aly 6000 | case study

Leveraging Big Data

predictive modeling to achive business goals

Manoj Kashyap

2020

# INTRODUCTION

Big data is a field that allows us to extract, transform and analyse huge amount of complex data with traditional data-processing application software. If we go couple of decades back, though huge amount of data was generated, it was not stored and processed to make use of it.

A screenshot of a social media post

Description automatically generated

Figure 1(Coronel, C & Morris, S)

Since 1970’s as the data storage and reporting systems started to evolve, people started to leverage huge amount of data that was collected upon converting data to information, information to knowledge and knowledge to wisdom. Introduction of desktops in 1980’s got people in touch with centralized data, then online analytical processing systems (OLAPs) boosted business intelligence process. Then came the smartphone generation since 2010, that is when majority of the people started to use social medias, online shopping were introduced, mobile banking made transaction process so comfortable. This gave rise to creative way of collection, organizing, processing and analysing the Big Data.

(reference: Coronel, C & Morris, S. (2017). Database Systems: Design, Implementation and Management. 13th Edition)

Since the industries start to get ample data, there were two options for the Business Intelligence team, either let it go waste and loose the market or make use of it to predict the customer needs and ace the market. So industries started to perform statistical operations on data and built models that can predict future outcomes, this process is called Predictive Modeling or Predictive Analytics.

PayPal leveraging Big Data Analytics

PayPal had about 80 million active users as per 2010, where it was raised exponentially to about 305 millions by the end of 2019. They process over 10+ million login every day, with more than 13 million transactions processing more than 1.1 PB of data. Around $250 billion worth of payments are processed every year with $315.3 million in payments per day. Today PayPal stands out as the top Payment Company in the world in more than 200 countries and supporting 25 currencies. 21 years old PayPal are extensively using Big Data Analytics and Data Science for all their processing .

Ample data generated in PayPal went in vain before the advent of Hadoop since traditional databases couldn’t manage them. Upon introduction of Hadoop with existing traditional database platforms, Data Scientists began to run exploratory queries for hypothesis testing on data stored in Hadoop. This helped Business Intelligence team to find answers to all of their business related questions. “PayPal’s data mining systems are built on machine learning algorithms that are written in Java and Python and run on top of Hadoop to mine complex data models for valuable insights”( dezyre.com). Data Science also helped PayPal for Fraudulent detection.

Since data is collected from smartphones, tablets, in-store and from other online sources including social medias, PayPal started to track customer’s requirements and their purchase history. PayPal used this data to send customized offers and discounts to individual customers from merchants. PayPal treat both Merchants and Customers as their own customers, they used big data analytics to analyse customer taste, purchase history and location to send personalized adds and relevant offers to increase their service quality. Predictive data models of PayPal predicted their customers future transactions with an accuracy of 69%. Natural Language Processing algorithms were used to enriching customer experience, since it uses text data to extract meaning out it in a statistical way using multidimensional vectors. Hadoop based text mining helped data scientists to understand customers review on particular brand, this data was used back to suggest recommendations.

(reference: [Big Data Use Cases: How PayPal leverages Big Data Analytics](https://www.dezyre.com/article/big-data-use-cases-how-paypal-leverages-big-data-analytics/231))

Amazon Predictive Modeling over Big Data

Amazon has over 2.5 million active sellers (as of 2019) selling around 3 billion products across 11 marketplaces throughout the world. Amazon's market grabs about 50% of the entire e-commerce retail market's gross merchandise volume (GMV).

Amazon processes about 1,000,000,000 GB of data on more than 1,400,000 servers for the sake of predictive analysis to improve their sales. Amazon uses its own tool in the Amazon web server(AWS) called Elastic MapReduce (EMR) to process big data using the Hadoop framework and analyses it. They use a web service interface called ‘Amazon S3’ to store and retrieve data. Amazon focuses on 2 types of customers – ‘Amazon-buying customers’ and ‘Sellers on the Amazon Marketplace’.

Amazon uses purchase history, browsing history, social media searches, and others to track their customer requirements and interests. We can see that the home page of Amazon never looks the same. Based on the individual account management system, they provide various recommendations such as ‘Inspired by Your Wish List’, ‘Recommendations for You’, ‘Inspired by Your Browsing History’, ‘Related to Items You Have Viewed’, ‘Customers Who Bought This Item Also Bought’. They use graph theory for Dynamic price optimization which boosted their profit by 25%. As per the survey conducted by CapGemini, 89% of the Amazon customers won’t buy the same product in Amazon for the next time if they face any issues on their first purchase, so Amazon builds an effective supply chain by providing one-day delivery. With the help of a big data system, Amazon predicts the number of data warehouses they need and the capacity each warehouse should have.

Now Amazon knows customer needs before they buy it, i.e. Anticipatory Shipping. Amazon also owns a patent for Anticipatory Shipping technology. Amazon also knows what their customers will buy and when they are going to buy, this helps Amazon to maintain the required stock.

(reference: [How Amazon Uses Its Own Cloud to Process Vast, Multidimensional Datasets](https://dzone.com/articles/big-data-analytics-delivering-business-value-at-am))

What we see in our devices

Every one of us might have experienced that, when we open our shopping-related apps or social media, the page is never the same. Every one of us has Personalized adds, which keep changing based on our web search, cookies, and browsing history. Retail and Banking related companies like Amazon, PayPal use individual data based on account sign-in, to track the needs and interests of the customers. They build Predictive modelling to anticipate user habits using data as fuel.

Ethical issues in leveraging Big Data(Privacy vs Utility)

“Collecting data is not wrong. However, applying that data in unethical ways is wrong”, said Mark Moorman the senior director of statistical analysis system customer advocacy in The Economist Group(jackcentral.org). Optimally tracking users' data with permission is beneficial. Some of the applications help elders in health care, they monitor their heartbeat, remind medicine, and control household appliances even remotely. While a lot of advertising applications within the app collect separate data other than the app itself which is ethically concerned and a matter of trustworthiness.

Collecting users' data without permission is wrong. The right companies can use the data if it clearly granted in their terms and conditions agreement page. However, most people won’t have the patience to read multipage agreements. So is it fair to theoretically compel users to give consent on their terms? Though its legally Yes, but morally No, utilizing users' ignorance to take advantage of their data since the given consent is meaningless if they didn’t understand and well informed of the consequences of their decision.

Moreover tracking users' information is not alarming to all in all the times. Though many of the apps collect the required data on consent, sharing it with the third party just because users have agreed to their policies is not respect in data collection.

“Optimally, tracking would be for the users’ benefit, but that can become disturbing”, (jackcentral.org). Many of the users won’t be aware that, someone could be listening to them, their screens could be monitored and their locations are tracked 24X7. The General Data Protection Regulation (GDPR) in Europe has implemented certain policies and regulations, regarding companies using users' cookies as data, and also created awareness among people about enabling cookies on sites before being able to access them(Intersoft Consulting, n.d.). This has forced many companies to re-evaluate their data privacy policies in accordance with GDPR and do business with European citizens. Coming to social media like Facebook and Instagram, they are exceeding the limits of ethical concerns, to help advertisers to advertise their product, this also includes trapping their conversations(finance.yahoo.com: Why people think their phones are listening to them).

Many of the Predictive analytical tools have increased their customers' profit at the same time gave beneficiary luxuries to their uses. By the end of the day, it’s all about respect. The applications should make use of mandatory users' data to improve their service and technology, but they should not use their customer's behaviors as a selling commodity to the third person, disrespecting users’ privacy. Organizations must ensure that their Predictive models help users to enjoy their technology, without disturbing their privacy.

References and Citations:

Coronel, C & Morris, S. (2017). Database Systems: Design, Implementation and Management. 13th Edition.

[Big Data Use Cases: How PayPal leverages Big Data Analytics](https://www.dezyre.com/article/big-data-use-cases-how-paypal-leverages-big-data-analytics/231)

[PayPal](https://en.wikipedia.org/wiki/PayPal)

[Big Data Analytics: You Have The Right To Remain Private. Or Do You?](https://newsroom.cisco.com/feature-content?type=webcontent&articleId=1170577)

[Privacy vs utility: respect in data collection](http://www.jackcentral.org/opinion/privacy-vs-utility-respect-in-data-collection/article_47964f86-bc2f-11e6-b906-1b73f1e70a19.html)

[Big Data](https://en.wikipedia.org/wiki/Big_data)

[7 Interesting and Unexpected Uses of Business Intelligence Tools](https://interestingengineering.com/7-interesting-and-unexpected-uses-of-business-intelligence-tools)

[How Amazon Uses Its Own Cloud to Process Vast, Multidimensional Datasets](https://dzone.com/articles/big-data-analytics-delivering-business-value-at-am)

[Why people think their phones are listening to them](https://finance.yahoo.com/news/facebook-amazon-google-smartphone-spying-170721118.html)

Intersoft Consulting. (n.d.). General Data Protection Regulation. Retrieved from https://gdpr-info.eu