**USER GENERATED CONTENT ANALYTICS**

**MIS 381N Unique: 04045 Fall 2018**

**T Th 2:00 – 3:30 p.m. (RRH 5.420)**

**Instructor:** Professor Anitesh Barua

**Office:** CBA 5.236

**Email:** aniteshb@gmail.com

**Tel:** 512-471-7895

**Office hours:** M W 9-10:30 a.m.

**Teaching Assistant:** Vasundhara Sharma

**Email:** [vasundhara383@gmail.com](mailto:vasundhara383@gmail.com)

**TA Office:** CBA 1.308E

**TA Office hours:** T Th 4:00 – 5:30 p.m.

**Course Overview**

We have witnessed an unprecedented rise in quantity of user generated content in recent times. The rapid proliferation of social media has led this explosion of unstructured data (e.g., text, pictures, audio and video) and a unique opportunity for enterprises to engage in real-time interactions with customers, and to enhance brand, customer loyalty, competitiveness, growth and profitability. Facebook statistics from 2016, which include over a billion users, 150 billion connections, 500 million daily photo uploads and 4.1 billion daily likes, suggest that companies can ill-afford to ignore the potential of social media platforms and the unstructured data they create. According to one source (in 2016), LinkedIn adds two new users every second, Instagram gets over 900 likes per second, and as a country, Twitter would be the 10th largest in the world. Apart from social media, unstructured data come from numerous sources in the physical world, including maintenance reports, notes scribbled (or dictated) by doctors and nurses, open ended interviews, etc. Such numbers and the sheer volume of user generated content make it imperative to develop strategies to create and extract value from this rapidly growing phenomenon. This course is designed to showcase the virtually unlimited opportunities that exist today to leverage the power of user generated content analytics. It focuses on a gamut of questions ranging from strategic to operational matters pertaining to a firm’s social media initiatives, metrics to capture relevant outcomes, and predictive analysis to link social media chatter to business performance.

**Learning Objectives**

This course is especially valuable to students contemplating careers in information management, business analytics, marketing, prediction modeling, business consulting and general management. Students taking this course will develop expertise in the following areas:

1. How to access user generated content through APIs and crawlers/scrapers (in Python).
2. How to analyze, derive insights from, and dash-board social media chatter using Python scripts
3. How to create prediction models from social media mentions.
4. How to analyze sentiments using Python and other tools
5. How to measure resonance between campaign message and target audience
6. How to develop real world social media applications
7. How to develop strategic aspects of investing in user generated content analytics, including ROI.

Students are not required to have a deep knowledge of statistics (though a basic understanding is absolutely necessary), data mining or technical ability in programming languages and software applications. The content of this course is presented in an intuitive format with emphasis on the connection between social media and business strategies. A key feature of this course is the use of hands-on software tools for collecting and analyzing social media interactions.

**Technical skills**

In this course we not only “talk the talk” but also “walk the talk” by actually trying hands on new concepts and ideas to leverage user generated content analytics in business and related areas. This means we actually run (Python) scripts to collect and analyze unstructured data, and also use other tools for text classification, visualization and sentiment analysis. However, there is NO coding required in the course (I provide all the necessary Python scripts along with an extensively documented primer for reference and guidance). While no coding background is needed, what is essential is a positive attitude toward using actual technologies. For instance, there may be some difficulties in installing Python on your computer (e.g., some errors in installing the various packages and libraries are often encountered); the most important thing is to be patient, and to Google (or Bing, if you prefer) the error message to find out solutions to the problem. Remember you are not the first in the world to have gotten that error message, and certainly not the last either.

**Course Material**

Given the nascent and dynamic nature of this field, where advances are happening so rapidly, there are no satisfactory textbooks that cover the topics we will study. As a result, I have created a list of readings and cases to highlight the latest advances in unstructured data analytics. There is a small Harvard course pack containing just three cases. In addition to this course pack, there are a set of readings and cases from a variety of sources, which will provide the bulk of the content we will cover. Some of these articles will be posted on Canvas one week before they will be discussed, while the ones with links should be downloaded by the students.

# Harvard course pack: To buy, please use this link:

https://hbsp.harvard.edu/import/563833

1. **Material outside of the Harvard course pack (you will have to download some of the articles with links on your own; others will be posted on Canvas as mentioned)**

**Articles**

# “Reducing Readmissions to Improve Care” (posted on Canvas)

# “Gaining Business Value from Unstructured Data” <https://www.crimsonhexagon.com/blog/finding-value-from-unstructured-data/>

(posted on Canvas)

* “The Customer Journey to Online Purchase” <http://metrictheory.com/using-googles-customer-journey-to-online-purchase/> (posted on Canvas)
* The Fundamentals of Social Media Analytics (posted on Canvas)
* “[Mine your own business: Market-structure surveillance through text mining](http://mktsci.journal.informs.org/content/31/3/521.short)” (posted on Canvas)
* “Applications of Text Classification Using Text Mining” (posted on Canvas)
* “Investigating Predictive Power of Stock Micro Blog Sentiment in Forecasting Future Stock Price Directional Movement” (to be posted on Canvas)
* “Product Comparison Networks” (posted on Canvas)
* “Introduction to Social Network Analysis” <http://www.orgnet.com/sna.html>
* “Analyzing Social Media Networks with NodeXL” <http://hcil2.cs.umd.edu/trs/2009-11/2009-11.pdf>
* “It is not the size of a customer’s network that matters; it’s what they do with it.”http://www.wired.co.uk/article/customer-network-lifetime-value
* “High Note” (posted on Canvas)
* “An Introductory Guide to Understand how ANNs Conceptualize New Ideas (using Embedding)”

<https://www.analyticsvidhya.com/blog/2018/04/introductory-guide-understand-how-anns-conceptualize-new-ideas/>

* “Masterfoods USA Case Study: Driving Innovation in R&D with Network Analysis” (to be posted on Canvas)
* “SHIFT Changes the Way Cemex Works”

**Grading**

Your course grade will be based on the following:

|  |  |  |
| --- | --- | --- |
| Item | Date(s) due | Weight |
| Individual assignments 1 & 2 | 9/20, 10/09 | 20% |
| Group assignments 3 & 4 | 10/25, 11/15 | 20% |
| Final group project | PowerPoint slides due on 12/04 by the beginning of class; the sequence of presentations will be decided by a random draw | 15% |
| Take home final (individual work) | Posted on Canvas on 12/07 by 11:59 p.m., due 12/11 by 11:59 p.m. on Canvas | 35% |
| Class participation |  | 10% |

**Class participation**

In this class much of the learning is dependent on the accessing the combined knowledge and experience of the group. It is everyone’s job to keep the discussion productive and moving forward. In evaluating your class participation grade, I take the following into consideration:

* useful arguments expressed coherently and succinctly
* good analysis supported by case facts or your own experience
* relevance to previous contributions, i.e. ability to listen and build on what others say
* constructive disagreement
* regard, respect and acknowledgment of others’ contributions
* readiness to contribute to class discussions

**Assignments 1-4 (40% of course grade)**

Assignments 1 and 2 involve individual work, and are designed to get everybody started with unstructured data analytics. Assignment 3 and 4 require group work. Students will be responsible for creating their own groups. The ideal group size is 5. The group membership should remain unchanged throughout the semester, unless there is a truly compelling reason to do so.

**Final group project (15% of final grade)**

Throughout the course, groups (5 students per group) will work on a social media analytics project dealing with real world data. Topics can vary widely depending on student experience and interest, and can include how social media affects brand, sales and other business outcomes, what type of messages are effective in social media, and analysis of customer network value. Groups will make their final presentations on 12/04 and 12/06.

**Course Agenda**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **#** | **Date** | **Topic** | **Readings** | **Learning outcomes** |
| 1 | 08/30 | Introduction: The importance of user generated content | “Reducing Readmissions to Improve Care” (posted on Canvas)“Gaining Business Value from Unstructured Data”https://www.crimsonhexagon.com/blog/finding-value-from-unstructured-data/ | Learn about the course structure, focus, tools, assignments, and take home exam  Understand the opportunities and challenges of leveraging various forms of user generated content in business applications |
| 2 | 09/04 | Introduction to Social Media Analytics | “The Customer Journey to Online Purchase” <http://metrictheory.com/using-googles-customer-journey-to-online-purchase/>  (posted on canvas)  The Fundamentals of Social Media Analytics (posted on Canvas) | Learn about what is unique about social media and the type of analytics we can perform on social data |
| 3 | 09/06 | Knowledge discovery and insights from social media I | “[Mine your own business: Market-structure surveillance through text mining](http://mktsci.journal.informs.org/content/31/3/521.short)” (posted on Canvas) | Learn how social mentions can lead to the discovery of new facts and actionable insights |
| 4 | 09/11 | Knowledge discovery and insights from social media II | “[Mine your own business: Market-structure surveillance through text mining](http://mktsci.journal.informs.org/content/31/3/521.short)” (posted on Canvas) | Learn how social mentions can lead to the discovery of new facts and actionable insights |
| 5 | 09/13 | Discoveries and insights III | “[Mine your own business: Market-structure surveillance through text mining](http://mktsci.journal.informs.org/content/31/3/521.short)” (posted on Canvas) | Learn how social mentions can lead to the discovery of new facts and actionable insights |
| 6 | 09/18 | How to access unstructured data from online platforms | “Social Media Analytics Primer 2018” (posted on Canvas) | Learn how to collect data from social media using Python scripts and crawlers/scrapers and APIs. |
| 7 | 09/20 | Sentiment Analysis Part I  **Assignment #1 due on Canvas by 11:59 p.m.** | Social Media Analytics primer (posted on Canvas) | Learn about different approaches to sentiment analysis (unsupervised methods), hands-on sentiment analysis with SentiStrength and custom Python scripts |
| 8 | 09/25 | Sentiment analysis Part II | Youtube videos  <http://www.youtube.com/watch?v=ytUHvMNnzZk>  <http://www.youtube.com/watch?v=0JsHvXmU0dA>  “Applications of Text Classification Using Text Mining” (posted on Canvas) | Learn how to perform supervised sentiment analysis with training data  Learn how to predict outcomes through basic classification of text with LightSIDE and Python scripts. |
| 9 | 09/27 | Predictions using social media Part I | “Investigating Predictive Power of Stock Micro Blog Sentiment in Forecasting Future Stock Price Directional Movement” (to be posted on Canvas) | Learn how social mentions and sentiments can be used to predict business outcomes |
| 10 | 10/02 | Predictions using social media Part II | “Product Comparison Networks” (posted on Canvas) | Learn how to create product comparison networks to capture relative preferences of customers, and how to predict business outcomes from such comparison networks. |
| 11 | 10/04 | Measuring resonance from social mentions Part I | Annotated slides (to be posted on Canvas) | Learn how to assess if a campaign is achieving resonance with its target audience, and how to apply the underlying metrics to a broad range of applications. |
| 12 | 10/09 | Designing Crowdsourced Recommendation Systems  **Assignment #2 due on Canvas by 11:59 p.m.** | Crowdsourced Recommendation Systems (posted on Canvas) | Learn how to leverage the wisdom of the crowd in building personalized recommendation systems |
| 13 | 10/11 | Clustering with text & topic modeling Part I | Annotated slides (to be posted on Canvas) | Learn how to group unstructured data and extract themes and topics of interest |
| 14 | 10/16 | Topic modeling Part II | Annotated slides (to be posted on Canvas) | Learn how to group unstructured data and extract themes and topics of interest |
| 15 | 10/18 | Measuring social attention and influence | Annotated slides (to be posted on Canvas) | Learn how to find individuals who are influential in social media |
| 16 | 10/23 | Social network analysis Part I | “Introduction to Social Network Analysis” http://www.orgnet.com/sna.html | Learn key metrics in a social network, and how to utilize such metrics to find important people or “nodes” in a large network. |
| 17 | 10/25 | Social network analysis Part II  **Assignment 3 (Group work) due on Canvas by 11:59 p.m.** | “Analyzing Social Media Networks with NodeXL” <http://hcil2.cs.umd.edu/trs/2009-11/2009-11.pdf> (to be posted on Canvas) | Learn how to represent different types of relationships (e.g., activity or interest based) as a social network |
| 18 | 10/30 | Image Analytics | Fundamentals of Image Analytics guide by Crimson Hexagon (posted on Canvas) | Learn how images can provide deep insights and predict business outcomes. |
| 19 | 11/1 | Case studies of social influence | “Ford Fiesta Movement” (in HBSP course packet) | Learn how Ford leveraged social media in promoting the Fiesta, and how companies can leverage social media for brand and product advocacy.  Predict social influence from Twitter data |
| 20 | 11/06 | New Customer Acquisition Through Social Influence | “High Note” (to be posted on Canvas) | Learn how to leverage social influence in acquiring new customers |
| 21 | 11/08 | Discuss final project ideas | | |
| 22 | 11/13 | ROI of social media  Network value of a customer | “Meteor Solutions: Measuring the Value of Social Media Marketing” (in HBSP course pack)  “It is not the size of a customer’s network that matters; it’s what they do with it.”  http://www.wired.co.uk/article/customer-network-lifetime-value  “How valuable is word of mouth” (in HBSP course packet) | Learn how digital word-of-mouth and influence can be tracked and how such influence culminates in actual purchases  Learn about best practices in maximizing returns from social media  Understand the difference between customer lifetime value (CLV) and customer network lifetime value (CNLV); how to measure CNLV, and why CNLV may be more important than CLV today. |
| 23 | 11/15 | Internal uses of social media  **Assignment 4 (Group work)** | “Masterfoods USA Case Study: Driving Innovation in R&D with Network Analysis” (to be posted on Canvas) “SHIFT Changes the Way Cemex Works” | Understand how social network analysis can be used to assess and improve communication patterns and outcomes in organizations.  Understand how social network analysis can be used to assess and improve communication patterns and outcomes in organizations  Learn about the latest advances in UGCA including embedding and Word2Vec |
| 24 | 11/20 | Work on final project | | |
| 25 | 11/27 | The cutting edge in UGCA | “An Introductory Guide to Understand how ANNs Conceptualize New Ideas (using Embedding)” https://www.analyticsvidhya.com/blog/2018/04/introductory-guide-understand-how-anns-conceptualize-new-ideas/  “Masterfoods USA Case Study: Driving Innovation in R&D with Network Analysis” (to be posted on Canvas)  “SHIFT Changes the Way Cemex Works” | Learn about the latest advances in UGCA including embedding and Word2Vec |
| 26 | 11/29 | Discussion of progress on final project | | |
| 27 | 12/04 | **Final Project presentations** | | |
| 28 | 12/06 | **Final Project presentations** | | |
|  | **12/07** | **Take home exam, handed out 12/07 by 11:59 p.m., due 12/11 on Canvas by 11:59 p.m.** | | |