

How is the data-driven approach used by a recommendation system?

A video streaming service uses a recommendation system to suggest to a user the films and TV shows best suited to them.

Machine learning uses a data-driven approach. The first step in any data-driven approach is to collect the data.

a) What data is collected for this recommendation system?

Once the data is collected, it is prepared for the model.

b) Does this involve labelling the data? Why / why not?

c) What is the third stage in a data driven approach?

After this, the model will be evaluated.

d) Describe two different ways in which this could be done.

Finally, the model will be used by the video streaming service. This happens once the model has been refined and is evaluated to be performing well.

Guidance for supervisors:

- a) The data collected will be the user's viewing history. It can include information such as which shows they watched all the way through, which sort of shows they started watching and then lost interest in etc. Some streaming services allow users to rate how much they enjoyed a programme, so this can be taken into account too if available.
- b) Preparing the data does not involve labeling the data in this case. This is because recommendation systems use clustering algorithms, which is a type of unsupervised learning. Data in unsupervised learning systems is not labeled.
- c) Training the model
- d) The model could be evaluated using metrics such as accuracy, precision and recall