



## PROPOSTA DE SOLUÇÃO DE ANÁLISE DE CRÉDITO BASEADA NO USO DE TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL

Vinícius Alves May<sup>1</sup>

Yuri Crotti<sup>2</sup>

**Resumo:** Com o aumento das operações de crédito, a necessidade de avaliar as informações dos clientes antes de efetuar uma venda é essencial. Neste processo, o histórico dos clientes será importante para obter uma resposta mais rápida e segura. A utilização de técnicas de inteligência artificial para analisar estas informações e fornecer o apoio necessário, passa a ser uma possibilidade. Com estas técnicas, é possível treinar as inteligências artificiais para auxiliarem nas tomadas de decisões e efetuarem tarefas até então realizadas por pessoas. Buscou-se com este estudo, desenvolver um sistema de análise de crédito que utilize técnicas de inteligência artificial para fornecer decisão rápida e segura, minimizando os riscos de inadimplência. A simulação do sistema foi realizada através de ferramentas como: Python e Sklearn. Para encontrar o melhor modelo, foi realizada a utilização do método *GridSearch*. Os resultados da simulação indicaram que o uso destas técnicas aumenta as chances de identificar clientes candidatos à inadimplência. A média de acerto do modelo analisando as métricas de classificação foram de 75% para acurácia, precisão 76%, *recall* 75% e *f1-score* 75%. A análise parte da comparação entre o percentual de acerto dos clientes que eram ou não inadimplentes e foram classificados corretamente pelo algoritmo. Com a inclusão de informações mais detalhadas sobre os clientes, a precisão será ainda maior, possibilitando que em cenários mais extremos ela seja decisiva para a tomada de decisão do analista de crédito.

**Palavras-Chave:** Inteligência artificial; Análise de crédito; Machine learnig; Python; Sklearn;

### 1 INTRODUÇÃO

A concessão de crédito é decisiva para fomentar as empresas e mantê-las ativas. Decidir qual será o limite de crédito de um cliente não é uma tarefa simples, precisam ser considerados alguns fatores, por exemplo, o cenário da economia no momento da análise e o histórico do cliente no mercado. Para que a análise seja mais

---

<sup>1</sup> Graduando em Engenharia de Computação. Ano 2022-1. E-mail: [viniciusmay11@gmail.com](mailto:viniciusmay11@gmail.com)

<sup>2</sup> Professor do Centro Universitário UniSATC. E-mail: [yuri.crotti@satc.edu.br](mailto:yuri.crotti@satc.edu.br)



crateriosa, o analista de crédito necessita de tempo, o que muitas vezes não é possível em virtude de negociações efetuadas no ato da compra.

De acordo com Carneiro (2017), com a necessidade de agilizar o parecer, o desenvolvimento de softwares voltados para esta área temática tornou-se imprescindível. Os softwares lançados até 1960 eram muito simples, onde o foco era agrupar as informações dos clientes em um único local, para facilitar a análise dos dados. Esses programas facilitam a avaliação das informações, mas continuam exigindo esforço por parte dos responsáveis pela análise de crédito (CARNEIRO, 2017).

Automatizar este processo passa a ser desejo não só dos analistas, mas também dos desenvolvedores. Fornecer decisão de forma rápida e segura, para minimizar os riscos de inadimplência e aumentar as vendas, é o interesse de todos. Alguns provedores de informações analisando o movimento do mercado começaram a sugerir limites de crédito, com base nas informações que possuem. Esta ação abasteceu as empresas com informações dos clientes, mas ainda não é garantia de veredito seguro.

Zerar o risco de inadimplência não é possível, mas pode-se usar técnicas para minimizar até um valor aceitável. Desta forma é necessário ter conhecimento sobre o assunto para diferenciar quais informações são relevantes para tomada de decisão.

De acordo com o Banco Central do Brasil (2021), as operações de crédito do Sistema Financeiro Nacional concederam no total R\$ 313,9 bilhões em fevereiro de 2021. Este valor representa um aumento de 6% em comparação com o mesmo mês em 2020 (BANCO CENTRAL DO BRASIL, 2021).

Um dos fatores que possibilitam o aumento da concessão de crédito é o avanço da tecnologia. Com mais informações disponíveis em tempo real sobre os clientes e a análise dos dados realizada por inteligência artificial, a redução do índice de inadimplência e o aumento do volume de negócios fazem parte do resultado obtido pelos analistas de crédito.

Para entender como a inteligência artificial auxilia na análise de crédito, primeiro é necessário entender o funcionamento dela. A inteligência artificial é um conjunto de técnicas que simulam a inteligência humana. Uma das capacidades dela é estudar os dados e aprender de forma autônoma ampliando o seu conhecimento.



Com isso é possível treiná-las para efetuarem tarefas até então realizadas por pessoas (TOTVS, 2019).

Tendo em vista o cenário, a utilização de inteligência artificial para auxiliar no parecer da análise passa a ser uma rota possível. Dentre os objetivos que ela possui, pode-se citar a entrega de resultado baseado no aprendizado de máquina, obtido através dos dados fornecidos a ela.

Assim, o objetivo deste trabalho está no desenvolvimento de sistema de análise de crédito que utilize técnicas de inteligência artificial para fornecer decisão de risco de inadimplência.

## **2 REVISÃO BIBLIOGRÁFICA**

### **2.1 ANÁLISE DE CRÉDITO**

Ao oferecer crédito a um cliente, a empresa precisa se certificar de que ele possui confiabilidade. Segundo Marques (2020), “Para oferecer crédito a um consumidor, às instituições financeiras e estabelecimentos comerciais precisam ter certeza de que vão receber o valor de volta.”.

Constatar se o cliente possui condições de honrar o compromisso assumido é princípio básico na hora de avaliar o crédito oferecido. Durante o processo de análise algumas informações básicas e financeiras do cliente deverão ser levadas em consideração. A partir destes dados, a empresa conseguirá estipular um valor mínimo de compra e quais serão as condições de pagamento. Considerar estes pontos auxilia a evitar fraudes e gerar estabilidade para a empresa (SERASA EXPERIAN, 2020).

Segundo a Deps Tecnologia (2018), gerir este conjunto de critérios é algo complexo. Devido às constantes mutações que o cenário sofre, as técnicas pessoais que levam em consideração apenas o conhecimento humano gastam mais tempo. Sendo assim, exercer gestão de crédito de qualidade que seja rápida e eficaz coloca a empresa à frente da concorrência.

Diminuir o tempo da análise é uma das principais vantagens oferecidas pelos softwares de gestão de crédito. Com acesso aos dados que as empresas possuem e levando em consideração as políticas exercidas, eles conseguem



automatizar etapas. Dentre as razões para a diminuição do tempo pode-se citar: gestão integrada, automação do processo decisório, vantagem na gestão dos recursos humanos e sugestão de limite e classificação (DEPS TECNOLOGIA, 2018).

Os softwares de análise de crédito são capazes de cruzar as informações que as empresas já possuem com as informações de órgãos oficiais, bancos e outros estabelecimentos e ao final da análise gerar um perfil para o cliente. Com base no resultado obtido e levando em consideração os critérios definidos pela empresa, o software determina se é possível conceder o crédito e qual o seu valor (DEPS TECNOLOGIA, 2019).

## 2.2 INTELIGÊNCIA ARTIFICIAL

De acordo com a Oracle (2021c), a inteligência artificial (IA) trata-se de sistemas ou máquinas capazes de executar funções até então executadas por humanos, isto deve-se ao fato de que ela é capaz de imitar a inteligência humana e aprimorar os seus conhecimentos através dos dados que coleta. O objetivo da IA não é substituir os seres humanos, mas sim auxiliar e melhorar as funções exercidas por eles.

A IA é composta por diversas tecnologias, que ao trabalharem em conjunto são capazes de fornecerem os resultados esperados, dentre elas: *machine learning*, *deep learning*. Cada uma desempenha um papel diferente no processo, mas isso não significa que será necessário utilizar todas elas (SAS, 2021c).

Segundo a SAS, “*Deep learning* é um tipo de machine learning que treina computadores para realizar tarefas como seres humanos, o que inclui reconhecimento de fala, identificação de imagem e previsões.”. O *deep learning* utiliza redes neurais artificiais consideradas profundas, para que ela possa identificar padrões e extrair características em suas diversas camadas de processamento (SAS, 2021c).

O *machine learning* (em português, aprendizado de máquina) é voltado para ideia de que os sistemas podem identificar padrões existentes nos dados e aprender com eles, assim sendo capazes de tomar decisões com a menor intervenção humana possível. Diferente do deep learning, em machine learning se faz necessário a extração de características durante o desenvolvimento. Na IA ele será responsável por ensinar a máquina sobre como aprender com dados (SAS, 2021c).



Cada uma das tecnologias apresentadas possui funções específicas e juntas formam o que hoje é conhecido como inteligência artificial. Tendo como objetivo a identificação de padrões nos dados e o aprendizado a partir deles, o uso de *machine learning* é o ideal. Empresas de e-commerce utilizam esta tecnologia para recomendar, de forma personalizada, produtos e serviços a seus clientes (Medes et al, 2019).

### 2.2.1 Machine Learning

Segundo a IBM, “Um aplicativo ou software com Machine Learning é um tipo de programa que melhora automaticamente e gradualmente com o número de experiências em que ele é colocado para treinar.”.

O aprendizado de máquina pode ocorrer de três formas: supervisionado, não supervisionado e por reforço. Para cada forma existe uma abordagem específica, tudo vai depender do objetivo que está buscando (IBM, 2021c).

No aprendizado supervisionado o resultado que será encontrado já é conhecido e isto ocorre por conta de que os dados utilizados possuem a resposta desejada. Para este processo é utilizado uma lista de variáveis independentes para prever uma variável dependente. Dentre as técnicas utilizadas para resolver problemas desta forma de aprendizado tem-se: regressão linear, regressão logística, redes neurais artificiais, máquinas de vetores de suporte, árvores de decisão, k-vizinhos mais próximos e Bayes ingênuo (HONDA et al, 2017).

Já os problemas que utilizam o aprendizado não supervisionado, não é possível ter uma clareza de qual será o resultado encontrado. Diferente do aprendizado supervisionado, neste os utilizados não são rotulados, ou seja, será necessário utilizar algoritmos para classificar os dados com base nos padrões e clusters encontrados (IBM, 2021c). Em outras palavras, ao utilizar este tipo de aprendizado espera-se entender melhor os dados que estão sendo analisados. As técnicas mais utilizadas no aprendizado não supervisionado são: redes neurais artificiais, máquinas kernel, florestas isoladoras e algoritmos de clusterização (HONDA et al, 2017).

Oposto aos outros dois aprendizados, o aprendizado por reforço não é treinado com um conjunto de dados, ele irá aprender através do método de tentativa

e erro. Desta forma, uma sucessão de decisões positivas irá resultar em processo reforçado, pois será mais eficaz na resolução do problema em questão (HONDA et al, 2017).

A Figura 1 contém no círculo central o aprendizado de máquina, que depois é dividido em mais três círculos que representam os tipos de algoritmos. Nos aprendizados supervisionado e não supervisionado, temos uma nova divisão que representam as classificações que os algoritmos possuem. Os círculos menores trazem exemplos de aplicações que utilizam os respectivos algoritmos.

Figura 1: Três tipos de aprendizagem



Fonte: Adaptado de Pant (2019)

A linguagem de programação mais utilizada entre os desenvolvedores para aplicações sobre este tema é o Python, isto devido a sua sintaxe relativamente simples e de fácil compreensão. Dentre as vantagens que ela possui, pode-se citar o grande número de bibliotecas nativas e de terceiros (ROVEDA, 2020).



O scikit-learn ou sklearn como é conhecida, é uma biblioteca de Python desenvolvida especificamente para aplicações de *machine learning*. A biblioteca foi construída com base nos pacotes NumPy, SciPy e matplotlib. A sklearn possui ferramentas simples e eficientes para predição de dados, pode ser utilizada em diferentes situações, possui código aberto e está disponível para todos (DIDÁTICA TECH, 2022c).

Organizada em módulos, a sklearn possui cada um desenvolvido para uma funcionalidade diferente. Conforme a documentação oficial, para entender melhor o funcionamento dela, podemos dividi-la em cinco módulos: Estimadores básicos, pré-processamento e transformadores, avaliação de modelos, *pipelines* e busca automática de parâmetros (PACHECO, 2021).

Os estimadores básicos são os vários algoritmos e modelos de *machine learning* que a biblioteca possui. O pré-processamento e transformadores possuem como objetivo normalizar os dados de entrada, visto que para treinar um modelo de predição, é essencial normalizarmos estes dados. A avaliação de modelos, como o próprio nome já diz, é responsável por avaliar o desempenho e a qualidade do modelo. Com o objetivo de agrupar alguns dos passos essenciais para treinamento do modelo, a pipeline exerce esta função e possibilita que seja realizada apenas uma chamada para todas as funções agrupadas. Como todos os modelos possuem parâmetros, que necessitam serem definidos antes da execução do treinamento, a busca automática de parâmetros auxilia na identificação dos melhores parâmetros a serem utilizados (PACHECO, 2021).

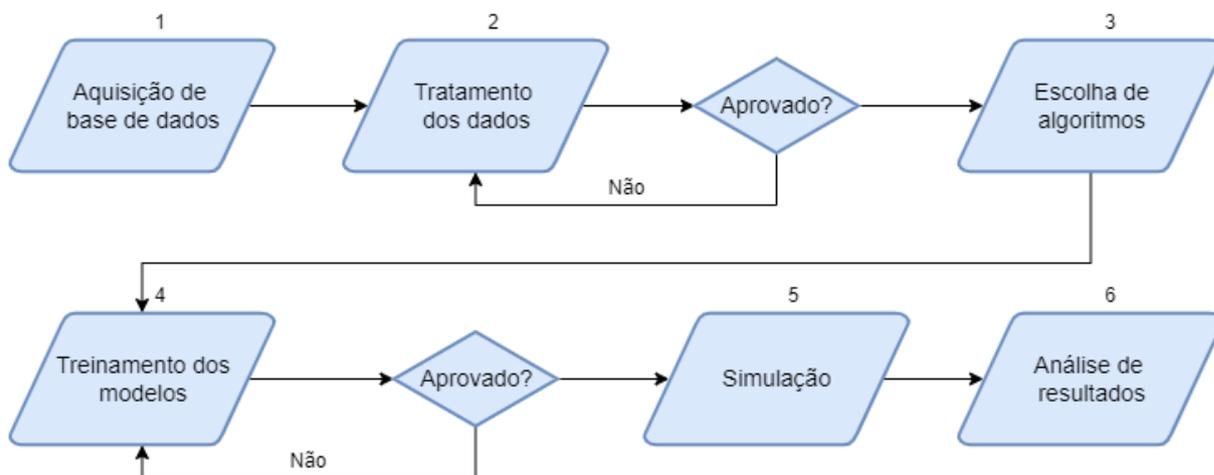
Como cada módulo foi desenvolvido para uma funcionalidade específica, é grande o número de aplicações que a biblioteca possui. Dentre possíveis aplicações, vale destacar as principais: Pré-processamento, classificação, regressão, clusterização, redução de dimensionalidade e ajuste de parâmetros (DIDÁTICA TECH, 2022c).

### 3 PROCEDIMENTOS EXPERIMENTAIS

A Figura 2, apresenta o fluxograma do desenvolvimento do trabalho, que foi dividido em 6 etapas principais: aquisição de base dados (1), tratamento de dados

(2), escolha de algoritmos (3), treinamento dos modelos (4), simulações (5) e análise dos resultados (6).

Figura 2: Fluxograma de atividades



Fonte: Do autor (2021)

A aquisição da base de dados com informações financeiras de pessoas físicas e jurídicas, foi a primeira etapa do desenvolvimento da solução proposta. A aquisição desta base ocorreu por intermédio da Deps Tecnologia, que cedeu as informações já anonimizadas em acordo com a Lei Geral de Proteção de Dados Pessoais (LGPD), Lei nº 13.709/2018. Após adquirir os dados foi constatado que estes já eram suficientes para o desenvolvimento da pesquisa, devido à grande quantidade de informações, sendo ao todo mais de 59 mil linhas com informações financeiras de pessoas físicas e jurídicas.

As informações que a base continha eram as seguintes: código do cliente, atraso matriz, prazo médio, total de cheques pagos, total de cheques vencidos, total de cheques a vencer, valor da última compra, data da última compra, valor da maior compra, data da maior compra, total de nota de crédito, total de nota de debito, total de títulos pagos, total de títulos vencidos, total de títulos a vencer, total de pedidos em aberto, valor total de protesto, quantidade total de protesto, valor total de pendências financeiras com instituições financeiras (refin), data do último pendências financeiras com instituições financeiras (refin), quantidade total de cheque sem fundo (ccf), data do último cheque sem fundo (ccf), valor total de pendências financeiras (pefin),



quantidade total de pendências financeiras (pefin), data da última pendência financeira (pefin) e se a pessoa era ou não inadimplente.

Para o desenvolvimento do projeto, foi utilizado a ferramenta Google Colab. Este é um serviço online disponibilizado de forma gratuita pela Google®. Ao acessar o site com a conta da Google®, o ambiente virtual é criado para o desenvolvimento dos projetos. Assim, foi possível seguir para o processo de engenharia de recurso da base.

O tratamento dos dados obtidos contemplou uma seleção, limpeza, enriquecimento e normalização. No processo de seleção, as informações foram agrupadas com base no formato dos dados que elas possuem. Assim, foram obtidos três grupos: data, quantidade e valores em reais. Como as informações referentes a data não eram precisas e possuíam muitas distorções, foi optado por desconsiderá-las e permanecer apenas com os outros dois grupos.

Com as informações selecionadas, a etapa de limpeza foi responsável por identificar quais informações possuíam dados ausentes e necessitavam atenção. Nesta etapa foram identificadas quatro informações que possuíam 48441 dados nulos, são elas: valor total de protesto, quantidade total de protesto, valor total de refin e quantidade total de cheque sem fundo. Ao analisar, foi constatado que as informações ausentes significavam que as pessoas relacionadas aos dados não possuíam, ou seja, poderiam ser alteradas de nulas para zero. Desta forma, na etapa de enriquecimento foram realizadas as substituições destes valores para zero.

O processo de normalização foi responsável por colocar os dados em intervalos de valores comuns, onde o intervalo utilizado foi de 0 a 1. Este processo foi essencial para que fosse possível trabalhar com todas as informações em conjunto e melhorar a acurácia modelo.

Desta forma, estes procedimentos foram responsáveis por avaliar as informações que deveriam ser utilizadas, dispensando as desnecessárias ou que não possuíam dados suficientes. As informações utilizadas foram: atraso matriz, prazo médio, valor de última compra, valor maior compra, total de nota de crédito, total de títulos pagos, total de títulos vencidos, total de títulos a vencer, total de pedidos em aberto, valor total de protestos, quantidade total de protestos, valor total de refin, quantidade total de ccf, valor total de pefin, quantidade total de pefin e inadimplência. Já as informações descartadas foram: código do cliente, data última compra, data



maior compra, data último pefin, data último ccf, data último protesto, data último refin, total de cheques pagos, total de cheques vencidos, total de cheques a vencer e total de notas de débito.

Com a etapa anterior já concluída, foi necessário efetuar a seleção dos algoritmos para busca de padrões e inferência dos dados. Os modelos escolhidos foram *LogisticRegression*, *SVC*, *KNeighborsClassifier* e *RandomForestClassifier*. O *LogisticRegression* tem como objetivo estimar a probabilidade de um evento ocorrer, já o *SVC* tem como foco se adaptar aos dados fornecidos e retornar um hiperplano que melhor categorize os dados. Assim como o *LogisticRegression*, o *KNeighborsClassifier* também tem como objetivo identificar o padrão da ocorrência dos dados e posteriormente classificá-lo, semelhante a esses, o *RandomForestClassifier* tem como foco identificar se um evento irá ou não ocorrer, mas diferente dos demais, ele utiliza um conjunto de árvores de decisão para posteriormente selecionar por meio de votação a melhor solução para a classificação da informação.

Dado a escolha dos modelos, selecionados com base no conhecimento sobre o assunto, foi utilizado o *GridSearch* para busca de melhores hiper parâmetros para cada um dos modelos. A utilização do *Gridsearch* facilitou a identificação dos ajustes dos parâmetros, visto que ele realiza uma série de combinações dos parâmetros e avalia os resultados, para na sequência armazená-los em um único objeto.

Para a solução atingir os resultados esperados, foi necessário efetuar a separação dos dados. Neste processo, o conjunto de dados foi dividido em treino e teste. Para essa divisão foi utilizado o método *hould-out*, que consiste em dividir a base de dados em tamanhos diferentes para treino e teste. A divisão ocorreu da seguinte forma: 70% para treino e 30% para teste. Isso significa dizer que os algoritmos treinaram com uma parcela dos dados para posteriormente testar os resultados com a outra parcela.

O passo seguinte consistiu em simular o processo completo para verificar a eficácia dos possíveis modelos e garantir que seria possível obter resultados. Nesta etapa foram realizadas em média dez simulações em diferentes cenários, ou seja, foi realizado diversos treinamentos e validações para analisar qual seria a melhor da



solução proposta. Este procedimento exigiu uma atenção especial para garantir que não houvesse falhas no processo.

Concluídas as etapas anteriores, foi realizada uma análise nos resultados obtidos para aferir se o objetivo do trabalho foi alcançado, ou seja, se foi possível obter uma informação segura referente a chance de a pessoa analisada tornar-se inadimplente. Para chegar a esta conclusão, foi utilizado métricas de classificação variando de 0 a 100%, onde zero significa o pior resultado possível e cem o melhor resultado. As métricas utilizadas foram acurácia, precisão, *recall* e *f1-score*. Esta análise foi essencial para verificar a necessidade de alterações nos processos e quais ações deveriam ser tomadas.

#### 4 RESULTADOS E DISCUSSÕES

Com os dados da base devidamente selecionados e tratados, foi necessário identificar qual o algoritmo que mais se adequa aos objetivos. Com a utilização do *GridSearch* foi possível não só identificar o melhor algoritmo como também suas configurações. Assim, o algoritmo que apresentou o melhor resultado foi *RandomForestClassifier* com os hiper parâmetros *max depth* (profundidade máxima das árvores) de 20 e *n estimators* (número total de árvores) de 1400, com o valor médio para acurácia 75%, precisão 76%, *recall* 75% e *f1-score* 75%.

Para uma análise mais profunda em relação aos resultados obtidos com o melhor modelo, foi necessário avaliar os resultados obtidos com as métricas de avaliação de modelos para classificação, analisando classes e respectivas métricas. O resultado mais detalhado é possível visualizar na Tabela 1.

Tabela 1: Resultados das métricas detalhadas.

	Precisão	Recall	F1-Score	Acurácia
Não inadimplente	72%	82%	77%	75%
Inadimplente	79%	69%	73%	
Média	76%	75%	75%	

Fonte: Do autor (2022)



Os resultados obtidos na métrica precisão, Equação 1, indicam a porcentagem de previsões positivas corretas em relação ao total de previsões positivas. Já os valores obtidos na métrica *Recall*, Equação 2, representam a porcentagem de previsões positivas corretas em relação ao total de positivos reais, enquanto a métrica *F1-Score*, Equação 3, é uma média harmônica ponderada de precisão e *Recall*, onde quanto mais perto de 100%, melhor é o modelo. A métrica acurácia, Equação 4 mede a acurácia de todas as previsões, sejam elas positivas ou negativas.

$$P = \frac{VP}{VP+FP} \quad (1)$$

$$R = \frac{VP}{VP+FN} \quad (2)$$

$$F1 = \frac{2 \times P \times R}{P+R} \quad (3)$$

$$AC = \frac{VP+VN}{VP+VN+FP+FN} \quad (4)$$

Onde:

AC = acurácia

P = precisão

R = Recall

F1 = F1-score

VP = verdadeiros positivos

VN = verdadeiros negativos

FP = falsos positivos

FN = falsos negativos

Ao analisar a métrica precisão, é possível concluir que todas as pessoas que o modelo previu que não seriam inadimplentes, 72% realmente não eram. Em relação aos clientes que ele previu que seriam inadimplentes, temos um acerto um pouco maior, visto que o modelo acertou 79% das análises. Assim, é possível concluir que a média destes dois resultados foi de 76% e a média ponderada de 76%.



Com os resultados da métrica *Recall*, é possível identificar que de todos os clientes que são inadimplentes, o modelo só conseguiu prever corretamente para 69% deles. Já os clientes que são não inadimplentes, o sistema conseguiu prever corretamente 82%. Com isso, a média dos resultados é de 75% e a média ponderada é de 75%.

Assim, com as análises das métricas precisão e *Recall*, é possível avaliar os resultados da métrica *F1-Score*. Para os clientes inadimplentes, o modelo apresentou um resultado de 73% e para os clientes não inadimplentes o resultado foi de 77%. Como o objetivo era identificar inadimplentes, o resultado que possui maior relevância é 73%.

Para analisar o resultado obtido na métrica acurácia, primeiro é necessário entender o cálculo que ela realiza. Conforme é possível verificar na Equação 4, que representa o cálculo da acurácia, quanto maior for o valor obtido, melhor será a acurácia. Assim, é possível avaliar que o resultado de 75% é positivo para o modelo.

Para entender melhor os resultados obtidos, podemos observar os resultados contidos na matriz de confusão representados na Tabela 2. A matriz de confusão é uma tabela que permite extrair métricas que auxiliam na avaliação de modelos de *machine learning* para classificação. As duas colunas representam os resultados obtidos por evento, já as linhas representam os resultados esperados por evento. Através destas informações, é possível evidenciar a quantidade de itens classificados de forma incorreta e qual informação gera este erro.

Tabela 2: Resultados obtidos através da Matriz de Confusão.

		Valor Predito	
		Não inadimplente	Inadimplente
Real	Não inadimplente	Verdadeiro Negativo 41,09%	Falso Positivo 9,02%
	Inadimplente	Falso Negativo 15,71%	Verdadeiro Positivo 34,18%

Fonte: Do autor (2022)

Ao analisar as informações contidas na Tabela 2, é possível verificar que temos a comparação do valor previsto pelo modelo com o valor real, conforme os nomes sugerem. Os verdadeiros negativos são aqueles que o valor real era negativo

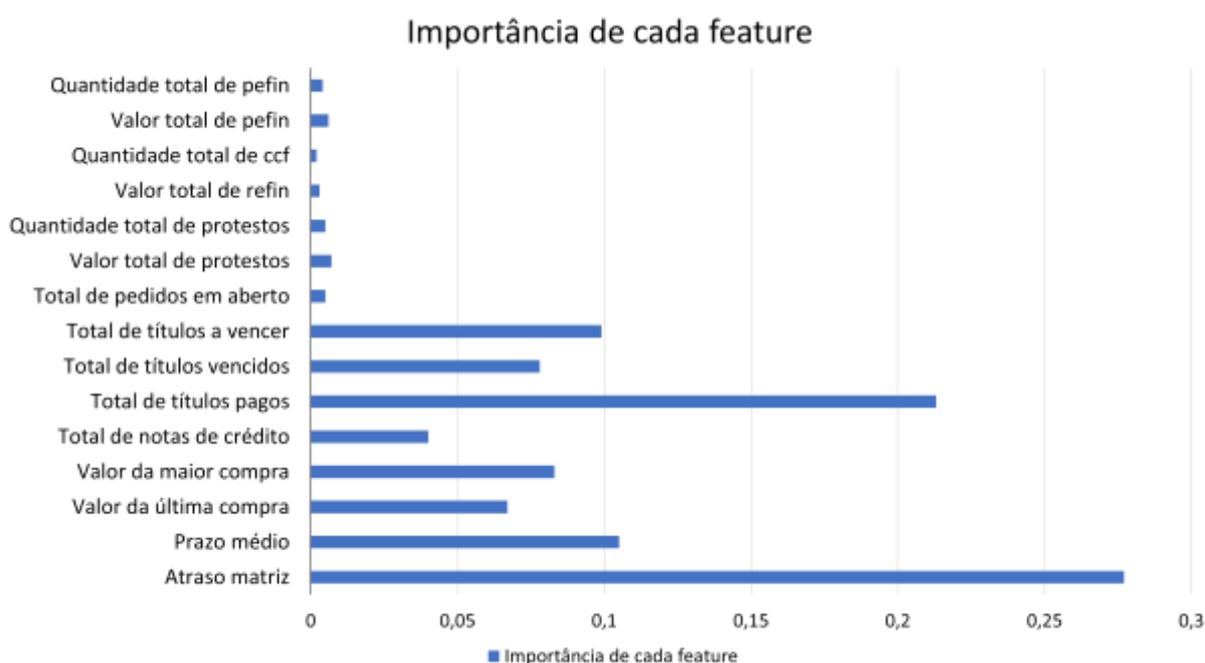


e o valor previsto também foi negativo, ou seja, o modelo classificou corretamente. Já os verdadeiros positivos são aqueles que o valor real era positivo e o modelo também classificou como positivo, isto é, o modelo acertou. Os falsos negativos são aqueles que o valor real era positivo, mas o modelo classificou como negativo. Assim, os falsos positivos são aqueles que o resultado correto era negativo, mas o modelo os classificou como positivos, ou seja, assim como nos falsos negativos, o modelo errou.

A partir da análise das informações da tabela, fica evidente o espaço para melhorias na abordagem proposta, visto que em médias as métricas apontam cerca de 75% de acerto. Um dos caminhos possíveis para um aumento na performance do modelo, é a aquisição de uma base de dados mais balanceada, onde o número de informações sobre clientes inadimplentes e não inadimplentes seja mais próximo.

Para obter o resultado se o cliente será ou não inadimplente, todas as informações são levadas em consideração, mas elas possuem graus de importância diferentes entre elas. Assim, para avaliar a importância de cada uma, é necessário levar em consideração os resultados obtidos através do método Importância de Características do modelo utilizado para classificação, representados na Figura 3.

Figura 3: Resultados obtidos através da Importância de Características do modelo.



Fonte: Do autor (2022)



Conforme é possível identificar na Figura 3, o atraso matriz dos clientes é a informação que possui mais peso no momento de identificar se o cliente irá tornar-se inadimplente ou não. Esta informação faz muito sentido ao entender o cálculo realizado para obter este valor, visto que ele é o resultado da divisão do valor de títulos vencidos pelo valor dos títulos pagos multiplicado pela quantidade de dias que estão vencidos, ou seja, clientes inadimplentes possuem títulos vencidos à uma grande quantidade de dias, o que faz com que o atraso matriz seja maior.

Outra informação relevante, é o total de títulos pagos. Assim como no atraso matriz, esta informação apresenta uma relação direta com a possibilidade de inadimplência. Clientes com uma maior quantidade de títulos pagos apresentam uma possibilidade menor de tornarem-se inadimplentes, visto que a tendência deles é manter o comportamento de pagamento com as empresas.

O prazo médio representa a média de dias que o cliente leva para pagar os títulos que possui. Assim, clientes com histórico de inadimplência possuem um prazo médio elevado, visto que levam um período maior para quitarem suas dívidas. Esta informação não possui uma importância grande quando comparada com atraso matriz e total de títulos pagos, porque alguns clientes possuem o hábito de pagarem seus títulos com um prazo maior, ou seja, mesmo com um prazo médio alto, não chegam a serem considerados inadimplentes.

O valor total de títulos a vencer também possui uma importância alta nas análises realizadas. Clientes com um alto valor de títulos a vencer possuem uma possibilidade maior de tornarem-se inadimplentes, mas como clientes que não são inadimplentes podem possuir o mesmo valor de títulos a vencer, este cenário acaba não tornando-se uma regra. Assim, o valor de títulos a vencer continua relevante na análise, mas com uma importância menor em relação às outras informações analisadas.

## 5 CONCLUSÕES

A utilização de técnicas de inteligência artificial para identificar clientes com potencial a tornarem-se inadimplentes, apresentou resultados consideráveis. Para isso, entender os critérios de classificação de análise de crédito foi essencial para identificar as informações importantes e obter resultados melhores.



Identificar padrões de inadimplência na base de dados possibilitou que relações entre o atraso matriz do cliente e a chance de inadimplência fossem estabelecidas. Clientes com um atraso matriz maior são mais propensos a virar inadimplentes com a empresa, isso não ocorre na relação entre valor de títulos pagos com inadimplência, onde a relação entre eles é inversamente proporcional. Assim, foi possível modelar um cenário de aprendizado de máquina para análise de crédito mais preciso e seguro em suas análises, extraindo resultados expressivos e grande importância para as empresas.

De fato, substituir as pessoas por inteligências artificiais nas tomadas de decisões não é um objetivo, mas sim provar que ferramentas baseadas em inteligência artificial podem auxiliar nas análises, identificando padrões e comportamentos que passariam despercebidos. Os resultados demonstram um grande potencial do uso das técnicas de inteligência artificial e mostram que quanto mais informações detalhadas as empresas possuem de seus clientes, maior será a precisão destas ferramentas.

Ao refletir sobre a evolução deste trabalho, podem ser apontados para trabalhos futuros utilizar técnicas de balanceamento de classes, ou até mesmo, buscar novas fontes de dados para a classe dita como “inadimplente”. Outro ponto para melhorias, seria a utilização de modelos considerados profundos (*Deep Learning*) para a avaliação do problema.

## REFERÊNCIAS

CARNEIRO, Éric Cavalcante. **Análise de crédito no agronegócio: um estudo de caso**. 2017. 114 f. Trabalho de Conclusão de Curso (graduação em Administração) - Universidade Federal do Ceará, Faculdade de Economia, Administração, Atuária e Contabilidade, Fortaleza, 2017.

Dantas, G., Godoj, L. F., & Neto, S. P. (2017). FERRAMENTA PARA AUTOMATIZAÇÃO DO PROCESSO DE CONCESSÃO DE CRÉDITO UTILIZANDO DE CONCEITOS DE INTELIGÊNCIA ARTIFICIAL. *Ensaio USF*, 1(1), 117–129. Disponível em: <<https://doi.org/10.24933/eusf.v1i1.63>> Acessado em: 15 Mar 2022.

HONDA, Hugo; et al. Os três tipos de aprendizado de máquina. **LAMFO**, 2017. Disponível em: <<https://lamfo-unb.github.io/2017/07/27/tres-tipos-am/>> Acessado em: 16 Mai 2021.



MARQUES, Flávia. Análise de crédito: o que é e como funciona o processo?. **Creditas**, 2020. Disponível em: <<https://www.creditas.com/exponencial/analise-de-credito/>> Acessado em: 16 Mai 2021.

Mendes, P., Machiavelli, J., & Gusmão, C. (2019). Avaliação de tecnologias de recomendação de conteúdo baseadas em Inteligência Artificial numa visão educacional. In 25º CIAED Congresso Internacional ABED de Educação a Distância. Disponível em: <<http://www.abed.org.br/congresso2019/anais/trabalhos/31352.pdf>> Acessado em: 15 Mar 2022.

PANT, Ayush. Introduction to Machine Learning for Beginners. **Towards data science**, 2019. Disponível em: <<https://towardsdatascience.com/introduction-to-machine-learning-for-beginners-eed6024fdb08>> Acessado em: 05 Jun 2021.

PACHECO, André. Introdução ao Scikit-learn - Parte 1: uma visão geral. **Computação Inteligente**, 2021. Disponível em: <<http://computacaointeligente.com.br/outros/intro-sklearn-part-1/>> Acessado em: 18 Jun 2022.

ROVEDA, Ugo. O que é Python, para que serve e por que aprender. **Kenzie**, 2020. Disponível em: <<https://kenzie.com.br/blog/o-que-e-python/>> Acessado em: 05 Jun 2021.

Estatísticas Monetárias e de Crédito. **Banco Central do Brasil**, 2021. Disponível em: <[https://www.bcb.gov.br/content/estatisticas/docs\\_estatisticasmonetariascredito/Nota%20para%20a%20imprensa%20-%20Estat%20C3%ADsticas%20Monet%20e%20de%20Cr%20C3%A9dito.pdf](https://www.bcb.gov.br/content/estatisticas/docs_estatisticasmonetariascredito/Nota%20para%20a%20imprensa%20-%20Estat%20C3%ADsticas%20Monet%20e%20de%20Cr%20C3%A9dito.pdf)> Acessado em: 15 Abr 2021.

Demora na análise de crédito. **Deps Tecnologia**, 2018. Disponível em: <<https://deps.com.br/demora-na-analise-de-credito/>> Acessado em: 16 Mai 2021.

Software de análise de crédito: o que você tem que saber antes de escolher?. **Deps Tecnologia**, 2019. Disponível em: <<https://deps.com.br/software-de-analise-de-credito-o-que-voce-tem-que-saber-antes-de-escolher/>> Acessado em: 05 Jun 2021.

A biblioteca scikit-learn – Python para machine learning. **Didática Tech**, 2022c. Disponível em: <<https://didatica.tech/a-biblioteca-scikit-learn-pyhton-para-machine-learning/>> Acessado em: 18 Jun 2022.

Machine Learning e Ciência de dados com IBM Watson. **IBM**, 2021c. Disponível em: <<https://www.ibm.com/br-pt/analytics/machine-learning>> Acessado em: 16 Mai 2021.

O que é inteligência artificial – IA?. **Oracle**, 2021c. Disponível em: <<https://www.oracle.com/br/artificial-intelligence/what-is-ai/>> Acessado em: 16 Mai 2021.



5 tecnologias de IA que você precisa conhecer. **SAS**, 2021c. Disponível em: <[https://www.sas.com/pt\\_br/insights/articles/analytics/five-ai-technologies.html](https://www.sas.com/pt_br/insights/articles/analytics/five-ai-technologies.html)> Acessado em: 16 Mai 2021.

Análise de crédito: um guia completo para desenvolver análises precisas. **Serasa Experian**, 2020. Disponível em: <<https://empresas.serasaexperian.com.br/blog/con1-analise-de-credito-precisa/>> Acessado em: 16 Mai 2021.

O que é Inteligência artificial? Como funciona, exemplos e aplicações. **TOTVS**, 2019. Disponível em: <<https://www.totvs.com/blog/inovacoes/o-que-e-inteligencia-artificial/>> Acessado em: 05 Mai 2021.

### AGRADECIMENTOS

Gostaria de agradecer primeiramente a minha família por todo o apoio prestado durante a realização deste trabalho, sem eles nada disso seria possível. Agradeço também o apoio do coordenador e orientador para realização do projeto. Por último, mas não menos importante, gostaria de agradecer a SATC e a Deps Tecnologia por terem me incentivado e fornecido todo o apoio necessário para a conclusão do trabalho.