

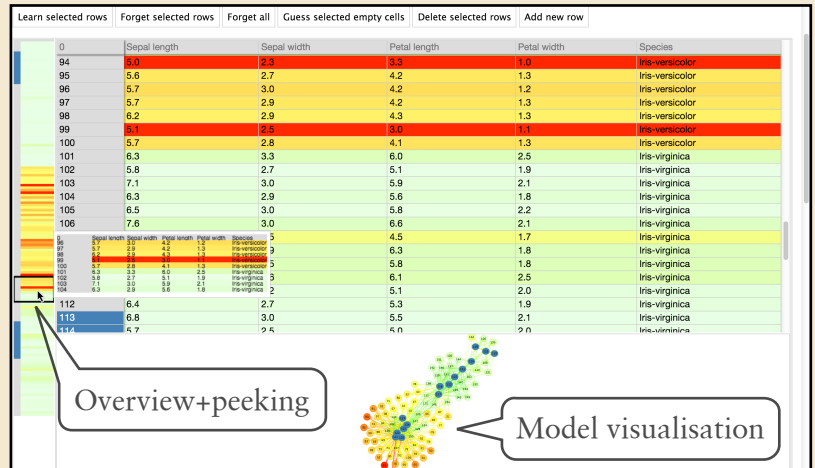
Visual machine learning in spreadsheets

Advait Sarkar, Mateja Jamnik, Alan Blackwell, & Martin Spott

Spreadsheets are a natural fit for machine learning

We augment the spreadsheet, a versatile yet familiar data manipulation tool, to enable non-expert end-users to build and apply machine learning models.

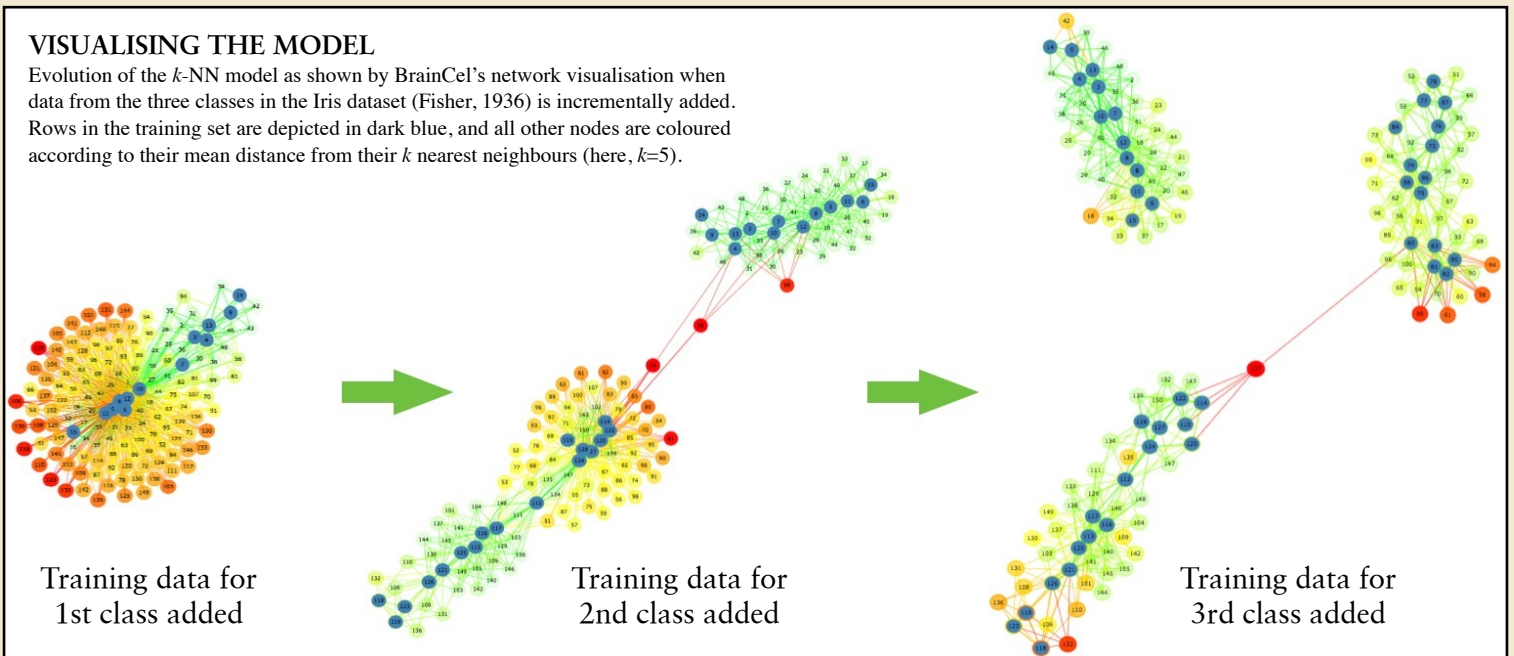
The user selects rows of complete, correct data to “teach” the software, which implicitly trains a statistical model. The user can then apply this model to guess values for empty cells, or validate the contents of filled cells.



Part of the BrainCel interface. Rows with blue row numbers have been added by the user to the training set. Rows are coloured according to the model’s confidence: green for high confidence, red for low.

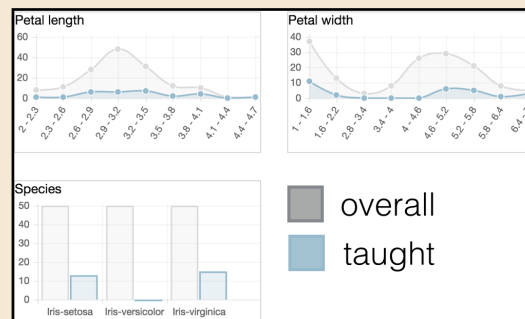
VISUALISING THE MODEL

Evolution of the k -NN model as shown by BrainCel’s network visualisation when data from the three classes in the Iris dataset (Fisher, 1936) is incrementally added. Rows in the training set are depicted in dark blue, and all other nodes are coloured according to their mean distance from their k nearest neighbours (here, $k=5$).



Visualisations improve end-user engagement

Through visualisations of the model along with meta-information such as class representation, the user can perform more nuanced activities, like judging the quality of the model, identifying good training examples, and querying why and how the model makes certain predictions.



CLASS REPRESENTATION

These graphs show how the distribution of taught data differs from that of the overall distributions of data in the spreadsheet. Here, *Iris-versicolor* appears to be under-represented.

Further reading

- [1] Sarkar, A., Jamnik, M., Blackwell, A. F., & Spott, M. (2015). Interactive visual machine learning in spreadsheets. In 2015 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC) (pp. 159–163).
- [2] Sarkar, A., Blackwell, A. F., Jamnik, M., & Spott, M. (2014). Teach and try: a simple interaction technique for exploratory data modelling by end users. In 2014 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC) (pp. 53–56). IEEE. doi:10.1109/VLHCC.2014.6883022
- [3] Sarkar, A. (2015). Spreadsheet interfaces for usable machine learning. In 2015 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC) (pp. 283–284).