

World Happiness Report

**Ana Flávia Dias Rodrigues, Camilla Damasceno da Silva,
Jorge Luiz Mendes Moreira Junior, Luiza Ávila Defranco Gonçalves,
Stefany Gaspar Xavier França**

¹Instituto de Ciências Exatas e Informática
Pontifícia Universidade Católica - (PUC-MG)
Belo Horizonte – MG – Brasil

1. Introdução

A felicidade, o bem estar, são temas cada vez mais recorrentes em nossa sociedade. As pessoas estão desesperadas em busca de serem felizes, de viver em paz. Ter o poder de analisar o índice de felicidade em um determinado país, levando em consideração opiniões de moradores e diversos outros fatores vivenciados, pode auxiliar e muito essa busca por um bem estar maior. Especialistas de diversas áreas, descrevem como as medições de bem-estar podem ser usadas de forma eficaz para avaliar o progresso das nações, auxiliando os governantes a entenderem quais são as áreas mais precárias, e que precisam de mais atenção, para que sua população seja mais feliz.

Nesse trabalho é apresentado um problema de regressão, utilizando a base *World Happiness Report*, que contém 156 instâncias, 9 atributos, sendo eles 8 atributos numéricos e 1 nominal: Overall rank, Country or region, Score, GDP per capita, Social support, Healthy life expectancy, Freedom to make life choices, Generosity, e Perceptions of corruption. Foram utilizados os algoritmos *K-means* e *Multilayer Perceptron* com a intenção de julgar os melhores atributos que contribuem para o elevado índice de felicidade dos países.

2. Descrição da base de dados.

A coluna Overall rank mostra a posição em que o país ficou após a análise. Country or region o nome do país ou região, o Score é o total de pontos do país/região. As colunas: GDP (PIB) per capita, Social support, Healthy life expectancy, Freedom to make life choices, Generosity e Perceptions of corruption descrevem até que ponto esses fatores contribuem na avaliação da felicidade em cada país, seus números estimam como cada um dessas escolhas contribuem para o valor dado em uma avaliação tipo-escada, em que cada país teve representantes avaliando suas condições de vida. Por exemplo, se os habitantes acham que de 1 a 10 a qualidade de vida é 7, os números estimam quanto cada um contribui para essa escala.

As classificações são baseadas nas respostas à principal pergunta de avaliação de vida feita na enquete. Isso é chamado de escada Cantril: pede aos entrevistados que pensem em uma escada, com a melhor vida possível para eles sendo 10 e a pior vida possível sendo 0. Eles então são solicitados a avaliar suas próprias vidas atuais de 0 a 10. As classificações são de amostras representativas nacionalmente, para os anos 2015-2019. Eles são baseados inteiramente nas pontuações da pesquisa, usando os pesos do *Gallup* para tornar as estimativas representativas. As sub-barras mostram a extensão estimada em que cada um dos seis fatores - níveis de PIB, expectativa de vida, generosidade, apoio social, liberdade e corrupção - contribui para tornar as avaliações de vida mais altas em cada

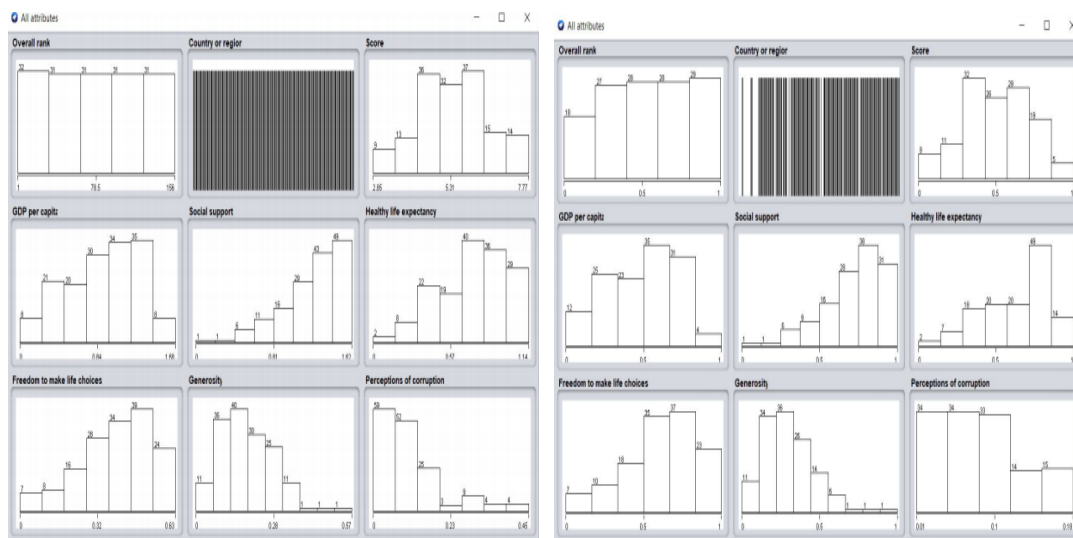
país do que na Distopia, um país hipotético que tem valores iguais às médias nacionais mais baixas do mundo para cada um dos seis fatores. As sub barras não têm impacto na pontuação total relatada para cada país, mas são apenas uma forma de explicar para cada país as implicações do modelo estimado em. As pessoas costumam perguntar por que alguns países têm classificação mais alta do que outros - as sub barras (incluindo os resíduos, que mostram o que não é explicado) são uma tentativa de dar uma resposta a essa pergunta.

3. Etapas de Pré Processamento

Para as etapas de pré processamento foram utilizados os filtros do *Weka*.

Inicialmente foram eliminados os dados inconsistentes e redundantes, em atributos, utilizando o filtro *remove*. Nessa etapa o número de atributos permaneceu o mesmo, ou seja, não ocorreram atributos redundantes.

Logo em seguida foi utilizado o filtro *RemoveMisclassified* para remoção de dados inconsistentes e redundantes em instâncias. Nessa etapa o número de instâncias reduziu de 156 para 130. E então foi aplicado o filtro *normalize*.



(a) Dados antes do pré-processamento

(b) Dados depois do pré-processamento

4. Algoritmos rodados

4.1. K-means

Inicialmente foi utilizado o algoritmo *K-means*, para realizar o agrupamento em clusters com o objetivo de analisá-los. Porém os resultados obtidos não foram muito satisfatórios, pois o algoritmo falhou em separar em clusters de forma precisa.

Métrica utilizada: *Silhouette Index*

Para $k = 4$:

- Cluster 0: 0.1907, verdict: a non substancial structure was found!
- Cluster 1: 0.1229, verdict: a non substancial structure was found!
- Cluster 2: 0.0861, verdict: a non substancial structure was found!
- Cluster 3: 0.1981, verdict: a non substancial structure was found!

Mean: 0.1494, verdict: a non substantial structure was found!

Concluímos, nesse caso, que nosso modelo não é bom, não apresenta uma estrutura substancial e, provavelmente, não apresenta a realidade. Isso ocorreu provavelmente porque o *K-means* não é o melhor algoritmo dadas as características da base de dados.

4.2. Multilayer Perceptron

Utilizando o *Multilayer Perceptron* obtivemos melhores resultados, com cerca de 60% a 70% de acerto. Nosso objetivo com esse algoritmo é predizer o *score* de cada país.

Especificamos a nossa rede da seguinte forma:

- hiddenLayers = a, que é *atributo + classe/2*
- batchSize = 30
- learningRate = 0.2
- trainingTime = 1000

```
Time taken to build model: 25.22 seconds
```

```
=== Cross-validation ===
```

```
=== Summary ===
```

Correlation coefficient	0.9
Mean absolute error	0.0553
Root mean squared error	0.0946
Relative absolute error	30.7589 %
Root relative squared error	43.6581 %
Total Number of Instances	130

5. Conclusão

Concluímos que a WHO (*world health organization*) fez uma seleção eficaz de atributos escolhidos para a classificação de felicidade em diversos países. A rede neural *Multilayer Perceptron* conseguiu aprender de forma satisfatória, com os atributos fornecidos, mesmo sendo poucos.

References

A. A. Alharbi, H. A. Alotebii and A. A. AlMansour, "Towards Measuring Happiness in Saudi Arabia based on Tweets: A research proposal," 2018 1st International Conference on Computer Applications & Information Security (ICCAIS), Riyadh, 2018, pp. 1 4, doi:10.1109/CAIS.2018.8442024.

Esteban Ortiz Ospina (2013) "Happiness and Life Satisfaction". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/happiness and life satisfaction' [Online R esource]

Carlsen, L. How Happy Are we Actually? A Posetic Analysis of the World Happiness Index 2016 2019 Denmark as an Exemplary Case. *Int. Journal of Com. WB* 3, 311 322 (2020). <https://doi.org/10.1007/s42413-019-00044-8>

Bhinde, Hetal, "An outline of Happiness Index", 2017 *CnR's International Journal of Social & Scientific Research*, India. Published online at [researchgate.net](https://www.researchgate.net) Retrieved from: 'https://www.researchgate.net/publication/333236662_An_outline_of_Happiness_