Fraud Detection

Introduction

"Merchants in the United States are losing approximately $190 billion a year to credit card fraud - much of it online", "The annual cost of fraud in the UK has been estimated at £193bn", "China Arrests Ezubo Executives in $7.6 Billion P2P Fraud Case "– common running theme on the biggest trading countries in the world and the most repeated word on them is "Fraud".

Fraud is not only affecting the biggest economies in the world but it is also being a barrier to some of the most exciting, emerging markets in the world. Emerging markets bring in new products to the world and opens new products to trade but it is being tethered because countries with emerging markets are losing billions of cash because of 'Fraud', it is also affecting the countries with biggest markets and the victims are the poor people who live in it. For example, in a study recently conducted by 'FTI Journal', they found out that 150 North American and European –based companies with emerging markets were making more than $1 billion but along with the growth, they were also losing more than $1 billion for the last five years. This was not because of poor product, marketing or supply chain decisions but from 'fraud'.

For these reasons, I have chosen to do Fraud detection domain. I want to find out how 'fraud' happens despite big companies investing money into military grade fraud detection technologies which affects their income turning into profits. I want to find out what systems are being used by the big companies in countries with big markets and in emerging markets to prevent 'fraud' which is affecting the economies around the world.

A description of how Big Data/Data Science is being used within the fraud detection, including an overall description of the specific techniques and technologies that are used (showing evidence of research)

Corporations around the world are using several innovative technologies to detect fraud detection. Before these technologies are even used there needs to be some data, and if you want to find patterns that predicts future trends then you are going to need a Big data. When you have a big data, you are always going to find something with it. "It's not unreasonable to think that the more data you can get access to that you might discover something of predictive value."- said Fred Cate, director of the Centre for Applied Cyber Security Research. When the data is big and ready, these are the innovative technologies used by the big corporations to detect fraud:

Use of Machine Learning and Analytics to predict Online Fraud – "The cyber security arm RSA of the US big data company EMC uses machine learning big data analytics methodologies to prevent online fraud. 500,000 attacks detected in 8 years"

IBM Watson - IBM Watson API is a cognitive service that simplifies the process of preparing data and makes it easier to run predictive analysis through machine translation, message resonance, question & answers and user modelling.

Microsoft Azure Machine Learning API – ability to create configurable R module so own R language code can be incorporated to train or predict tasks. Allows python scripts to be included.

Google Prediction API - lets data scientists tap into Google's machine learning algorithms to crunch big data and give possible results to make analysis such as spam detection, purchase prediction, intelligent routing and more.

Amazon Machine Learning API – simplifies the process of making predictions that require lots of expertise around model building, data cleansing and statistical analysis by letting data scientists or analysts predict if a user will pay by the first week or first month, detect fake users, bots or spammers in the systems.

BigML - BigML helps create a descriptive model to understand the relationships between the various attributes in the complex data and the predicted attributes so that business analysts can play with the what-if scenarios.

These are the machine learning methodologies that are used by the companies around the world, all the machine learning API are unique to each other, however they are not perfect and thus I think why there are still fraudulent activities going on around the world, but these methodologies help reduce the damage to the minimum and control bigger loss. These machines take in big data and predicts probable fraudulent activities and it does it better than a human can, even though it is not perfect. However, I think in the future along with the advancements of technologies machine learning will grow and lead to more innovations, making it close to perfect and assist people from being victims of 'fraud'.

Reflection on whether Big Data/Data Science solutions are successfully meeting business objectives

Big Data/Data science solutions are not perfect but it is helping businesses meet their objectives. For example, one of the biggest online website PayPal, "is beating the bad guys with machine learning". PayPal uses Amazon Machine learning API and it is standing out and retuning 'fraud' with immediate payback. The smart API has fraud detection and it would not have been possible if machine learning didn't exist. PayPal users are targeted with many numbers of scams via email, fake weblinks etc. however with the help of Machine learning API, PayPal is a pioneer in risk management. According to Dr. Hui Wang, senior director for PayPal – PayPal protects it's user from two levels of security, first he says "we can separate the good from the bad with one straight line", and if this fails then PayPal has another advanced layer of security on top of it and it is called multiples lines or curve the lines, which works like neural network and it imitates how neurons work in the human world therefore making judgements like if it is raining, then the user would say I'll take an umbrella. This is an advanced big data solution and is successfully helping business like PayPal meet their business objectives which is to protect their valuable users from scam and fraud and in return PayPal is getting back the best reputation they can which helps not only attract new customers but also keep their existing ones, this means they are meetings their business objective by protecting the business from fraud.

Conclusion and closing remarks

In conclusion, I think that big data and its techniques has helped big businesses prevent fraud and sustain longer in the trading market because of their financial power to have a good enough fraud detection system such as PayPal. However, it is not the same for emerging businesses as they do not have the income and customers as the market leaders and they cannot afford to have a system that is probably worth more than their whole business combined. The use of API's has made the fraud system more fast and intelligent through smart algorithms and predictions. I think that the fraud system has evolved and become more efficient than the past because of technology advancements and with it, it's going to keep getting more and more intelligent in-terms of getting the predictions more accurate and reliable. It is a technological era and I think it is only going to be better in the future.

A description on how you think that Big Data/Data Science can be further used within that fraud detection

The amount of data we generate everyday through our smart devices are large but when we produce large data everyday over the period of weeks, months and years, it becomes big data and it is all being uploaded and stored in a virtual location that we do not know. Data volumes we produce will continue to grow which means more data about us uploaded every day, such as the locations we went to, the amount we spent running, walking, standing and sitting etc. This data can be further used in Fraud detection by monitoring these data in real-time for every single user. This will mean if there is an unexpected behaviour of the user/users compared to the data collected from other times then the system would warn the data scientists about the behaviour, displaying where the user went off track from the usual road/pattern, user sprinted from than he/she should have or the user stalled in one area more than he/she should have. This will give the system real-time response time and therefore also giving the system more faster time to react if there is a fraudulent behaviour. However, to collect such information would mean that the devices would have to be uploading data in real-time and the user will not have any choice about giving the data or not. This is breaching their privacy rights which is protected by law. This also breaches ethical issue as the data scientists themselves might not want to get their data uploaded in real-time. These data would also be uploaded in a virtual location and if therefore the virtual location is breached the data can be stolen and used for the wrong reasons, such as the bigger the data, the bigger the potential and therefore from these data passwords, personal information could be extracted and this is a danger not only to the user but also to the business and the business would fall into lawsuits from it's user if the data is stolen because it is violating their privacy.

Analysis of how the Big Data/Data Science ideas and solutions in fraud detection could be expanded to other Security domains and how knowledge and experience can be transferred.

I know that API are used to detect fraud, however it could also be expanded to other security domains such as to prevent Terrorism. The idea will be that the Big data would have a system where it would screen the data that from all around the world into their respective government databases which would be an enormously big data. Then the system could use Microsoft Azure Machine Learning API to sort the data through python script and another API such as Google Prediction API to combine the data and build a statistical data model and normalise the data. The data could be display/predicted by the same API to show the next probable terrorist attack. However, this collides with privacy laws of different countries as most countries have different privacy laws and it will be against the people's rights. It would be also unethical to give data of people from one country to another because no one would know how the data will be handled into another country which has different set of laws and jurisdictions.

Another way for Fraud detection's knowledge and experience can be transferred and used into other security domain is identity theft. As I have mentioned above that we upload big data of ourselves to every year to a virtual location through our smart devices. These data contain our personal information such as our full name, dob, sex, house address, credit/debit cards and family details.

This can be expanded as to verify a person's identity real-time. Usually when a person is asked to verify it's identity, It is asked to enter their personal information voluntarily and when the information is entered, it is then checked with the information stored in virtual servers. This can be expanded to checking the data real-time which means the data being checked is not the one from the server but from their smart devices which they are entering the data from in real-time. This will help verify them to check if it is the right person. It's could be more effective this way as the data being checked is real-time which means it's the most up-to-date information there is. However, again this raises privacy and ethical issues as it is not allowed to track a person in real-time and the person would not want to get tracked in real-time as well.

References:
http://www.bbc.co.uk/news/uk-36379546
https://www.bloomberg.com/news/articles/2016-02-01/china-arrests-ezubo-executives-in-7-6-billion-p2p-fraud-case
http://www.forbes.com/sites/haydnshaughnessy/2011/03/24/solving-the-190-billion-annual-fraud-scam-more-on-jumio/#2051de9b7db4
https://hbr.org/2015/09/what-companies-have-learned-from-losing-billions-in-emerging-markets
http://www.kdnuggets.com/2015/12/big-data-science-security-fraud-detection.html
http://www.kdnuggets.com/2015/11/machine-learning-apis-data-science.html
http://www-03.ibm.com/software/products/en/category/data-security
http://www.infoworld.com/article/2907877/machine-learning/how-paypal-reduces-fraud-with-machine-learning.html
http://www.csoonline.com/article/2855641/big-data-security/the-5-worst-big-data-privacy-risks-and-how-to-guard-against-them.html