

Trabalho final - Classificação do Risco Cardiovascular

Universidade Federal de Pernambuco

Tatiana Alejandra Moreno Avila

Introdução

No contexto das doenças cardiovasculares, estas representam a principal causa de mortalidade no mundo. A prevalência dos diferentes fatores de risco para o desenvolvimento de doenças cardiovasculares tem aumentado ao longo dos anos.

Este estudo busca caracterizar o perfil cardiovascular de um banco de dados proveniente do projeto “Corações de Baependi”, desenvolvido na Universidade de São Paulo.

Através de um modelo preditivo de rede neural totalmente conectada, busca-se identificar padrões e correlações entre variáveis antropométricas e bioquímicas associadas ao risco cardiovascular.

Enfermedades Cardiovasculares



Metodologia

Os modelos de redes neurais binárias buscam prever uma variável categórica, sendo no nosso caso o risco cardiovascular. Essa variável é definida pelos parâmetros apresentados em diversas pesquisas, levando em conta variáveis como colesterol total (CTOTAL), LDL (CLDL), HDL (CHDL), triglicerídeos (Triglic) e idade, as quais possuem correlação associada ao risco cardiovascular.

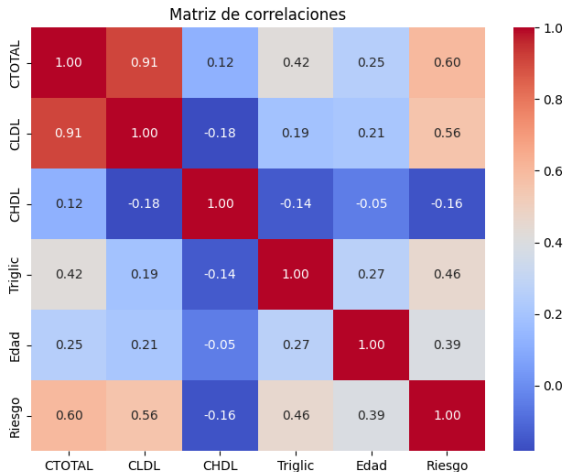


Figure 1: Matriz de Correlação

Esta base de dados conta com um número de 1587 registros e possui 12 variáveis explicativas, das quais algumas são utilizadas para determinar a variável de risco cardiovascular, conforme estipulado pela Organização Mundial da Saúde.

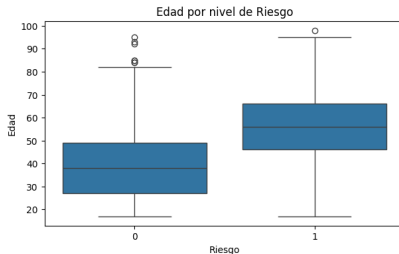
Para realizar a aplicação, primeiro realizamos uma análise exploratória da base de dados, pois ela conta com as seguintes variáveis de medidas antropométricas e bioquímicas.

	IID	Sexo	Edad	IMC	BAI \
count	1587.000000	1587.000000	1587.000000	1.587000e+03	1.587000e+03
mean	61134.683680	1.567738	44.097669	2.136145e+14	2.218316e+14
std	35385.313165	0.495546	16.999219	9.103282e+13	9.974663e+13
min	1101.000000	1.000000	17.000000	2.080000e+01	1.936997e+11
25%	28926.500000	1.000000	30.000000	1.961328e+14	1.707279e+14
50%	61702.000000	2.000000	43.000000	2.291299e+14	2.213687e+14
75%	91919.500000	2.000000	56.000000	2.650850e+14	2.758850e+14
max	122401.000000	2.000000	98.000000	4.919502e+14	9.645401e+14

	Cintura	Cadera	CVLDL	Triglic	CTOTAL \
count	1587.000000	1587.000000	1587.000000	1587.000000	1587.000000
mean	87.263390	97.830498	26.628859	131.672968	180.768179
std	12.317006	9.988689	18.842513	70.381810	47.410663
min	54.000000	51.000000	3.200000	34.200000	84.700000
25%	78.000000	92.000000	16.700000	83.600000	145.100000
50%	86.000000	97.000000	22.400000	112.300000	174.700000
75%	95.000000	103.000000	31.750000	158.500000	210.400000
max	144.000000	160.000000	526.000000	639.000000	435.000000

	CLDL	CHDL
count	1587.000000	1587.000000
mean	98.570762	55.901701
std	43.763440	15.753750
min	9.700000	13.000000
25%	67.050000	45.100000
50%	94.300000	53.900000
75%	125.800000	65.950000
max	390.100000	128.400000

Nossa variável de risco divide nossa base de dados em: Alto risco (1): 429 pacientes, Baixo risco (0): 1158 pacientes e Proporção: 27,03% de alto risco.



No gráfico anterior, podemos observar uma diferença entre as idades de ambos os grupos, sendo mais comum que os pacientes com baixo risco sejam mais jovens do que os pacientes com alto risco.

Estrutura do Modelo

Depois de realizar o respectivo escalonamento e padronização das variáveis, o modelo de rede neural binária é definido da seguinte maneira:

```
Model: "sequential"
```

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 64)	512
dropout (Dropout)	(None, 64)	0
dense_1 (Dense)	(None, 32)	2,080
dropout_1 (Dropout)	(None, 32)	0
dense_2 (Dense)	(None, 16)	528
dense_3 (Dense)	(None, 1)	17

```
Total params: 3,137 (12.25 KB)
```

```
Trainable params: 3,137 (12.25 KB)
```

```
Non-trainable params: 0 (0.00 B)
```

Avaliação

Após treinar o modelo, obtemos o seguinte desempenho em sua respectiva avaliação do modelo.

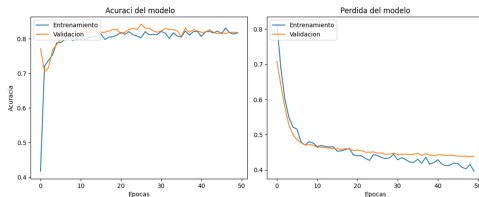
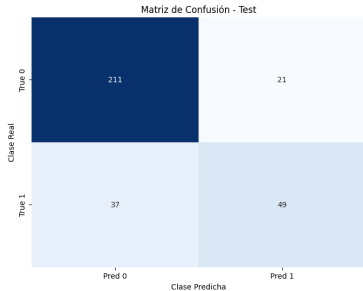


Figure 2: Avaliação - Accuracy e Loss

Notemos que este modelo apresenta um bom desempenho, o que é evidenciado pelas curvas de accuracy e loss, com valores de 0,8176 e 0,4348.

matriz de confusão



	precision	recall	f1-score	support
Clase 0	0.85	0.91	0.88	232
Clase 1	0.70	0.57	0.63	86
accuracy			0.82	318
macro avg	0.78	0.74	0.75	318
weighted avg	0.81	0.82	0.81	318

Este modelo apresenta um melhor desempenho na identificação de pacientes com baixo risco e, para os pacientes de alto risco, possui um desempenho moderado. Além disso, o modelo demonstra capturar padrões relevantes das variáveis.

Conclusão

Este modelo de rede neural apresenta um comportamento razoavelmente bom e, com base nos resultados obtidos, mostra-se uma proposta satisfatória para determinar padrões na classificação dos pacientes, permitindo assim identificar quais outros fatores podem influenciar no risco de desenvolver alguma doença cardiovascular. Sendo assim, constitui uma possível ferramenta de apoio em pesquisas de saúde e genética, capaz de classificar de maneira aceitável, a partir de medidas antropométricas e bioquímicas, todos os padrões que levam a um risco elevado, possibilitando análises exploratórias clínico-descritivas mais rigorosas.

Referências

- ▶ Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.
- ▶ Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning.
- ▶ Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Springer.
- ▶ Organização Mundial da Saúde (OMS). Relatórios sobre doenças cardiovasculares.