



UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS SOCIAIS APLICADAS
DEPARTAMENTO DE FINANÇAS E CONTABILIDADE
CURSO DE CIÊNCIAS ATUARIAIS

JESSICA EVELIN SILVA DAMACENA

PREVISÃO DE INSOLVÊNCIA DE SEGURADORAS BRASILEIRAS UTILIZANDO
MODELOS DE APRENDIZAGEM DE MÁQUINAS

JOÃO PESSOA - PB

2022

JESSICA EVELIN SILVA DAMACENA

**PREVISÃO DE INSOLVÊNCIA DE SEGURADORAS BRASILEIRAS UTILIZANDO
MODELOS DE APRENDIZAGEM DE MÁQUINAS**

Trabalho de Conclusão de Curso para o curso de Ciências Atuariais do Centro de Ciências Sociais Aplicadas, da Universidade Federal da Paraíba, como requisito para a obtenção do título de Bacharel em Ciências Atuariais.

Orientador: Prof.º Me. Filipe Coelho de Lima Duarte.

JOÃO PESSOA – PB

2022

JESSICA EVELIN SILVA DAMACENA

**PREVISÃO DE INSOLVÊNCIA DE SEGURADORAS BRASILEIRAS UTILIZANDO
MODELOS DE APRENDIZAGEM DE MÁQUINAS**

Trabalho de Conclusão de Curso para o curso de Ciências Atuariais do Centro de Ciências Sociais Aplicadas, da Universidade Federal da Paraíba, como requisito para a obtenção do título de Bacharel em Ciências Atuariais.

BANCA EXAMINADORA



Prof. Me. Filipe Coelho de Lima Duarte.

(Orientador)

UNIVERSIDADE FEDERAL DA PARAIBA (UFPB)



Profª. Dra. Vera Lúcia Cruz

(Avaliadora)

UNIVERSIDADE FEDERAL DA PARAIBA (UFPB)



Prof. Dr. Herick Cidarta Gomes De Oliveira

(Avaliador)

UNIVERSIDADE FEDERAL DA PARAIBA (UFPB)

Catálogo na publicação
Seção de Catalogação e Classificação

D154p Damacena, Jessica Evelin Silva.

Previsão de insolvência de seguradoras brasileiras
utilizando modelos de aprendizagem de máquinas /
Jessica Evelin Silva Damacena. - João Pessoa, 2022.
67 f.

Orientação: Filipe Coelho de Lima Duarte.
TCC (Graduação) - UFPB/CCSA.

1. Previsão de insolvência. 2. Aprendizagem de
máquinas. 3. Seguradoras. 4. Indicadores contábeis. I.
Duarte, Filipe Coelho de Lima. II. Título.

UFPB/CCSA

CDU 368(02)

AGRADECIMENTOS

Aos meus pais, **Agnalva** e **Edinaldo**, pelo amor, conselhos e incentivos que sempre me ofereceram.

Aos meus primos e avós, em especial a minha avó **Berenice** que estará eternamente em meu coração.

Ao meu orientador **Prof. Me. Filipe Coelho de Lima Duarte**, pelo auxílio, atenção e confiança e ensinamentos.

Ao meu coordenador **Prof. Dr. Luiz Carlos Santos Júnior**, pelo apoio, paciência e ensinamentos.

A todos os meus amigos e colegas, que sempre estiveram comigo.

A todos os professores e professoras que um dia me ensinaram e contribuíram para a minha formação.

Muito obrigada!

“Arrisque-se! Toda vida é um risco. O homem que vai mais longe é geralmente aquele que está disposto a fazer e a ousar. O barco da segurança nunca vai muito além da margem.”

(Dale Carnegie)

RESUMO

As companhias de seguros, mediante recebimento de prêmios, assumem riscos de terceiros e se tornam responsáveis pela reposição de bens à sociedade. A partir das demonstrações contábeis extraídas do banco de dados do Sistema de Estatística (SES) da SUSEP, desenvolver modelos de aprendizagem de máquina que permitam classificar adequadamente as seguradoras brasileiras em solventes e insolventes. O estudo foi produzido em programação *Python*, com a importação do conjunto de dados disponibilizado pelo SES onde foram extraídos as demonstrações contábeis anuais auditadas do período de todos os anos disponibilizados pela SUSEP, assim como todas as seguradoras. Com os dados brutos foi feito o pré-processamento das informações para iniciar a análise exploratória dos dados, calcular os indicadores e obter o índice de solvência de acordo com as normas da SUSEP e então classificar as demonstrações contábeis em solvente ou insolvente. Aplicando os dados aos classificadores escolhidos foi possível encontrar os modelos que são mais adequados a esse tipo de informação, assim como os que tiveram menos capacidade de se adaptar ao conjunto de dados, esse ranqueamento foi feito a partir da matriz de confusão e dos indicadores de desempenho do modelo, principalmente a acurácia. Foi concluído que o algoritmo de Random Forest foi capaz de prever 4045 das 4063 previsões realizadas, obteve 99,56% de Accuracy, Recall de 98%, Precision de 98%. Sendo assim, confirmando ser o modelo que conseguiu trazer mais resultados satisfatórios. Já o algoritmo que menos se adaptou ao conjunto de dados foi a Análise Discriminante que conseguiu prever corretamente 3881 das 4063 previsões realizadas, obteve 95,52% de Accuracy, Recall de 58%, Precision de 98% e F1-Score de 73%.

Palavras-Chave: Previsão de Insolvência. Aprendizagem de Máquina, Seguradoras, Indicadores Contábeis

ABSTRACT

Insurance companies, upon receipt of premiums, assume third-party risks and become responsible for the replacement of assets to society. Based on the financial statements extracted from the SUSEP Statistical System (SES) database, develop machine learning models that allow for the proper classification of Brazilian insurers into solvent and insolvent. The study was produced in Python programming, with the import of the data set made available by SES, where the audited annual financial statements for the period of all years made available by SUSEP, as well as all insurance companies, were extracted. With the raw data, the pre-processing of the information was carried out to start the exploratory analysis of the data, calculate the indicators and obtain the solvency index according to SUSEP standards and then classify the financial statements into solvent or insolvent. Applying the data to the chosen classifiers, it was possible to find the models that are most suitable for this type of information, as well as those that had less capacity to adapt to the data set, this ranking was made from the confusion matrix and the indicators of model performance, mainly the accuracy. It was concluded that the Random Forest algorithm was able to predict 4045 of the 4063 predictions performed, it obtained 99.56% of Accuracy, Recall of 98%, Precision of 98%. Thus, confirming to be the model that managed to bring more satisfactory results. The algorithm that least adapted to the data set was the Discriminant Analysis, which was able to correctly predict 3881 of the 4063 predictions made, obtaining 95.52% of Accuracy, Recall of 58%, Precision of 98% and F1-Score of 73%.

Keywords: Forecast of Insolvency. Machine Learning, Insurance, Accounting Indicators

LISTAS DE FIGURAS

Figura 1 - Empresa Solvente e Insolvente	17
Figura 2 - Divisão de dados para treino e teste com k-fold	39
Figura 3 - Exemplo de previsão com Regressão Logística	43
Figura 4 - Representação da classificação por meio do algoritmo KNN	45
Figura 5 - Exemplo de Árvore de Decisão	46
Figura 6 - Formação de uma Random Forest a partir de Árvores de Decisões	48
Figura 7 - Esquema de classificação através do algoritmo SVM	48
Figura 8 - Arquitetura de uma rede neural artificial	51
Figura 9 - Representação da matriz das variáveis explicativas que permaneceram na análise	55
Figura 10 - Matriz de confusão de todos os algoritmos otimizados	59
Figura 11 - Comparando os classificadores	61

LISTA DE QUADROS

Quadro 1 - Resumo dos Capitais definidos na resolução	22
Quadro 2- Resumo das deduções definidas da resolução.....	23
Quadro 3 - Resumo das provisões técnicas	25
Quadro 4 - Limite percentual permitido pelo BCB para cada tipo de investimento nas aplicações dos recursos das reservas e das provisões.....	26
Quadro 5 - Resolução CNSP nº 321/2015.....	26
Quadro 6 - Exigências da SUSEP depois de 3 meses de insuficiência	27
Quadro 7 - Índices Financeiros.....	34
Quadro 8 - Índices Econômicos	35
Quadro 9 - Resumo dos hiper-parâmetros otimizados	39
Quadro 10 - Matriz de Confusão	52
Quadro 11 - Métricas de avaliação para modelos de classificação	53

LISTA DE TABELAS

Tabela 1 - Frequência de demonstrações financeiras classificadas em solventes e insolventes com base no Capital Mínimo Requerido e no Patrimônio Líquido Ajustado das seguradoras selecionadas.....	58
---	----

LISTA DE ABREVIATURAS E SIGLAS

ADM	Análise Discriminante Múltipla
BCB	Banco Central do Brasil
CMR	Capital Mínimo Requerido
CNSP	Conselho Nacional de Seguros Privados
CR	Capital de risco
CRcred	Capital de risco de crédito
CRmerc	Capital de risco de mercado
CRoper	Capital de risco operacional
CRsubs	Capital de risco de subscrição
DPVAT	Danos Pessoais Causados por Veículos Automotores de Vias Terrestres
FIP	Formulários de Informações Periódicas
IBNR	Provisão de Sinistros Ocorridos e Não Avisados
LT	Limite de retenção
MARS	Multivariate Adaptive Regression Splines
MLP	Rede Multilayer Perceptron
PCC	Provisão Complementar de Cobertura
PDR	Provisão de Despesas Relacionadas
PEF	Provisão de Excedentes Financeiros
PET	Provisão de Excedentes Técnicos
PIB	Produto Interno Bruto
PLA	Patrimônio Líquido Ajustado
PMBAC	Provisão Matemática de Benefícios a Conceder
Continua	

Continuação

LISTA DE ABREVIATURAS E SIGLAS

PMBC	Provisão Matemática de Benefícios Concedidos
PPNG	Provisão de Prêmios Não Ganhos
PRS	Plano de Regularização de Solvência
PSL	Provisão de Sinistros a Liquidar
PVR	Provisão de Resgates e Outros Valores a Regularizar
RNA	Redes Neurais Artificiais
SERASA	Serviços de Assessoria S.A.
SES	Sistema de Estatística
SUSEP	Superintendência de Seguros Privados
SVM	Support Vector Machine

SUMÁRIO

1 INTRODUÇÃO.....	16
1.1. Objetivo	18
1.1.1. Objetivo Geral	18
1.1.2. Objetivo Específico	18
1.2. Justificativa	18
2 REVISÃO DA LITERATURA	21
2.1. Controle Da Solidez Econômico-Financeira Das Seguradoras	21
2.1.1. Capital Mínimo Requerido (CMR)	22
2.1.2. Patrimônio Líquido Ajustado (PLA)	23
2.1.3. Limite Técnico	24
2.1.4. Constituição de Provisões Técnicas.....	25
2.1.5. Ativos Garantidores das Provisões Técnicas	25
2.1.6. Plano de regularização de solvência	26
2.2. Modelos De Previsão/Classificação De Insolvência	27
2.3. Aprendizagem De Máquina	29
3 METODOLOGIA.....	31
3.1. Coleta dos dados	31
3.2. População e Amostra	31
3.3. Etapas da Pesquisa	31
3.3.1. Métodos de Avaliação da Solvência (Susep)	33
3.3.1.1. Análise Econômico-Financeira por Quocientes.....	33
3.3.1.1.1. Análise Financeira	33
3.3.1.1.2. Análise Econômica.....	35
3.3.1.2. Margem de Solvência.....	37
3.3.2. Métodos de Classificação Propostos	38
3.3.2.1. Validação Cruzada	38
3.3.3. Hiper-parâmetros dos modelos de aprendizagem de máquinas.....	39
3.3.4. Modelos de Aprendizagem de Máquinas Propostos	41
3.3.4.1. Análise Discriminante Linear.....	41
3.3.4.2. Regressão Logística	42
3.3.4.3. K-Nearest Neighbor (KNN)	44
3.3.4.4. Árvore de decisão.....	45
3.3.4.5. Extreme Gradient Boosting (XGBoost).....	46
3.3.4.6. Random Forest	47

3.3.4.7. Support Vector Machine - SVM.....	48
3.3.4.8. Redes Neurais Artificiais.....	49
3.4. Medidas de Avaliação de Eficácia	52
3.4.1. Matriz de confusão.....	52
3.4.2. Métricas de avaliação.....	53
4 RESULTADOS	55
4.1. Análise Exploratória dos Dados.....	55
4.2. Validação dos modelos.....	59
4.2.1. Matrizes de Confusão.....	59
4.2.2. Resultado dos modelos com validação cruzada.....	60
5 CONCLUSÃO.....	63
REFERENCIAS	64

1 INTRODUÇÃO

As companhias de seguros, mediante recebimento de prêmios, assumem riscos de terceiros e se tornam responsáveis pela reposição de bens à sociedade. Ao aceitar um risco, a seguradora deve respeitar os limites e as exigências legais que visam garantir as condições de pagamento aos segurados (SILVA, 2017).

O mercado de seguros nacional fechou o ano de 2019 com arrecadação de R\$ 270,1 bilhões, excluindo saúde suplementar e o seguro obrigatório de Danos Pessoais Causados por Veículos Automotores de Vias Terrestres (DPVAT). O setor no Brasil está em constante crescimento, muitos fatores contribuem para ampliação do setor, alguns desses fatores são: a criação de nichos de produtos para atender as necessidades e expectativas do público-alvo, a fomentação da cultura de seguros em grandes segmentos da população e a divulgação da necessidade da proteção contra riscos que podem afetar o seu bem-estar e competitividade das empresas. O crescimento nominal do mercado em relação ao ano de 2018 foi de 12,1%, enquanto o aumento real foi de 8,1%. De 2009 a 2014, o setor de seguros brasileiro cresceu mais de dois dígitos, superando o aumento do Produto Interno Bruto (PIB), com maior crescimento, de 21,7%, em 2012 (GANDRA, 2020).

Os países da América Latina que se consolidaram com a maior participação do setor em relação ao PIB, o Chile com 4,9%, o Brasil com 4,1% e a Colômbia com 2,9%. No entanto, estão bem longe dos números de economias desenvolvidas como Coreia do Sul, Reino Unido ou Japão que atingem 11,6%, 9,6% e 8,6%, respectivamente. Em 2018, o setor latino-americano registrou um volume total de prêmios próximo a US\$ 159.217 bilhões, 8,6% a mais que no ano anterior, segundo dados da Mapfre em 2019. Nesse sentido, os números do segmento vida destacam-se com um aumento de 9%, enquanto os gerais mostram crescimento de 8,2%, dentre os quais o seguro automóvel é o mais importante, com 19,2% do total de prêmios, seguidos pelo seguro de saúde com 10,5% (ANALITIK, 2019).

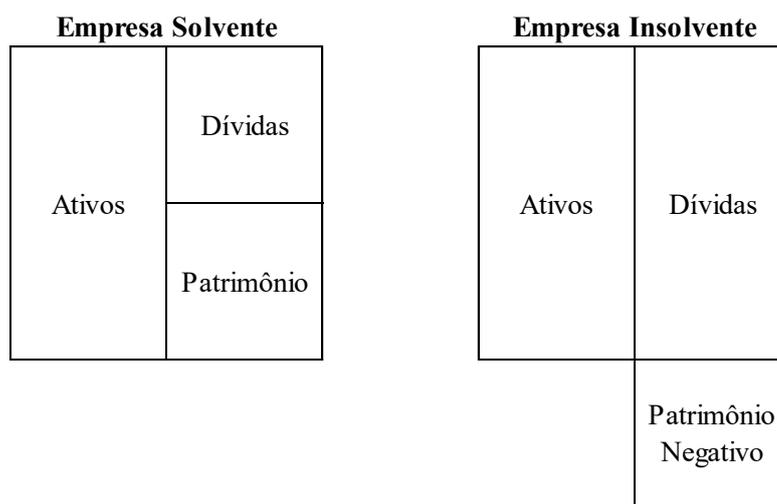
O setor de seguros vem apresentando crescimento e têm participação relevante no Produto Interno Bruto do país, por isso, torna-se necessário o desenvolvimento de métodos adequados para a análise econômico-financeira das empresas de seguros (DALCIN, 2021).

Conforme Alencar (2018), quando uma seguradora decreta falência por má gestão, e ou pela não observância da regulamentação do setor, o segurado fica no prejuízo, por essa razão a necessidade de analisar bem a seguradora, observar se ela está em dia com suas obrigações e se tem efetivamente condições de oferecer os produtos adequados.

De acordo com Silva (2009), o estado de insolvência de uma empresa pode ser definido como a incapacidade para pagar as suas obrigações financeiras na data de seu vencimento, bem como quando seus ativos forem inferiores ao valor dos seus passivos.

Uma empresa solvente é aquele em que o estado do devedor que possui seu ativo pelo menos igual ao passivo. Portanto, do ponto de vista econômico, uma empresa é solvente quando está em condições de fazer frente a suas obrigações correntes e ainda apresentar uma situação patrimonial e uma expectativa de lucros que garantam sua sobrevivência no futuro. Já uma empresa insolvente ocorre quando o patrimônio se torna negativo. Isso acontece quando os valores das obrigações superam a soma de todos os ativos de uma empresa. Dito em outras palavras, com o patrimônio líquido negativo, a empresa possui um passivo maior do que o ativo (SUNO, 2018).

Figura 1- Empresa Solvente e Insolvente



Fonte: Silva (2017)

Os modelos de previsão são relevantes para os gestores, pois a partir dos seus resultados é possível tomar decisões e medidas com antecedência e evitar maiores problemas para a companhia (PROCENGE, 2018).

As principais contribuições no estudo de previsão de insolvência foram de: Fitzpatrick (1932) constituídos por 19 empresas insolventes e solventes e os que os principais índices utilizados pelo autor foram as relações entre Patrimônio Líquido sobre Passivo e o Lucro Líquido em relação ao Patrimônio Líquido, Beaver (1966) utilizando os demonstrativos financeiros das empresas e uma amostra com 79 empresas que apresentaram problemas de solvência durante os anos 1954 a 1968 e analisou as características de uma empresa insolvente, sendo seguido por Altman (1968) que fez seu estudo de insolvência utilizando-se a análise

discriminante multivariada, esses pesquisadores influenciaram diversas pesquisas no Brasil como Elizabetsky (1976), Kanitz (1974), Matias (1978) e Silva (1982).

Para automatizar os modelos de previsão que são complexos e seus algoritmos de predições pode ser utilizado o aprendizado de máquina, que envolve o estudo da teoria do reconhecimento de padrões, a construção de algoritmos e aprendizagem computacional em inteligência artificial, que pode treinar, aprender e fazer predições sobre determinados dados (ZHU et al. 2016). No que se refere a aprendizagem de máquinas concentrados na análise de insolvência, estudos como o de: Scarpel (2008), onde foi comparado o desempenho do modelo *support vector machine* com diferentes abordagens de análise discriminante (linear e quadrática) em previsão de insolvência de empresas. Onusic, Casa Nova e Almeida (2017), comparando os resultados das técnicas Análise por Envoltória de Dados e Redes Neurais utilizadas para prever a insolvência em uma base de dados fornecida pela Serviços de Assessoria S.A. (SERASA). Horta et al. (2010) nesse realizaram a comparação de seleção de atributos aplicados em previsão de insolvência empresas brasileiras não financeiras de capital aberto.

Diante o exposto, o presente estudo tem a finalidade de responder ao seguinte problema de pesquisa: **Qual o modelo de aprendizagem de máquinas que pode permitir a classificação adequada das seguradoras brasileiras em solventes e insolventes?**

1.1. Objetivo

1.1.1. Objetivo Geral

A partir das demonstrações contábeis extraídas do banco de dados do Sistema de Estatística (SES) da SUSEP, desenvolver modelos de aprendizagem de máquina que permitem classificar adequadamente as seguradoras brasileiras em solventes e insolventes.

1.1.2. Objetivo Específico

- Explorar a classificação dos indicadores de solvência definidos pela SUSEP.
- Mensurar o nível de solvência das companhias seguradoras brasileiras pelos indicadores da SUSEP.
- Desenvolver uma análise comparativa dos modelos de aprendizagem de máquinas.

1.2. Justificativa

Conforme Silva (2017), no Brasil, a fiscalização do setor de seguros é de responsabilidade da SUSEP, a qual possui personalidade jurídica de direito público e executa as políticas do Conselho Nacional de Seguros Privados (CNSP). Cabe a essa autarquia fiscalizar, organizar e controlar as companhias de seguros, previdência privada aberta, capitalização e resseguro, além de zelar pelos interesses dos consumidores desse mercado. Dentre as atribuições da SUSEP, ressalta-se o acompanhamento da liquidez e solvência das sociedades que integram o mercado, a fim de garantir controle da solidez econômico-financeira.

Segundo Affonso Silva (2009), torna-se imperioso para todos aqueles que, na estrutura da empresa, detenham o poder de decisão, não perder de vista a total obediência aos parâmetros instituídos pelas normas vigentes, no sentido de resguardar, de todas as formas, a solvência da seguradora, desde a constituição do seu capital social até a declaração de falência e o encerramento de suas atividades.

Conforme Mario (2016), a grande maioria das seguradoras que entra em Processo de Direção Fiscal não consegue se reerguer financeiramente, conforme informações obtidas junto aos profissionais da SUSEP e liquidantes, tendo como consequência a instauração do Processo de Liquidação Extrajudicial ou solicitada sua Falência. Esse procedimento, embora não contemplado totalmente em normas, é corroborado por profissionais da área: “na prática é exatamente isso, todas (99,999%) as seguradoras que entram em Direção Fiscal são descontinuadas, quer via processo administrativo (Liquidação Extrajudicial), quer por processo judicial (Falência)”.

Ainda segundo o autor, ao se instaurar o Processo de Direção Fiscal, faz-se a comunicação do fato ao mercado via Diário Oficial da União, além de comunicação aos Corretores de Seguros, além de atualmente os fatos serem divulgados publicamente no site da SUSEP. Estes, por sua vez, desaconselham seus clientes a contratarem com a seguradora que se encontre nessa situação. Portanto, a decretação da Direção Fiscal é uma ampla sinalização ao mercado de que determinada seguradora está com viés de descontinuidade

Cabe observar que não há transferência de carteira para uma outra seguradora, já que as apólices vincendas não sofrem renovação. Caso persistam apólices em vigor quando da liquidação, elas serão canceladas, sendo o prêmio restituído ao segurado, ficando a seu critério procurar outra seguradora para firmar novo contrato (MARIO, 2016).

Diante da relevância do tema, e ao considerar que o seguro é importante tanto para pessoas como para empresas de diversos segmentos, percebe-se a necessidade de que assegadoras tenham condições de proteger os bens e direitos segurados, repondo a perda

financeira provocada por eventos não programados e, dessa forma, reestabelecendo o equilíbrio econômico (SILVA, 2017).

Os custos envolvidos em uma eventual falência de uma empresa provocam grandes perdas a uma série de pessoas e empresas. As técnicas de previsão de insolvência podem ser utilizadas como ferramentas auxiliares para os gestores a serem capazes de controlar o desempenho empresarial e administrativo da companhia antes que ela ultrapasse uma situação irrecuperável, por isso a existe a necessidade do desenvolvimento de técnicas de previsão de insolvência.

A importância da pesquisa se dá tanto em nível empresarial como acadêmico. No nível empresarial, um dos grandes desafios é a criação de uma relação mútua de lealdade entre segurado, corretor e seguradora, com informações claras, serviços de qualidade e ambiente que mantem o mercado de seguros competitivos, para isso é necessário instrumentos e ferramentas de gestão e controle que permitam a tomada de decisão rápida e com informações de qualidade. A pesquisa para a academia traz uma nova abordagem do tema de insolvência de seguradoras, e ampliando a discussão de aprendizado de máquinas na universidade.

As seguradoras prestam servidões importantes de utilidade pública a comunidade e são normalmente concentrados e fontes institucionais de fundos de investimentos necessários para o desenvolvimento do país. Sendo assim, o setor de seguros ocupa um espaço estratégico para o desenvolvimento do Brasil, hoje e no futuro, e todas as ações que auxiliem o desenvolvimento desse setor devem ser consideradas (PEDRO, 2016).

2 REVISÃO DA LITERATURA

Neste capítulo serão abordados alguns trabalhos correlatos à pesquisa, em seguida diversos conceitos necessários para o entendimento do trabalho desenvolvido.

2.1. Controle Da Solidez Econômico-Financeira Das Seguradoras

Uma atuação fundamental da SUSEP é a realização da análise da capacidade da seguradora de cumprir todos os compromissos assumidos. Ela tem que garantir que todas as seguradoras, mesmo que haja uma ocorrência de alto volume de sinistros, consigam arcar com os valores envolvidos nas indenizações. As seguradoras são obrigadas a manter fundo de reservas para arcar com esses gastos e situações inesperadas e catástrofes (MINUTO SEGUROS, 2018).

De acordo com a SUSEP (2020), O monitoramento da situação econômico-financeira e de solvência das sociedades e entidades supervisionadas pela Susep, as coordenações responsáveis por esse acompanhamento é a Coordenação de Monitoramento de Riscos (CORIS) e a Coordenação de Monitoramento de Solvência e Contabilidade (COMOC) essas realizam várias atividades. Entre essas atividades, podemos citar: (I) acompanhamento dos relatórios contábil-financeiros, regulatórios e de auditoria contábil; (II) monitoramento dos passivos, exceto as provisões técnicas, as quais são monitoradas por outra Coordenação; (III) monitoramento da situação econômico-financeira e de solvência com produção de relatórios periódicos, inclusive refletindo os efeitos relativos aos grupos e/ou conglomerados econômicos aos quais pertençam; (IV) estabelecimento e atualização de protocolo de classificação e de sinalização antecipada, objetivando também a definição da priorização e do escopo da fiscalização prudencial; (V) solicitação e monitoramento de ações regulatórias em relação ao nível de capital requerido; e (VI) proposição e instrução da aplicação do regime repressivo. Ademais, no que se refere às regulações de contabilidade e de auditoria a serem observadas pelas entidades e sociedades supervisionadas pela Susep, as seguintes atividades são realizadas: (I) proposição de normas e padrões de contabilidade, bem como de normas e procedimentos de auditoria contábil independente; (II) elaboração de pareceres e documentos de orientação sobre assuntos contábeis e de auditoria contábil; e (III) participação no desenvolvimento dos padrões internacionais de contabilidade junto aos organismos competentes no Brasil e no exterior.

Corroborando Rodrigues (2018), afirma que para ingressar e permanecer autorizada a atuar no mercado brasileiro de seguros, qualquer sociedade seguradora deve se submeter às regras de controle da solidez econômico-financeira estabelecida na regulação do setor, tais

como: capital mínimo, parâmetros de solvência, limites técnicos, constituição de provisões técnicas e manutenção de ativos garantidores das provisões técnicas.

2.1.1. Capital Mínimo Requerido (CMR)

De acordo com Silva (2017), para atuar no mercado de seguros brasileiro, as seguradoras devem atender aos requisitos de capital exigidos pela SUSEP. Esses requisitos vêm sendo alterados desde 2008, com o objetivo de convergir com os padrões internacionais de capital para as empresas seguradoras, visando à solvência do setor. Desde então, o capital exigido passou a considerar a exposição aos riscos de subscrição, de crédito, operacional e de mercado.

A resolução CNSP ° 321/2016 define os tipos de capitais e suas finalidades, conforme pode ser visto no quadro 1.

Quadro 1 - Resumo dos Capitais definidos na resolução

Nº	Tipo de Capital	Descrição
I	Capital base	Montante fixo de capital que a supervisionada deverá manter, a qualquer tempo, conforme disposto nos anexos XXIII a XXV, sendo que para as supervisionadas que operem exclusivamente em micros seguro será de 20% (vinte por cento) do valor definido no anexo XXIII.
II	Capital de risco (CR)	Montante variável de capital que a supervisionada deverá manter, a qualquer tempo, para garantir os riscos inerentes à operação.
II.a	Capital de risco de subscrição (CRsubs)	Montante variável de capital para garantir o risco de subscrição.
II.a	Capital de risco de crédito (CRcred)	Montante variável de capital para garantir o risco de crédito a que está exposta.
II.c	Capital de risco operacional (CROper)	Montante variável de capital para garantir o risco operacional a que está exposta.
II.d	Capital de risco de mercado (CRmerc)	Montante variável de capital para garantir o risco de mercado a que está exposta.
III	Capital mínimo requerido (CMR)	Capital total que a supervisionada deverá manter para operar, sendo equivalente ao maior valor entre o capital base, definido nos anexos XXIII a XXV e o capital de risco.

IV	Ativos líquidos	São todos os ativos aceitos pelo Conselho Monetário Nacional em 100% (cem por cento) na cobertura das provisões técnicas.
V	Liquidez em relação ao CR	Situação caracterizada quando a supervisionada apresentar montante de ativos líquidos, em excesso à necessidade de cobertura das provisões técnicas, superior a 20% (vinte por cento) do CR.

Fonte: 321 (2015)

Ainda segundo a autora, o cálculo do Capital de Risco (CR) contempla riscos como o de errar a precificação do seguro ou o cálculo das provisões técnicas (CRsubs), do não recebimento dos prêmios emitidos (CRcred), de perdas decorrentes de falhas nos processos operacionais (CROper) e da exposição à volatilidade do mercado (CRmerc). Assim, verifica-se uma melhor adequação dos requerimentos de capital, que considera o fluxo de operações de cada seguradora, visando à solvência das seguradoras.

2.1.2. Patrimônio Líquido Ajustado (PLA)

Segundo Costa (2005), ajustar o patrimônio líquido contábil das seguradoras foi a maneira encontrada pelos supervisores de seguros para torná-lo mais adequado às operações específicas do mercado.

A resolução CNSP ° 321/2015 relaciona a definição das deduções, conforme pode ser visto no quadro 2.

Quadro 2- Resumo das deduções definidas da resolução

Nº	Descrição
I	Valor das participações societárias em sociedades financeiras e não financeiras classificadas como investimentos de caráter permanente, nacionais ou no exterior, considerando a mais-valia e o Goodwill, bem como a redução ao valor recuperável.
II	Despesas antecipadas não relacionadas a resseguro.
III	Créditos tributários decorrentes de prejuízos fiscais de imposto de renda e bases negativas de contribuição social.
IV	Ativos intangíveis.
V	Imóveis urbanos e fundos de investimentos imobiliários com lastros em imóveis urbanos, considerando reavaliações, redução ao valor recuperável e depreciação, que excedam 14% do ativo total ajustado.

VI	Imóveis rurais e fundos de investimentos imobiliários com lastro em imóveis rurais, considerando reavaliações, redução ao valor recuperável e depreciação.
VII	Ativos diferidos.
VIII	Direitos e obrigações relativos a operações de sucursais no exterior.
IX	Obras de arte.
X	Pedras preciosas.
XI	Créditos oriundos da alienação de ativos elencados nos incisos anteriores, respeitada a regra de dedução do inciso V, em caso de alienação de imóveis urbanos.

Fonte: 321 (2015)

De acordo com a Resolução CNSP nº 321/2015, o patrimônio líquido contábil (PLA) ou patrimônio social contábil, conforme o caso, ajustado por adições e exclusões, para apurar, mais qualitativa e estritamente, os recursos disponíveis que possibilitem às supervisionadas executarem suas atividades diante de oscilações e situações adversas, devendo ser líquido de elementos incorpóreos, de ativos de elevado nível de subjetividade de valoração ou que já garantam atividades financeiras similares, e de outros ativos cuja natureza seja considerada pelo órgão regulador como impróprias para resguardar sua solvência.

2.1.3. Limite Técnico

Limite de retenção (LT) é o valor máximo de responsabilidade que as sociedades supervisionadas podem reter em cada risco isolado, determinado com base no valor dos respectivos patrimônios líquidos ajustados (SUSEP, 2017).

Conforme Silva (2017), para atuar no mercado de seguros brasileiro, a SUSEP exige que as seguradoras mantenham um limite de retenção para cada linha de risco em que a seguradora atue. Assim, de acordo com a Resolução CNSP nº 321/2015, o limite de retenção é o valor máximo de responsabilidade que as seguradoras podem reter em cada risco isolado, e deve ser calculado pelo atuário responsável em nota técnica, com base no PLA da seguradora.

De acordo com a Afonso (2009), não será fixado limite de retenção para a sociedade seguradora quando o valor dos prejuízos contabilizados for superior à soma do capital realizado mais reservas ou quando a seguradora não possuir o capital mínimo exigido. A SUSEP poderá fixar limite de retenção em valores diversos dos calculados pela sociedade seguradora, desde que devidamente justificados. A sociedade seguradora não poderá operar no ramo em que não obtiver valor positivo para seu limite de retenção.

2.1.4. Constituição de Provisões Técnicas

Segundo Fernández (2015), as provisões são formadas pela diferença dos valores pagos pelos segurados e os riscos/custos assumidos pela empresa seguradora. Tem a finalidade de arcar com compromissos futuros, na maioria das vezes, estimados estatisticamente. Esses valores são alocados no passivo da empresa e afetam diretamente a distribuição de lucros, a base para tributação e avaliação financeira da empresa.

A resolução CNSP ° 321/2015 relaciona a definição as seguintes provisões técnicas, conforme pode ser visto no quadro 3.

Quadro 3 - Resumo das provisões técnicas

Nº	Descrição
I	Provisão de Prêmios Não Ganhos (PPNG)
II	Provisão de Sinistros a Liquidar (PSL)
III	Provisão de Sinistros Ocorridos e Não Avisados (IBNR)
IV	Provisão Matemática de Benefícios a Conceder (PMBAC)
V	Provisão Matemática de Benefícios Concedidos (PMBC)
VI	Provisão Complementar de Cobertura (PCC)
VII	Provisão de Despesas Relacionadas (PDR)
VIII	Provisão de Excedentes Técnicos (PET)
IX	Provisão de Excedentes Financeiros (PEF)
X	Provisão de Resgates e Outros Valores a Regularizar (PVR)

Fonte: CNSP (2015)

2.1.5. Ativos Garantidores das Provisões Técnicas

De acordo com Afonso (2009), considerando o risco e a incerteza inerente à operação das seguradoras, é necessário estabelecer critérios para alocação dos investimentos realizados por essas empresas. Assim, a SUSEP exige que as seguradoras mantenham investimentos que excedam o valor das provisões técnicas constituídas. As provisões técnicas têm por objetivo garantir a estabilidade econômico-financeira das seguradoras, para cumprimento de obrigações futuras. Nesse sentido, a legislação obriga as seguradoras a possuir ativos no mesmo montante das provisões, como lastro e garantia para fazer frente a obrigações futuras.

Conforme a resolução nº 3.308 do Banco Central do Brasil (BCB), que define sobre a aplicação dos recursos das reservas e das provisões. Art. 2º Observadas as limitações e as

demais condições estabelecidas neste regulamento, os recursos devem ser alocados nos seguintes segmentos: I - De renda fixa; II - De renda variável e III - de imóveis.

O autor Costa (2014), relaciona os limites percentuais para cada classe de investimentos destinado as reservas, conforme pode ser visto no quadro 4.

Quadro 4 - Limite percentual permitido pelo BCB para cada tipo de investimento nas aplicações dos recursos das reservas e das provisões.

Ativos	Exemplos	Máximo
Títulos Públicos de Renda Fixa	Letras Financeiras do Tesouro (LTF), Notas do Tesouro (NTN) e Letras do Tesouro Nacional (LTN)	100%
Títulos Privados de Renda Fixa	Certificado de Deposito Bancário (CDB), Letras Hipotecarias (LH), Debentures, Caderneta de Poupança	80%
Títulos de renda Variável (por nível de Governança da B3)	Ações de Companhias do Segmento Novo Mercado	49%
	Ações de Companhias do Segmento Nível II	40%
	Ações de Companhias do Segmento Nível I e Bovespa Mais	35%
	Ações de Companhias do Segmento - Tradicionais	30%
Imóveis Urbanos	Terrenos e Edificações	8%

Fonte: Costa (2014)

De acordo com Costa (2004), em caso de insuficiência de ativos garantidores das provisões técnicas, a SUSEP pode nomear, por tempo indeterminado, às expensas da seguradora, um diretor fiscal com atribuições e vantagens que lhe forem indicadas pelo CNSP.

2.1.6. Plano de regularização de solvência

Em 2015, a SUSEP estabeleceu um plano de regularização para as seguradoras que não estejam em adequação aos requerimentos de capital, de ativos garantidores e de provisões técnicas, denominado Plano de Regularização de Solvência (PRS). O quadro 5 resume a Resolução CNSP nº 321/2015, sobre quando as seguradoras devem encaminhar o PRS à SUSEP:

Quadro 5 - Resolução CNSP nº 321/2015

Nº	Descrição
-----------	------------------

1	Na hipótese de insuficiência de PLA em relação ao CMR de até 50% (cinquenta por cento) ou de insuficiência de liquidez em relação ao CR. ou
2	Liquidez em relação ao CR situação caracterizada quando a seguradora apresentar montante de ativos líquidos, em excesso à necessidade de cobertura das provisões técnicas, superior a 20% (vinte por cento) do CR.

Fonte: 321 (2015)

A SUSEP exige que as seguradoras submetam o PRS quando for apurada insuficiência por 3 meses consecutivos ou nos meses de junho e dezembro. Ainda de acordo com Resolução CNSP nº 321/2015.

Quadro 6 - Exigências da SUSEP depois de 3 meses de insuficiência

Nº	Descrição
1	As supervisionadas estarão sujeitas ao regime especial de direção-fiscal, conforme dispõe a legislação vigente, quando a insuficiência de PLA, em relação ao CMR, for maior que 50% (cinquenta por cento) e menor ou igual a 70% (setenta por cento).
2	As supervisionadas estarão sujeitas à liquidação extrajudicial, conforme dispõe a legislação vigente, quando a insuficiência de PLA, em relação ao CMR, for superior a 70% (setenta por cento).

Fonte: 321 (2015)

2.2. Modelos De Previsão/Classificação De Insolvência

O primeiro trabalho sobre o tema de previsão de insolvência é o de Fitzpatrick (1932), no estudo foi comparado 19 empresas solventes com 19 empresas que faliram durante os anos de 1920 a 1929 nos Estados Unidos. A pesquisa teve três momentos, primeiro foi o cálculo e análise os indicadores de desempenho da empresa e em seguida classificou com bom ou ruim de acordo com um valor que foi determinado como padrão ideal e para concluir observou a tendência e o comportamento desses indicadores ao longo do tempo, o autor concluiu que os índices mais significativos na diferenciação das empresas foram o patrimônio líquido sobre o passivo e o lucro líquido sobre o patrimônio líquido.

Já Beaver (1967) utilizou um método diferente invés de compara empresas solventes e insolventes, ele caracterizava as empresas insolventes como aquelas que não cumpriam as suas obrigações financeiras até a data de vencimento. Os dados das empresas eram de 1954 e 1964.

Foram adotados e analisados 30 indicadores financeiros por 5 anos, assim analisando os índices um a um, Beaver chegou à conclusão de que o índice que melhor classificava uma empresa solvente da insolvente era o (fluxo de caixa / total de endividamento). A pesquisa concluiu que esta metodologia é capaz de prever a ocorrência de insolvência com até 5 anos de antecedência.

Outros estudos foram elaborados e o desenvolvimento de modelo de previsão de inadimplência, mas o que trouxe relevância ao tema foi Altman (1968) que propôs usar a técnica da análise discriminante multivariada para a previsão de falência das empresas. Ele estava interessado especificamente em identificar os índices com maior poder de previsão. Altman (1977) propôs o seguinte modelo Escore-Z, cujo ponto crítico é o Zero, este modelo tentou superar as deficiências das análises que eram baseadas em um único índice. Para o estudo, o autor decidiu eleger a análise discriminante múltipla (ADM), como técnica estatística por considerar a mais adequada ao problema já que a utilização da análise discriminante leva a um grupo de índices com capacidade de separar as empresas solventes das empresas insolventes e ao mesmo tempo determina o peso relativo de cada índice, sem que para determinação desses pesos prevaleçam critérios arbitrários. O modelo, no geral tem uma capacidade preditiva de 83%. A pesquisa também levantou trabalhos com a mesma temática, realizados no Brasil.

Um dos primeiros trabalhos realizados no Brasil foi o modelo desenvolvido por Elizabetsky (1976) foi feito a partir de análise realizada em um grupo de 373 empresas, sendo 99 com problemas financeiros e 274 sem problemas em suas finanças. A pesquisa iniciou com 60 índices financeiros e finalizou com apenas 38. Elizabetsky propôs um modelo cujo valor padrão ideal é de 0,5, ou seja, depois de calculados os índices e aplicado ao modelo o resultado acima de 0,5 diz que a empresa é solvente e abaixo de 0,5 a empresa é insolvente.

Por sua vez, Matias (1978), analisou um grupo de cem empresas, sendo metade delas insolventes e metade em boas condições financeiras. Neste modelo, o valor padrão ideal que separa as empresas solventes das insolventes é o 0 (zero). Kanitz (1978), se espelhou nos trabalhos de Altman e com base a análise discriminante, construiu um modelo chamado termômetro de insolvência, a partir do cálculo do Fator de insolvência (FI), o modelo conseguiu identificar e classificar empresas de acordo com a situação financeira que ela se encontra.

Guimarães e Moreira (2008), desenvolveram um modelo de previsão de insolvência baseado em indicadores contábeis com o uso da análise discriminante. A amostra utilizada são 116 empresas de capital aberto de 17 setores diferentes com dados coletados, no período de 1994 a 2003. A partir dos indicadores houve a separação de empresas solventes e insolventes.

A pesquisa concluiu que o modelo proposto tem robustez estatística e tem bom índice de acerto nas predições.

Prado (2020), propôs um modelo capaz de prever a insolvência de empresas por meio da aplicação do modelo de redes neurais artificiais. Com os resultados, obtiveram-se dois modelos um contendo apenas as variáveis do modelo tradicional e outro com as variáveis do modelo tradicional e do modelo dinâmico de análise financeira. A comparação entre estes dois modelos de análise de crédito possibilitou verificar a contribuição das variáveis do modelo dinâmico para o modelo final. Os índices que mais contribuíram para acurácia do modelo proposto foram Índice de Rentabilidade com 100% de exatidão. Índice de Estrutura de Capitais com 98,9% de acerto. e Índice do Modelo Dinâmico com 91% de precisão.

Ao longo do tempo Inúmeros modelos estatísticos foram desenvolvidos com a intenção de prever a insolvência e, conseqüentemente a falência das organizações.

2.3. Aprendizagem De Máquina

Segundo Mariana González (2018), o aprendizado de máquinas é a área da inteligência artificial relacionada à busca de um conjunto de regras e procedimentos para permitir que as máquinas possam agir e tomar decisões baseadas em dados, ao invés de serem explicitamente programadas para realizar uma determinada tarefa.

Dessa forma, ao analisar um grande volume de informações com a construção de algoritmos e aprendizagem computação, elas são capazes de identificar padrões e de tomar decisões com o auxílio de modelos. Isso torna as máquinas capazes de fazer predições por meio do processamento de dados (GONZALEZ, 2018).

Tipicamente subdivide-se *machine learning* quanto ao tipo de aprendizado, podendo ser supervisionado ou não-supervisionado. No caso de aprendizado supervisionado se tem um conjunto de dados previamente rotulados, isto é, o aprendizado se dá por exemplos. No aprendizado não-supervisionado, se possui um conjunto de dados sem nenhuma informação acerca desses dados e deseja-se encontrar padrões (HINTON, 1999).

E a partir de estudos do cérebro e do sistema nervoso humano desenvolveu-se uma tecnologia de processamento de informações conhecida como Redes Neurais Artificiais (RNA) O uso das RNA no tratamento de dados para análise de falência ainda é recente. Francisco Henrique (2003), O estudo das RNA é originário do esforço em simular o funcionamento do cérebro humano (obviamente que em uma escala muito menos complexa). Elas são

especialmente úteis para simular inteligência em atividades como reconhecimento de padrões, em atividades de associação e classificação (KLIMASAUSKAS, 1996).

Odom e Sharda (1990), pioneiros em comparar os resultados da análise discriminante (ADM), com a habilidade de predição de Redes Neurais Artificiais no risco de falência. Os pesquisadores chegaram à conclusão de que as redes neurais são mais robustas estatisticamente do que apenas a análise discriminante em amostra de tamanho reduzido. Altman (1994), também fez comparações entre as duas técnicas. Os resultados indicaram um grau equilibrado de precisão e a sugestão de combinação da análise discriminante e a aprendizagem de máquinas.

Tam e Kiang (1992), utilizando as redes neurais, análise discriminante, regressão logit e DT, em 59 bancos em crise e 59 bancos em situação normal nos Estados Unidos utilizando dados durante o período de 1985 e 1987, com base em 19 índices financeiros e comparando 19 índices financeiros-econômico. De acordo com os resultados desses estudos, os métodos com melhor capacidade preditiva das condições financeiras das empresas foram a rede neurais artificiais e a análise discriminante.

Lee e Chen (2005), avaliaram um modelo de crédito modelado com redes neurais artificiais e *Multivariate Adaptive Regression Splines* (MARS). O MARS foi utilizado para construir o modelo de avaliação de crédito. As variáveis significantes obtidas foram utilizadas no modelo utilizando rede neural. Segundo Duarte (2019), muitas pesquisas utilizam combinações de técnicas e classificadores que integram os diversos métodos, fazendo comparações entre eles. Essa combinação de classificadores, técnicas e métodos tem trazido melhor desempenho na predição, quando a hipótese nula é verdadeira e você a rejeita, comete um erro do tipo I. Quando a hipótese nula é falsa e você não a rejeita, comete um erro de tipo II.

3 METODOLOGIA

Nesta seção será descrito o processo de coleta e os processos utilizados para treinar os algoritmos de *Machine Learning*.

3.1. Coleta dos dados

Este trabalho desenvolveu um modelo de aprendizagem de máquina na linguagem de programação *Python*, conforme a regulação que permite classificar as companhias de seguros em solventes e insolventes, a partir dos principais indicadores contábeis utilizados para avaliar a saúde financeira das seguradoras disponibilizados pela SUSEP. Esses indicadores foram construídos com as informações contábeis das seguradoras divulgadas obrigatoriamente no Sistema de Estatística da SUSEP. Os dados do SES são extraídos diretamente dos Formulários de Informações Periódicas (FIP), enviados à SUSEP pelas companhias seguradoras, resseguradoras, entidades abertas de previdência privada e sociedades de capitalização, em atendimento às normas vigentes. Cabe ressaltar que as informações referentes aos meses de junho e dezembro são auditadas, de forma não sejam divergentes dos balanços publicados em jornais de grande circulação.

3.2. População e Amostra

A população foi composta pelas informações contábeis anuais e auditadas de todas as sociedades seguradoras brasileiras autorizadas pela SUSEP e disponibilizada pelo SES, esses dados estão disponíveis a partir de 2016, o *dataset* tem um total de 363652 registros no total e 314 seguradora. A amostra que foi usada na pesquisa inicia no período de 2016 e finaliza em 2019 com 193 seguradoras, 10695 registros de cada conta contábil.

3.3. Etapas da Pesquisa

Quanto aos seus objetivos, este trabalho classifica-se como descritivo e quanto à natureza, como pesquisa aplicada; com relação aos procedimentos técnicos utilizados, segue o delineamento de uma pesquisa correlacional e, portanto, de abordagem quantitativa.

Desde 2008 a SUSEP vem alterando a forma de regulação para convergir com os padrões internacionais de capital para as empresas seguradoras chegando a um consenso com a Resolução CNSP nº 321/2015, por isso a decisão de analisar os dados a partir de 2016. Quando uma seguradora tem uma relação de insuficiência de PLA em relação ao CMR de até 50%

(cinquenta por cento) ou de insuficiência de liquidez em relação ao CR, durante 3 meses seguidos, a companhia é considerada insolvente para a SUSEP e então ela é imediatamente submetida ao Plano de regularização de solvência. A partir dos índices contábeis e da variável proxy que é resultado da relação PLA e CMR, é possível classificar e separar em dois grupos de seguradoras, em solventes. Então a partir dessa separação será possível empregar a análise discriminante de modo que tenha a realização dos testes estatísticos.

Conforme Johnson & Wichern (1999), A discriminação ou separação é a primeira etapa, sendo a parte exploratória da análise e consiste em se procurar características capazes de serem utilizadas para alocar objetos em diferentes grupos previamente definidos. A classificação ou alocação pode ser definida como um conjunto de regras que serão usadas para alocar novos objetos.

De acordo com Gentile (2017) Poderíamos utilizar a totalidade dos dados históricos no processo acima, criando assim um modelo de machine learning pronto para receber novos dados e realizar suas previsões, porém desta forma não saberíamos o real desempenho deste modelo. O algoritmo poderia aprender perfeitamente a relação existentes nos dados apresentados e com isso criar um modelo que sofre de Overfitting e só descobriríamos esse problema após as previsões desastrosas geradas por este modelo.

Ainda conforme o autor, após a realização do pré-processamento, iremos separar a totalidade dos dados históricos existentes em dois grupos, sendo o primeiro responsável pelo aprendizado do modelo, e o segundo por realizar os testes. Os dados de treino são os dados que serão apresentados ao algoritmo de machine learning para criação do modelo. Estes dados costumam representar cerca de 70% da totalidade dos dados. Serão apresentados ao modelo após a sua criação, simulando previsões reais que o modelo realizará, permitindo assim que o desempenho real seja verificado. Estes dados costumam representar cerca de 30% da totalidade dos dados.

O conjunto de dados foi separados em conjuntos de treinamento e teste para avaliar capacidade preditiva dos modelos, utilizando dados anuais para treinamento dos seguintes anos 2016, 2017, 2018 e para testes dados de 2019.

O tipo de corte do conjunto de dados é transversal que consiste em uma amostra de uma unidade de análise, tomada em um determinado ponto no tempo. Após separação dos dados, será realizado um pré-processamento dados com o objetivo de padronizá-los. Desse modo, o pré-processamento escolhido será a normalização *min-max*, a qual padroniza os dados no

intervalo 0 – 1. Em seguida, utilizar-se-á a validação cruzada 10-fold como mecanismo para obter os hiper-parâmetros ajustados dos modelos candidatos.

A classificação e separação de seguradoras solventes e insolventes serão feitas a partir dos métodos de avaliação regulados pela SUSEP que serão destacados na seção 3.3.

3.3.1. Métodos de Avaliação da Solvência (Susep)

De acordo com Camila Menezes da Silva (2017), utilizam-se as demonstrações contábeis para calcular os principais indicadores que devem ser analisados para verificar o desempenho econômico-financeiro das seguradoras, segregados em indicadores de estrutura de capital, liquidez, rentabilidade e outros indicadores pertinentes à análise financeira de seguradoras. Há também outros indicadores utilizados pela SUSEP para monitoramento à distância, dentre os quais se destacam a exigência de suficiência do PLA em relação ao CMR ou o excesso de ativos líquidos em relação às provisões técnicas, conforme destacado na resolução CNSP 321/2015.

3.3.1.1. Análise Econômico-Financeira por Quocientes

Conforme Affonso Silva (2009), é o método de análise que compara “valores de naturezas heterogêneas”, relacionando itens e grupos do balanço patrimonial e da Demonstração do Resultado do Exercício – DRE, obtendo a informação sobre quantas vezes um item contém o outro. Propicia ao analista extrair tendências, comparando os índices obtidos com os “índices padrão” apurados para todo o mercado, divulgados através de revistas técnicas, órgãos oficiais e privados. Preliminarmente, serão abordados os quocientes comuns a todas as atividades, para, em seguida, focar aqueles específicos utilizados para as seguradoras.

3.3.1.1.1. Análise Financeira

A análise financeira está voltada para as relações entre disponibilidades e exigibilidades. Envolve os ativos realizáveis da empresa, tanto a curto quanto a longo prazo, e os seus exigíveis, também a curto e longo prazo. Ou seja, tudo do que a seguradora dispõe para cumprir suas obrigações, bem como o que tem a receber e o que tem a pagar. Trata do aspecto da liquidez e da solvência da sociedade (AFONSO, 2019).

No quadro 7 será exposto os indicadores financeiros capazes de mensurar a relação entre disponibilidades e exigibilidades em vários prazos.

Quadro 7 - Índices Financeiros

Índice	Sigla	Fórmula	Descrição
Liquidez Geral	LG	$\frac{\text{Realizável Total}}{\text{Exigível Total}}$	Este quociente mede a capacidade da empresa de cumprir seus compromissos perante terceiros, exigíveis com qualquer prazo
Liquidez corrente	LC	$\frac{\text{Ativo Circulante}}{\text{Passivo Circulante}}$	É o quociente utilizado para medir a capacidade da seguradora de saldar seus compromissos a curto prazo.
liquidez seca	LS	$\frac{\text{Ativo Circ.} - \text{Desp. Comerc.} - \text{Desp. Antec.}}{\text{Passivo Circulante}}$	Considera, somente, os itens de fácil conversibilidade em dinheiro dos quais a empresa dispõe para pagar cada real de dívidas vencíveis a curto prazo.
Solvência Geral	SG	$\frac{\text{Ativo Total}}{\text{Exigível Total}}$	Índice que mede a capacidade total da seguradora de cumprir a terceiros, tanto a curto quanto a longo prazo. obrigações assumidas junto
Endividamento	E	$\frac{\text{Exigível Total}}{\text{Ativo Total}}$	Representa o quanto de capital de terceiros existe para cada real aplicado no ativo total.
Garantia de Capitais de Terceiros	GCT	$\frac{\text{Patrimônio Líquido}}{\text{Exigível Total}}$	Quociente que demonstra a proporção dos capitais próprios em relação aos capitais de terceiros.
Imobilização do Capital Próprio	ICP	$\frac{\text{Ativo Permanente}}{\text{Patrimônio Líquido}}$	Este quociente apura o quanto dos capitais próprios (patrimônio líquido) está aplicado no ativo permanente da seguradora.
Independência Financeira	IF	$\frac{\text{Patrimônio Líquido}}{\text{Ativo Total}}$	A independência financeira avalia a proporção do patrimônio líquido sobre o ativo total, demonstrando o quanto de recursos próprios a empresa aplicou no seu ativo.
Capital Circulante Líquido	CCL	Ativo Circulante – Passivo Circulante	Também chamado de capital de giro próprio, representa os recursos próprios da empresa aplicados no ativo circulante.

Liquidez Operacional	LO	$\frac{\text{Créditos Operacionais com Seguros}}{\text{Débitos Operacionais com Seguros}}$	É o índice que apura o grau de liquidez entre os subgrupos “créditos operacionais com seguros”, pertencentes ao grupamento do ativo circulante, e os “débitos operacionais com seguros”, integrantes do grupamento passivo circulante.
Cobertura Vinculada	CV	$\frac{\text{Prov. Tec. N/ Comp. (+) Comp.}}{\text{Ativo Total}}$	Índice que estabelece o nível de comprometimento das aplicações da seguradora oferecidas como garantia para cobertura de suas provisões técnicas.

Fonte Afonso (2019)

3.3.1.1.2. Análise Econômica

Tem como finalidade avaliar o desempenho econômico da seguradora, considerando o capital investido e o volume monetário das receitas oriundas dos prêmios ganhos no período. Para esta análise tomamos por base a demonstração do resultado do exercício (Afonso, 2019).

No quadro 8 serão expostos os indicadores econômicos, tratando sobre os resultados de faturamento e lucro da empresa:

Quadro 8 - Índices Econômicos

Índice	Sigla	Fórmula	Conceito
Margem Bruta	MB	$\frac{\text{Resultado Bruto}}{\text{Prêmios Ganhos}}$	Representa a relação percentual entre o resultado bruto e os prêmios ganhos.
Margem Operacional	MO	$\frac{\text{Resultado das Operações de Seguros}}{\text{Prêmios Ganhos}}$	Também conhecida por margem de lucro ou lucratividade operacional, mede a relação entre o resultado líquido de prêmios produzida no período.
Margem Líquida	ML	$\frac{\text{Lucro Líquido do Exercício}}{\text{Prêmios Ganhos}}$	Mensura a relação entre o lucro líquido do exercício, depois de deduzidos a contribuição social e o imposto de renda, e os prêmios ganhos, ou seja, a receita líquida de prêmios.
Taxa de Retorno do Capital Próprio	TRC P	$\frac{\text{Lucro Líquido do Exercício}}{\text{Patrimônio Líquido Médio}}$	É o índice que mede a relação entre o lucro líquido após o imposto de renda e

Índice	Sigla	Fórmula	Conceito
			os capitais próprios (patrimônio líquido).
Retenção Própria	RP	$\frac{\text{Prêmio Retido}}{\text{Prêm. Emitidos} - \text{Prêm. Restituídos}}$	Trata-se de um quociente que indica o nível de retenção própria da seguradora sobre o montante de prêmios de sua emissão, líquido de restituições.
Retenção de Terceiros	RT	$\frac{\text{Prêm. Cosseg. Ced.} + \text{Prêm. Res. Ced.}}{\text{Prêm. Emiti.} - \text{Prêm. Rest. os}}$	A quota dos riscos não assumidos pela seguradora em cada contrato, repassados às Congêneres ou ao IRB, é expressa pelo quociente apurado na relação entre os prêmios de sua emissão e os prêmios cedidos a terceiros.
Sinistralidade	S	$\frac{\text{Sinistro Retido}}{\text{Prêmio Ganho}}$	Da relação entre sinistro retido e prêmio ganho surge o índice de sinistralidade, cuja finalidade é medir, comparativamente, o nível da despesa líquida de sinistros com a receita líquida de prêmios.
Custo de Comercialização	CC	$\frac{\text{Despesas de Comercialização}}{\text{Prêmio Ganho}}$	Quociente destinado a mensurar a proporcionalidade entre as despesas de comercialização e as receitas líquidas de prêmios
Custo Administrativo	CA	$\frac{\text{Despesas Administrativas}}{\text{Prêmio Ganho}}$	O resultado da relação das despesas administrativas com a receita líquida de prêmios é o índice que espelha o nível do custo administrativo da seguradora.
Prêmio Margem	PM	$\frac{\text{Prêmio Retido}}{\text{Patrimônio Líquido}}$	Esse indicador afere a razão prêmio margem, ou seja, verifica se a seguradora terá problemas futuros de solvência, considerando que o patrimônio líquido, embora ajustado, é, também, a base para o cálculo da margem de solvência da seguradora.
Resultado Patrimonial	RPA T	$\frac{\text{Resultado Patrimonial}}{\text{Lucro Líquido}}$	Este índice mede o quanto do lucro da seguradora é composto por resultados em coligadas, controladas e aluguéis de imóveis (este último item insignificante frente aos primeiros). Quando o

Índice	Sigla	Fórmula	Conceito
			resultado for negativo, a avaliação será de quanto do lucro da seguradora foi perdido, em função destes investimentos.
Resultado Combinado	RC	$\frac{\text{Sinis. Ret.} + \text{Desp. Comer.} + \text{Desp. Adm.}}{\text{Prêmios Ganhos}}$	É o indicador que reflete o resultado das operações básicas de seguro, estabelecendo a relação entre receitas e despesas da seguradora de natureza exclusivamente operacional.
Resultado Combinado – Ampliado	RCA	$\frac{\text{Sinis. Ret.} + \text{Desp. de C.} + \text{Desp. Adm.}}{\text{Prêmios Ganhos} + \text{Resultado Financ.}}$	Trata-se da ampliação do índice combinado, mediante a agregação do resultado financeiro aos prêmios ganhos da seguradora.

Fonte: Afonso (2019)

3.3.1.2. Margem de Solvência

Conforme Campello (2005), alerta que para a aferição da solvência é necessária a investigação de uma série de fatores, como a constituição correta de provisões técnicas, a gestão apropriada dos ativos dados em garantia às provisões técnicas, e a necessidade de capital para cobertura de perdas eventuais, entre outros.

A margem de solvência é a regra que visa à manutenção de capital próprio em volume suficiente para fazer frente a eventuais perdas decorrentes de cenários adversos (prejuízos). Por esse motivo, o capital requerido é sempre proporcional ao volume de contraprestações e eventos registrados pela operadora. Em outras palavras: um ponto percentual de prejuízo em um faturamento de R\$ 1 bilhão representa um valor muito superior a um ponto percentual de prejuízo em um faturamento de R\$ 100 mil (SERRA, 2016).

Segundo Affonso Silva (2009), as operações das seguradoras têm como objeto o risco. Por isso, a seguradora deve aumentar a quantidade de contratos sob sua responsabilidade com a finalidade de reduzir seus riscos, mediante a ampliação da mutualidade. A ideia básica é de que a solvência da empresa não está suficientemente garantida apenas com a adequação das provisões técnicas, porque há algo mais a ser considerado a margem de solvência, que se constitui num fator adicional de garantia, além das provisões técnicas.

A margem de solvência, método adotado legalmente pelo mercado segurador brasileiro, conforme determinação da resolução CNSP nº 321/2015, é calculada em relação ao PLA da seguradora, comparando-o com o Capital Mínimo Requerido pela SUSEP. A margem de

solvência (MS) corresponderá à suficiência do Patrimônio Líquido Ajustado (PLA) para cobrir montante igual ou maior do que o capital mínimo requerido (CMR) pela SUSEP para cobrir valores dos prêmios e sinistros das operações de retrocessões, líquidos de resseguro e cosseguro cedidos, anulações, restituições, cancelamentos, salvados e ressarcimentos.

Uma vez calculada a MS, Na hipótese de insuficiência de PLA em relação ao CMR de até 50% (Limite de margem) ou de insuficiência de liquidez em relação ao CR (Situação caracterizada quando a seguradora apresentar montante de ativos líquidos, em excesso à necessidade de cobertura das provisões técnicas, superior a 20%), As supervisionadas estarão sujeitas ao regime especial de direção-fiscal, conforme dispõe a legislação vigente, e se essa insuficiência passar de 70% As supervisionadas estarão sujeitas à liquidação extrajudicial, conforme a resolução CNSP nº 321/2015.

3.3.2. Métodos de Classificação Propostos

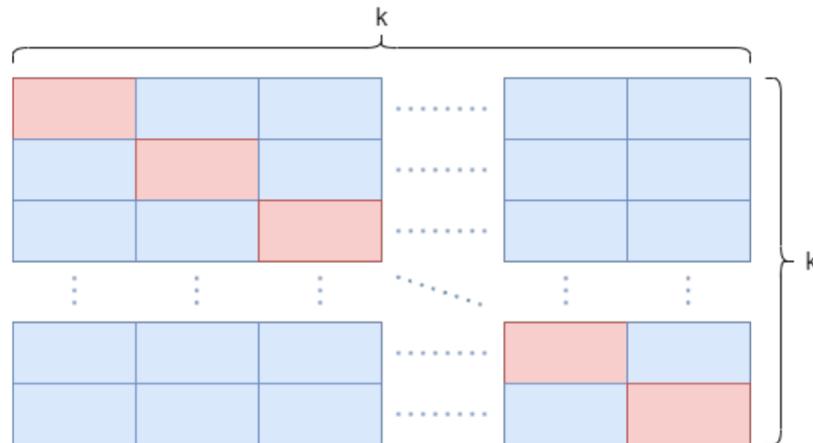
Nesta seção serão abordados algumas ferramentas e modelos de aprendizagem de máquinas que foram utilizados na fase de desenvolvimento.

3.3.2.1. Validação Cruzada

Segundo Murphy (2012), na primeira etapa de avaliação, os modelos foram treinados e testados com o mesmo conjunto de dados. Nesta etapa, é verificada a capacidade do modelo em aprender sobre a base de dados, onde uma maior acurácia implica em melhor aprendizado. Na segunda etapa de avaliação foi utilizada a validação cruzada *k-fold* com 10 folds estratificados. A padronização dos dados ocorreu a cada lote dentro da validação cruzada de forma a prevenir contaminação dos conjuntos de treino e teste. A validação cruzada é fundamental para avaliar os modelos de forma justa e realista.

A Figura 2 ilustra a divisão de dados feita validação cruzada *k-fold*. Nesta técnica, o *k* representa o número de subconjuntos em que a base de dados será dividida.

Figura 2 - Divisão de dados para treino e teste com *k-fold*



Fonte: Lopes (2022)

Desta forma, o modelo é treinado k vezes utilizando um subconjunto para teste, representado em vermelho, e $k - 1$ subconjuntos para treino, em azul. Comumente são utilizados 5 ou 10 subconjuntos. A estratificação tem a finalidade de manter a proporção entre as classes para treino e teste, quando possível. Assim, caso a base de dados possua 40% inadimplentes e 60% adimplentes, a distribuição dos dados permanecerá aleatória entre os subconjuntos, porém visando manter a razão da base original. Tanto um *k-fold* de 10 folds e a estratificação visam amenizar o viés causado por uma divisão aleatória dos dados vantajosa ou desvantajosa. (MURPHY, 2012).

3.3.3. Hiper-parâmetros dos modelos de aprendizagem de máquinas

Em machine learning (ML), a maioria dos algoritmos apresenta um ou mais parâmetros que controlam a complexidade (ou o equilíbrio entre viés e variância) do modelo ajustado. Tais parâmetros, denominados parâmetros de sintonização ou hiper-parâmetros, podem ser diretamente especificados antes do ajuste do modelo preditivo, pois não são diretamente estimados pelos dados de treinamento, ou otimizados por validação cruzada, pois não há uma fórmula analítica disponível para o cálculo do seu valor apropriado. São exemplos de hiper-parâmetros o número k de vizinhos mais próximos, utilizados pelo algoritmo KNN, e o número m de árvores de decisão presentes no algoritmo *Boosting* (KUHLM, 2013).

O quadro 9 apresenta os hiper-parâmetros dos classificadores que foram otimizados.

Quadro 9 - Resumo dos hiper-parâmetros otimizados

Algoritmo	Hiper-parâmetros Otimizados
-----------	-----------------------------

Análise Discriminante Linear	n_components: Número de componentes para redução de dimensionalidade.
Regressão Logística	C: Inverso da força de regularização, como nas máquinas de vetor de suporte, valores menores especificam uma regularização mais forte.
	fit_intercept: Especifica se uma constante (também conhecida como polarização ou interceptação) deve ser adicionada à função de decisão.
K-Nearest Neighbor	n_neighbors: Número de vizinhos a serem usados.
	weights: Função de peso usada na previsão.
Árvore de Decisão	max_depth: profundidade da árvore, relacionada ao número de divisões presentes em cada árvore.
Multilayer Perceptron	alpha: Força do prazo de regularização L2. O termo de regularização L2 é dividido pelo tamanho da amostra quando somado à perda.
	hidden_layer_sizes: refere-se ao número de unidades latentes da camada intermediária (oculta) de uma rede neural
Random Forest	n_estimators: O número de árvores na floresta.
	max_depth: A profundidade máxima da árvore.
Extreme Gradient Boosting	n_estimators: número de árvores presentes no modelo final.
	max_depth: profundidade da árvore, relacionada ao número de divisões presentes em cada árvore.
Support Vector Machine	penalty: Especifica a norma utilizada na penalização.
	max_iter: O número máximo de iterações a serem executadas.

Fonte: Adaptado de Santos (2018)

Como os hiper-parâmetros apresentam relação com a complexidade (flexibilidade) de um modelo preditivo, escolhas inadequadas para o seu valor podem resultar, por exemplo, em sobre ajuste (*overfitting*) e performance ruim do modelo em novas observações. Na prática,

uma métrica é escolhida para avaliar o erro de predição, como o EQM (em problemas de regressão) e a AUC ROC (em problemas de classificação) e, para um dado algoritmo, essa métrica é avaliada em uma lista de valores candidatos ao hiper-parâmetros por validação cruzada, com o objetivo de selecionar aquele que resulte em um modelo que minimiza o erro de predição (KUHN; JOHNSON, 2013).

3.3.4. Modelos de Aprendizagem de Máquinas Propostos

Será apresentada nesta seção os modelos de aprendizado de máquina e suas principais características.

3.3.4.1. Análise Discriminante Linear

Para Hair Jr. et al. (2005), a análise discriminante múltipla é a técnica multivariada adequada para estudar problemas em que a variável estatística (combinação linear de variáveis com pesos determinados empiricamente) é dicotômica e, portanto, não métrica. Então, ela é indicada para construir modelos de previsão de inadimplência, cujo objetivo principal é a classificação de um aspirante a crédito em um determinado grupo de:

- (a) Provável adimplente (solvente) ou
- (b) Provável inadimplente (insolvente).

De acordo com Guimarães (2008), A técnica procura, basicamente, responder se um elemento pertence ou não a uma determinada categoria. Tratando-se da análise de risco de crédito, a análise discriminante, ou especificamente a função discriminante, indicará se uma empresa pertence a um grupo de firmas solventes ou ao conjunto das insolventes.

Conforme Castro Junior (2003) A ADL é capaz de lidar com dois ou mais grupos. Esta técnica envolve a derivação de uma variável que é a combinação linear de duas ou mais variáveis independentes que melhor discriminam os grupos definidos a priori. Essa discriminação é atingida com a atribuição de pesos para as variáveis de forma a maximizar a variância entre os grupos em relação à variância dentro dos grupos. O resultado dessa combinação linear produz a chamada função discriminante, que é uma função que segue a seguinte forma

$$Z_{jk} = a + W_1X_{1k} + W_2X_{2k} + \dots + W_nX_{nk} \quad (1)$$

Onde

Z_{jk} = escore Z discriminante da função discriminante j para o objeto k

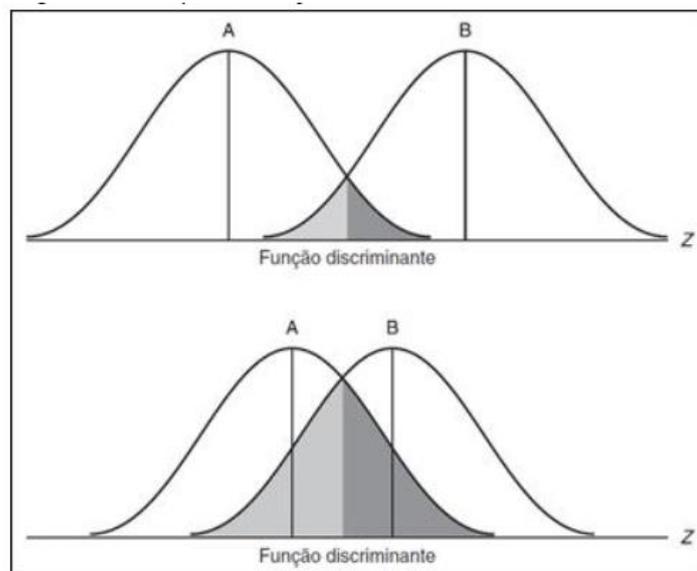
a = intercepto

W_i = peso discriminante para a variável independente i

X_{ik} = variável independente i para o objeto k

Para Castro Junior (2003), O teste para significância estatística da função discriminante é uma medida generalizada da distância entre os centroides dos grupos. Esse cômputo é feito através da comparação entre a distribuição do escores discriminantes para cada um dos grupos. O teste de significância para a função discriminante se dá através da distância entre os centroides (médias) dos grupos, comparando as distribuições dos escores discriminantes deles. Na Figura 2, apresentada abaixo mostra duas distribuições de escores discriminantes.

Figura 2 - Representação univariada de escores Z discriminantes



Fonte: Corrar (2014)

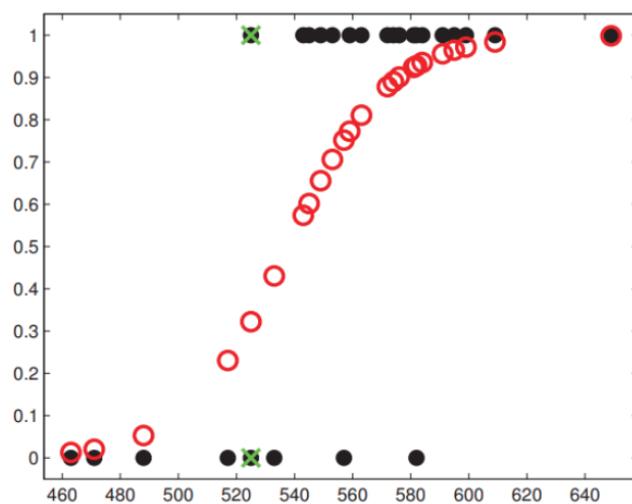
No primeiro diagrama a sobreposição das distribuições é pequena, significando que a função discriminante separa bem os grupos (A e B), enquanto no segundo a sobreposição é grande, significando que a função discriminante Z não é um bom discriminador entre o grupo A e B.

3.3.4.2. Regressão Logística

Conforme Giolo (2012), a regressão logística se constitui em um dos principais modelos utilizados quando se deseja analisar dados em que a variável resposta é binária ou dicotômica. Mesmo quando a resposta de interesse não é originalmente binária, é usual que esta seja dicotomizada de modo que a probabilidade de sucesso possa ser estimada por meio de um

modelo de regressão logística. Embora existam outros modelos para analisar dados em que a resposta é binária, a regressão logística se tornou popular por ser flexível do ponto de vista matemático, de fácil utilização e por apresentar interpretação simples de seus parâmetros. A Figura 3 exemplifica o funcionamento da predição com a Regressão Logística. Os pontos pretos são os dados e os círculos vermelhos são as probabilidades previstas. É possível observar que os círculos vermelhos seguem a tendência do formato “S” da função.

Figura 3 - Exemplo de previsão com Regressão Logística



Fonte: Murphy (2012)

Para Hosmer (2020), o objetivo ao ajustar um modelo de regressão logística é o de descrever a relação entre uma variável resposta e um conjunto de variáveis explicativas. Para isso, a regressão logística modela o valor esperado da variável resposta condicionado aos valores de p variáveis explicativas $z = (z_1, \dots, z_p)$, isto é, $E(Y | z) = \pi(z)$, (em que $E(Y | z)$ pertence ao intervalo $[0,1]$). O modelo de regressão logística é expresso por

$$\pi(z) = \frac{\exp\{\beta'z\}}{1 + \exp\{\beta'z\}} = \frac{\exp(\beta_0 + \sum_{k=1}^p \beta_k z_k)}{1 + \exp(\beta_0 + \sum_{k=1}^p \beta_k z_k)} \quad (2)$$

Em que $z = (z_1, z_2, \dots, z_p)$ denota o vetor de valores observados das variáveis explicativas, β_0 corresponde a uma constante e os componentes β_k são os p parâmetros ou coeficientes de regressão.

Conforme Gonzales (2018), para ajustar um modelo de regressão, é necessário estimar os parâmetros β_0 e β_1 do modelo. Para isso utiliza-se o método de estimação da máxima

verossimilhança. A partir dos dados da amostra, ou seja, o conjunto de observações, este método irá procurar os estimadores para o modelo de regressão logística, que são os valores de β_0 e β_1 que maximizam o logaritmo da função de máxima verossimilhança. A estimação por máxima verossimilhança, permite encontrar os estimadores dos parâmetros do modelo, que tem maior probabilidade de replicar o padrão de observações, nos dados da amostra.

$$L(\beta) = \prod_{i=1}^n \pi(z_i)^{y_i} (1 - \pi(z_i))^{1-y_i} \quad (3)$$

Seja $\beta = (\beta_0, \beta_1)$ o vetor de coeficientes, e sejam as probabilidades $P(y_i = 1|z_i) = \pi(z_i)$ e $P(y_i = 0|z_i) = 1 - \pi(z_i)$. Então, para os pares (z_i, y_i) tais que $y_i = 1$, a contribuição para a função de verossimilhança é $\pi(z_i)$, e para os pares tais que $y_i = 0$, a contribuição para a função de verossimilhança é $1 - \pi(z_i)$, onde $\pi(z_i)$ denota o valor de $\pi(z)$ avaliado em z_i , z é o preditor linear $z = x \beta$.

$$\frac{\partial \ln[L(\beta)]}{\partial \beta_0} = \sum_{i=1}^n [y_i - \pi(z_i)] \quad (4)$$

$$\frac{\partial \ln[L(\beta)]}{\partial \beta_1} = \sum_{i=1}^n z_i [y_i - \pi(z_i)] \quad (5)$$

Para Gonzales (2018), os estimadores de (β_0, β_1) , denotados por $(\widehat{\beta}_0, \widehat{\beta}_1)$, são as soluções das equações (4) e (5) quando igualadas a 0. Estes estimadores dos parâmetros, medem a taxa de variação do logit para uma unidade de variação na variável independente, isto significa que eles são de fato, a inclinação da linha de regressão entre a variável dependente y_i e a sua variável independente z_i .

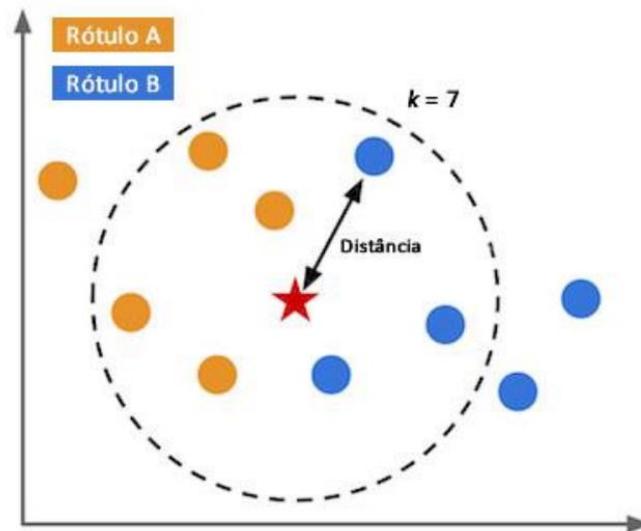
3.3.4.3. K-Nearest Neighbor (KNN)

O algoritmo KNN é baseado na aprendizagem em distância, isto significa que os dados de treinamento são armazenados e a classificação de um novo item é realizado através da comparação entre as similaridades do item a ser classificado com os dos dados de teste (BUANI et al., 2009).

O KNN, é um método de aprendizagem supervisionada, do tipo classificador, não-paramétrico, que utiliza *Lazy Learning* e possui três elementos principais: um conjunto de exemplos rotulados (por exemplo, um conjunto de registros armazenados), uma métrica de distância, e o valor de k (o número de vizinhos mais próximos) (OLIVEIRA, 2016).

Na Figura 4, é apresentado por Pacheco (2017), percebe-se a classificação de um novo item perante os 7 vizinhos mais próximos ($k=7$). Por existir 4 vizinhos do rótulo A e apenas 3 vizinhos do rótulo B a nova instância será classificada pelo rótulo A.

Figura 4 - Representação da classificação por meio do algoritmo KNN



Fonte: Oliveira (2016)

As vantagens do KNN, estão na sua simplicidade de implementação, o seu desempenho é bastante eficaz em diversas situações e áreas (engenharia, saúde, educação, entre outras), possui fácil interpretação, e é ideal para bancos de dados pequenas ou médios. Além disso, constrói diretamente a regra de decisão sem estimar as densidades condicionadas às classes, sendo uma boa escolha para problemas de classificação em que padrões próximos no espaço de características possivelmente pertencem à mesma classe (OLIVEIRA, 2016).

3.3.4.4. Árvore de decisão

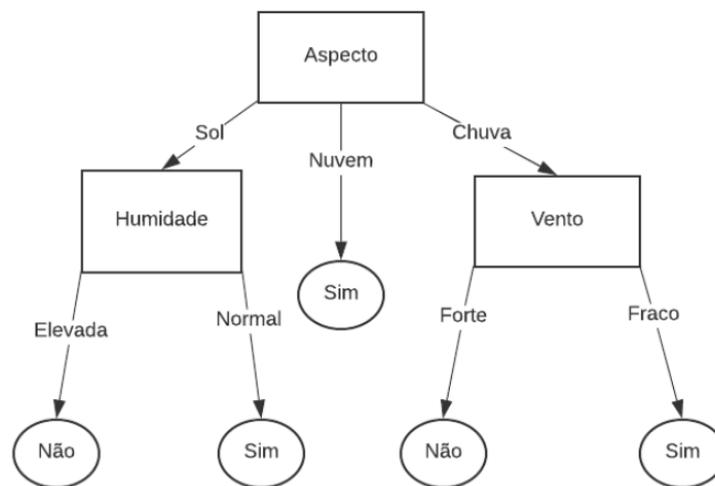
Conforme Sarro e Ferreira (2019), uma árvore de decisão é uma ferramenta de suporte à tomada de decisão que usa um gráfico no formato de árvore e demonstra visualmente as condições e as probabilidades para se chegar a resultados. Uma árvore de decisão é uma forma de visualizar as regras de negócio que levam a determinados grupos de indivíduos, construídos com base em uma variável alvo. Caracteriza de maneira satisfatória o ambiente de projetos exploratórios, sendo bastante aplicada na modelagem de decisões que englobam os seguintes estágios:

- Modelar, minuciosamente, um contexto, no qual há inúmeras decisões sequenciais e eventos que possam variar;

- Calcular o valor esperado de diferentes possibilidades de decisão, tendo como objetivo optar pela que minimiza os custos;
- Obter a distribuição de probabilidade para os pay-offs da decisão ótima;
- Através das probabilidades dos eventos, obter uma análise de sensibilidade que mostre alternativas à decisão considerada ótima.

Segundo Santos (2021), no geral as árvores de decisão podem ser explicadas como uma série de declarações *if-elses*, que quando aplicados a um registro de uma base de dados, resultam na classificação daquele registro. O mais interessante sobre o programa de árvores de decisão não é a sua construção a partir de classificação de um conjunto de treinamento, e sim a sua habilidade de aprendizado. Quando o treinamento é finalizado, é possível armazenar os dados na árvore de decisão construída a partir de exemplos com novos casos a fim de classificá-los. Na Figura 5, está um exemplo de árvore de decisão que demonstra a decisão de se jogar ou não tênis, baseado nas características representadas.

Figura 5 - Exemplo de Árvore de Decisão



Fonte: Machine Learning (MITCHELL, 1997)

Cada retângulo representa um nó que será um atributo. Esse nó que pode ser dividido para gerar outros nós. Essas divisões são os possíveis valores que ele pode ter. Para demonstrar as divisões temos as setas representando cada possível valor do atributo. Já os círculos são os nós-folhas que não possuem nós filhos e representam classe/resultado da amostra (SANTOS, 2021).

3.3.4.5. Extreme Gradient Boosting (XGBoost)

O XGBoost utiliza uma implementação diferente que busca aumentar a acurácia do algoritmo e elimina modelos menos eficientes. Ele possui diversos hiper-parâmetros que podem ser otimizados a fim de proporcionar um maior desempenho que o *Gradient Boosting* clássico (CHEN; GUESTRIN, 2016).

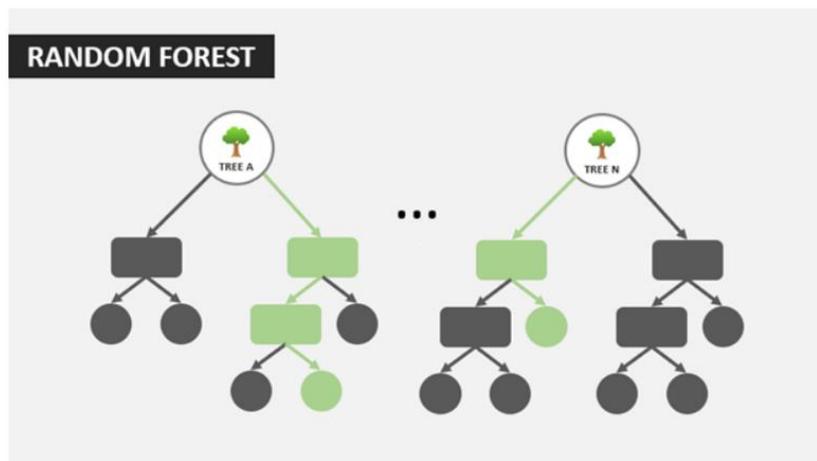
Os conjuntos são construídos a partir da combinação entre dois modelos: árvore de decisão e de impulso (*Boosting*). Esse último modelo tem a capacidade de ponderar as variáveis de entrada de acordo com a força de sua correlação com a saída do algoritmo. Ou seja, parâmetros previsores mais promissores recebem um impulso, sendo assim, preponderantes nas decisões do modelo. As árvores são adicionadas uma de cada vez ao conjunto e ajustadas para corrigir os erros de previsão feitos pelos modelos anteriores. Esse é um tipo de modelo de aprendizado de máquina de conjunto conhecido como *boosting*. (DHALIWAL; NAHID; ABBAS, 2018, p 2-3).

3.3.4.6. Random Forest

Segundo Breiman (2001), Random Forest é um algoritmo que consiste na combinação de diversos modelos de forma a aumentar a acurácia da classificação.

Conforme Gentile (2017), o Random Forest divide a amostra em diversas amostras menores, e a partir de cada uma das pequenas amostras seleciona exemplos de uma maneira específica, chamada *bootstrapping*. A partir desses exemplos, e da escolha aleatória de características (variáveis independentes), são geradas e treinadas árvores de decisão (CART). Dessa forma, o algoritmo gera diversas árvores de decisão, cada uma treinada com uma parte da amostra e considerando algumas das características disponíveis. Por isso, as árvores geradas pelo classificador são todas distintas entre si. Quando um exemplo é submetido ao classificador, a classificação é executada por todas as árvores. Com base na experiência obtida com as árvores de decisão, estas árvores são algoritmos de aprendizado supervisionado capazes de tratar dados discretos e contínuos, então é possível afirmar que as árvores podem classificar corretamente a maioria dos registros, e que diferentes modelos têm diferentes erros de classificação. Essas são duas premissas que o Random Forest usa para a classificação. o registro é submetido às diversas árvores, são verificados os resultados da classificação do exemplo pelas diferentes árvores de decisão, e o algoritmo classifica o registro com o resultado mais frequente. Na figura 6, mostra o formato de um *Random Forest* e suas árvores.

Figura 6 - Formação de uma Random Forest a partir de Árvores de Decisões



Fonte: Pessanha (2019)

O algoritmo seleciona subconjuntos de *features* e gera várias mini árvores de decisão. Basicamente o modelo possui 4 passos:

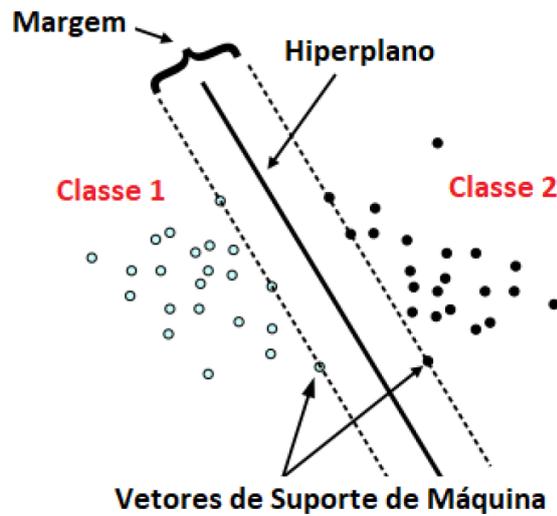
- Seleção aleatória de algumas *features*;
- Seleção da *feature* mais adequada para a posição de nó raiz;
- Geração dos nós filhos;
- Repete os passos acima até que se atinja a quantidade de árvores desejada.

Depois que o modelo é gerado, as previsões são feitas a partir de “votações”. Cada mini árvore toma uma decisão a partir dos dados apresentados. A decisão mais votada é a resposta do algoritmo.

3.3.4.7. Support Vector Machine - SVM

Segundo Lorena e Carvalho (2007), as Máquinas de Vetores de Suporte (Support Vector Machine) são algoritmos de aprendizagem de máquina que determinam os limites de decisão de classificação, resultando em uma separação das classes analisadas. A divisão é realizada pela determinação de um hiperplano, o qual busca encontrar a maior distância entre os pontos mais próximos de duas classes distintas. A distância entre o hiperplano e o ponto mais próximo com relação a cada classe é denominado margem. A figura 7, demonstra o esquema de classificação do modelo *Support Vector Machine*.

Figura 7 - Esquema de classificação através do algoritmo SVM



Fonte: Lorena e Carvalho (2007, p. 17)

Conforme Santos (2002), SVM destaca-se por pelo menos duas características possui sólida fundamentação teórica e pode alcançar alto desempenho em aplicações práticas. A teoria de aprendizagem pode identificar precisamente os fatores que devem ser considerados para a aprendizagem ser bem-sucedida e construir modelos que são bastante complexos.

Para Kwok (1998) O treinamento de SVM envolve a otimização de uma função quadrática convexa, que é um problema de Otimização Matemática. SVM envolve poucos parâmetros livres que precisam ser ajustados pelo usuário e não há uma dependência, pelos menos de uma forma explícita, na dimensão do espaço de entrada do problema, o que sugere que SVM pode ser útil em problemas com muitas entradas. Essa técnica pode ser aplicada ao Reconhecimento de Padrões (estimar funções indicadores), Regressão (estimar funções de valores reais) e Extração de Características.

3.3.4.8. Redes Neurais Artificiais

Redes neurais artificiais (RNA) são sistemas que mimetizam o comportamento e a estrutura do cérebro humano, todavia, possuindo um conjunto muito limitado de neurônios. Esses neurônios, por sua vez, processam paralelamente os dados e os propagam através de uma complexa malha de interconexão. Analogamente ao cérebro humano, as RNA têm a capacidade de interagir com o meio externo e adaptar-se a ele. Essas características conferem as RNA uma importância multidisciplinar, razão pela qual essa ferramenta vem ganhando destaque em diferentes áreas do conhecimento, tais como engenharia, matemática, física, informática etc. Particularmente, na área de transformadores, as RNAs vem sendo aplicadas na previsão de

manutenção, no controle e monitoramento da operação e, como nos casos desse trabalho e outros, na modelagem (FINOCCHIO, 2014).

De acordo Finocchio (2014), com diversas estruturas neurais e algoritmos de aprendizagem propostos por vários pesquisadores, as redes neurais possuem certas características exclusivas de sistemas biológicos. Sistemas de computação baseados em redes neurais têm a capacidade de receber ao mesmo tempo várias entradas e distribuí-las de maneira organizada. Geralmente, as informações armazenadas por uma rede neural são compartilhadas por todas as suas unidades de processamento. Essa característica contrasta com os atuais esquemas de memória dos computadores convencionais, onde a informação ficar confinada em um determinado endereço.

Conforme Morais (2010), as redes neurais artificiais foram concebidas de forma a simular em um ambiente computacional, a estrutura e a funcionalidade de redes neuronais. Suas principais vantagens são tolerância a falhas. capacidade de aprendizagem. capacidade de auto adaptação. e capacidade de resolver problemas práticos sem a necessidade da definição de lista de regras ou de modelos precisos.

Segundo Zanchettin (2015), as redes neurais artificiais oferecem melhores abordagens para problemas que requeiram reconhecimentos de padrões, classificação de padrões, associação de padrões, identificação, resistência ao ruído, aproximação de funções e aprendizado. Pelas razões descritas acredita-se que redes neurais artificiais podem ser perfeitamente aplicadas na predição de insolvência.

De acordo com Morais (2010), quando Redes Neurais Artificiais de uma só camada são utilizados os padrões de treinamento apresentados à entrada são mapeados diretamente em um conjunto de padrões de saída da rede, ou seja, não é possível a formação de uma representação interna. Neste caso, a codificação proveniente do mundo exterior deve ser suficiente para implementar esse mapeamento. Tal restrição implica que padrões de entrada similares resultem em padrões de saída similares, o que leva o sistema à incapacidade de aprender importantes mapeamentos.

Para Morais (2010), as Redes Perceptron Multicamadas (*Multilayer Perceptron - MLP*), é uma rede neural com uma ou mais camadas ocultas com um número indeterminado de neurônios. A camada oculta possui esse nome porque não é possível prever a saída desejada nas camadas intermediárias. As redes são treinadas com o algoritmo *BackPropagation*, consiste no ajuste dos pesos da rede baseado na minimização do erro médio quadrático pelo algoritmo do gradiente descendente. As RNA do tipo MLP estão constituídas de três partes: a primeira

parte é denominada camada de entrada (*input layer*), a qual é constituída de um conjunto de unidades sensoriais; a segunda parte é constituída de uma ou mais camadas escondidas (*hidden layers*); e a terceira parte é constituída de uma camada denominada camada de saída (*output layer*). Com exceção da camada de entrada, todas as outras camadas estão constituídas por neurônios, as quais implicam em um esforço computacional.

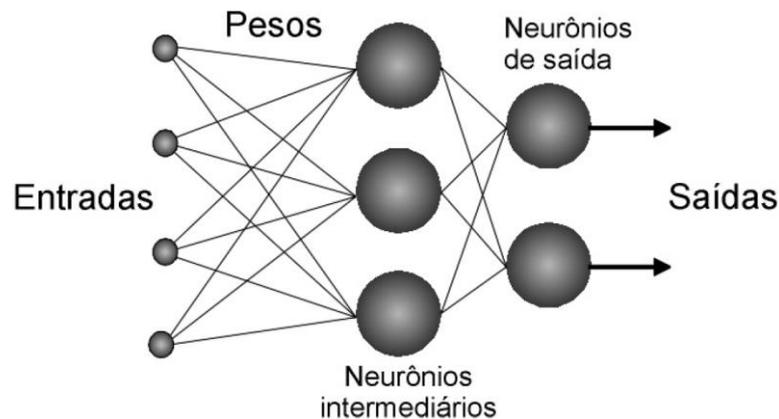
Na Figura 8, apresenta-se a arquitetura de uma rede neural MLP com uma camada de entrada constituída por unidades sensoriais, camadas intermediária e “escondidas”, e uma camada de saída constituída de neurônios, formando a arquitetura MLP.

A presente pesquisa utilizará uma rede neural com três camadas

- a) Camada de entrada Inserção dos padrões dos dados das seguradoras a rede.
- b) Camada Intermediária Local onde é feita a maior parte do processamento, através das conexões ponderadas, podem ser consideradas como extratoras ou moduladores de características.
- c) Camada de Saída Onde o resultado com padrões é apresentado.

A figura 8, mostra a arquitetura neural são tipicamente organizadas em camadas, com unidades que podem estar conectadas às unidades da camada posterior.

Figura 8 - Arquitetura de uma rede neural artificial



Fonte Oliveira (2011)

De acordo com H.G Santos, (2018), em muitos casos, alguns modelos serão equivalentes em termos de performance, logo é possível ponderar os benefícios dos diferentes algoritmos, como complexidade computacional, facilidade de implementar a função de predição estimada e interpretar. A seguir, uma sugestão de comparação de modelos para escolha do modelo final é apresentada:

1. Avaliar, inicialmente, modelos menos interpretáveis e mais flexíveis, como por exemplo, *Support Vector Machine*, *Random Forest*. Para muitos problemas, estes modelos irão apresentar resultados mais acurados;

2. Posteriormente, investigar modelos mais simples que os do passo anterior, como o MARS e as regressões penalizadas (*Lasso e Ridge*);

3. Se os modelos forem equivalentes, utilize o modelo mais simples que aproxime de modo razoável a performance de modelos mais complexos. Nesse contexto, o próximo tópico apresenta as principais características de algoritmos utilizados na etapa de aprendizado de modelos preditivos, ajustados para os problemas de classificação e regressão que compõem a tese.

3.4. Medidas de Avaliação de Eficácia

3.4.1. Matriz de confusão

Segundo Marques (2017), a matriz de confusão é uma ferramenta importante pois permite analisar de forma rápida o desempenho de cada sistema. Os valores que compõem a matriz são obtidos fornecendo os segmentos do conjunto de teste ao método de classificação e comparando sua predição com a classe correta de cada segmento.

Na matriz confusão, as linhas representam valores preditos pelo modelo e as colunas representam os valores conhecidos e previamente anotados. A diagonal principal mostra os valores corretamente classificados pelo algoritmo, sendo eles verdadeiros positivos (VP) e verdadeiros negativos (VN). As demais células da matriz contêm os valores classificados incorretamente, os falsos positivos (FP) e os falsos negativos (FN). A matriz confusão utilizada neste estudo possui apenas duas linhas e duas colunas, pois a classificação é binária, onde o indivíduo será classificado como possível bom pagador ou possível mau pagador. Em sistemas com mais de uma classe, o tamanho da matriz é maior (LOPES, 2022). No quadro 10, está a tabela que permite a visualização do desempenho de um algoritmo de classificação.

Quadro 10 - Matriz de Confusão

	Positivos	Negativos
Positivos	Verdadeiro Positivo (VP)	Falso Positivo (FP)
Negativos	Falso Negativo (FN)	Verdadeiro Negativo (VN)

Fonte: Adaptado de Marques (2017)

Após a comparação os valores são classificados em quatro possíveis opções:

1. Verdadeiro Positivo (VP) segmentos que pertencem a classe positiva e foram classificados como positivos. No caso desse trabalho, são segmentos que contêm ovos do esquistossomo e foram classificados corretamente.

2. Falso Positivo (FP) segmentos que pertencem a classe negativa e foram classificados como positivos.

3. Falso Negativo (FN) segmentos que pertencem a classe positiva e foram classificados como negativos. São segmentos que contêm ovos, porém foram erroneamente classificados como negativos.

4. Verdadeiro Negativo (VN) segmentos que pertencem a classe negativa e foram corretamente classificados como negativos.

Uma matriz de confusão fornece informações para o cálculo de outras métricas importantes. Algumas delas estão dispostas na seção 3.4.2.

3.4.2. Métricas de avaliação

As métricas descritas abaixo são popularmente usadas na comparação de modelos de classificação. Cada uma delas busca avaliar um aspecto diferente do modelo. No quadro 11, estão as principais medidas de qualidade da previsão de um modelo de classificação.

Quadro 11 - Métricas de avaliação para modelos de classificação

Modelo	Fórmula	Conceito
Precisão	$\frac{\text{Verdadeiro Positivo}}{\text{Verdadeiro Positivo} + \text{Falso Positivo}}$	Quantidade de segmentos que, dentre todos os segmentos classificados como positivos, são pertencem a classe positiva.
Recall	$\frac{\text{Verdadeiro Positivo}}{\text{Verdadeiro Positivo} + \text{Falso Negativo}}$	Quantidade de segmentos positivos que foram classificados corretamente.
Acurácia	$\frac{\text{Verdadeiro Positivo} + \text{Verdadeiro Negativo}}{\text{Total}}$	Quantidade de segmentos tanto da classe positiva da negativa que foram classificados corretamente.

Fonte: Marques (2017)

O uso de métricas apropriadas em um problema de classificação é crucial para o sucesso de um projeto de Machine Learning. A escolha da métrica deve levar em conta o objetivo do modelo no mundo real, o custo de cada tipo de erro, o quão interpretável ela deve ser, dentre

outros fatores. É sempre importante ter uma visão crítica da avaliação de um modelo, e questionar se a escolha de métricas de fato reflete a definição de valor que a sua aplicação de Machine Learning necessita (KUNUMI, 2022).

4 RESULTADOS

Neste capítulo serão discutidos a análise dos dados, a validação dos modelos e os resultados obtidos neste trabalho

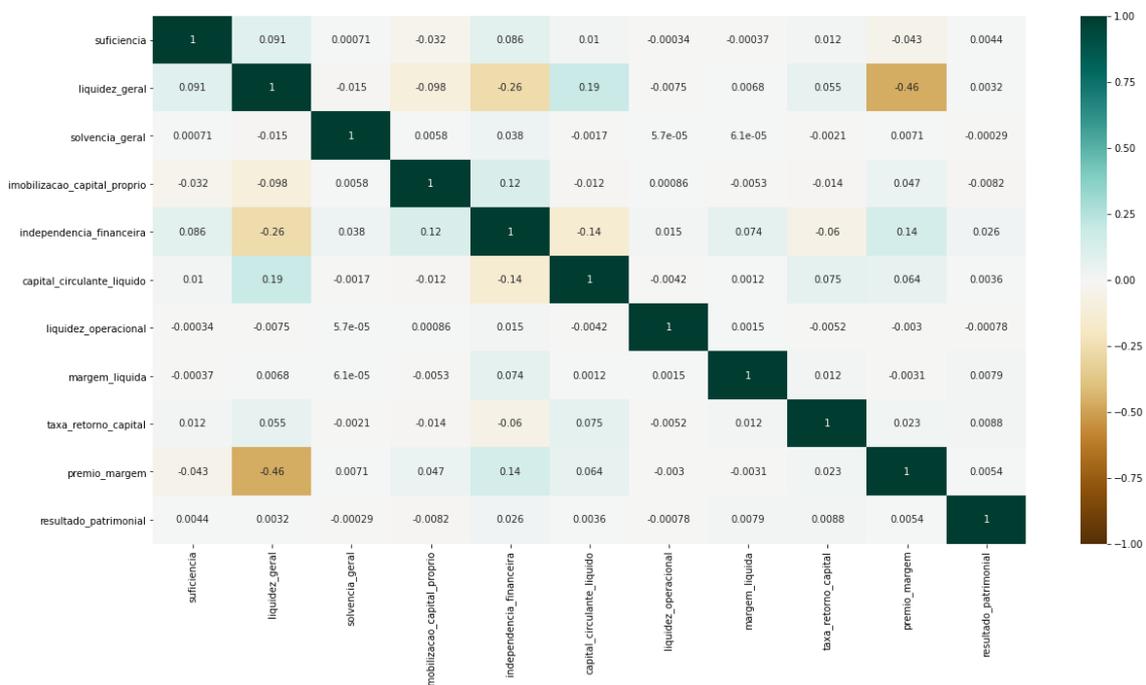
4.1. Análise Exploratória dos Dados

Após o pré-processamento dos dados com a devida transformação dos dados brutos em formatos eficientes, pode-se fazer estatísticas descritivas, gráficos, histogramas, testes de covariância e então a partir dessa exploração entender os dados, descobrir como estão distribuídos, calcular as médias, mediana, dentre outras métricas da estatística descritiva.

O conjunto de informações disponibilizado é referente a cada seguradora no Brasil e as contas contábeis requeridas pela SUSEP por meio da legislação. O *dataset* tem um total de 363652 registros no total e 314 seguradoras, O conjunto de dados depois do pré-processamento têm 193 seguradoras, 10695 registros de cada conta contábil em seu respectivo mês e 24 variáveis com a criação dos indicadores.

Depois de aplicado a análise de correlação nas variáveis brutas foram removidas aquelas que poderiam causar problemas de multicolinearidade como a Liquidez Corrente, Endividamento, Garantia de Capitais de Terceiros, Margem Operacional, Sinistralidade, Custo Administrativo. A figura 9 é intuitiva no sentido de que cada quadrado apresenta o coeficiente de correlação de Spearman do par de variáveis mostrando o nível de semelhança e dependência entre elas, tal que quanto mais verde e escuro, mais próximo de 1.

Figura 9 - Representação da matriz das variáveis explicativas que permaneceram na análise



Fonte: Dados da pesquisa (2022)

A figura 10 mostra as estatísticas descritivas de cada indicador contábil ao longo dos anos analisados na pesquisa.

Figura 10 - Estatística descritiva dos dados

2016	Observações	Média	Desvio	Mínimo	1º Quartil	Mediana	3º Quartil	Máximo
Suficiência	2064	-51,647	597,538	-23429,492	0,005	0,023	0,112	20,003
Liquidez Geral	2064	0,356	0,297	0,000	0,000	0,456	0,626	0,810
Liquidez Corrente	2064	0,008	0,022	0,007	0,007	0,007	0,007	1,000
Solvência Geral	2064	0,004	0,022	0,004	0,004	0,004	0,004	1,000
Endividamento	2064	0,349	0,128	0,000	0,308	0,387	0,432	0,557
Garantia de Capitais de Terceiros	2064	0,004	0,022	0,004	0,004	0,004	0,004	1,000
Imobilização de Capital de Próprio	2064	0,005	0,047	0,000	0,000	0,000	0,000	0,790
Independência Financeira	2064	0,505	0,147	0,335	0,408	0,479	0,564	0,997
Capital Circulante Líquido	2064	0,036	0,068	0,013	0,026	0,027	0,027	0,661
Liquidez Operacional	2064	0,893	0,000	0,893	0,893	0,893	0,893	0,894
Margem Operacional	2064	0,711	0,035	0,000	0,712	0,712	0,712	1,000
Margem Líquida	2064	0,699	0,035	0,000	0,700	0,700	0,700	1,000
Taxa de Retorno de Capital	2064	0,496	0,022	0,000	0,494	0,495	0,499	0,559
Sinistralidade	2064	0,155	0,027	0,135	0,153	0,154	0,154	1,000
Custo Administrativo	2064	0,950	0,042	0,000	0,953	0,953	0,953	0,953
Prêmio Margem	2064	0,890	0,068	0,396	0,867	0,920	0,936	1,000
Resultado Patrimonial	2064	0,823	0,002	0,816	0,823	0,823	0,823	0,892

2017	Observações	Média	Desvio	Mínimo	1º Quartil	Mediana	3º Quartil	Máximo
Suficiência	2044	-58,748	702,668	-24848,579	0,005	0,022	0,091	56,737
Liquidez Geral	2044	0,373	0,298	0,000	0,000	0,477	0,632	0,867
Liquidez Corrente	2044	0,007	0,000	0,007	0,007	0,007	0,007	0,010
Solvência Geral	2044	0,004	0,000	0,004	0,004	0,004	0,004	0,005
Endividamento	2044	0,357	0,122	0,000	0,321	0,386	0,435	0,585
Garantia de Capitais de Terceiros	2044	0,004	0,000	0,004	0,004	0,004	0,004	0,005
Imobilização de Capital de Próprio	2044	0,006	0,063	0,000	0,000	0,000	0,000	1,000
Independência Financeira	2044	0,505	0,142	0,317	0,415	0,485	0,548	0,997
Capital Circulante Líquido	2044	0,038	0,082	0,004	0,026	0,027	0,027	0,746
Liquidez Operacional	2044	0,893	0,020	0,000	0,893	0,893	0,893	0,894
Margem Operacional	2044	0,713	0,006	0,706	0,712	0,712	0,712	0,870
Margem Líquida	2044	0,701	0,007	0,693	0,700	0,700	0,700	0,868
Taxa de Retorno de Capital	2044	0,497	0,017	0,199	0,494	0,496	0,500	1,000
Sinistralidade	2044	0,155	0,021	0,128	0,153	0,154	0,154	0,620
Custo Administrativo	2044	0,953	0,001	0,929	0,953	0,953	0,953	0,962
Prêmio Margem	2044	0,896	0,061	0,416	0,878	0,922	0,936	0,972
Resultado Patrimonial	2044	0,822	0,019	0,000	0,823	0,823	0,823	0,914

2019	Observações	Média	Desvio	Mínimo	1º Quartil	Mediana	3º Quartil	Máximo
Suficiência	2045	-40,273	533,904	-18562,921	0,005	0,025	0,105	4,727
Liquidez Geral	2045	0,383	0,306	0,000	0,000	0,469	0,657	0,886
Liquidez Corrente	2045	0,007	0,000	0,007	0,007	0,007	0,007	0,010
Solvência Geral	2045	0,004	0,000	0,004	0,004	0,004	0,004	0,005
Endividamento	2045	0,361	0,134	0,000	0,328	0,395	0,441	0,952
Garantia de Capitais de Terceiros	2045	0,004	0,000	0,004	0,004	0,004	0,004	0,005
Imobilização de Capital de Próprio	2045	0,002	0,017	0,000	0,000	0,000	0,000	0,491
Independência Financeira	2045	0,505	0,155	0,000	0,410	0,476	0,542	0,997
Capital Circulante Líquido	2045	0,041	0,102	0,000	0,026	0,027	0,027	0,936
Liquidez Operacional	2045	0,894	0,001	0,893	0,893	0,893	0,893	0,955
Margem Operacional	2045	0,712	0,003	0,678	0,712	0,712	0,712	0,838
Margem Líquida	2045	0,700	0,003	0,664	0,700	0,700	0,700	0,812
Taxa de Retorno de Capital	2045	0,498	0,008	0,450	0,494	0,496	0,500	0,556
Sinistralidade	2045	0,153	0,005	0,000	0,153	0,154	0,154	0,160
Custo Administrativo	2045	0,952	0,014	0,458	0,953	0,953	0,953	1,000
Prêmio Margem	2045	0,896	0,081	0,000	0,890	0,925	0,936	0,967
Resultado Patrimonial	2045	0,823	0,005	0,736	0,823	0,823	0,823	0,952

2018	Observações	Média	Desvio	Mínimo	1º Quartil	Mediana	3º Quartil	Máximo
Suficiência	2038	-43,408	578,323	-21504,413	0,005	0,022	0,093	2,059
Liquidez Geral	2038	0,380	0,304	0,000	0,000	0,467	0,649	0,872
Liquidez Corrente	2038	0,007	0,000	0,007	0,007	0,007	0,007	0,013
Solvência Geral	2038	0,004	0,000	0,004	0,004	0,004	0,004	0,012
Endividamento	2038	0,360	0,127	0,000	0,325	0,392	0,436	0,757
Garantia de Capitais de Terceiros	2038	0,004	0,000	0,004	0,004	0,004	0,004	0,012
Imobilização de Capital de Próprio	2038	0,001	0,007	0,000	0,000	0,000	0,000	0,102
Independência Financeira	2038	0,500	0,144	0,335	0,406	0,479	0,535	1,000
Capital Circulante Líquido	2038	0,039	0,092	0,001	0,026	0,027	0,027	0,818
Liquidez Operacional	2038	0,893	0,000	0,893	0,893	0,893	0,893	0,895
Margem Operacional	2038	0,713	0,004	0,711	0,712	0,712	0,712	0,799
Margem Líquida	2038	0,700	0,004	0,699	0,700	0,700	0,700	0,788
Taxa de Retorno de Capital	2038	0,497	0,009	0,435	0,494	0,495	0,499	0,559
Sinistralidade	2038	0,153	0,001	0,123	0,153	0,154	0,154	0,161
Custo Administrativo	2038	0,953	0,000	0,944	0,953	0,953	0,953	0,953
Prêmio Margem	2038	0,896	0,065	0,370	0,882	0,924	0,936	0,988
Resultado Patrimonial	2038	0,823	0,006	0,703	0,823	0,823	0,823	1,000

2019	Observações	Média	Desvio	Mínimo	1º Quartil	Mediana	3º Quartil	Máximo
Suficiência	2045	-40,273	533,904	-18562,921	0,005	0,025	0,105	4,727
Liquidez Geral	2045	0,383	0,306	0,000	0,000	0,469	0,657	0,886
Liquidez Corrente	2045	0,007	0,000	0,007	0,007	0,007	0,007	0,010
Solvência Geral	2045	0,004	0,000	0,004	0,004	0,004	0,004	0,005
Endividamento	2045	0,361	0,134	0,000	0,328	0,395	0,441	0,952
Garantia de Capitais de Terceiros	2045	0,004	0,000	0,004	0,004	0,004	0,004	0,005
Imobilização de Capital de Próprio	2045	0,002	0,017	0,000	0,000	0,000	0,000	0,491
Independência Financeira	2045	0,505	0,155	0,000	0,410	0,476	0,542	0,997
Capital Circulante Líquido	2045	0,041	0,102	0,000	0,026	0,027	0,027	0,936
Liquidez Operacional	2045	0,894	0,001	0,893	0,893	0,893	0,893	0,955
Margem Operacional	2045	0,712	0,003	0,678	0,712	0,712	0,712	0,838
Margem Líquida	2045	0,700	0,003	0,664	0,700	0,700	0,700	0,812
Taxa de Retorno de Capital	2045	0,498	0,008	0,450	0,494	0,496	0,500	0,556
Sinistralidade	2045	0,153	0,005	0,000	0,153	0,154	0,154	0,160
Custo Administrativo	2045	0,952	0,014	0,458	0,953	0,953	0,953	1,000
Prêmio Margem	2045	0,896	0,081	0,000	0,890	0,925	0,936	0,967
Resultado Patrimonial	2045	0,823	0,005	0,736	0,823	0,823	0,823	0,952

2020	Observações	Média	Desvio	Mínimo	1º Quartil	Mediana	3º Quartil	Máximo
Suficiência	2018	-42,555	629,519	-21860,910	0,005	0,024	0,095	1,419
Liquidez Geral	2018	0,327	0,294	0,000	0,000	0,385	0,593	1,000
Liquidez Corrente	2018	0,007	0,001	0,000	0,007	0,007	0,007	0,037
Solvência Geral	2018	0,004	0,001	0,000	0,004	0,004	0,004	0,064
Endividamento	2018	0,342	0,143	0,000	0,270	0,388	0,440	1,000
Garantia de Capitais de Terceiros	2018	0,004	0,001	0,000	0,004	0,004	0,004	0,064
Imobilização de Capital de Próprio	2018	0,003	0,026	0,000	0,000	0,000	0,000	0,922
Independência Financeira	2018	0,532	0,174	0,335	0,412	0,483	0,619	0,997
Capital Circulante Líquido	2018	0,042	0,109	0,000	0,026	0,027	0,027	0,997
Liquidez Operacional	2018	0,894	0,003	0,893	0,893	0,893	0,893	1,000
Margem Operacional	2018	0,712	0,000	0,708	0,712	0,712	0,712	0,715
Margem Líquida	2018	0,700	0,000	0,695	0,700	0,700	0,700	0,702
Taxa de Retorno de Capital	2018	0,496	0,012	0,271	0,494	0,495	0,499	0,553
Sinistralidade	2018	0,154	0,005	0,034	0,153	0,154	0,154	0,207
Custo Administrativo	2018	0,953	0,001	0,930	0,953	0,953	0,953	0,955
Prêmio Margem	2018	0,902	0,067	0,235	0,897	0,932	0,936	0,994
Resultado Patrimonial	2018	0,823	0,003	0,780	0,823	0,823	0,823	0,898

Fonte: Dados da pesquisa (2022)

Analisando a estatística descritiva dos dados ao longo dos anos, percebe-se que a Liquidez Geral das demonstrações com a média de 0,363, um valor considerado baixo quando o ideal é 1, outro ponto importante é que as seguradoras mais capacidade de honrar as suas obrigações de curto e longo prazo estão localizadas no 3º quartil ou seja, em 25% do conjunto de dados a média desse valor o período analisado é de 0,6312. Já o Endividamento médio é de 0,2949, e esses valores se alteram muito pouco com o passar dos anos, o que demonstra que as seguradoras dependem mais do capital próprio do que de terceiros. A Sinistralidade média de 0,1539 valida que a despesa líquida é menor que a receita líquida dos prêmios e com baixo desvio da média no período. A partir dessas informações posso inferir que as seguradoras nacionais em sua maioria são sociedades líquidas, pouco endividadas e capazes de honrar os pagamentos futuros.

A seguir será tratado a frequência de seguradoras classificadas como solvente e insolvente pelos indicadores da SUSEP.

Tabela 1 - Frequência de demonstrações financeiras classificadas em solventes e insolventes com base no Capital Mínimo Requerido e no Patrimônio Líquido Ajustado das seguradoras selecionadas

Ano	Total	Solventes		Insolventes	
		Quantidade	Frequência	Quantidade	Frequência
2016	2064	1768	85,66%	296	14,34%
2017	2044	1780	87,08%	264	12,92%
2018	2038	1799	88,27%	239	11,73%
2019	2045	1818	88,90%	227	11,10%
2020	2018	1819	90,14%	199	9,86%

Fonte: Dados da pesquisa (2022)

Com base na tabela 2 de frequência, houve uma melhora no índice de insolvência no período de análise. Em 2016 e 2017, os anos com maior quantidade de demonstrações insolventes que pode ser explicado pela queda de atividade econômica. Em 2018 e 2019, apesar de juros mais baixos, a recuperação econômica com que o lucro volte a se recuperar. Em 2020, a pandemia é o principal responsável para a queda dos lucros, como, por exemplo, resultado da diminuição da receita das seguradoras em alguns ramos ao mesmo tempo que buscam conforto não só para atravessar o momento como para fazer frente aos efeitos esperados com o fim de estímulos, como o auxílio emergencial, e uma eventual segunda onda da covid-19 no País (RONCARATI, 2020).

4.2. Validação dos modelos

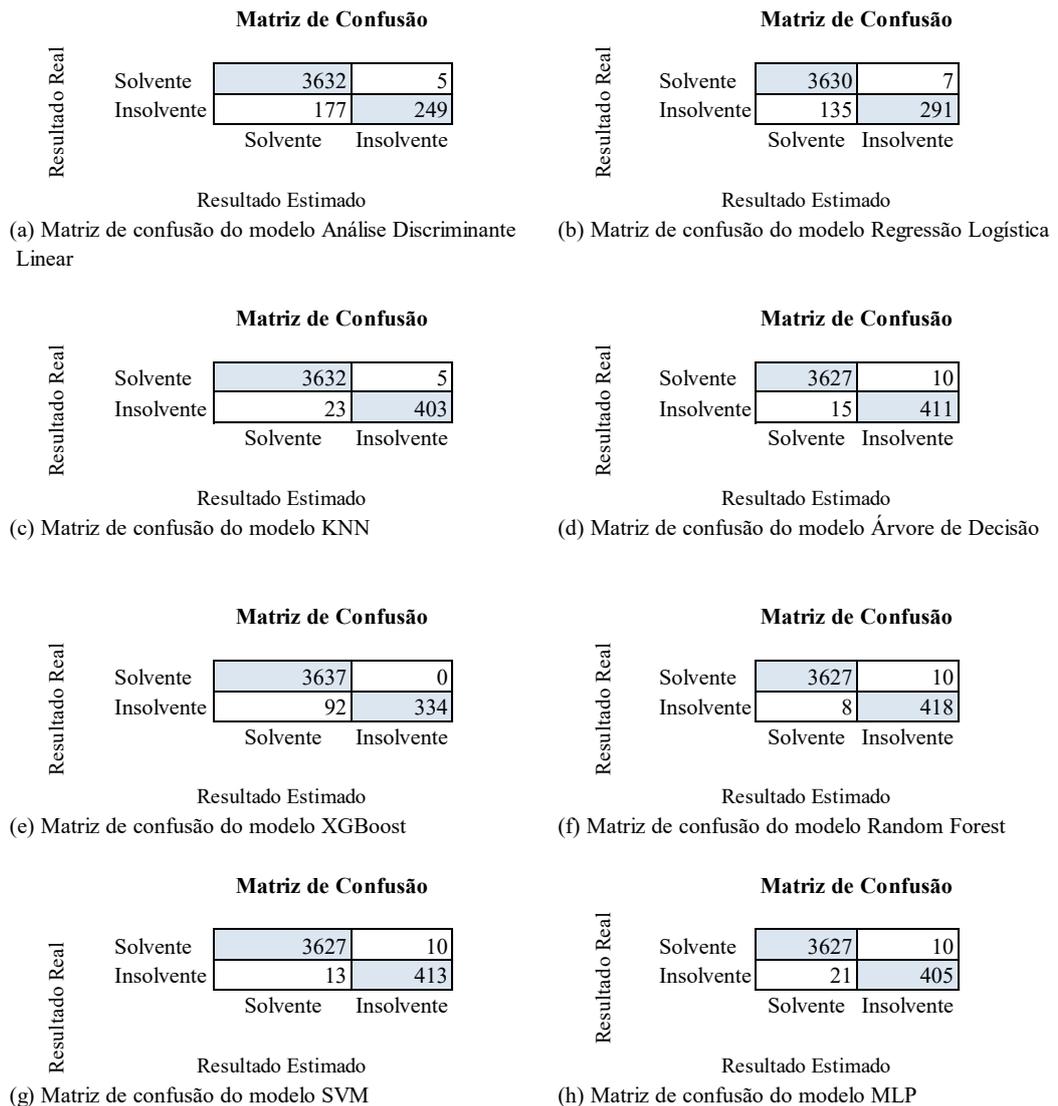
Nesta seção, serão expostos e discutidos os resultados obtidos a partir da avaliação usando a validação cruzada. A validação cruzada tem o propósito de treinar o modelo com um subconjunto de dados da base e testar seu aprendizado em um subconjunto de dados previamente desconhecidos. Ao usar novos dados para serem classificados, esta técnica avalia a generalização do modelo, ela estima como seria o desempenho do modelo em uma situação real com hiper-parâmetros otimizados visando o aumento do potencial do modelo ser um bom previsor do conjunto de dados.

Os modelos foram treinados utilizando o conjunto de dados de treinamento dos anos de 2016 a 2018. Após concluído o treino de todas as técnicas propostas, utilizou-se as amostras de dados dos anos de 2019 e 2020 para a validação dos modelos gerados a partir da acurácia, matriz de confusão das predições dos modelos de aprendizagem de máquina, a fim de descobrir a qualidade da predição desses modelos e encontrar o melhor entre eles. As métricas apresentadas são Matriz de confusão, *Accuracy*, *Precision*, *Recall* e *F1-Score*.

4.2.1. Matrizes de Confusão

A figura 11 a seguir, demonstra a matriz de confusão para cada modelo na validação cruzada. As linhas representam os dados reais e as colunas, predições realizadas pelos modelos. A classe de solventes e insolventes representa os clientes inadimplentes.

Figura 11 - Matriz de confusão de todos os algoritmos otimizados



Fonte: Dados da pesquisa (2022)

O algoritmo com melhor desempenho na matriz de confusão foi o *Random Forest*, pois ele errou menos comparado com outros modelos:

- Previu as demonstrações solventes 3627 vezes corretamente;
- Previu as demonstrações insolventes 418 vezes corretamente;
- Previu as demonstrações solventes 10 vezes incorretamente;
- Previu as demonstrações solventes 8 vezes incorretamente.

4.2.2. Resultado dos modelos com validação cruzada

A Quadro 12, estão os resultados obtidos através da validação cruzada no conjunto de dados das seguradoras. Na validação cruzada, o método *Random Forest* teve a maior Accuracy 99,56%, ou seja, foi o modelo com resultados de previsão mais próximo da referência real,

seguido pelo SVM com 99,43% e Árvore de Decisão com 99,41%. O modelo com menor acurácia foi a Análise Discriminante com 95,52%, seguido por Regressão Logística com 96,51% e XGBoost 97,74%. A diferença entre os modelos com maior e menor acurácia foi de 4,04%.

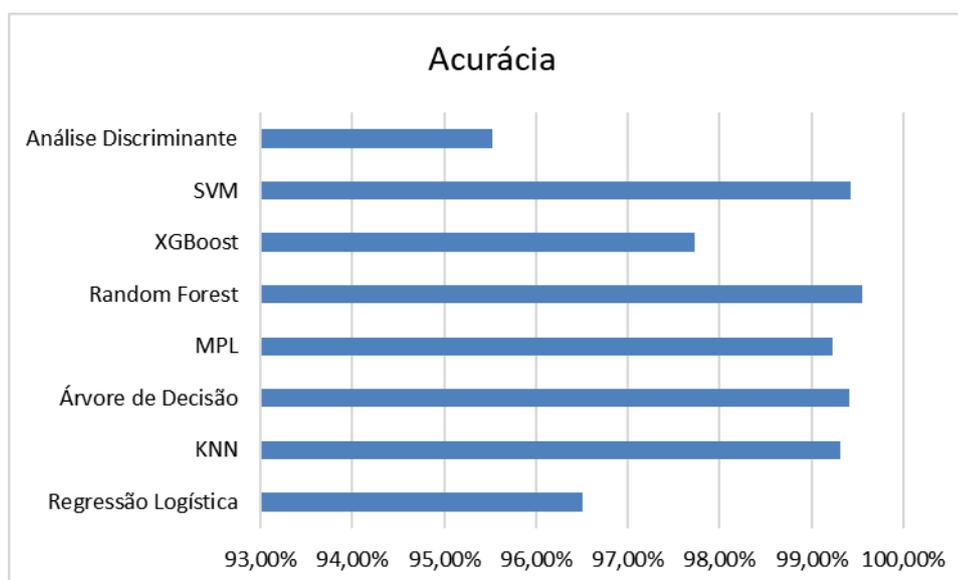
Quadro 12 - Resultados dos modelos com validação cruzada

Modelo	Accuracy	Recall	Precision	F1-Score
Regressão Logística	96,51%	68,00%	98,00%	80,00%
KNN	99,31%	95,00%	99,00%	97,00%
Árvore de Decisão	99,41%	97,00%	98,00%	97,00%
MPL	99,24%	95,00%	98,00%	96,00%
Random Forest	99,56%	98,00%	98,00%	98,00%
XGBoost	97,74%	78,00%	100,00%	88,00%
SVM	99,43%	97,00%	98,00%	97,00%
Análise Discriminante	95,52%	58,00%	98,00%	73,00%

Fonte: Dados da pesquisa (2022)

Na figura 12, mostra os resultados dos classificadores obtidos por meio do uso dos algoritmos descritos obtiveram taxas de exatidão próximas a 98,34%. Na figura 10, é estabelecida uma comparação gráfica entre estes valores.

Figura 12 - Comparando os classificadores



Fonte: Dados da pesquisa (2022)

A técnica de Análise Discriminante, Regressão Logística e XGBoost produziram os menores desempenhos de classificação dentre todas as outras técnicas. Já o modelo baseado em

Redes Neurais Artificiais como o *Multilayer Perceptron* mostrou-se robusto no quesito acurácia. Por fim, algoritmos de classificação embasados em árvores de decisão obtiveram as maiores taxas de desempenho entre as outras técnicas. Com a aplicação do algoritmo havendo um aumento notável na taxa de verdadeiros positivos, assim como uma redução nos falsos negativos, características tipicamente conservadoras.

5 CONCLUSÃO

A partir das demonstrações contábeis extraídas do banco de dados do Sistema de Estatística (SES) da SUSEP, foi desenvolvido modelos de aprendizagem de máquina que permitem classificar adequadamente as seguradoras brasileiras em solventes e insolventes.

O estudo foi produzido em programação *Python*, com o conjunto de dados disponibilizado pelo SES onde foram extraídos as demonstrações contábeis anuais auditadas do período de 2016 a 2019. Com os dados brutos foi feito o pré-processamento das informações para iniciar a análise exploratória dos dados, calcular os indicadores e obter o índice de solvência de acordo com as normas da SUSEP e então classificar as demonstrações contábeis em solvente ou insolvente. Aplicando os dados aos classificadores escolhidos foi possível encontrar os modelos que são mais adequados a esse tipo de informação, assim como os que tiveram menos capacidade de se adaptar ao conjunto de dados, esse ranqueamento foi feito a partir da matriz de confusão e dos indicadores de desempenho do modelo, principalmente a acurácia.

A experimentação por meio de outros tipos de algoritmos de classificação permitiu interpretações diversas, de acordo com as características próprias a cada técnica estudada. Baseado nas métricas de avaliação apresentadas, o algoritmo de *Random Forest* foi o que conseguiu prever 4045 das 4063 previsões realizadas, obteve 99,56% de *Accuracy*, *Recall* de 98%, *Precision* de 98%. Sendo assim, confirmando ser o modelo que conseguiu trazer mais resultados satisfatórios. Já o algoritmo que menos se adaptou ao conjunto de dados foi a Análise Discriminante que conseguiu prever corretamente 3881 das 4063 previsões realizadas, obteve 95,52% de *Accuracy*, *Recall* de 58%, *Precision* de 98% e *F1-Score* de 73%.

Com esses classificadores é possível identificar e reproduzir padrões de um grande volume de dados para compor um modelo de aprendizado. A consequência disso são resultados mais precisos e praticamente em tempo real, contribuindo para tomadas de decisão mais certas, diminuindo as intervenções humanas e os riscos de falhas de interpretação.

Algumas dificuldades foram enfrentadas no trabalho, como as limitações do computador interferindo diretamente na quantidade de hiper-parâmetros a serem otimizados, reduzindo a quantidade de combinações que aumentem a qualidade da previsão.

Para trabalhos futuros é recomendado aplicar técnicas mais robustas de otimização de hiper-parâmetros com a menor limitação da capacidade da máquina possível, uma dessas técnicas é o uso da otimização Bayesiana, que pode aumentar a *Accuracy* dos modelos, ao definir parâmetros com melhor desempenho para cada classificador em cada base de dados.

REFERENCIAS

- AFONSO. Solvência das seguradoras / Affonso Silva. – Rio de Janeiro **Funenseg**, 2009. 72
- ALTMAN, E. I. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. **The Journal of Finance**, v. 23, n.4, p. 589-609, sep. 1968.
- ALENCAR. Saiba o que acontece após uma seguradora decretar falência. **CQSC**. 2018. Disponível em: <https://www.cqcs.com.br/noticia/saiba-o-que-acontece-apos-seguradora-decretar-falencia/>. Acesso em: 30 mai. 2022.
- ANALITIK, Valora. A penetração de seguros na América Latina ainda é baixa, mas tem potencial para crescer. **Alsum**. Set. 2019. Disponível em: <https://alsum.co/pb/noticias/penetracion-de-seguros-en-latinoamerica-aun-es-baja-pero-cuenta-con-potencial-para-crecer>. Acesso em 19 set. 2020.
- BEAVER, W. H. Financial Ratios as Predictors of Failure, Journal of Accounting Research - Empirical research in accounting selected studies. **Journal of Accounting Research 4** (Suppl.), p. 71–111, 1967.
- BREIMAN, L. **Random Forests**. Machine Learning, 45, pp. 5-32. 2001
- Brito, Renata. **Por proteção contra imprevistos, brasileiros buscam por seguros residenciais**. Disponível em: <https://www.segs.com.br/seguros/179064-por-protecao-contrainprevistos-brasileiros-buscam-por-seguros-residenciais>. Acesso em: 02 Jan 2022.
- CAMARGO, Y. B. L. **Abordagem Linguística na Classificação Automática de Textos em português**. 2007. Rio de Janeiro, 2007. Disponível em: <http://www.pee.ufrj.br/index.php/pt/producao-academica/dissertacoes-de-mestrado/2007-1/2007062502-2007062502/file>. Acesso em: 24 mar.2021.
- CASTRO JUNIOR, Francisco Henrique Figueiredo de. **Previsão de insolvência de empresas brasileiras usando análise discriminante, regressão logística e redes neurais**. 2003. São Paulo, 2003. doi10.11606/D.12.2003.tde-16092004-121634. Disponível em: <https://www.teses.usp.br/teses/disponiveis/12/12139/tde-16092004-121634/pt-br.php>. Acesso em: 24 nov. 2020.
- COSTA, Jorge Andrade. **Contabilidade de seguros as experiências no Brasil e no Mercosul em comparação com as normas propostas pelo IASB**. Rio de Janeiro **Funenseg**, 2005. Disponível em: <http://www.docvirt.no-ip.com/funensegbiblioteca/home.htm>. Acesso em: 02 set. 2022.
- CORREA, R. F. **Categorização de Documentos Utilizando Redes Neurais Análise**

Comparativa Com Técnicas Não-Conexionistas. 2002. 136f. Recife, 2002. Disponível em: <http://pee.ufrj.br/teses/textocompleto/2007122001.pdf>. Acesso em: 10 jul. 2021.

DALCIN, Ronaldo. Razões que norteiam o crescimento do mercado de seguros em 2021. **Revista Apólice**, 2022. Disponível em: <https://www.revistaapolice.com.br/2021/03/razoes-que-norteiam-o-crescimento-do-mercado-de-seguros-em-2021/>. Acesso em: 30 mai. 2022.

DHALIWAL, Sukhpreet Singh; NAHID, Abdullah-Al; ABBAS, Robert. Effective intrusion detection system using XGBoost. **Information**, v. 9, n. 7, p. 149, 2018.

DUARTE, Denize Lemos. **Previsão de insolvência corporativa no Brasil considerando a regionalidade.** 2020. Uberlândia, 2020. Disponível em: <https://repositorio.ufu.br/handle/123456789/28779?mode=full>. Acesso em: 23 ago. 2021.

ELISABETSKY, Roberto. **Um modelo matemático para decisões de crédito no Banco Comercial.** Dissertação (Mestrado) – Instituto Politécnico da Universidade de São Paulo, 1976. Disponível em: <https://periodicos.ifpb.edu.br/index.php/rgo/article/view/4958>. Acesso em: 14 fev. 2020.

FERNANDEZ, Raquel. Entenda as provisões técnicas atuariais. **Atuarial e Tecnologia.** Rio de Janeiro, 2015. Disponível em: <http://www.raquelfernandez.com.br/provisoes-tecnicas-atuariais/> acesso em: 20 nov. 2020.

FITZPATRICK P.J. **A Comparison of ratios of successful industrial enterprises with those of failed firms.** Certified Public Accountant, October, November and December, p.598-605, p. 656-662, p. 727-731, 1932

GANDRA, Alana. Mercado de seguros cresceu 12,1% em 2019. **Agência Brasil**, Rio de Janeiro, 13/02/2020. Disponível em: <https://agenciabrasil.ebc.com.br/economia/noticia/2020-02/mercado-de-seguros-cresceu-121-em-2019>. Acesso em: 04 de setembro de 2020.

GENTILLE, Nicolas Leonardo. **Aprendizado de máquina e caracterização de aterosclerose subclínica um estudo de caso.** São Paulo. 2017. Disponível em: https://www.ipt.br/pos_graduacao_ipt/solucoes/dissertacoes/982-aprendizado_de_maquina_e_caracterizacao_de_aterosclerose_subclinica:_um_estudo_de_caso.htm. Acesso em: 09 abr. 2021.

GIOLO, S. R. **Introdução à Análise de Dados Categóricos com Aplicações.** Paraná, 2012. Disponível em: http://people.ufpr.br/~giolo/CE073/Material/Suely_Giolo.pdf. Acesso em: 28 abr. 2013.

GONZALES, Mariana. O que é machine learning e quais são as 7 etapas do aprendizado das máquinas. **Idblog**. 2018. Disponível em: <https://blog.idwall.co/o-que-e-machine-learning/>. Acesso em: 19 nov. 2020.

GUIMARAES, Ailton. MOREIRA, Tito Belchior Silva. Previsão de insolvência um modelo baseado em índices contábeis com utilização da análise discriminante. **Rev. econ. contemp.**, Rio de Janeiro, v. 12, n. 1, p. 151-178, 2008. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-98482008000100006&lng=pt&nrm=iso. Acesso em: 23 nov. 2020.

HAYKIN, S. Neural Networks Comprehensive Foundation. 2nd. ed. **Upper Saddle River, NJ, USA Prentice Hall PTR**, 1998. ISBN 0132733501.

HAIR, J.F., ANDERSON, R.E., TATHAM, R.L. & BLACK, W.C. Multivariate Data Analysis with readings. 5ª edição. **New Jersey Prentice Hall**, 1998. HAWLEY, D.D.,

HOSMER, D. W. LEMESHOW, S. **Applied Logistic Regression**. New York John Wiley & Sons, Inc., 2000.

JOHNSON, J.D. & RAINA, D. Artificial Neural Systems

KANITZ, S. C. Como prever falências, São Paulo **Mc Graw-Hill do Brasil**, 1978.

KLIMASAUSKAS, C.C. Applying Neural Networks. In TRIPPI, R.R. & TURBAN, E. (Eds.), Neural Networks in Finance and Investing using artificial intelligence to improve real-world performance, Edição **Revista, Estados Unidos Irwin**, pp.45-69, 1996.

LEMES, Neslon H.T. Redes Neurais o Perceptron. 2005. **Unicor**. Disponível em: <http://www.usuarios.unicor.br/nhtlemes/cpu/perceptron.pdf>. Acesso em: 1 dez. 2020.

LEWIS, D. D. Evaluating Text Categorization. In **WORKSHOP ON SPEECH AND NATURAL LANGUAGE**, 1991, Massachusetts. Massachusetts, Amherst MA University of Massachusetts, 1991. p. 312-318. Disponível em: <http://clair.si.umich.edu/clair/anthology/query.cgi?type=Paper&id=H91-1061>. Acesso em: 01 nov. 2009.

LOPES. Renato de Santana. **Comparação de métodos de aprendizado de máquina para análise de risco de crédito**. Serra, 2022. Disponível em: https://repositorio.ifes.edu.br/bitstream/handle/123456789/1755/2021__Renato__TCC__Credit__Scoring.pdf?sequence=2&isAllowed=y. Acesso em: 06 jun. 2022.

MATIAS, Alberto Borges. **Contribuição às técnicas de análise financeira um modelo de concessão de crédito**. 1978. 106 f. São Paulo. Disponível em:

<https://silo.tips/download/contribuicao-as-tecnicas-de-analise-financeira-um-modelo-de-concessao-de-credito>. Acesso em: 1 mar. 2021.

MARQUES, Victor Gutemberg Oliveira. **Avaliação do desempenho das redes neurais convulsionais na detecção de ovos de esquistossomose**. 2017. Disponível em: https://www.cin.ufpe.br/~tg/2017-1/vgom_tg.pdf. Acesso em: 5 set. de 2021.

MITCHELL, T. M. Machine Learning. New York: **McGraw-Hill**, 1997. ISBN 978-0-07-042807-2.

MORAIS, Emerson Cordeiro. **Reconhecimento de Padrões e Redes Neurais Artificiais em Predição de Estruturas Secundárias de Proteínas**. 2010. Disponível em: <https://www.cos.ufrj.br/uploadfile/1277729485.pdf>. Acesso em 27 out. 2021.

NEVES. Samuel Antônio das. **Técnicas de aprendizado de máquina aplicadas a classificação da qualidade de pavimentos asfálticos utilizando smartphones**. Ouro Preto. 2018. Disponível em: https://www.monografias.ufop.br/bitstream/35400000/799/1/MONOGRAFIA_T%C3%A9cnicasAprendizadoM%C3%A1quina.pdf. Acesso em 07 jun. 2022.

ODOM, M. SHARDA, R. “A neural network model for bankruptcy prediction”. In Proceedings of the international joint conference on neural networks, Vol. 2, **IEEE Press**, Alamitos, CA, 1990, p. 163–168.

ONUSIC L.M., Casa Nova, S.P.C, Almeida, F.C. **Modelos de previsão de insolvência utilizando a análise por envoltória de dados aplicação a empresas brasileiras**. Revista de Administração Contemporânea 11 (SPE2), 77-97. Disponível em: <https://www.scielo.br/j/rac/a/bZNtQcwsDBvJ6PCfLpFSSMk/?lang=pt#:~:text=Os%20modelos%20de%20previs%C3%A3o%20de,previs%C3%A3o%20de%20insolv%C3%Aancia%20de%20empresas>. Acesso em: 7 nov. de 2021.

PEDRO, Paulo Sergio. **ABC em companhias de seguros**. Porto Alegre. 2006. Disponível em: <https://lume.ufrgs.br/bitstream/handle/10183/7887/000559814.pdf?sequence=1&isAllowed=y>. Acesso em: 30 jun. 2022.

PESSANHA, Cinthia. Random Forest: como funciona um dos algoritmos mais populares de ML. **Medium**. 2019. Disponível em: <https://medium.com/cinthiabpessanha/random-forest-como-funciona-um-dos-algoritmos-mais-populares-de-ml-cc1b8a58b3b4>. Acesso em: 5 jul. de 2021.

PRADO, José Willer do et al. **Previsão Da Insolvência Empresarial Utilizando Redes Neurais Artificiais**. *Revista Gestão e Desenvolvimento*, Novo Hamburgo, v. 17, n. 2, p. 136-162, may 2020. ISSN 2446-6875. Disponível em:

<https://periodicos.feevale.br/seer/index.php/revistagestaoedesenvolvimento/article/view/1777>. Acesso em: 24 nov. 2020.

PROCENGE. A tomada de decisão baseada no planejamento estratégico e resultados de longo prazo. **PROCENGE**. Disponível em <https://procenge.com.br/blog/tomada-de-decisao-planejamento-estrategico/>. Acesso em: 30 maio de 22.

MARIO, Poueri do Carmo, Aversari Martins, Vinícius , Martins, Eliseu , Lopes Cardoso, Ricardo Insolvência, Regulação e Valor de Liquidação no Mercado de Seguros Brasileiro. **Contabilidade Vista & Revista**. 2006, 17(4), 73-95. ISSN: 0103-734X. Disponível em: <https://www.redalyc.org/articulo.oa?id=197014773005>. Acesso em: 28 abr. 2022.

OLIVEIRA, Jeymisson. Redes Neurais Artificiais. **Computação UFCG**. Disponível em: <http://www.dsc.ufcg.edu.br/~pet/jornal/setembro2011/materias/informatica.html>. Acesso em: 01 dez. 2020.

RASCHKA, S. Python Machine Learning. 2. Ed. Birminigham: **Packt Publishing Ltd**, 2017

RODRIGUES, Adriano. **Gerenciamento da informação contábil e regulação evidências no mercado brasileiro de seguros**. 2008. Tese (Doutorado) - São Paulo FEA-USP, 2008. 2008. Disponível em <http://www.teses.usp.br/teses/disponiveis/12/12136/tde-17032008-123836/>. Acesso em: 2020-10-26.

RONCARATI. Evolução do lucro das seguradoras em 2020. 2021. **Revista Roncarati**. Disponível em: <https://www.editoraroncarati.com.br/v2/Artigos-e-Noticias/Artigos-e-Noticias/Evolucao-do-Lucro-das-Seguradoras-em-2020.html#:~:text=Em%202018%20e%202019%2C%20apesar,lucro%20volte%20a%20se%20recuperar.&text=Em%202020%2C%20a%20pandemia%20%C3%A9,das%20seguradoras%20em%20alguns%20ramos>. Acesso em: 4 jan. 2021.

RUSSELL, S.; NORVIG, P. Artificial intelligence: a modern approach. 1995.

SANTOS. Helen Geremias do. **Comparação da performance de algoritmos de machine learning para a análise preditiva em saúde pública e medicina**. 2018. São Paulo, 2018. Disponível em: <https://teses.usp.br/teses/disponiveis/6/6141/tde-09102018-132826/pt-br.php>. Acesso em: 19 nov. 2019.

SANTOS, Eulanda Miranda dos. **Teoria e aplicação de support vector machines à aprendizagem e reconhecimento de objetos baseado na aparência**. 2002. 121 f. Paraíba, Brasil, 2002. Disponível em: <http://dspace.sti.ufcg.edu.br:8080/jspui/handle/riufcg/4265>. Acesso em: 20 nov. 2020.

SANTOS, Karolayne Rodrigues Ramos. **Classificação com arvores de decisão em paralelo**. Urutaí, 2021. Disponível em: https://repositorio.ifgoiano.edu.br/bitstream/prefix/1723/1/Monografia_Karolayne.pdf. Acesso em: 10 jun. 2021.

SARRO, L.R.A, FERREIRA, T. S. **Árvore de Decisão Aplicada à Exploração e Produção de Petróleo - Teoria e Prática, com Estudo de Caso**. Rio de Janeiro, 2019. Disponível em: <http://repositorio.poli.ufrj.br/monografias/monopoli10029313.pdf>; Acesso em: 09 jun. de 2022.

SERRA, C. Entendendo a regulação prudencial no mercado de saúde suplementar. **Cadernos de Seguro**. Disponível em: <http://cadernosdeseguro.funenseg.org.br/pdf/cadernos186artigo-cesarserra.pdf>>. Acesso em: 8 mai. 2020.

SCARPEL, R. A. Previsão de insolvência de empresas utilizando support vector machine. **Revista de Economia e Administração**, v. 7, p. 281-295, 2008.

SILVA, Camila Menezes da. **O uso dos indicadores contábeis para análise de solvência das seguradoras brasileiras**. São Paulo, 2017. Disponível em: <https://tede2.pucsp.br/handle/handle/19887>. Acesso em: 21 nov. 2020.

SUNO. Patrimônio Líquido entenda o que é e como analisar esse indicador. 2019. **SUNO**. Disponível em: <https://www.suno.com.br/artigos/patrimonio-liquido/#~:text=Patrim%C3%B4nio%20l%C3%ADquido%20negativo,-Existe%20uma%20situa%C3%A7%C3%A3o&text=Isso%20acontece%20quando%20os%20valores,maior%20do%20que%20o%20ativo.>> Acesso em: 30 de maio de 2022.

SUSEP. **Limite de Retenção - Orientações da Susep ao Mercado**, 2017. Disponível em: http://www.susep.gov.br/setores-susep/cgsoa/copra/arquivos-copra/orientacoes/Limite%20de%20Retencao%20-%20Versao%20-%202017_08.pdf> . Acesso em: 30 de maio de 2022.

ZHU, Y., XIE, C., WANG, G.-J., YAN, X.-G. Comparison of individual, ensemble, and integrated ensemble machine learning methods to predict China's SME credit risk in supply chain finance. **Neural Computing and Applications**, 28S1, 41–50, 2016.