

Data Mining Case Study: Information Analytics

SI310 Case 9

Fall 2019

Written by Kevyn Collins-Thompson

September, 2018

You have been hired as a consultant for a small but fast-growing investment bank that has an online Web portal for customers. Using the Web portal site, users can create investment accounts with the bank, transfer funds, do research on different stocks, get real-time news on companies, watch videos teaching them investment and retirement planning strategies, and so on. The bank has hired you to create a programmatic analysis tool that will help them understand how customers are using their Web site, which features are in high or low demand, and whether customers are successful or unsuccessful at completing important investing-related tasks using the site. This is the first time the company has attempted to gain insight from their web traffic.

To do this, your program will analyze the Web server logs of the portal site, which capture all interactions that a user has with the site, stored in a central database. This database contains tables that include: demographic information on each customer (e.g. reported age, net worth, location), the customer's account transactions (e.g. buying and selling securities; funds transfers including deposits and withdrawals), and the customer's online interactions with the website.

Specifically, the online interactions table stores timestamped user interaction events, which include when a user started their visit to the site, the URL of every click they made on specific site links, how long they viewed a particular page, any text they entered into online forms, the queries they typed into the site's search box, which videos they watched and for how long, and so on.

Your liaison for this consulting work is the Chief Technical Officer at the company. She tells you that the output from your analysis program will be shared broadly, starting with the Executive Management Board. This group includes the Vice-President of Customer Service, who wants to show that the large investment was worth the expense (several million US dollars): in an email, he expresses interest in evidence that users are engaged with the new features they added, and that users are successfully completing important tasks, like signing up for a new account. Also on the EMB is the Director of Software Development. She also emails you stating she is curious to see if you find any evidence of actual bugs that customers may be encountering with the site. The Web development team would ideally like to run online experiments to test which features were likely to be of most interest to their customers. The EMB also includes the company's VP of Legal Affairs, the VP of International Investments, and VP of Human Resources.

1. Identify who the key stakeholders are, and why. For each, what are examples of questions about the web site analysis that each might want you to answer?
2. How might the new information infrastructure you're providing to analyze web traffic integrate with the company's existing information infrastructure and likely information flows?
3. Different people in an organization can have very different perspectives on the same underlying data – and very different biases. Given the stakeholders you identified in (1), characterize the nature of biases that each one might have that could influence either their requirements for your tool, or the conclusions they might reach from its output.
4. You have six months and a budget for a small team of two other software developers to build your analysis tool. This is one of the constraints that limits the scope and nature of stakeholder questions and requirements you can realistically address. Identify the tradeoffs that you face in trying to make the best possible system according to different stakeholder requirements.
5. What privacy and ethics issues should you consider when (a) developing the tools to perform your analysis and (b) writing up a report for stakeholders summarizing your results?
6. Write a memo of no more than 500 words – a “pithy” summary of the key points. The people mentioned above will read and consider this memo. If they find the document useful they will probably give it to other key people, important people who can influence your welfare over time. This isn't many words to address a complicated topic. The text of the memo above is over 400 words, and does not include the answers to the questions raised! The accompanying reading is nearly four times longer than the number of words you have available. You must know more than you talk about to make sensible selections about what to include. You also have to think about the stakeholders – what do *they* want to know about? If it seems you know, they will come back and ask. This is a *good thing*.