

Análise de Propensão de Cartões de Crédito para Pessoas Classe Média nos Estados Unidos da América.

Ariel Jentof, Eliseu Fernandes de Oliveira, Lucas Pereira Galves, Matheus Aparecido Ribeiro De Oliveira, Pedro Gabriel B. Perobelli, Roberto dos Santos

Faculdade Impacta de Tecnologia – São Paulo – SP, Brasil

Resumo

Este trabalho apresenta uma aplicação real de análise de propensão de cartões de crédito para Norte Americanos de classe média-baixa, a partir das principais variáveis identificadas. O objetivo é permitir aos bancos um controle para ofertar corretamente um cartão de crédito ao público com renda inferior a USD 50.000/ano. Foi utilizada uma base com cerca de 400.000 registros, foram induzidas algumas regras e a metodologia de Regressão Logística com KNN Vizinhos para melhor entendimento das características dos indivíduos. Tanto os desempenhos da metodologia quanto das regras induzidas atingiram resultados de alto nível e uma acurácia acima do desejado.

Abstract

This paper presents a real-world application of credit card propensity analysis for low to middle-income North Americans based on key identified variables. The objective is to enable banks to accurately target credit card offerings to individuals with an annual income below USD 50,000. A dataset of approximately 400,000 records was utilized, and rules were induced using the Logistic Regression methodology with KNN neighbors to better understand the characteristics of individuals. Both methodology's performance and the induced rules achieved high-level results and accuracy above the desired threshold.

1. Introdução

Os cartões de crédito desempenham um papel significativo na economia moderna, oferecendo aos consumidores uma forma conveniente de pagamento e às empresas uma ferramenta para impulsionar o consumo. Nos Estados Unidos da América, onde o crédito desempenha um papel importante no estilo de vida dos cidadãos, compreender a propensão de pessoas de renda média a possuírem cartões de crédito é de grande importância. Nos últimos anos, o acesso a cartões de crédito tem se tornado cada vez mais importante na vida das pessoas de classe média-baixa nos Estados Unidos. Com a oferta de diversos tipos de cartões de crédito, tornou-se crucial para os consumidores

entenderem suas opções e fazerem escolhas informadas. A partir de 2021, a taxa de utilização de cartão de crédito nos EUA para população classe média-baixa, considerada população que o salário anual é inferior a USD 50.000, vem crescendo aproximadamente 28% e enquanto a taxa de utilização para famílias que ganham até USD 125.000 subiu apenas 23%. (White, 2023).

A análise de dados se apresenta como uma ferramenta poderosa para entender a propensão de cada indivíduo em obter um cartão de crédito, leva-se em consideração sua situação financeira, histórico de crédito e outros fatores relevantes. Neste artigo, realizou-se análise de dados aprofundada para entender quais são as principais características que influenciam a propensão de pessoas de classe média nos EUA em obterem cartões de crédito e como essa informação pode ser utilizada para auxiliar as empresas do setor financeiro. Não é um objetivo deste estudo realizar a análise de risco de inadimplência e score de concessão de crédito.

Nesse contexto, a análise de dados surge como uma ferramenta valiosa para ajudar as instituições financeiras a tomar decisões mais informadas sobre a propensão de cartão de crédito a pessoas de classe média baixa nos Estados Unidos da América. Por meio da análise de dados, é possível avaliar uma variedade de informações sobre o candidato, como salário anual, informações de emprego e renda, da família, entre outros fatores relevantes.

Este artigo visa discutir o uso da análise de propensão de cartões de crédito para pessoas de classe média baixa nos Estados Unidos da América, abordando as principais técnicas utilizadas, os desafios enfrentados e as oportunidades de aplicação dessa abordagem. A tecnologia utilizada para gerar o modelo foi a linguagem Python 3.9, o ambiente de desenvolvimento Google Collaboratory, bibliotecas para Machine Learning, modelos estatísticos como Regressão Logística e KNN.

2. Objetivo

O objetivo deste artigo de análise de dados é identificar as principais características que influenciam a propensão de pessoas de classe média baixa nos Estados Unidos em obterem cartões de crédito. A base de dados retirada do site Kaggle (Kaggle, 2022). Para isso, utilizou-se técnicas estatísticas avançadas para analisar dados de diversos indicadores socioeconômicos, como histórico familiar, possuir carro, casa e/ou celular próprio, nível de escolaridade, histórico de crédito, entre outros. Com os resultados obtidos, propõe-se estratégias e recomendações que possam auxiliar empresas do

setor financeiro a tomar decisões mais informadas e assertivas em relação à obtenção e concessão de cartões de crédito para essa parcela da população.

Para este estudo considerou-se informações de classe social dos Estados Unidos segundo (White, 2023), sendo ela , a classe média baixa, que inclui famílias com renda anual entre cerca de US\$ 30.000 e US\$ 50.000. No próximo segmento, apresentaremos a metodologia adotada para realizar a análise da propensão de cartões de crédito para pessoas de renda média nos Estados Unidos, incluindo a descrição do conjunto de dados utilizado, as variáveis selecionadas e as técnicas estatísticas aplicadas.

3. Metodologia

Para realizar a análise de propensão à cartões de crédito para pessoas de classe média baixa nos Estados Unidos, utilizou-se a KNN e Regressão Logística como preditor do modelo. Esta técnica estatística é amplamente utilizada em estudos que envolvem a previsão de uma variável categórica, como é o caso da propensão à obtenção de cartões de crédito.

```
#-----  
## Ajustando - Aprendizado supervisionado  
#-----  
# Regressão logística com dados de treinamento  
from sklearn.linear_model import LogisticRegression  
LogisticReg = LogisticRegression()  
LogisticReg.fit(X_train, y_train)
```

```
▼ LogisticRegression  
LogisticRegression()
```

```
#-----  
## Separando em dados de treinamento e teste  
#-----  
y = df['ALVO'] #Carrega a variável alvo  
X = df.iloc[:,1:9].values #Carrega as colunas  
  
from sklearn.model_selection import train_test_split  
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.25, random_state = 25)
```

```

#-----
## Calculando a KNN - Aprendizado supervisionado
#-----
from sklearn.neighbors import KNeighborsClassifier
Classifier_kNN = KNeighborsClassifier(n_neighbors=30, weights='uniform', algorithm='brute', p=2)
Classifier_kNN.fit(X_train, y_train)

# Árvore de decisão com dados de treinamento
from sklearn.tree import DecisionTreeClassifier
dtree = DecisionTreeClassifier(class_weight=None, criterion='entropy', max_depth=None,
                              max_features=None, max_leaf_nodes=None,
                              min_impurity_decrease=0.0, #min_impurity_split=None,
                              min_samples_leaf=100, min_samples_split=200,
                              min_weight_fraction_leaf=0.0,
                              random_state=0, splitter='best')
dtree.fit(X_train, y_train)

```

```

DecisionTreeClassifier
DecisionTreeClassifier(criterion='entropy', min_samples_leaf=100,
                      min_samples_split=200, random_state=0)

```

```

#-----
## Previsão usando todos os conjuntos de teste
#-----
# Rede Neural
y_pred_test_RLog = LogisticReg.predict(X_test)
y_pred_test_RNA_P = LogisticReg.predict_proba(X_test)

# Knn
y_pred_test_KNN = Classifier_kNN.predict(X_test)
#y_pred_test_KNN_P = Classifier_kNN.predict_proba(X_test)

# Árvore
y_pred_test_DT = dtree.predict(X_test)

```

Segundo a BBC (Luis Farjado, 2022), para ser considerado classe média-baixa americana, a renda família não deve ultrapassar o valor de US\$ 26 mil anual. Porém, induzimos as regras que, pessoas sem carro próprio, casa própria, telefone celular, valor anual inferior a USD 50.000, valor por pessoa na casa inferior a USD 50.000, tipo de educação “Lower Secondary”, pessoas casadas e que sejam do seguinte grupo de ocupação: “Motoristas de aplicativo, profissionais da segurança, limpeza, cozinha, operacional e garçons/barmens”.

Inicialmente, foi utilizada uma base existente do Kaggle, que inclui informações como histórico de cartões (se a pessoa possui ou não), histórico familiar, informações de emprego e renda, entre outros fatores relevantes para a análise. Os dados foram obtidos a partir de fontes públicas e/ou privadas, como registros de órgãos governamentais, informações fornecidas pelos próprios candidatos ao cartão de crédito, entre outras fontes.

Em seguida, foi realizado a análise exploratória dos dados, com o objetivo de identificar possíveis outliers, variáveis com alta correlação, valores faltantes, entre outros problemas que possam afetar a qualidade da análise. Nessa etapa, foram utilizadas técnicas de visualização de dados, como histogramas, gráficos de dispersão, entre outras.

Após a análise exploratória, utilizou-se de técnicas de estatísticas para análise de propensão de cartão de crédito incluindo a técnica de regressão logística, foi escolhida por ser adequada para modelos de classificação binária, como no caso do estudo em questão ou não à posse de cartão de crédito. Com isso se permitiu avaliar a relação entre as variáveis coletadas e a probabilidade de propensão a cartões de crédito. Empregadas métricas de avaliação de modelos, como a acurácia e a precisão, para avaliar a qualidade do modelo.

É importante destacar algumas limitações do estudo. Primeiramente, a análise foi baseada em dados secundários, o que implica na dependência da qualidade e abrangência desses dados. Além disso, a análise se limitou a pessoas de renda média-baixa nos Estados Unidos, não sendo representativa de outras faixas de renda ou contextos culturais diferentes. Por fim, a análise não considerou fatores contextuais, como mudanças econômicas ou políticas, que poderiam influenciar a propensão de posse de cartões de crédito.

Por fim, avaliou-se o desempenho do modelo, como acurácia e erro de classificação a fim de verificar sua capacidade de prever com precisão a propensão à obtenção de cartões de crédito para pessoas de classe média baixa nos Estados Unidos. Os resultados obtidos foram interpretados e utilizados para propor estratégias e recomendações que possam auxiliar empresas do setor financeiro na tomada de decisões mais informadas e assertivas.

Todo o processo de análise demandou ferramentas de programação, como Python e suas bibliotecas específicas para análise de dados, Pandas¹ e Scikit-Learn². As análises e resultados estão presentes em forma de gráficos e tabelas, permitindo uma melhor visualização e interpretação dos resultados obtidos.

¹ <https://pandas.pydata.org/docs/>

² <https://scikit-learn.org/stable/>

4. Desenvolvimento

A análise dos resultados revela que a propensão de pessoas de renda média-baixa nos Estados Unidos a possuírem cartões de crédito é influenciada por vários fatores-chave, como idade, renda anual. Essas descobertas têm implicações teóricas e práticas significativas.

No que diz respeito à renda anual, observou-se que uma renda mais alta está relacionada a uma maior propensão à posse de cartões de crédito. Isso pode ser explicado pela capacidade de pagamento mais elevada e pelo acesso a benefícios e recompensas oferecidos pelos cartões de crédito.

As implicações teóricas dessas descobertas estão relacionadas à compreensão dos fatores que influenciam a posse de cartões de crédito. O estudo reforça a importância de considerar não apenas fatores demográficos, como gênero e renda, mas também o status familiar, nível acadêmico e profissão ao analisar a propensão à posse de cartões de crédito. Do ponto de vista prático, os resultados fornecem insights para instituições financeiras e empresas de cartões de crédito. Essas organizações podem usar essas informações para segmentar seu público-alvo de forma mais eficaz, desenvolver produtos e serviços adequados às necessidades e características dos consumidores de renda média-baixa e adaptar suas estratégias de marketing para atrair e reter clientes.

```
## Tratamento de nulos e normalização --- Variáveis de entrada numéricas
# Faz o tratamento da variável de AMOUNT INCOME
df['AMT_INCOME_TOTAL'] = df['AMT_INCOME_TOTAL'].apply(lambda x: x.strip())
df['AMT_INCOME_TOTAL'] = df['AMT_INCOME_TOTAL'].apply(lambda x: x.replace(',', ''))
df['AMT_INCOME_TOTAL'] = df['AMT_INCOME_TOTAL'].apply(lambda x: x.replace('$', '').replace('.', '').replace('-', ''))
# Converte para FLOAT
df['AMT_INCOME_TOTAL'] = df['AMT_INCOME_TOTAL'].astype(float)

##--- Dummies - transformação de atributos categóricos em numéricos e tratamento de nulos -----
df['CODE_GENDER'] = [1 if x == 'F' else 0 for x in df['CODE_GENDER']]
df['FLAG_OWN_CAR'] = [1 if x == 'Y' else 0 for x in df['FLAG_OWN_CAR']]
df['FLAG_OWN_REALTY'] = [1 if x == 'Y' else 0 for x in df['FLAG_OWN_REALTY']]
df['CNT_CHILDREN'] = [1 if x <= 1 else 0 for x in df['CNT_CHILDREN']]
df['AMT_INCOME_TOTAL'] = [1 if x <= 50000 else 0 for x in df['AMT_INCOME_TOTAL']]

df['PRE_WORKING'] = [1 if x == 'Working' else 0 for x in df['NAME_INCOME_TYPE']]
df['PRE_COMMERCIAL_ASSOCIATE'] = [1 if x == 'Commercial associate' else 0 for x in df['NAME_INCOME_TYPE']]

df['PRE_EDUCATIO_HIGHER'] = [1 if x == 'Higher education' else 0 for x in df['NAME_EDUCATION_TYPE']]
df['PRE_EDUCATION_INCOMPLETE'] = [1 if x == 'Incomplete higher' else 0 for x in df['NAME_EDUCATION_TYPE']]

df['NAME_FAMILY_STATUS'] = [1 if x == 'Marriage' else 0 for x in df['NAME_FAMILY_STATUS']]

# Atribui 0 para os valores nulos para a variável ALVO
df['HAS_CREDIT_CARD'].fillna(0, inplace=True)
df['ALVO'] = df['HAS_CREDIT_CARD']
df['ALVO'] = df['ALVO'].astype(int)
df = df.drop('HAS_CREDIT_CARD', axis = 1)
```

No entanto, é importante destacar que cada contexto pode ter particularidades que influenciam esses resultados, e estudos adicionais devem ser conduzidos para validar e aprofundar essas associações. Para pesquisas futuras, recomenda-se explorar outras variáveis que possam influenciar a propensão de pessoas de renda média-baixa a possuírem cartões de crédito. Por exemplo, idade, a cultura financeira e o comportamento de gastos podem ser considerados como fatores adicionais que podem afetar a propensão à posse de cartões. Além disso, estudos longitudinais podem fornecer uma visão mais abrangente das mudanças ao longo do tempo na propensão à posse de cartões de crédito e nos fatores subjacentes. Em suma, este estudo contribui para a compreensão dos fatores que influenciam a propensão de pessoas de renda média-baixa nos Estados Unidos a possuírem cartões de crédito. Fornecem um ponto de partida para futuras pesquisas que podem aprofundar o conhecimento nessa área.

5. Conclusão

Para os americanos, ainda existe incerteza sobre a contratação de cartão de crédito, algumas razões para a aquisição de cartão, como proteção contra compras fraudulentas e delitos praticados por terceiro no âmbito das operações bancária, segundo a seção 75 do acordo cível, o consumidor pode recorrer a uma compensação entre compras ou *chargeback*. Existem cartões de crédito sem nenhuma anuidade ou taxa de manutenção, possíveis ganho monetário com milhas, proteção do próprio dinheiro. Algumas empresas precisam de um cartão de crédito adicionado como pagamento. (Money, Totally, n.d.). Logo, é importante a aquisição de cartão de crédito. Entretanto, ainda há forte relutância entre a obtenção do cartão.

```

##-----
## Cálculo dos erros da classificação e Matriz de confusão da RNA
##-----
Erro_RLog_Classificacao = np.mean(np.absolute(y_pred_test_RLog - y_test))
Erro_KNN_Classificacao = np.mean(np.absolute(y_pred_test_KNN - y_test))
Erro_DT_Classificacao = np.mean(np.absolute(y_pred_test_DT - y_test))

print()
print('-----')
print('Regressão Logística - Erro de Classificação:', Erro_RLog_Classificacao)
print('Regressão Logística - Acurácia:', 1-Erro_RLog_Classificacao)
print('KNN - Erro de Classificação:', Erro_KNN_Classificacao)
print('KNN - Acurácia:', 1-Erro_KNN_Classificacao)
print('Árvore - Erro de Classificação:', Erro_DT_Classificacao)
print('Árvore - Acurácia:', 1-Erro_DT_Classificacao)
print('-----')

```

```

-----
Regressão Logística - Erro de Classificação: 0.4846771251368114
Regressão Logística - Acurácia: 0.5153228748631886
KNN - Erro de Classificação: 0.49139000364830354
KNN - Acurácia: 0.5086099963516965
Árvore - Erro de Classificação: 0.4848048157606713
Árvore - Acurácia: 0.5151951842393288
-----

```

Alguns americanos citam a possibilidade de descontrolo de crédito, utilização indevida e a dificuldade da impulsividade de gastos. Quando a fatura do cartão de crédito não é integralmente paga, a taxa, mais conhecida como “*interest rates*” é aplicada e o juros gerados são abusivos. A utilização desenfreada do cartão de crédito pode levar a gastos e mais gastos e a população tende a poupar menos, logo, também é possível levar ao ‘*bankruptcy*’³, pois quando há o pagamento em débito ou dinheiro, existe um limite do atual valor moeda em posse da pessoa, enquanto o crédito, se não houver limite, pode conduzir a falência.

A acurácia do modelo retornar exatamente esse equilíbrio entre a dificuldade do americano entre a decisão de aquisição de cartão de crédito ou se manter somente com pagamento de papel moeda e débito, evitando o acúmulo de dívida.

³ Bankruptcy: Falência

6. Referências Bibliográficas

- [1] Fontinelle, A. (2023, february 22). Retrieved from Investopedia:
<https://www.investopedia.com/articles/younginvestors/08/purchase-financing.asp>
- [2] <https://pandas.pydata.org/docs/>. (24 de April de 2023). Fonte: Pandas Org:
<https://pandas.pydata.org/docs/>
- [3] <https://scikit-learn.org/stable/>. (2023). <https://scikit-learn.org/stable/>. Fonte: Scikit Learn.
- [4] Kaggle. (01 de julho de 2022). Fonte: Kaggle:
<https://www.kaggle.com/datasets/lucaspgalves/applicationamericancreditcard>
- [5] Luis Farjado. (8 de novembro de 2022). Fonte: BBC:
<https://www.bbc.com/portuguese/internacional-63559606>
- [6] Money, Totally. (n.d.). Retrieved from Totally Money:
<https://www.totallymoney.com/credit-cards/credit-cards/>
- [7] White, M. C. (2023, September 22). Retrieved from America's dependence on credit cards is growing. The Fed's rate hike will make it more painful:
<https://edition.cnn.com/2022/09/22/economy/fed-rate-hike-credit-card-debt/index.html>