

Método

Utilizamos o *método de aprendizagem baseada no exemplo*, ou *método do vizinho mais próximo*, onde personas são os exemplos e associados os rótulos que desejamos classificar.

(1) Fizemos uma seleção de características para determinar quais critérios seriam incluídos no modelo e (2) classificamos os 981 associados entrevistados na pesquisa quantitativa através da norma-L1, ou *distância de Manhattan**, (3) atribuindo pesos interpretados a partir do conhecimento do negócio. Para evitar que um atributo numérico ganhe um peso maior que outro, (4) normalizamos todas as dimensões através da média arredondada das distâncias. Em seguida, (5) aplicamos um cálculo de análise de sensibilidade para selecionar apenas as perguntas relevantes, reduzindo o número de dimensões. Partimos de um modelo com 33 dimensões e para um modelo otimizado de 10 dimensões.

$$*d_{Manhattan}(X,Y) = ||X - Y||_1 = |x_1 - y_1| + |x_2 - y_2| + \dots$$

É chamada de *distância de Manhattan* porque representa a distância total das ruas que teríamos que viajar entre dois pontos (persona - associado).

Resumo Algoritmo

A distância é a soma de todas as dimensões, multiplicadas pelo peso de negócio, divididas pela média arredondada individual.

Peso * Dimensão / Média Arredondada

3*Idade/13

1*Renda/3000

2*Ocupação/1

0,5*Maturidade Profissional/0,63

2*Escolaridade/1

1*Com quem mora/2

1*Produtos/10

1*Velocidade Crescimento Profissional/3

1*Conhecimento em Investimentos/3

1*Guarda (poupança)/3

Dados de entrada

- CSV "Personas Associadas" - exemplos. Este arquivo não deve ser modificado ou alterado.
- CSV "Base Integrada" - coleta quantitativa aplicada pela UCS + informações do SIG fornecidas pela Sicredi Serrana

| | | |
|----|---------------------------|--|
| | ID | Chave que conecta entrevistado com SIG |
| | NOME | - |
| 1 | IDADE | Idade |
| 2 | FUNDAMENTAL_BÁSICO | Escolaridade (0= Não / 1 = Sim) - Marcar apenas um |
| | MÉDIO_2º GRAU | |
| | SUPERIOR_INCOMPLETO | |
| | SUPERIOR_COMPLETO | |
| | PÓS_GRADUAÇÃO | |
| 3 | RENDA | Renda |
| 4 | CRESCIMENTO_PROFISSIONAL | Pergunta 32 pesquisa quanti UCS - Escala Likert (1 a 5) |
| 5 | CONHECIMENTO_INESTIMENTOS | Pergunta 39 pesquisa quanti UCS - Escala Likert (1 a 5) |
| 6 | GUARDA_OU_INVESTE | Pergunta 41 pesquisa quanti UCS - Escala Likert (1 a 5) |
| 7 | ESTUDANTE | Ocupação (0= Não / 1 = Sim) - Pode marcar mais de um |
| | DESEMPREGADO | |
| | CLT | |
| | MEI | |
| | EMPRESÁRIO/EMPREENDEDOR | |
| | FUNCIONÁRIO_PUBLICO | |
| | APOSENTADO | |
| | AUTONOMO | |
| | DO_LAR | |
| | AGRICULTOR | |
| 8 | MATURIDADE_PROFISSIONAL | Pergunta 28 e 29 pesquisa quanti UCS - (0 = não trabalha/ 1 = menos de um ano/ 2 = entre 1 e 5 anos/ 3 = mais de 5 anos) |
| 9 | PAIS/IRMÃOS | Com quem mora (0= Não / 1 = Sim) - Pode marcar mais de um |
| | CONJUGE | |
| | SOZINHO | |
| | AMIGOS/COLEGAS | |
| | FILHOS | |
| 10 | PROD_CRED_FINANC | Produtos SIG (consórcios e seguros em uma única coluna) |
| | PROD_CREDITO_GERAL | |
| | PROD_CHEQUE_ESPECIAL | |
| | PROD_DESC_RECEB | |
| | PROD_POUPANCA | |
| | PROD_DEPOSITO_A_PRAZO | |

| | | |
|--|---------------------------|--|
| | PROD_FUNDOS | |
| | PROD_PREVIDENCIA | |
| | PROD_DEBITO_CONTA | |
| | PROD_COBRANCA | |
| | PROD_PAGTO_FORN | |
| | PROD_CUSTODIA_CHEQUE | |
| | PROD_PAGAMENTOS | |
| | PROD_CARTAO_DEBITO | |
| | PROD_CARTAO_CREDITO | |
| | PROD_CESTA_RELACIONAMENTO | |
| | CONSORCIO | |
| | SEGUROS | |
| | PROD_CANAIS | |
| | PROD_ADQ | |
| | PROD_CAMBIO | |
| | PROD_CTASAL | |
| | PROD_LCA | |

Códigos

Código principal:

principal.R - busca arquivos distancias.R, analisa.R e relatorios.R.

Importa dados de entrada, calcula as distâncias dos 981 associados contras as 14 personas, seleciona as 3 primeiras personas mais próximas, calcula probabilidades e chama relatórios.

Códigos secundários:

distancias.R: funções que calculam as distâncias de cada dimensão;

analisa.R: funções secundárias que calculam as probabilidades, incertezas e análises internas do algoritmo;

relatorios.R: gera o arquivo de saída do algoritmo CSV "Relatório";

Dados de saída

- CSV "Relatório" - tabela de probabilidade, distância e classificação do alvo em relação as 3 primeiras personas mais próximas nos 981 associados;

| ID | Prob.1 | Prob.2 | Prob.3 | Distancia.1 | Distancia.2 | Distancia.3 | Persona.1 | Persona.2 | Persona.3 |
|----|--------|--------|--------|-------------|-------------|-------------|---------------|-------------|---------------|
| 1 | 0,4234 | 0,3076 | 0,2690 | 8,0444 | 11,0752 | 12,6624 | O consolidado | A servidora | O Constante |
| 2 | 0,3811 | 0,3107 | 0,3083 | 14,0261 | 17,2056 | 17,3389 | O consolidado | A servidora | A estável |
| 3 | 0,3473 | 0,3359 | 0,3168 | 6,2755 | 6,4885 | 6,8807 | A estável | A lutadora | A eficiente |
| 4 | 0,3929 | 0,3392 | 0,2679 | 5,9058 | 6,8391 | 8,6610 | O pacato | O Constante | O promissor |
| 5 | 0,4828 | 0,2894 | 0,2278 | 3,8894 | 6,4897 | 8,2440 | A singela | O pacato | O consolidado |
| 6 | 0,3587 | 0,3357 | 0,3055 | 6,8949 | 7,3667 | 8,0951 | O promissor | A Indecisa | A eficiente |
| 7 | 0,5237 | 0,2703 | 0,2061 | 1,7667 | 3,4231 | 4,4897 | A singela | O pacato | A lutadora |

Análise de Desempenho do Algoritmo

10 Dimensões X 3 Dimensões

9% dos associados mantiveram a *mesma classificação* quando algoritmo rodado com 10 dimensões ou as 3 dimensões do SIG.

55,04% dos associados foram classificados com a *mesma primeira persona* no algoritmo aplicado para 10 dimensões ou para as 3 dimensões disponíveis no SIG.

80,32% tiveram a *primeira* persona das 10 dimensões entre as 3 primeiras personas quando aplicado apenas as 3 dimensões.

32,72% tiveram a *segunda* persona das 10 dimensões entre as 3 primeiras personas quando aplicado apenas as 3 dimensões.

29,86% tiveram a *terceira* persona das 10 dimensões entre as 3 primeiras personas quando aplicado apenas as 3 dimensões.

Frequência de classificação por distâncias:

90,21% da base classificada em até distância 6:

| | | | | |
|-------------------------|--------|----------|----------|--------|
| Freq acumulada relativa | 32,93% | 90,21% | 98,88% | 100% |
| Distância | 0 a 3 | 3,01 a 6 | 6,01 a 9 | 9,01 + |

Formação da base de entrevistados:

