DATA MINING AND ANALYSIS FOR THE CHILDREN'S INN AT THE NATIONAL INSTITUTE OF HEALTH

THE EFFECTS OF DATA MINING ON NON-PROFIT DONATION STRATEGIES

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virgnia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Systems Engineering

By
Sofia Ponos

December 11, 2020

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

Signed	Date:
Sofia Ponos Sofia Ponos	12/11/2020
Approved:	Date:
William T. Scherer, Department of Engineering Systems and Environment	
Approved:	Date:
S. Travis Elliott, Department of Engineering and Society	

Introduction

The availability of big data and data mining has affected every sector, including the non-profit space. In recent years, analyzing big data has become a major competitive tool in all business sectors. "'Data Mining', often referred to as 'Knowledge Discovery in Databases' (KDD), is a young sub-discipline of computer science aiming at the automatic interpretation of large datasets." (Kriegal 87). The technical focus of this project is to use data-driven analysis to recommend changes to improve targeted outreach for the Children's Inn's at the National Institute of Health's current fundraising strategy. The Science, Technology & Society (STS) component will be to explore the effects that data mining and big data has had on fundraising in the non-profit space using the Social Construction of Technology framework. The goal is to recognize how fundraising has changed over time and learn if any of these changes can be attributed to overall changes in big data.

Technical Topic

The Children's Inn at the National Institute of Health (NIH) is a residential "Place Like Home" for families with children participating in clinical research studies at the NIH. A unique quality at the Inn is that families stay there for absolutely no cost. The Inn provides room and board, meals and activities free of charge through their own fundraising efforts. The Inn has many fundraising events throughout the year, such as a golf tournament and a gala. They also solicit donations from regular mail and email appeals which are sent out 10-12 times a year. Currently, the Inn spends 15% of their budget on fundraising, while an overall goal for non-profits is to spend less than 10% of their budget on fundraising (Charity Navigator).

One of the key goals of the Inn is to expand and diversify its base of supporters to ensure they can continue to provide important services to the families and patients of the NIH. The Inn has a database with over 60,000 records and approximately 11,000 active donors. My Capstone team's job is to mine the database and additional resources to build models to help with the overall goal of increasing donations to the Inn. We are approaching the project from two sides, one is to learn how to increase the dollar amount of donations from current donors to make them more profitable and the other is to determine how to efficiently solicit new donors. The goal of the project is to use statistical methods and data analysis to gain insight into the types of donors the Inn currently has and to create a marketing and recruitment plan to gain new donors optimize donations from current donors.

In order to learn about the current donors, it is important to identify what unique traits donors possess that cause them to be more likely to donate. Some metrics that non-profits use to measure donors are recency (R), frequency (F) and monetary (M) value. Recency refers to their latest donation, frequency measures how often they donate, and monetary measures how much they donated. There are different ways to interpret these measurements so it is crucial to define metrics. My team has explored scoring each donor with a score from 1-5 (1 being lowest and 5 being highest) in each category of R, F, and M. These scores were appended and each donor was given a three point numerical score with each number representing recency, frequency, and monetary respectively. For example, using the cutoffs my team created based on historical donor data, a donor who has given in the past year, who on average 3 times a year, and who gives $5 on average would have an RFM score of 541. These scores can then be used to segment donors into different states of giving. The goal is to eventually be able to map a donor's journey transitioning through the different states of giving and identify what causes donors to transition. For example,

perhaps when a donor's children graduate college they often transition into a higher frequency of giving, due to more disposable income. This will allow The Children's Inn to solicit donations more effectively by approaching donors in a more personalized way. The different RFM traits that donor's possess can also separate them into different "user groups" some examples would be "those who give annually from a mutual fund" or "college students who donate online".

In addition to the data that the Inn provided, 5W Strategists, a strategic marketing firm, appended mined data to using the names and addresses of donors. This data includes over 500 demographic attributes per person including things like race, income level, and political affiliation. This data is used to get a clearer picture on the profile of a donor so that the Inn can find "look-alikes" who are likely to become new donors. 5W Strategists also provided a national sample that was used to determine how current Inn donors differ from the overall population of the U.S.

Some of the direct competitors for the Inn's donation dollars are other children's medical causes like Ronald McDonald House and St. Jude's Children's Hospital. In the past, the Inn has purchased lists of donors who have donated to similar causes to use when sending out appeals. This is often very costly and the Inn is concerned that the return on their investment is negative. Using the information from the data, it will be possible to predict which of the people on those lists are most likely to give. The Inn can lower costs and raise their response rates by only soliciting those individuals.

By working with this data and making observations the goal is to create actionable recommendations on how to appeal specifically to different donors, which will increase the Inn's overall donations and their return on investment.

STS Topic

The purpose of the STS research paper will be to explore how has the availability of big data and data mining algorithms changed the way that non-profit organizations go about their fundraising efforts. The STS Framework being used for this analysis is the Social Construction of Technology (SCOT). I will explore how a variety of social factors from the nonprofit space caused technological development in the way donors are valued and solicited. The Social Construction of Technology Framework posits that human action drives technological development. I will use the SCOT framework to explore 4 main tenants, relevant social groups, interpretive flexibility, closure and stabilization, and problems and conflicts. My analysis will explore how trends in big data have been affected by human actions overall and specifically in the non-profit space.

When it comes to relevant social groups, there are multiple stakeholders to consider when viewing data mining as a whole and in the non-profit space. An important stakeholder is the user whose data is being collected or specifically, the donors. Some other stakeholders are the data collectors, data analysts, and those who consume the analysis, people who sell data mining software and the recipients of the charity.

Data mining has a lot of interpretive flexibility, because different social groups have very different outlooks on the technology. The main concerns of those whose data is collected is privacy. Many users want to know what data is being stored and how it will be used. Using data mining to increase charitable giving may bring up privacy concerns from donors, but can also bring up privacy concerns from the organization's standpoint. "Not all participants in a joint activity of data mining would like to share all of their collected data, possibly in order to protect the privacy of their customers. Thus, there is a growing need for distributed, privacy preserving

data mining algorithms for complex data." (Kriegal 90). The need for privacy is one way that human action drives changes in data mining technology. Some concerns that those participating in data mining activities may have are: a) the accuracy of the data, and b) the strength of the analysis. A misinterpretation of data can cause problems in analysis and lead to false conclusions. For example, "Historical data can be analyzed for trends but has comparatively little value when used for real-time decisions … As one study found, "90% of the data in the world today has been created in the last two years…70% of unstructured data is stale after only 90 days," (Tucker 4). A common problem that non-profits run into is having outdated information, on which they are wasting time and resources this presents a possible negative externality of data mining.

When it comes to closure and stabilization it is important to keep in mind that data mining is only relevant to charities if we assume trackable factors, such as demographics affect charitable giving. There has been much research done around this topic. A paper about factors that affect charitable giving states, "Some research suggests that people are rational beings and are likely not to make large contributions towards a public good (Landesman, 1995)." However, he also states, "Past studies have suggested that several factors can impact charitable contributions. Hughes (2002) found that personal experience with an organization was a motivating factor for charitable giving (Hughes, 2002). Likewise, in their study of 49 British givers, Radley and Kennedy (1995) found that personal experience with the charity, either directly or indirectly, was the major reason people elected to donate money." (Robin 73). Based off of these statements, one can see why data mining could be extremely useful to soliciting donors if certain traits are available and identified.

Finally, data mining has the potential to create some problems in the non-profit space. A paper on the ethical issues of data mining brings up, "Web mining does pose a threat to some important ethical values like privacy and individuality. Web mining makes it difficult for an individual to autonomously control the unveiling and dissemination of data about his/her private life," (Van Wel 1). In most situations, there is need for clear communication between the organization and the donor about the data being collected to avoid a conflict between different stakeholders. There could be extremely negative repercussions, like a decrease in donations, if donors lose trust in an organization because they believe the organization is invading their privacy.

Using the lens of SCOT, I will research the actions of the stakeholders through case studies and analyze how their actions affect how data mining efforts have changed and are developing.

Research Methods

For the technical part of the project, I will be using Excel, Minitab, R and Tableau to draw conclusions from the data provided by the Children's Inn on donor demographics and transaction activity and appended data from 5W Strategies. In addition, we will be looking at outside case studies of methods that other non-profits have used to make donations. For the STS paper, I will be looking into changes overtime of how non-profits have fundraised, and then look into how this has shaped the use of data mining as a fundraising technique. I want to look into specific case studies of non-profits that have used data mining and how this has affected their donations. In addition, I want to observe what some possible negative externalities of data mining in non-profits exist.

Conclusion

The conclusion of the technical part of this project will be to recommend changes to the Children's Inn's current fundraising strategy. The conclusion of the STS component will be to present the effects that data mining and big data has had on fundraising in the non-profit space using the lens of the Social Construction of Technology framework. I explore different case studies and past research to guide my analysis.

References

Your Guide To Intelligent Giving: Home. (n.d.). Retrieved October 19, 2020, from https://www.charitynavigator.org/

Kriegel, H., Borgwardt, K.M., Kröger, P. et al. Future trends in data mining. Data Min Knowl Disc 15, 87–97 (2007). https://doi.org/10.1007/s10618-007-0067-9

Robin L. Snipes and Sharon L. Oswald (2010). Charitable giving to not-for-profit organizations: factors affecting donations to non-profit organizations. Innovative Marketing , 6(1)

Tucker, Darren S. and Wellford, Hill, Big Mistakes Regarding Big Data (December 1, 2014). Antitrust Source, American Bar Association, December 2014

Van Wel, L., Royakkers, L. Ethical issues in web data mining. Ethics and Information Technology 6, 129–140 (2004). https://doi.org/10.1023/B:ETIN.0000047476.05912.3d