Group 4 Poster Project – Milestone 2

Title: As The World Churns

Customer Data, Business Models, and Data Science: Predicting Customer Trends and Behaviors

Introduction:

As our economy and corporations begin to operate in a global context, there have been increasing efforts to retain customers. Frequent acquisition and loss of customers is defined as customer churn and has been a particular area of focus in data science. It is important to businesses as "it is directly tied to firm profitability" [4]. The costs of keeping a customer are usually less than the costs of recruiting new customers [21]. This is why it is becoming increasingly important to use data science techniques and advanced analytics to predict which customers are vulnerable to leaving. It can be difficult to differentiate between customers who will respond to interventions and those who will not [4]. In addition, excessive customer turnover can be a sign of potential fraudulent activity. This is complicated by the fact that technology can serve two purposes to become closer to customers as well as alienate them [21].

There is also the risk of customer churn on customers that were won back after churning originally. This makes the situation even more complex to further analyze [16]. There are multiple fields to study in this; some organizations use predictive modeling by studying customer behavior while others focus on more traditional demographics (behavioral attributes and financial churn prediction source).

Traditionally, the data science technique of k-means clustering is used to determine risk of customer churn (clustering prediction techniques source). However, there are other methods available to help predict this risk. In some instances, other methods like decision tree analysis are more valid and the field continues to become more diverse.

The risk can be more than financial; in certain insurance industries, customer churn can signify loss of critical healthcare coverage and can significantly impact a person's health. In fact, data science technique and predictive analytics in particular are being applied to treat cancer and impact healthcare outcomes [17,20]. Therefore, it benefits us all both economically and personally to obtain further insight into customer churn, its prediction, and its avoidance (if at possible). This project aims to critically evaluate the current state of customer churn and customer behavior in the financial and insurance industry, propose a data science framework and algorithm to ascertain customer churn, and reflect on the future direction of this field.

Why is this Data Science?

"Data is the new oil for all industries, and data science is the power that drives the industry."

Data Science transforms raw data into useful information. Industries need data to help them make careful decisions and is used in almost every industry. Some of the main sectors are health, finance, banking, business, startups, etc. Companies use the data to analyze their marketing strategies and create better ads. The industry needs data scientists to help them make smarter decisions.

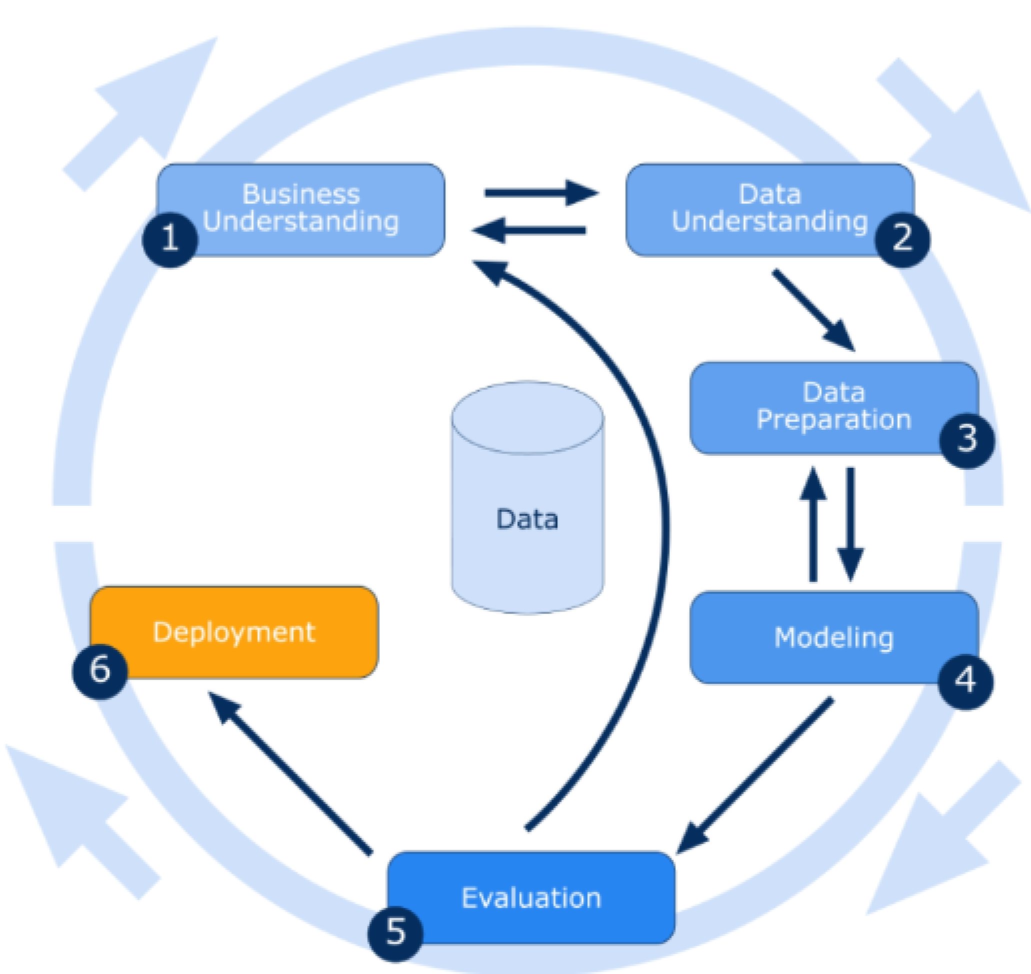
Let us understand the importance of data science in our lives. Getting a ride with Uber is easy. We simply open the app, set your pick-up and drop-off location, book a taxi, get picked up and pay with your phone. Each time you book a taxi through Uber, you will receive an estimated fare and the time it takes to travel the route. How can these applications display all of the information they do? The answer is the data science. Using data science predictive analytics, Uber can determine the pick-up, drop-off location and arrival time.

Technology giants such as Facebook, Amazon and Google are constantly working in the field of machine learning and data science. Data science encompasses processes such as purging, processing, and analyzing data. A data scientist collects data from multiple sources, e.g. From surveys and physical data plots. Then, data was passed through strict algorithms to extract important information from the data and create a record. This record could also be used to parse algorithms to make more sense.

According to DOMO research, "More than 2.5 billion bytes of data are created every day and will only grow from there, and by 2020 an estimated 1.7 MB of data will be generated per second for every human being on Earth."

Predicting customer attrition and churn is just another way for data science to flex its capabilities in a modern world.

The Crisp-DM Process:



Deliverables:

Main Goal: Reduce customer Churn by 50%

Goals has been divided into 3 main states namely short, medium- and short-term goals.

· The short-term goal is to reduce churn by 20%

· The medium-term goal is to reduce churn by 20%

· The Long-term goal is to reduce churn by 10%

Build models using various machine learning algorithms like Neural Networks, Decision Tree, K-means, Random Forest and Logistic Regression to understand root causes of churn and identify at-risk customers.

Create dashboards and visualizations on the fly to analyze customer data with the Tableau to make explanation easy to the stake holders

Test and analyze ideas gathered to enable company focus resources on customers that are high leaving. Adopt customer focus strategies to retain and reduce the rate of customer churn.

Refine and repeat the process quarterly to measure progress

Data mining Section
Data mining using Tableau(shown with examples and graphs/charts)
- Working wit aliases
- Adding reference line
- Looking for anomalies
- Validating approach
- Some advance topics in tableau

Modeling
- A short description
- Data types- a sort description(relates to the models below)
- Logistic regression modeling
    - Description
    - Simple Example with sample data set, gretl model building and analysis
    - False positive & false negatives-
    - Confusion matrix
    - Interpreting coefficients of a logistic regression
- Geodemographic segmentation modeling using logistic regression modeling-
    - a more complex example with bank customer data with many independent variables and one dependent variable with value stayed/exited
    - description
    - Example with sample data set, gretl model building and analysis

Conclusion:

Customer churn for any company is costly, but it is especially expensive in customer service areas. The financial and insurance sector has a large amount of competition and with new digital-only institutions coming into the arena the amount of competition has only grown. The biggest reason sighted for leaving a bank was “poor service” and high fees. High fees are intrinsic and dependent upon each company’s profitability. Each institution attempts to show its value by offering the most economic service for the customer while remaining profitable. The poor service was the driver for 56% of the individuals that changed banks [13]. Financial institutions and insurance companies struggle with customer churn as the institutions do not usually see the customer leaving before they have closed all their account and left. Normally, they never get a chance to try and attempt to retain the customer.

There are challenges to predicting customer attrition at a higher rate. The most obvious issue is in the data being used. The data usually require cleansing or preparation. The most common decision tree analysis usually lacks flexibility. Decision trees are based on expectations and if the data arrives with unexpected inputs the model has a lower accuracy rate [19]. The above being just one example most model suffer from incorrectly selected variables and that can have a big influence on the type of model being run. The logit-model is one example were selecting the correct variables have a big influence on the success rating of the model.

Many companies have implemented different techniques to manage customer attrition like neuro-fuzzy, k-means, spatio-temporal, liner regression, decision tree, logit-model (logistics regression) or some combination of two or more techniques. These are all used to increase the ability to predict with the highest accuracy which customer will defect to other institutions. One item does standout is the type of data being used. Many of the models used some type of tenure information or social behavioral input. Due to the variations in the type of customer data customer churn is still hard to predict with a high rate of accuracy. Most human decisions are emotional and usually follow some type of adverse experience with a service the institution provides. Some sources advocate a behavioral analysis technique to predict which customers will respond positively to attempts to retain them as customers to better concentrate resources [4].

Conclusion Version for Poster:

Customer churn in the financial and insurance sector is high. Companies struggle to identify customers who are likely to leave before they have left. Surveys are infrequent and a poor service might not show up on a survey. To increase customer lifetime value (CLTV), organizations need to understand the correct behavioral attributes and build predictive models using new and traditional data science techniques like k-means or spatio-temporal algorithms [14]. This helps in selecting the correct behavioral traits based on transactions and other demographic behaviors to identify customer churn and determine if a customer is a good candidate to be retained. Once identified, measures can be undertaken to prevent customer churn before it is too late.

References Cited:

1. Abbasimehr, H., Setak, M., & Soroor, J. (2013). A framework for identification of high-value customers by including social network-based variables for churn prediction using neuro-fuzzy techniques. International Journal of Production Research, 51(4), 1279–1294. https://doi-org.ezproxy.bellevue.edu/10.1080/00207543.2012.707342

2. Al-Shboul, B., Faris, H., & Ghatasheh, N. (2015). Initializing Genetic Programming Using Fuzzy Clustering and Its Application in Churn Prediction in the Telecom Industry. *Malaysian Journal of Computer Science*, *28*(3), 213–220. https://doi-org.ezproxy.bellevue.edu/10.22452/mjcs.vol28no3.3
3. Amin, Adnan & Anwar, Sajid & Adnan, Awais & Nawaz, Muhammad & Aloufi, Khalid & Hussain, Amir & Huang, Kaizhu. (2016). Customer Churn Prediction in Telecommunication Sector using Rough Set Approach. Neurocomputing. 10.1016/j.neucom.2016.12.009. https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0191-6.ris
4. Ascarza, E. (2018). Retention Futility: Targeting High-Risk Customers Might be Ineffective. *Journal of Marketing Research*, *55*(1), 80–98. doi: 10.1509/jmr.16.0163
5. Brugnoli-Ensin, I., & Mulligan, J. (2018). Instability in Insurance Coverage: The Impacts of Churn in Rhode Island, 2014-2017. *Rhode Island Medical Journal*.
6. Della Torre, E., Zatzick, C. D., Sikora, D., & Solari, L. (2018). Workforce churning, human capital disruption, and organisational performance in different technological contexts. *Human Resource Management Journal*, *28*(1), 112–127. https://doi-org.ezproxy.bellevue.edu/10.1111/1748-8583.12167
7. Farquhar, J. D. (2005). Retaining customers in UK financial services: The retailers tale. *The Service Industries Journal*, *25*(8), 1029–1044. doi: 10.1080/02642060500237478
8. Ferreira, P., Telang, R., & Matos, M. G. D. (2019). Effect of Friends' Churn on Consumer Behavior in Mobile Networks. *Journal of Management Information Systems*, *36*(2), 355–390. doi: 10.1080/07421222.2019.1598683

9. Flores-Méndez, M. R., Postigo-Boix, M., Melús-Moreno, J. L., & Stiller, B. (2018). A model for the mobile market based on customers profile to analyze the churning process. *Wireless Networks (10220038)*, *24*(2), 409–422. https://doi-org.ezproxy.bellevue.edu/10.1007/s11276-016-1334-8
10. Gunthera, C-C., Tvetea, I., Aasa, K., Sandnesb, G., & Rorganc O. (2014). Modelling and predicting customer churn from an insurance company. Scandinavian Actuarial Journal 2014 Vol. 2014, No. 1, 58–71
11. Imarticus.org. (2018, October 8). Why is Data Science So Famous? - Imarticus Learning. Retrieved from https://imarticus.org/why-is-data-science-so-famous/.
12. Jennings, A., & Stratagree. (2015, December 25). The 4 D's of Customer Attrition. Retrieved from https://thefinancialbrand.com/55772/banking-customer-attrition-analysis/.
13. Kaemingk, D. (2018, August 29). Reducing customer churn for banks and financial institutions. Retrieved from https://www.qualtrics.com/blog/customer-churn-banking/.
14. Kaya, E., Dong, X., Suhara, Y., Balcisoy, S., Bozkaya, B., & Pentland, A. "S. (2018). Behavioral attributes and financial churn prediction. *EPJ Data Science*, *7*(1). doi: 10.1140/epjds/s13688-018-0165-5
15. Keramati, A., Ghaneei, H., & Mirmohammadi, S. M. (2016). Developing a prediction model for customer churn from electronic banking services using data mining. *Financial Innovation*, *2*(1). doi: 10.1186/s40854-016-0029-6
16. Kumar, V., Leszkiewicz, A., & Herbst, A. (2018). Are you Back for Good or Still Shopping Around? Investigating Customers Repeat Churn Behavior. *Journal of Marketing Research*, *55*(2), 208–225. doi: 10.1509/jmr.16.0623

17. Menden, M. P., Iorio, F., Garnett, M., Mcdermott, U., Benes, C. H., Ballester, P. J., & Saez-Rodriguez, J. (2013). Machine Learning Prediction of Cancer Cell Sensitivity to Drugs Based on Genomic and Chemical Properties. *PLoS ONE*, *8*(4). doi: 10.1371/journal.pone.0061318

18. Mousavirad, S. J., & Ebrahimpour-Komleh, H. (n.d.). A Comparative Study on Medical Diagnosis Using Predictive Data Mining. *Data Mining and Analysis in the Engineering Field Advances in Data Mining and Database Management*, 327–360. doi: 10.4018/978-1-4666-6086-1.ch017

19. Nayab, N. (2019). A Review of Decision Tree Disadvantages https://www.brighthubpm.com/project-planning/106005-disadvantages-to-using-decision-trees/

20. Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the Future — Big Data, Machine Learning, and Clinical Medicine. *New England Journal of Medicine*, *375*(13), 1216–1219. doi: 10.1056/nejmp1606181

21. Panaggio, T. (2015, May). The Customer Comes First - Always. *USA Today*, pp. 52–54.

22. Quora. (2017, October 25). Why Data Science Is Such A Hot Career Right Now. Retrieved from https://www.forbes.com/sites/quora/2017/10/25/why-data-science-is-such-a-hot-career-right-now/#72bf6ee9106b.

23. Rachid, A. D., Abdellah, A., Belaid, B., & Rachid, L. (2018). Clustering Prediction Techniques in Defining and Predicting Customers Defection: The Case of E-Commerce Context. *International Journal of Electrical and Computer Engineering*, *8*(4), 2367–2383.

24. Roman, O. (2019). Churn prediction. https://towardsdatascience.com/churn-prediction-770d6cb582a5

25. Saha, J. M. (2011). Business sustainability amidst global churning - paradoxes and dilemmas. *GMJ*, *5*(1-2), 19–24.

26. Syahida Binti, M. Z., & Ameer, R. (2010). Turnaround prediction of distressed companies: Evidence from malaysia. Journal of Financial Reporting and Accounting, 8(2), 143-159. doi:http://dx.doi.org.ezproxy.bellevue.edu/10.1108/19852511011088398

27. THE FINANCIAL BRAND FORUM 2020 — Discover the big ideas disrupting banking and explore the latest trends redefining the future of financial marketing at the FORUM 2020. The world's most elite conference on marketing. (2017, November 2). Plug Those Leaks: Stop Attrition From Stalling Your Growth Strategy. Retrieved from https://thefinancialbrand.com/68371/banking-customer-acquisition-attrition-growth-strategy/.

28. Thompson, R. (2017, February 27). Understanding Data Science and Why It's So Important. Retrieved from https://blog.alexa.com/know-data-science-important/.

29. Vijaya, J., & Sivasankar, E. (2018). Computing efficient features using rough set theory combined with ensemble classification techniques to improve the customer churn prediction in telecommunication sector. Computing, 100(8), 839–860. https://doi-org.ezproxy.bellevue.edu/10.1007/s00607-018-0633-6.

30. Zoric, A. B. (2016). Predicting Customer Churn in Banking Industry using Neural Networks. *Interdisciplinary Description of Complex Systems*, *14*(2), 116–124. doi: 10.7906/indecs.14.2.1