

## **PREDIÇÃO DE INSOLVÊNCIA DE EMPRESAS POR MEIO DA INTELIGÊNCIA ARTIFICIAL – TÉCNICAS DE APRENDIZADO DE MÁQUINA**

Rafael de Almeida, FIPECAFI/SP, [rafael.informat@gmail.com](mailto:rafael.informat@gmail.com)

George Sales, FIPECAFI/SP, [george.sales@fipecafi.org](mailto:george.sales@fipecafi.org)

Rodolfo Nunes, IFPR/PR, [rodolfonunes@usp.br](mailto:rodolfonunes@usp.br)

### **RESUMO**

Qualquer empresa que possua um processo de tomada de decisões em relação a uma concessão de crédito mais ágil e com um modelo confiável, servindo de apoio para as análises de insolvência das organizações, terá vantagens no mercado. Assim, o objetivo deste artigo foi desenvolver e testar um modelo de predição de insolvência com base no aprendizado em máquina. A metodologia do artigo enquadra-se por meio de uma pesquisa bibliográfica, descritiva e quantitativa. Para a amostra da pesquisa, foram selecionadas 24 empresas de capital aberto do setor de comércio, e a base de dados utilizada para a coleta foi a Economatica, no período entre 2014 e 2019. O modelo preditivo desenvolvido está estruturado baseado em 16 variáveis preditoras, nas quais existem informações financeiras e não financeiras, e a verificação do modelo se faz por meio de uma análise ferramental *CheckUp+*. As evidências mostram que o modelo embasado na aprendizagem de máquina conseguiu alcançar bons patamares de acerto e previsibilidade nas verificações. Os valores são semelhantes aos modelos de previsão tradicionais, baseados em análise discriminante e regressão. O trabalho é relevante, pois contribui para a teoria e a prática, com o aprimoramento dos modelos de previsões de risco, e fornece suporte para as instituições financeiras tomarem decisões mais assertivas sobre concessões de crédito.

**Palavras-chave:** Insolvência. Aprendizado de Máquina. Análise de Risco. Ferramenta *CheckUp+*.

**Data de recebimento:** 23/07/2022

**Data do aceite de publicação:** 20/04/2023

**Data da publicação:** 30/04/2023

## PREDICTION OF BUSINESS INSOLVENCE THROUGH ARTIFICIAL INTELLIGENCE – MACHINE LEARNING TECHNIQUES

### ABSTRACT

Any company that has a more agile decision-making process in relation to granting credit and with a reliable model serving as support for the insolvency analysis of organizations will have advantages in the market. Thus, the purpose of this article was to develop and test a machine learning-based insolvency prediction model. The methodology of the article fits through a bibliographical, descriptive and quantitative research. For the research sample, 24 publicly traded companies in the commerce sector were selected, and the database used for the collection was Economatica in the period between 2014 and 2019. The predictive model developed is structured based on 16 predictor variables, where there are financial and non-financial information, and the model verification is done through a CheckUp+ tool analysis. Evidence shows that the model based on machine learning managed to reach good levels of accuracy and predictability in the checks, the values are similar to traditional prediction models based on discriminant analysis and regression. The work is relevant, as it contributes to theory and practice with the improvement of risk forecast models, and provides support for financial institutions to make more assertive decisions on credit concessions.

**Keywords:** Insolvency. Machine Learning. Risk Analysis. CheckUp+ Tool.

### 1 INTRODUÇÃO

Os investidores e o mercado financeiro de crédito têm-se apoiado fortemente em índices financeiros como uma ferramenta para a avaliação da perspectiva de sucesso de uma empresa, enquanto estudiosos de finanças e contabilidade também os utilizam para prever falências (Souza, 2010). Conhecer melhor um cliente e saber se ele honrará seus compromissos são informações cruciais no momento de uma tomada de decisão, principalmente no tocante à concessão de crédito (Selau; Ribeiro, 2009).

Todas as instituições financeiras e empresas estão vulneráveis às diversas formas de riscos, seja durante o ciclo operacional ou relacionadas às atividades de investimento e financiamento. Compreender os riscos é um pilar central sobre como fazer a gestão correta dentro das diretrizes de cada organização e o seu mercado de atuação (Pinto, 2021).

Uma análise da situação financeira sempre se fará necessária. Para tanto, é importante que, em benefício da própria análise financeira, possamos criar instrumentos que sejam capazes não somente de criar índices financeiros, mas também de aprender por meio do aprendizado indutivo, criando, assim, generalizações, para serem submetidas a questionamentos e testes, para que se conheça a capacidade de nos possibilitar a fazer previsões.

Quaisquer modelos preditivos, dos mais simples ou complexos, que avaliam o crédito são fundamentais para conhecer a situação financeira de uma organização, já que uma classificação equivocada pode acarretar prejuízos. Exemplo: classificar um cliente ruim como bom, ou, então, prejudicar os ganhos, classificando um cliente bom como ruim (Gouvêa; Gonçalves; Mantovani, 2012).

Nos últimos anos, temos visto termos como aprendizado de máquina (machine learning), redes neurais, lógica *fuzzy*, etc., que tem sua força sendo expandida em praticamente

todas as áreas de conhecimento e mercados, e o mercado que mais tem se aproveitado delas ou desse conjunto de tecnologias é o financeiro. Segundo Facelli, Lorena, Gama, Almeida e Carvalho (2011), o aprendizado de máquina é um método de análise de dados que automatiza o desenvolvimento de modelos analíticos, fazendo, assim, a interpretação dos dados e reconhecendo padrões, de modo a ter seu aprendizado efetivo e reconhecer esses padrões em situações futuras, ou seja, predizer algo a partir de dados existentes.

O mercado financeiro dispõe de uma quantidade enorme de dados, sendo gerado cada vez mais rápido devido aos sistemas e às interações no mercado. Temos visto a utilização desse conjunto de tecnologias em diversos ecossistemas da área financeira, tais como aprovação de crédito, gerenciamento de ativos de avaliação, bem como gerenciamento de risco, mitigação das fraudes, carteira de investimento e previsão de *tradings* de curto prazo (de Souza; dos Santos, 2013).

Diante do cenário econômico do país, as empresas devem buscar ferramentas para analisar qual caminho estão seguindo e, se necessário, ajustar e controlar melhor seu custo, bem como seus investimentos. Nesse sentido, essa pesquisa irá tentar responder à seguinte questão: como é possível aplicar e prever a situação financeira futura de uma empresa, utilizando modelos de previsão de insolvência, por meio da abordagem de inteligência artificial com a técnica de aprendizado de máquina?

Assim, o cerne do artigo é prever a situação das empresas utilizando métricas de assertividade para demonstrar acurácia ou grau de precisão dessa inteligência de aprendizado em máquina. Com isso, tentaremos identificar se é possível treinar uma inteligência com base no histórico de classificação de empresas como solventes ou insolventes, para que a inteligência aprenda os padrões e faça inferências, considerando-se uma nova situação de uma empresa que não fazia parte, mediante dados históricos previamente utilizados para treinamento.

Esse estudo irá servir de contribuição para a área financeira, utilizando a abordagem tecnológica, bem como gerar uma ferramenta capaz de ajudar profissionais no mercado financeiro, contábil e jurídico, para diagnosticar a saúde financeira das empresas. Também pretende servir como suporte acadêmico para a criação de um novo produto financeiro, visando à mitigação de risco de insolvência, fornecimento de crédito, mercado, entre outras. Assim, pretende gerar uma ferramenta útil acessível para pequenas e médias empresas brasileiras.

## **2 REFERENCIAL TEÓRICO**

### **2.1 Definição de Solvência e Insolvência das Empresas**

A abordagem do presente estudo se dará na análise e no diagnóstico da saúde financeira das empresas. Para tanto, é necessário definir o que seria o conceito de uma empresa em estado de solvência no entendimento do mercado. Para Soares e Rebouças (2015), solvência, em finanças, bem como na contabilidade, é estado devedor (empresa) que possui seu ativo maior do seu passivo. Ou seja, a empresa tem capacidade de cumprir seus compromissos com recursos que constituem seu patrimônio ou seu ativo. 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 (Horta; Borges; Carvalho; Alves, 2011; Gozer; Gimenes; Menezes; de Albuquerque, 2014).

Conceitua-se a insolvência como a incapacidade de cumprir com suas obrigações econômicas à medida que os vencimentos ocorrem (Minatto; Borba, 2022). Assim, para Nunes e Sales (2020), o pressuposto básico da insolvência é a incapacidade de honrar as dívidas, ou seja, a insolvência de fluxo ocorre pela falta de capacidade de honrar as obrigações, de modo

que a geração de receita das operações não paga as dívidas correntes. Entretanto, a situação de insolvência é um aspecto momentâneo, pois qualquer companhia pode tornar-se insolvente por determinado período e, posteriormente, tornar-se solvente (Szymanski, 2017).

As características de uma situação de insolvência ou não vão além de simples definição de concordatas e/ou falências para as empresas. Por isso, os índices de insolvência são comumente conhecidos no mercado financeiro como índices de endividamento, ou seja, são os índices que trazem à tona a liquidez de uma determinada empresa. Conforme Pereira, Domínguez e Ocejó (2007), os índices visam fornecer uma medida, ou melhor, um indicador da capacidade de a empresa pagar suas dívidas, partir da comparação entre os direitos realizáveis e as exigibilidades.

Vale destacar que existe diferença entre insolvência e falência, já que esta se estabelece quando o processo legal para encerrar as atividades da empresa foi realizado e finalizado. É um processo formal perante os órgãos públicos com o propósito de liquidação dos ativos da organização (Silva; Wienhage; Souza; Bezerra; Lyra, 2012).

Para o artigo, vamos utilizar a definição anterior de insolvência. Para prever esse estado, mediante uma abordagem computacional, iremos reunir os dados, conforme os demonstrativos contábeis, e observar sistemicamente ou inferir se determinada empresa está no estado de insolvência. Assim sendo, a possível solução do problema pode estar nos modelos de previsão, que permitem a identificação da insolvência, seja de saldo ou de fluxo, antes que a situação seja irreversível.

### **2.1.1 Risco de Crédito e Modelos de Previsão**

A gestão de risco nas últimas décadas passou a ter um destaque dentro das atividades da administração financeira. Isso se deve em razão da expansão do crédito e do crescimento desse mercado. A atividade de concessão de crédito é um elemento básico das instituições financeiras. A função de uma análise de crédito nada mais é que prever o risco envolvido para a concessão ou não dele. Portanto, o risco de crédito tem um papel relevante na composição dos riscos de uma instituição e pode ser encontrado em operações nas quais há liberação de recurso com a possibilidade do uso para os clientes (Gouvêa; Gonçalves; Mantovani, 2012).

Já a definição de avaliação de risco de crédito é vista como um processo de mensurar a probabilidade de que os fluxos de caixa esperados com as operações de crédito não se confirmem, gerando, assim, inadimplência (Soares; Coutinho; Camargos, 2012).

Em muitas empresas, a avaliação da concessão de crédito é baseada em uma variedade de informações vindas de fontes financeiras e não financeiras. Analisam-se essas informações de maneira subjetiva quando envolvem uma análise mais qualitativa e utilizam-se modelos de avaliação para uma verificação mais quantitativa. Muitas vezes, essas mensurações não são capazes de explicar os processos de tomada de decisão, embora consigam apontar os fatores que influenciam as decisões (Selau; Ribeiro, 2009).

O uso dos modelos de previsão de risco é vital: já que são baseados em dados recentes de clientes com a empresa, geram uma pontuação para as características que levam à criação de um padrão de comportamento em relação à inadimplência. Na verdade, o método estatístico utilizado para a construção do modelo leva em consideração o histórico da instituição com seus clientes, possibilitando a identificação das características capazes de diferenciar o bom do mau pagador (Pinto, 2021). No Brasil, é relativamente recente o uso de modelos de previsão, as instituições financeiras passaram a utilizar os modelos de *credit scoring* apenas após a estabilidade alcançada com a implantação do Plano Real (Camargos; Araújo; Camargos, 2012).

Segundo Silva (2017), uma das principais barreiras para desenvolvimento dessas análises no Brasil é a relativa escassez de pesquisas desenvolvidas como o propósito de

encontrar parâmetros para previsão de insolvência ou servirem de guias para concessão de crédito. Alguns dos principais trabalhos de pesquisa, nos quais foi utilizado o modelo de análise discriminante, ficou conhecido como termômetro da insolvência de Kanitz e o Z-score de Altman no Brasil. Ainda assim, encontramos poucos trabalhos de pesquisa no âmbito acadêmico fazendo junção da teoria financeira com abordagem tecnológica (Horta; Alves; Carvalho, 2014; Soares; Rebouças, 2015).

## **2.2 Inteligência Artificial e Aprendizagem de Máquina**

Conforme Facelli *et al.* (2011), por meio da experiência, aperfeiçoa-se a capacidade de melhorar o desempenho na realização de alguma tarefa. O objetivo de um algoritmo de aprendizado de máquina utilizado em tarefas de aprendizado, a partir de um subconjunto de dados, denominado conjunto de treinamento, que é denominado modelo ou hipótese, é capaz de relacionar os valores dos atributos de entrada de um objeto do conjunto de treinamento ao valor do seu atributo de saída.

Uma abordagem de aprendizado de máquina preditiva diz respeito às tarefas de previsão utilizadas no aprendizado de máquina, cuja meta é encontrar uma função também chamada de modelo ou hipótese, que, a partir dos dados de treinamento, possa ser utilizada para prever ou rotular um novo exemplo, com base nos valores que foram fornecidos na entrada (Scarpel, 2005).

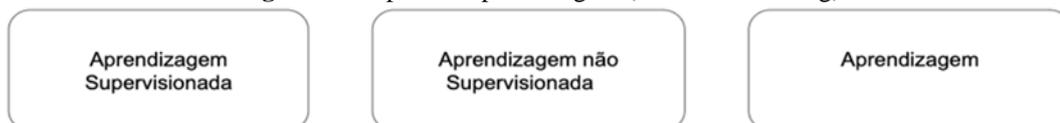
No processo de aprendizagem em si, assim como nós não precisamos ver todas as laranjas nem todos os limões do mundo para aprender a reconhecer essas frutas, os algoritmos de aprendizagem de máquina podem usar o poder computacional e ampla disponibilidade de dados, sobre os mais diversos tipos de fenômenos, para resolver muitos problemas, sem a necessidade de ter acesso a todos os dados.

No entendimento de Ribeiro, Neto, Merlo e Mello (2005), no aprendizado de máquina voltado para problema de insolvência ou mesmo solvência, de uma ou várias empresas, não será necessário olharmos todas as empresas, que faliram ou não. A partir de um conjunto de informação, um analista financeiro, por meio de observação constante, pode aprender as características necessárias, que apontam para uma possível falência de uma empresa, em um intervalo de tempo, aliando, assim, o poder computacional com a quantidade de dados disponíveis atualmente. É possível aprender de maneira mais rápida essas características e padrões, tornando o resultado mais rápido e efetivo, pois o computador consegue observar uma gama muito de grande de informações, em velocidade exponencial – um aspecto no qual nós, seres humanos, somos limitados.

### **2.2.1 Tipos de Aprendizagem de Máquina**

Segundo Haykin (2007) em processo de aprendizagem de máquina, temos alguns tipos de aprendizagem, representados na Figura 1 a seguir. A premissa é que estamos simulando o processo de aprendizagem humana, por meio da abordagem computacional.

**Figura 1** – Tipos de Aprendizagem (Machine Learning)



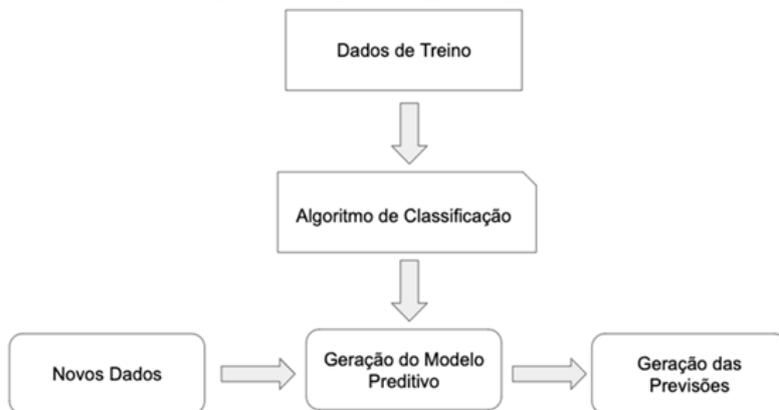
**Fonte:** Elaborado pelos autores, 2021.

### **2.2.2 Aprendizagem Supervisionada**

## PREDIÇÃO DE INSOLVÊNCIA DE EMPRESAS POR MEIO DA INTELIGÊNCIA ARTIFICIAL – TÉCNICAS DE APRENDIZADO DE MÁQUINA

A aprendizagem supervisionada, conforme a Figura 2, é o termo usado sempre que o programa ou algoritmo é treinado sobre um conjunto pré-definido de dados. Nesse sentido, os algoritmos fazem previsões com base em um conjunto de exemplos fornecidos e marcados previamente (Haykin, 2007; Facelli *et al.*, 2011).

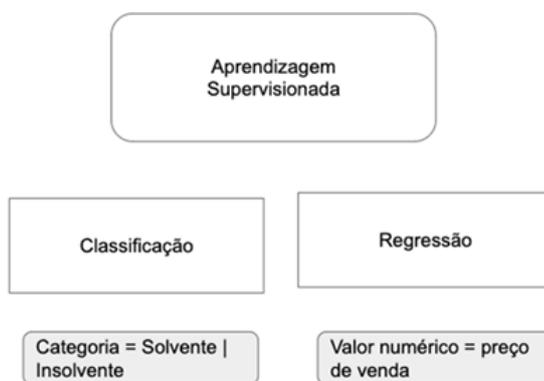
**Figura 2** – Aprendizagem Supervisionada



**Fonte:** Elaborado pelos autores, 2021.

Esse tipo de aprendizagem está subdividido de duas maneiras: aprendizagem de classificação, cujo objetivo é prever uma classe ou categoria, e aprendizagem de regressão, cuja intenção é prever o valor numérico.

**Figura 3** – Divisão da Aprendizagem Supervisionada

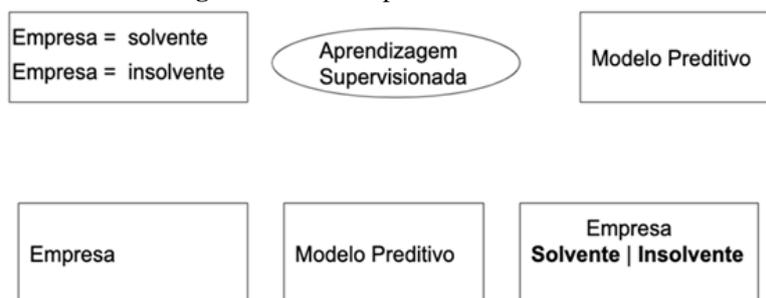


**Fonte:** Elaborado pelos autores, 2021

Classificação é processo de identificar a qual conjunto de categorias uma nova observação pertence, com base em um conjunto de dados de treino contendo observações (ou instância), com associação conhecida. Neste estudo, utilizaremos o tipo de aprendizagem supervisionada na sua subdivisão de classificação, o problema de previsão de insolvência para uma empresa, que consiste em classificar a empresa como solvente ou insolvente.

## PREDIÇÃO DE INSOLVÊNCIA DE EMPRESAS POR MEIO DA INTELIGÊNCIA ARTIFICIAL – TÉCNICAS DE APRENDIZADO DE MÁQUINA

Figura 4 – Passo a passo do modelo selecionado



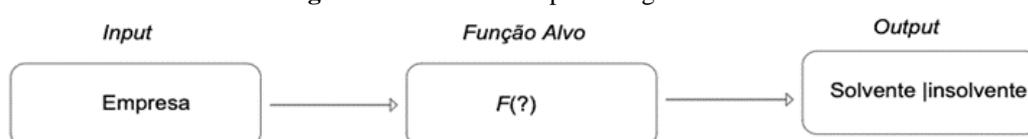
Fonte: Elaborado pelos autores, 2021.

Para determinar o diagnóstico da saúde financeira de uma empresa, temos a Figura 4, na qual se observam as características similares em outros grupos de empresas. Para tanto, iremos selecionar dentro do conjunto de observações, o que é chamado de variáveis preditoras, ou seja, o conjunto de característica que o algoritmo deve aprender. E iremos gerar a classe ou categoria, que é chamado de variável *target*, a qual pode assumir os valores de solvente ou insolvente nesse conjunto observável (Coelho; de Lima Amorim; de Camargos, 2021).

### 2.2.3 Processo de Aprendizagem

Um algoritmo de *machine learning* funciona de maneira semelhante ao método de aprendizagem humana. Na prática, estamos reproduzindo a inteligência humana por meio de computadores, e esse aprendizado acontece usando o poder de abstração da matemática (Scarpel, 2005). Em aprendizado de máquina, nós construímos as capacidades de previsões de um modelo de aprendizado de máquina, mediante uma formulação de uma função matemática, que inclui todas as características dadas de um determinado fenômeno. O algoritmo, a partir dos dados, irá criar uma formulação matemática, que tem como nome função alvo,  $f(x)$ , que possui as características do fenômeno de uma empresa, sejam elas de uma empresa solvente ou insolvente. À medida que o algoritmo for exposto a novos dados, ele determinará se a função matemática gerada é suficiente para explicar os dados observados (Haykin, 2007; Jabeur; Gharib; Mefteh-Wali; Arfi, 2021). A Figura 5 a seguir resume o processo de aprendizagem otimizado para a área financeira.

Figura 5 - Processo de aprendizagem otimizado



Fonte: Elabora pelos autores, 2021.

### 2.2.4 Processo de Generalização de Aprendizagem

Após o processo de aprendizagem, o algoritmo deve ser capaz de generalizar uma função matemática que faça previsões sobre novos conjuntos de dados. Nesse estudo, o processo de aprendizagem consiste em criar uma função matemática que aprenda as características das empresas solventes, bem como característica das empresas não solventes que, portanto, faliram. Desse modo, é possível fazer previsões sobre o conjunto de dados de uma nova empresa. Se o algoritmo não for capaz de fazer essa previsão, entendemos que o processo de aprendizagem memorizou os dados das empresas, mas não aprendeu realmente as

## PREDIÇÃO DE INSOLVÊNCIA DE EMPRESAS POR MEIO DA INTELIGÊNCIA ARTIFICIAL – TÉCNICAS DE APRENDIZADO DE MÁQUINA

suas características. Para generalizar a função que melhor resolve o problema, os algoritmos de aprendizagem de máquina baseiam-se em três componentes, conforme a Figura 6 a seguir.

**Figura 6** – 3 Componentes



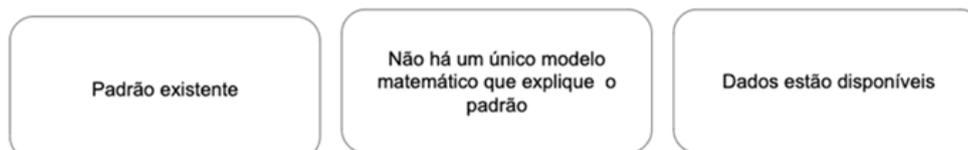
**Fonte:** Elaborado pelos autores, 2021.

Na visão de Scarpel (2005), os algoritmos de aprendizagem possuem diversos parâmetros internos. Assim, os núcleos dos algoritmos mudam constantemente os valores, fazendo uma otimização para representar a melhor função que resolve o problema dos dados, gerando um espaço de hipótese com várias hipóteses geradas, no intuito de encontrar a melhor hipótese que resolve o problema de previsão de insolvência para empresas.

### 2.2.5 Elementos do Processo de Aprendizagem

Na Figura 7, vemos os três elementos do processo de aprendizagem. No primeiro, espera-se que exista um padrão no conjunto de dados observados. Ou seja, é necessário que exista, previamente, um padrão no conjunto de dados. Caso não exista, não será possível observá-los. No segundo elemento para aprendizagem, não existe um modelo matemático que explique fielmente o padrão a ser observado. E, por fim, para que ocorra o processo de aprendizagem, é fundamental que os dados referentes ao problema de negócio estejam disponíveis, sendo possível, assim, observar um conjunto de informações para gerar o modelo.

**Figura 7** - Elementos do Processo de Aprendizagem



**Fonte:** Elaborado pelos autores, 2021.

O problema que temos como objetivo exploratório resolver, é criar um modelo preditivo para previsão de insolvência das empresas, observando os três elementos anteriores de aprendizagem, as condições estão satisfeita para modelagem preditiva, sendo que ao observar os diversas pesquisas, referente ao problema de insolvência, nota-se que existe um padrão entre as empresas que faliram, no que se refere aos dados financeiros, bem como, existe um padrão nas empresas que permaneceram solvente, também não existe uma modelo único que garante essa predição, várias tentativas de modelos foram gerados ao longo do tempo, cabendo ao analista financeiro dado sua experiência por observação constante julgar ou inferir se empresa irá falir ou não no próximo período, e por último os dados estão disponíveis em diversas fontes, seja empresas de capital aberto, seja em empresas privadas, que guardam suas informações contábeis em seus bancos de dados, para análise posterior internamente (Al-Shayea; El-Refae, 2012; Coelho; de Lima Amorim; de Camargos, 2021; Kezelj; Gruenbichler, 2021).

### 2.3 Indução do Espaço de Hipótese

Para ilustrar a relação entre o aprendizado de máquina e indução do espaço de hipóteses, teremos, neste estudo, um conjunto de dados financeiros das empresas, denominado como

## PREDIÇÃO DE INSOLVÊNCIA DE EMPRESAS POR MEIO DA INTELIGÊNCIA ARTIFICIAL – TÉCNICAS DE APRENDIZADO DE MÁQUINA

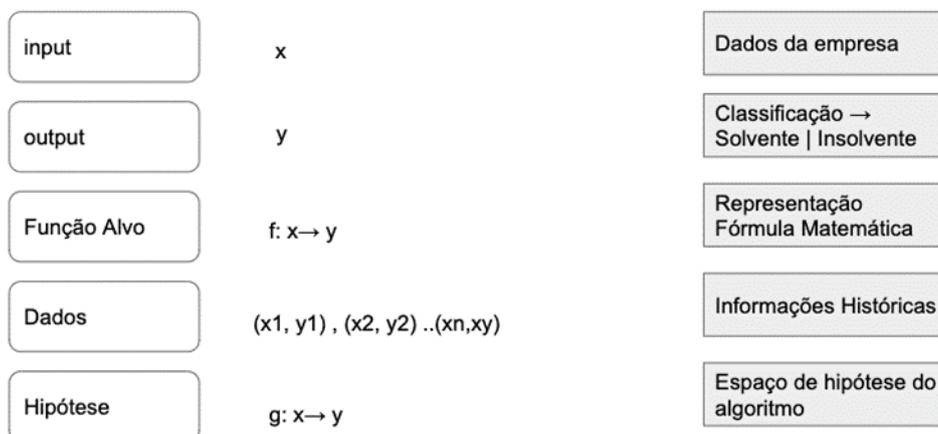
dados financeiros, de modo que cada um (exemplo ou registro) corresponda a uma empresa e uma dupla formada pelos valores e características ou atributos referentes à empresa, que descrevem seus principais aspectos financeiros. Os atributos utilizados para cada empresa podem ser, por exemplo, seu nome, período de avaliação (trimestral, semestral, anual), ativo total, ativo circulante total (exigido em curto ou em longo prazo), passivo circulante total, receita total, passivo total, que são dados disponíveis nos demonstrativos contábeis e financeiros das empresas.

Dentro do conjunto de informações das empresas, um dos atributos é considerado o atributo de saída. Neste caso, temos o de classificação, que pode receber o valor de solvente ou insolvente. Tal valor será estimado utilizando os demais atributos (os de entrada). Assim, o objetivo do aprendizado de máquina é aprender, a partir de um subconjunto de dados, denominado conjunto de treinamento, um modelo ou hipótese capaz de relacionar os valores de atributos de entrada de um objeto de conjunto de treinamento ao valor de seu atributo de saída (Horta *et al.*, 2014).

### 2.3.1 Modelagem da Aprendizagem Financeira

Nas palavras de Scarpel (2005), o algoritmo de aprendizagem usa o espaço de hipótese para gerar a melhor hipótese, que irá resolver o problema de negócio proposto, conforme vemos na imagem a seguir, Figura 8, na qual temos a representação dos componentes de aprendizagem financeira para o problema de previsão de falência.

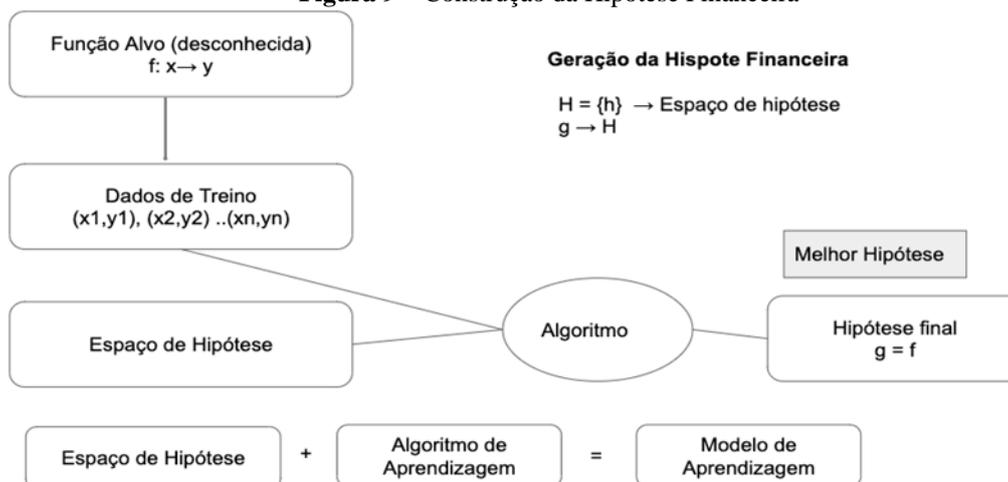
Figura 8 – Componentes da Aprendizagem Financeira



Fonte: Elaborado pelos autores, 2021.

Para visualização em termos abstratos, em resumo na Figura 9, temos como é gerada a melhor hipótese para conjunto de dados financeiros.

Figura 9 – Construção da Hipótese Financeira



Fonte: Elaborado pelos autores, 2021.

### 2.3.2 Criação do Modelo Preditivo

O problema de insolvência das empresas é dado, comumente, como um problema de classificação. Ou seja: temos que, ao analisar as informações financeiras da empresa, podemos classificá-la como empresa solvente, isto é, a instituição não irá falir rapidamente, pois sua situação financeira é favorável e, portanto, boa, mas podemos observar os dados e chegar à conclusão de que a empresa está passando por dificuldades financeiras para pagar os seus credores, classificando-a em situação de insolvência. Isso significa que a empresa passa por grande problema financeiro e poderá falir se nenhuma atitude for tomada, para corrigir-lhe a situação atual (Thomaz; Vellasco, 2005).

Com essa situação, o modelo predito, que é prever uma função de aproximação do fator solvência da empresa, será construído utilizando os dados financeiros históricos, com base nos principais demonstrativos financeiros, sendo eles: balanço patrimonial do ativo, balanço patrimonial do passivo, demonstração do resultado do exercício e demonstração do fluxo de caixa.

Assim, ao observar as informações contidas nos demonstrativos financeiros, nota-se que existe um conjunto de contas, que, ao longo do tempo, são utilizadas para se fazer a análise do fator de solvência. Nesse sentido, vamos trabalhar com um conjunto de variáveis preditoras, ou seja, que serão utilizadas para que o modelo predito possa aprender ou observar o padrão existente. E, como se trata de um problema de classificação, iremos trabalhar com uma variável chamada de Classe ou Variável *Target* ou também chamada de variável alvo. Neste caso, nossa variável *target* será a classificação da empresa como solvente ou como insolvente.

O objetivo do aprendizado de máquina é aprender a aproximação da função  $f$ , que melhor representa a relação entre os atributos de entrada, conhecida como variáveis preditoras, com a variável de saída, que recebe o nome de variável *target*.

### 2.3.3 Método Probabilístico de Aprendizagem de Máquina

A estatística é uma parte da matemática que fornece métodos para a coleta, organização, descrição, análise e interpretação de dados que viabilizam a tomada de decisões. Probabilidade consiste em modelos matemáticos que explicam os fenômenos estudados pela estatística em condições normais de experimentação.

Um modelo muito utilizado é o modelo Bayesiano, baseado no teorema de Bayes, que viveu na Inglaterra, estudou teologia na Escócia e publicou o livro A doutrina dos fluxos. O

## PREDIÇÃO DE INSOLVÊNCIA DE EMPRESAS POR MEIO DA INTELIGÊNCIA ARTIFICIAL – TÉCNICAS DE APRENDIZADO DE MÁQUINA

teorema de Bayes é um dos métodos probabilísticos mais utilizados, ou *naive bayes* (aprendizados nocivos), que calcula as probabilidades de os eventos independentes ocorrerem. Cada exemplo de treinamento pode decrementar ou incrementar a probabilidade de uma hipótese ser correta.

Exemplo: a probabilidade de a empresa estar solvente ou insolvente é dada como:

$$P(\text{Empresa} = \text{solvente}) = 0.08$$

$$P(\text{Empresa} = \text{insolvente}) = 0.96$$

Conforme Facelli *et al.* (2011) observaram, a partir das experiências passadas da utilização do teste de falência, é possível inferir as probabilidades condicionais para a variável Teste de Falência.

Aplicamos o teste de falência para saber a probabilidade de o teste ser verdadeiro.

$$P(\text{Teste de Falência} = \text{sim} \mid \text{Empresa} = \text{solvente}) = 0.75$$

$$P(\text{Teste de Falência} = \text{não} \mid \text{Empresa} = \text{insolvente}) = 0.96$$

Nessas probabilidades condicionais, o “Teste de Falência = sim” representa a probabilidade de o resultado de um teste ou diagnóstico de insolvência de uma empresa ser positivo. O oposto vale para o “Teste de Falência = não”. Os valores das probabilidades condicionais serão ilustrados na matriz de confusão a seguir.

Uma matriz de confusão (também conhecida como tabela de confusão) é uma matriz de dimensão 2x2, podendo ser maior, dependendo da quantidade de classes que se pretende prever. Portanto, é uma métrica voltada para modelos de classificação e tem como objetivo calcular a quantidade de falso positivo, falso negativo, assim como verdadeiro positivo, verdadeiro negativo, e fornecer acurácia geral do modelo.

Quando o modelo prevê um caso positivo corretamente, o classificador prevê uma empresa insolvente corretamente. Temos, neste caso, um verdadeiro positivo (True Positive). Caso uma empresa seja classificada como solvente, mas deveria tê-lo sido como insolvente, temos um falso positivo (False Positive), que é conhecido com *type I error*, ou tipo I de erro. Na avaliação do modelo, pode ocorrer o oposto do tópico anterior: quando o modelo previr que uma empresa não é insolvente, quando, na verdade, ela é insolvente. Neste contexto, temos, então, um falso negativo (False Negative), conhecido como *type II error*, ou tipo II de erro. Ainda, quando o modelo previr que a empresa está insolvente e o fez corretamente, temos verdadeiro negativo (True Negative).

### 3 METODOLOGIA

O artigo foi desenvolvido por meio da modalidade de estudo bibliográfico. Segundo Santos (2012), qualquer que seja o campo a ser pesquisado, sempre será necessária uma pesquisa bibliográfica, que proporciona um conhecimento prévio do estágio em que se encontra o assunto, já que procura explicar o problema a partir de referências teóricas publicadas em documentos.

Podemos entender a pesquisa descritiva como sendo aquela na qual não há interferência do pesquisador. Já para Marconi e Lakatos (2021), utiliza-se este tipo de pesquisa para conseguir informação ou acontecimentos acerca de uma questão para a qual se procura uma hipótese ou descobrir novos fenômenos. Cervo e Bervian (2007) salientam que, em relação à forma de abordagem do problema, esta pesquisa é classificada como quantitativa por tratar de

## PREDIÇÃO DE INSOLVÊNCIA DE EMPRESAS POR MEIO DA INTELIGÊNCIA ARTIFICIAL – TÉCNICAS DE APRENDIZADO DE MÁQUINA

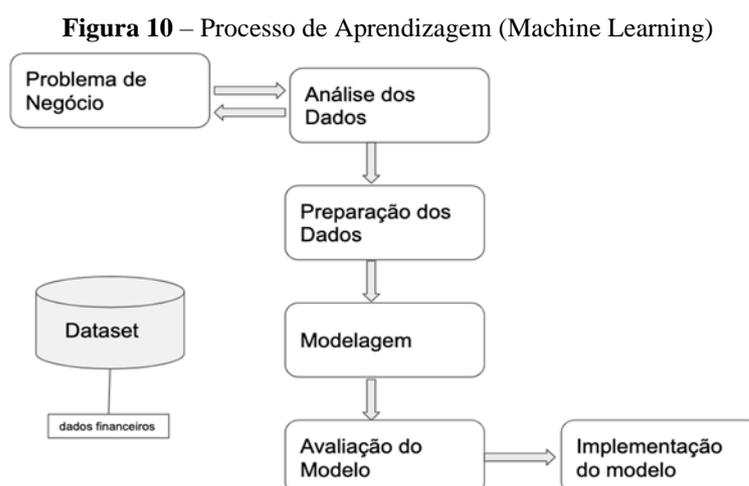
dados numéricos financeiros de empresas, setores e mercado.

A necessidade de diagnosticar ou prever a situação financeira de solvência da empresa sempre será crucial para o mercado financeiro em geral. Nesse sentido, acreditamos que, ao aplicarmos as técnicas de inteligência artificial, por meio de uma abordagem de aprendizado de máquina, usando métodos de aprendizado probabilístico, com o classificador *Naive Bayes*, temos uma solução simples, para coleta de dados reais das empresas. De fato, o aprendizado bayesiano irá calcular a probabilidade antes ou depois de a solvência ou a insolvência ocorrer. Temos, assim, uma informação relevante no tocante à questão de insolvência.

### 3.1 Análise e Classificação de Dados

Dada a característica de classificação das empresas, como solventes ou insolventes, iremos procurar questionar as principais informações contábeis e financeiras, para, por meio da aplicação de técnicas de aprendizagem de máquina, fazer classificação rotulando os resultados e, assim, prever se a empresa está solvente ou insolvente. Desse modo, iremos gerar o que é conhecido, em ciência de dados, como *dataset*, a representação dos conjuntos de informações históricas que pretendemos analisar.

A Figura 10 a seguir resume, de forma simplificada, o processo de aprendizado e análise de dados, para um problema a ser resolvido por aprendizagem de máquina.



Fonte: Elaborado pelos autores, 2021.

Neste trabalho de pesquisa fizemos uso da base dados da Economatica, que é referência no desenvolvimento de sistemas para análise de investimento, desde sua fundação em 1986, a empresa detém cem por cento do seu foco na coleta e gerenciamento de dados de altíssima confiabilidade. Desse modo a população objeto desta pesquisa é formada por empresas de capital listadas na Brasil Bolsa Balcão (B3).

Conforme há uma diversidade de informações disponíveis, bem como a natureza desse estudo, assumimos alguns parâmetros para coletar as informações na base de dados da Economatica, sendo eles: (1) O setor econômico selecionado é o comércio, (2) O período selecionado está em anos de 2014 a 2019, (3) A análise dos dados foram realizadas utilizando o Excel.

O conjunto de dados de treinamento inclui a amostra de números casos, cada um contendo valores no intervalo das variáveis de entrada e saída respectivamente. Um dos grandes desafios em ciência de dados e aprendizado de máquina, é decidir quais variáveis utilizar diante de universo muito grande possibilidades. Para este estudo exploratório, utilizamos os índices

## PREDIÇÃO DE INSOLVÊNCIA DE EMPRESAS POR MEIO DA INTELIGÊNCIA ARTIFICIAL – TÉCNICAS DE APRENDIZADO DE MÁQUINA

de solvência, bem como, indicadores financeiros, juntamente com algumas informações que representam as contas das empresas nos demonstrativos contábeis ou financeiros.

Usamos uma amostra de 24 empresas derivadas do setor de comércio no Brasil, onde os dados selecionados são dos últimos cinco anos sendo eles: 2014 a 2019. Este período foi escolhido devido a potencialização da crise econômica no país, onde observamos o advento da operação Lava Jato e todos os desdobramentos políticos e culminaram na instabilidade financeira em diversas empresas nos diversos setores da economia. Os dados classificam as empresas em duas categorias sendo uma como solvente e outra categoria com insolvente.

Todos os dados necessários para aplicação do modelo foram coletados através das demonstrações financeiras das empresas disponíveis na base de dados da Economatica, também foi feito o levantamento das informações, referente aos indicadores financeiros, tais como: liquidez seca, liquidez geral e liquidez seca da própria também da base Economatica de acordo com o período especificado.

### 3.2 Variáveis do Modelo

O aprendizado de máquina requer que dividamos o *dataset*, ou conjunto de dados (que consiste nas variáveis selecionadas para medir a saúde financeira das empresas), em duas partes: variável dependente e variável independente, para que, assim, o algoritmo possa aprender as características dos dados.

#### 3.2.1 Variável Dependente

As variáveis dependentes são os índices de solvência, bem como indicadores financeiros, juntamente com algumas informações que representam as contas das empresas nos demonstrativos contábeis ou financeiros, normalmente usados pelos analistas financeiros e no universo econômico em geral, bem como nas abordagens tradicionais de avaliação de solvência das empresas, nas quais essa seleção foi inspirada.

Nesse experimento, foi selecionado um conjunto de 16 variáveis, consistindo algumas delas em indicadores de solvência da empresa e outras, em informações gerais, bem como indicadores financeiros, conforme mostra a Tabela 1.

**Tabela 1** - Variáveis Depoentes do *dataset*

Variáveis Predictoras   Dataset	Nome	Descrição
1	Empresa	Identificação da empresa
2	Descrição	Descrição da empresa
3	Setor Economatica	Setor conforme a base economatica
4	Situação CMV	Status da empresa na CVM
5	Data [Ano]	Ano referencia das informações
6	Ativo Total	Total consolidado do ativo da empresa
7	Passivo+Patrimônio Líquido	Total Passivo + PL consolidado
8	Liquez Geral	Índice de Solvencia
9	Liquidez Seca	Índice de Solvencia
10	Liquidez Corrente	Índice de Solvencia
11	Ativo Circulante	Valor do ativo circulante
12	Ativo não circulante	Valor do ativo não circulante
13	Capital de Giro	Indicador Financeiro
14	EBTIDA	Resultado operacional
15	Alavagem financeira	Índice Finceiro
16	Patrimônio Líquido	Valor do patrimonio liquido

Fonte: Elaborado pelos autores, 2021.

# PREDIÇÃO DE INSOLVÊNCIA DE EMPRESAS POR MEIO DA INTELIGÊNCIA ARTIFICIAL – TÉCNICAS DE APRENDIZADO DE MÁQUINA

## 3.2.2 Variáveis preditoras do modelo

Algumas informações na Tabela 1, tais como nome da empresa, descrição e setor, não são relevantes para treinamento do modelo preditivo e podem até mesmo ser removidas do conjunto de dados. No entanto, foram exibidas para a identificação visual dos dados

Um dos grandes desafios, na área de ciência de dados, é a seleção das variáveis preditoras do modelo de predição. Para resolver ou mitigar esse problema, na pesquisa de campo, foi feita a seleção mediante entrevista, com profissional especializado em finanças que atua no mercado como CFO (*Chief Executive Officer*) de uma grande consultoria de tecnologia. Deste modo, foi possível entender o contexto de mercado e a relevância de cada variável selecionada.

### Variável Independente

A variável independente é a situação que queremos prever, ou seja, a probabilidade de uma empresa estar solvente ou insolvente. Neste estudo, usaremos o conceito de variável *dummy*, que é a transformação de uma variável texto para variável numérica. Em nosso *dataset*, temos a variável Classificação, que pode conter dois valores: solvente, para empresas que estão com boa saúde financeira, e insolvente, para aquelas com problemas ou risco de insolvência. Vejamos, na Tabela 2, como ficará o resultado dessa transformação.

- Se classificação = solvente será transformado em 1
- Se classificação = insolvente será transformado em 0

**Tabela 2 – Variável alvo do modelo**

Variável Alvo   Target   Dataset	Nome	Descrição
1	Classificação	Classificação da empresa em Solvente   Insolvente

Fonte: Elaborado pelos autores, 2021.

## 3.2.3 Variável alvo do modelo

A Tabela 3 a seguir resume o conjunto de informações no *dataset*, bem como alguns exemplos de registros, capturados de várias empresas, antes de serem processados pela solução proposta, que será definida no próximo tópico.

**Tabela 3 - Dataset final**

Empresa	Descrição	Setor Economico	Situação CVM	Data	Ativo Total	Passivo+Patrimônio Líquido	Liquidez Geral	Liquidez Corrente	Liquidez Seca	Ativo Circulante	Ativo Não Circulante	Capital de Giro	EBITDA	Alavancagem Financeira	Patrimônio Líquido	Classificação
Wim Industria	Concessionárias de outros veículos motorizados	Comércio	ATIVO	2019	526.430	526.430	2,1	3,2	2,1	200.420	326.010	126.884	52.028	1,3	415.044	Solvente
Viarejo	Loja de departamentos	Comercio	ATIVO	2019	21.858.000	21.858.000	0,8	0,7	0,4	10.770.000	11.088.000	-3.076.000	425.000	-154,8	1.474.000	Solvente
Vivara S.A	Loja de jóias, malas e artigos de couro	Comércio	ATIVO	2019	1.265.365	1.265.365	1,5	2,2	1,1	725.112	538.274	395.536	-	-	604.529	Solvente
RaisDrogasil	Loja de artigos para saúde e cuidados pessoais	Comércio	ATIVO	2019	12.562.528	12.562.528	0,7	1,4	0,5	5.526.455	7.036.071	1.704.767	1.953.376	2,5	4.305.550	Solvente
Dimed	Loja de artigos para saúde e cuidados pessoais	Comércio	ATIVO	2019	1.352.517	1.352.517	1,0	1,7	0,8	813.547	538.970	341.933	204.725	1,9	498.579	Solvente
Profarma	Comércio atacadista de remédios	Comércio	ATIVO	2019	3.148.602	3.148.602	1,0	1,3	0,8	2.021.043	1.128.559	459.559	123.621	-0,2	984.286	Solvente
Grupo Natura	Comércio atacadista de bens não duráveis variados	Comércio	ATIVO	2019	17.625.734	17.625.734	0,6	1,5	1,1	6.209.929	11.415.805	2.168.000	2.269.281	2,5	3.057.544	Solvente
Imc S/A	Loja de comida e bebida	Comércio	ATIVO	2019	2.148.733	2.148.733	0,5	1,9	1,8	572.634	1.577.099	272.830	174.334	0,3	954.646	Solvente
Lojas Renner	Loja de roupas	Comércio	ATIVO	2019	11.791.735	11.791.735	1,0	1,4	1,2	6.656.209	5.135.526	1.887.812	2.373.745	2,2	4.764.614	Solvente
Lojas Araçua	Outras lojas de mercadorias variadas	Comércio	ATIVO	2019	100.009	100.009	0,0	0,0	0,0	24.305	75.704	-7.027.571	-177.731	0,0	-8.194.069	Insolvente
Lojas Héring	Loja de roupas	Comércio	CANCELADA	2019	-	-	-	-	-	-	-	-	-	-	-	Insolvente
Lojas Americ	Loja de departamentos	Comércio	ATIVO	2019	32.121.409	32.121.409	0,9	2,2	1,7	18.654.206	13.477.203	9.982.695	3.070.369	0,9	7.032.299	Solvente
Dutty AG	Outras lojas de mercadorias variadas	Comércio	ATIVO	2019	39.697.263	39.697.263	0,4	1,2	0,6	9.231.574	30.465.689	1.525.429	3.919.801	1,2	14.456.612	Solvente
Carrefour BR	Loja de departamentos	Comércio	ATIVO	2019	37.406.000	37.406.000	1,0	1,0	0,6	15.940.000	21.467.000	424.000	3.721.000	1,9	14.948.000	Solvente
Centrao	Loja de artigos esportivos e instrumentos musicais	Comércio	ATIVO	2019	3.323.218	3.323.218	0,7	1,3	0,6	1.146.940	2.176.278	229.804	-	-	991.151	Solvente
Graziotin	Loja de roupas	Comércio	ATIVO	2019	776.108	776.108	1,5	2,7	1,9	303.446	472.661	102.866	99.497	3,4	530.043	Solvente
Cea Modas	Loja de roupas	Comércio	ATIVO	2019	5.547.552	5.547.552	0,8	1,8	1,3	2.155.084	3.392.468	970.209	-	-	1.912.021	Solvente
Battistella	Concessionárias de outros veículos motorizados	Comércio	ATIVO	2019	147.739	147.739	0,6	1,0	0,8	66.959	80.780	-242	10.207	-0,0	8.541	Solvente
B2W Digital	Vendas por correio ou meio eletrônico	Comércio	ATIVO	2019	16.726.818	16.726.818	1,2	2,5	2,2	10.210.791	6.516.027	6.999.474	483.414	-7,8	5.637.459	Solvente
BR Pharma	Loja de artigos para saúde e cuidados pessoais	Comércio	ATIVO	2019	70.490	70.490	0,2	0,1	0,1	11.206	59.284	-93.272	-26.199	-2,4	-262.204	Solvente
Advanced-On	Comércio atacadista de bens não duráveis variados	Comércio	ATIVO	2019	410	410	0,0	0,0	0,0	128	282	-15.294	-2.204	-0,1	-16.742	Insolvente
BR Towers Spa 1 SA	Loja de autopeças, acessórios e pneus	Comércio	CANCELADA	2019	-	-	-	-	-	-	-	-	-	-	-	Insolvente
Imprend Pague Menos S	Loja de mercadorias variadas	Comércio	ATIVO	2019	4.669.527	4.669.527	0,7	1,3	0,9	1.362.677	3.306.850	489.480	467.594	0,9	1.918.554	Solvente
Inbramós SA	Loja de roupas	Comércio	ATIVO	2019	1.178.987	1.178.987	0,5	1,7	0,8	365.412	814.575	147.380	99.539	-23,5	322.401	Solvente

Fonte: Elaborado pelos autores, 2021.

### 3.3 Acurácia e Precisão do Modelo Preditivo

No presente estudo, utilizaremos algumas técnicas de avaliação de performance do modelo, tais como matriz de confusão explicada em tópicos anteriores. A precisão, que é a proporção de casos positivos, e a acurácia, que é a proporção de casos positivos e negativos identificados corretamente, para, assim, medir o quanto o algoritmo de aprendizado de máquina está aprendendo as características do conjunto de dados, resumidos e apresentados no *dataset*.

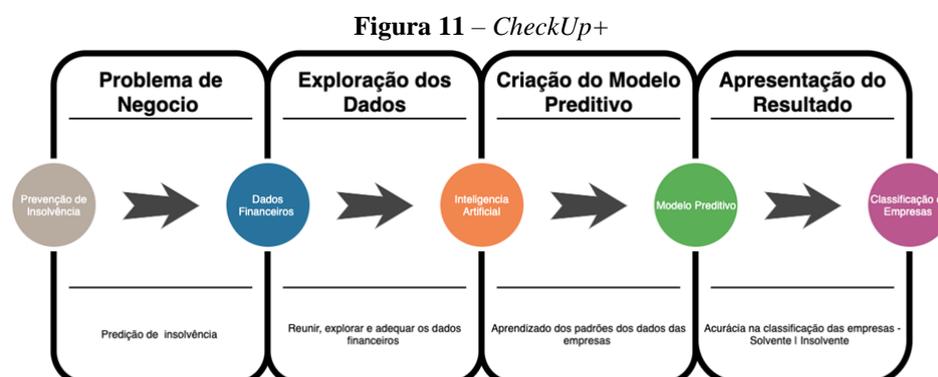
#### 3.3.1 Abordagem Experimental Proposta de Solução

Para esta abordagem procuramos desenvolver uma solução própria utilizando as características de uma ferramenta de análise de dados (*data analytics*), unificando o universo tecnológico, no que se refere às áreas de ciência da computação, ciência de dados, juntamente com o universo das finanças, no que tange análise de solvência das empresas, sendo o nome proposto para a ferramenta de *CheckUp+*, trazendo assim a alusão, a análise, medição e diagnóstico da saúde financeira das empresas observadas, o adendo (+) refere-se a intenção que o resultado seja positivo, embora seja sabido previamente que há empresas que estão em situação de saúde financeira ruim, e portanto podem estar no estado de insolvência no curto ou longo prazo.

O ambiente de alta competitividade das empresas, tem acirrado a busca pela sobrevivência, o autor caracteriza como competição analítica, como sendo uma forma de empresas extraírem valor das informações transformando e melhorando produtos e serviços para os seus clientes, de modo que é necessário construir ferramentas que utilizam o poder de grande volume de dados (*big data*) no sentido de tirar valor dos dados para as empresas. (Davenport; Barth; Bean, 2012).

Nesse sentido, procuramos nos apoiar em técnicas de *data visualization* (visualização de dados), bem como em boas práticas de *User Interface* (UI), de modo a trazer uma boa experiência para os utilizadores do produto. A escolha da ferramenta de apresentação também foi uma decisão importante, pois a partir dela as informações são acessadas e consumidas, para tomada de decisão. Na camada de visualização de dados, foi utilizado o Power BI, de acordo com a Microsoft, define o Power bi como um serviço de análise de negócios, lançado em 2015, com objetivo de fornecer visualizações interativas e recursos de *business intelligence* (BI) com uma interface simples para os usuários.

Para tanto, foi definido um fluxo de desenvolvimento da solução, em algumas etapas, conforme representado na Figura 11 a seguir.



Fonte: Elaborado pelos autores, 2021.

## PREDIÇÃO DE INSOLVÊNCIA DE EMPRESAS POR MEIO DA INTELIGÊNCIA ARTIFICIAL – TÉCNICAS DE APRENDIZADO DE MÁQUINA

Em relação ao problema de negócio, foi definido que iremos tentar solucionar, por meio da prevenção de insolvência de empresas, que é um problema que afeta o mercado como um todo. E, como já visto ao longo desse estudo, várias pesquisas e processos foram cunhados na tentativa de propor uma solução. Já na etapa de exploração dos dados, em ciência de dados, essa fase é também chamada de pré-processamentos dos dados. Utilizamos a base de dados da Economatica, para reunir, explorar e adequar os dados para o processamento computacional. Nesta etapa, é comum gastar a maior parte do tempo. Estima-se que 80% do tempo de um cientista de dados (o profissional habilitado para fazer as técnicas de aprendizagem de máquina serem aplicadas) seja gasto na fase de pré-processamento e tratamentos dos dados.

Uma vez que temos as informações devidamente tratadas, pode-se iniciar a criação do modelo preditivo, sendo gerado o *dataset*, conforme mencionado em tópicos anteriores, que nada mais é que os dados históricos das empresas, representando, assim, a situação de cada empresa. Desse modo, esse *dataset* é dividido em dois conjuntos de informações, sendo uma parte chamada de conjunto de treino, e outra parte é chamada de conjunto de teste. Uma divisão comum, amplamente utilizada na comunidade de ciência de dados, é dividir os dados em 70% para treino, e 30% dos dados para teste. Nesse momento, as informações são passadas para algoritmo de inteligência artificial, ocorrendo o aprendizado dos padrões observados nos dados, e serão feitas as previsões com base nas características das empresas. O resultado será um modelo de previsão de insolvência, bem como de solvência treinado, e pronto para ser utilizado com novos conjuntos de dados.

Após as etapas anteriores serem concluídas, fazemos uma apresentação e avaliação do modelo preditivo gerado, no intuito de verificar o quão assertivo é esse modelo, observando se o modelo está, de fato, aprendendo as características das empresas solventes e insolventes. São questões importantes a serem respondidas. Para o presente estudo, estabelecemos uma acurácia de ordem entre 70% a 85% de acerto na previsibilidade. Este valor foi proposto devido à análise dos vários modelos já cunhados em pesquisas e métodos anteriores, que se basearam na análise estatística discriminante. A solução segue o fluxo de desenvolvimento, de um projeto tipo de tecnologia, no que tange à área de ciência de dados, que utiliza a metodologia científica aplicada para resolver um problema de mercado e, assim, obter referências na sua abordagem. Mas também procura utilizar os métodos aplicados à academia, para inovar no mercado competitivo, com o foco em trazer um *insight*, no intuito de resolver um problema comum que afeta todas as empresas.

### 4 ANÁLISE E DISCUSSÃO DOS DADOS

A proposta deste trabalho foi utilizar uma solução que faça uso de uma abordagem de aprendizado de máquina. Para tanto, foi utilizado o modelo de aprendizagem baseado no teorema Bayes, conforme abordado em tópicos anteriores, e todo o processamento dos dados, foi feito por meio da solução proposta *CheckUp+*. Assim, um classificador bayesiano pode prever as probabilidades de associação de uma determinada classe, utilizando a probabilidade particular em uma amostra de dados. Sendo assim, o aprendizado nocivo assume que o efeito de um valor de atributo em uma determinada classe é independente de valores de outros atributos, essa suposição é chamada de independência condicional de classe.

Nesse experimento, foi utilizado um *dataset* com 16 variáveis independentes, ou variáveis preditoras, e uma variável dependente ou variável *target*, para que, assim, o algoritmo aprendesse as características internas do conjunto de dados. Também foi feita uma divisão dos

## PREDIÇÃO DE INSOLVÊNCIA DE EMPRESAS POR MEIO DA INTELIGÊNCIA ARTIFICIAL – TÉCNICAS DE APRENDIZADO DE MÁQUINA

dados em um conjunto de treino representando 70% dos dados, bem como um conjunto de teste representando 30% dos dados, respectivamente.

No primeiro cenário proposto para teste, foram utilizados os dados das 24 empresas, com base histórica de 2014 até 2019, sendo todos os dados consolidados em cada ano, bem como a devida classificação anual da empresa como solvente ou insolvente. Na execução da solução *CheckUp+*, com o algoritmo de aprendizagem de máquina utilizando classificador *naive bayes*, tivemos uma acurácia de, aproximadamente, 91%, com um nível de precisão de 70% para empresas que estão classificadas como insolventes. Ao analisarmos o *recall* para empresas insolventes, temos 83%. Essa métrica significa o quão bom é o modelo para acertar empresas insolventes. Já para empresas classificadas como solventes, obtivemos um *recall* de 90%, também com um alto nível de reconhecimento das empresas classificadas como solventes neste teste, que fica evidente nas Tabela 4 e 5.

**Tabela 4** – Matriz de Confusão e para o cenário de teste I

Matriz Confusão Cenário I	Insolvente	Solvente
Empresa Insolvente	26	0
Empresa Solvente	11	100

**Fonte:** Elaborado pelos autores, 2021.

**Tabela 5** - Relatório de Classificação Geral do Modelo cenário I

Classe	Precision	Recall
0 = insolvente	0.70	0.83
1 = solvente	1.00	0.90

**Fonte:** Elaborado pelos autores, 2021.

Para fazer a validação do modelo, foi feito um segundo cenário de teste, realizado para 24 empresas, somente com os dados consolidados, referentes ao ano de 2019. Assim, tivemos acurácia do modelo de, aproximadamente, 96%, com nível de precisão de 80% para empresas que estão realmente insolventes. Ao analisarmos o *recall*, verificamos que a proporção de positivos encontrados corretamente pelo modelo foi de 100%, o que significa que o modelo consegue reconhecer muito bem as empresas classificadas como insolventes. Já para empresas classificadas com solventes, o modelo teve um *recall* de 95%, também com um alto nível de reconhecimento das empresas classificadas como solventes, segundo demonstrado nas Tabelas 6 e 7.

**Tabela 6** – Matriz de Confusão para o cenário de teste II

Matriz Confusão Cenário II	Insolvente	Solvente
Empresa Insolvente	4	0
Empresa Solvente	1	19

**Fonte:** Elaborado pelos autores, 2021.

**Tabela 7** – Relatório de Classificação Geral do Modelo cenário II

Classe	Precision	Recall
0 = insolvente	0.80	1.00
1 = solvente	1.00	0.95

**Fonte:** Elaborado pelos autores, 2021.

## PREDIÇÃO DE INSOLVÊNCIA DE EMPRESAS POR MEIO DA INTELIGÊNCIA ARTIFICIAL – TÉCNICAS DE APRENDIZADO DE MÁQUINA

Verificamos também que, no segundo cenário, com menor quantidade de dados, o modelo pode estar aprendendo demais as características dos dados, o que é conhecido como *overfitting*, e, portanto, não ser tão preciso como esperado. Observamos que, no primeiro cenário, em que a base histórica foi maior (apresentando, assim, maior pluralidade das informações), o modelo reduziu a precisão, bem como sua acurácia, o que faz mais sentido no mundo real.

### 5 CONSIDERAÇÕES FINAIS

O objetivo deste artigo foi prever a situação de insolvência ou solvência das empresas, por meio da abordagem de aprendizado de máquina, usando, assim, um algoritmo de inteligência artificial, para criar um modelo preditivo com base nos resultados obtidos no treinamento do modelo.

Como resultado, vemos que foi viável utilizar o aprendizado de máquina, para ter um alto nível de acurácia, bem como gerar um modelo predito, utilizando-o, assim, com outros dados, conforme os cenários de teste feitos. Quanto à proposta de solução CheckUp+ para diagnóstico da saúde financeira das empresas, na ordem de 70% de acerto e previsibilidade, verificamos que foi possível alcançar este resultado. Diante dos trabalhos feitos por outros pesquisadores, que usaram os modelos de estatísticas discriminantes para calcular o fator de insolvência, verificou-se um percentual de acerto de 70% a 85% nos casos de previsão de insolvência. Desse modo, temos evidências, conforme Selau e Ribeiro (2009), Gouvêa, Gonçalves e Mantovani (2012), Soares e Rebouças (2015) e Pinto (2021), que atestam ser possível automatizar, de maneira efetiva e com melhor desempenho, a previsão de falências, trabalhando com métodos estatísticos automatizados juntamente com as técnicas de aprendizado de máquina.

Porém, alguns problemas são comuns ao aplicarmos métodos estáticos, sendo um deles problema com o tamanho da massa de dados. Para um aprendizado de máquina, quanto maior o conjunto de dados, melhor o modelo é treinado, sendo, assim, o modelo mais preciso em seu aprendizado. Outra dificuldade encontrada neste trabalho foi fazer a classificação corretamente das empresas em solvente e insolvente. Para mitigar esse problema, foi feita uma análise dos dados e indicadores de solvência, bem como indicadores financeiros, determinando a classe da empresa (Ribeiro *et al.*, 2005; Thomaz; Vellasco, 2005; Horta; Borges; Dos Santos Alves, 2015).

Pesquisas futuras podem ser feitas para melhorar a performance deste modelo, bem como a solução proposta pelo CheckUp+, por meio do ajuste dos dados, no que se refere à coleta e ao tratamento das informações, utilizando novas tecnologias, como o Big Data, que amplia a disponibilidade de informação no mercado financeiro. Seria importante utilizar outros modelos de aprendizado de máquina, tal como Support Vector Machine, sendo um espaço de hipótese mais complexo, com alto nível de abstração dos dados. Também é possível utilizar métodos Ensemble de aprendizagem, que nada mais é que uma aplicação de um conjunto de algoritmos para, assim, corrigir erros nos modelos e garantir uma melhor precisão, como a acurácia desejada. E, por fim, redes neurais podem ser uma boa ferramenta para a criação de modelos preditivos em tempo real, ajudando, assim, as empresas a cuidarem de sua saúde financeira no mercado altamente competitivo.

### 6 REFERÊNCIAS

**PREDIÇÃO DE INSOLVÊNCIA DE EMPRESAS POR MEIO  
DA INTELIGÊNCIA ARTIFICIAL – TÉCNICAS DE APRENDIZADO DE MÁQUINA**

- Al-Shayea, Q. K., & El-Refae, G. A. (2012). Evaluation of banks insolvency using artificial neural networks. In: *Proceedings of the 11th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases (AIKED'12)*, Cambridge, United Kingdom.
- Camargos, M. A.; Araújo, E. A. T.; & Camargos, M. C. S. (2012). A inadimplência em um programa de crédito de uma instituição financeira pública de Minas Gerais: uma análise utilizando regressão logística. *REGE - Revista de Gestão*, 19(3), 473-492.
- Cervo, A. L.; & Bervian, P. A. (2007). *Metodologia Científica*. 6ª ed. São Paulo: Pearson Prentice Hall.
- Coelho, F. F., de Lima Amorim, D. P., & de Camargos, M. A. (2021). Analisando métodos de machine learning e avaliação do risco de crédito. *Revista Gestão & Tecnologia*, 21(1), 89-116.
- Davenport, T. H., Barth, P. & Bean, R. (2012). How “big data” is different. *MIT Sloan Management Review*, 54(1).
- De Souza, C. A.; & Dos Santos, D. H. A. (2013). A situação econômica e financeira das empresas com negociação suspensa na BOVESPA e os modelos de previsão de falência. *Revista Mineira de Contabilidade*, 4(52), 6-14.
- Facelli, K.; Lorena, A. C., Gama, J., Almedia, T. A.; & Carvalho, A. C. P. F. L. (2011). *Inteligência Artificial Uma Abordagem de Máquina*. 2ª ed. São Paulo: LTC.
- Gouvêa, M. A.; Gonçalves, E. B.; & Mantovani, D. M. N. (2013). Análise de risco de crédito com aplicação de regressão logística e redes neurais. *Contabilidade Vista & Revista*, 24(4), 96-123.
- Gozer I. C.; Gimenes, R. M. T.; Menezes, E. A.; De Albuquerque, A. R. P. L.; & Isotani, S. (2014). Avaliação de insolvência em cooperativas de crédito: uma aplicação de redes neurais artificiais e do sistema PEARLS. *Informe Gepec*, 18(1), 6-30.
- Haykin, S. (2007). *Redes neurais: princípios e prática*. 2ª ed. São Paulo: Bookman, 2007.
- Horta, R. A. M.; Borges, C. C. H.; Carvalho, F. A. A. D.; & Alves, F. J. D. S. (2011). Previsão de insolvência: Uma estratégia para balanceamento da base de dados utilizando variáveis contábeis de empresas brasileiras. *Sociedade, Contabilidade e Gestão*, 6(2).
- Horta, R. A. M.; Alves, F. J. S.; & Carvalho, F. A. A. (2014). Seleção de atributos na previsão de insolvência: aplicação e avaliação usando dados brasileiros recentes. *RAM - Revista de Administração Mackenzie*, 15, 125-151.
- Horta, R. A. M.; Borges, C. C. H.; & Dos Santos Alves, F. J. (2015). Previsão de Insolvência no Setor de Materiais Básicos Aplicando Mineração de Dados. *Revista Universo Contábil*, 11(3), 43-62.
- Jabeur, S. B., Gharib, C., Mefteh-Wali, S., & Arfi, W. B. (2021). CatBoost model and artificial intelligence techniques for corporate failure prediction. *Technological Forecasting and Social Change*, 166, 120658.
- Kezelj, T., & Gruenbichler, R. (2021). A Systematic Literature Review on Corporate Insolvency Prevention Using Artificial Intelligence Algorithms. *Journal of Strategic Innovation & Sustainability*, 16(4).
- Marconi, M. A.; & Lakatos E. M. (2021). *Fundamentos da Metodologia Científica*. 9ª ed. São Paulo: Atlas.
- Minatto, F.; & Borba, J. A. (2022). Insolvência em clubes de futebol brasileiros: proposição de modelos baseados em redes neurais. *BBR - Brazilian Business Review*, 18, 624-642.
- Nunes, R. V.; & Sales, G. A. W. (2020). Quais as Condições de Liquidez e Insolvência das Empresas Estatais S/A não Financeiras? *Revista de Contabilidade e Gestão Contemporânea UFF*, 3(1), 76-87.

**PREDIÇÃO DE INSOLVÊNCIA DE EMPRESAS POR MEIO  
DA INTELIGÊNCIA ARTIFICIAL – TÉCNICAS DE APRENDIZADO DE MÁQUINA**

- Pereira, J. M.; Domínguez, M. Á. C.; & Ocejo, J. L. S. (2007). Modelos de previsão do fracasso empresarial: aspectos a considerar. *Tékhné-Revista de Estudos Politécnicos*, 7, 111-148.
- Pinto, A. C. (2021). O poder preditivo dos modelos com aprendizado de máquina é superior aos modelos tradicionais para análise do risco de crédito? *Revista Debates em Economia Aplicada*, 1(1).
- Ribeiro, E. M. S.; Neto, J. D. D. O.; Merlo, E. M.; & Mello, C. R. G. D. C. (2005). Aplicação das Redes Neurais na Concessão de Crédito - Um Estudo de Caso em uma Empresa de Consórcio. *Congresso USP de Controladoria e Contabilidade*, VI, Anais. São Paulo.
- Santos, J. A. D. (2012). *Metodologia Científica*. 2ª ed. São Paulo: Cengage Learning.
- Scarpel, R. A. (2005). Utilização de Support Vector Machine em previsão de insolvência de empresas. *Simpósio Brasileiro de Pesquisa Operacional*, XXXVII, Anais, 671-677.
- Silva, J. O. da; Wienhage, P.; Souza, R. P. S. de; Bezerra, F. A.; & Lyra, R. L. W. C. de. (2012). Capacidade Preditiva De Modelos De Insolvência Com Base Em Números Contábeis E Dados Descritivos. *Revista de Educação e Pesquisa em Contabilidade (REPeC)*, 6(3), 246–261.
- Silva, J. P. (2017). *Gestão e Análise de Risco de Crédito*. 9ª ed. São Paulo: Cengage Learning.
- Selau, L. P. R.; & Ribeiro, J. L. D. (2009). Uma sistemática para construção e escolha de modelos de previsão de risco de crédito. *Gestão & Produção*, 16, 398-413.
- Soares, G. O. G.; Coutinho, E. S.; & Camargos, M. A. (2012). Determinantes do Rating de Crédito de Companhias Brasileiras. *Revista Contabilidade Vista & Revista*, 23(3), 109–143.
- Soares, R. A.; & Rebouças, S. M. D. P. (2015). Avaliação do desempenho de técnicas de classificação aplicadas à previsão de insolvência de empresas de capital aberto brasileiras. *Revista ADM MADE*, 18(3), 40-61.
- Souza, C. R. (2010). *Modelos para Previsão do Risco de Crédito*. 105 f. 2010. Dissertação de Mestrado, Universidade de Campinas, Campinas.
- Szymanski, S. (2017). Entry into exit: insolvency in English professional football. *Scottish Journal of Political Economy*, 64(4), 419-444.
- Thomaz, C. E.; & Vellasco, M. M. (2005). Análise de Tendências de Mercado por Redes Neurais Artificiais. *Congresso Brasileiro de Redes Neurais*, VII, Anais, Natal.