APPLICATION DETAILS

Host university: Cal Poly

Project name: Tweetonomics

First name: Sunjay

Last name: Dhama

Field/major: Computer Science

Email: sdsunjay73@yahoo.com

Club/team name: Tweetonomics Co

Faculty sponsor/advisor: Eriq Augustine

Project description:
Tweetonomics is software as a service. The service aggregates Tweets from Twitter and analyzes them to provide real-time feedback to companies. Currently, surveys for customer feedback require a company to send out massive amounts of snail mail and e-mail asking customers to fill out a form and send it back. For example, Kaiser Permanente sent out 50,000 surveys by mail and received 900-1400 responses, according to Greg Matthews, the Director of Social Media Marketing at WCG World, whom our team spoke to over the phone. We know this process can be dramatically changed and updated for the 21st century.
During March of 2011 around a billion tweets were sent each week. What this means is there is an abundance of valuable freely accessible information on Twitter.
How our service will work is we would contract with a company, spend time understanding their needs, use that information to personalize the application, and then provide the company with real-time feedback from customers and potential customers.
Our application is still in development, but when finished it will be able to classify tweets, show trends in sentiment, provide alerts if a large number of customers are complaining or having problems with the company or service, and various other functionality depending on a company's needs. Our service is an expedited way of obtaining customer feedback and will be available at a substantially lower cost.
When we have completed the application, we will work to make it scalable. We

believe it can get to a point where any company, regardless of industry will be able to use the software with only minimal modification to the source code.

What do we get to see at the end of the project:
By June 30th, we will have a working prototype that successfully gathers and accurately categorizes tweets around a certain business. Furthermore, we will have an intuitive graphical representation of the data, such as pie charts and bar graphs, for easy interpretation by a marketing team through a web page. We also hope to have some type of time series graph. This graph will show consumer sentiment toward a company over time. As a hypothetical example, Starbucks might want to know whether recent changes to their advertising has been promoting a more positive image for the company. We could show whether customers have been viewing Starbucks more positively or negatively and then show a breakdown with reasons. Perhaps customers have been viewing the company more positively, but it is because they recently introduced a new size of drink. This kind of analysis would allow companies to see the immediate effects of their marketing campaigns or product launches.

Key concept that must be validated, proved, prototyped:
We must prove that we can get a sufficient accuracy rate of our analysis algorithm to where a business would be willing to pay for such information. Because analysis of language is such a difficult problem, it is a challenge to teach a machine to categorize phrases and sentences accurately. This will be the largest hurdle to overcome. We need to validate what accuracy rate would be sufficient for a company to pay
for our service. We already have a working prototype of the software that can gather tweets based around a particular keyword and categorize them. However, we have yet to check the accuracy of the categorizations. Once we have a sufficient accuracy level, then we have proof of concept. Companies are already conducting marketing surveys. That's what our system does. Tweetonomics just does it faster and cheaper.

Essential innovation:
The innovation of our service is that we are applying the field of Natural Language Processing to analyze tweets in order to survey a specific company's consumer market*. Analysis of tweets has been around in various forms for a couple of years. There are a number of companies, most recently Bloomberg, that attempt to use Twitter to predict stock prices. Countries, such as Japan, are using it as a warning system for Tsunamis. One may think that Tweetonomics faces competition from Google and Salesforce, but this would be incorrect. Google Analytics provides

companies with a standard analysis API that is not customized for the company's particular information needs. Using Google's classification API as a stand alone tool would produce little to no valuable information. It would need to be implemented into a larger system by a development team. Our team spoke with Brendan Wood, the sole creator of Salesforce's tweet aggregator. He confirmed that aggregation of Tweets is all they currently do. What all of this means is that Tweetonomics is innovative in its approach to tweet analysis.
*This is not a novel innovation, but one which was suggested to the team by our mentor, Eriq Augustine, who worked on a similar project as his master's thesis, which can be found athttp://dl.acm.org/citation.cfm?id=2187983

Commercial potential:
The commercial potential for Tweetonomics is immense. On average, marketing companies spend about $10,000 for 200 responses via snail mail surveys or about $5,000 for 300 responses via email survey. This money is used mostly to pay a third party to analyze the responses in order to obtain meaningful feedback. Tweetonomics can not only do this automatically, but can also increase the number of people being surveyed. The number of new accounts on Twitter per day as of March 2011 is on average 460,000. The feedback our software will generate will also be much more uptodate. Since tweets can be pulled from Twitter as soon as they are posted and then analyzed immediately, the customer sentiment being generated is only a minute or two old. Some may argue that a 140 character tweet could never replace a page long survey, but that would only demonstrate a lack of understanding of Twitter's potential. The 140 characters that are allowed per tweet should not be viewed as a limitation, but an opportunity. Typing out 140 characters is not overly burdensome to people, which means a much greater percentage of people are tweeting about an interaction with a company as opposed to filling out a survey or calling customer service. The 140 character limit does present a challenge to obtaining a high accuracy in classification, but it is not something that cannot be overcome. The only company we were able to find that does analysis of tweets for marketing survey purposes is WCG World. According to Greg Matthews, WCG World's client base is made up of 75% physicians, which means there are plenty of untapped markets. This software has the potential to transform the way a company in almost any industry interacts with their customers. Companies are always looking to become more competitive, which is what this service would allow them to do.

Market addressed:
The hot market we are addressing is that of customer feedback. Companies are always trying to figure out what their customers want.

Our team will focus on analyzing tweets regarding the company Starbucks. Starbucks is the largest coffeehouse company in the world. With the number of people that regularly visit Starbucks and make a purchase, there are more than enough tweets being sent about Starbucks for us to derive trends in customer sentiment that Starbucks will find useful.

Benefit to mankind/sustainability:
The benefits to mankind are that we allow companies to gather more accurate information about their customers allowing them to better serve them and address their needs. Businesses will get feedback directly from their customers, so if there is a problem a company can quickly remedy the situation.

Team description:
Our team consists of 3 Cal Poly students, Dallas Brown, Sunjay Dhama, and Sam Kligman. Dallas Brown is a 3rd year undergraduate student in Economics. Sunjay Dhama is a 4th year undergraduate in Computer Science. Sam Kligman is a graduate student in both Mechanical Engineering and Business Administration.The team also has a mentor, Eriq Augustine. Eriq recently graduated from Cal Poly with a Masters degree in Computer Science. While Eriq is not participating in IQ, his contributions to the team cannot go unmentioned. Eriq allowed the team access (with no stipulations) to over 50,000 lines of twitter sentiment analysis code that is currently being using by Netflix to detect outages in their systems. Eriq's code, which he maintains ownership of, already has 95% accuracy in classifying Netflix tweets (which is considered phenomenal) has allowed us to leap forward to the late prototyping phase, bringing us closer to a launch.

Go-to-market plan:
We are currently exploring many opportunities. Since we are still working on the software, we have yet decide on the best gotomarket plan. We plan to initially offer our service for free to one or two small or medium sized businesses, such as Woodstock's Pizza. Once we get good feedback from these companies, we would then have the credibility to approach larger companies, such as Starbucks, to sign up for our service.

Name(s) of other participants:
Dallas Brown, Economics, iamdallas2001@gmail.com
Sunjay Dhama, Computer Science, sdsunjay73@yahoo.com
Sam Kligman, MBA and MS in Mechanical
Engineering, samkligman@gmail.com
Eriq Augustine, MS in Computer Science, eriq.public@gmail.com