

Raciocínio Baseado em Casos Empréstimo Bancário

Inteligência Artificial | 2024.2
Prof. Rudimar Dazzi

Acâdemicos:
Alexandre Machado Azevedo,
Daniel Battisti,
Matheus Passold Carelli

Sumário

- *Contexto*
- *Definição*
- *Metodologia*
- *Atributos, pesos e métricas*
- *Objetivo*
- *Resultados*
- *Código*
- *Referencias*

Contexto

O problema

Fazer um programa em que dada uma base de dados e uma entrada, informe a similaridade em relação aos dados, e qual a provável saída da variável alvo.

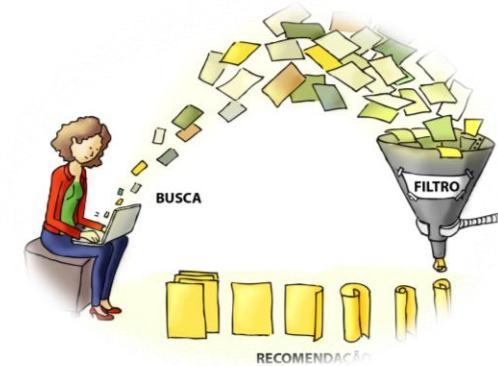
Solução

Utilização da técnica de raciocínio lógico baseado em casos para obter se o cliente conseguirá o empréstimo ou não, desenvolvido na linguagem python e javascript.



Raciocínio Baseado em Casos

- *O Raciocínio Baseado em Casos (RBC) é uma técnica de Inteligência Artificial que resolve novos problemas utilizando experiências anteriores. Ao invés de criar soluções do zero, o RBC busca em uma base de dados casos similares já resolvidos e adapta essas soluções para o problema atual. Esse processo envolve a recuperação de casos relevantes, a adaptação das soluções, a aplicação no novo contexto, e o armazenamento do novo caso para futuros usos.*
- *RBC é especialmente útil em domínios onde o conhecimento completo das regras não está disponível ou onde as soluções dependem fortemente de situações anteriores. Aplicações típicas incluem diagnósticos médicos, tomadas de decisão, e sistemas de recomendação.*



Definição

Definição do contexto

O tema relacionado à concessão de empréstimos é altamente relevante no contexto atual, onde instituições financeiras buscam maneiras eficientes e seguras de tomar decisões sobre crédito

Com o crescimento do mercado de crédito e o aumento da demanda por empréstimos, a necessidade de sistemas que possam avaliar rapidamente a viabilidade de aprovação, com base em dados históricos, tornou-se crucial. Utilizar um modelo de Raciocínio Baseado em Casos (RBC) para comparar solicitações de empréstimos com casos anteriores é uma solução prática que pode melhorar significativamente a precisão das decisões e reduzir o risco de inadimplência. Além disso, a automação desse processo oferece uma maneira mais ágil de lidar com grandes volumes de solicitações, garantindo um atendimento mais rápido e personalizado.

Definição

Definição do problema

O problema em questão envolve a análise de um dataset relacionado à concessão de empréstimos, contendo diversas informações sobre os solicitantes e o status de aprovação ou rejeição do pedido.

Definição do objetivo

O objetivo é utilizar esse conjunto de dados para implementar um protótipo de Raciocínio Baseado em Casos (RBC), no qual novos pedidos de empréstimo são comparados com casos anteriores para determinar a probabilidade de aprovação ou rejeição.

Amostragem dos dados



	loan_id	no_of_dependents	education	self_employed	income_annum	loan_amount	loan_term	cibil_score	residential_assets_value	commercial_assets_value	luxury_assets_value	bank_asset_value	loan_status
3628	3629	2	Graduate	Yes	8400000	25400000	4	758	14800000	600000	18000000	6900000	Approved
1731	1732	2	Graduate	No	4100000	12000000	16	472	4100000	6000000	16300000	4500000	Rejected
1719	1720	1	Graduate	No	4100000	11800000	8	519	4100000	16700000	17300000	6000000	Rejected
23	24	0	Not Graduate	Yes	200000	500000	2	551	0	600000	1100000	300000	Approved
1803	1804	2	Graduate	Yes	4300000	12400000	2	309	4400000	1000000	21600000	7600000	Approved
1530	1531	0	Graduate	Yes	3700000	10700000	8	712	3500000	11300000	19300000	11100000	Approved
1297	1298	3	Graduate	Yes	3200000	9200000	12	476	2800000	6600000	16400000	9000000	Rejected
2335	2336	5	Not Graduate	Yes	5500000	15900000	16	697	6400000	1200000	7700000	3700000	Approved
3445	3446	3	Graduate	No	8000000	23600000	10	841	13200000	12100000	14800000	6000000	Approved
3092	3093	4	Not Graduate	Yes	7200000	20800000	16	449	10600000	10600000	28700000	5400000	Rejected

Métodos para pesos

Correlação

A correlação mede a força e a direção da relação entre dois atributos, sendo útil para definir pesos. Através dela, é possível identificar quais atributos têm maior influência sobre o resultado final, atribuindo pesos maiores àqueles que têm uma relação mais forte com a variável-alvo, como o status de aprovação de empréstimos

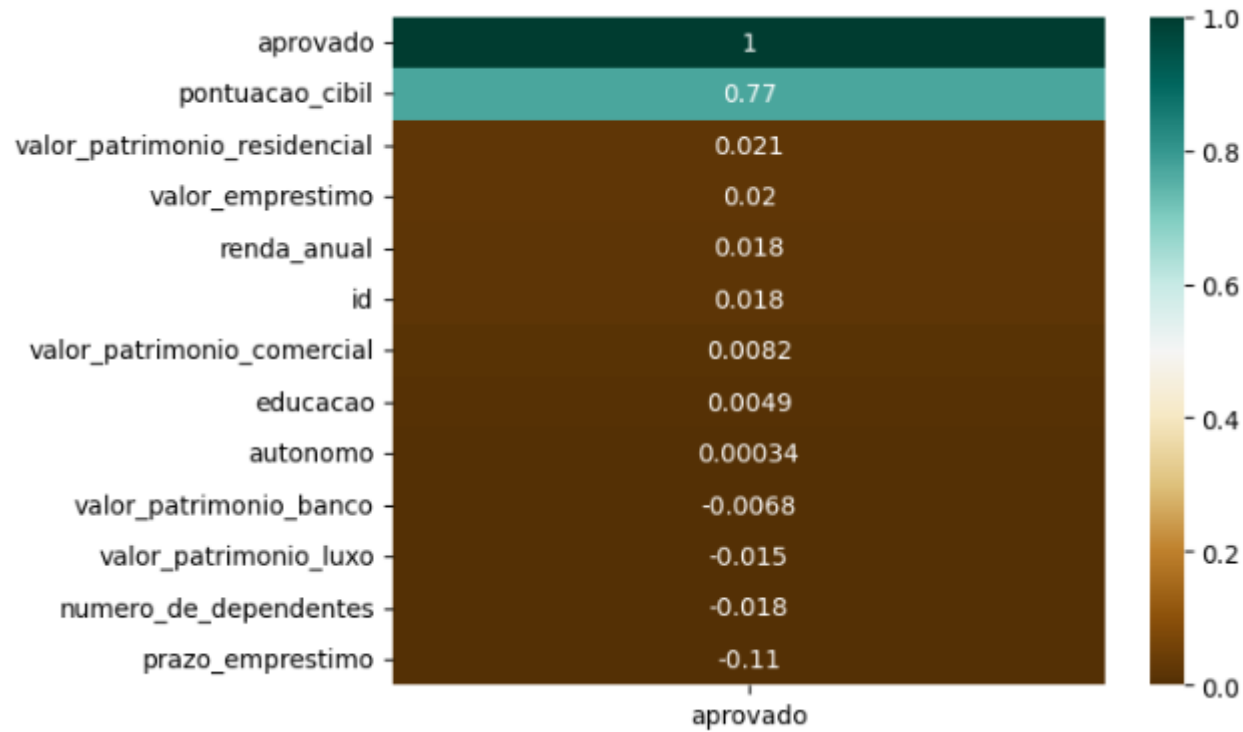
Regressão logística

A regressão logística, estima as probabilidades de um determinado resultado ocorrer com base nos atributos, ajustando pesos de acordo com a contribuição de cada variável no modelo

Profissional da área

Um especialista, como um analista de crédito ou um economista, pode identificar nuances nos dados que um algoritmo sozinho pode não captar, como a importância de certos atributos em diferentes cenários de concessão de empréstimos. Além disso, sua expertise permite ajustar os pesos com base em fatores externos, como tendências econômicas ou mudanças nas políticas de crédito.

Pesos



$$\text{peso}_i = \frac{|\text{correlação}(y, x_i)|}{\sum_{j=1}^n |\text{correlação}(y, x_j)|}$$

Dicionario de dados

Variáveis de entrada

- *ID – Identificação do requerente – numérica*
- *Numero de dependentes - Quantidade de dependentes do requerente – numérica*
- *Educação – Educação do requerente - categórica*
- *Autonomo - Situação profissional do requerente - categórica*
- *Renda anual - Renda Anual do Requerente - numérica*
- *Valor do empréstimo - Valor do empréstimo - numérica*
- *Prazo do empréstimo - Prazo do empréstimo em anos - numérica*
- *Pontuação cibil – Pontuação de crédito - numérica*
- *Valor do patrimonio residencial – Valor imóveis residenciais do requerente - numérica*
- *Valor do patrimonio comercial – Valor dos imóveis comerciais do requerente - numérica*
- *Valor de patrimonio de luxo – Bens de alto valor do requerente - numérica*
- *Valor do patrimonio de banco – Montante de ativos mantidos em contas do requerente - numérica*

Variáveis de saída

- *Aprovado – Define se o empréstimo foi concedido ou não – categórica*

Distribuição dos dados

```
numero_de_dependentes: ['0-1', '1-2', '2-3', '3-4', '4-5']
educacao: ['0-0', '0-1']
autonomo: ['0-0', '0-1']
renda_anual: ['0-2625000', '2625000-5050000', '5050000-7475000', '7475000-9900000']
valor_emprestimo: ['0-10100000', '10100000-19900000', '19900000-29700000', '29700000-39500000']
prazo_emprestimo: ['0-6', '6-11', '11-16', '16-20']
pontuacao_cibil: ['0-450', '450-600', '600-750', '750-900']
valor_patrimonio_residencial: ['0-7275000', '7275000-14550000', '14550000-21825000', '21825000-29100000']
valor_patrimonio_comercial: ['0-4850000', '4850000-9700000', '9700000-14550000', '14550000-19400000']
valor_patrimonio_luxo: ['0-10025000', '10025000-19750000', '19750000-29475000', '29475000-39200000']
valor_patrimonio_banco: ['0-3675000', '3675000-7350000', '7350000-11025000', '11025000-14700000']
```

bank_asset_value(Valor dos Ativos Bancários)	valor	1	2	3	4
0 até 3675000	1	1	0,75	0,5	0,25
3675001 até 7350000	2	0,75	1	0,75	0,5
7350001 até 11025000	3	0,5	0,75	1	0,75
11025001 até 14700000+	4	0,25	0,5	0,75	1

$$\frac{1 - |a - b|}{n}$$

```
{'numero_de_dependentes': [[1.0, 0.8, 0.6, 0.4, 0.2],
[0.8, 1.0, 0.8, 0.6, 0.4],
[0.6, 0.8, 1.0, 0.8, 0.6],
[0.4, 0.6, 0.8, 1.0, 0.8],
[0.2, 0.4, 0.6, 0.8, 1.0]],
'educacao': [[1.0, 0.5], [0.5, 1.0]],
'autonomo': [[1.0, 0.5], [0.5, 1.0]],
'renda_anual': [[1.0, 0.75, 0.5, 0.25],
[0.75, 1.0, 0.75, 0.5],
[0.5, 0.75, 1.0, 0.75],
[0.25, 0.5, 0.75, 1.0]],
'valor_emprestimo': [[1.0, 0.75, 0.5, 0.25],
[0.75, 1.0, 0.75, 0.5],
[0.5, 0.75, 1.0, 0.75],
[0.25, 0.5, 0.75, 1.0]],
'prazo_emprestimo': [[1.0, 0.75, 0.5, 0.25],
[0.75, 1.0, 0.75, 0.5],
[0.5, 0.75, 1.0, 0.75],
[0.25, 0.5, 0.75, 1.0]],
'pontuacao_cibil': [[1.0, 0.75, 0.5, 0.25],
[0.75, 1.0, 0.75, 0.5],
[0.5, 0.75, 1.0, 0.75],
[0.25, 0.5, 0.75, 1.0]],
'valor_patrimonio_residencial': [[1.0, 0.75, 0.5, 0.25],
[0.75, 1.0, 0.75, 0.5],
[0.5, 0.75, 1.0, 0.75],
[0.25, 0.5, 0.75, 1.0]],
'valor_patrimonio_comercial': [[1.0, 0.75, 0.5, 0.25],
[0.75, 1.0, 0.75, 0.5],
[0.5, 0.75, 1.0, 0.75],
[0.25, 0.5, 0.75, 1.0]],
'valor_patrimonio_luxo': [[1.0, 0.75, 0.5, 0.25],
[0.75, 1.0, 0.75, 0.5],
[0.5, 0.75, 1.0, 0.75],
[0.25, 0.5, 0.75, 1.0]],
'valor_patrimonio_banco': [[1.0, 0.75, 0.5, 0.25],
[0.75, 1.0, 0.75, 0.5],
[0.5, 0.75, 1.0, 0.75],
[0.25, 0.5, 0.75, 1.0]]}
```

Resumo dos pesos

numero_de_dependentes: ['0-1', '1-2', '2-3', '3-4', '4-5']
educacao: ['0-0', '0-1']
autonomo: ['0-0', '0-1']
renda_anual: ['0-2625000', '2625000-5050000', '5050000-7475000', '7475000-9900000']
valor_emprestimo: ['0-10100000', '10100000-19900000', '19900000-29700000', '29700000-39500000']
prazo_emprestimo: ['0-6', '6-11', '11-16', '16-20']
pontuacao_cibil: ['0-450', '450-600', '600-750', '750-900']
valor_patrimonio_residencial: ['0-7275000', '7275000-14550000', '14550000-21825000', '21825000-29100000']
valor_patrimonio_comercial: ['0-4850000', '4850000-9700000', '9700000-14550000', '14550000-19400000']
valor_patrimonio_luxo: ['0-10025000', '10025000-19750000', '19750000-29475000', '29475000-39200000']
valor_patrimonio_banco: ['0-3675000', '3675000-7350000', '7350000-11025000', '11025000-14700000']

	aprovado
numero_de_dependentes	0.018174
educacao	0.004934
autonomo	0.000346
renda_anual	0.018060
valor_emprestimo	0.020121
prazo_emprestimo	0.113407
pontuacao_cibil	0.773049
valor_patrimonio_residencial	0.021320
valor_patrimonio_comercial	0.008273
valor_patrimonio_luxo	0.015515
valor_patrimonio_banco	0.006801

```
{'numero_de_dependentes': [[1.0, 0.8, 0.6, 0.4, 0.2],  
[0.8, 1.0, 0.8, 0.6, 0.4],  
[0.6, 0.8, 1.0, 0.8, 0.6],  
[0.4, 0.6, 0.8, 1.0, 0.8],  
[0.2, 0.4, 0.6, 0.8, 1.0]],  
'educacao': [[1.0, 0.5], [0.5, 1.0]],  
'autonomo': [[1.0, 0.5], [0.5, 1.0]],  
'renda_anual': [[1.0, 0.75, 0.5, 0.25],  
[0.75, 1.0, 0.75, 0.5],  
[0.5, 0.75, 1.0, 0.75],  
[0.25, 0.5, 0.75, 1.0]],  
'valor_emprestimo': [[1.0, 0.75, 0.5, 0.25],  
[0.75, 1.0, 0.75, 0.5],  
[0.5, 0.75, 1.0, 0.75],  
[0.25, 0.5, 0.75, 1.0]],  
'prazo_emprestimo': [[1.0, 0.75, 0.5, 0.25],  
[0.75, 1.0, 0.75, 0.5],  
[0.5, 0.75, 1.0, 0.75],  
[0.25, 0.5, 0.75, 1.0]],  
'pontuacao_cibil': [[1.0, 0.75, 0.5, 0.25],  
[0.75, 1.0, 0.75, 0.5],  
[0.5, 0.75, 1.0, 0.75],  
[0.25, 0.5, 0.75, 1.0]],  
'valor_patrimonio_residencial': [[1.0, 0.75, 0.5, 0.25],  
[0.75, 1.0, 0.75, 0.5],  
[0.5, 0.75, 1.0, 0.75],  
[0.25, 0.5, 0.75, 1.0]],  
'valor_patrimonio_comercial': [[1.0, 0.75, 0.5, 0.25],  
[0.75, 1.0, 0.75, 0.5],  
[0.5, 0.75, 1.0, 0.75],  
[0.25, 0.5, 0.75, 1.0]],  
'valor_patrimonio_luxo': [[1.0, 0.75, 0.5, 0.25],  
[0.75, 1.0, 0.75, 0.5],  
[0.5, 0.75, 1.0, 0.75],  
[0.25, 0.5, 0.75, 1.0]],  
'valor_patrimonio_banco': [[1.0, 0.75, 0.5, 0.25],  
[0.75, 1.0, 0.75, 0.5],  
[0.5, 0.75, 1.0, 0.75],  
[0.25, 0.5, 0.75, 1.0]]}
```

Resultados



A depender do caso de entrada o resultado final pode ser variado devido as características específicas de cada dado, entretanto após a análise o programa informará se o empréstimo será concedido ao requerente ou não, tomando como base os dados já obtidos em seu Dataset anteriormente e efetuando a análise do raciocínio baseado em casos.

	numero_de_dependentes	educacao	autonomo	renda_anual	valor_emprestimo	prazo_emprestimo	pontuacao_cibil	valor_patrimonio_residencial	valor_patrimonio_comercial	valor_patrimonio_luxo	valor_patrimonio_banco	aprovado	similaridade
172	3	0	1	500000	1500000	16	748	200000	2400000	9700000	2600000	1	99.753299
134	3	0	0	500000	1300000	16	717	200000	800000	4400000	1300000	1	99.736017
302	4	1	1	800000	2400000	12	633	500000	800000	4800000	2000000	1	99.636522
973	2	1	1	2400000	7000000	14	641	1900000	900000	5500000	2400000	1	99.636522
427	4	1	1	1100000	3200000	14	711	700000	500000	3600000	600000	1	99.636522
...
3776	1	1	0	8800000	27100000	4	313	16500000	15300000	23000000	10900000	1	49.770114
4203	4	1	1	9800000	34900000	6	437	24500000	10000000	24600000	4100000	0	49.491683
4069	0	0	0	9500000	31200000	2	331	21500000	11200000	23700000	8000000	1	49.227204
3909	1	0	1	9100000	28800000	4	304	18400000	16700000	30200000	9600000	1	49.152807
4254	0	0	1	9900000	37700000	6	372	27000000	14900000	33700000	6100000	0	48.286793

4269 rows x 13 columns

Referências

- COSTA, R. et al. **Risk management in credit systems**. Journal of Financial Analysis, v. 5, n. 2, p. 34-49, 2020.
- SILVA, A. **Key determinants in credit approval**. AI in Finance, v. 8, n. 1, p. 22-31, 2021.
- FERREIRA, J. **The impact of credit scores on loan approval**. Economic Perspectives, v. 12, n. 4, p. 56-67, 2019.
- MARTINS, L. **Logistic regression in credit risk models**. Data Science Review, v. 7, n. 3, p. 89-101, 2020.
- RAMOS, P. **Euclidean distance in multi-attribute models**. Expert Systems Journal, v. 14, n. 5, p. 112-124, 2018.
- PEREIRA, C. **Advanced similarity metrics in case-based reasoning**. Journal of AI, v. 9, n. 2, p. 45-60, 2021.
- GOMES, F. **Best practices in case-based reasoning for financial applications**. AI Conference, p. 142-156, 2019.
- ALMEIDA, M. **Optimizing weights in RBC for financial risk**. Financial Systems Review, v. 4, n. 1, p. 74-86, 2020.
- OLIVEIRA, T. **Case-based reasoning in loan approvals: A comprehensive study**. Journal of AI and Finance, v. 6, n. 3, p. 105-123, 2021.

Obrigado