Classifying Accounting Journals by Nature

Vikki Richardson, MSc in Big Data

UNIVERSITY of STIRLING

PROBLEM STATEMENT

Audit Scotland are responsible for auditing the accounts of many public agencies in Scotland including local councils and NHS Boards.

During an audit, sample journals are extracted from the accounts for further inspection. Currently, an auditors judgement is used to determine which samples are chosen. This can be time consuming and restrictive with the risk of introducing bias or missing potentially relevant information.

Audit Scotland have created an Excel tool, GLiQ (General Ledger Audit Intelligence Query tool) which maps transactions in the accounts to risk analysis look up tables reliant on debit and credit transactions within the journal.

Positive results are being recorded in terms of extracting risky journals and automating the sample choosing process .

They would like to investigate whether Machine Learning, more specifically, clustering the journals can improve the results and highlight potentially 'interesting' journals.

CRISP-DM APPROACH

Business Understanding:

Undergo a review of auditor workflow to ensure understanding of final outcome required.

Data Understanding:

Analyse the given data along with GLiQ developer ensuring full understanding of the current look up table procedure.

Data Preparation:

Aggregate transaction level data to journal level incorporating a newly created 'character' variable showing the structure of that journal in terms of debits and credits at different levels of account code. Higher levels being more specific.

Modelling:

Using R; specifically the tidyverse, cluster and dbscan packages compare results from K-means with K-modes and Density clustering on the prepared data.

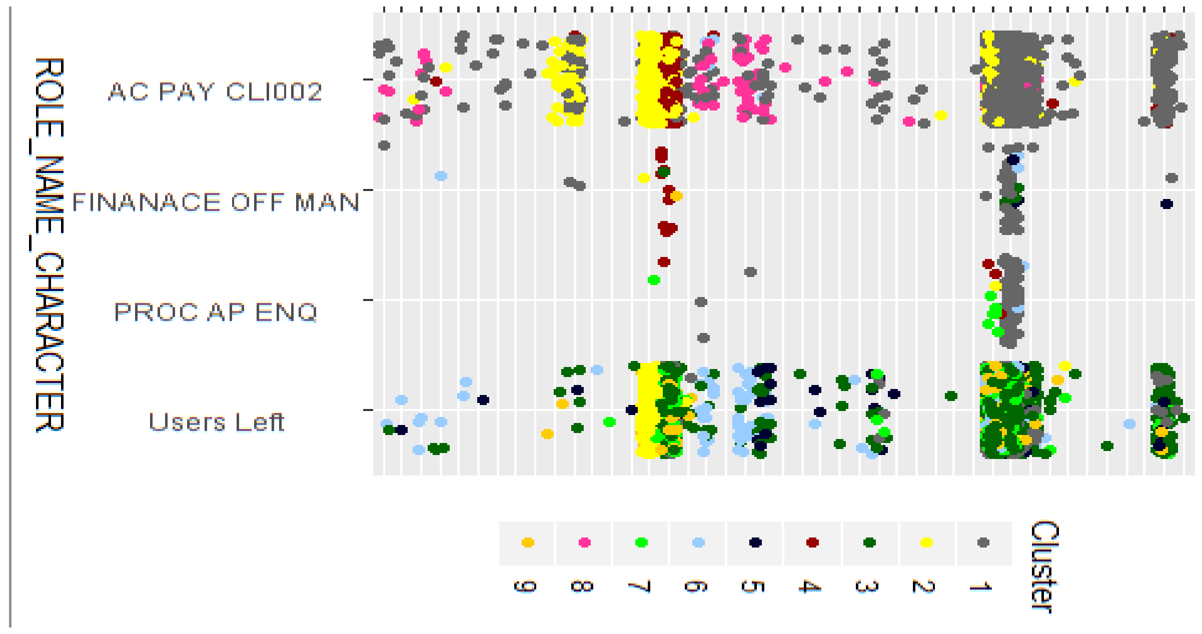
Evaluation:

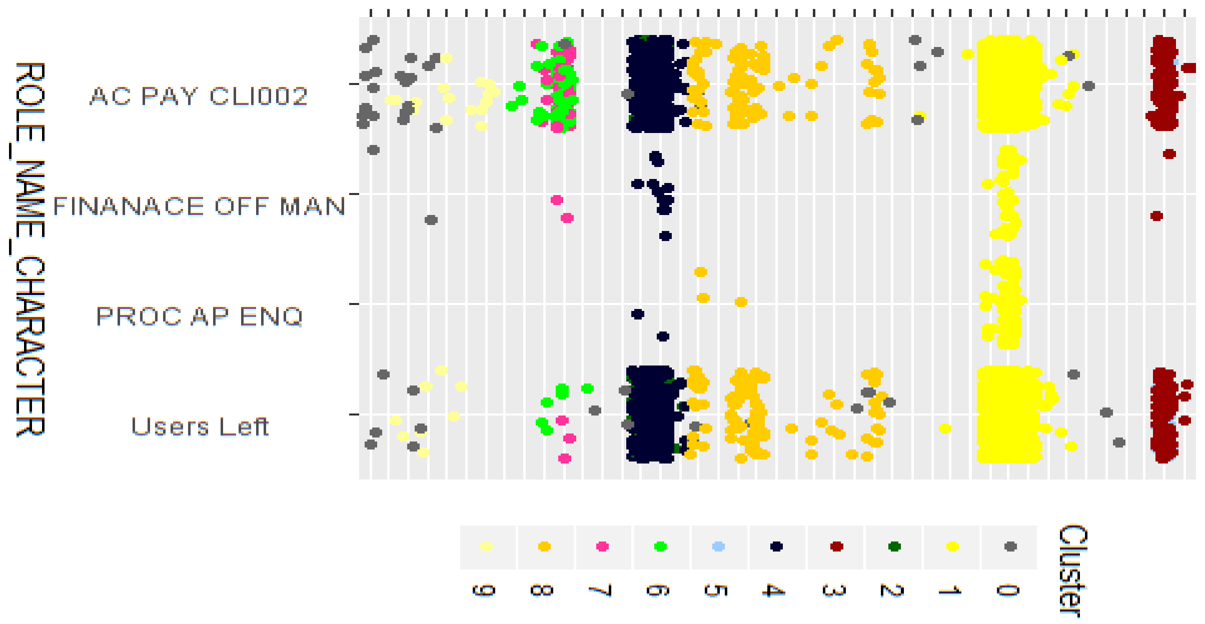
Evaluate the results against the current GLiQ tool for comparison and possible revelation of new insight.

PRELIMINARY RESULTS

Preliminary results show potential for new insights currently unavailable with GLiQ using density clustering. Journals identified as 'unusual' through GLiQ are starting to form separate clusters within the model.

K-MODES CLUSTERING vs DENSITY CLUSTERING





AUDIT SCOTLAND

stir.ac.uk

For further information contact: *Vikki Richardson*

vip00001@students.stir.ac.uk

BE THE DIFFERENCE