenho do modelo ao aplicá-lo em MPEITs diferentes daquelas em que foi treinado.

1.4 METODOLOGIA

Nesta seção, será apresentada a metodologia adotada para o desenvolvimento do presente trabalho, por meio da classificação da pesquisa e da descrição das etapas seguidas a fim de atingir os objetivos propostos.

1.4.1 Classificação da Pesquisa

As pesquisas podem ser classificadas de acordo com quatro critérios distintos: quanto à abordagem, natureza, objetivos e procedimentos (SILVEIRA; CÓRDOVA, 2009).

Com relação à abordagem, uma pesquisa pode ser definida como qualitativa ou quantitativa. Na primeira, busca-se explicar um problema a partir avaliação de aspectos

subjetivos e de fontes de dados não-numéricas. Por outro lado, nas pesquisas quantitativas, utiliza-se dados numéricos e recorre-se à linguagem matemática com o objetivo de descrever um fenômeno e as relações entre as variáveis.

No que se refere à natureza, a pesquisa poderá ser avaliada como básica ou aplicada. A primeira classificação é utilizada para pesquisas que visam gerar novos conhecimentos sem aplicação prática prevista. Já a segunda é adotada em pesquisas que possuem aplicação prática e são direcionadas à solução de um problema específico.

Quanto aos objetivos, a pesquisa será classificada como exploratória caso o seu propósito seja tornar um problema mais explícito, por meio da elaboração de hipóteses. Já a pesquisa descritiva apresenta os fatos e fenômenos de determinada realidade, enquanto a explicativa busca identificar fatores que contribuem para a ocorrência de fenômenos (GIL, 2007).

Para categorizá-la com relação ao procedimento, deve-se avaliar a técnica aplicada no seu desenvolvimento. No âmbito da Engenharia de Produção, entre as técnicas mais utilizadas, inclui-se o levantamento, o estudo de caso, a modelagem matemática, a simulação, o experimento, e a teórico-conceitual (FILIPPINI, 1997).

Diante dos conceitos apresentados, pode-se classificar a presente pesquisa como quantitativa, aplicada, explicativa e de modelagem matemática, uma vez que são utilizadas técnicas estatísticas para desenvolver um modelo preditivo.

Para realizar a modelagem, utilizou-se as bibliotecas *Pandas*, *Scikit-Learn*, *Scipy.Stats*, *Sqlite3* e *Numpy*, da linguagem de programação *Python*. Estas bibliotecas possuem funções que possibilitam a extração e manipulação dos dados, além da aplicação e avaliação dos modelos. O *software* Microsoft Excel também foi utilizado para complementar as análises e desenvolver visualizações gráficas.

1.4.2 Etapas da Pesquisa

Visando cumprir os objetivos definidos para o presente trabalho, foram definidos os seguintes passos metodológicos:

- Revisão bibliográfica sobre características das micro e pequenas empresas da indústria da transformação e fatores relacionados à sua elevada taxa de mortalidade, além de modelos utilizados para prever o seu encerramento e as respectivas técnicas de avaliação;

- Levantamento de dados que contenham informações públicas sobre empresas ativas e encerradas;
- Disponibilização das informações em um banco de dados local que possibilite a extração e manipulação delas;
- Definição das premissas e critérios que serão adotados para o desenvolvimento do modelo, definidos a partir das características do conjunto de dados e de revisão bibliográfica;
- Tratamento da base de dados para corrigir possíveis inconsistências e adequá-la aos requisitos do algoritmo escolhido;
- Seleção dos atributos da base de dados que serão utilizados para construção do modelo, além da criação de novas variáveis a partir do processo de engenharia de atributos;
- Aplicação e avaliação do desempenho do modelo de predição de encerramento das MPEITs, de acordo com os critérios definidos a partir de revisão bibliográfica.

1.5 ESTRUTURA DO TRABALHO

O presente trabalho está dividido em cinco capítulos que visam explorar o tema e o modelo aplicado.

- I. O primeiro capítulo traz uma introdução do estudo e contém a contextualização do problema, a justificativa, os objetivos gerais e específicos e a metodologia adotada.
- II. O segundo capítulo apresenta o referencial teórico sobre o tema, contendo conceitos sobre as micro e pequenas empresas da indústria de transformação, fatores relacionados ao seu sucesso ou fracasso e métodos utilizados para predição de seu encerramento.
- III. O terceiro capítulo apresenta, de forma detalhada, todas as atividades realizadas a fim desenvolver o modelo de predição de encerramento das micro e pequenas empresas de transformação, abordando desde o processo de coleta dos dados, até a avaliação de desempenho do modelo.

- IV. O quarto capítulo reúne os resultados obtidos no presente estudo. Ele está dividido entre a análise dos fatores selecionados e a avaliação das predições realizadas.
- V. O quinto e último capítulo apresenta a conclusão, as limitações e as recomendações para futuros trabalhos.

2 REFERENCIAL TEÓRICO

A revisão da literatura é um processo fundamental no desenvolvimento de uma pesquisa científica. De acordo com Gil (2007), ela permite ampliar a compreensão sobre o problema estudado, aprimorar as ideias previamente levantadas e, até mesmo, levar à descoberta de novas intuições.

Dessa forma, o presente capítulo tem como objetivo apresentar a fundamentação teórica que embasou o desenvolvimento do projeto. Para isso, serão abordados os principais conceitos relativos a este estudo, desde as classificações das empresas, até o levantamento de fatores apontados na literatura como relevantes para o seu encerramento e os modelos utilizados para prevê-lo.

2.1 INDÚSTRIA DE TRANSFORMAÇÃO

De acordo com o Instituto Brasileiro de Geografia e Estatística (IBGE, 2022), a Indústria de Transformação compreende “atividades que envolvem a transformação física, química e biológica de materiais, substâncias e componentes com a finalidade de se obterem produtos novos”.

Estas atividades podem ser divididas então em três grupos, de acordo com a sua finalidade. As indústrias de bens de capital produzem produtos que serão utilizados no processo produtivo de outras empresas, como máquinas e ferramentas. Já as de bens intermediários fabricam produtos que serão transformados novamente por outras indústrias, enquanto as de bens de consumo encaminham a produção diretamente ao mercado consumidor (CNI, 2022).

De maneira geral, estas atividades são desenvolvidas em plantas industriais ou fábricas, por meio da utilização de máquinas e outros equipamentos que possibilitam a manipulação e transformação dos materiais. No entanto, mesmo aquelas realizadas em domicílios, como as produções manuais e artesanais, são enquadradas neste ramo da indústria (IBGE, 2022).

No Brasil, existem 24 subsetores diferentes que fazem parte da Indústria de Transformação. Juntos, eles são responsáveis por 11,3% do Produto Interno Bruto (PIB) brasileiro, por quase 15% do número total de empregos formais e por 46,2% das exportações de bens e serviços (CNI, 2022). No Quadro 1, observa-se o número de estabelecimentos por subsetor da Indústria de Transformação, conforme dados disponibilizados pelo Ministério do Trabalho e Emprego (2016).

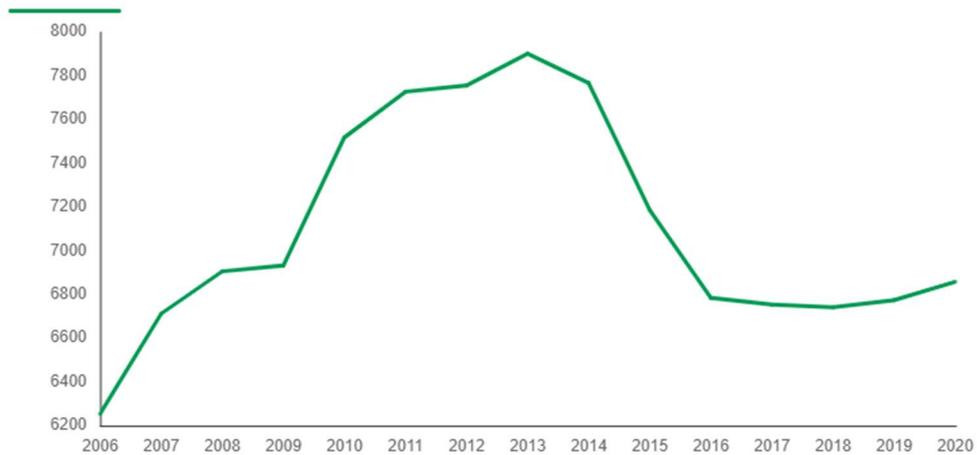
Quadro 1 – Número de estabelecimentos por subsetor da Indústria de Transformação

Subsetor da Indústria de Transformação	Total de Estabelecimentos	Participação Percentual
Confecção de Artigos do Vestuário e Acessórios	49.133	14,9%
Produtos Alimentícios	47.811	14,5%
Produtos de Metal, exc. Máquinas e Equipamentos	38.479	11,6%
Produtos de minerais não-metálicos	26.945	8,1%
Manutenção de Máquinas e Equipamentos	22.215	6,7%
Móveis	21.157	6,4%
Produtos de Madeira	13.811	4,2%
Produtos de Borracha e de Material Plástico	13.736	4,2%
Máquinas e Equipamentos	13.695	4,1%
Produtos Diversos	13.341	4,0%
Impressão e Reprodução de Gravações	13.309	4,0%
Artefatos de Couro, Artigos para Viagem e Calçados	10.958	3,3%
Produtos Têxteis	10.153	3,1%
Produtos Químicos	8.990	2,7%
Veículos Automotores, Carrocerias e Autopeças	6.120	1,9%
Máquinas, Aparelhos e Materiais Elétricos	4.475	1,4%
Celulose, Papel e Produtos de Papel	4.265	1,3%
Metalurgia	3.584	1,1%
Equipamentos de Informática, Produtos Eletrônicos e Ópticos	3.246	1,0%
Bebidas	2.656	0,8%
Outros Equipamentos de Transporte, exc. Veículos Automotores	1.164	0,4%
Produtos Farmoquímicos e Farmacêuticos	808	0,2%
Coque, Derivados do Petróleo e Biocombustíveis	547	0,2%
Produtos do Fumo	203	0,1%
Total da Indústria de Transformação	330.801	100%

Fonte: Adaptado de Ministério do Trabalho e Emprego (2016)

Apesar destes dados mostrarem a importância da Indústria de Transformação para a economia brasileira, alguns dos seus indicadores mais relevantes apresentaram quedas significativas nos últimos anos. De 2013 a 2020, o número de trabalhadores formais se reduziu em mais de 13% o que, em termos absolutos, representa 1.042.600 empregados a menos no setor (CNI, 2022), conforme indicado no Gráfico 1.

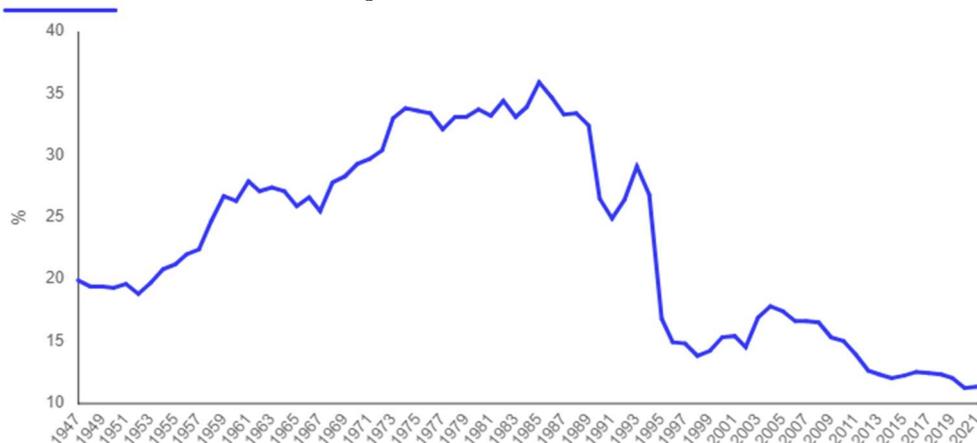
Gráfico 1 – Número de trabalhadores formais da indústria de transformação (mil)



Fonte: CNI (2022)

Além disso, mesmo com o PIB desta indústria apresentando crescimentos significativos nos últimos anos – e chegando a um total de R\$ 837,2 bilhões em 2021 – a sua participação percentual no PIB brasileiro diminuiu de forma considerável (CNI, 2022), conforme apresentado no Gráfico 2.

Gráfico 2 – Participação da Indústria de Transformação no PIB



Fonte: CNI (2022)

Os resultados destes indicadores são alarmantes e impactam significativamente o Brasil. De acordo com o Instituto de Estudos para o Desenvolvimento Industrial (2018), há uma forte relação entre os resultados econômicos de um país e o desenvolvimento de suas indústrias. O artigo ainda destaca que a melhoria de aspectos sociais e os ganhos de produtividade para todo o sistema econômico são consequências diretas desta evolução.

Por fim, Dadush (2015) afirma que, independentemente de seu porte, as Indústrias de Transformação se mantêm essenciais para o desenvolvimento econômico e social e, por isso, devem ser valorizadas e incentivadas.

2.2 CÓDIGO CNAE

Com o objetivo de classificar as organizações quanto às atividades desempenhadas, o Governo Federal oficializou, no ano de 1995, a tabela de código e denominações da Classificação Nacional de Atividades Econômicas (CNAE).

Ela é monitorada e padronizada pela Comissão Nacional de Classificação (CONCLA) e, de acordo com a Receita Federal (2001), deve ser “aplicada a todos os agentes econômicos que estão engajados na produção de bens e serviços, podendo compreender estabelecimentos de empresas privadas ou públicas, estabelecimentos agrícolas, organismos públicos e privados, instituições sem fins lucrativos e agentes autônomos (pessoa física)”.

Desse modo, o código permite a classificação e identificação de todos os estabelecimentos brasileiros, sendo composto por 21 seções. Cada uma delas contém divisões com os seus respectivos grupos, que por sua vez são divididos em classes e subclasses. De acordo com esta definição, todos os estabelecimentos que compreendem a Indústria de Transformação fazem parte da seção “C” do código CNAE. Na Figura 1, pode-se observar um exemplo de busca no site do IBGE pela subclasse “Fabricação de margarina e outras gorduras vegetais e de óleos não comestíveis de animais”.

Figura 1 – Exemplo de código CNAE da indústria de transformação

Hierarquia	
Seção:	C INDÚSTRIAS DE TRANSFORMAÇÃO
Divisão:	10 FABRICAÇÃO DE PRODUTOS ALIMENTÍCIOS
Grupo:	10.4 Fabricação de óleos e gorduras vegetais e animais
Classe:	10.43-1 Fabricação de margarina e outras gorduras vegetais e de óleos não comestíveis de animais
Subclasse:	1043-1/00 Fabricação de margarina e outras gorduras vegetais e de óleos não comestíveis de animais

Notas Explicativas:
Esta classe compreende:

- a fabricação de margarina e de outras gorduras vegetais, comestíveis ou não
- a fabricação de preparações à base de creme vegetal
- a fabricação de óleos não-comestíveis de origem animal
- a extração de óleos de peixe e de mamíferos marinhos

Fonte: IBGE (2022)

Cabe ressaltar que um estabelecimento pode desempenhar mais de uma atividade econômica e, por isso, ter mais de um código CNAE. Entretanto, uma dessas atividades precisa ser definida como a principal, enquanto as outras serão classificadas como secundárias. De acordo com a Receita Federal (2006), a atividade principal é aquela que possui maior receita auferida ou esperada.

2.3 MICRO E PEQUENAS EMPRESAS

No Brasil, existem dois critérios distintos que podem ser utilizados para definir o porte de uma empresa que são o número de funcionários e o faturamento. O primeiro critério é definido pelo Sebrae (2013) e, de acordo com a sua classificação, as Microempresas (ME) podem ter até 9 empregados, caso sejam de comércio ou serviços, e até 19, se fizerem parte do setor industrial. Da mesma forma, as Empresas de Pequeno Porte (EPP) de comércio ou serviço podem ter entre 10 e 49 empregados, enquanto as indústrias de 20 a 99.

No entanto, como o critério de faturamento está previsto no Estatuto Nacional da Microempresa e da Empresa de Pequeno Porte, é o mais utilizado na identificação das MPEs. Além disso, esta é a classificação oficial adotada por importantes órgãos públicos, como a Receita Federal do Brasil, e pode ser observada com mais detalhes no Quadro 2 (CNI, 2022).

Quadro 2 – Classificação das empresas com base em seu faturamento

Classificação	Receita Bruta Anual
Microempresa	Até R\$ 360 mil
Pequena Empresa	Maior que R\$ 360 mil e menor ou igual a R\$ 4,8 milhões
Empresa de Médio Porte	Maior que R\$ 4,8 milhões e menor ou igual a 300 milhões
Empresa de Grande Porte	Maior que 300 milhões

Fonte: Adaptação do CNI (2022)

Apesar de possuírem um faturamento inferior ao das empresas de médio e grande porte, as MPEs desempenham um papel significativo no cenário nacional. Juntas, elas são responsáveis por 52% de todos os empregos formais existentes no Brasil. Ademais, representam 99% do total de empresas privadas, respondem por 27% do Produto Interno Bruto Brasileiro e geram 40% da massa de salários pagos aos trabalhadores (SEBRAE, 2019).

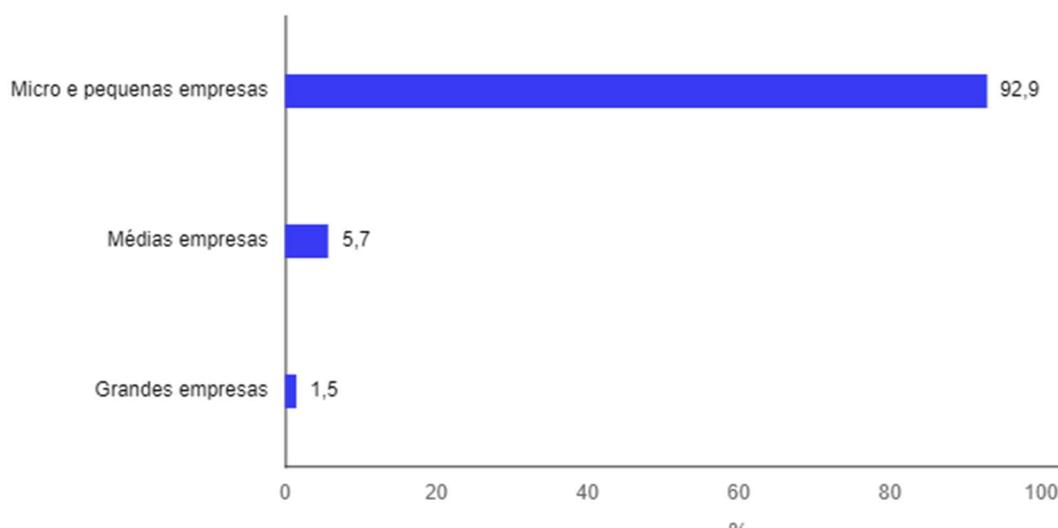
Deve-se destacar ainda que, segundo dados oficiais do Governo Federal (2022), foram abertas 579,5 mil microempresas no ano de 2020, um acréscimo de mais de 7% em comparação com 2018. Por outro lado, as empresas de pequeno porte tiveram um crescimento de quase 26%, chegando a um total de 94,3 mil em 2020.

2.4 MICRO E PEQUENAS EMPRESAS DA INDÚSTRIA DE TRANSFORMAÇÃO

Conforme abordado nos tópicos anteriores, são consideradas Indústrias de Transformação as empresas cuja atividade principal está contida na seção “C” do código CNAE. Já o porte de uma empresa é definido com base em sua receita anual, sendo que as ME devem ter até R\$ 360 mil de faturamento, enquanto as EPP devem faturar no máximo R\$ 4,8 milhões. Ao combinar estes dois critérios, chega-se à definição das Micro e Pequenas Empresas da Indústria de Transformação (MPEITs) (SEBRAE, 2017).

De acordo com um levantamento realizado pela Confederação Nacional da Indústria (2022), havia 268.006 MPEITs ativas no território nacional no ano de 2020, distribuídas nos 26 estados e no Distrito Federal. Este valor representa mais de 90% dos estabelecimentos da Indústria de Transformação e destaca o papel das micro e pequenas empresas para o setor, conforme apresentado no Gráfico 3.

Gráfico 3 – Percentual de estabelecimentos da Indústria de Transformação de acordo com o porte



Fonte: CNI (2022)

As MPEITs compartilham muitas de suas características com aquelas associadas às micro e pequenas empresas em um contexto mais geral. Via de regra, a administração dessas indústrias é feita de maneira mais intuitiva pelos gestores que, além de desempenharem esse papel, são seus donos (SARAIVA et al., 2020). Ressalta-se ainda o expressivo número de empregos que as micro e pequenas indústrias da transformação geram e o seu importante papel no desenvolvimento local, uma vez que grande parte dos postos de trabalho são ocupados por pessoas que vivem no entorno do empreendimento.

Assim, por possuírem características semelhantes e estarem inseridas em contextos próximos, inclusive com relação aos tipos de desafios enfrentados, entende-se que as micro e pequenas indústrias de transformação estão contidas no universo das micro e pequenas empresas (SARAIVA et al., 2020). Portanto, ao longo deste estudo, deve-se assumir que qualquer menção feita às MPes também abrange as MPEITs.

2.5 ÍNDICES DE MORTALIDADE DAS MPES

A alta taxa de mortalidade das micro e pequenas empresas é citada por diferentes autores como uma das principais dificuldades enfrentadas por este segmento. Simplificadamente, esta medida pode ser obtida ao calcular, do total de empresas constituídas em um determinado período, quantas delas encerraram as suas atividades. (FERREIRA et al., 2012; PINHEIRO; NETO, 2019).

No Brasil, o Sebrae realiza pesquisas sobre a sobrevivência e mortalidade das empresas desde 1999. Inicialmente, elas eram feitas somente no Estado de São Paulo e traziam dados relevantes sobre as empresas da região. Em um destes estudos, foram avaliadas as empresas abertas entre 2003 e 2007 e descobriu-se que 27% delas fechavam em até um ano de operação e, no terceiro ano de atividade, este percentual chegava a 46% (SEBRAE-SP, 2010).

Recentemente, as pesquisas do Sebrae passaram a compreender empresas localizadas em todo o território nacional e, com isso, se tornaram mais abrangentes e fidedignas à realidade brasileira. Dessa forma, foi possível constatar que, apesar de ser um problema comum a todas as empresas, os índices de mortalidade são muito superiores naquelas de micro e pequeno portes (SEBRAE, 2016). Em 2020, a taxa de mortalidade de microempresas de 1 a 5 anos chegou a 21,6% e a de empresas de pequeno porte foi de 17% (SEBRAE, 2020).

Além disso, ao analisar a sobrevivência das empresas por setor, observa-se que a indústria de transformação apresenta números preocupantes. Segundo o mesmo levantamento desenvolvido pelo SEBRAE (2020), 27,3% das empresas desse segmento fecham em até 5 anos.

2.5.1 Comprovante de Inscrição e de Situação Cadastral

A fim de mensurar os índices de mortalidade das empresas, diferentes estudos se baseiam no Comprovante de Inscrição e de Situação Cadastral. Este é um documento oficial do Governo Federal, aprovado pela Instrução Normativa nº 1.863, e reúne informações sobre as empresas registradas no sistema público (RECEITA FEDERAL DO BRASIL, 2018). Um exemplo deste documento preenchido pode ser observado na Figura 2.

Figura 2 – Exemplo de Comprovante de Inscrição e de Situação Cadastral

 REPÚBLICA FEDERATIVA DO BRASIL CADASTRO NACIONAL DA PESSOA JURÍDICA			
NÚMERO DE INSCRIÇÃO 00.038.174/0001-43 MATRIZ	COMPROVANTE DE INSCRIÇÃO E DE SITUAÇÃO CADASTRAL		DATA DE ABERTURA 29/12/1966
NOME EMPRESARIAL FUNDAÇÃO UNIVERSIDADE DE BRASÍLIA			
TÍTULO DO ESTABELECIMENTO (NOME DE FANTASIA) *****			PORTE DEMAIS
CÓDIGO E DESCRIÇÃO DA ATIVIDADE ECONÔMICA PRINCIPAL 85.31-7-00 - Educação superior - graduação			
CÓDIGO E DESCRIÇÃO DAS ATIVIDADES ECONÔMICAS SECUNDÁRIAS Não informada			
CÓDIGO E DESCRIÇÃO DA NATUREZA JURÍDICA 113-9 - Fundação Pública de Direito Público Federal			
LOGRADOURO ST CAMPUS UNIVERSITARIO	NÚMERO S N	COMPLEMENTO *****	
CEP 70.310-500	BARRIO/DISTRITO A SA NORTE	MUNICÍPIO BRASÍLIA	UF DF
ENDEREÇO ELETRÔNICO		TELEFONE	
ENTE FEDERATIVO RESPONSÁVEL (EFR) UNIÃO			
SITUAÇÃO CADASTRAL ATIVA		DATA DA SITUAÇÃO CADASTRAL 03/11/2005	
MOTIVO DE SITUAÇÃO CADASTRAL			
SITUAÇÃO ESPECIAL *****		DATA DA SITUAÇÃO ESPECIAL *****	

Fonte: Receita Federal do Brasil (2022)

Entre as informações presentes neste comprovante, pode-se destacar a situação cadastral da empresa, o código CNAE de suas atividades, seu porte, data de abertura, natureza jurídica, se é matriz ou filial, sua localização, entre outras. A seguir, alguns destes atributos serão explicados com mais detalhes.

2.5.1.1 Situação Cadastral

Uma das informações presentes no documento é a situação cadastral em que a empresa se encontra. De acordo com a RFB (2018), existem cinco categorias distintas em que os estabelecimentos podem ser enquadrados, conforme apresentado a seguir.

Quadro 3 – Tipos de situações cadastrais das empresas

Código	Situação Cadastral	Descrição
01	Nula	Para ter sua situação cadastral declarada como nula, uma empresa deve se enquadrar em algum dos casos a seguir: 1) Atribuição de mais de um CNPJ para o mesmo estabelecimento; 2) Alegação de falsidade ou simulação de participação de um sócio; 3) Constatação de ações ilícitas no ato de registro.
02	Ativa	Negócio regularizado e sem pendências.
03	Suspensa	Há dois principais motivos previstos para que um CNPJ tenha sua situação suspensa: 1) Interrupção temporária de suas atividades, declarada pela própria empresa; 2) Descobrimto de qualquer tipo de inconsistência nos seus dados cadastrais, como capital social, CEP, representante legal, entre outros; 3) Indícios de interposição fraudulenta de sócio ou titular.
04	Inapta	Existem três situações em que uma empresa pode ser declarada como inapta: 1) Se não apresentar, em dois anos consecutivos, qualquer declaração e demonstrativo solicitados pela RFB; 2) Se não for localizada; 3) Caso apresente irregularidades em operações no comércio exterior.
08	Baixada	Para ter sua situação cadastral baixada, uma empresa precisa realizar uma solicitação de baixa e esperar que ela seja deferida.

Fonte: Adaptado de Receita Federal do Brasil (2018)

É importante ressaltar, além das características descritas no Quadro 3, que empresas com a situação cadastral avaliadas como nula, suspensa ou inapta podem sofrer alterações nesta classificação. Por exemplo, caso uma empresa inapta apresente as suas declarações e regularize sua situação na RFB, ela voltará a ser classificada como ativa. Por outro lado, se permanecer sem apresentar as declarações devidas por cinco períodos consecutivos, terá sua situação cadastral alterada para baixada (RECEITA FEDERAL DO BRASIL, 2018). Dessa forma, a situação cadastral baixada é a única que não pode sofrer nenhum tipo de alteração.

2.5.1.2 Matriz e Filial

No documento também há uma identificação do estabelecimento como matriz ou filial. Segundo a Receita Federal do Brasil (2014), o estabelecimento matriz é aquele constituído no

momento de abertura da empresa e é considerado como a sua sede legal. Por outro lado, a filial é uma outra unidade do negócio, subordinada à matriz.

Dessa forma, uma empresa pode ter diversas filiais, porém somente uma matriz. Como elas constituem a mesma pessoa jurídica, as filiais e as matrizes possuem os oito primeiros dígitos do CNPJ iguais, diferenciando-se somente nos últimos caracteres. Além disso, de maneira geral, as matrizes possuem a ordem do CNPJ igual a 0001, por terem sido o primeiro estabelecimento constituído no momento de abertura da empresa. No entanto, como há a possibilidade de alteração da matriz, ela pode ter uma ordem diferente.

Por fim, deve-se ressaltar que, caso um estabelecimento filial seja encerrado, a situação cadastral da empresa não será alterada. Assim, uma empresa só é considerada encerrada quando a sua matriz tiver a situação cadastral baixada.

2.5.1.3 Natureza Jurídica

A empresa pode ainda ser classificada de acordo com a sua natureza jurídica. Segundo a Comissão Nacional de Classificação (Concla), os códigos de natureza jurídica têm como objetivo identificar e padronizar a constituição jurídico-institucional das entidades públicas e privadas. Assim, desde 1995, a Concla desenvolve e publica as diretrizes referentes a esta classificação (IBGE; CONCLA, 2021).

Existem 89 categorias distintas de natureza jurídica. No entanto, a maioria das empresas pode ser enquadrada em duas delas: Empresa Individual de Responsabilidade Limitada de Natureza Empresária (EIRELI) e Sociedade Empresária Limitada (RECEITA FEDERAL DO BRASIL, 2021).

A EIRELI é um formato empresarial constituído por apenas um sócio. Nela, é necessário integralizar um capital social de, pelo menos, 100 vezes o salário mínimo vigente, no momento de sua abertura. Além disso, o titular possui responsabilidade limitada com as obrigações da empresa, de forma que seu patrimônio pessoal não poderá ser afetado pelas dívidas contraídas pela pessoa jurídica (IBGE; CONCLA, 2021).

Por outro lado, a Sociedade Empresária Limitada consiste em uma associação que possui normas específicas com base no valor investido por cada associado. Por isso, há a necessidade de que a empresa tenha no mínimo 2 sócios. Este tipo de sociedade ainda permite que a empresa tenha um administrador externo e garante que a responsabilidade dos sócios seja limitada às cotas que possuem da empresa (IBGE; CONCLA, 2021).

2.6 FATORES RELEVANTES PARA O FECHAMENTO DAS MPES

Há inúmeras causas possíveis para o fechamento de uma empresa e elas podem estar relacionadas com fatores internos ou externos ao empreendimento. De maneira geral, os fatores internos são capazes de ser previstos com maior antecedência, enquanto os externos tendem a ser menos previsíveis. Além disso, é muito raro que exista apenas um motivo responsável pelo encerramento. Na maioria das vezes, a sua causa é uma complexa combinação de fatores (ATSAN, 2016).

Justamente por essa complexidade, existe uma extensa literatura sobre o tema, desenvolvida por profissionais de diversas áreas e sob diferentes perspectivas. Assim, alguns autores buscaram agrupar as pesquisas desenvolvidas de acordo com o seu foco e, dessa forma, chegaram a três principais abordagens utilizadas para explicar o fechamento de uma empresa, que são a determinista, a voluntarista e a emotiva (KHELIL, 2015).

A perspectiva determinista sugere que as empresas fracassam devido a fatores externos, sobre os quais os empreendedores possuem pouco ou nenhum controle. Por isso, ela afirma que o ambiente em que a empresa está inserida é o fator que mais impacta em seu sucesso ou fracasso (MELLAHI; WILKINSON, 2004).

Por outro lado, a abordagem voluntarista rejeita a hipótese de que os empreendedores são meros agentes passivos. Segundo essa perspectiva, as ações e decisões tomadas pelos líderes de uma empresa estão diretamente relacionadas com o seu sucesso ou fracasso e, por isso, são mais relevantes que os fatores externos (CARDON et al., 2011).

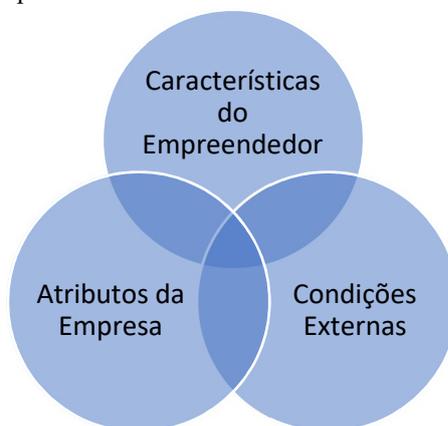
Com o objetivo de realizar uma análise mais robusta, Mellahi e Wilkinson (2004) propuseram que os estudos referentes ao fechamento de empresas sejam baseados na complementaridade das abordagens determinista e voluntarista, de forma a considerar tanto os aspectos externos quanto os internos.

Novas pesquisas têm apontado os fatores psicológicos como críticos para o desempenho do negócio. De acordo com esta perspectiva, conhecida como emotiva, o comprometimento, a ambição e a motivação dos empreendedores desempenham um papel fundamental no resultado dos negócios. Além disso, estas características ajudam a explicar porque algumas pessoas, mesmo diante das situações mais adversas, realizam todos os esforços possíveis para manter sua empresa (KHELIL, 2015). Apesar da relevância destas características, há poucas pesquisas que adotam esta abordagem, principalmente devido à dificuldade de mensurar estes atributos.

Diante disso, os fatores relacionados ao encerramento das micro e pequenas empresas podem ser divididos em três grupos intercomplementares, contendo as características do

empreendedor, os atributos da empresa e as condições externas (CAMPOS, 2021), conforme destacado na Figura 3.

Figura 3 – Grupos de fatores relacionados ao encerramento das MPEs



Fonte: Adaptação de Campos (2021)

Nas seções a seguir, cada um dos fatores que compõem estes grupos será explicado de forma detalhada.

2.6.1 Características do empreendedor

Diversos estudos apontam as características pessoais do empreendedor como fatores relevantes para o desempenho de uma empresa. De acordo com o Sebrae (2016), a capacitação do dono em gestão empresarial e o tipo de ocupação do empresário antes de abrir o negócio (se estava desempregado ou não) são dois fatores que, conjuntamente, estão associados com o fechamento de micro e pequenas empresas no Brasil.

Alves e Lisboa (2014) ressaltam a relevância do papel da capacitação em gestão empresarial. Em sua pesquisa, os autores constataram que as principais dificuldades enfrentadas pelos empresários estão relacionadas com o planejamento financeiro do negócio e com a falta de estratégia. Além deles, Guerra e Teixeira (2010) e Borges e Oliveira (2014) também apontaram fatores de natureza estratégica como os mais importantes para o fechamento de um pequeno negócio

Outro fator frequentemente relacionado ao desempenho das micro e pequenas empresas é a experiência prévia do empreendedor, tanto em termos técnicos, quanto práticos. De acordo com uma pesquisa realizada com MPEs localizadas no Estado de São Paulo, o baixo nível de escolaridade está diretamente relacionado com a falta de competência gerencial e, por isso, é

um fator preponderante para a mortalidade de empresas. O estudo ainda destaca que, em um contexto de elevada complexidade do ambiente de negócios, esta habilidade se torna ainda mais relevante (FERREIRA et al., 2012).

A mesma conclusão pode ser observada em outros estudos. Segundo o Sebrae (2014), o grau de escolaridade do empresário exerce forte influência em seu processo decisório, principalmente em momento críticos do ciclo de vida da empresa. Além disso, ter uma experiência anterior no contexto empresarial lhes possibilita superar as dificuldades de gerenciamento da empresa e lidar com possíveis consequências negativas geradas pelo cenário econômico.

Esta experiência muitas vezes está relacionada com a idade do empreendedor. Em um modelo desenvolvido por Halabí (2010), cujo objetivo era prever o fechamento das empresas, a idade se mostrou como uma variável relevante para o resultado do negócio. A pesquisa ainda concluiu que gestores com idade superior a 35 anos possuem maior probabilidade de garantir que a empresa não feche, possivelmente por causa da experiência adquirida ao longo de suas vidas.

Contudo, estudos sugerem que pode haver uma associação negativa entre o desempenho de uma empresa e o aumento da idade de seu administrador. Santos et al. (2018) destacam que, quanto mais jovem é o gestor de uma empresa, maior a probabilidade de utilizar ferramentas modernas de controle gerenciais, como a curva de custos ABC e o *benchmarking*. Estes instrumentos, por sua vez, foram apontados pelos autores como mecanismos relevantes para o auxílio à sobrevivência das organizações.

É necessário ainda atentar-se ao fato de que pode haver casos de fraude envolvendo o comportamento dos sócios das empresas. Segundo Paiva (2020), uma das fraudes mais frequentes é a sucessão empresarial. Nela, uma empresa nova é constituída para continuar com as atividades de uma inadimplente, visando se desobrigar das dívidas e afastar a má reputação.

2.6.2 Atributos da empresa

Mesmo contribuindo significativamente para os resultados de uma organização, as características pessoais do empreendedor não são capazes de explicar todo o contexto de um negócio. Dessa forma, fatores como a disponibilidade de recursos financeiros, o ponto de localização e o setor de atuação da empresa são frequentemente apontados pelos empresários como relevantes para o fechamento ou sucesso de seu empreendimento (SANTINI et al., 2015).

No contexto das micro e pequenas empresas, alcançar a saúde financeira do negócio, principalmente em seus primeiros anos, é um dos desafios centrais enfrentados pelos empresários. Assim, Hammes e Mallman (2021) defendem a importância do capital social, no momento de abertura da empresa, para conquistar este objetivo. Este capital pode ser definido como os recursos que os sócios investem sem uma garantia de retorno e, segundo as autoras, traz credibilidade junto a possíveis investidores, auxilia no sustento financeiro da operação e possibilita a exclusão dos sócios que não cumprirem com as suas obrigações legais.

Dessa forma, observa-se que relacionamento entre os sócios também impacta no desempenho de uma organização. Em São Paulo, mais de 8% dos donos de micro e pequenas empresas afirmaram que fecharam o seu negócio devido a problemas com os sócios (SEBRAE-SP, 2010). Por outro lado, em uma análise mais quantitativa, Dutra e Previdelli (2005) mostraram que MPEs com um maior número de sócios se mantêm abertas por mais tempo. A maior disponibilidade de recursos financeiros é a principal hipótese para este fenômeno, de acordo com os autores.

A opção por determinado tipo de pessoa jurídica também parece estar relacionada com as taxas de mortalidade das micro e pequenas empresas. Conforme explicado anteriormente, cada categoria de natureza jurídica está sujeita a determinadas regras, que envolvem desde o número de sócios até o capital social. Dessa forma, têm-se demonstrado que os índices de encerramento são diferentes para cada tipo jurídico (REBONATTO et al., 2020). Em Santa Catarina, as Sociedades Limitadas são as que possuem o maior índice, seguidas pelos Empresários Individuais.

Ribeiro e Oliveira (2016) ainda destacam que um fator crítico de sucesso das pequenas empresas é a variedade dos produtos que elas oferecem. Segundo os autores, uma maior diversificação em seu *mix* de produtos pode garantir acesso a novos mercados consumidores e, até mesmo, aumentar as vendas com os clientes atuais. Portanto, esta estratégia é capaz de aumentar o faturamento das empresas e torná-las mais sólidas.

Outro atributo relevante na análise das MPEs é a sua localização. Seja devido à proximidade ao mercado consumidor, aos custos de matéria-prima, energia e mão-de-obra, ou mesmo por conta de outras condições, o local de estabelecimento de uma empresa pode estar diretamente relacionado com o seu desempenho (HOLANDA, 1975). No Rio de Janeiro, Campos (2021) mostrou que há diferenças significativas na probabilidade de fechamento de uma empresa, dependendo do bairro em que ela foi estabelecida.

Apesar da relevância da localização para o resultado dos negócios, os empreendedores possuem pouco conhecimento técnico sobre como realizar esta escolha. Segundo pesquisa do

Sebrae (2014), 37% dos donos de micro e pequenas empresas não sabem como escolher um ponto comercial. Araújo et al. (2017) ainda mostraram que 57% dos empresários participantes de sua pesquisa não fizeram uma análise prévia para selecionar a localização de seu estabelecimento.

Ressalta-se ainda que a probabilidade de fechamento de uma empresa é superior em seus primeiros anos de atividade, principalmente devido à dificuldade de se estabelecer no mercado. Porém, depois de ultrapassar esta barreira inicial, há uma tendência de que consigam sobreviver por mais anos. Campos (2021) chegou a esta conclusão em seu estudo, no qual demonstrou que as MPEs ativas possuem idade média de 12,68 anos, enquanto a das fechadas é de apenas 4,10 anos.

2.6.3 Condições Externas

Ainda que uma empresa tenha um planejamento robusto e consiga desempenhar suas atividades internas de maneira muito eficiente, é impossível desprezar o efeito que o ambiente externo exerce nos seus resultados. Em um estudo exploratório sobre as MPEs brasileiras, observou-se que, entre os anos de 2010 e 2015, houve um aumento na taxa de mortalidade de aproximadamente 344%. Estes números foram então comparados com a inflação do período e descobriu-se que as duas variáveis apresentam uma correlação positiva. Dessa maneira, sugere-se que o aumento na inflação está relacionado com o maior número de empreendimentos que encerraram as suas atividades (FABRES et al., 2015).

Associado a esse cenário, o aumento na taxa básica de juros torna a obtenção de linhas de crédito ainda mais difícil para as micro e pequenas empresas, o que reduz as suas chances de sobrevivência (RORATTO et al., 2017). La Rovere (2002) ainda ressalta que a complexidade no acesso ao crédito faz com que elas usualmente não tenham recursos suficientes para expandir sua capacidade produtiva, tornando-as menos competitivas.

Outro fator frequentemente apontado como importante para o sucesso ou fracasso de um negócio é o seu setor de atuação. De acordo com um levantamento realizado pelo Sebrae (2020), 30,2% das MPEs do setor de comércio fecham em até cinco anos. Por outro lado, esta taxa é de apenas 14,3% para a indústria extrativa. Dessa forma, dependendo do setor de atividade da empresa, a sua probabilidade de fechamento pode apresentar diferenças consideráveis. Marques et al. (2021) apontam que uma das possíveis razões para esta diferença é o tamanho da concorrência existente.

A possibilidade de participar de um programa de incentivo à produtividade também parece exercer um papel relevante no desempenho das empresas, principalmente as de micro e pequeno portes. Segundo Sbaraini e Oliveira (2021), empresas que participaram de um programa de orientação, oferecido pelo Sebrae, tiveram uma redução significativa em suas taxas de mortalidade nos anos subsequentes. Entretanto, os autores ressaltaram que, devido ao caráter pontual do programa, esta tendência deixou de ser observada com o passar dos anos.

Diante do exposto, pode-se observar que os fatores externos são, direta ou indiretamente, influenciados por ações tomadas pelo governo. No Brasil, uma das principais políticas públicas desenvolvidas no contexto empresarial foi o Imposto Simples Nacional. Criado em 2007 com o objetivo de estimular a formalização e a sobrevivência das MPes, ele prevê um tratamento tributário simplificado, com o pagamento de oito impostos por meio de uma única guia. Além disso, quando comparado com outros regimes de impostos, apresenta uma alíquota mais baixa para a maioria das atividades empresariais (OECD, 2020).

Por causa dessas vantagens, pode-se supor que empresas que adotam este regime tributário possuem menor probabilidade de fechar. No entanto, não há um consenso na literatura sobre esta afirmação. Enquanto Campos (2021) mostrou que MPes optantes pelo Simples Nacional apresentaram menor taxa de mortalidade no estado do Rio de Janeiro, Castro (2010) ressaltou um possível efeito negativo na sua implementação. Segundo o autor, a tributação progressiva de acordo com o faturamento pode desestimular o crescimento das empresas.

2.7 PREVISÃO DE ENCERRAMENTO DAS MPES

A área de pesquisa relacionada à predição de encerramento de pequenos negócios tem ganhado força nos últimos anos, principalmente no âmbito internacional (POMPE; BILDERBEEK, 2005). No entanto, o interesse sobre o tema não é recente. Em uma pesquisa desenvolvida no início dos anos 1970, Edmister (1972) aplicou ferramentas estatísticas com o objetivo de prever, dentro de um grupo de pequenas empresas, quais delas iriam à falência. A partir deste estudo, o autor concluiu que é possível prever o encerramento com até cinco anos de antecedência, considerando certa margem de erro.

Desde então, pesquisas realizadas com o intuito de prever o fechamento das empresas evoluíram e passaram a adotar novas técnicas e abordagens, valendo-se do aumento do poder computacional e da facilidade em acessar e manipular grandes volumes de dados. Apesar disso, os modelos estatísticos convencionais se mantêm relevantes e, muitas vezes, são a base para a aplicação de novas técnicas, como redes neurais, aprendizagem de máquina, até mesmo,

aprendizagem profunda (do inglês, *deep learning*) (STEVENSON et al., 2021 & BALCAEN, OOGHE, 2004).

De maneira geral, estes modelos se enquadram em um problema de classificação binomial, cujo objetivo é prever se a empresa encerrará ou não as atividades a partir de diferentes atributos (LI et al., 2010). Entre os fatores mais utilizados, destacam-se os indicadores financeiros das empresas, que podem ser divididos em quatro categorias principais: índices de liquidez, lucratividade, alavancagem e atividade (HORTA; CAMANHO, 2011).

Estudos recentes mostram que a utilização de variáveis de outras naturezas pode melhorar significativamente o desempenho das previsões. Em uma pesquisa com pequenas empresas, Horta e Camanho (2011) desenvolveram dois modelos distintos: um com variáveis financeiras e outro com a adição de indicadores estratégicos. Essa alteração foi responsável por aumentar em cerca de 10% a acurácia do modelo.

Além da seleção das variáveis, o desenvolvimento de um modelo de previsão de encerramento das micro e pequenas empresas envolve diferentes etapas, desde o processamento e tratamento dos dados, até a escolha das métricas adequadas para avaliar o seu desempenho.

2.7.1 Pré-processamento dos dados

Segundo Batista (2003), a fase de pré-processamento se inicia a partir do momento em que os dados são coletados e organizados. Seus principais objetivos são identificar e tratar os dados corrompidos, faltantes ou irrelevantes para o modelo. O autor ainda destaca que este é um processo semiautomático, pois depende de uma pessoa com domínio sobre o tema para identificar os problemas e escolher os métodos mais adequados para solucioná-los.

Carvalho et al. (2011) ressaltam que, dessa forma, o pré-processamento é capaz de minimizar ou, até mesmo, eliminar os problemas relacionados aos dados, tornando-os mais adequados para a aplicação de um modelo. Ainda de acordo com os autores, esta fase pode ser subdividida em sete grupos de tarefas, que devem ser executadas na ordem avaliada como mais adequada diante do contexto do problema.

- *Integração dos dados* – esta atividade visa reunir os dados oriundos de diferentes fontes em um mesmo local, de forma a facilitar a sua manipulação. Além disso, busca-se identificar atributos em comum nas tabelas, de forma a correlacioná-las e, assim, possibilitar a realização de análises mais robustas;

- *Eliminação manual dos atributos* – consiste em remover os atributos que não possuem nenhuma relação com o problema de trabalho e, por isso, não serão utilizados no modelo;
- *Redução da dimensionalidade* – a partir desta atividade, busca-se reduzir os atributos, dado que alguns algoritmos possuem dificuldade em lidar com um número elevado deles. Para isso, existem dois tipos de abordagem: agregação, que combina os atributos ao aplicar funções lineares ou não-lineares, e seleção, que identifica e mantém aqueles que são mais relevantes para o problema;
- *Limpeza dos dados* – esta atividade envolve a detecção e correção de problemas existentes no conjunto de dados, evitando que se propaguem no desenvolvimento do modelo. Dados inconsistentes, ruidosos, redundantes e incompletos são os problemas mais comumente encontrados;
- *Transformação dos dados* – alguns algoritmos exigem que todas as suas variáveis sejam de um determinado tipo. Outros possuem seu desempenho excessivamente influenciado pela variação dos atributos numéricos. Sendo assim, o objetivo desta atividade é ajustar os dados de acordo com os requisitos do modelo selecionado. Para isso, são aplicadas diferentes técnicas, tais como: a conversão simbólico-numérico, em que uma variável categórica é transformada em numérica, a transformação logarítmica, que torna a escala de todos os dados mais próxima, e a tradução, em que um valor é convertido por outro mais facilmente manipulável;
- *Amostragem dos dados* – devido ao tempo computacional ou à saturação de memória, alguns algoritmos podem apresentar dificuldade para lidar com um número elevado de objetos. Dessa forma, uma possível solução é a utilização de um subconjunto dos dados originais, método conhecido como amostragem. Dentre as técnicas mais utilizadas, destaca-se: a amostragem aleatória simples, que seleciona aleatoriamente um grupo, a estratificada, que mantém a mesma proporção de objetos para cada classe, e a progressiva, que aumenta o tamanho da amostra enquanto a acurácia do modelo apresentar melhoras;
- *Balanceamento dos dados* – muitos algoritmos são incapazes de lidar com um conjunto de dados desbalanceados. Ou seja, quando uma das classes possui um número muito maior de registros do que a outra e, com isso, tende a ser favorecida. Sendo assim, esta atividade visa balancear artificialmente o conjunto

de dados. Existem diferentes técnicas que podem ser utilizadas com este fim e elas serão explicadas na seção a seguir.

2.7.1.1 Técnica de Balanceamento dos Dados

Em diferentes aplicações reais, o objetivo de um modelo é identificar eventos anormais, também chamados de interessantes, em uma população contendo um elevado número de eventos comuns. Apesar disso, por assumirem que as distribuições são equilibradas, grande parte deles apresenta dificuldade em lidar com dados desbalanceados (BRAGA; CASTRO, 2011).

Diante disso, diferentes técnicas foram desenvolvidas com o objetivo de mitigar este efeito. Uma das mais utilizadas é a subamostragem, que envolve a eliminação de exemplos da classe majoritária. Entretanto, seu principal problema é a perda de informação causada pela eliminação destes registros (BRAGA; CASTRO, 2011).

Assim, a fim de evitá-lo, Chawla et al. (2002) desenvolveram uma técnica de sobreamostragem conhecida como SMOTE (*Synthetic Minority Oversampling Technique*). Nesse método, o equilíbrio entre as classes se dá por meio da criação de registros sintéticos. Dessa forma, para cada exemplo pertencente à classe minoritária, novos exemplos artificiais são criados entre os segmentos de reta que o ligam aos seus vizinhos mais próximos.

A aplicação do método SMOTE e a criação de exemplos artificiais permite que o modelo construa uma região de decisão mais ampla e menos específica. Assim, as características da classe minoritária passam a ter maior relevância e, por consequência, a capacidade de generalização do modelo aumenta. Em testes desenvolvidos com diferentes bases de dados, foram obtidos resultados satisfatórios com a aplicação desta metodologia (CHAWLA et al., 2012).

2.7.2 Engenharia de Atributos

Outra atividade essencial para o desenvolvimento do modelo é a engenharia de atributos (do inglês, *feature engineering*). Segundo Nargesian et al. (2017), esta técnica pode ser definida como a prática de criação de novas variáveis, a partir das existentes, com o objetivo de melhorar a sua performance.

De maneira geral, este processo envolve a aplicação de funções de transformação – principalmente as aritméticas e de agregação – nas variáveis presentes na base de dados

(NARGESIAN et al., 2017). Destaca-se ainda que estas manipulações são necessárias para extrair o máximo dos atributos, dado que muitas de suas informações estão implícitas e precisam desse tratamento para serem reveladas (GRUS, 2015).

Assim, a etapa de engenharia de atributos é, majoritariamente, manual e requer o conhecimento do conjunto de dados e das especificidades do algoritmo adotado (HEATON, 2016). Além disso, é um processo iterativo de tentativa e erro, no qual avalia-se ciclicamente o desempenho do modelo após a criação das novas variáveis (NARGESIAN et al., 2017).

Deve-se destacar que a aplicação desta técnica possibilita uma melhoria significativa no desempenho do modelo. Nascimento et al. (2020) exemplificam este cenário em seu estudo, cujo objetivo era prever as vendas de micro e pequenas empresas. Ao aplicar a engenharia de atributos, os autores criaram dez novas variáveis e, com isso, obtiveram um acréscimo expressivo na acurácia do modelo.

2.7.3 Modelos de Classificação

A seleção do modelo que será aplicado é mais uma etapa essencial para o processo de desenvolvimento. Dentre os diferentes tipos existentes, os modelos de classificação figuram entre as aplicações mais conhecidas de aprendizado de máquina. De maneira simplificada, seu objetivo é identificar a classe a que determinado exemplo não rotulado pertence, com base em suas características (LIU, 2011).

Para isso, estes modelos recorrem a um tipo de aprendizado conhecido como supervisionado. Nele, para cada exemplo fornecido ao algoritmo, deve-se informar os seus atributos (variáveis explicativas) e o rótulo de sua classe (variável resposta). Dessa forma, o modelo consegue identificar uma relação entre estas variáveis e, com isso, é capaz de determinar a classe de um novo registro sem rótulo (LUDERMIR, 2021).

Existem diferentes algoritmos que podem ser aplicados com este fim. Entre os mais populares, pode-se destacar: *Naive-Bayes*, *Árvore de Decisão*, *Random Forest*, *Support Vector Machine*, *K-Nearest-Neighbors* e *Regressão Logística*. No entanto, não é possível afirmar que um deles apresenta, consistentemente, resultados melhores que os outros. A opção por um determinado modelo deve-se basear nas características dos dados disponíveis e no contexto do problema em que ele será aplicado (SARKER et al., 2019).

Com relação à predição de encerramento das empresas, pode-se encontrar na literatura um elevado número de técnicas e metodologias que, desde o início da década de 1970, vêm sendo utilizadas pelos pesquisadores. Em uma revisão sistemática destes estudos, Shi e Li

(2019) mostraram que, entre as técnicas mais aplicadas, figura o algoritmo de regressão logística.

Diante deste contexto, destaca-se o trabalho de Ohlson (1980), que foi o primeiro a utilizar a técnica com o objetivo de prever quais empresas iriam à falência. Outro estudo importante foi o de Wood (2006), que aplicou o modelo de regressão logística no segmento das pequenas empresas. Mais recentemente, Zizi et al. (2020) obtiveram um excelente resultado ao aplicá-lo em pequenos negócios. Com 2 anos de antecedência, o modelo foi capaz prever corretamente a situação de 84,4% das empresas analisadas.

Duas importantes características que ainda ajudam a explicar a relevância da regressão logística neste contexto são a sua capacidade em lidar com dados categóricos e o fato de sua natureza ser não-paramétrica (ZIZI et al., 2020). Por estes motivos, o algoritmo será explicado com mais detalhes na seção seguinte.

2.7.3.1 Regressão Logística

A regressão logística é um dos métodos de modelagem estatística mais utilizado no desenvolvimento de análises discriminatórias. Por relacionar um conjunto de variáveis independentes, com uma variável dependente, ela pode ser enquadrada na classe de métodos estatísticos multivariados de dependência. Além disso, quando mais de uma variável explicativa é utilizada, classifica-se a regressão como múltipla (PAULA, 2013).

O uso deste modelo é recomendado nas situações em que a variável resposta é de natureza dicotômica, isto é, problemas que podem ser divididos em duas categorias. O objetivo da regressão logística é estimar a probabilidade associada à ocorrência destes eventos, a partir das informações contidas nas variáveis explicativas.

Em um cenário cuja variável dependente Y assuma os valores de 0 ou 1 e que possua n variáveis independentes (X_1, X_2, \dots, X_n), a probabilidade de ocorrência do evento pode ser descrita da seguinte forma (MINUSSI et al., 2002):

$$P(Y = 1) = \frac{1}{1 + e^{-g(x)}} \quad (1)$$

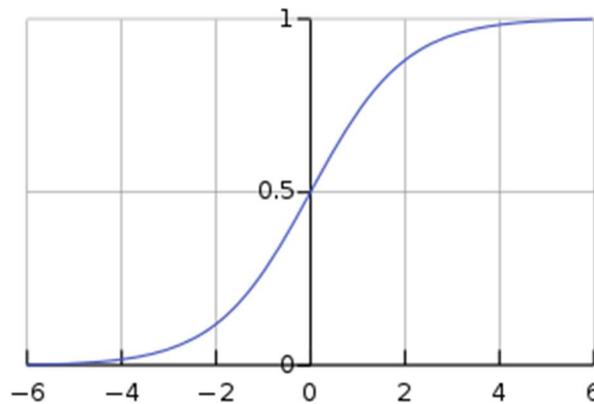
Onde,

$$g(x) = B_0 + B_1 * X_1 + B_2 * X_2 + \dots + B_n * X_n \quad (2)$$

Estes coeficientes B_0, B_1, \dots, B_n são estimados com base no conjunto de dados, com o método de verossimilhança, no qual obtém-se a combinação dos coeficientes que maximiza a

probabilidade de que a amostra tenha sido observada (HOSMER; LEMESHOW, 1989). Assim, a partir destes resultados, é possível obter a função sigmoide da curva logística. Ela apresenta um formato de S e, com isso, confere ao modelo um alto grau de generalidade, o que pode ser observado na Figura 4.

Figura 4 – Curva Logística



Fonte: Jannedy et al. (2003)

Deve-se ressaltar ainda que este modelo assume que o logaritmo da razão entre as probabilidades de ocorrência e não ocorrência do evento é linear, de forma que sua equação pode ser descrita da seguinte maneira:

$$\ln\left(\frac{P(Y = 1)}{P(Y = 0)}\right) = B_0 + B_1 * X_1 + B_2 * X_2 + \dots + B_n * X_n \quad (3)$$

Além disso, de acordo com a sua aplicação, a regressão logística apresenta algumas particularidades. Em um problema de discriminação de dois grupos, também conhecido como classificação, as seguintes regras devem ser seguidas:

- Se a probabilidade de ocorrência do evento for maior do que o 0,5, considera-se que ele ocorreu. Ou seja, Y será classificado como 1;
- Caso contrário, classifica-se Y como 0.

2.7.4 Avaliação do desempenho do modelo

A avaliação do desempenho do modelo é uma etapa imprescindível para garantir o sucesso da pesquisa. Na literatura, encontra-se uma extensa variedade de metodologias que

podem ser utilizadas com este fim, o que torna importante compreender quais delas se adequam melhor ao contexto do estudo e às características dos dados disponíveis.

Tradicionalmente, a métrica mais utilizada na avaliação e na seleção dos modelos de classificação é a acurácia, que mensura o percentual de acerto do algoritmo. Sua popularidade justifica-se pela facilidade em interpretá-la e pelo fato de que o objetivo padrão do aprendizado supervisionado é minimizar o erro global. Entretanto, para problemas com classes altamente desbalanceadas, a acurácia pode fornecer uma informação incompleta e, assim, não deve ser analisada isoladamente (CASTRO; BRAGA, 2011).

Isso pode ser observado em um exemplo no qual apenas 2% dos registros fazem parte de uma classe, enquanto o restante é de outra. Ao simplesmente classificar todos como pertencentes à classe majoritária, o modelo obtém 98% de acurácia, porém, caso o objetivo seja identificar os exemplos raros, ele não terá nenhuma utilidade (CHAWLA et al., 2002; CASTRO; BRAGA, 2011).

Dessa forma, faz-se necessário buscar métricas que consigam avaliar o desempenho do classificador para cada grupo. Assim, Fawcett (2006) propôs a adoção da matriz de confusão. Nela, os elementos da diagonal principal representam o número de exemplos que foram corretamente classificados, os verdadeiros negativos (VN) e verdadeiros positivos (VP). Por outro lado, aqueles que estão fora desta diagonal representam os erros cometidos pelo modelo e são os falsos positivos (FP) e falsos negativos (FN). Esta relação pode ser observada a seguir, no Quadro 4.

Quadro 4 – Matriz de Confusão

Matriz de Confusão	Predição ($y = 0$)	Predição ($y = 1$)
Real ($y = 0$)	Verdadeiros Negativos	Falso Positivo
Real ($y = 1$)	Falso Negativo	Verdadeiros Positivos

Fonte: Adaptado de Fawcett (2006)

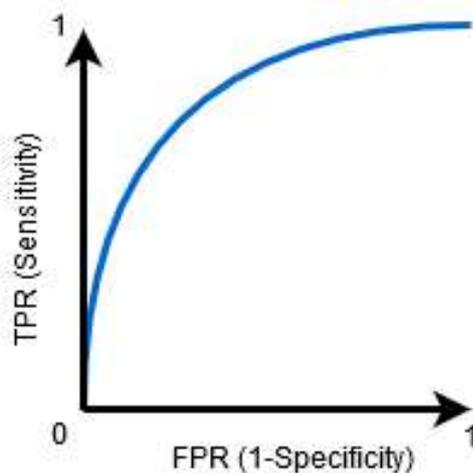
Entretanto, apesar de ser eficaz na avaliação de classificadores com dados desbalanceados, a matriz de confusão não permite uma comparação simples e direta entre dois modelos. Assim, visando superar esta limitação, grande parte das pesquisas utilizam a curva *Receiver Operating Characteristic* (ROC) e a *Area Under the ROC Curve* (AUC) como medidas da capacidade de discriminação do classificador (CASTRO; BRAGA, 2011).

A curva ROC traz uma comparação entre a taxa de verdadeiros positivos e a de falsos positivos. O primeiro deles avalia, do total de registros positivos, quantos foram classificados corretamente. Já o segundo mensura o número de casos negativos que foram identificados como

positivos, dividido pelo total de registros da classe negativa. Graficamente, esta curva é gerada a partir da variação de um limiar de decisão θ . Quanto menor o valor de θ , mais exemplos são classificados como positivos (PRATI et al., 2008).

Já a área sob a curva ROC (AUC) (HANLEY; MCNEIL, 1982) representa uma medida geral da capacidade de discriminação do classificador. Por ser independente do valor selecionado para o limiar de decisão θ , esta métrica é amplamente aceita como critério de comparação entre dois modelos. Ela pode assumir valores entre 0 e 1 e, quanto mais próxima de 1, melhor é o seu desempenho. Na Figura 5, pode-se observar a representação gráfica desta medida.

Figura 5 – Gráfico da Área Sob a Curva ROC



Fonte: Bhandari (2020)

De maneira geral, as pesquisas de predição de encerramento das empresas se enquadram em um problema com classes desbalanceadas, o que ocorre porque o número de empresas ativas costuma ser significativamente superior ao de encerradas. Assim, as métricas apresentadas são comumente aplicadas neste contexto. Horta et al. (2013), Zizi et al. (2020) e Stevenson et al. (2021) são alguns dos autores que utilizaram a acurácia, combinada com a AUC, para avaliar o desempenho de seus classificadores.

3 DESENVOLVIMENTO DO MODELO

O presente capítulo visa descrever todo o processo de desenvolvimento do modelo de predição de encerramento das micro e pequenas empresas da indústria de transformação. Ele será dividido de acordo com as suas diferentes etapas e abordará desde a coleta dos dados até a avaliação de seu desempenho.

3.1 COLETA DOS DADOS

O processo de coleta dos dados é necessário para o desenvolvimento de um modelo. O conjunto de informações obtidas durante esta etapa direcionarão todas as atividades seguintes. Conforme destacou Batista (2003), a qualidade do conhecimento gerado é amplamente influenciada pela qualidade dos dados de entrada.

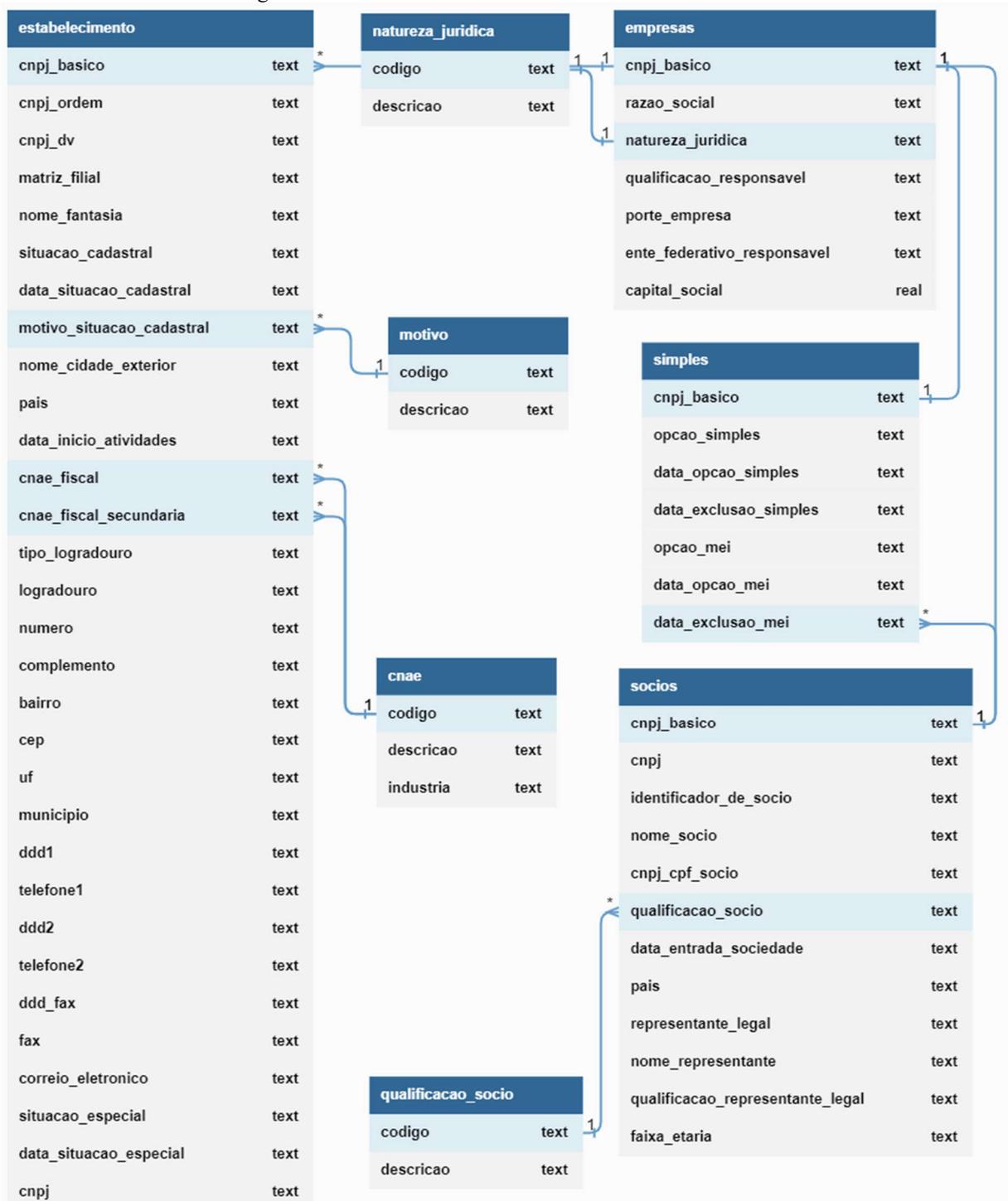
Dessa forma, para que fosse possível criar um modelo preditivo de encerramento das MPEITs, foram definidos alguns critérios mínimos que o conjunto de dados precisaria atender. Eles foram baseados em outros trabalhos desenvolvidos sobre o tema, como o de Zizi et al. (2020) e Campos (2021), e adaptados para o contexto do presente estudo:

- Conter informações no âmbito da empresa;
- Possibilitar a seleção de MPEITs;
- Ter registros de empresas ativas e encerradas, com as respectivas datas de início e fim das atividades.

A partir disso, foram realizadas buscas na literatura para encontrar a base que melhor se adequasse a estes critérios. Assim, chegou-se aos dados públicos do Cadastro Nacional de Pessoa Jurídica (CNPJ), disponibilizados em arquivos no formato *Comma Separated Files* (CSV) (RECEITA FEDERAL DO BRASIL, 2022). Neles, estão reunidas todas as informações contidas no documento de Comprovante de Inscrição e de Situação Cadastral, para cada estabelecimento formal registrado no sistema público. No presente estudo, utilizou-se a versão atualizada no dia 15/05/2022.

Ao consolidar todos os arquivos em que estas informações estão presentes, foram geradas oito tabelas com características complementares das empresas. Na Figura 6 estão dispostos os campos disponíveis em cada uma destas tabelas, bem como o relacionamento entre elas.

Figura 6 – Tabelas do Cadastro Nacional de Pessoa Jurídica



Fonte: Adaptado de Receita Federal do Brasil (2022)

De acordo com este esquema, observa-se que o conjunto de dados é composto por tabelas com diferentes propósitos e níveis de complexidade. Aquelas que estão dispostas no centro da imagem, – “qualificacao_socio”, “cnae”, “motivo” e “natureza_juridica” –, contêm

informações complementares das tabelas principais. A tabela “natureza_juridica”, por exemplo, traz o significado dos códigos presentes em “empresas”.

Por outro lado, as tabelas principais reúnem diferentes informações sobre os empreendimentos cadastrados no sistema da RFB e podem ser correlacionadas pelo campo “cnpj_basico”. Dentre elas, a tabela “empresas” informa o seu porte, natureza jurídica, capital social, entre outras características. Já a tabela “estabelecimento” contém dados dos estabelecimentos que compõem uma empresa, como a sua localização, código CNAE das atividades, situação cadastral, etc. As últimas duas, “simples” e “sócios”, possuem, respectivamente, informações sobre a opção pelo Sistema Simples Nacional e sobre os seus sócios.

Além disso, é importante destacar que esta base de dados possibilita que seja selecionado somente o público-alvo da pesquisa: as micro e pequenas empresas da indústria de transformação. Para tanto, foram escolhidas as empresas com porte igual a micro ou pequeno, e com o código de classificação CNAE pertencente a algum dos subsetores da indústria de transformação (RECEITA FEDERAL DO BRASIL, 2022; CONCLA, 2021).

3.2 PREMISSAS ADOTADAS

A fim de garantir que os resultados obtidos no presente estudo sejam transparentes e compreensíveis, serão explicadas as premissas adotadas no desenvolvimento deste modelo. Esta é uma importante etapa no alinhamento dos conceitos utilizados e servirá como referência para a elaboração das atividades seguintes.

Em uma revisão das principais pesquisas realizadas sobre predição de encerramento das empresas (*business failure prediction*), Balcaen e Ooghe (2004) observaram que não há um consenso sobre o marco que define o seu encerramento. Os autores ainda destacaram que, muitas vezes, este conceito depende das leis vigentes no país.

Sendo assim, o presente estudo utilizou como base a situação cadastral informada pela Receita Federal do Brasil (2018) para definição deste critério. Com isso, as empresas com a sua situação igual a “ativa” foram classificadas como “em atividade” e aquelas com a situação “baixada” compuseram o grupo de encerradas, o que pode ser observado na Figura 7.

Figura 7 – Análise da situação cadastral das empresas



Fonte: Autor (2022)

Sendo assim, a taxa de mortalidade das MPEITs foi calculada dividindo-se o número de empresas baixadas pelo total de empresas, ativas ou baixadas, em um determinado período. Vale ressaltar ainda que as empresas com a situação cadastral nula, suspensa ou inapta ainda podem sofrer mudanças nesta classificação. Dessa forma, por haver a possibilidade de que sua situação seja alterada, tanto para ativa, quanto para baixada, elas não serão consideradas neste estudo. Critério semelhante foi adotado por Sbairini et al. (2021) e visa garantir que empresas ativas não sejam erroneamente apontadas como encerradas, e vice-versa.

Ainda com relação à definição de encerramento, deve-se destacar que uma empresa pode ser composta por diferentes estabelecimentos filiais. Com isso, caso um deles seja encerrado, a situação cadastral da empresa não será alterada. Uma empresa somente será considerada encerrada quando a matriz, sua sede legal, for baixada (RECEITA FEDERAL DO BRASIL, 2014).

Por este motivo, outra premissa adotada no estudo foi a de avaliar as características de uma empresa de acordo com as de seu estabelecimento matriz. Dessa forma, considerou-se que uma empresa iniciou as atividades na data em que a sua matriz foi estabelecida. Esta regra também foi aplicada para os diferentes atributos presentes na tabela “estabelecimentos”.

Dessa forma, fez-se necessário selecionar apenas as empresas que não tiveram uma mudança de matriz, o que pode ser verificado por meio do campo “cnpj_ordem” e é fundamental para garantir que as características analisadas se referem ao primeiro estabelecimento da empresa. Com este critério, identificou-se que menos de 0,02% das empresas realizaram uma alteração de matriz.

Além disso, conforme destacado pelo Sebrae (2020), os primeiros cinco anos após a abertura de uma MPE são os mais críticos. Por ainda buscarem se estabelecer no mercado e enfrentarem dificuldades financeiras, muitas empresas não conseguem suportar este cenário e

acabam encerrando as suas atividades. Diante disso, o modelo proposto adotará este período de cinco anos como padrão para realizar as suas predições.

Aliado a isso, para que o modelo de predição possa ser aplicado no momento de abertura da empresa, foram avaliadas as características que ela possuía naquela época. Por exemplo, só foram contabilizados os sócios que estavam na empresa desde o princípio, ou seja, cuja data de entrada na sociedade fosse igual à data de início das atividades.

Devido ao método de atualização da tabela adotado pela RFB (2018), no entanto, não foi possível garantir este critério para toda a base. Em alguns dos campos, a tabela sobrescreve o registro antigo pelo novo, de forma que não há um histórico das atualizações. Assim, diante da indisponibilidade deste histórico, o valor presente na base de dados será considerado para o desenvolvimento do modelo. Esta premissa foi adotada para os seguintes campos: capital social, CNAEs secundários, porte da empresa e natureza jurídica. Em sua pesquisa, Campos (2021) trabalhou com a mesma base de dados e seguiu o critério de utilizar as informações disponíveis no momento da pesquisa.

Ademais, os Microempreendedores Individuais (MEI) não serão considerados no presente estudo. Apesar da relevância destes profissionais, eles apresentam características muito diferentes das outras micro e pequenas empresas. Por exemplo, não é permitido que suas empresas tenham sócio, nem que ele seja sócio de outras. Por este motivo, não há nenhuma informação sobre os MEIs na tabela de “sócios” da base da pesquisa. Além disso, eles são automaticamente enquadrados no Simples Nacional (SEBRAE, 2022). Como estes atributos serão relevantes para o desenvolvimento do modelo, optou-se por desconsiderar os microempreendedores individuais.

Deve-se ressaltar ainda que o modelo será construído com base nos dados das MPEITs fundadas em um mesmo ano. O objetivo ao considerar este critério é evitar que empresas submetidas a diferentes condições externas sejam avaliadas da mesma forma e compreender o impacto que outras variáveis possuem neste cenário. A título de exemplo, mesmo que a inflação esteja elevada no início das atividades de uma empresa, todos os outros empreendimentos avaliados estarão sujeitos a este mesmo fator. Logo, caso um deles feche e outro permaneça aberto, deve haver outras características que expliquem esta diferença (FABRES et al. 2015).

No Brasil, observou-se um aumento no índice de mortalidade das MPEs a partir de 2015. Segundo o Sebrae (2020), a crise econômica foi uma das principais responsáveis por isso. O estudo ainda destaca que, com a chegada da pandemia da Covid-19, este cenário tornou-se mais incerto e difícil de avaliar. Portanto, visando reduzir o impacto destas externalidades no

resultado do modelo, optou-se por utilizar somente empresas abertas durante o ano de 2011 ao longo de sua construção.

É importante destacar que o presente estudo utilizará o modelo de regressão logística para prever o encerramento das MPEITs. Segundo Shi et al. (2019), esta é uma das técnicas mais utilizadas nas pesquisas sobre o tema, devido à sua eficácia e simplicidade. Zizi et al., (2020), ainda ressaltam que a regressão logística não exige que os dados numéricos apresentem uma distribuição normal e possibilita a inclusão de variáveis categóricas. Dessa modo, avaliou-se que este algoritmo é o que melhor se adequa ao presente contexto.

3.3 TRATAMENTO INICIAL DOS DADOS

A etapa de pré-processamento dos dados é essencial no desenvolvimento de um modelo. Entre os seus principais benefícios pode-se citar a adequação da base aos requisitos dos algoritmos e a redução da complexidade computacional. Para isso, sua aplicação envolve um conjunto de diferentes tarefas, desde a disponibilização dos dados em um mesmo local, até a redução da dimensionalidade deles (CARVALHO et al., 2011).

Ao todo, esta fase é composta por sete grupos de atividades. Entretanto, conforme ressaltam Carvalho et al. (2011), não há uma ordem correta para executá-las, de maneira que cada contexto deve ser analisado individualmente. Sendo assim, no presente estudo, optou-se pela seguinte divisão: realização de um tratamento inicial nos dados; na sequência, criação e seleção das variáveis; finalmente, a execução das atividades restantes de pré-processamento. Destaca-se ainda que estas etapas foram executadas por meio de códigos desenvolvidos nas linguagens *Python* e *SQL*, que podem ser encontrados no Apêndice A.

Diante disso, iniciou-se o tratamento dos dados por meio de sua integração. Conforme explicado na seção anterior, este estudo utilizou como base os dados públicos de CNPJs, fornecidos pela RFB por meio de 37 arquivos no formato CSV. Dessa forma, a atividade consistiu em disponibilizá-los em um local que proporcionasse a manipulação e consulta adequadas, inclusive com a possibilidade de correlacionar campos de diferentes tabelas.

Identificou-se, assim, um repositório no software *GitHub* no qual este banco de dados já havia sido criado (RICTOM, 2022). De maneira simplificada, o código foi desenvolvido em *Python* e agrupou os arquivos CSV que continham informações sobre uma mesma tabela e as disponibilizou em um banco de dados local, por meio da biblioteca *SQLite*. Após executá-lo, foi possível acessar o banco diretamente da máquina e correlacionar as diferentes tabelas com o uso da linguagem *SQL*.

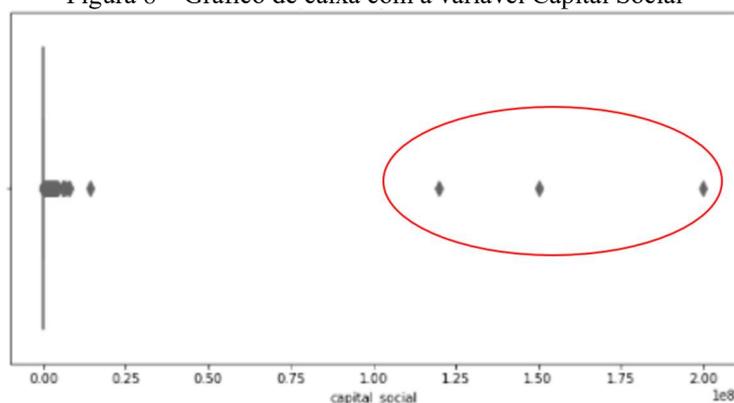
A partir da consolidação destes dados, pôde-se iniciar o processo de limpeza da base, cujo principal objetivo é identificar e corrigir os problemas nela existentes (CARVALHO et al., 2011). Inicialmente, realizou-se uma verificação dos valores nulos em cada uma das tabelas. Com isso, foram identificados campos que não possuíam todas as informações preenchidas. Muitas empresas não informam o número de telefone, o nome fantasia e o tipo de logradouro do endereço no seu cadastro. No entanto, por não serem dados relevantes no contexto do estudo, estas colunas puderam ser descartadas sem prejuízos para o resultado final.

Em sequência, as oito tabelas disponíveis foram correlacionadas de forma a reunir, em uma única base, todas as informações das empresas e de seus sócios. Com isso, verificou-se se havia dados faltantes em alguma delas. Ao unir a tabela de empresas com a de estabelecimentos, por meio do campo “cnpj_base”, foi possível garantir que cada estabelecimento presente no banco de dados possuía uma empresa associada a ele, condição necessária para obter as informações de porte e de capital social. Este mesmo processo foi realizado com as outras tabelas e nenhum problema foi identificado.

Ainda com relação à atividade de limpeza dos dados, buscou-se avaliar a consistência dos atributos presentes no modelo. Dessa maneira, observou-se que, na coluna de capital social, alguns dos registros se distanciavam significativamente dos demais, o que indica a possibilidade de serem *outliers*. Segundo Carvalho (2011), a presença de casos extremos pode fazer com que o modelo se atenha às suas especificidades e, por isso, obtenha um pior desempenho.

Assim, a fim de analisar este cenário, aplicou-se o gráfico de caixa (do inglês, *boxplot*), uma técnica qualitativa que fornece uma visualização da dispersão dos dados e possibilita a identificação de pontos extremos (SCHWERTMAN, 2004). A partir dele, foram removidos três registros cujo capital social era superior a 100 milhões de reais e, portanto, muito distantes dos demais. Além disso, foram identificadas duas empresas com capital social menor que R\$ 500,00, que também foram retiradas. O resultado obtido com a aplicação da técnica do gráfico de caixa pode ser observado na Figura 8.

Figura 8 – Gráfico de caixa com a variável Capital Social



Fonte: Autor (2022)

Por fim, realizou-se a eliminação manual dos atributos, atividade que visa remover aqueles que não possuem nenhuma relação com o problema analisado. Como o conjunto de dados utilizado na pesquisa não foi gerado com o intuito de prever o encerramento de MPEITs, grande parte de suas informações não se aplicam neste contexto e, portanto, puderam ser retiradas.

No total, mais de 20 colunas foram excluídas nesta etapa, entre elas o nome fantasia, o complemento do endereço e outras informações secundárias de identificação das empresas e dos sócios. Deve-se ressaltar que esta é apenas uma eliminação inicial dos campos e, posteriormente, na etapa de seleção das variáveis, será realizada uma nova verificação dos remanescentes.

Após a realização destas atividades, foi possível reduzir o tamanho do conjunto de dados da pesquisa e torná-lo mais assertivo para o desenvolvimento da solução do problema. Ao final, a base processada passou a conter dados de 13.892 micro e pequenas empresas da indústria da transformação, abertas em 2011. Destas, 2.385 encerraram as atividades após cinco anos, o que representa uma taxa de mortalidade de 17,2%.

3.4 SELEÇÃO E CRIAÇÃO DE ATRIBUTOS

Nesta seção, serão apresentadas as variáveis escolhidas para compor o modelo de previsão de encerramento das MPEITs. Estas características foram selecionadas a partir de diferentes pesquisas científicas que as apontaram como relevantes para o desempenho das empresas. Entre estes estudos, pode-se destacar aqueles desenvolvidos pelo Sebrae (2013, 2014, 2016), que é a instituição referência no contexto dos pequenos negócios brasileiros. Em adição,

a fim de aprimorar o modelo, foram criadas novas variáveis, em um processo conhecido como engenharia de atributos.

Deve-se ressaltar que algumas das características citadas na literatura precisaram ser adaptadas ao contexto da pesquisa, de forma que fosse possível mensurá-las a partir do conjunto de dados. Como exemplo, pode-se citar que o número de CNAEs secundários foi a métrica adotada para avaliar a variedade dos produtos oferecidos pelas empresas, fator apontado como crítico para o seu sucesso (RIBEIRO et al., 2016).

Além disso, seguindo a metodologia definida por Mellahi e Wilkinson (2004), considerou-se a complementariedade entre as abordagens determinista e voluntarista na seleção destas variáveis. Dessa forma, ambos os fatores, internos e externos, são considerados relevantes para o encerramento das empresas. No entanto, conforme explicado anteriormente, buscou-se minimizar o efeito das externalidades ao selecionar somente empresas abertas no ano de 2011.

Mesmo assim, os fatores externos presentes no conjunto de dados serão avaliados na construção do modelo. Dessa maneira, os atributos foram divididos em três grupos complementares: características do empreendedor, atributos da empresa e condições externas (CAMPOS, 2021) e serão apresentados a seguir.

Quadro 5 – Variáveis selecionadas para o modelo de predição de encerramento das MPEITs

Variável	Grupo	Tema	Significado
Situação Cadastral Avaliada	–	–	Situação cadastral da empresa avaliada 5 anos após sua abertura
Capital Social	Atributo da Empresa	Capital Social	Capital social declarado pela empresa
Número de CNAEs Secundários	Atributo da Empresa	Diversificação da Empresa	Número de atividades secundárias declaradas pela empresa
Região da Empresa	Atributo da Empresa	Localização	Região do Brasil em que a matriz da empresa está localizada
Número de Estabelecimentos Ativos no CEP	Atributo da Empresa	Localização	Número de estabelecimentos ativos, no mesmo CEP, no momento de abertura da empresa
Número de Estabelecimentos Encerrados no CEP	Atributo da Empresa	Localização	Número de estabelecimentos encerrados, no mesmo CEP, no momento de abertura da empresa

População do Estado	Atributo da Empresa	Localização	População total do Estado em que a matriz da empresa está localizada
Número de Sócios	Atributo da Empresa	Número de sócios	Número de sócios que a empresa possuía no momento de abertura
Porte da Empresa	Atributo da Empresa	Porte da Empresa	Porte da empresa conforme classificação da RFB
Qualificação Sócio mais Velho	Característica do Empreendedor	Gestão da Empresa	Qualificação do sócio mais velho presente na abertura da empresa, conforme classificação adotada pela RFB
Qualificação Sócio mais Novo	Característica do Empreendedor	Gestão da Empresa	Qualificação do sócio mais novo presente na abertura, conforme classificação adotada pela RFB
Faixa Etária Sócio mais Velho	Característica do Empreendedor	Idade do Sócio	Faixa etária do sócio mais velho presente na abertura da empresa
Faixa Etária Sócio mais Novo	Característica do Empreendedor	Idade do Sócio	Faixa etária do sócio mais novo presente na abertura da empresa
Número de Empresas Ativas dos Sócios	Característica do Empreendedor	Experiência sócio	Número de empresas ativas que os sócios possuíam no momento de abertura da empresa analisada
Número de Empresas Encerradas dos Sócios	Característica do Empreendedor	Experiência sócio	Número de empresas encerradas que os sócios possuíam no momento de abertura da empresa analisada
Opção Avaliada Simples	Fator Externo	Imposto Simples Nacional	Avaliação se a empresa era optante pelo Imposto Simples Nacional no momento de sua abertura
Natureza Jurídica	Fator Externo	Natureza Jurídica	Natureza Jurídica da empresa, de acordo com classificação da Concla

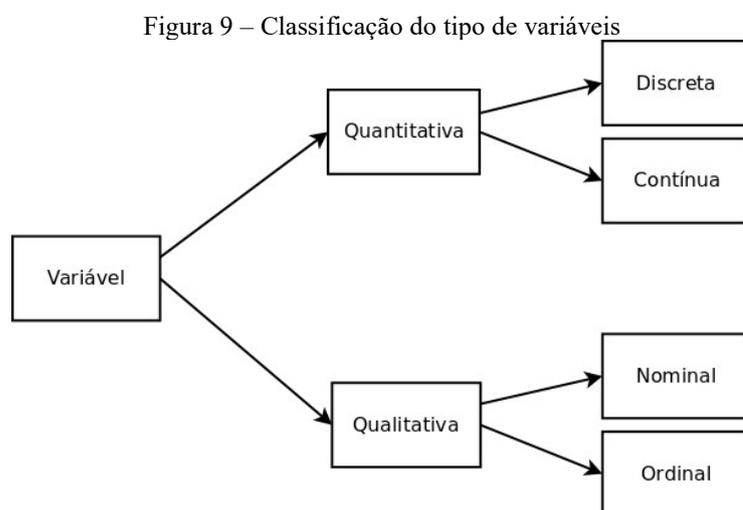
Fonte: Autor (2022)

Cabe ressaltar que cada uma das variáveis apresentadas no Quadro 5 foi analisada individualmente, por meio da comparação entre os grupos de empresas ativas e encerradas. Este processo foi essencial para seleção das características adequadas para o desenvolvimento do

modelo de predição de encerramento das MPEITs e para as análises serão apresentadas no Capítulo 4.

3.5 TRATAMENTO FINAL DOS DADOS

Após a criação e seleção das variáveis que irão compor o modelo, fez-se necessário realizar os últimos ajustes nos dados, de forma a torná-los adequados para a aplicação do algoritmo de regressão logística. Muitos destes ajustes estão diretamente relacionados com o tipo destes dados, que podem ser classificados como numéricos (quantitativos) ou categóricos (qualitativos), conforme destaca a Figura 9.



Fonte: Mayer (2022)

As variáveis numéricas, quando assumem apenas valores inteiros, são consideradas discretas. O número de sócios das empresas é um exemplo de atributo deste tipo. Por outro lado, quando podem assumir qualquer valor no intervalo dos números reais, são descritas como contínuas. Já as variáveis categóricas podem ser subdivididas entre nominais, se não possuírem uma ordem natural, e ordinais, caso possam ser ordenadas, como é o caso da faixa etária dos sócios.

O algoritmo de regressão logística possibilita a utilização de variáveis qualitativas em seu desenvolvimento. No entanto, para que possam ser aplicadas, é necessário transformá-las em dados numéricos. Sendo assim, aplicou-se a conversão simbólico-numérica nas variáveis nominais, utilizando a técnica de codificação distribuída (CARVALHO et al., 2011). Nela, cada

valor distinto assumido por um atributo é convertido em uma coluna, que terá como entrada apenas 0 ou 1, indicando, respectivamente, a ausência ou presença deste valor.

Assim, a variável de opção pelo Simples Nacional será convertida em outras duas colunas, contendo a informação se a empresa optou ou não por este sistema. Dessa forma, caso a MPEIT tenha aderido ao Simples, o valor da primeira coluna será igual a 1, enquanto a segunda terá o valor 0. Além desta, as variáveis de qualificação dos sócios mais velhos e mais novos, da natureza jurídica da empresa e da região em que está localizada passaram pela conversão simbólico-numérica.

Os atributos numéricos também precisaram passar por transformações. Ao analisá-los, observou-se que a ordem de grandeza de seus valores diferia significativamente. Enquanto a população média dos estados em que as empresas estão localizadas é de cerca de 18 milhões, o número de sócios por empresa é de 1,67. Diante disso, para evitar que um atributo predomine sobre o outro e para permitir a comparação entre eles, optou-se por realizar a normalização logarítmica destas variáveis (CARVALHO et al., 2011).

Assim, aplicou-se o logaritmo natural em todas as variáveis numéricas, discretas ou contínuas, presentes no conjunto de dados. Ao todo, 8 atributos foram normalizados: o capital social, o número de sócios, o número de estabelecimentos, ativos ou encerrados, no mesmo CEP, o número de empresas, ativas ou encerradas, que os sócios possuíam, a população do estado e o número de CNAEs secundários.

Em seguida, realizou-se uma amostragem estratificada destes registros por meio da seleção de um subconjunto dos dados originais, mantendo a mesma proporção de empresas ativas e encerradas (PAULA, 2013). Esta atividade foi realizada com o objetivo melhorar a eficiência computacional do modelo e deve-se ressaltar que os resultados alcançados com a aplicação desta técnica foram semelhantes aos obtidos a partir da base completa.

Por fim, realizou-se o balanceamento dos dados, a última atividade da fase de pré-processamento. Conforme citado anteriormente, após cinco anos, há mais empresas que permanecem ativas do que encerradas. No presente estudo, cerca de 82,8% das empresas se mantiveram abertas neste período. Entretanto, o modelo de regressão logística assume que as distribuições das classes são equilibradas, de forma que a majoritária é favorecida quando não há nenhum tratamento (BRAGA; CASTRO, 2011).

Diante deste cenário, aplicou-se a técnica de sobreamostragem SMOTE, desenvolvida por Chawla et al. (2002). Nesta abordagem, o equilíbrio entre as classes ocorre por meio da criação de novos registros sintéticos para o grupo minoritário, de maneira que ambos passam a ter o mesmo número de exemplos. Com isso, destaca-se que há um aumento na capacidade de

generalização do modelo. No presente estudo, utilizou-se a biblioteca SMOTE, do *Python*, para aplicação desta ferramenta.

3.6 APLICAÇÃO E AVALIAÇÃO DO DESEMPENHO DO MODELO

Após coletar os dados, criar e selecionar as variáveis e realizar todos os tratamentos necessários na base, foi possível aplicar o modelo de regressão logística para predição do encerramento das micro e pequenas empresas da indústria de transformação. Para tanto, o primeiro passo foi realizar a divisão dos atributos das empresas em dois tipos: de resposta ou explicativa.

A variável de resposta é aquela que terá a probabilidade de ocorrência estimada pelo modelo. No presente estudo, esta variável é a situação cadastral da empresa, avaliada cinco anos após a sua abertura. Se o empreendimento tiver encerrado as suas atividades durante este período, ela assumirá o valor de 0. Caso contrário, terá o valor 1. Por outro lado, as variáveis explicativas são independentes e reúnem as características das empresas que serão utilizadas como insumo para predição.

Em seguida, os registros foram divididos, de forma aleatória, em conjuntos distintos de treino e de teste, contendo 70% e 30% da base, respectivamente. Esta etapa é necessária para garantir que o modelo seja validado com dados diferentes daqueles que foi treinado (HORTA; CAMANHO, 2013). Deve-se destacar ainda que a técnica de balanceamento dos dados, SMOTE, foi aplicada somente na construção do modelo, ou seja, no conjunto de dados de treino.

Além disso, visando melhorar a sua capacidade de generalização e evitar o problema de *overfitting*, em que os dados se ajustam muito bem ao conjunto de teste, porém não ao de treino, aplicou-se uma técnica de regularização do modelo (REINEKING; SCHRÖDER, 2006). Com esse método, buscou-se otimizar o parâmetro do termo de erro, conhecido como C, ao testar diferentes valores. Para isso, o desempenho do modelo foi avaliado com C assumindo valores de uma sequência exponencial ($C = 2^{-5}, 2^{-3}, \dots, 2^{13}$) e escolheu-se aquele que resultou na melhor performance. Este processo, conhecido como validação cruzada, foi aplicado por Horta e Camanho (2013) no desenvolvimento de um modelo preditivo.

A partir destas definições, foi possível aplicar o modelo de regressão logística por meio de um código desenvolvido em *Python* e disponível no Apêndice B. Para isso, foram utilizadas as bibliotecas *Scikit-learn* e *Scipy.Stats*, que calculam a probabilidade de ocorrência do evento,

de acordo com o conjunto de dados e os parâmetros informados. Caso esta probabilidade seja superior a 0,5, a situação cadastral da empresa será classificada como 1.

Dessa forma, iniciou-se um processo cíclico de aplicação do modelo com diferentes combinações de parâmetros, seguida por uma avaliação de seu desempenho. Esta avaliação foi feita com base na acurácia e na área sob a curva ROC, métricas que são tipicamente utilizadas para mensurar o resultado de modelos de regressão logística (ZIZI et al., 2020).

Vale ressaltar ainda que buscou-se obter um equilíbrio entre estas métricas. De maneira que, caso um modelo apresente uma boa acurácia, porém tenha um desempenho ruim nas outras duas, ele não será avaliado positivamente. Assim, após a realização de diversos testes, foram selecionados os parâmetros descritos no Quadro 6.

Quadro 6 – Parâmetros para aplicação do modelo de regressão logística

Parâmetro	Valor
Tamanho da Amostra	5000
Termo de Erro (C)	0,15
Tamanho do Conjunto de Treino	3500 (70%)
Tamanho do Conjunto de Teste	1500 (30%)
Remoção de Outliers	Sim

Fonte: Autor (2022)

A fim de tornar a avaliação ainda mais robusta, realizou-se um teste para mensurar o desempenho do modelo a partir das MPEITs abertas ao longo do ano de 2012. Considerando que a sua construção se deu com base nos dados de empreendimentos fundados em 2011, esta validação foi essencial para medir a performance e a capacidade de generalização do modelo. Todos estes resultados serão apresentados no capítulo seguinte.

4 RESULTADOS

Neste capítulo, serão apresentados os resultados obtidos a partir do desenvolvimento do modelo de regressão logística para predição de encerramento das MPEITs. Para tanto, ele será dividido em duas seções: análise dos atributos das empresas e avaliação das predições.

Na primeira seção, serão explicadas as análises feitas sobre cada uma das variáveis que compõem o modelo. No segundo momento, serão descritos os resultados da predição do modelo, a partir das métricas definidas previamente.

4.1 ANÁLISE DOS ATRIBUTOS DA EMPRESA

Conforme destacado anteriormente, as variáveis escolhidas para compor o modelo de predição passaram por um processo prévio de análise. Nele, os fatores apontados na literatura como relevantes para o encerramento das MPEITs foram testados, com base nos dados da pesquisa. Para as variáveis categóricas, calculou-se a taxa de mortalidade das empresas em cada um de seus grupos. Já para as quantitativas, avaliou-se a média que a variável assumia nos grupos de empresas ativas e encerradas e comparou-se esta diferença.

Seguindo os mesmos conceitos definidos durante o desenvolvimento do modelo, as variáveis foram avaliadas a partir de micro e pequenas empresas da indústria de transformação abertas em 2011 divididas em três grupos: características do empreendedor, atributos da empresa e condições externas.

4.1.1 Características do Empreendedor

A fim de avaliar a experiência prévia do sócio no ramo empresarial, foram criadas as variáveis com o número de empresas ativas e encerradas que os sócios possuíam no momento de abertura do empreendimento. Apesar do fator experiência ser considerado como positivo na maioria das pesquisas, ter várias empresas encerradas pode ser um indício de falta de capacidade de gestão ou, até mesmo, de fraude (DIAS et al., 2014; PAIVA, 2020; SEBRAE, 2016).

Ao analisar estes atributos a partir da base da pesquisa, identificou-se que mais da metade dos sócios não possuía nenhuma outra empresa, ativa ou encerrada, no momento de abertura do empreendimento avaliado. Cerca de 29,4% deles tinham ao menos uma empresa

ativa, enquanto 22,9% possuíam no mínimo uma encerrada. A seguir, será calculado o número médio de empresas anteriores destes sócios.

Tabela 1 – Comparação do número de empresas ativas e encerradas por sócio

Situação Cadastral da MPEIT avaliada	Média de Empresas Ativas Constituídas Anteriormente	Média de Empresas Encerradas Constituídas Anteriormente
Encerrada	1,97	1,80
Ativa	1,88	1,64
Diferença Percentual	4,55%	10,21%

Fonte: Autor (2022)

De acordo com a Tabela 1, pode-se observar que, em média, os sócios que encerraram as MPEITs avaliadas no estudo possuíam mais empresas, tanto ativas quanto fechadas, no momento de abertura desta MPEIT. Vale destacar que os números apresentados consideram somente os empreendedores que já possuíam uma ou mais empresas abertas ou encerradas anteriormente.

Em sequência, foi analisado o impacto da idade dos sócios neste cenário. Enquanto alguns autores defendem que o aumento da idade traz mais experiência para o gerenciamento do negócio, outros afirmam que gestores mais novos são mais abertos à inovação e a oportunidades (HALABÍ, 2010; SANTOS et al., 2018). Na base de dados da pesquisa, esta informação é disponibilizada com a faixa etária de cada um dos sócios, dividida em intervalos de dez anos. No entanto, optou-se por agrupar todas as faixas etárias em duas: pessoas com idade igual ou menor a 50 anos e com mais de 50.

Dessa maneira, foram avaliadas as variáveis de faixa etária do sócio mais novo e do mais velho da empresa. Caso o empreendimento tenha apenas um sócio, a idade dele foi informada em ambas. Diante disso, foi possível calcular a taxa de mortalidade das MPEITs para estas variáveis, conforme a Tabela 2.

Tabela 2 – Taxa de mortalidade de acordo com a faixa etária dos sócios

Faixa Etária	Taxa de Mortalidade Empresa - Sócio Mais Velho	Taxa de Mortalidade Empresa - Sócio Mais Novo
Menor ou igual a 50 anos	14,61%	18,20%
Maior que 50 anos	20,03%	13,90%

Fonte: Autor (2022)

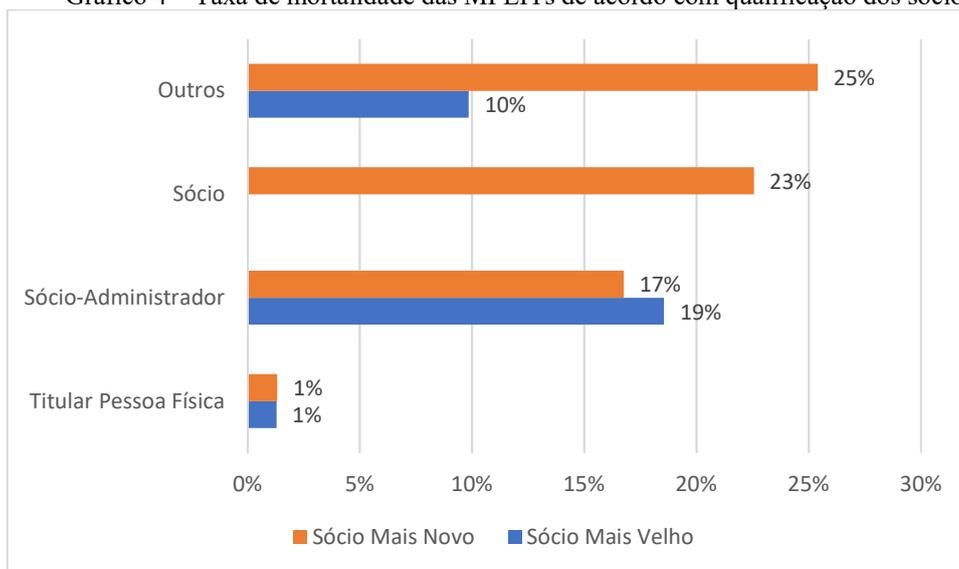
Assim, observa-se que há uma diferença entre as duas categorias. Ao analisar somente o grupo dos sócios mais velhos, descritos na primeira coluna, a taxa de encerramento é maior para as empresas cujo sócio possui mais de 50 anos. Porém, a tendência é contrária para os

sócios mais novos, de forma que mais empresas encerraram suas atividades quando o seu sócio mais novo tinha menos de 50 anos.

Analisou-se, ainda, a qualificação dos sócios mais velhos e mais novos, conforme classificação da Receita Federal do Brasil (2018). Esta qualificação é informada no ato de constituição da empresa e deverá ser escolhida de acordo com o papel e a responsabilidade que serão assumidos pelo sócio. A partir destas variáveis, buscou-se mensurar, mesmo que indiretamente, um fator crítico para o sucesso das MPEITs: a gestão empresarial realizada pelos seus sócios (GUERRA; TEIXEIRA, 2010).

Para tanto, foram comparadas as taxas de mortalidade das empresas, de acordo com a qualificação de seu sócio mais novo e do mais velho. A fim de evitar o excesso de informações, as qualificações presentes em menos de 2% dos registros foram agrupadas em uma mesma categoria, chamada “Outros” e estes resultados estão dispostos a seguir.

Gráfico 4 – Taxa de mortalidade das MPEITs de acordo com qualificação dos sócios



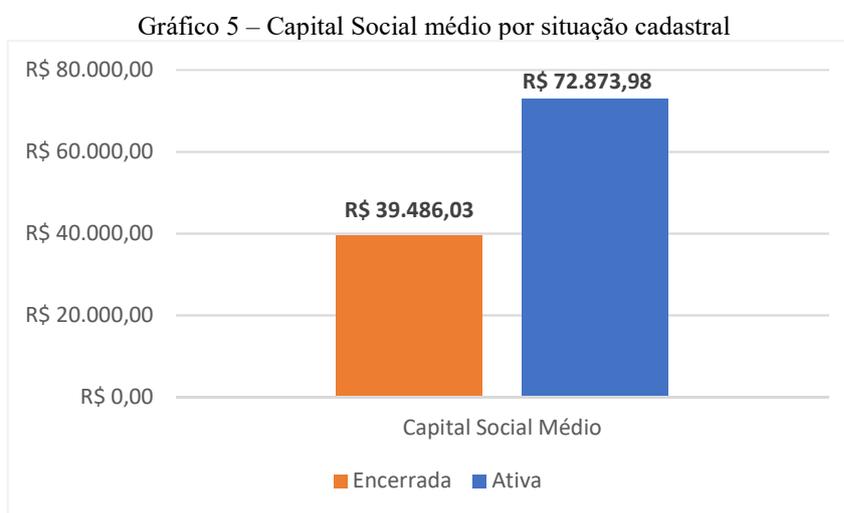
Fonte: Autor (2022)

No Gráfico 4, pode-se observar uma tendência interessante. Não há nenhum sócio mais velho cuja qualificação seja apenas “Sócio”, todos possuem alguma função de administração. Por outro lado, 3.310 sócios mais novos têm esta qualificação e o índice de encerramento de suas empresas é mais elevado. Isso revela que, quando o sócio mais novo não possui um cargo administrativo, a taxa de mortalidade da empresa é mais alta, ao passo que, quando possui, é mais baixa.

4.1.2 Atributos da Empresa

Além dos fatores relativos aos empreendedores, sabe-se que as características das empresas desempenham um papel importante nos seus resultados. No contexto das micro e pequenas, o recurso financeiro é um dos atributos mais relevantes. Principalmente nos anos iniciais, quando a empresa ainda busca se estabelecer no mercado e, muitas vezes, não é rentável (SANTINI et al., 2015).

Observa-se, então, a relevância do capital social. Por serem os recursos investidos pelos sócios sem garantia de retorno, ele traz mais segurança para a operação e auxilia no seu sustento (HAMMES; MALLMAN, 2021). No Gráfico 5, pode-se observar o seu valor médio de capital social nos grupos de empresas ativas e encerradas.



Fonte: Autor (2022)

A partir deste gráfico, nota-se que há uma diferença significativa entre o capital social das MPEITs ativas e encerradas. Em média, o capital das ativas foi cerca de 84,6% superior. Porém, deve-se ressaltar que esta variável apresenta uma variância elevada, ou seja, os dados estão bem dispersos. Dessa forma, 50% das empresas analisadas possuem um capital social de até R\$ 30.000,00.

Em seguida, buscou-se avaliar o efeito da diversificação das atividades desempenhadas pelas MPEITs. Este é um fator crítico para a captação e manutenção de clientes e, consequentemente, para o sucesso da empresa (RIBEIRO et al., 2016). No presente estudo, este atributo foi mensurado de forma indireta, ao contabilizar o número de CNAEs secundários declarados pela empresa, e os resultados obtidos estão dispostos na Tabela 3.

Tabela 3 – Número médio de CNAEs secundários, por situação cadastral

Situação Cadastral Avaliada	Média de CNAEs Secundários
Encerrada	1,99
Ativa	2,99
Diferença Percentual	50,12%

Fonte: Autor (2022)

As informações revelam que empresas que permanecem abertas após cinco anos de atividade possuem, em média, um número de CNAEs secundários 50,12% superior às encerradas, o que corrobora os estudos sobre o tema. Além disso, é interessante notar que cerca de 22,0% das empresas possuem apenas um CNAE secundário, enquanto outras 18,3% têm dois.

O porte da empresa é mais um fator que está relacionado com o seu encerramento. Devido à menor disponibilidade de recursos, as pequenas empresas possuem mais dificuldades em sobreviver nos anos iniciais. Portanto, dentro do grupo de MPEs, a taxa de mortalidade das microempresas costuma ser superior à das pequenas (SEBRAE, 2020).

Tabela 4 – Taxa de mortalidade de acordo com o porte da empresa

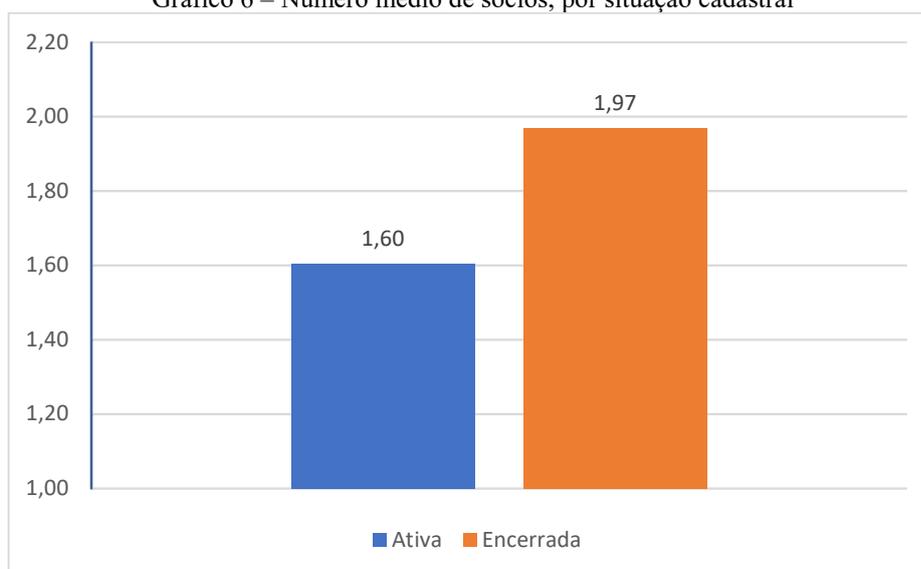
Porte da Empresa	Número de Empresas	Taxa de Mortalidade
Microempresa	11.413	18,16%
Pequena Empresa	2.479	12,59%

Fonte: Autor (2022)

Sendo assim, esta taxa foi calculada para as MPEITs analisadas, conforme a Tabela 4, e os resultados ocorreram conforme esperado, de forma que o índice de encerramento nos primeiros cinco anos foi superior nas empresas de menor porte. Todavia, também é interessante destacar que o número de microempresas é quase cinco vezes superior ao de empresas de pequeno porte.

Outra característica avaliada foi o número de sócios. Na literatura, diferentes pesquisas abordam este tema. Por um lado, problemas entre os sócios são um dos motivos que mais levam ao encerramento de MPEs (SEBRAE-SP, 2010). Por outro, com o aumento no número de sócios, pode-se ter uma disponibilidade maior de recursos financeiros (DUTRA; PREVIDELLI, 2015). No Gráfico 6, calculou-se o número médio de sócios por empresa, de acordo com a sua situação cadastral.

Gráfico 6 – Número médio de sócios, por situação cadastral

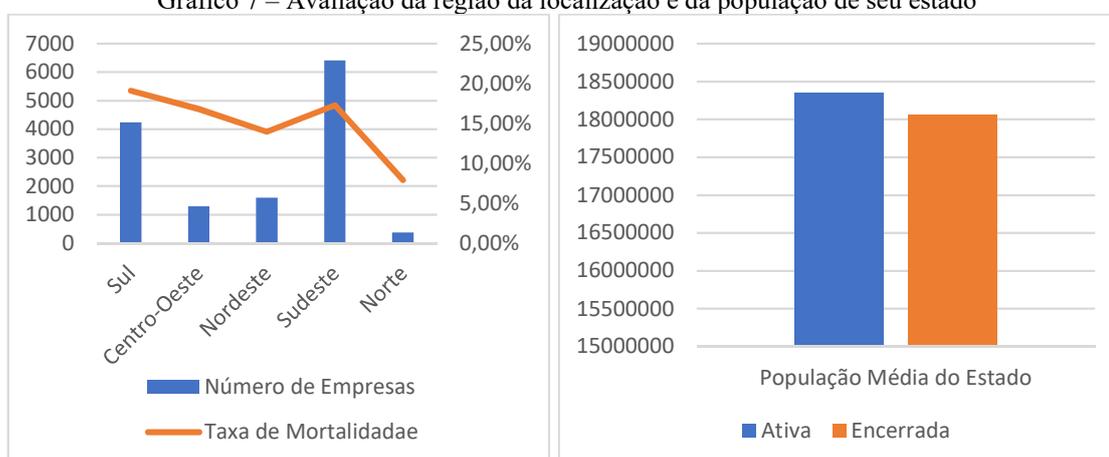


Fonte: Autor (2022)

Diante disso, os dados do presente estudo revelam que, em média, as empresas ativas possuíam menos sócios do que as encerradas, no seu momento de abertura. É importante destacar ainda que o grupo das MPEITs com apenas um ou dois sócios correspondem à cerca de 95,8% dos empreendimentos avaliados.

Em sequência, analisou-se a localização das MPEITs. A importância desta característica está relacionada a diferentes fatores, como: o tamanho do mercado consumidor local, os custos para a operação e a concorrência (HOLANDA, 1975). Para avaliá-los, foram utilizadas duas variáveis encontradas na literatura e criadas outras duas. Com as primeiras, explorou-se os impactos da região da empresa e do tamanho da população de seu estado (IBGE, 2011), conforme apresentado no Gráfico 7.

Gráfico 7 – Avaliação da região da localização e da população de seu estado

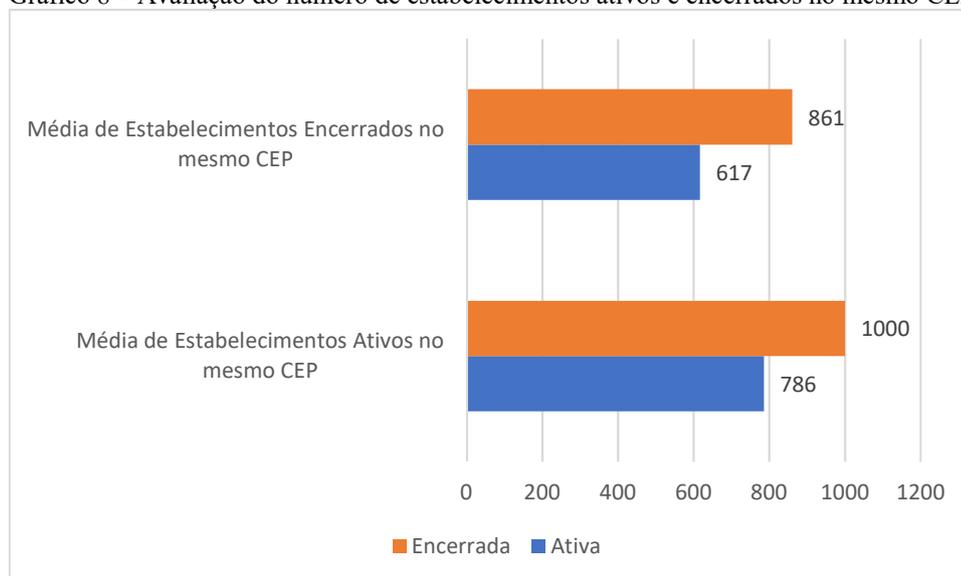


Fonte: Autor (2022)

Os resultados mostram que a região com a maior taxa de mortalidade das empresas foi a Sul, enquanto a Norte apresentou a menor. Apesar disso, é importante destacar o número de empresas em cada uma delas. Enquanto a região Sul é a segunda com mais empreendimentos, na Norte foram abertas apenas 431 MPEITs em 2011. Vale destacar ainda que a população da cidade aparenta ter um pequeno impacto no encerramento das empresas. Em média, as empresas que permaneceram ativas estão localizadas em cidades com maior população.

Além disso, a fim de avaliar de maneira mais específica o efeito do ponto de localização das empresas, foram analisados o número de estabelecimentos, ativos ou encerrados, que havia no mesmo CEP, no momento de abertura da MPEIT.

Gráfico 8 – Avaliação do número de estabelecimentos ativos e encerrados no mesmo CEP



Fonte: Autor (2022)

Diante do Gráfico 8, é possível observar um comportamento interessante. As MPEITs que se mantiveram ativas após cinco anos foram abertas em locais que havia menos estabelecimentos, encerrados ou ativos, no início de suas atividades. Portanto, ao considerar estas duas métricas, pode-se inferir que lugares menos explorados por outras empresas se mostram mais favoráveis para abrir uma micro e pequena empresa da indústria de transformação.

4.1.3 Fatores Externos

Com relação às externalidades, foram avaliados dois fatores, que são a natureza jurídica da empresa e a sua opção pelo Imposto Simples Nacional. Apesar de cada empreendimento ter a liberdade de escolhê-los, estes atributos foram considerados como externos, pois suas regras são definidas por terceiros e precisam ser seguidas. Dessa forma, mesmo existindo diferentes tipos de natureza jurídica, as empresas devem se enquadrar em alguma delas e seguir as suas diretrizes e obrigações (GOVERNO FEDERAL, 2020).

A seguir, pode-se observar as naturezas jurídicas das MPEITs avaliadas no estudo, ressaltando que aquelas com menos de 1% dos registros foram agrupadas na categoria “Outras”.

Tabela 5 – Avaliação do impacto da natureza jurídica

Natureza Jurídica	Número de Empresas	Taxa de Mortalidade
Sociedade Empresária Limitada	12720	18,58%
Outras	80	8,75%
Empresa Individual de Responsabilidade Limitada de Natureza Empresária (EIRELI)	1092	1,37%

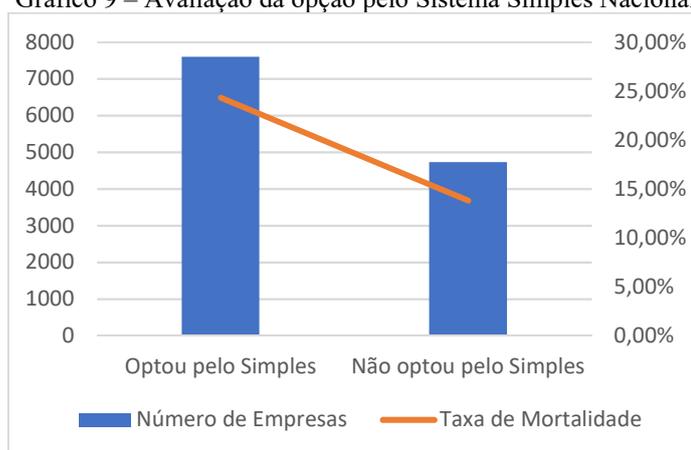
Fonte: Autor (2022)

Conforme pode-se observar na Tabela 5, cerca de 91,6% das MPEITs avaliadas se enquadram como Sociedade Empresária Limitada. No entanto, ao compará-las com aquelas de natureza jurídica EIRELI, identificou-se que a sua taxa de mortalidade foi muito superior. Para entender se este foi um resultado atípico, ou uma tendência, foram avaliadas as MPEITs fundadas nos anos seguintes.

Com isso, descobriu-se que, apesar desta ter sido a maior diferença de taxa de mortalidade entre os dois grupos, as EIRELI abertas nos cinco anos seguintes também tiveram um índice de encerramento significativamente menor. Uma possível explicação para este resultado é que, por ter apenas um sócio, as decisões nas EIRELI podem ser tomadas de forma mais eficiente e objetiva.

No que se refere ao Imposto Simples Nacional, foi verificado se a empresa era optante pelo Sistema no momento em que iniciou as suas atividades. Com isso, buscou-se avaliar se esta escolha resulta em algum impacto, positivo ou negativo, no desempenho das MPEITs, o que foi checado por meio da taxa de encerramento e está disponível no Gráfico 9.

Gráfico 9 – Avaliação da opção pelo Sistema Simples Nacional



Fonte: Autor (2022)

Os resultados mostram que 7.602 das MPEITs analisadas optaram pelo Sistema Simples Nacional em sua abertura, o que representa cerca de 61,6% do total. Apesar disso, a taxa de mortalidade deste grupo foi de 24,34%, superior à das empresas que não fizeram esta opção. Um dos motivos para isso pode ser a tributação progressiva deste sistema que, segundo Castro (2010), é capaz de desestimular o crescimento do faturamento das empresas e, por isso, gerar um efeito negativo.

4.2 AVALIAÇÃO DAS PREDIÇÕES

Em seguida, serão apresentados os resultados da predição de encerramento das MPEITs, obtidos ao aplicar o modelo de regressão logística desenvolvido. Conforme destacado no capítulo anterior, esta avaliação baseou-se nas métricas de acurácia e da área sob a curva ROC, que são amplamente utilizadas em trabalhos desta natureza. Enquanto a primeira avalia o percentual de acerto do modelo, a segunda visa mensurar a sua capacidade em identificar ambos os grupos de empresas ativas e encerradas, por meio da comparação entre as taxas de verdadeiros e falsos positivos.

Além disso, visando alcançar resultados mais robustos, testou-se o modelo em duas bases distintas: nas MPEITs abertas em 2011 que foram reservadas para este fim, e naquelas fundadas durante o ano de 2012. Dessa maneira, foi possível avaliar o desempenho do modelo em um conjunto de empresas com características diferentes daquelas que ele foi construído e, assim, analisar a sua capacidade de generalização. A Tabela 6 apresenta os resultados da aplicação do modelo na primeira base.

Tabela 6 – Resultados alcançados com o primeiro teste do modelo

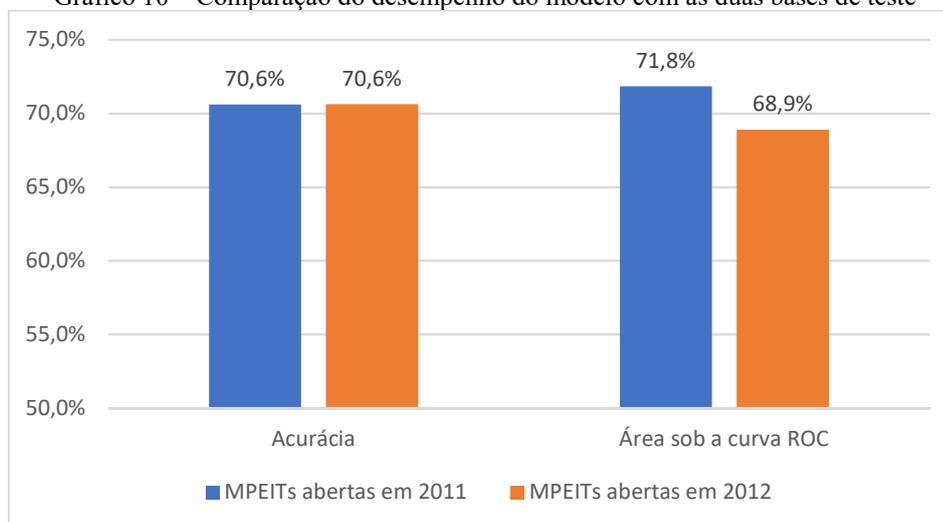
Métrica	Resultado
Empresas Previstas Corretamente	1.059
Acurácia	70,60%
Área sob a curva ROC	71,84%

Fonte: Autor (2022)

Dessa forma, observa-se que, de um total de 1.500 MPEITs, o modelo foi capaz de prever corretamente a situação cadastral de 1.059 delas, em um período de cinco anos após o início de suas atividades. Isso representa uma acurácia de 70,6% e revela um desempenho satisfatório nas predições. Além disso, a área sob a curva ROC de 71,8% mostra a capacidade de discriminação do modelo em conseguir identificar tanto empresas ativas quanto encerradas.

A fim de avaliar a capacidade em manter esta performance com um conjunto de dados diferente, testou-se o modelo com as MPEITs abertas no ano de 2012. No total, este grupo é composto por 12.339 empresas e os resultados da predição foram comparados com aqueles obtidos na etapa anterior.

Gráfico 10 – Comparação do desempenho do modelo com as duas bases de teste



Fonte: Autor (2022)

A partir do Gráfico 10, observa-se que o modelo conseguiu manter um nível de desempenho muito similar quando aplicado no segundo grupo de MPEITs. Este resultado destacou a sua capacidade de generalização, uma vez que a aplicação ocorreu em um conjunto de dados com características diferentes daquelas em que foi treinado. Tendo em vista que um dos principais objetivos de modelos desta natureza é conseguir prever a situação cadastral de novas empresas, este resultado se mostrou muito positivo.

5 CONCLUSÃO

Neste capítulo será realizada uma avaliação do cumprimento dos objetivos estabelecidos e serão apresentadas as limitações e recomendações para trabalhos futuros.

5.1 ATENDIMENTO DOS OBJETIVOS

Este trabalho teve como objetivo desenvolver um modelo de predição de encerramento de micro e pequenas empresas brasileiras da indústria de transformação. A fim de alcançá-lo, foram definidos cinco objetivos específicos, sendo o primeiro referente ao levantamento de fatores relacionados à mortalidade das MPEITs, o segundo relacionado ao conjunto de dados da pesquisa, o terceiro sobre a seleção e criação de variáveis a partir destes dados, o quarto quanto à programação do modelo preditivo e o quinto ligado à aplicação e avaliação deste modelo.

O primeiro objetivo foi atingido a partir de revisão bibliográfica, na qual os fatores relevantes para o encerramento das MPEITs foram identificados e classificados quanto à sua natureza, conforme apresentado no tópico 2.6.

O segundo objetivo foi alcançado no tópico 3.1 por meio da definição, extração e consolidação dos dados das empresas em um banco de dados relacional, o que possibilitou os desenvolvimentos posteriores.

O terceiro foi cumprido por meio da criação e da seleção de variáveis do modelo, considerando os fatores levantados na literatura e a disponibilidade dos dados, conforme apresentado no tópico 3.4.

Por fim, o quarto e o quinto objetivos foram concretizados no tópico 3.6, por meio de um processo cíclico de construção do modelo com diferentes parâmetros e a validação de seu desempenho com testes em grupos de empresas.

Diante do exposto, pode-se concluir que o objetivo geral do trabalho foi alcançado, assim como os seus respectivos objetivos específicos. A análise sobre fatores relacionados ao encerramento das MPEITs trouxe maior compreensão sobre o tema, além de possibilitar uma avaliação quantitativa do impacto destas variáveis. Ademais, os resultados do modelo revelaram a possibilidade de predizer quais MPEITs encerrarão as atividades em até cinco anos após a sua abertura, com uma acurácia média de 70,6%.

5.2 LIMITAÇÕES DO TRABALHO E OPORTUNIDADE PARA FUTURAS PESQUISA

Ainda que os objetivos definidos para esta pesquisa tenham sido cumpridos, foram identificadas algumas limitações durante a sua execução, bem como oportunidades para elaboração de futuros trabalhos relacionados ao tema.

A escassez de pesquisas nacionais sobre a predição de encerramento de micro e pequenas empresas se mostrou como uma limitação para o trabalho, uma vez que as particularidades do país constituem informação relevante para o seu desenvolvimento. Sendo assim, a escolha dos métodos utilizados foi adaptada a partir de revisão bibliográfica de estudos internacionais sobre o tema.

Deve-se destacar ainda que pesquisas nesta área se baseiam, majoritariamente, em dados sensíveis das empresas, como os indicadores financeiros de receitas, despesas e lucratividade. No entanto, considerando que o presente trabalho foi desenvolvido a partir de dados públicos, não houve a possibilidade de acesso a estas informações e fez-se necessário selecionar as variáveis de acordo com a disponibilidade do conjunto de dados. Dessa forma, entende-se que a adição de atributos financeiros ao modelo desenvolvido pode representar um acréscimo significativo em seu desempenho. Esta é uma sugestão para continuidade e aprimoramento do trabalho.

Além disso, devido aos seus requisitos e por ser amplamente utilizada neste contexto, o presente estudo aplicou a técnica de regressão logística para predizer o encerramento das MPEITs. Entretanto, o modelo poderia basear-se também em outros métodos, como a máquina de vetores de suporte, a árvore de decisão ou, até mesmo, redes neurais. Assim, sugere-se que futuros trabalhos comparem o desempenho das predições ao utilizar diferentes técnicas.

É necessário evidenciar também que o modelo foi desenvolvido a partir de micro e pequenas empresas de todos os setores da indústria de transformação, localizadas em qualquer região do território nacional. Da mesma forma que isso possibilitou uma compreensão mais abrangente deste cenário, ressalta-se que há espaço para realizar pesquisas mais específicas, referentes a algum setor ou localidade particulares.

De qualquer modo, tendo em vista os resultados alcançados, entende-se que o trabalho ofereceu contribuições pertinentes para o contexto das micro e pequenas empresas da indústria de transformação. Assim, espera-se que seja capaz de incentivar a elaboração de novas pesquisas sobre o tema.

REFERÊNCIAS

- ALVES, F. P.; LISBOA, W. T. **Vida e Morte da Pequena Empresa no Brasil: Aspectos Conceituais e Contextuais com Vistas ao Reposicionamento Estratégico**. Revista Cesumar Ciências Humanas e Sociais Aplicadas, v. 19, n. 2, p. 479–500, 2014.
- ARAÚJO, F.; MORAIS, F.; PANDOLFI, E. **A Fábula dos Mortos-Vivos: Determinantes da Mortalidade Empresarial Presentes em Micro e Pequenas Empresas Ativas**. XX SEMEAD Seminários em Administração, nov. 2017.
- ATSAN, N. **Failure Experiences of Entrepreneurs: Causes and Learning Outcomes**. Procedia - Social and Behavioral Sciences, v. 235, p. 435–442, nov. 2016.
- BALCAEN, S.; OOGHE, H. **35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems**. The British Accounting Review, v. 38, n. 1, p. 63–93, mar. 2006.
- BARRETTO, Luiz; OLIVEIRA, Valdir. **As pequenas empresas como solução para o desenvolvimento do Brasil**. Correio Braziliense. Disponível em: <<https://www.correiobraziliense.com.br/opiniao/2022/06/5014005-artigo-as-pequenas-empresas-como-solucao-para-o-desenvolvimento-do-brasil.html>>. Acesso em: 20 ago. 2022.
- BATISTA, G. E. DE A. P. A. **Pré-processamento de dados em aprendizado de máquina supervisionado**. Tese (Mestrado) – Universidade de São Paulo, São Paulo, 2003.
- BORGES, R. P.; OLIVEIRA, D. M. DE. **Sobrevivência e Mortalidade das Micro e Pequenas Empresas: estudo dos fatores determinantes e condicionantes**. Enciclopédia Biosfera, Centro Científico Conhecer, v. 10, n. 19, p. 506–512, 2014.
- CAMPOS, N. **Fatores determinantes para o fechamento de micro e pequenas empresas no município do Rio de Janeiro**. 2021. Tese (Mestrado) – Fundação Getúlio Vargas, Rio de Janeiro, 2021. Disponível em: <https://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/31418/Disserta%c3%a7%c3%a3o_Final.pdf?sequence=1&isAllowed=y>. Acesso em: 10 jun. 2022.
- CARDON, M. S.; STEVENS, C. E.; POTTER, D. R. **Misfortunes or mistakes?. Cultural sensemaking of entrepreneurial failure**. Journal of Business Venturing, v. 26, n. 1, p. 79–92, 2011.
- CASTRO, A. **Uma Análise de Impactos do Simples Nacional no DF**. Tese (Mestrado) – Universidade de Brasília, Brasília, 2003.
- CASTRO, C. L.; BRAGA, A. P. **Aprendizado supervisionado com conjuntos de dados desbalanceados**. Sba: Controle & Automação Sociedade Brasileira de Automatica, v. 22, p. 441–466, 1 out. 2011.
- CHAWLA, N. V. et al. **SMOTE: Synthetic Minority Over-sampling Technique**. Journal of Artificial Intelligence Research, v. 16, n. 16, p. 321–357, 1 jun. 2002.

CNI – COMISSÃO NACIONAL DA INDÚSTRIA. **Portal de Informações Estatísticas da Indústria**. Disponível em: <<https://www.portaldaindustria.com.br/cni/estatisticas/>>. Acesso em 15 mai. 2022.

DUTRA, I. S.; PREVIDELLI, J. J. **Fatores condicionantes da mortalidade de empresas: um estudo dos empreendedores de micro e pequenas empresas paranaenses**. Revista Capital Científico - Eletrônica, v. 3, n. 1, p. 0-0, 2005.

EDMISTER, R. O. **An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction**. Journal of Financial and Quantitative Analysis, v. 7, n. 2, p. 1477-1493, 1972.

FABRES, S. F. C.; SILVA, K. DE L.; CAVALCANTI, R. G. **A Correlação entre a mortalidade das Micro e Pequenas Empresas e o índice de Inflação no Brasil**. Congresso Internacional de Administração, 2015.

FACELI, K. et al. **Inteligência artificial: uma abordagem de aprendizado de máquina**. Rio de Janeiro: LTC, 2011.

FAWCETT, T. **An introduction to ROC analysis**. Pattern Recognition Letters, v. 27, n. 8, p. 861–874, jun. 2006.

FERREIRA, F.; OLIVA, F.; SANTOS, S.; GRISI, C. **Análise quantitativa sobre a mortalidade precoce de micro e pequenas empresas da cidade de São Paulo**. Gestão & Produção, v. 19, n. 4, p. 811–823, dez. 2012.

FGV. Fundação Getúlio Vargas. **10 anos do simples nacional**. Cadernos FGV Projetos, ano 11, n. 29, 2016.

FILIPPINI, R. **Operations management research: some reflections on evolution, models and empirical studies in OM**. International Journal of Operations and Production Management, v. 17, n. 7, p. 655-70, 1997.

GIL, A. **Como elaborar projetos de pesquisa**. Atlas: São Paulo, 2007.

GRUS, J. **Data science from scratch: first principles with Python** /. [s.l.] O'Reilly, 2015.

GUERRA, O.; TEIXEIRA, F. **A Sobrevivência Das Pequenas Empresas No Desenvolvimento Capitalista**. Revista de Economia Política, v. 30, n. 117, p. 124–139, 2010.

HAMMES, E. D.; MALLMANN, L. M. **A importância do capital social na abertura de empresa**. Revista Brasileira de Direito Empresarial, v. 7, n. 1, p. 41–61, 12 ago. 2021.

HEATON, J. **An empirical analysis of feature engineering for predictive modeling**, SoutheastCon 2016, 2016, pp. 1-6, 30 mar. 2016.

HOLANDA, N. **Planejamento e Projetos (uma introdução às técnicas de planejamento e elaboração de projetos)**. Rio de Janeiro: APEC. 1975.

HORTA, I. M.; CAMANHO, A. S. **Company failure prediction in the construction industry**. Expert Systems with Applications, v. 40, n. 16, p. 6253–6257, nov. 2013.

HOSMER, D.; LEMESHOW, S. **Applied logistic regression**. New York: John Wiley & Sons, 1989.

IBGE. CONCLA – COMISSÃO NACIONAL DE CLASSIFICAÇÃO. **Busca Online CNAE**. Disponível em: <<https://concla.ibge.gov.br/busca-online-cnae.html?secao=C&tipo=cnae&view=secao>>. Acesso em: 05 mai. 2022.

IBGE. CONCLA – COMISSÃO NACIONAL DE CLASSIFICAÇÃO. **Tabela de Natureza Jurídica**. 2021. Disponível em: <<https://concla.ibge.gov.br/estrutura/natjur-estrutura/natureza-juridica-2021>>. Acesso em: 05 mai. 2022.

KRUGER, S. D.; REBONATTO, C.; MORESCHI, F. **Análise entre as constituições e extinções de empresas na região Sul do Brasil**. Revista Contabilidade e Controladoria, v. 13, n. 2, 24 ago. 2021.

LA ROVERE, R.L. **Sistematização do Painel: Micro, Pequena e Médias Empresas**, IE/UFRJ, p.75-115, 2002.

LIU, B. **Supervised Learning**. *Web Data Mining*, p. 63–132, 2011.

LUDERMIR, T. B. **Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências**. Estudos Avançados, v. 35, n. 101, p. 85–94, abr. 2021.

MAROM, S.; LUSSIER, R. N. **A Business Success Versus Failure Prediction Model for Small Businesses in Israel**. Business and Economic Research, v. 4, n. 2, p. 63, 2014.

MELLAHI, K.; WILKINSON, A. **Organizational failure: A critique of recent research and a proposed integrative framework**. International Journal of Management Reviews, v. 5–6, n. 1, p. 21–41, 2004.

MTE – MINISTÉRIO DO TRABALHO E EMPREGO. **Programa de Disseminação das Estatísticas do Trabalho**. Relação Anual de Informações Sociais (RAIS) Brasília, DF, 2016.

MINUSSI, J. A.; DAMACENA, C.; NESS JR, W. L. **Um modelo de previsão de solvência utilizando regressão logística**. Revista de Administração Contemporânea, v. 6, p. 109–128, 1 dez. 2002.

NARGESIAN, F. et al. **Learning Feature Engineering for Classification**. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, ago. 2017.

NASCIMENTO, A. M. et al. **Artificial intelligence applied to small businesses: the use of automatic feature engineering and machine learning for more accurate planning**. Revista de Contabilidade e Organizações, v. 14, p. e171481, 14 out. 2020.

OECD. **SME and Entrepreneurship Policy in Brazil**. 2020, OECD Studies on SMEs and Entrepreneurship Paris OECD. Disponível em: <<https://www.oecd.org/publications/sme-and-entrepreneurship-policy-in-brazil-2020-cc5feb81-en.htm>>. Acesso em: 30 mai. 2022.

PAIVA, F. H. G. **A desconsideração expansiva da personalidade jurídica na sucessão empresarial ilícita**. Revista Tributária e de Finanças Públicas, v. 144, p. 91–112, 20 out. 2020.

PAULA, G. A. **Modelos de regressão: com apoio computacional**. São Paulo: IME-USP, 2013.

PEREIRA, R.; SOUSA, P. **Fatores de mortalidade de micro e pequenas empresas: um estudo sobre o setor de serviços**. Simpósio de Excelência em Gestão e Tecnologia, 2019.

PINHEIRO, J.; NETO, M. **Fatores que contribuem para mortalidade das micro e pequenas empresas no Brasil**. Brazilian Journal of Development, Curitiba, v. 5, n. 7, p. 11107-11122. 2019.

POMPE, P. P. M.; BILDERBEEK, J. **The prediction of bankruptcy of small- and medium-sized industrial firms**. Journal of Business Venturing, v. 20, n. 6, p. 847–868, nov. 2005.

PRATI, R. C.; BATISTA, G. E. A. P. A.; MONARD, M. C. **Evaluating Classifiers Using ROC Curves**. IEEE Latin America Transactions, v. 6, n. 2, p. 215–222, jun. 2008.

RFB – RECEITA FEDERAL DO BRASIL. **Ato Declaratório Executivo Cocad nº 1/2018**. 2018. Disponível em: <<https://www.gov.br/receitafederal/pt-br/assuntos/orientacao-tributaria/cadastros/cnpj/consultas/ato-declaratorio-executivo-cocad-no-1-2018>>. Acesso em 29 abr. 2022.

RFB – RECEITA FEDERAL DO BRASIL. **Código CNAE Principal**. 2006. Disponível em: <https://www38.receita.fazenda.gov.br/cadsincnac/jsp/coleta/ajuda/topicos/Codigo_CNAE_Principal.htm#:~:text=Codigo%20CNAE%20Principal,maior%20receita%20auferida%20ou%20esperada.>. Acesso em 15 jun. 2022.

RFB – RECEITA FEDERAL DO BRASIL. **Inscrição do Primeiro Estabelecimento (Matriz)**. 2014. Disponível em: <<https://www.gov.br/receitafederal/pt-br/assuntos/orientacao-tributaria/cadastros/cnpj/solicitacao-de-atos-perante-o-cnpj-por-meio-da-internet/inscricao-de-primeiro-estabelecimento-matriz>>. Acesso em 02 jul. 2022.

RFB – RECEITA FEDERAL DO BRASIL. **Portal de Informações de Dados Públicos de CNPJ**. 2022. Disponível em: <<https://www.gov.br/receitafederal/pt-br/assuntos/orientacao-tributaria/cadastros/consultas/dados-publicos-cnpj>>. Acesso em: 25 jul. 2022.

RFB – RECEITA FEDERAL DO BRASIL. **Roteiro da Codificação em CNAE Fiscal**. 2001. Disponível em: <http://subcomissaoacnae.fazenda.pr.gov.br/UserFiles/File/CNAE/Roteiro_CNAE-Fiscal.pdf>. Acesso em 12 jun. 2022.

RIBEIRO, D. P.; OLIVEIRA, U. R. **Fatores críticos de sucesso em uma pequena empresa**. Revista Brasileira de Administração Científica, v. 7, n. 3, p. 139–153, 30 nov. 2016.

RICTOM. **CNPJ-SQLITE**. Disponível em: <<https://github.com/rictom/cnpj-sqlite>>. Acesso em: 20 ago. 2022.

RORATTO, R.; DIAS, E. D.; ALVES, E. B. **Mortalidade em micro e pequenas empresas: Um estudo de caso na Região Central o Rio Grande do Sul.** Revista Espacios, Caracas, v. 38, n. 28, p.1-11, 2017.

SANTINI, S. et al. **Fatores de mortalidade em micro e pequenas empresas: um estudo na região central do Rio Grande do Sul.** Revista Eletrônica de Estratégia & Negócios, 8(1), 145–169. 2015.

SANTOS, V. et al. **Características das empresas e dos gestores que influenciam o uso de controles gerenciais.** Advances in Scientific and Applied Accounting, p. 003-023, abr. 2018.

SARAIVA, A. et al. **Micro e Pequenas Empresas da Indústria de Transformação no desenvolvimento de Imperatriz-Ma: caracterização das indústrias e dos proprietários.** Revista Competitividade E Sustentabilidade, vol. 7, no. 3, pp. 621–637, 13 Dec. 2020.

SARKER, I. H.; KAYES, A. S. M.; WATTERS, P. **Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage.** Journal of Big Data, v. 6, n. 1, 1 jul. 2019.

SBAIRINI, J.; OLIVEIRA, R. **Um estudo sobre o impacto da orientação empresarial na taxa de sobrevivência das microempresas da cidade de Campinas-SP.** Revista da Micro e Pequena Empresa; Sao Paulo Vol. 15, Ed. 2, ago. 2021.

SEBRAE, DIEESE. **Anuário do Trabalho na Micro e Pequena Empresa.** 6. ed. Brasília: SEBRAE, 2013. 17 p. Disponível em: <http://www.sebrae.com.br/Sebrae/Portal%20Sebrae/Anexos/Anuario%20do%20Trabalho%20Na%20Micro%20e%20Pequena%20Empresa_2013.pdf>. Acesso em 20 jun. 2022.

SEBRAE. **Causas Mortis: o sucesso e fracasso das empresas nos primeiros 5 anos de vida.** São Paulo: 2014. Disponível em: <https://www.sebrae.com.br/Sebrae/Portal%20Sebrae/UFs/SP/Anexos/causa_mortis_2014.pdf>. Acesso em: 20 jun. 2022.

SEBRAE. **Estudo Sobre Sobrevivencia Das Empresas No Brasil.** 2016. Disponível em: <<https://www.sebrae.com.br/Sebrae/Portal%20Sebrae/Anexos/sobrevivencia-das-empresas-no-brasil-102016.pdf>>. Acesso em: 20 mai. 2022.

SEBRAE. **Tudo o que você precisa saber sobre o MEI.** 2022. Disponível em: <<https://www.sebrae.com.br/sites/PortalSebrae/artigos/tudo-o-que-voce-precisa-saber-sobre-o-mei,caa7d72060589710VgnVCM100000d701210aRCRD#:~:text=MEI%20significa%20Microempendedor%20Individual%2C%20ou,direitos%20de%20uma%20pessoa%20jur%C3%AAdica.>>>. Acesso em: 20 jul. 2022.

SEBRAE. **Doze anos de monitoramento da sobrevivência e mortalidade de empresas.** São Paulo, 2010.

SEBRAE-SP. **Quem são as micro e pequenas indústrias.** São Paulo: SEBRAE-SP, 2017. Disponível em: <<https://www.sebrae.com.br/sites/PortalSebrae/ufs/sp/bis/quem-sao-as-micros-e-pequenas-industrias,4ba39a6cf8bbf510VgnVCM1000004c00210aRCRD>>. Acesso em 20 jun. 2022.

SHI, Y.; LI, X. **An overview of bankruptcy prediction models for corporate firms: A Systematic literature review.** Intangible Capital, v. 15, n. 2, p. 114, 14 out. 2019.

SILVEIRA, D. T.; CÓRDOVA, F. P. **Métodos de pesquisa.** 1. ed. Porto Alegre: Universidade Federal do Rio Grande do Sul, 2009.

SILVEIRA, F. C. N. et al. **Uma análise da relação das extinções empresarias e indicadores econômicos, frente aos cenários de crises econômicas no brasil.** Revista Expressão Católica, v. 5, n. 1, 30 ago. 2017.

STEVENSON, M.; MUES, C.; BRAVO, C. **The value of text for small business default prediction: A Deep Learning approach.** European Journal of Operational Research, mar. 2021.

ZIZI, Y.; OUDGOU, M.; EL MOUDDEN, A. **Determinants and Predictors of SMEs' Financial Failure: A Logistic Regression Approach.** Risks, v. 8, n. 4, p. 107, 16 out. 2020.

APÊNDICE A – CÓDIGO UTILIZADO PARA DISPONIBILIZAÇÃO E TRATAMENTO DA BASE DE DADOS DA PESQUISA

```
#Importação das bibliotecas que serão utilizadas

import sqlite3
import pandas as pd
import numpy as np
from datetime import datetime
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings("ignore")

# Função para acessar o banco de dados em Sqlite via Python, utilizando a biblioteca
sqlite3

def df_query_result(query):
    conn = sqlite3.connect('cnpj.db') # conecta com o banco de dados utilizado
    conn.row_factory = sqlite3.Row
    cur.execute(query) # executa a query
    rows = cur.fetchall() #retorna o resultado da query, mas é necessário formatar
    data=[]
    try:
        for row in rows:
            data.append(dict(row))
        return pd.DataFrame(data) # cria um DataFrame com os resultados da query
    except:
        return rows
    cur.close_connection() # fecha a conexão com o banco de dados

#Arquivo em Excel com todos os nomes dos subsetores da indústria de transformação

df = pd.read_excel('cnae_names_final.xlsx')

# Código para filtrar somente as indústrias de transformação, de acordo com o código CNAE

df['codigo'] = df['codigo'].apply(lambda x: str(x)+'%')

def create_column(table_name, column_name, type='TEXT'):
    conn = sqlite3.connect('cnpj.db')
    cur = conn.cursor()
    cur.execute (f'ALTER TABLE {table_name} ADD {column_name} {type}')
    conn.commit()
    cur.close

create_column('cnae', 'industria')
```

```

def insert_data(df):
    conn = sqlite3.connect('cnpj.db')
    cur = conn.cursor()
    for codigo, nome in zip(df.codigo, df.nome):
        cur.execute(f'UPDATE cnae SET industria = "{nome}" WHERE codigo LIKE "{codigo}";')
    conn.commit()
    cur.close()
    conn.close()

```

```
insert_data(df)
```

Query para gerar uma base com os dados brutos que serão utilizados na análise

```

query = '''
    select
        -- estabelecimento
        estabelecimento.cnpj,
        estabelecimento.cnpj_basico,
        case
            when estabelecimento.situacao_cadastral = '02' then 'ativa'
            else 'encerrada' end as situacao_cadastral,
        estabelecimento.situacao_cadastral as situacao_cadastral_raw,
        estabelecimento.motivo_situacao_cadastral,
        estabelecimento.data_situacao_cadastral,
        estabelecimento.data_inicio_atividades,
        estabelecimento.cnae_fiscal,
        estabelecimento.cnae_fiscal_secundaria,
        estabelecimento.uf,
        estabelecimento.bairro,
        estabelecimento.cep,
        estabelecimento.tipo_logradouro,
        estabelecimento.nome_fantasia,
        -- empresas
        empresas.capital_social,
        empresas.natureza_juridica,
        empresas.qualificacao_responsavel,
        empresas.porte_empresa,
        -- simples
        case
            when
                simples.data_opcao_simples <= estabelecimento.data_inicio_atividades
                and simples.data_opcao_simples is not null then 'S'
            else
                'N' end as opcao_simples_avalizada,
        simples.opcao_simples,
        simples.data_opcao_simples,
        simples.data_exclusao_simples,
        -- socios

```

```

socios.qualificacao_socio,
socios.qualificacao_representante_legal,
socios.data_entrada_sociedade,
socios.faixa_etaria,
socios.nome_socio,
socios.cnpj_cpf_socio,
-- cnae
cnae.descricao,
cnae.industria
from estabelecimento
left join empresas on estabelecimento.cnpj_basico = empresas.cnpj_basico
inner join socios on estabelecimento.cnpj_basico = socios.cnpj_basico
left join simples on estabelecimento.cnpj_basico = simples.cnpj_basico
inner join cnae on estabelecimento.cnae_fiscal = cnae.codigo
where
    cnae.industria is not null
    and porte_empresa in ('01', '03')
    and cast(data_inicio_atividades as int) >= 20100101
    and socios.nome_socio is not null
    and socios.data_entrada_sociedade <= data_inicio_atividades
    and estabelecimento.matriz_filial = '1'
    and estabelecimento.cnpj_ordem = '0001'
    ...

```

```
df_base = df_query_result(query)
```

Função para acrescentar uma coluna com o número de empresas ativas que o sócio possuía na época em que abriu a empresa

```

def num_empresas_ativas(df):
    conn = sqlite3.connect('cnpj.db')
    cur = conn.cursor()
    num_empresas_ativas = []
    for nome_socio, cnpj_cpf_socio, data_inicio_atividades in zip(df['nome_socio'],
df['cnpj_cpf_socio'], df['data_inicio_atividades']):
        query = f'''
            select
                socios.cnpj_cpf_socio || socios.nome_socio,
                count(distinct socios.cnpj)
            from
                socios
            left join
                estabelecimento on socios.cnpj = estabelecimento.cnpj
            where
                (socios.cnpj_cpf_socio = '{cnpj_cpf_socio}'
                and socios.nome_socio = '{nome_socio}'
                and estabelecimento.data_inicio_atividades < {data_inicio_atividades}
                and estabelecimento.matriz_filial = '1')
                and (estabelecimento.situacao_cadastral = '02'

```

```

        or estabelecimento.situacao_cadastral != '02'
        and estabelecimento.data_situacao_cadastral > {data_inicio_atividades})
    group by 1
    ...
cur.execute(query)
row = cur.fetchall()
if len(row)==0:
    num_empresas_ativas.append(0)
else:
    num_empresas_ativas.append(row[0][-1])

cur.close()
conn.close()

df.loc[:, 'num_empresas_ativas'] = num_empresas_ativas

return df

# Função para acrescentar uma coluna com o número de empresas encerradas que o sócio possui

def num_empresas_encerradas(df):
    conn = sqlite3.connect('cnpj.db')
    cur = conn.cursor()
    num_empresas_encerradas = []
    for nome_socio, cnpj_cpf_socio, data_inicio_atividades in zip(df['nome_socio'],
df['cnpj_cpf_socio'], df['data_inicio_atividades']):
        query = f'''
            select
                socios.cnpj_cpf_socio || socios.nome_socio,
                count(distinct socios.cnpj)
            from
                socios
            left join
                estabelecimento on socios.cnpj = estabelecimento.cnpj
            where
                socios.cnpj_cpf_socio = '{cnpj_cpf_socio}'
                and socios.nome_socio = '{nome_socio}'
                and estabelecimento.data_inicio_atividades < {data_inicio_atividades}
                and estabelecimento.matriz_filial = '1'
                and estabelecimento.situacao_cadastral != '02'
            group by 1
            ...
cur.execute(query)
row = cur.fetchall()
if len(row)==0:
    num_empresas_encerradas.append(0)
else:
    num_empresas_encerradas.append(row[0][-1])

```

```

cur.close()
conn.close()

df.loc[:, 'num_empresas_encerradas'] = num_empresas_encerradas

return df

# Estabelecimentos ativos que existiam no mesmo CEP, no momento de abertura da empresa

def estabelecimento_cep_ativas(df):
    conn = sqlite3.connect('cnpj.db')
    cur = conn.cursor()
    num_estabelecimento_cep_ativas = []
    tempo_medio_cep_ativas = []
    for cep, data_inicio_atividades in zip(df['cep'], df['data_inicio_atividades']):
        query = f'''
            select
                count(estabelecimento.cnpj)
            from estabelecimento
            where
                (estabelecimento.cep = "{cep}"
                 and estabelecimento.data_inicio_atividades <
"{data_inicio_atividades}")
                and (estabelecimento.situacao_cadastral = '02'
                    or estabelecimento.situacao_cadastral != '02'
                    and estabelecimento.data_situacao_cadastral >
"{data_inicio_atividades}")
            '''
        cur.execute(query)
        row = cur.fetchall()
        num_estabelecimento_cep_ativas.append(row[0][0])

    cur.close()
    conn.close()

    df.loc[:, 'num_estabelecimento_cep_ativas'] = num_estabelecimento_cep_ativas

    return df

#Estabelecimentos encerrados que existiam no mesmo CEP, no momento de abertura da empresa

def estabelecimento_cep_encerradas(df):
    conn = sqlite3.connect('cnpj.db')
    cur = conn.cursor()
    num_estabelecimento_cep_encerradas = []
    tempo_medio_cep_encerradas = []
    for cep, data_inicio_atividades in zip(df['cep'], df['data_inicio_atividades']):
        query = f'''

```

```

        select
            count(estabelecimento.cnpj)
        from estabelecimento
        where
            estabelecimento.cep = "{cep}"
            and estabelecimento.data_situacao_cadastral <=
"{data_inicio_atividades}"
            and estabelecimento.situacao_cadastral != '02'
        ...
        cur.execute(query)
        row = cur.fetchall()
        num_estabelecimento_cep_encerradas.append(row[0][0])

    cur.close()
    conn.close()

    df.loc[:, 'num_estabelecimento_cep_encerradas'] = num_estabelecimento_cep_encerradas

    return df

# Cálculo do tempo de funcionamento da empresa

data_bd = datetime(2022,2,12).date()

df.loc[:, 'tempo_empresa'] = np.where(df['situacao_cadastral'] == 'ativa',
                                     data_bd - df['data_inicio_atividades'],
                                     df['data_situacao_cadastral'] -
df['data_inicio_atividades'])

# Contagem do número de CNAEs fiscais secundários por empresa

df.loc[:, 'num_cnae_sec'] = np.where(df['cnae_fiscal_secundaria'] == '',
                                     0,
                                     df['cnae_fiscal_secundaria'].apply(lambda x: len(x.split(','))))

# Cálculo do número de sócios por empresa

num_socios =
df.groupby('cnpj')[['qualificacao_socio']].count().rename(columns={'qualificacao_socio': 'nu
m_socios'}).reset_index()

df = df.merge(num_socios, how='left')

# Seleção de um registro de cada empresa, criando colunas com a faixa etária do sócio mais
velho (faixa_etaria_o) e do mais novo (faixa_etaria_y)

df_y = df.sort_values(['cnpj', 'faixa_etaria'],
ascending=True).drop_duplicates('cnpj', keep='first').reset_index(drop=True)
df_y.rename(columns={'faixa_etaria': 'faixa_etaria_y'}, inplace=True)

```

```
df_o = df.sort_values(['cnpj', 'faixa_etaria'],
ascending=False).drop_duplicates('cnpj', keep='first').reset_index(drop=True)
df_o.rename(columns={'faixa_etaria': 'faixa_etaria_o'}, inplace=True)

df = df_o.merge(df_y[['cnpj', 'faixa_etaria_y']], on='cnpj')

# Criação de nova coluna com a população de cada estado

tabela_estados =
pd.read_html('https://www.estadosecapitaisdobrasil.com/')[0][['Sigla', 'Região', 'Estado']]

df = pd.merge(df, tabela_estados, how='left', left_on='uf', right_on='Sigla')

# Salvar o arquivo no formato CSV

df.to_csv('base_modelo_completa.csv')
```

APÊNDICE B – CÓDIGO UTILIZADO PARA O DESENVOLVIMENTO E PARA A AVALIAÇÃO DO MODELO DE PREDIÇÃO DE ENCERRAMENTO

```
# Importação das bibliotecas que serão utilizadas

import sqlite3
import pandas as pd
import numpy as np
from datetime import datetime, date
import seaborn as sns
import matplotlib.pyplot as plt
from dateutil.relativedelta import relativedelta
import statistics
import warnings
warnings.filterwarnings("ignore")
import scipy.stats as stats
from sklearn.preprocessing import StandardScaler
from imblearn.over_sampling import SMOTE
from sklearn.linear_model import LogisticRegression
import statsmodels.api as sm
from sklearn.model_selection import train_test_split

# Ler o arquivo CSV com a base das MPEs da Indústria de Transformação Tratada na Etapa
Anterior

df = pd.read_csv('base_modelo_completa.csv', converters = {'cnpj':str,
'cep':str,'ano_inicio_atividades':str})

# Convertendo as datas que estavam em formato de texto, como datas

df.loc[:, 'data_inicio_atividades'] = df['data_inicio_atividades'].apply(lambda x:
datetime.strptime(x, "%Y-%m-%d").date())

df.loc[:, 'data_situacao_cadastral'] = np.where(df['data_situacao_cadastral']=='0',
df['data_inicio_atividades'],
df['data_situacao_cadastral'])

df.loc[:, 'data_situacao_cadastral'] = df['data_situacao_cadastral'].apply(lambda x:
datetime.strptime(x, "%Y-%m-%d").date() if type(x) is str else x)

# Função para gerar base das empresas fundadas no ano selecionado e avaliar se estavam
ativas ou encerradas

def base_empresa(ano_inicio_atividades='2017', somente_encerrada='N', codigo_setor='todos',
anos_avaliacao = 'todos', uf = 'todas', motivo_encerramento = 'todos',
regiao = 'todas', cols='todas', df = df, remo_out=False):
```

```

# Avaliação do ano de início das atividades
try:
    if len(ano_inicio_atividades) > 1:
        df = df[df['ano_inicio_atividades'].isin(ano_inicio_atividades)]
except:
    df = df[df['ano_inicio_atividades'].isin([str(ano_inicio_atividades)])]

# Avaliação situação cadastral
if somente_encerrada == 'N':
    pass
else:
    df = df[df['situacao_cadastral_raw'].isin([2,8])]

# Avaliação setor
if codigo_setor == 'todos':
    pass
else:
    df = df[df['codigo_cnae_setor']==str(codigo_setor)]

# Seleção dos estados

if uf == 'todas':
    pass
else:
    df = df[df['uf']==uf]

# Seleção das regiões

if regioao == 'todas':
    pass
else:
    df = df[df['regiao']==regiao]

# Motivo encerramento

if motivo_encerramento == 'todos':
    pass
elif motivo_encerramento == 'v':
    df = df[df['motivo_situacao_cadastral'].isin([0,1])]
else:
    df = df[df['motivo_situacao_cadastral'].isin([0,63])]

# Seleção dos anos

if anos_avaliacao == 'todos':
    df.loc[:, 'data_avaliacao'] = df['data_inicio_atividades'].max()
else:
    df.loc[:, 'data_avaliacao'] = df['data_inicio_atividades'].apply(lambda x: x +
relativedelta(years=anos_avaliacao))

```

```

# Avaliação se a empresa estava aberta ou fechada no período selecionado

situacao_cadastral_avalizada = []

for i in range(len(df)):
    if df['situacao_cadastral'].iloc[i] == 'ativa':
        situacao_cadastral_avalizada.append(1)
    else:
        if(df['data_situacao_cadastral'].iloc[i] > df['data_avalizacao'].iloc[i]):
            situacao_cadastral_avalizada.append(1)
        else:
            situacao_cadastral_avalizada.append(0)

df.loc[:, 'situacao_cadastral_avalizada'] = situacao_cadastral_avalizada

# Remoção das empresas que declararam um capital social inferior a R$ 500,00 ou
superior a R$ 15.000.000,00

if remo_out == True:
    df = df[(df['capital_social'] >= 500) & (df['capital_social'] <=
15000000)].reset_index(drop=True)
else:
    pass

# Seleção das colunas informadas

if cols == 'todas':
    pass
else:
    df = df[cols]

print('Total de empresas:',df.shape[0])

display(df['situacao_cadastral_avalizada'].value_counts(normalize=True))

return df.reset_index(drop=True)

# Divisão da base em treino e teste e aplicação de dummies nas colunas

def get_dummies(df,non_dummy_cols):
    """Retorna um DataFrame com os dummies das colunas desejadas."""

    dummy_cols = list(set(df.columns) - set(non_dummy_cols))
    df = pd.get_dummies(df, columns=dummy_cols)
    return df

def train_test(df, columns, non_dummy_cols, sample_size=12000, test_size=0.3, smote=True):

```

```

# Seleção das colunas que serão utilizadas no modelo

if columns == 'todas':
    pass
else:
    df = df[columns]

# Aplicar a função get_dummies nas variáveis explicativas categóricas

df = get_dummies(df, non_dummy_cols)

# Seleção de uma amostra aleatória conforme o número informado. Se for igual a 0, serão
selecionados todos os registros

if sample_size == 0:
    df_sample = df
else:
    df_sample = df.sample(sample_size, random_state=42)

# Divisão do DataFrame entre a variável dependente (X) e das independentes (y)

X = df_sample.loc[:, df_sample.columns != "situacao_cadastral_avalizada"]

y = df_sample['situacao_cadastral_avalizada']

# Divisão do modelo em dados de treino e teste. Por padrão, 70% e 30%, respectivamente.

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=test_size,
random_state=42)

# Aplicação do método SMOTE para lidar com dados desbalanceados

if smote==True:

    oversample = SMOTE(random_state=42)

    X_train, y_train = oversample.fit_resample(X_train, y_train)

else:
    pass

return X_train, X_test, y_train, y_test

# Função para aplicação do modelo de regressão logística

def scipy_models(X_train, X_test, y_train, y_test, class_weight='balanced', C=0.1, model =
'LR'):

    if model == 'LR':

```

```

        clf = svm.SVC(random_state=42, C=C, gamma='auto', class_weight=class_weight,
kernel='rbf', probability=True)

        clf.fit(X_train, y_train)

        y_pred = clf.predict(X_test)

        y_pred_prob = clf.predict_proba(X_test)[:, 1]

        print('score:',clf.score(X_test, y_test))

        print('roc_auc_score:', roc_auc_score(y_test,y_pred_prob))

        print('confusion_matrix:', confusion_matrix(y_test,y_pred))

        return clf

# Aplicação da função para gerar a base de empresas abertas em 2011

df_base = base_empresa(ano_inicio_atividades = '2011', anos_avaliacao=5,
somente_encerrada=True, cols='todas',
                        codigo_setor='todos', regiao='todas', df=df, remo_out=True)

# Aplicação do modelo de regressão logística na base selecionada e avaliação do resultado

X_train, X_test, y_train, y_test = train_test(df=df_base, columns=columns,
sample_size=5000,
                                            non_dummy_cols = non_dummy_cols, smote=True)

clf = scipy_models(X_train = X_train, X_test = X_test, y_train = y_train, y_test = y_test,
                   C=0.15, class_weight='balanced', model='LR')

# Aplicação do modelo desenvolvido em empresas abertas no ano de 2012

df_base_2 = base_empresa(ano_inicio_atividades = '2012', anos_avaliacao=5,
somente_encerrada=True, cols='todas',
                        codigo_setor='todos', regiao='todas', df=df, remo_out=True)

apply_model(df=df_base_2, columns=columns, non_dummy_cols=non_dummy_cols, model='LR')

```