Personalising Leadership Learning with Artificial Intelligence – a proof-of-concept study

Emma Mi, Ella Mi

Emma is a doctor and currently a National Medical Director's Clinical Fellow working in the Leadership and Lifelong Learning Team of the People Directorate of NHS England and NHS Improvement.

Ella is a doctor and Clinical Machine Learning researcher, with experience in developing models for medical imaging and EHR data, and currently a Computer Science MSc student at the University of Oxford on a Google DeepMind scholarship. She is also a member of the British Medical Association Technology and Innovation Committee.

Data

This analysis is based on a dataset of 1000 participants who previously enrolled on Leadership Academy programmes. The data fields available are demographic characteristics and programme enrolment information as follows:

- Programme enrolled on
- Current status
- Cohort start year
- Organisation type
- Local delivery partner (region)
- Band
- Role
- Age
- Gender
- Ethnic Origin
- Sexual Orientation
- Religion
- Nationality
- Disability

Aim of the study

To generate insights from programme participant data using machine learning techniques, both supervised and unsupervised

Machine learning models

Supervised learning

Outcome: A model that predicts the programme choice of participants from their demographic characteristics

Data processing

Characteristics most likely to be useful in prediction were selected to be used in modelling. They are:

- Age
- Gender
- Ethnic origin – in groups of White, Black, Asian and Other
- Disability
- Band – in groups of 2 to VSM
- Role – in groups of Organisational, ANP, Nurse/Midwife, Doctor, Executive

In some fields, the participants' data entries were classified into broader groups (e.g. in ethnicity *White Irish* classified simply as *White*) in order to reduce the number of categories for each feature, and therefore increase number of people in each category, to produce an adequate sample size for machine learning.

Any participants who had missing values in any of the above fields was omitted from the analysis. This consisted of 33 participants (out of 1000) in total.

Modelling methods

We divided the participant data into a **training set (used to develop the algorithm)** and a **test set (used to independently evaluate the algorithm)**. This was in a ratio of 80:20 – that is 774 participants in the training set and 193 participants in the test set.

Coding was done in the **Python programming language**, using libraries: numpy, pandas, scipy, scikit-learn and matplotlib.

A **Random Forest classifier** from the scikit-learn library, a type of machine learning classification algorithm, was used.

- Advantages of this algorithm are that it is one of the most robust and versatile classification algorithms, able to process categorical, ordinal and continuous variables, which our dataset contains.

Results

N.B. **Accuracy** is equal to $\frac{True\ Positive + True\ Negative}{True\ Positive + False\ Positive + True\ Negative + False\ Negative}$

On the training set – Accuracy of model: 74%

On the test set – Accuracy of model: 70%

- Accuracy on the training and test set are similar indicating the model is neither under- nor over-fit.

Interpretation

- Based on a participant's demographic characteristics, **the model can accurately predict the programme choice of the participant 70% of the time**
- What is "good" accuracy depends on the intended use of the model and the amount of data that has been used to develop the model
- For our purposes and given the size of the current dataset (which is fairly small), 70% accuracy is reasonable

Limitations

- More datapoints are needed to improve accuracy
- A more balanced dataset, in terms of spread across programmes, is needed to improve accuracy. In this dataset nearly 70% of participants were on the Mary Seacole programme.

Opportunities for further work

- If we had programme outcome data, e.g. participant assessment/satisfaction scores, supervised learning, including classification or regression algorithms, could be applied to predict outcomes based on demographic characteristics.

Unsupervised learning

Outcome: A model that identifies distinct clusters of programme participants according to their demographic characteristics

Data Processing

Same as for supervised learning

Modelling methods

Coding was again done in the **Python programming language**

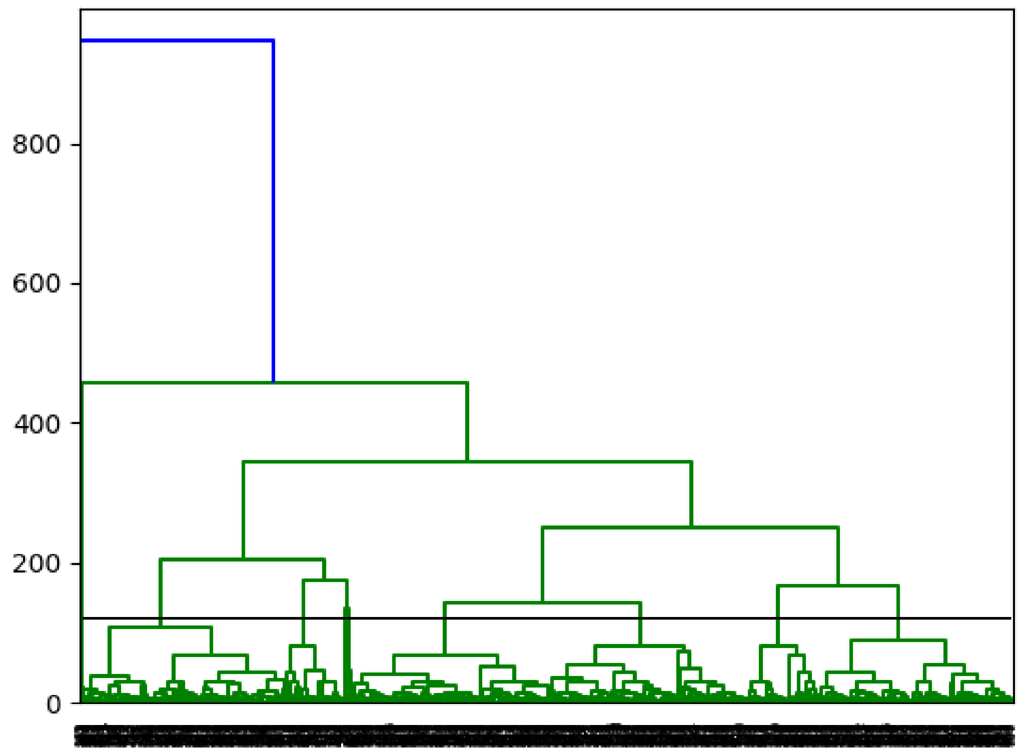
We used **hierarchical clustering**, specifically the **agglomerative clustering algorithm**

- The advantages of this algorithm is that it handles categorical data and cluster number can be inferred from the dataset

N.B. Clustering seeks to minimise the difference between individuals in the same cluster and maximise the difference between individuals in different clusters

Results

Firstly, we created a dendrogram of the data (a hierarchical diagram which builds natural clusters from the individual data points upwards) shown below and identified that 7 clusters is the most appropriate, indicated by the horizontal line on the dendrogram (at which position there are 7 clusters).



Then the algorithm was run to generate 7 clusters. The results are shown in the table:

Cluster	Number	Band	Age	Gender	Disability	Role_AHP	Role_Doctor	Role_Executive	Role_Nurse	Role_Organisational	Ethnic_Asian	Ethnic_Black	Ethnic_Other	Ethnic_White
0	24	6.541667	39.75	0.333333	0.208333	1	0	0	0	0	0	0.625	0.375	0
1	74	6.175676	40.32432	0.351351	0.067568	0	0	0	0	1	0.527027	0.418919	0.054054	0
2	64	6.34375	44.15625	0.328125	0.03125	0	0	0	1	0	0.328125	0.671875	0	0
3	673	6.763744	41.06835	0.17682	0.049034	0.23477	0	0.17682	0.369985	0.218425	0	0	0.005944	0.994056
4	48	7.291667	38.54167	0.541667	0.020833	0.645833	0	0.354167	0	0	1	0	0	0
5	24	7.625	46.16667	0.291667	0	0	0	1	0	0	0	0.791667	0.208333	0
6	60	8.15	41.28333	0.45	0	0	1	0	0	0	0.616667	0.016667	0.033333	0.333333

Interpreting this table, the general characteristics of each cluster are:

- Cluster 0 – Band 6-7, age 38, female (67%), disability (21%), AHP, Black ethnicity (63%)
- Cluster 1 – Band 6, age 40, female (65%), no disability (93%), organisational role, Asian (53%) and Black (42%) ethnicity
- Cluster 2 – Band 6, age 44, female (67%), no disability (97%), nurse/midwife, Black (67%) and Asian (33%) ethnicity
- Cluster 3 – Band 7, age 41, female (82%), no disability (95%), nurse/midwife (37%), AHP (23%) and organisational (22%) roles, White ethnicity

- Cluster 4 – Band 7, age 39, male (54%), no disability (98%), AHP (65%) and executive (35%) roles, Asian ethnicity
- Cluster 5 – Band 7-8, age 46, female (71%), no disability, executive role, Black ethnicity (79%)
- Cluster 6 – Band 8, age 41, male (45%), no disability, doctor, Asian (62%) and White (33%) ethnicity

Interpretation

- **Programme participants fall into characteristic groups**
- The largest group is characterised by middle-aged white females who are nurse/midwives/AHPs at Band 6-7
- There are small niche groups, such as senior Black females in executive roles, and Black females who have a disability and are in AHP roles

Opportunities for further work

- If we had programme outcome data, e.g. participant assessment/satisfaction scores, we could apply clustering to outcome data and explore the demographics of people in different clusters.