



Anais do Simpósio Acadêmico de Engenharia de Produção (SAEPRO) da EEL-USP

IX SAEPRO – 25 e 26 de novembro de 2025

APRENDIZADO DE MÁQUINA PARA ANÁLISE DE CRÉDITO DE RISCO EM UMA INSTITUIÇÃO FINANCEIRA

Ana Maria Ramos Da Silva

Universidade Estadual Paulista (ana.r.silva@unesp.br)

José Roberto Dale Luche

Universidade Estadual Paulista (dale.luche@unesp.br)

Fernando Augusto Silva Marins

Universidade Estadual Paulista (fernando.marins@unesp.br)

Aneirson Francisco da Silva

Universidade Estadual Paulista (aneirson.silva@unesp.br)

Resumo

A aprendizagem de máquina tem sido amplamente empregada em diferentes domínios para extrair padrões e apoiar processos de tomada de decisão baseados em dados. No contexto financeiro, o risco de crédito representa um dos principais desafios enfrentados pelas instituições, exigindo métodos cada vez mais robustos para a avaliação da inadimplência. Este estudo tem como objetivo aplicar e comparar algoritmos de aprendizagem de máquina na análise de crédito de risco, visando aprimorar a gestão e a precisão das previsões. Foram implementados modelos supervisionados e avaliados por meio da matriz de confusão e da métrica de acurácia. Os resultados indicam que o modelo de Máquina de Vetores de Suporte (SVM) apresentou desempenho superior em relação aos demais algoritmos testados, evidenciando seu potencial para apoiar a classificação de clientes e otimizar decisões de concessão de crédito.

Palavras-chave: Aprendizagem de Máquina, Crédito de risco, Análise de Crédito de Risco, Máquina de Vetores de Suporte (SVM).



Anais do Simpósio Acadêmico de Engenharia de Produção (SAEPRO) da EEL-USP

IX SAEPRO – 25 e 26 de novembro de 2025

Abstract

Machine learning has been widely applied across various domains to extract patterns and support data-driven decision-making processes. In the financial sector, credit risk stands out as one of the major challenges faced by institutions, demanding increasingly robust methods for default assessment. This study aims to apply and compare machine learning algorithms for credit risk analysis in order to enhance risk management and prediction accuracy. Supervised models were implemented and evaluated using the confusion matrix and accuracy metrics. The results indicate that the Support Vector Machine (SVM) model achieved superior performance compared to the other algorithms tested, highlighting its potential to support client classification and optimize credit granting decisions.

Keywords: Machine Learning, Credit Risk, Credit Risk Analysis, Support Vector Machine (SVM).

Resumen

El aprendizaje automático se ha aplicado ampliamente en diversos ámbitos para extraer patrones y respaldar procesos de toma de decisiones basados en datos. En el sector financiero, el riesgo crediticio constituye uno de los principales desafíos que enfrentan las instituciones, lo que exige métodos cada vez más sólidos para la evaluación de la morosidad. Este estudio tiene como objetivo aplicar y comparar algoritmos de aprendizaje automático en el análisis del riesgo crediticio, con el fin de mejorar la gestión y la precisión de las predicciones. Se implementaron modelos supervisados y se evaluaron mediante la matriz de confusión y la métrica de exactitud. Los resultados muestran que el modelo de Máquina de Vectores de Soporte (SVM) presentó un desempeño superior en comparación con los demás algoritmos analizados, lo que evidencia su potencial para apoyar la clasificación de clientes y optimizar las decisiones de concesión de crédito.

Palabras clave: Aprendizaje automático, riesgo crediticio, análisis de riesgo crediticio, máquina de vectores de soporte (SVM).



1. INTRODUÇÃO

A análise de risco de crédito ganhou destaque após a crise financeira global de 2008. Segundo Leo et al. (2019), há um foco aprimorado na detecção, mensuração e reporte dos riscos, impulsionado pela significativa influência da aprendizagem automática, com soluções já implementadas e outras ainda em fase de exploração.

Instituições financeiras e bancos enfrentam uma gama diversificada de riscos, incluindo risco de crédito, risco de mercado, risco operacional, risco de liquidez e risco de taxa de juros (Vashisht; Sarva; Mundi, 2022).

O risco de crédito destaca-se como um tema de grande relevância, uma vez que, ao conceder um empréstimo, torna-se essencial realizar uma avaliação criteriosa das probabilidades de recuperação dos recursos financeiros (Mhlanga, 2021). Nesse contexto, o principal objetivo da análise de risco de crédito é classificar os clientes em dois grupos distintos: bons e maus pagadores (Ha; Nguyen, 2016).

A análise de risco de crédito é realizada por meio de métodos que utilizam dados provenientes de diversas fontes. No entanto, essa abordagem apresenta algumas limitações, como questões relacionadas à privacidade dos dados e aos custos de comunicação (Oualid; Maleh; Moumun, 2023).

Pesquisas também exploraram a aplicação de técnicas de aprendizado de máquina em diferentes contextos. Namioka et al. (2025) aplicaram modelos supervisionados para prever a rotatividade de clientes no setor bancário, enquanto Fioreto et al. (2025) investigaram a predição da temperatura do rotor em motores PMSM com o uso de regressões polinomiais e aprendizado profundo. Esses estudos reforçam a abrangência



Anais do Simpósio Acadêmico de Engenharia de Produção (SAEPRO) da EEL-USP

IX SAEPRO – 25 e 26 de novembro de 2025

das aplicações de aprendizado de máquina e sua relevância tanto para a gestão organizacional quanto para sistemas industriais.

Diante desse contexto, este estudo propõe o uso de técnicas de aprendizagem de máquina como uma alternativa para aprimorar a análise e a gestão do risco de crédito. O estudo teve a seguinte questão de pesquisa como direcionador: Como o aprendizado de máquina pode transformar a análise de crédito de risco para melhorar a precisão na identificação de bons e maus pagadores?

1.1.Objetivos

Este trabalho possui o objetivo de aplicar o aprendizado de máquina na análise de crédito de risco em instituições financeiras, a fim de melhorar a gestão de risco. Especificamente, comparar diferentes algoritmos de aprendizado de máquina para identificar o método mais eficaz na previsão de inadimplência e analisar o impacto de variáveis demográficas na precisão da previsão da inadimplência.

1.2. Justificativas

A análise de risco de crédito é fundamental para as instituições financeiras, pois contribui para a prevenção de perdas e a maximização de lucros. Como a análise de risco depende da tomada de decisões baseadas em informações, a aprendizagem de máquina surge como um caminho promissor para o desenvolvimento de novos métodos e tecnologias que auxiliem nesse processo decisório e, além disso, nos últimos anos, tem-se observado uma crescente adoção de métodos de aprendizado de máquina para diversas tarefas relacionadas ao gerenciamento de risco (Mashrur et al., 2020).



Anais do Simpósio Acadêmico de Engenharia de Produção (SAEPRO) da EEL-USP

IX SAEPRO – 25 e 26 de novembro de 2025

A Inteligência Artificial já está incluída no cotidiano e está se desenvolvendo em todas as áreas. Diante disso, com a competitividade das empresas, as instituições financeiras buscam melhorar a experiência do cliente com o uso da análise de dados para gerenciar os riscos. Portanto, a aprendizagem de máquina vem para reduzir a incerteza dos riscos, pois permite a análise de grandes volumes de dados (Noriega et al., 2023).

Foi realizada uma pesquisa bibliométrica dos artigos publicados entre os anos de 1995 e 2024 na base *Scopus* por meio das expressões de busca na Figura 1.

Figura 1 – Pesquisa de artigos de “*Machine Learning*” AND “*risk credit*” na base Scopus.

Search within Article title, Abstract, Keywords	Search documents * "machine learning"
AND	
Search within Article title, Abstract, Keywords	Search documents "credit risk"

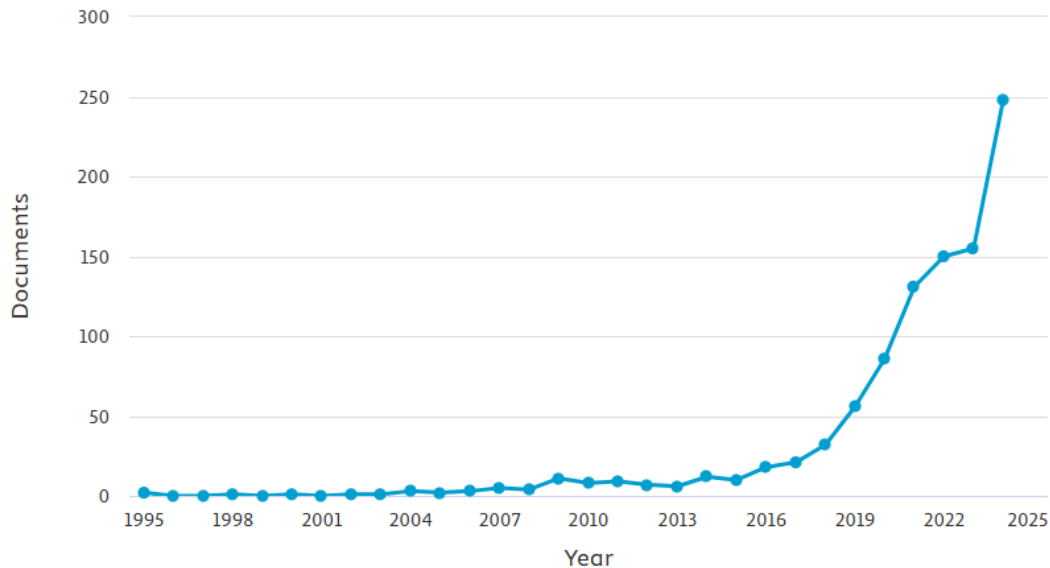
Na Figura 2, as publicações encontradas com os filtros utilizados ocorreram entre os anos de 1995 e 2024. No ano de 2024 foram encontrados 250 documentos.

Figura 2 – Documentos por ano.



Anais do Simpósio Acadêmico de Engenharia de Produção (SAEPRO) da EEL-USP

IX SAEPRO – 25 e 26 de novembro de 2025



Fonte: Scopus(2025).

Este trabalho está estruturado da seguinte forma: A Seção 2 apresenta os conceitos básicos de aprendizagem de máquina, e crédito de risco. A Seção 3 descreve a base de dados utilizada, métodos e ferramentas utilizadas. Na Seção 4 é apresentado o desenvolvimento do projeto. Na Seção 5 encontram-se os resultados e discussões, por fim, a Seção 6 traz a conclusão e recomendações de pesquisas futuras.

2. FUNDAMENTAÇÃO TEÓRICA

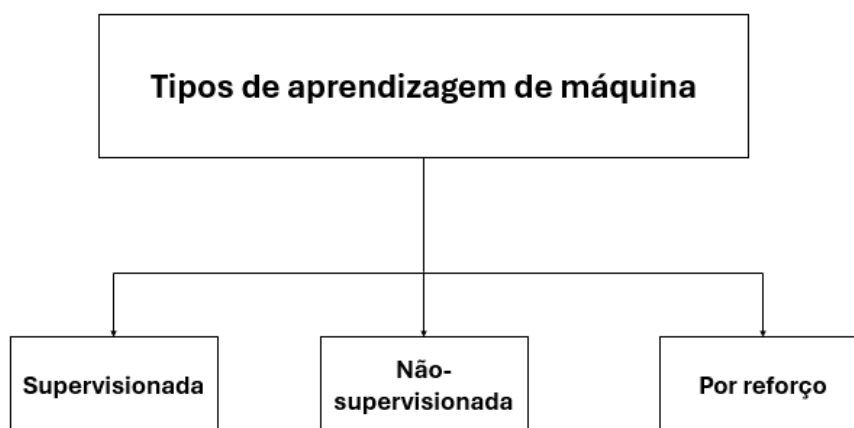
2.1. Aprendizagem de Máquina

A aprendizagem de máquina é utilizada em diversas áreas como, por exemplo, publicidade, sistemas de recomendação, visão computacional, processamento de linguagem natural e análise de comportamento do usuário. O motivo da ampla aplicabilidade da aprendizagem de máquina é devido ao seu alto desempenho de análise de dados (Yang; Shami, 2020).

De forma simplificada, o objetivo do aprendizado de máquina é aprender, de maneira automática, relacionamentos e padrões a partir de exemplos e observações. Essa área do conhecimento visa automatizar a construção de modelos analíticos por meio da aplicação de algoritmos que aprendem com dados específicos de um problema. Dessa forma, o computador é capacitado a identificar padrões complexos por conta própria (Janiesch et al., 2021).

Os três principais paradigmas de aprendizado de máquina são: o supervisionado, o não supervisionado e o por reforço, como está ilustrado na Figura 3.

Figura 3: Tipos de Aprendizagem de Máquina



2.2. Risco de Crédito

Após a crise financeira global de 2008, a gestão de riscos ganhou maior destaque no setor bancário. Diariamente, os bancos enfrentam diversas categorias de risco, como: risco de taxa de juros, risco de mercado, risco de crédito, risco cambial, risco de liquidez, risco de insolvência, risco-país (ou soberano), risco operacional e tecnológico, e riscos fora do balanço patrimonial. Dentre esses, o risco de crédito representa a maior ameaça para um banco, uma vez que se refere à possibilidade de perdas decorrentes do não cumprimento



de obrigações contratuais por parte de seus clientes. Para mensurá-lo, cada instituição financeira pode desenvolver seu próprio modelo, cujos principais parâmetros a serem estimados são a Probabilidade de Inadimplência (PD), a Perda dada a Inadimplência (LGD) e a Exposição na Inadimplência (EAD) (Leo et al., 2019).

3. MATERIAIS E MÉTODOS

Para a realização deste estudo, foi necessário o uso de diversas ferramentas e plataformas tecnológico-computacionais. A seguir, detalham-se os recursos utilizados:

a) A plataforma Kaggle foi utilizada na fase inicial da pesquisa. Trata-se de um repositório público de dados e competições de ciência de dados. Por meio dela, obteve-se o dataset de análise bancária utilizado no trabalho, além de ter sido possível acessar fóruns e discussões que auxiliaram na execução do estudo.

b) A linguagem de programação Python foi escolhida para o desenvolvimento deste trabalho devido ao seu amplo reconhecimento e aplicação nas áreas de ciência de dados e aprendizado de máquina. A codificação envolve etapas de pré-processamento dos dados, aplicação de modelos de classificação e análise dos resultados.

c) O Jupyter Notebook foi o ambiente computacional utilizado para criar e executar os códigos. Trata-se de um aplicativo web de código aberto que permite criar e compartilhar documentos interativos, suportando diversas linguagens de programação, incluindo Python (JUPYTER, 2025).

d) A biblioteca Scikit-learn foi empregada para a modelagem e análise dos dados. Esse framework de aprendizado de máquina para Python forneceu as ferramentas necessárias para implementar e avaliar os modelos de classificação: Regressão Logística, Máquina de Vetores de Suporte (Support Vector Machine - SVM) e Floresta Aleatória (Random Forest).

e) O Excel será usado para realizar a análise de correlação da base de dados.

f) Modelos de Classificação: Foram utilizados e comparados os seguintes modelos de aprendizado de máquina supervisionado, com o objetivo de prever a variável-alvo e comparar suas performances com base na métrica de acurácia:

- Regressão Logística: Modelo linear utilizado para problemas de classificação binária.
- Máquina de Vetores de Suporte (SVM): Algoritmo que encontra um hiperplano ótimo para separação entre classes.
- Floresta Aleatória (Random Forest): Método de ensemble que combina múltiplas árvores de decisão para melhorar a previsão.

3.1 SELEÇÃO E EXPLORAÇÃO DA BASE DE DADOS

Para o estudo da análise de risco de crédito, foi escolhida uma base de dados contendo informações sobre 30.000 clientes. O conjunto de dados contém uma série de variáveis essenciais para a compreensão do risco de crédito.

A base é binária e inclui as seguintes informações, sendo NT as notas taiwanesas e inclui as informações a seguir.

- **LIMIT_BAL**: Valor do crédito fornecido em dólares NT (inclui crédito individual e familiar/suplementar).
- **SEXO**: Gênero (1=masculino, 2=feminino).
- **EDUCAÇÃO**: (1=pós-graduação, 2=universidade, 3=ensino médio, 4=outros, 5=desconhecido, 6=desconhecido).
- **CASAMENTO**: Estado civil (1=casado, 2=solteiro, 3=outros).
- **IDADE**: Idade em anos.
- **PAY_0**: Situação do pagamento em setembro de 2005 (-1=pagar em dia, 1=atraso no pagamento por um mês, 2=atraso no pagamento por dois meses, ... 8=atraso no pagamento por oito meses, 9=atraso no pagamento por nove meses ou mais).

- **PAY_2:** Situação de reembolso em agosto de 2005.
- **PAY_3:** Situação de reembolso em julho de 2005.
- **PAY_4 :** Situação de reembolso em junho de 2005.
- **PAY_5:** Situação de reembolso em maio de 2005.
- **PAY_6:** Situação de reembolso em abril de 2005.
- **BILL_AMT1:** Valor da fatura em setembro de 2005 (dólar NT).
- **BILL_AMT2:** Valor da fatura em agosto de 2005.
- **BILL_AMT3:** Valor da fatura em julho de 2005.
- **BILL_AMT4:** Valor da fatura em junho de 2005.
- **BILL_AMT5:** Valor da fatura em maio de 2005.
- **BILL_AMT6:** Valor da fatura em abril de 2005.
- **PAY_AMT1:** Valor do pagamento anterior em setembro de 2005 (dólar NT).
- **PAY_AMT2:** Valor do pagamento anterior em agosto de 2005.
- **PAY_AMT3:** Valor do pagamento anterior em julho de 2005.
- **PAY_AMT4:** Valor do pagamento anterior em junho de 2005.
- **PAY_AMT5:** Valor do pagamento anterior em maio de 2005.
- **PAY_AMT6:** Valor do pagamento anterior em abril de 2005.

Foi necessário realizar a limpeza da base de dados. A biblioteca Pandas é disponibilizada gratuitamente e é amplamente utilizada na análise de dados, sendo reconhecida por sua praticidade e eficiência. Ela oferece funcionalidades como limpeza, transformação e análise de conjuntos de dados.

3.2. BASE DE TREINO E DE TESTE

O próximo passo na aplicação dos modelos de aprendizado de máquina é dividir a base de dados em dois conjuntos: treino e teste. Cada um desempenha um papel fundamental no processo. O conjunto de treino é utilizado para ajustar o modelo, enquanto o conjunto de teste avalia seu desempenho.

Os modelos de aprendizado de máquina utilizados neste estudo foram implementados em Python, e seus resultados são comparados para identificar o mais eficiente.

3.3. MATRIZ DE CONFUSÃO

A matriz de confusão é uma ferramenta fundamental em aprendizado de máquina para avaliar o desempenho de modelos de classificação, permitindo a comparação entre os valores previstos pelo algoritmo e os valores reais observados (ZENG, 2018).

A matriz de confusão possui algumas medidas de precisão do modelo de previsão. A usada neste trabalho será a acurácia que representa a razão entre as instâncias previstas corretamente (Chicco e Jurman, 2021). É dada pela fórmula a seguir.

$$\text{Acurácia} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

4. EXPERIMENTO

4.1 ANÁLISE DE CORRELAÇÃO

Para o impacto de variáveis demográficas e comportamentais na precisão da previsão de risco de crédito foi realizada uma análise de correlação. Esta análise envolveu a construção de uma matriz de correlação utilizando a ferramenta Microsoft Excel. O resultado desta análise é apresentado na Figura 4.

Figura 4 - Matriz de Correlação

Nomes	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	PAY_0	PAY_2	PAY_3	PAY_4	PAY_5	PAY_6	default.payment.next.month
LIMIT_BAL	1,000	0,025	-0,219	-0,108	0,145	-0,271	-0,296	-0,286	-0,267	-0,249	-0,235	-0,154
SEX	0,025	1,000	0,014	-0,031	-0,091	-0,058	-0,071	-0,066	-0,060	-0,055	-0,044	-0,040
EDUCATION	-0,219	0,014	1,000	-0,143	0,175	0,105	0,122	0,114	0,109	0,098	0,082	0,028
MARRIAGE	-0,108	-0,031	-0,143	1,000	-0,414	0,020	0,024	0,033	0,033	0,036	0,034	-0,024
AGE	0,145	-0,091	0,175	-0,414	1,000	-0,039	-0,050	-0,053	-0,050	-0,054	-0,049	0,014
PAY_0	-0,271	-0,058	0,105	0,020	-0,039	1,000	0,672	0,574	0,539	0,509	0,475	0,325
PAY_2	-0,296	-0,071	0,122	0,024	-0,050	0,672	1,000	0,767	0,662	0,623	0,576	0,264
PAY_3	-0,286	-0,066	0,114	0,033	-0,053	0,574	0,767	1,000	0,777	0,687	0,633	0,235
PAY_4	-0,267	-0,060	0,109	0,033	-0,050	0,539	0,662	0,777	1,000	0,820	0,716	0,217
PAY_5	-0,249	-0,055	0,098	0,036	-0,054	0,509	0,623	0,687	0,820	1,000	0,817	0,204
PAY_6	-0,235	-0,044	0,082	0,034	-0,049	0,475	0,576	0,633	0,716	0,817	1,000	0,187
BILL_AMT1	0,285	-0,034	0,024	-0,023	0,056	0,187	0,235	0,208	0,203	0,207	0,207	-0,020
BILL_AMT2	0,278	-0,031	0,019	-0,022	0,054	0,190	0,235	0,237	0,226	0,227	0,227	-0,014
BILL_AMT3	0,283	-0,025	0,013	-0,025	0,054	0,180	0,224	0,227	0,245	0,243	0,241	-0,014
BILL_AMT4	0,294	-0,022	0,000	-0,023	0,051	0,179	0,222	0,227	0,246	0,272	0,266	-0,010
BILL_AMT5	0,296	-0,017	-0,008	-0,025	0,049	0,181	0,221	0,225	0,243	0,270	0,291	-0,007
BILL_AMT6	0,290	-0,017	-0,009	-0,021	0,048	0,177	0,219	0,222	0,239	0,263	0,285	-0,005
PAY_AMT1	0,195	0,000	-0,037	-0,006	0,026	-0,079	-0,081	0,001	-0,009	-0,006	-0,001	-0,073
PAY_AMT2	0,178	-0,001	-0,030	-0,008	0,022	-0,070	-0,059	-0,067	-0,002	-0,003	-0,005	-0,059
PAY_AMT3	0,210	-0,009	-0,040	-0,004	0,029	-0,071	-0,056	-0,053	-0,069	0,009	0,006	-0,056
PAY_AMT4	0,203	-0,002	-0,038	-0,013	0,021	-0,064	-0,047	-0,046	-0,043	-0,058	0,019	-0,057
PAY_AMT5	0,217	-0,002	-0,040	-0,001	0,023	-0,058	-0,037	-0,036	-0,034	-0,033	-0,046	-0,055
PAY_AMT6	0,220	-0,003	-0,037	-0,007	0,019	-0,059	-0,037	-0,036	-0,027	-0,023	-0,025	-0,053
default.payment	-0,154	-0,040	0,028	-0,024	0,014	0,325	0,264	0,235	0,217	0,204	0,187	1,000

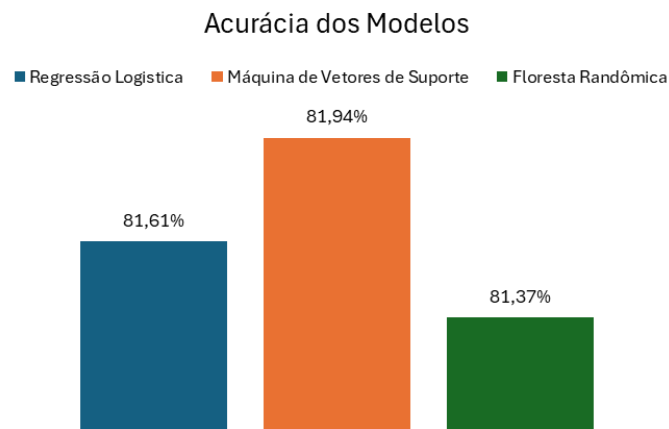
Devido ao volume de dados, foi necessário limitar a exibição da matriz de correlação para apenas as colunas *Limit_Bal*, *Sex*, *Education*, *Marriage*, *Age*, *Pay_0* a *Pay_6* e *default.payment.next.month*.

A partir da Figura 4, percebe-se uma baixa correlação entre todas as variáveis. Portanto, nenhuma das variáveis demográficas apresentam correlações fortes com a inadimplência. O que significa que tais variáveis não são determinantes no comportamento dos clientes da instituição financeira.

5. RESULTADO E DISCUSSÃO

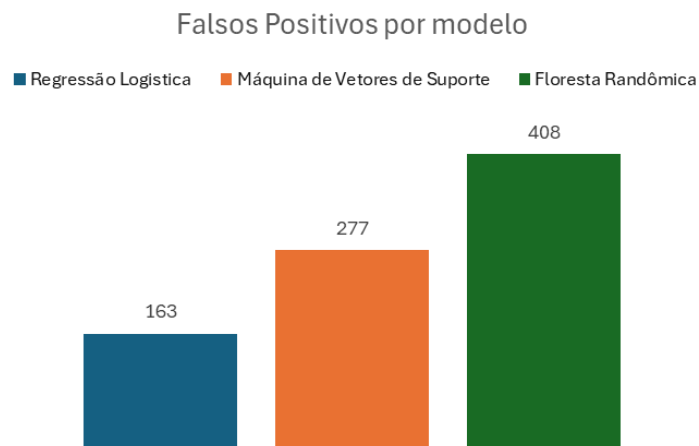
As matrizes de confusão constituíram a principal ferramenta para avaliação dos modelos de aprendizado de máquina neste estudo. A partir delas, foi calculada a métrica de acurácia de cada modelo, cujos resultados são apresentados na Figura 6. Os dados da figura evidenciam a superioridade do modelo Support Vector Machine (SVM) em relação aos demais. Todos os algoritmos analisados obtiveram acurácia superior a 80%, com uma variação de apenas 0,57 ponto porcentual entre o modelo de menor desempenho, Random Forest, e o SVM, que registrou a maior taxa.

Figura 5: Gráfico de barras das Acurácias



A Figura 5 exibe uma visão ampliada dos erros de cada modelo, com maior enfoque nos falsos positivos. Essa análise ajuda a identificar padrões ou tendências comuns nos erros.

Figura 6: Gráfico de barras dos Falsos Positivos



O modelo de Regressão Logística apresentou o menor valor de falsos positivos, apenas 163, ou seja, errou menos nas previsões ao classificar clientes como inadimplentes quando, na verdade, não eram.



Anais do Simpósio Acadêmico de Engenharia de Produção (SAEPRO) da EEL-USP

IX SAEPRO – 25 e 26 de novembro de 2025

O modelo *Random Forest* apresentou a maior quantidade, totalizando 408 falsos positivos. Isso sugere que esse modelo pode ser mais rigoroso na concessão de crédito. O que pode minimizar riscos para a instituição financeira, mas, em contrapartida, pode significar a recusa de crédito para clientes que pagaram sua dívida.

O SVM foi o meio termo em relação aos dois modelos testados, mostrando que sua análise pode estar mais balanceada.

Ao comparar o SVM, que obteve a maior acurácia, com o *Random Forest*, que apresentou a menor, algumas diferenças devem ser consideradas. O SVM se destaca por sua capacidade de regularização, ou seja, sua habilidade de generalizar para novos dados (Bhavsar; Panchal, 2012), além de exigir a padronização das variáveis para um melhor desempenho. Por outro lado, o *Random Forest* não necessita dessa padronização, pois seu modelo, baseado em árvores de decisão, segmenta os dados em múltiplos subconjuntos e combina previsões individuais para gerar um resultado mais robusto (Liaw; Wiener, 2002). Além disso, o *Random Forest* é amplamente utilizado devido à sua eficiência no processamento de grandes volumes de dados e à sua alta precisão, reduzindo o risco de *overfitting* em comparação com modelos de árvores individuais.

No entanto, este estudo apresenta uma limitação, pois, embora seja possível identificar o modelo com melhor desempenho entre os três analisados, a decisão seria mais precisa com a inclusão de métricas como F1-score, precisão e recall, que permitem uma avaliação mais equilibrada entre os erros do modelo e sua capacidade preditiva.

6. CONCLUSÃO



Após análises extensas e comparativas entre os modelos de Regressão Logística, *Random Forest* e SVM. Pode-se perceber que cada modelo possui suas deficiências e seus destaques, mas o modelo que se sobressaiu foi o SVM em termos de acurácia e falsos positivos, embora todos os modelos tenham apresentado desempenhos comparativamente bons.

Para as empresas, esse estudo é de suma importância, porque tem como objetivo principal analisar a concessão de crédito para reduzir o risco de inadimplência e tornar a tomada de decisão da instituição financeira mais precisa, assim, diminuindo os riscos envolvidos na análise.

Para futuras pesquisas, recomenda-se um aprofundamento nas técnicas avançadas de modelagem, isso inclui testar mais modelos e a avaliação por meio de outras métricas da matriz de confusão. Seria proveitoso implementar técnicas de processamento de linguagem natural para analisar a comunicação financeira dos clientes e melhorar a precisão na identificação de riscos potenciais em bases de dados que contenham essas interações. Também seria conveniente avaliar a eficácia de modelos de aprendizagem de máquina em diferentes segmentos de mercado, como crédito ao consumidor, hipotecas e empréstimos empresariais, para adaptar as abordagens de risco de crédito.

7. REFERÊNCIAS

BHAVSAR, Himani; PANCHAL, Mahesh H. A review on support vector machine for data classification. **International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)**, v. 1, n. 10, p. 185-189, 2012.

CHICCO, Davide; TÖTSCH, Niklas; JURMAN, Giuseppe. The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. **BioData mining**, v. 14, n. 1, p. 13, 2021.



Anais do Simpósio Acadêmico de Engenharia de Produção (SAEPRO) da EEL-USP

IX SAEPRO – 25 e 26 de novembro de 2025

DA LESSANDRO FIORETO, Vinícius; DE FREITAS, Claudia Regina; LUCHE, José Roberto Dale. Aplicação de modelos de aprendizado de máquina para a predição da temperatura do rotor em motores PMSM. **Revista de Gestão e Secretariado**, v. 15, n. 8, p. e3981-e3981, 2024.

HA, Van-Sang; NGUYEN, Ha-Nam. Credit scoring with a feature selection approach based deep learning. **In: MATEC web of conferences. EDP Sciences**, 2016. p. 05004.

JANIESCH, C.; ZSCHECH, P.; HEINRICH, K. Machine learning and deep learning. **Electronic Markets**. 2021.

LEO, Martin; SHARMA, Suneel; MADDULETY, Koilakuntla. Machine learning in banking risk management: A literature review. **Risks**, v. 7, n. 1, p. 29, 2019.

LIAW, Andy et al. Classification and regression by randomForest. **R news**, v. 2, n. 3, p. 18-22, 2002.

MASHRUR, Akib et al. Machine learning for financial risk management: a survey. **Ieee Access**, v. 8, p. 203203-203223, 2020.

MHLANGA, David. Financial inclusion in emerging economies: The application of machine learning and artificial intelligence in credit risk assessment. **International journal of financial studies**, v. 9, n. 3, p. 39, 2021.

NAMIOKA, Fernando Akira; LUCHE, José Roberto Dale; FREITAS, Claudia Regina de. Aprendizado de máquinas na previsão e análise de rotatividade de clientes. In: SIMPÓSIO DE EXCELENÇA EM GESTÃO E TECNOLOGIA – SEGeT, 2025, Resende, RJ. Anais. 2025.

NORIEGA, J. P., RIVERA, L. A., & HERRERA, J. A. (2023). Machine learning for credit risk prediction: A systematic literature review. *Data*, 8(11), 169.

OUALID, Adil; MALEH, Yassine; MOUMOUN, Lahcen. Federated learning techniques applied to credit risk management: A systematic literature review. **EDPACS**, v. 68, n. 1, p. 42-56, 2023.

VASHISHT, Shailja; SARVA, Mahesh; MUNDI, Hardeep Singh. Risks measurement in banking: a bibliometric and content analysis. **International Social Science Journal**, v. 72, n. 246, p. 955-977, 2022.

YANG, Li; SHAMI, Abdallah. On hyperparameter optimization of machine learning algorithms: Theory and practice. **Neurocomputing**, v. 415, p. 295-316, 2020.



Anais do Simpósio Acadêmico de Engenharia de Produção (SAEPRO) da EEL-USP

IX SAEPRO – 25 e 26 de novembro de 2025

ZENG, Guoping. On the confusion matrix in credit scoring and its analytical properties. **Communications in Statistics-Theory and Methods**, v. 49, n. 9, p. 2080-2093, 2020.