Data Analytics in Practice

Use Case

Laura Anderson, Data Scientist

Katherine Hutcheson, Graduate Data Scientist

What do we do ?

Categorisation-as-a-Service **(CaaS)**

Our machine learning categorisation engine

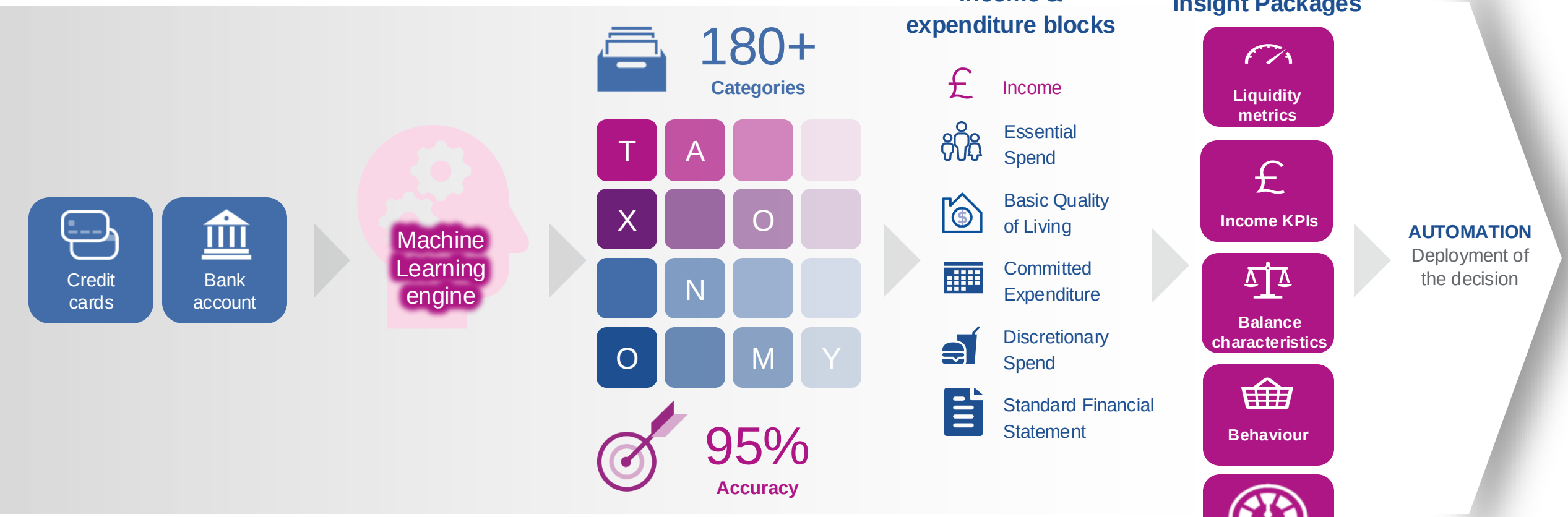
CaaS **identifies patterns and signals in a customer's financial behaviour**.

It **creates value by giving clients a deeper insight** on an **individual's finances,** as well as a broader understanding of the behaviour of **a portfolio of customers.**

It helps clients **anticipate risk, improve decision making, and be more responsive to changes in a customer's personal finances**.

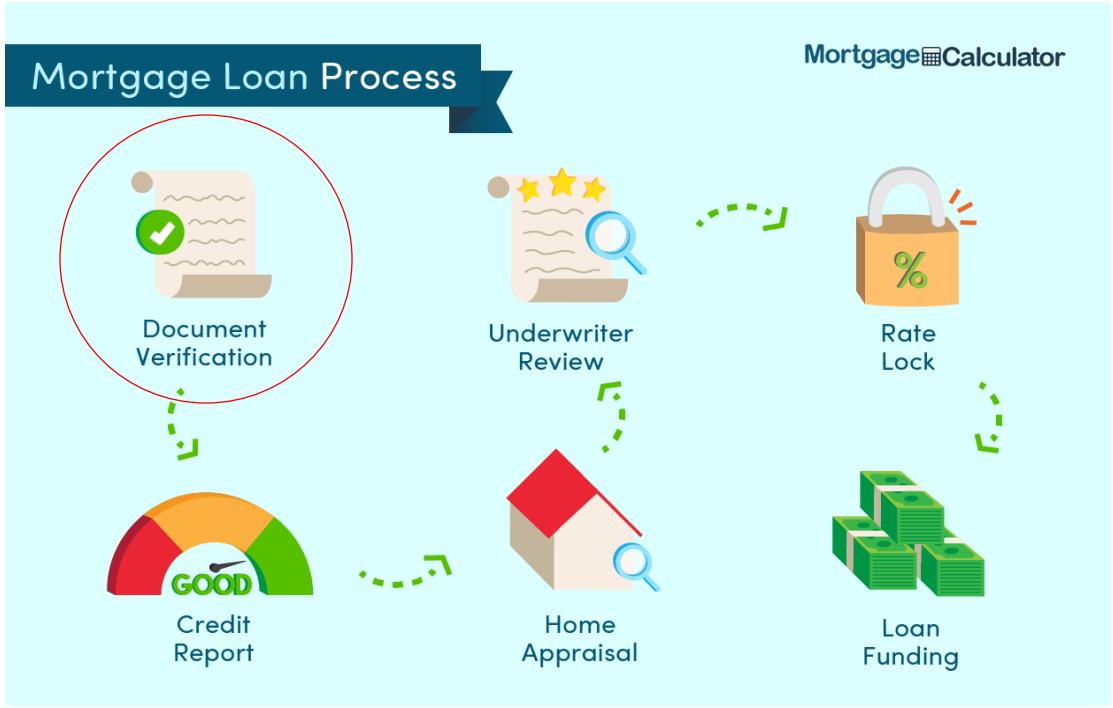
“Categorisation as a Service” (CaaS)

Our machine learning categorisation engine has processed over 10Bn transactions



The motivation

Mortgage Applications



You could see a typical mortgage application process on the left.

How can Experian help?

The first step for lenders is to begin verifying assets, income and employment etc. This could be a time-consuming process, with a time delay and a relatively primitive view of a consumer.

Faster and more Granular Process
Experian's income verification model calculates applicants' monthly income derived from their bank transactions and aim to speed up the mortgage application process.

The issue is that decisionmakers (lenders) use calculations based on annual gross income which cannot be derived from transactional data **(net income).** Therefore, we must calculate annual gross income from monthly net income.

Prevent Mortgage Fraud
Buyers provide fraudulent employment details and income values to support their application.

According to Nationwide Building Society 1 in 10 applicants think it's reasonable to exaggerate their income on a mortgage application

Lenders need to be aware of staged income. **Staged income** is where applicants create fake income streams regular payments throughout the application process which end after completion.

How could you help ?

Task 1 (50%)

We are asking for you to calculate annual gross income from given monthly net income. The dataset consists of two columns – monthly_net_income and location. Tax systems operate differently in these two regions.

monthly_net_income	location	
4587	rou	⟹ Rest of UK
1946	sco	⟹ Scotland

A jupyter notebook will be provided for you to code a means of achieving this, along with a sample of data for you to test your solution. A holdout dataset will be used to evaluate your final solution. The percentage of correct calculations within a margin of error will determine your grade for this task.

N.B. - Follow the format set out in the notebook, as it is an automatic grading system.
- Understand how the tax system works in the UK, strongly advised to write a *gross --> net* function before *net --> gross*.
- Assume a tax year of 2024/2025 (metrics on following slide)
- Assume the only deductions from gross to net are income tax and national insurance (no pension contributions, student loan repayments, etc)

Task 2 (50%)

Have a think about the limitations of your solution (what about pensions, student loan repayments etc), and how these can be addressed (what information can you infer from someone's bank transactions?) . Also think about how we could identify self-employed workers and location based on transactional data and write a short summary (400 word limit).

experian

What to return

- Your source code in the form of a python jupyter notebook environment. No notebook, no grade. We need to understand your code, so please ensure it is well commented understandable variable names, etc.

- The salary_holdout.csv with the *annual_gross_income* field filled.

- A pdf or word document for your task 2 solution.

Marking Breakdown

- Task 1 (50%):
  - Result from code (25%)
  - Explainability of code (10%)
  - Readability of code (10%)
  - Demonstration of data exploration (5%)
- Task 2 (50%) is dependent on your understanding of the weaknesses of your solution, and how well you have thought out potential solutions.

How the tax system works.

- The UK operates on a bracket tax band system. Your rate of tax only applies to the corresponding bracket!
- NI and Income Tax are both applied to the gross **(pre-tax)** salary.
- Following numbers based on tax year 2024/2025:

ROU Income Tax Table

Name	Income Bracket	Rate
Personal Allowance	£0 to £12,570	0%
Basic Rate	£12,571 to £50,270	20%
Higher Rate	£50,271 to £125,140	40%
Additional Rate	> £125,140	45%

Please note: You may notice there is an additional personal allowance taper for non-Scottish residents earning above £100,000. You **do not** need to consider this personal allowance taper for those in that bracket.

ROU National Insurance Table

Name	Income Bracket	Rate
No payment	£0 to £11,904	0%
NI lower	£11,905 to £50,270	8%
NI higher	> £50,271	2%

Scottish Income Tax Table

Name	Income Bracket	Rate
Personal Allowance	£0 to £12,570	0%
Starter Rate	£12,571 to £14,876	19%
Basic Rate	£14,877 to £26,561	20%
Intermediate Rate	£26,562 to £43,662	21%
Higher Rate	£43,663 to £75,000	42%
Advanced Rate	75,001 to 125140	45%
Top Rate	> £125,140	48%

Scottish National Insurance Table

Name	Income Bracket	Rate
No payment	£0 to £11,904	0%
NI lower	£11,905 to £50,270	8%
NI higher	> £50,271	2%

Questions ?

Email: Katherine.Hutcheson@experian.com