Evaluation of Feature Importance and Feature Dependencies

The traditional manner to present visualisations of feature importance is done through bar charts[1] [2] [3] [4] [5] or box plots [6], as depicted in Figure 1. These methods have been described as an efficient method to visualise information but have been often described as boring [7].

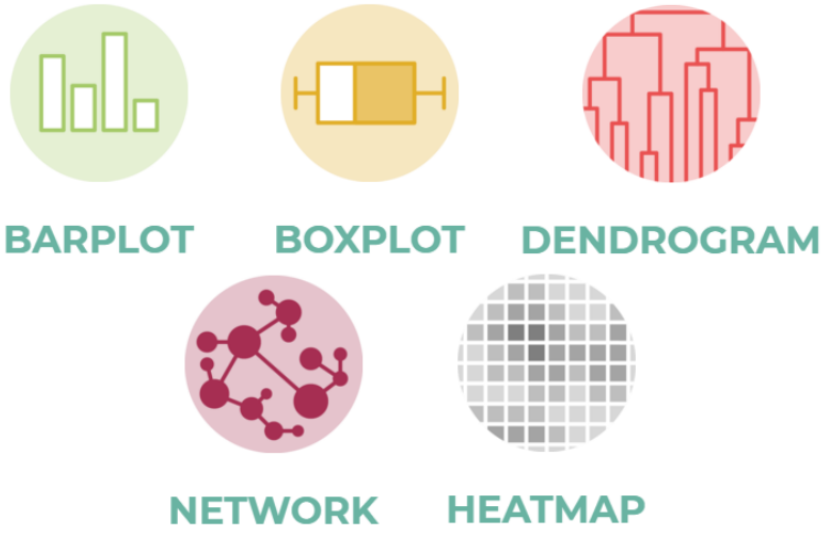

Figure 1: Different Data Visualisations [7]

Feature Dependencies are traditionally represented through correlation heatmaps, dendrograms or network graphs, as shown by [8] [9] [10]. These can also be seen in Figure 1. These types of graphs are useful to display a general view of the numerical data but do not extract specific data points. The above mentioned visualisations have the same caveats, if too much data is presented the figure gets cluttered and unreadable [7].

For this project, we have put together a combination of a dendrogram and a chord graph to display both relative feature importance and relative feature dependencies of a diagnostic breast cancer dataset. Only the top 5 feature levels are portrayed in the visualisation namely:

1) **Perimeter worst:** worst size of the core tumour
2) **Radius worst:** mean of the distances from the center to points on the perimeter
3) **Area worst:** worst area of the affected region
4) **Concave points worst:** largest value for number of concave portions of the contour
5) **Concave points mean:** mean for number of concave portions of the contour

If the above mentioned “traditional” methods are followed, the feature importance of the data would be represented with a bar chart as illustrated in Figure 2 and the correlations between the features would be represented in Figure 3.

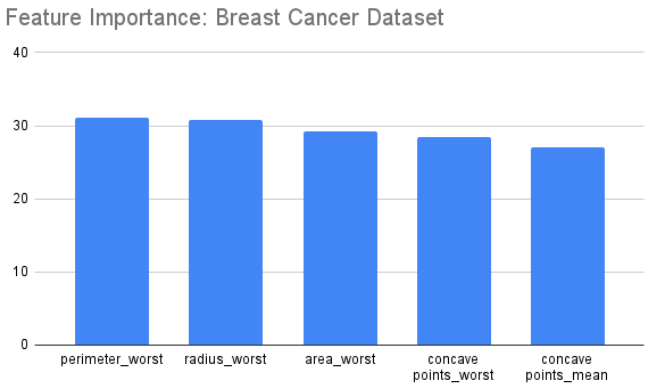

Figure 2: Feature Importance

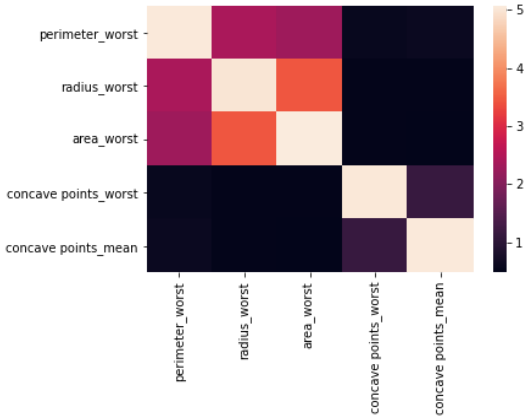

Figure 3: Feature Dependencies

To evaluate the effectiveness, learnability and efficiency of the visualisation, we request you to answer the following questions and then ask you to fill out a short questionnaire about your experience with the visualisation.

You may first head to the following URL to access the visualization and then use that to answer the survey that follows

Visualization: https://abhishek1ahuja.github.io/feature_importance/

Survey:
https://docs.google.com/forms/d/e/1FAIpQLSdF6HGYnCZXPb-NOqyqftSh0oZOGUTlhLM6aPj5-xF3_VhWkw/viewform?usp=sf_link

Thank you very much for participating!!

If you have any questions, please feel free to contact me!
Radhika Kapoor
+31 (0) 614035270

References

[1] How to calculate feature importance with python. https://machinelearningmastery.com/calculate-feature-importance-with-python/. (Accessed on 06/16/2021).

[2] G. Wei, J. Zhao, Y. Feng, A. He, and J. Yu. A novel hybrid feature selection method based on dynamic feature importance. Applied Soft Computing Journal, 93, 2020. doi: 10.1016/j.asoc.2020.106337. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85085240086&doi=10.1016%2fj.asoc.2020.106337&partnerID=40&md5=a181e025c4e3adfd8a22a538d9818df9. cited By 5

[3] Feature selection and data visualization | kaggle. https://www.kaggle.com/kanncaa1/feature-selection-and-data-visualization. (Accessed on 06/16/2021).

[4] Gunnar König, Christoph Molnar, Bernd Bischl, and Moritz Grosse-Wentrup.Relative feature importance. In 2020 25th International Conference on Pattern Recognition (ICPR), pages 9318–9325, 2021. doi: 10.1109/ICPR48806.2021.9413090.

[5] Huiting Zheng, Jiabin Yuan, and Long Chen. Short-term load forecasting using emd-lstm neural networks with a xgboost algorithm for feature importance evaluation. Energies, 10(8), 2017. ISSN 1996-1073. doi: 10.3390/en10081168. URL https://www.mdpi.com/1996-1073/10/8/1168.

[6] André Altmann, Laura Toloşi, Oliver Sander, and Thomas Lengauer. Permutation importance: a corrected feature importance measure. Bioinformatics, 26(10): 1340–1347, 04 2010. ISSN 1367-4803. doi: 10.1093/bioinformatics/btq134. URL https://doi.org/10.1093/bioinformatics/btq134.

[7] From data to viz | find the graphic you need. https://www.data-to-viz.com/ #barplot. (Accessed on 06/16/2021).

[8] T.R. Vilmansen. Feature evalution with measures of probabilistic dependence. IEEE Transactions on Computers, C-22(4):381–388, 1973. doi: 10.1109/T-C.1973

[9] Wei Zhang, Hong Mei, and Haiyan Zhao. A feature-oriented approach to modeling requirements dependencies. In 13th IEEE International Conference on Requirements Engineering (RE'05), pages 273–282, 2005. doi: 10.1109/RE.2005.6.

[10] Iran Rodrigues, Márcio Ribeiro, Flávio Medeiros, Paulo Borba, Baldoino Fonseca, and Rohit Gheyi. Assessing fine-grained feature dependencies. Information and Software Technology, 78:27–52, 2016. ISSN 0950-5849. doi: https://doi.org/10. 1016/j.infsof.2016.05.006. URL https://www.sciencedirect.com/science/article/pii/ S0950584916300921.