Austin Anderson

Report on Big Data and Machine Learning in the Banking Industry

Big data and machine learning are becoming more and more integral parts of the banking industry every day. Whether it's gathering large amounts of data in order to preserve records, providing customers with convenient access to their own data, or to making automated lending decisions to those who have no credit score and no way of obtaining one, machine learning and big data will be key factors in the progression of this industry.

First Utah Bank; Traditional Consumer banks usage

First Utah Bank is a traditional financial institution, in that they provide checking accounts, mortgages, physical branches, etc. Their perspective on the ability to utilize and analyze big data is a valuable one as it provides the view of big data from the position of a relatively small institution. I spoke to Amy Foulks, CIO at First Utah Bank, and the following information came from an interview with her.

First Utah Bank operates with two primary sets of data: Customer Data, which is stored by a third party vendor, and shared files, data used daily by employees of the bank that is stored either locally or in a secure server farm. The shared file set is a much smaller set of data, things like excel spreadsheets, loan documents, etc. This set of data is kept to a minimum for security purposes, as the more of this data there is the more risk there is that it is breached.

Customer Data is everything that the bank gathers from its customer's activity, which is primarily transactional. They gather data "every second" on their customer's activity. This is

primarily for record keeping; the bank needs to have a record of every transaction for the protection of both the customer, the bank, and the transaction partner. The necessity for this data ranges from fraud protection to government inquires, but this is one of the primary services modern banks serve; they offer detailed transaction records for customers to be able to access. Customers both implicitly and explicitly agree to having this data collected on them when they choose to open an account at the bank. There is no opt out of this type of data collection, as the bank needs the data, just as much as the consumer might need any transactional records the bank provides to them using this data. Clients have full view of the data that the bank is collecting on them, as they can go into their transactional account records and see for themselves.

First Utah banks primary usage of the data they gather beyond providing consumers with transactional records is "complete relationships" with businesses to which they provide loans. The bank looks at financial trends for businesses in order to know what kind of services they can offer at what time to clients in order to entice them into doing their business exclusively at First Utah Bank. For example, businesses doing well financially might be approached for expansion loans or some other kind of service which the bank would like to offer them. Data analysis is also utilized to determine who the most important customers are, through a "Relationship Profitability" program. The data garnered from these programs helps First Utah Bank to determine which clients are most profitable to work with and helps to determine who to dedicate Human Resources to. This is not a practice among personal accounts, because the bank is more small business focused than personal.

As the bank does not have as much of a consumer focus and is not marketing heavy, so they do not target consumer categories for services through marketing or data analysis. This is a common trend among other more personal account focused banks, though, and they would be "stupid not to" use data in this manner, as it has become a necessary practice in order for banks to stay in business in this day and age. Along with that, merchant/bank relationships are becoming more and more common. The example presented to me was a community bank offering a promotion where a bank customer using their Debit card at a local restaurant might receive 5% off their restaurant purchase. This kind of relationship is usually formed through analyzing local spending data and ascertaining which local businesses would be most effectively targeted by this type of campaign. Another usage of big data, and something that's been around for a while, is a database which all banks have access to keep track of "unbankable people", those who abuse bank accounts and are unwise for banks to do business with.

Another use of Data Analysis among banks that this is not exclusive to banks, is determining the need for different types of services. For example, data might suggest certain physical locations of a franchise might not need to exist, while certain digital functions might need to be improved as they are utilized more often than others. This kind of data is especially relevant to banking, as physical locations are becoming less and less useful as mobile applications are becoming more useful, and as the money used maintaining physical locations might be better used upgrading online functionality.

As for security, all data is encrypted and stored by a third party vendor. The bank does not share or sell data gathered on its customers with anybody except the government. Customer

Data is usually retained for 7 Years; however older data can sometimes be recovered if requested. While First Utah Bank isn't aware of any breeches of data, there is a sentiment among banks that "If you haven't been breached, you just don't know that you have." Even so, the bank takes their customers data security very seriously, and all customer data is encrypted. However, banking is one industry where the protection of data by not retaining it is not really an option.

According to Mrs. Foulks, the banking industry is in the middle of a "Netflix/Blockbuster" movement, that analogy referencing the rapid decline of Blockbuster when instant streaming Netflix became available. The Netflix in this situation is FinTech, an industry that has created much more convenient methods of obtaining loans online without requiring human approval or trips to the bank. Traditional banks are seeing large amounts of competition rise up from the FinTech industry, and must decide how to keep up and whether to incorporate or create their own types of FinTech in order to stay relevant. The factor that traditional banks have going for them is that "regulators expect banks to prove every machine learning model they use", as the FCRA requires that no discrimination was factored into credit decisions.

Lenddo, Mirador, and FinTech

Within the banking ecosystem, the sector with the greatest opportunity to benefit from big data and machine learning is the extending of credit. Whether it be small personal loans or large business loans, the ability to predict who will pay you back and who won't be a power banks are loathe to get their hands on. Big data and machine learning algorithms aren't crystal

balls, but they can run deep analysis on thousands of factors in order to help make better predictions on who will and won't repay loans.

One of the leading businesses in this field is Lenddo. Lenddo provides algorithmic decision making API's to banks in order to help them analyze and make lending decisions without requiring a credit score. I interviewed Lenddo's Florentine Lenoir (Marketing and Business Development Director at Lenddo), told me that their system can gather data from up to 12,000 data points, resulting in around 300 compounded weighted variables. Lenddo prides itself on being able to help offer loans to those without credit scores in less economically developed countries; they gather variables based on Facebook, Twitter, Email, Text messages, Phone Calls, smartphone metadata, etc. These data points are all funneled into an algorithm to determine behavioral patterns of the person, and match those behavioral patterns to those of clients most likely to repay loans. Those without one or more of these primary data sources have no disadvantage, as long as Lenddo's algorithm can come to a satisfactory decision about the client's behavior.

With regards to privacy concerns, Lenddo does not keep any data they analyze. Their system gathers the necessary data, processes it, makes a recommendation to the lending entity, and then does not store the data. Any data inputted into the system is encrypted in transit and known only by the algorithm, and not even the lender can see the raw inputted data, only the resultant decision the algorithm has made. Along with their practice of not keeping data they analyze is the philosophy that the data does not belong to them; Lenddo believes that data belongs to the end user that is benefitting from their services, and ultimately they have only the

permission to gather and use it, not to keep it. They do feed results from their analysis into their model, but their algorithm doesn't keep the data which trains it, only learns from it and adjusts it's model accordingly. The algorithm uses behavior patterns and not demographic information in order to make its decision, which negates its ability to accidently stray into biased decision making, but it is also manually tweaked in order to assure certain values are not given too much weight and the algorithm remains focused on more well researched variables.

The algorithm is customized for each bank that implements it, and then the decision on what info to share on how the algorithm makes its decisions is left to the bank to disclose if it so chooses. The methodology of the decision making process is explained to each bank as the algorithm is implemented, and the weight given to each variable is allowed to be adjusted by the bank. Once that is done, the bank can choose whether or not to disclose how it's decisions are made and how much of the methodology it wants to disclose to the end client. However, Lenddo is not meant to be the sole factor in the decision making process; it is meant to be a supplementary resource to help replace the lack of traditional credit scores or other methods of determining credit worthiness. Lenddo does not want to be a competitor to banks; it wants to be a compliment to them, providing another factor in the loan decision.

Lenddo is certainly not the only FinTech business gaining ground on traditional banks and credit institutions. Companies like Mirador are providing services similar to Lenddo's for Small businesses companies. Startups like Earnest are stepping into the student loan and refinancing market. The rise of FinTech is, as Amy Foulks put it, a "Netflix/Blockbuster" moment for the banking industry; whether banks choose to stick to their roots and rely on

traditional brick and mortar service that FinTech can't provide, or, as Amy indicated to me most banks are doing, adopt some sort of FinTech and incorporate it into their own business model.

Some concerns are raised by the sudden rise of FinTech, key among them being that the industry is widely unregulated at the moment and could be operating with models that promote discriminate impact. While most FinTech companies, such as Lenddo, are meant as a "recommendation", and not a determinative factor in the decision to lend, the likeliness that they will have to fall in line with federal regulation, and be able to prove their Algorithm doesn't make decisions utilizing discriminatory biases, is very likely. In what manner and how soon these actions will be taken will really depends on the FTC, and whether any high profile law suits come to public attention. Along with that, companies like Lenddo don't base their model on demographic or protected categories; they base recommendations on behavioral patterns, which they claim has no dramatic demographic bias. To go along with that, positive behavioral data patterns are adjusted for the climate in which they're being implemented; each implementation of the algorithm is customized for the bank it is being implemented at.

However, many FinTech firms are *not* being implemented at banks but lending directly to the client, which brings them from a "recommendation" to a "decision", and thus bringing them under closer scrutiny by the FTC. The industry as a whole has not received any major governmental inquiry, but it is likely to happen in the near future and, as Amy Foulks said, "the regulators expect each of these companies to be able to prove their model".

This brings the importance of explainable and accountable algorithms to the forefront of the FinTech industry. This also leads into the problems of transparency. While companies like

Kabbage state their algorithm utilizes "thousands of data points", it (from my research) gives no indication as to what those data points are until you go to apply for a loan, and doesn't present its methodology or any assurance of non-discriminatory practices to the consumer. While it may not be a requirement to disclose such things, the lack of such transparency may raise concerns among consumers and certainly among the FTC. And convincing the FTC that your algorithm is fair is only the first part; if consumers believe that your algorithm is discriminating against them, at very least the company will lose their business, and at worst they will be dragged into unwanted lawsuits.

Big data and machine learning have great potential in improving the banking industry. From quicker, easier loans, better investments for both customers and banks, and protecting the welfare of a bank and its customers. There are numerous issues that could result in incorrect implementation or design of these algorithms, but ultimately their ability and efficiency will provide great leaps forward in the ability and confidence of banks to provide loans and other useful services to their customers.