

**CSC105M**

Big Data Examples in Local and Foreign Companies

|  |  |
| --- | --- |
| **Section** | G01 |
| **Team Members** | Choo, Beatris Mariell T.  Fernandez, Ryan Austin  Lam, Janica Mae M.  Poblete, Clarisse Felicia M.  San Pedro, Marc Dominic  Tan, Johansson E.  Te, Robee Khyra Mae J. |
| **Date Submitted** | June 6, 2016 |

**Table of Contents**

I. Local Companies / 1

A. Senti / 1

B. Sulit.ph / 2

C. Talas / 3

II. Foreign Companies / 4

A. Google / 4

B. Netflix / 5

C. Starbucks / 6

D. Target / 7

III. References / 8

**I. Local Companies**

**A. Senti (Social Media)**

Senti is a local startup founded by CCS' very own Ralph Regalado. Their main offering is a social media analysis and visualization tool that gathers public sentiments on certain aspects or issues. The main thing that sets their tool apart from those of other companies is its ability to understand Tagalog, the preferred language of most Filipinos when using social media.

Senti is able to understand words and terms written in English, Filipino, Taglish, and other variations of the Filipino language like “j3j3mon”, “txt spk”, and even the “beki” language. Senti provides information about how their customers feel about their brand. Senti is also able to calculate and provide data visualizations in real-time, allowing their clients to react quickly and efficiently. Senti can also provide location-based information so its clients can see where their brand is popular, and where they need to focus more.

Senti gathers its data mainly from various social media streams like Facebook and Twitter. It crawls through data using keywords, and through Natural Language Processing and Machine Learning algorithms, it can identify the emotions associated with statements made by users. One of the most relevant events in recent history wherein Senti was able to display its tool was during the 2016 Philippine National Elections. Throughout the entire campaign period, and especially during the various presidential and vice presidential debates, they took the opportunities to provide visualizations on what the public sentiments are towards the different candidates and their statements, like how they feel about candidate the candidates’ stands on various topics.

Senti was also able to visualize the sheer popularity of the various candidates regardless of whether or not those people were talking about them positively or negatively. For example, during the Miting de Avance held last May 7, Senti analyzed nearly 500,000 tweets and found out that that more than half of these tweets were about candidate Duterte, meaning that by far he was the most popular candidate. To a certain extent, one can even think that Senti was able to predict the results of the elections because it was able to recognize that Duterte was the most popular among the five of them.

Senti also partnered up with Rappler, one of the most popular news websites that drives Filipinos to get involved in social issues, in order to provide credible information backed by social media data. The idea was that the public would use their social media accounts to voice out their opinions, Senti would gather and organize this data, and Rappler would present this back to the public. This partnership aimed to promote active civic involvement during the entire elections period by providing unbiased content via social media and to educate undecided voters by informing them of what the candidates' stands on various issues are and how the public feels about these stands.

Senti shows that social media is really a very powerful tool, and if used properly, can drive real change. Senti also shows that social media is in fact a giant repository of valuable data and that being able to organize, interpret, and analyze this data is a very important skill, and very relevant in the IT industry today.

**B. Sulit.com.ph (Shopping and Retail)**

Sulit.com.ph is a buy and sell website catering primarily to the Filipino market. It was founded in 2006 by the 26-year-old RJ David alongside with his wife, Arianne David, where they began an experiment in data gathering to find out what Filipino consumers preferred or wanted. RJ David created and developed the website on and running through updating features that the visitors liked and removed the features they didn’t. Soon enough, the site was getting popular as time goes by and after six years later, the couple’s brainchild became the most popular Filipino website.

Sulit.com.ph is an online marketplace for both buyers and sellers where a buyer can freely look and search for an item that they want to buy, hire, rent, lease or even exchange for. There are millions of advertisements posted on the site which gives the buyer a considerably number of options for them to choose from. Aside from products and items, buyers can also look for services, jobs and opportunities in the website as long as a seller provides. The seller, on the other hand, are the ones who posts advertisements about the product to sell, service to offer or brand to promote. The cost for the seller to put up an advertisement is free of charge which gives sellers the motivation to post whatever and whichever they like.

There are 10.4 million page views and 1.7 million visitors for the month of November back in 2007 and it rose to 200 million page views in 2012 which led Filipinos to easily bump into Sulit.com.ph in their search engines. Its immense popularity has gained them a huge user base and generated traffic from it as it is definitely a rising star in the Philippines.

As buyers and visitors go to Sulit.com.ph, data has been collected, gathered and stored to let them know which are the products or services most Filipinos would look for and want and which are not. Afterwards, they see to which category would be most appealing or strike interest to the consumers. In regards to this, the analysis of the data collected are segregated to age, gender and location in which items are more favored by the group of audience. This helps catch people’s attention as Sulit.com.ph aims to be service-oriented and community driven and they want to provide and advertise products and offers from different trusted sellers to tailor one’s need.

Through analyzing the data gathered, the Davids said that the average user, back in 2012, are in between 18 to 35 years old with 57% male and 43% female where Cars and real estates are mostly searched or looked up while most of the members are also from Metro Manila, followed by Cebu, Rizal and Davao. This let Sulit.com.ph post up advertisements with what is more favorable to the audience in a specific are or location which gives them a higher probability of having a new visitor click the advertisement and visit their website.

Even though Sulit.com.ph mostly focuses on Filipino consumers, it promises to develop new features to users through feedback and analysis by encoding and implementing new features to the website and then deciding to change or remove it depending on the resulting data analysis collected. This is to gain knowledge and understanding on what the users prefer to have as sulit.com.ph is a very data driven company.

**C. Talas (Telecommunications)**

PLDT CEO Ernesto Alberto once said that “Big Data drives our everyday lives”, and that paired with proper analytics tools, data can “become valuable insights that enable enterprises … to develop more effective products, services and programs for their target markets”. In an effort to bring big data to the Philippines, PLDT began the work needed to be done to launch Talas, its new big data analytics arm.

To prepare for Talas' launch, PLDT began storing data collected over the past year in a "data lake". At the time, they had been unable to analyze this data extensively because of the lack of any major big data services in the country, making Talas the first of its kind in the Philippines. The data stored in the data lake was used to test Talas’ platforms before the launch.

One of Talas’ major projects, and the undertaking Talas is most known for, is providing Smart subscribers with personalized offers and plans according to a subscriber's past usage. Talas would do this by taking a user's past activities and behaviors, such as the amount of mobile data consumed in a month, number of text messages sent, and length of calls, and sending the subscribers offers that cater directly to their usage needs. They also use location information based on cell sites a subscriber connects to in order to analyze a subscriber's travel activity. It is reported that Talas collects approximately 6TB worth of telecommunications data in a day, and Smart’s decisions are continuously being influenced by the analysis performed by Talas. For example, analysis on location has allowed Smart to better plan the locations of their cell sites around airports and frequently visited provinces.

Despite its affiliation with PLDT, Talas, headed by Silicon Valley’s Winston Damarillo, is meant to be a separate arm of the company and also offers big data services to other corporations, regardless of industry. Talas' platform was built in partnership with HortonWorks, and is based on industry leader Hadoop, which was not made purely to work for telecommunications-related operations. They also invested $5 million in Matrixx, a software provider with headquarters in California and in the United Kingdom. These partnerships allow Talas the flexibility it needs to perform data analytics operations on logistics, healthcare, and financial data, among other things. This is in line with PLDT’s vision to provide big data services to the nation. At present, Talas is also providing services to real estate company PrimaryHomes and Phinma. Talas also provides services such as sentiment analysis. When then presidential candidate Rodrigo Duterte’s controversial remarks about rape were reported to have sparked an outrage among citizens, Talas collected data from Twitter and said that the feedback about said comments on Twitter were actually positive.

Another of Talas’ major efforts is working to promote the use of big data in the Philippines, both by providing their services to various companies and by holding events to raise awareness and garner interest in the field. Since its launch in 2015, it has organized multiple events for the public where they have displayed the analytical powers of big data.

**II. Foreign Companies**

**A. Google (Search Engine et al.)**

Google is a multinational company specializing in Internet-related services and products such search, cloud computing and advertising technologies. The corporation is said to run about 900,000+ servers in data centers around the world, as of 2010. The March 2016 statistics on the average number of Google searches per second is at 2.3 million. Google processes about 3.5 billion requests per day, and each one queries a database of about 20 billion web pages. Google's bots copy down what they see and take it back for their index database. About 20 billion pages are indexed per day. Their aim is to make much of the world's information available to the public. As such, Google search is constantly revised and updated. Initially, they had the PageRank algorithm, which included additional elements to the keyword-based analysis that most search engines before used. Now, the aim was to move away from the then-usual keyword-based search and go towards semantic search.

In 2007, Google launched Universal Search, which pulls in data from many different sources including language databases, weather forecasts and historical data, financial data and many more. In 2012, this evolved into the Knowledge Graph, which displays information on the subject of the search from a wide range of resources directly into the search results. Knowledge Graph pages are given the ability to differentiate between words and phrases with different meanings to find out their relationship to each other. Knowledge Graph pages are given the ability to differentiate between words and phrases with different meanings to find out their relationship to each other. It could use things it knows about the user and their previous search history (if they are signed in). Words from the search phrase are analyzed by analyzes petabytes of web documents and historical search data to understand it.

Google first does a literal search on the search phrase wherein it looks for a match with some of or all of the search phrase. The root of the search phrase, when found, is examined and expanded to find better results. Then, using those Knowledge Graph pages, Google is able to understand the context of a phrase, basically, a semantic search. The results are then arranged to provide the most relevant outcome for the search query. Google also has factors to consider in determining the relevance of the websites it will give the user like site structure, page structure, etc.

Not only are they able to use big data analytics in order to improve on their search engine, they also are able to use the vast amounts of data about the people searching for their advertising technologies. They have an algorithm called Adsense algorithm which matches up companies with potential customers based on those data.

Google not only uses big data analytics but also gives big data services. They have many big data projects. In 2010, it launched BigQuery. BigQuery is a commercial service that allows companies to store and analyze big data sets on this cloud platforms. Another project that is known to many people is the self -driving car. This project uses and generates a lot of data from sensors, cameras and other related devices together with on-board and real-time data analysis from Google Maps, Streetview and other sources. Google is continuously using big data analytics for their search engine, their mapping service, their cloud services and many more.

**B. Netflix (Video Streaming)**

Netflix is a streaming service which allows people all over the world to watch various television shows and movies at their leisure. Due to the massive amount of users performing massive amounts of actions on the site that their system would consider as noteworthy “events” per second, along with the variety of actions considered as “events”, Netflix’s data mining model already satisfies the 3 V’s of Big Data. This sections aims to analyze the various data collected by Netflix and the predictive and descriptive analytics it performs on this data.

Netflix’s data collection process is extremely verbose. A subset of data collected by Netflix is what its users watch, search for and rate, as well as the time of day, date and device used, as well as the user’s browsing and scrolling patterns. (Poggi, 2013). In addition to this, data such as time of day that movies are viewed, time spent selecting movies, and how often playback was stopped, paused, rewound, and fast-forwarded (either by the user or due to network limitations) all became measurable. (Marr, 2015). This leads to a massive amount of unstructured data arriving at a large volume per second, which qualifies Netflix’s data as Big Data.

Based from this data, Netflix performs various predictive analytics. When multiple show pilots (a single episode of a show used to pitch a show), their analytics run the pilot against their massive data to estimate their audience. This allows them to select the most marketable shows. (Hegde, 2014). This also allows them to predict a user’s ratings of a particular show or movie based on their previous behavior (Marr, 2015). This helps in generating descriptive analytics.

Their analytics allows them to consider user groups with similar interests. For example, in their biggest experiment with Big Data Analytics, David Fincher’s “House of Cards” (Leonard, 2013), they discovered various groups in their user base that may be interested in watching the new show because they were interested in the 1990 BBC version. Also, based on user group interests, they produced multiple trailers. Fans of Kevin Spacey’s work were given trailers that centered on Kevin Spacey while most women were shown trailers that focused on the female characters. (Hegde, 2014) They also analyzed the posters that most frequently led to users viewing the show it was advertising, and used the same color scheme for their future posters. (Simon, 2014) When analyzing viewing patterns, it appeared that users preferred watching episodes in full bursts i.e. binge watching, so they released all their original content in that way ever since. Finally, they used viewer’s behavior to determine what shows they would like and when they would like to watch them. This is displayed on their recommendation in computed recommendation groups that are most likely to entice the user to watch the show at that time.

There is no doubt that big data has improved Netflix’s standing. In fact, 75% of their views come from recommendations alone. (Amatriain & Basilico, 2012) This, along with their predictions of which shows will do well and what advertisements would most likely garner viewers proves that in this respect, big data is extremely effective in improving Netflix’s business model.

**C. Starbucks (Food and Drink)**

Starbucks is one of the largest coffee shops in the world that has a total of 21,000 stores from 62 countries and offers 87,000 different flavored drinks varying from hot, cold and Frappuccino beverages. Currently, there is an approximate of 162,000 employees worldwide that had served 4 billion cups annually and recorded $13.3 billion net revenue last 2012. On a day-to-day basis, Starbucks sells around 1500 cups per branch; that would be 1 cup produced every 50 seconds.

Starbucks collects and analyzes data from each branch’s in-store coffee sales to decide what flavor they should add, retain or remove from their existing menu. This is applied when they learned that in a particular branch, there are some tea drinkers who don’t add sugar and some iced coffee drinkers don’t add milk in their beverages. Having to know these, Starbucks could possibly create new unsweetened drinks to satisfy these customers. Also, Starbucks introduces a limited-edition flavored beverages for every special occasion. For example, in fall 2003, the Pumpkin spice was famous in groceries. Thus, Starbucks’ very own Pumpkin Spice Latte was first announced in their stores. The orders of this beverage brought a huge profit to Starbucks, thus, they decided to do this again together with other Pumpkin spice drinks such as Pumpkin Spice Latte, instant Pumpkin Spice Latte packets, bottled Pumpkin Spice Frappuccino and bottled Iced Espresso with pumpkin spice flavoring the following years. This data can allow each branch to cater the needs of their customers, to improve customer’s contentment and to increase sales as well.

Starbucks also uses big data analysis for selecting their next branch. Data analytics use demographic, population density, income, businesses around the area including competitors, possible revenue, landmarks that attract customers, and even traffic situation in a location. These data are studied to provide the projected foot traffic and average income. Starbucks and Foursquare also team up to see the number of people checking-in in a location. These two companies provided the amount of people within an area and studied how people travel from one place to the other which will also be studied together with the data said earlier. Starbucks wanted to create as many branches as possible to provide for their customers without affecting the nearby branches negatively.

Lastly, in 2013, Starbucks recorded that there were 2.4 million users who activated their Starbucks cards. These cards can record data such as how often a customer buys, types of beverages ordered, bulk coffee purchases, and as well as food purchases. Through this new technology, Starbucks knows almost every Starbucks card user’s details. In order to analyze data more, Starbucks also gather other details like the location, time of the day, weather of the day a customer orders, inventory status, promotions within the day, and even local events happening on a certain location. These data could actually tell the daily routine of a customer. Thus, to improve customer’s satisfaction upon entering the shop, Starbucks employees may greet the customer by his/her name and the employees could also know and prepare his/her favorite beverage without even asking in the future.

As stated above, Starbucks had billions of data coming in annually which most of these was unstructured. This clearly shows that Starbucks’ data was difficult to analyze. Starbucks as one of the leading company showed how important analyzing data was. Thus, analyzing data certainly is one of the skills needed in the businesses today.

**D. Target (Shopping and Retail)**

Target is a discount retailer, so it is a company that focuses on convenience of shopping and competitiveness of prices, with its brand promise, “Expect More. Pay Less®”. The first Target store opened in 1962, in Roseville Minnesota. Today, there are 1,793 stores open in the United States. It is currently the second largest general merchandise retailer in America, while its website, Target.com, is the among the most visited retail websites.

Target collects data from its customers on a regular basis. It does this by assigning a unique Guest ID to each customer, and linking relevant data, from when customers fill out forms, use credit cards or coupons, open an email sent by Target, use Target’s website, etc. Aside from the data taken from direct interactions with Target, Target also links demographic information to this ID, such as age, address, salary, civil status, number of children, visited websites, ethnicity, educational attainment, product preferences, etc., taking this data from both direct interactions with Target, like through a form on their website, or indirectly, through Facebook posts, sometimes buying this data if necessary. Target gets every bit of data it can in order to form detailed customer profiles.

This data is used by members of Target’s Guest Marketing Analytics department to understand both shoppers’ shopping habits and shoppers’ personal habits, and to figure out how to use this information to better market products to these customers.

An example of how they use this is in predicting when a customer is pregnant. If Target could predict if a customer was pregnant, or even what trimester she is in her pregnancy, they could direct the buying of products such as baby furniture, diapers, clothes, and the like to their stores by sending coupons for the appropriate baby products at just the right time, encouraging to spend at Target rather than elsewhere. They’re able to predict this by analyzing data to find useful patterns. An example of a pattern found was how they noticed that women at the beginning of their second trimester of pregnancy tended to buy more unscented lotion. They also noticed that women bought more supplements in their first 20 weeks of pregnancy. Customers’ shopping habits, along with pattern analysis, allow Target to assign each customer a pregnancy prediction score as well as an estimate of her due date, so that they can benefit from this stage wherein customers would potentially do a lot of shopping.

It has been suggested the targeting of pregnant women and mothers specifically has had a great impact on the revenue increase of Target from $44 billion in 2002 to $67 billion in 2010, since 2002 was the year they hired Andrew Pole, a statistician working at Target who focused on predicting pregnancies. Many retailers do customer analytics, but Target is often regarded as one of the best at it. This has surely contributed to Target’s success over the years, by allowing Target to personalize each customer’s shopping experience based on his/her personal habits.

**III. References**

Amatriain, X. & Basilico, J. (2012). Netflix Recommendations: Beyond the 5 Stars (Part 1). Retrieved June

2, 2016, from The Netflix Tech Blog: http://techblog.netflix.com/2012/04/netflix

-recommendations-beyond-5-stars.html

Aquino, T. (2016). Talas: Despite Duterte rape remarks, sentiments on Twitter were more positive than

negative. Retrieved June 2, 2016 from: http://interaksyon.com/article/126615/talas-despite

-duterte-rape-remarks-sentiments-on-twitter-were-more-positive-than-negative

Boulton, C. (2016). Starbucks' CTO brews personalized experiences. Retrieved June 03, 2016, from

http://www.cio.com/article/3050920/analytics/starbucks-cto-brews-personalized

-experiences.html

Braue, D. (2015). For data analytics leaders in Asia Pacific, success comes from willingness to fail.

Retrieved June 2, 2016, from: http://www.computerweekly.com/news/4500259838/For-data

-analytics-leaders-in-Asia-Pacific-success-comes-from-willingness-to-fail

Careon, M. (2016). Rappler and Senti Forms Partnership for 2016 Philippine Elections. Retrieved June 2,

2016, from Senti Blog: http://blog.senti.com.ph/post/?rappler-and-senti-forms-partnership-for

philippine-elections

Ferraz, E. (2013). Nkk2wa! Meet the Social Media Tool that’s Fluent in Filipino and Text-Speak. Retrieved

June 2, 2016, from TechInAsia: https://www.techinasia.com/nkk2wa-meet-social-media-tool

-fluent-filipino-textspeak

Google (the company). (n.d.). SearchCIO. Retrieved June 5, 2016 from http://searchcio.techtarget.com

/definition/Google-The-Company

Hill, K. (2012, 2 16). *How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did*. Retrieved

June 5, 2016, from Forbes: http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target

-figured-out-a-teen-girl-was-pregnant-before-her-father-did/#77d7823f34c6

Marr, B. (2014). Wow! Big data at Google. Retrieved June 5, 2016 from http://www.smartdatacollective

.com/bernardmarr/202181/big-data-google

van Rijmenam, M. (n.d.). How Google applies big data to know you - infographic. Retrieved June 5, 2016

from https://datafloq.com/read/google-applies-big-data-infographic/385

Hegde, H. (2014). Netflix and Its Revolutionary Use of Big Data. Retrieved June 2, 2016, from

Dataconomy: http://dataconomy.com/netflix-and-its-revolutionary-use-of-big-data/

Lara, Tanya T. (2014). How P2,400 in Capital Became A Billion-Peso Business. Retrieved June 05, 2016,

from http://www.philstar.com/business-life/2014/09/29/1373787/how-p2400-in-capital

-became-a-billion-peso-business-

Leonard, A. (2013). How Netflix is Turning Its Viewers into Puppets. Retrieved June 2, 2016, from Salon:

http://www.salon.com/2013/02/01/how\_netflix\_is\_turning\_viewers\_into\_puppets/

Lipka, M. (2014, 1 23). *What Target knows about you*. Retrieved 6 5, 2016, from Reuters: http://www

.reuters.com/article/us-target-breach-datamining-idUSBREA0M1JM20140123

Marr, B. (2015). Big Data: How Netflix Uses It to Drive Business Success. Retrieved June 2, 2016, from

Smart Data Collective: http://www.smartdatacollective.com/bernardmarr/312146/big-data-how

-netflix-uses-it-drive-business-success

Meeting the Filipino Couple Behind Sulit.com.ph. (2013, April 22). *ABS-CBN News*. Retrieved from

http://news.abs-cbn.com/business/04/22/13/meet-filipino-couple-behind-sulitcomph

PLDT Group launches Big Data services in the Philippines. (2015). Retrieved June 2, 2016, from:

http://www.pldt.com/news-center/article/2015/05/13/pldt-group-launches-big-data-services-in

-the-philippines#.V1S0xPl97Dd

Poggi, J. (2013). Data-Mining Boosts Netflix’s Subscriber Base, Showbiz Clout. Retrieved June 2, 2016,

from: http://adage.com/article/special-report-marketer-alist-2013/data-mining-boosts-netflix-s

-subscriber-base-showbiz-clout/243759/

Rijmenam, M. (n.d.). Datafloq - Connecting Data and People. Retrieved June 03, 2016, from

https://datafloq.com/read/world-largest-coffee-shop-starbucks-grinds-lot-dat/440

Ruiz, J. (2012). Starbucks Customer Profile; Relationship Marketing Customer Analysis. Retrieved June 3,

2016, from http://www.socialmediatoday.com/content/starbucks-customer-profile-relationship

-marketing-customer-analysis

Sanchez, V. (2016). Miting de Avance Through the Social Media Lens. Retrieved June 2, 2016, from Senti

Blog: http://blog.senti.com.ph/post/?miting-de-avance-through-the-social-media-lens

Simon, P. (2014). Big Data Lessons from Netflix. Retrieved June 2, 2016, from Wired: http://www.wired

.com/insights/2014/03/big-data-lessons-netflix/

Target. (n.d.). Retrieved June 5, 2016, from Target Corporate: Social Responsibility, Careers, Press,

Investors: http://corporate.target.com

Thau, B. (2014). How Big Data Helps Chains Like Starbucks Pick Store Locations -- An (Unsung) Key To

Retail Success. Retrieved June 03, 2016, from http://www.forbes.com/sites/barbarathau/2014

/04/24/how-big-data-helps-retailers-like-starbucks-pick-store-locations-an-unsung-key-to-retail

-success/#23451fe821c3

Visconti, Katherine. (2012). A Look Inside Sulit.com.ph and Online Commerce Trends. Retrieved June 5,

2016, from http://www.rappler.com/business/16475-a-look-inside-sulit-and-online-commerce

Whitten, S. (2016). Starbucks Knows How You Like Your Coffee. Retrieved June 03, 2016, from

http://www.cnbc.com/2016/04/06/big-data-starbucks-knows-how-you-like-your-coffee.html