Musing 3

Alex Christensen

2/22/2017

Data analytics has been termed as the “sexiest job of the 21st century” (Morris, 2017). Almost every company, regardless of if they are a technology company or not, can benefit from data analytics. Some companies, like Amazon and Netflix, have much of their business model based off analytics, specifically predictive analytics. Amazon suggests products that you might like based on your browsing history, as well as that of other customers who have viewed similar products. Meanwhile, Netflix attributes 75% of customer content viewing to their suggestion algorithm (Albanesius, 2017). This algorithm depends heavily on analyzing your data, and that of customers like you. However, there is one big hurdle that keeps many companies from capitalizing on the potential benefits of data analytics: data quality.

Data is abundant in today’s world. There are countless sources of data, from traditional data, like census data, to more complex data, like a Twitter feed or YouTube channel. Additionally, companies have their own, internal data generated by website traffic, sales statistics, and transactional data. This brings up the question of how this data can be properly managed and processed. Dirty data will ruin even the best analytics model.

This dirty data can come from both the internal and external data. External data often has more of an unknown element. We don’t always know where/how the data was tracked, who has had access to modify the data, and if the data is complete and accurate. Internal data can sometimes poise an even greater risk. If we assume that we can trust our own data simply because our company generated it, we put ourselves at risk for data infected with user entry errors. Simply put, we cannot take any data as being 100% clean and accurate, without looking into it further.

“What's more, the tools for ensuring data quality, managing data, and storing data have increased in number and complexity, too” (Davis, 2017). With the amounts of data being so large, and always increasing, these tools help those in the data cleansing to clean large datasets.

In conclusion, if a company wants to make the most of what data analytics has to offer, they need to ensure the quality of the data is up to par. This isn’t some small task that can be overlooked. It is a vital part of ensuring the accuracy of a data model. While data cleansing might not sound as sexy as data analytics itself, it is just as important.

# References

Albanesius, C. (2017, 2 22). *75 Percent of Netflix Viewing Based on Recommendations*. Retrieved from PC Mag: http://www.pcmag.com/article2/0,2817,2402739,00.asp

Davis, J. (2017, 2 22). *How To Tackle Today's Data Quality Challenges*. Retrieved from Information Week: http://www.informationweek.com/big-data/how-to-tackle-todays-data-quality-challenges/d/d-id/1328119?

Morris, C. (2017, 2 22). *The Sexiest Job of the 21st Century: Data Analyst*. Retrieved from CNBC: http://www.cnbc.com/id/100792215