Public Domain Pictures 2016

How Big Data and Data Science are used to prevent fraud

The Problem

Fraud is a major problem in today's society. The Centre for Counter Fraud Studies' Annual Fraud Indicator 2016 estimates the annual UK fraud loss is as high as £193 billion. Business fraud losses estimated to be £144 billion a year, and fraud committed against individuals estimated at around £10 billion a year (Centre for Counter Fraud Studies 2016). The actual numbers are undoubtedly higher than these estimates, but are impossible to calculate as some instances of fraud can go unnoticed for long periods of time, potentially indefinitely. Without reliable and accurate estimates of fraud, it is difficult to determine which preventative measures, if any, work or do not work to protect victims. Current estimates vary widely, so it is difficult for anyone to conclude upon the full scope of the issue (Fraud Research Center 2014).

I have chosen to cover this issue as fraud is one of the largest crimes affecting the UK economy today, and without the current systems and solutions in place, fraudulent crimes would be undeniably crippling to the global economy. Fraud affects every level of society: individuals, families, companies, government organisations, even whole economies; everyone is at risk of fraud, no matter how secure or careful they are with their information. Individual victims of fraud can be left bankrupt, without important medical care, or have their identity in question due to identity theft. Businesses who are victims of fraud can be left financially damaged, potentially even bankrupt, and can result in a loss of customer confidence (ACPO 2016).

Because fraud is such a broad subject, with numerous different types of fraud ranging from large scale malicious to small accidental occurrences, it is impossible to stop fraud without radical societal and technological changes. Developing technologies and our advancement into the online age on the back of cloud computing presents new opportunities for fraudsters to target, and opens up new vulnerabilities. Traditional fraud detection techniques typically involve comparing IP addresses and login data with those previously associated with fraud The reported median time for fraud cases (the amount of time from when fraud commenced to when it was detected) was 18 months in 2014 (Bănărescu. A 2015). However big data can help to identify potential risks and prevent fraud in real time, and a reduce in detection and reaction times can mean a reduce in loss.

Big Data against Fraud

Fraud prevention usually can be classed as either reactive or proactive. Most businesses employ reactive investigations and penalties against the culprits, however these methods aren't always successful, and sometimes the culprit has successfully covered their tracks, leaving the victim at a significant loss. What's more, even if a business successfully reacts to a high-value act of fraud, it can leave their security in question, often times losing them business in the process (Analytics Magazine 2014).

Proactive fraud prevention uses big data, data sciences and predictive analytics to identify vulnerabilities and risks, and increase overall resilience to fraud by analysing huge sets of data. Big data is often defined using the '3Vs': variety, volume, and velocity. Volume is how much data you can store, variety is the variance in data types and formats which you have, and velocity is the speed at which you can process data (SQLAuthority 2013). In the field of fraud prevention, velocity is clearly the most important characteristic of big data, as the faster that fraud can be detected, the sooner it can be prevented. The general aim of analytics on big data is to look for patterns and regularities in the data, and once a pattern has been established, identify any aberrations or anomalies that occur in the data. If certain anomalous results occur more regularly, correlate to detected fraudulent activity, or follow a pattern outside of the regular data patterns, then they can be flagged and investigated further. Predictive analytics can also apply the same processes to individuals, analysing their typical behaviour and interaction with a system. An example of such would be in a customer typically only withdraws cash from their bank from cashpoints during the day, and never more that £100, then when someone withdraws £700 online in the middle of the night, the transaction will be flagged by the system as a potential instance of identity fraud. If this type of individual anomaly is quite prevalent, or the risks it exposes are large, then more recursive investigations are required. Typically referred to as repetitive (or continuous) analysis of data, where scripts are run against large volumes of data to identify the anomalies as they occur, recursive investigations are key in preventing anomalies as they occur. Anomalies will be determined by comparing the data to a set of calculated statistical parameters, and outliers that fall outside of standard deviations are will require investigation (ACL 2014)

Until the start of this decade, retaining large quantities of data wasn't economically viable, and as a result data was only retained for a short period. However now, with big data tools such as the Hadoop framework and database management systems like MongoDB, businesses can store data across large scale systems and clusters. Such systems allow companies to extend the scale and speed they can process data, which in turn expedites the storage and analysis of security information. (Cardenas, Manadhata, Rajan 2013).

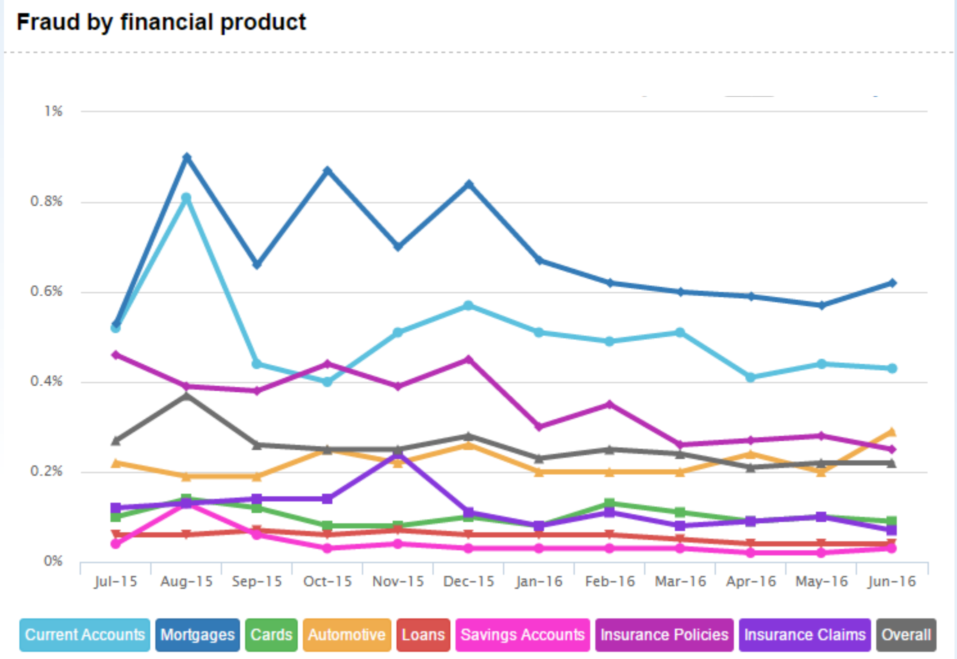

Figure 1: First-party fraud by financial product (Experian 2016)

Figure 1 shows the trends in different categories of first-party fraud between July 2015 and June 2016. The overall decrease over time seen in areas which are either financial or online clearly indicates that the increasing use of big data and data science is having a noticeable effect on combatting fraud. The only area that cannot be effectively analysed by big data, Automotive, is showing an increase, proving that without the use of prevention methods provided by big data, the occurrence of other types of fraud would also show an average increase. On average, the is a clear yet slow decrease in the levels of fraud presented on the graph. Cases of fraud undoubtedly rose in the climax and wake of the 2008 and 2014 financial crises following the bankruptcy of Lehman Brothers, who were involved in massive accounting fraud only detected in 2010 (AccountancyAge 2010).

Ethical aspects of Big Data

Mass storage of data about individuals comes with the responsibility of storing and using the data in compliance with laws and regulations. The Data Protection Act 1998 states that usage of personal information must follow the outlined data protection principles. One such principle is that personal information must only be used for limited, specifically stated purposes. This means that a business must state to the customer if they intend on using their data for security analytics. Another principle states that the information must be kept for no longer than absolutely necessary (Gov 2016). Data is often collected automatically by interactive technology from external sources such a social media, and then used to determine missing information, such as credit worthiness. When it is collected in this manner, data is not consciously provided by individuals, and so they may object certain uses of such data (although often it is stated in the user license agreement that their data may be used in such a way). It is debatable whether using the data for analytics is necessary or not, and if the data is fully not anonymised, potentially leaving an individual identifiable, then the business at fault is potentially breaking the law and breaching data protection rights. (ICO 2016)

Conclusion

It is clear that big data is having some effect on fraud, albeit rather slow and small in relation to the sheer amount and value of fraud-related incidents occurring globally. Over the coming years, more data will be collected, hypothetically increasing the effectiveness of big data and data science in fraud prevention by highlighting new patterns and as well as obscure anomalies previously undetected due to lack of sufficient data. Data analytics will also be aided by the inevitable increase in processing power in the coming years, as well as increased storage capacity.

Big data solutions in the field of fraud detection have clear potential in other fields, such as phishing and scam detection, as well as detection of malicious software using the same pattern-anomaly detection methods.

Flickr 2016

References

- Public Domain Pictures (2016) *Networking* [online] Available at: <http://www.publicdomainpictures.net/view-image.php?image=45242&picture=networking> [Accessed on: 4 Dec. 2016]
- Centre for Counter Fraud Studies (2016) *Annual Fraud Indicator 2016* [online] Available at: <http://www.port.ac.uk/media/contacts-and-departments/icjs/ccfs/Annual-Fraud-Indicator-2016.pdf> [Accessed on: 4 Dec. 2016]
- Fraud Research Center (2014) *The True Impact of Fraud* [online] Available at: <http://fraudresearchcenter.org/wp-content/uploads/2014/06/The-True-Impact-of-Fraud-Proceedings-Final.pdf> [Accessed on: 7 Dec. 2016]
- ACPO (2016) *Who is affected by fraud?* [online] Available at: <http://www.fraud-stoppers.info/about/who.html>[Accessed on: 8 Dec. 2016]
- Adrian Bănărescu (2015) *Detecting and Preventing Fraud with Data Analytics* [online] Available at: < http://www.sciencedirect.com/science/article/pii/S2212567115014859> [Accessed on: 7 Dec. 2016]
- Experian (2016) *Fraud Statistics* [online] Available at: <http://www.experian.co.uk/identity-and-fraud/fraud-statistics/> [Accessed on: 6 Dec. 2016]
- AccountancyAge (2010) *E&Y sued of Lehman's audit* [online] Available at: < https://www.accountancyage.com/aa/news/1934026/-sued-lehmans-audit> [Accessed on: 6 Dec. 2016]
- Analytics Magazine (2014) *Fighting Fraud: Employing big data and analytics to reduce fraud* [online] Available at: <http://analytics-magazine.org/fighting-fraud-employing-big-data-and-analytics-to-reduce-fraud/> [Accessed on: 4 Dec 2016]
- SQLAuthority (2013) *What is Big Data?* [online] Available at: < http://blog.sqlauthority.com/2013/10/02/big-data-what-is-big-data-3-vs-of-big-data-volume-velocity-and-variety-day-2-of-21/> [Accessed on
- ACL (2014) *Detecting and Preventing Fraud with Data Analytics* [online] Available at: <https://www.acl.com/portfolio-items/detecting-and-preventing-fraud-with-data-analytics/> [Accessed on: 6 Dec. 2016]
- Alvaro Cardenas, Pratyusa K. Manadhata, Sreeranga P.Rajan (2013) *Big Data Analytics for Security* [online] Available at: <https://www.computer.org/csdl/mags/sp/2013/06/msp2013060074-abs.html> [Accessed on: 6 Dec. 2016]
- Gov (2016) *Data Protection* [online] Available at: <https://www.gov.uk/data-protection/the-data-protection-act> [Accessed on: 7 Dec. 2016]
- ICO (2016) *Big data and data protection* [online] Available at: <https://ico.org.uk/media/1541/big-data-and-data-protection.pdf> [Accessed on 7 Dec. 2016]
- Flickr (2016) *Data Breach* [online] Available at: <https://www.flickr.com/photos/143601516@N03/29723649810/in/photostream/> [Accessed on: 7 Dec. 2016]