34
EXPERIAN

Using Big Data To Make Lending Decisions And To Crack Down On Identity Fraud

Background

Experian are best known for providing credit references, used by banks and financial services companies to assess risk when deciding whether to lend money.

They also provide a range of other services based around the data they have collected, such as fraud and identity theft protection. More recently, they have added specialized data-analytics-driven services aimed at helping business customers in the automobile trading, healthcare insurance and small business markets.

What Problem Is Big Data Helping To Solve?

Banks and insurance companies play a game of chance when they lend money or offer cover: they must be confident that their customer can afford to make the repayments, with interest, or that the value of their premiums will cover expenses incurred by paying out on their claims.

On top of that, cyber fraud and identity theft are both growing problems, with more and more financial transactions taking place online, and customers using online banking portals to manage their

accounts. Experian's own research has shown that, while five years ago online fraud was mostly a problem for the rich, today all sectors of society are being targeted by increasingly sophisticated hackers and scammers.[1]

How Is Big Data Used In Practice?

Experian hold around 30 petabytes of data on people all over the world in their credit bureau database, which is currently growing at a rate of 20% annually.

This data is used to build up a detailed picture of consumers and businesses. As well as holding detailed data on individuals such as their credit history, and demographic information such as age, location and income status, Experian group them into one of 67 types and 15 groups using their "socio-demographic clarification" tool Mosaic. These groups include "urban cool", successful city dwellers owning or renting expensive apartments in fashionable city locations; "professional rewards", experienced professionals with successful careers living in financial comfort in rural or semi-rural areas; and "global fusion", young working people in metropolitan terraces with a wide variety of ethnic backgrounds.

This data is used to segment customers for marketing purposes as well as to assess creditworthiness and insurability.

Experian offer their services to financial companies to help prevent fraud by matching incoming transactions against their fraud prediction model, which monitors 282 attributes – such as the value of the transaction, the geographical locations of those involved and their previous behaviour – to offer real-time fraud detection. Should a transaction show a similar profile to previous transactions that were known to be fraudulent, it can be flagged up for manual review or real-time intervention.

They also provide services to 10,000 healthcare organizations in the US alone, including 2900 hospitals, helping them to assess healthcare claims as well as to gain insights into their patients' financial situations and set up affordable payment plans.

Finally, their services aimed at used car buyers draw data from national databases on vehicle trades, insurance companies and government regulatory agencies to provide information such as whether a vehicle has been in a collision or has other issues, for example structural damage or mileage discrepancies on the odometer.

What Were The Results?

Experian have said that by integrating data analysis across the entirety of their operation, and treating all of their data as a centralized pool rather than as separate, segregated resources, they are enabling more people to buy homes, expand their businesses and manage their finances effectively.

What Data Was Used?

Experian collect their data on individuals from lenders, which give them details on how much people borrow and whether they make repayments as well as links between addresses that people have moved from and to, and aliases – when people have changed the name by which they are known. They also harvest large amounts of data from public records such as postal address databases, electoral registers, county court registers, birth and death records (to establish if fraud is being committed in the name of a deceased person) and national fraud prevention services such as the UK's CIFAS system.

What Are The Technical Details?

Experian host their 30-petabyte consumer reference database on a secure, Linux-powered computing cluster built around Hadoop

architecture. MapR Hadoop is used for distributed storage, and the server cores also contribute processing power towards the analytics operations – essential for the high-volume, high-velocity data processing required to provide their services in near real time.

Other technologies put to use include Apache Hive and data visualization tool Tableau to provide graphical feedback to analysts.

Any Challenges That Had To Be Overcome?

Drawing up a profile of the sort of people who were likely to be targeted was seen as key to identifying and preventing fraudsters from carrying out their operations.

In order to do this, Experian overlaid data from their Mosaic socio-demographic profiling tool over the UK National Fraud Database, showing where and when fraudulent attempts had been made to separate people from their hard-earned money.

What immediately became apparent was that, far from identity theft and online fraud being a problem facing the rich, scammers were targeting the far less financially secure at an equally high rate.

This led to a fundamental reassessment of the way fraud prevention initiatives are applied by banks, insurers and other financial institutions. Rather than giving priority to monitoring high-value business transactions, the majority of analytics and scrutiny is now applied to regular, small-value, everyday transactions, which previously may have slipped through the net.

What Are The Key Learning Points And Takeaways?

Although there is a lot of misunderstanding about the roles of credit agencies – they do not, for example, as many believe, "blacklist"

people who they feel are not worthy of credit – they have an important role to play in ensuring that efficient lending and insurance can take place.

The more data available to help them do that, the better (in theory) the situation for us – losing less money to bad debts, or incorrectly calculated insurance risk, inevitably leads to higher premiums and more expensive credit for all of us.

Cybercrime is no longer just a problem facing the rich and well off. Scammers and fraudsters worked out that security systems were generally geared towards detecting large-scale thefts, and a less risky path is to attempt to commit multiple smaller crimes. Security providers are adapting to this, and shifting focus to monitoring smaller transactions.

REFERENCES AND FURTHER READING

1. Evans, D. (2014) The changing identity of UK fraud victims, http://www.experian.co.uk/blogs/latest-thinking/the-changing-identity-of-uk-fraud-victims/, accessed 5 January 2016.

For more information on Sprint's use of Big Data, see:

http://www.experian.co.uk/assets/consumer-information/white-papers/Harnessing%20the%20power%20of%20Big%20Data.pdf