Data Management Systems

Course Number: 610

Course Term: Summer, 2020

Assignment number: 1

Project title: Decision Framing

Contact Information:

Melissa Hunfalvay

Email: melissa@righteye.com

Phone: 240-357-0026

Instructors Names:

Dr. Moretto

Dr. Goldberg

Introduction

The purpose of this assignment is to identify a problem within my organization that requires a decision. This decision must need to be made repeatedly, must be non-trivial, measurable and a possible candidate for future automation.

The problem selected is an on-going, product enhancement for RightEye eye tracking products. Eye tracking data is not widely understood and has little guidance on how it should be interpreted by customers. Therefore, one of the first points of reference after completing the test is for customers to view their patients results (reports). In doing so, the customers need to determine where the patient falls compared to others in the same age group. In other words, they look for a normative comparison to make sense of the testing results.

RightEye is a cloud-based platform that has approximately 1000 systems deployed throughout the world to date. Hence, data is collected, 24 hours a day, seven days a week, from anywhere in the world increasing our data volume exponentially with every new sale.

RightEye needs to consistently monitor and update the normative guidelines as more data becomes available. This is important for our customers and the business. The payment model is Software-As-A-Service (SAAS), and customers have the option to renew each year. The eye tracking data, especially at such volume, is a unique business, product and data advantage to the company and is therefore a key selling point for renewals. It provides customers an up-to-date and relevant enhancement to their product which increases the likelihood of their continued use.

Therefore, the main decision to be made is "***When*** *should the normative data be updated?"*

The objective is to provide guidelines and thresholds that frame the decision. Furthermore, when the decision to move ahead is agreed upon, the objective is to have a high likelihood of success in order for the business to have a Return On Investment (ROI). Success is more likely to occur if the decision to move forward has been carefully thought through. Thoughtful Exploratory Data Analysis (EDA); realistic timelines for the project; available resources to work on the project; and evidence-based statistical guidelines, all contribute to the objective of completing the project successfully.

Decision Identification and Descriptions

The main decision is: When should the normative data be updated? This decision is complex and spans across all areas of the company.

Supporting decisions from the business context include decisions based on customer feedback and renewals. Customers buy the RightEye product at any time throughout the year. At the time of purchase, the yearly renewal date is set for 12 months in the future. However, there seems to be a seasonality associated with purchase time and the Customer Success Manager will need to run an EDA to determine if that assumption is, in fact, correct. If correct, we would know when *most* customers are due to renew and plan to update the normative data ahead of most renewals.

From an organizational context, RightEye is a small business and therefore there is a constant struggle between a long list of product priorities and a short list of resources. Therefore, as an organization we need to determine what are the product priorities? Where does this decision fall within those priorities? What resources are needed to update normative data? How

long will it take? The organizational context is also related to the business context as length of time relates to money for resources to complete the project.

From an application context, decisions need to be triggered based on certain data and statistical thresholds and standards. Furthermore, new eye tracking algorithms are consistently being added to the database and require normative ranges to be added to the results. Data thresholds help the statistician, data analyst, researcher and Chief Science Officer (the Research and Development team; R&D), determine if there is enough additional data, at any point in time, that could add further insights. For instance, more data may allow for a smaller range in age groups derived from a cluster analysis. More data, may allow additional insights in terms of adding other demographic or epidemiological insights such as gender groupings or ethnicity, both of which are captured fields but currently do not have enough sensitivity to create a separate grouping of data.

Information and Knowledge Sources

Information sources are those that require content or some unknown input to further the decision-making process (Taylor, 2015). Information sources are needed to know when to update the normative data and include the delta between prior versions, database totals and current versions. This helps determine if we have enough new data to upgrade to the next version of normative output.

An EDA is an informational source that provides the R&D team input that shows if a) new algorithms are "stable" enough to include in the new version of normative data; b) if there is a visual change in data that could yield new and interesting insights. The EDA includes

distribution charts, box plots, measures of central tendency such as means, and measures of variability, such as standard deviations.

Statistical power analysis is an informational source that provides guidance as to whether the sample size is large enough. If the sample size is too small, the results may be inconclusive or even inaccurate (known as type I or Type II errors; Kasneci, Kübler, Broelemann, Kasneci, 2017).

Another information source is the customer. The customers view is represented within RightEye through the Customer Success (CS) manager. This person reports on what customers a) say is important to them, such as product features like normative data, b) when the customer is due to renew, c) what product enhancements they would like to see in the future. This information source influences priorities of the product. On a related note, it is very important for the R&D and Technology Team (TT) of engineers, to clearly communicate time, resources and project deadlines to the CS team, for two primary reasons: 1) these set customer expectations, 2) priorities should not be changed during the product development or analytical lifecycle. Scope changes or increases in the product development cycle require a reset of timelines and are not necessarily 1-to-1 comparisons with prior timelines due to context switching between tasks.

Another information source that influences priorities of the product are revenue opportunities for the business. For instance, a recent change in priorities occurred as a customer agreed to purchase three RightEye products if changes were made to the English only content and Italian was added. The revenue opportunity was great enough to support this change in priority.

Knowledge sources are the know-how or expertise needed to inform the decisions. As previously mentioned, the CS manager is the knowledge source for the customer needs. The R&D team are the knowledge sources for statistical and data guidelines, standards, and thresholds. The Chief Executive Officer (CEO) and Chief Revenue Officer (CRO) are knowledge sources for sales, new business opportunities and business priorities.

<div align="center">Analytic Questions/ Key Performance Indicators</div>

Due to the nature of the main decision, many of the questions are of an analytical nature. Key Performance Indicators (KPI's) are quantifiable measures used to evaluate success of a project or organization.

The primary KPI for this project is that value is added to the product as measured by high levels of annual renewals.

More specifically, when normative data updates show something new, either new age group(s) or new insights into demographic or epidemiological constructs that were previously unknown, then unique value is added to the product. For example, should differences in male and female eye movements result from the normative analysis, then this is a KPI that adds value in the form of new and unique information that is valuable, adding insights for the doctor to provide the patient.

Another related KPI is the increased specificity of the data output. As previously mentioned, there is little known in scientific literature about the normative expectations of eye movement behavior. Therefore, as more data is acquired, and further iterations and normative versions of analysis are conducted the more specific and sensitive the output becomes. For example, the results of the first normative analysis included a sample size of 3000 that resulted in

five age-based clusters (Murray, et al., 2019). The second analysis had a sample size of 61,000

and resulted in eleven age-based clusters. Furthermore, the eleven age clusters showed less

variability within the groups. This enables the R&D team to display a measure of central

tendency (mean values) and a measure of variability (standard deviation) after the second

analysis. This was difficult for customers to understand in the first version as the standard

deviations would overlap between age groups due to the smaller sample size and higher

variability. Therefore, in the first analysis, the standard deviations were removed from the

reports. The second analysis, with the standard deviations, gives doctors a smaller age range,

with less variability within each group. Such results allow for more specific normative

comparisons and therefore more value to the patient.

Related to the increased data specificity is the increased accuracy of the data. Data

accuracy is a KPI. It is well understood that, if our data is not trusted we do not have a product,

or a business. Accuracy of normative data is measured by a low type I error (alpha α), also

known as a false positive and a low type II error (Beta β), also known as a false negative. To

mitigate such errors, we ensure the sample size is sufficient for adequate power (Galgani, Sun,

Lanzi & Leigh, 2009). Data accuracy gives customers confidence in the reports.

In conclusion, the main decision as to "***When*** *should the normative data be updated?"*

has wide ranging implications across the business. In addition, the value-add of updating

normative data and showing new insights are critical KPIs directly impacting retention of

customers via annual SAAS renewals. Renewals are key to RightEyes business longevity,

growth, and success.

References

Galgani, F., Sun, Y., Lanzi, P. L. & Leigh, J. (2009). Automatic analysis of eye tracking data for medical diagnosis. *IEEE Symposium on Computational Intelligence and Data Mining*, Nashville, TN, 195-202. http://doi.org/10.1109/CIDM.2009.4938649.

Kasneci, E., Kübler, T., Broelemann, K., Kasneci, G. (2017). Aggregating physiological and eye tracking signals to predict perception in the absence of ground truth, *Computers in Human Behavior, 68*, 450-455. https://doi.org/10.1016/j.chb.2016.11.067.

Murray, N., Kubitz, K., Roberts, C., Hunfalvay, M., Bolte, T., & Tyagi, A. (2019). An examination of the oculomotor behavior metrics within a suite of digitized eye tracking tests. *Vision Development and Rehabilitation, 5*(4), 269-284. https://doi.org/10.31707/VDR2019.5.4.

Taylor, J. (2015). *Improving Analytics results using decision modeling.* Retrieved from https://www.slideshare.net/jamet123/improve-analytic-results-with-decision-modeling
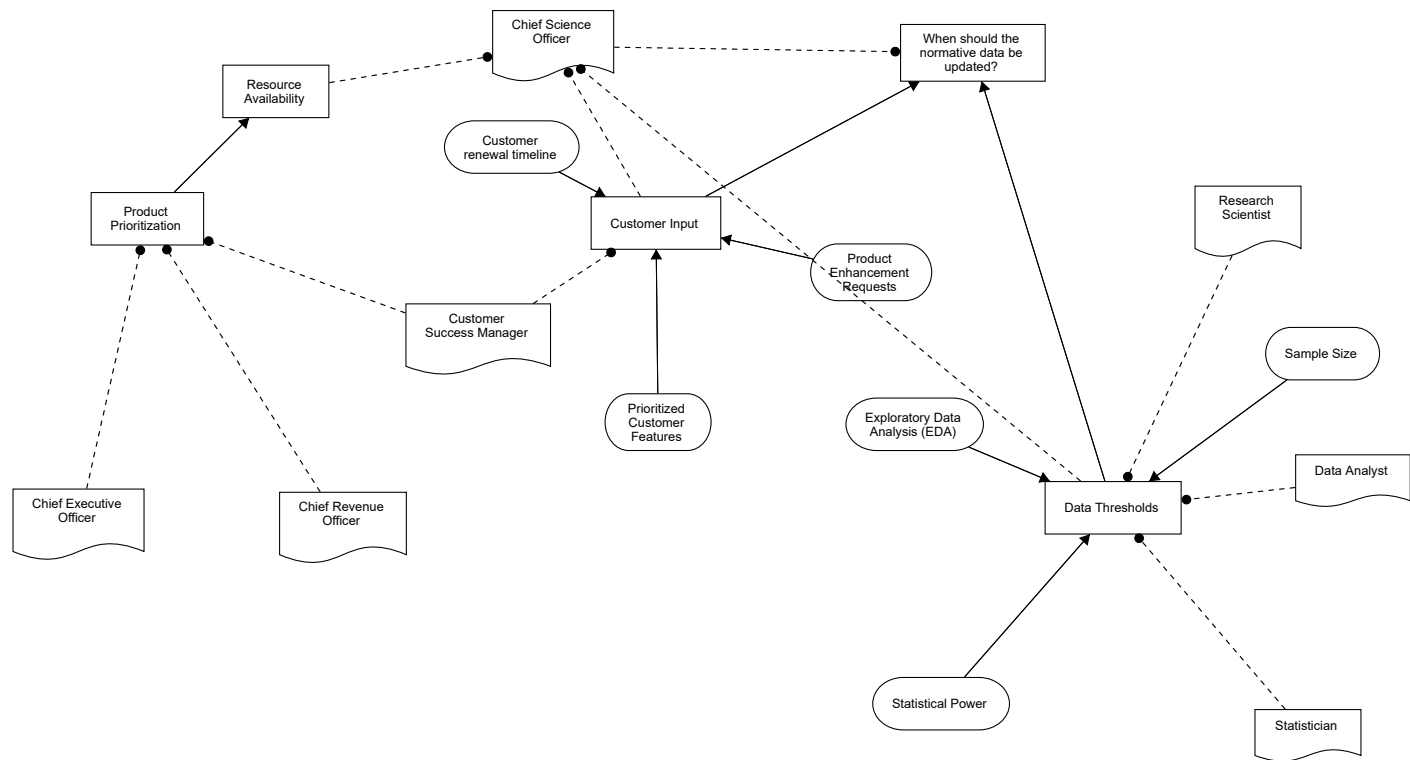
Appendix A

Analytics Requirement Document (ARD)

MH - Normative data decision making

This diagram is designed to provide guidance for determining when to update the normative data for RightEye products

Decisions

Customer Input

Input from the customer is managed by the Customer Success Manager. Several factors are considered in the input and include: 1) when the majority of customers are going to renew; 2) what other enhancements are being requested, this is a large bucket of internal and external features, requests, enhancements and bugs 3) more specifically, the prioritized customer features specifically related to new sales or renewal opportunities. All these inputs affect the decision of where to prioritize customer feedback and ultimately to prioritize when to update the normative data.

Question: Is the normative data updating within the top three tasks prioritized by customers for renewal?

Yes / No: Yes

Supporting information: Salesforce is used to collect all customer input. Support Ninja is an external contracted agency used as a level one, basic, customer support process for listening and learning to customer needs.

Diagrams

- MH - Normative data decision making

Required By Decisions

- When should the normative data be updated?

Required By Knowledge Sources

- Chief Science Officer

Requires Input Data

- Customer renewal timeline
- Prioritized Customer Features
- Product Enhancement Requests

Requires Knowledge Sources

- Customer Success Manager

Required By Knowledge Sources

- Chief Science Officer

Data Thresholds

Data thresholds considers many components related to the success of the project from a data analytics perspective. Some of these components, if not satisfactory stop the project from going forward as subsequent efforts maybe time and resources wasted the baseline structure of the data is not met. Components include: both knowledge sources and informational sources. Generally speaking the process includes the Data Analyst conducting an Exploratory Data Analysis (EDA), a power analysis and sample size data. This is then summarized in an initial report and sent to the Chief Science Officer (CSO), Statistician, and Research Scientist. This makes up the Research and Development Team. Once the initial report is reviewed the team meets to decide whether the Data Thresholds are met and if this component of the decision making process allows the project to move forward.

Status Level: Completed

Question: Are the data thresholds met?

Yes / No: Yes

Supporting information: Sample size is adequate. Power analysis is high. EDA shows good stability as measured via standard deviations and variance measures. Visual inspection shows potential for additional insights via mean group comparisons.

Diagrams
- MH - Normative data decision making

Required By Decisions
- When should the normative data be updated?

Required By Knowledge Sources
- Chief Science Officer

Requires Input Data
- Exploratory Data Analysis (EDA)
- Sample Size
- Statistical Power

Requires Knowledge Sources
- Data Analyst
- Research Scientist
- Statistician

Required By Knowledge Sources
- Chief Science Officer

Product Prioritization

As a small start-up, both time and money are key items for the business to balance. There are always more product enhancements than time or resources can achieve. Therefore prioritization is key. Prioritization is informed, in part, by customer feedback. Other considerations are business opportunities and revenue goals. The executive team (Chief Revenue Officer, Chief Executive Officer and Chief Science Officer) and the Customer Success Manager determine product priorities.

Status Level: In process

Question: Is updating the normative data now within the top 3 product prioritization tasks?

Yes / No: No

Supporting information: If the answer is no, do not proceed with updating the normative data.

Diagrams
- MH - Normative data decision making

Required By Decisions
- Resource Availability

Requires Knowledge Sources
- Chief Executive Officer
- Chief Revenue Officer
- Customer Success Manager

Resource Availability

The Research and Development (R&D) team is a small team. It is lead by the Chief Science Officer and contains a full-time data analyst, a contracted statistician and researcher. There are several other external non-paid researchers who also contribute to peer reviewed articles. Resource availability for updating the normative data falls mostly on the data analyst.

Status Level: In process

Question: Are resources available?

Yes / No: No

Supporting information: If adequate resources are not available a decision needs to be made by the Chief Science Officer to either a) hold on the project, b) obtain more resources.

Diagrams
- MH - Normative data decision making

Required By Knowledge Sources
- Chief Science Officer

Requires Decisions
- Product Prioritization

Required By Knowledge Sources
- Chief Science Officer

When should the normative data be updated?

RightEye data is continually expanding, and new insights can be made from more data. This unique data is informative to customers and their patients. Insights from normative data allow doctors and patients to compare themselves to others and therefore better understand their results. Customer renewal rates of software are tied to product enhancements and to understanding the reports. Updating normative data is not a trivial task, given our current lack of automation in some areas it takes several months to push the changes to production for customers to see. This costs the company money. Prioritization is informed, in part, by customer feedback. Updates to normative data are high on customer feedback requests.

Status Level: In process

Question: Should normative data be updated?

Yes / No: Yes

Diagrams
- MH - Normative data decision making

Requires Decisions
- Customer Input
- Data Thresholds

Requires Knowledge Sources
- Chief Science Officer

Input Data

Customer renewal timeline

As the RightEye product is a Software-As-A-Service (SAAS) product their are annual renewals. The renewals are tied to when the customer begins their first test. using the product. The industries RightEye sells into engage in seasonal buying periods. The Customer Success Manager (CSM) has renewal dates and can provide insights into which customer groups are collectively renewing at certain times throughout the year. Furthermore, there is more than one product, and different customers use different products (for instance sports vision practices use the Sports Vision Test and not the Brain Health test). Therefore, the CSM can provide both the number of customers renewing per month AND the types of tests those customers most frequently use.

Diagrams
- MH - Normative data decision making

Required By Decisions
- Customer Input

Exploratory Data Analysis (EDA)

EDA comprises of a summary report that provides high level information one component of success for the project. EDA comprises of descriptive statistics including, measures of central tendency (mean, median, mode) and measures of variability (standard deviation, variance, ranges). It also includes box plots and distribution graphs for visual inspection. It may include Analysis of Variance (ANOVA's) or percentiles or percent differences from past versions. These latter statistics are included IF the prior statistics are not conclusive to the decision making process. Such statistics are conducted on each test metric (algorithm). To date, this includes approximately 1000 metrics. Metrics include both eye tracking data and demographic data. They may also include test related parameters to assist in high level cleaning efforts.

Diagrams

- MH - Normative data decision making

Required By Decisions

- Data Thresholds

Prioritized Customer Features

These are product features that customers have shown to be important to them.

Diagrams

- MH - Normative data decision making

Required By Decisions

- Customer Input

Product Enhancement Requests

These are future product improvements, known bugs, internal and external features or enhancements needed on the product. Examples include front end changes, back end analytics, bugs, features, UI changes, UX changes, security updates. Anything at all to do with the product.

Diagrams

- MH - Normative data decision making

Required By Decisions

- Customer Input

Sample Size

Sample size, which is the total number of people within certain group categories, is viewed as part of the statistical power and also independently. It is pulled out of the statistical power information because, often times the data is growing at such a rapid rate that it is important to view this component by itself. In some cases, no further EDA is required to make the decision to update the normative data, if the sample size has grown by 100% from the prior analysis. An EDA will still be conducted, however, it will serve a different purpose and move to later in the analytical process.

Diagrams

- MH - Normative data decision making

Required By Decisions

- Data Thresholds

Statistical Power

Statistical Power Analysis is used to determine if there is a likelihood of differences being found. In other words, are we going to find new insights that will add value to the customer. What is the sensitivity of a statistical procedure to differences being sought? Provides the basis for correctly rejecting the null hypothesis. That is it helps reduce Type 2 Error. Power is increased by an increase in sample size. It is reported in a range from 0-1. .8 refers to a good power level. Power analysis gives confidence in probability of statistical significance.

A prospective power analysis can be used to determine the minimum sample size (N) given prior expectations regarding the effect size, the alpha significance criterion, and the desired level of statistical power.

For example, if you hope to detect an effect of size $r = .40$ using a two-tailed test, you can look up a table to learn that you will need a sample size of at least $N = 46$ given conventional alpha and power levels.

To detect a smaller effect of $r = .20$ under the same circumstances, you will need a sample of at least $N = 193$.

Diagrams

- MH - Normative data decision making

Required By Decisions

- Data Thresholds

Knowledge Sources

Chief Executive Officer

The Chief Executive Officer (CEO) has oversight of the business and priorities of the business including new opportunities, emerging markets and trends. The CEO is also responsible to the Board of Directors (which includes the Chief Science Officer) for company vision. He is important in helping prioritize product related tasks due to his high level oversight of the company strategy, business opportunities and vision.

Diagrams

- MH - Normative data decision making

Required By Decisions

- Product Prioritization

Chief Revenue Officer

The Chief Revenue Officer (CRO) is responsible for all revenue opportunities throughout the business. He has oversight into customer feedback and renewals related to that feedback. He also has insight into sales opportunities and upgrades of software to new products.

Diagrams

- MH - Normative data decision making

Required By Decisions

- Product Prioritization

Chief Science Officer

After reviewing all the information gathered, the Chief Science Officer (CSO) is responsible for making the ultimate decision to go forward and update the normative data. The CSO has insights into all company activities, tasks, priorities and markets. She makes this decision in coordination with the Chief Revenue Officer (CRO) and Chief Executive Officer (CEO). She is also a Subject Matter Expert (SME) in the scientific applications of eye tracking technology. She reviews the customer input, data thresholds and resource availability to make the "go, or-no-go" decision for the project.

Diagrams

- MH - Normative data decision making

Requires Decisions

- Customer Input
- Data Thresholds
- Resource Availability

Required By Decisions

- When should the normative data be updated?

Customer Success Manager

The Customer Success Manager is responsible for all aspect of customer relations. Primarily that includes two main areas: 1) customer support; 2) customer success. Customer support relates to problems a person may have with the software. This could be technical problems such as logging in. Customer Success relates to ensuring customers are happy and likely to renew. Updating the normative data would fall under this per-view. For example, recently a customer asked for the Italian language to be added. S/he would not renew without the new language addition. This information came from the Customer Success Manager and was directly reported to the CRO.

Diagrams

- MH - Normative data decision making

Required By Decisions

- Customer Input
- Product Prioritization

Data Analyst

The Data Analyst is the hands on expert of the RightEye data. He pulls the data, conducts the analysis and presents the results. He has input and is required to provide his understanding and recommendations for next steps.

Diagrams

- MH - Normative data decision making

Required By Decisions

- Data Thresholds

Research Scientist

The Research Scientist, like the Chief Science Officer has some domain specific knowledge. He is required to view the Data Thresholds both from an analysts, statistical and research level perspective. His goal is to ensure that when the data analysis process is complete, that he can publish the data in a peer reviewed, scientific journal.

Diagrams

- MH - Normative data decision making

Required By Decisions

- Data Thresholds

Statistician

The statistician views the data reports and provides input on the likelihood of success of the project based solely on the data available. He provides input on the risks and benefits of the data process based on the EDA and Data Thresholds. If the project moves forward the Statistician is involved in choosing the correct methodology and interpreting the results from a statistical stand point only.

Diagrams

- MH - Normative data decision making

Required By Decisions

- Data Thresholds

Appendix B

Insights into Regulatory Requirements, Company Culture and Communication

Regulatory Requirements: FDA 510K

The RightEye product is an FDA cleared 510K device. This means that is considered safe and effective and is substantially equivalent to a legally marketed device. The substantial equivalent is a nystagmogram.

Regulatory considerations are important to be aware of when analyzing data as FDA clearance is tied to a certain version of data. This does not mean that data cannot be changed. Instead, it means that as data (or other changes to the product) occur over time then an updated submission to FDA is required. If the update is not submitted within a certain time, the clearance can be revoked. FDA clearance is important for our product as it provides comfort to customers that there is regulatory oversight and standards being met.

Company Culture and Communication

Within the next 8-12 weeks, RightEye will be launching version two of the normative data. There has been considerable thought on my end of how to communicate and position this data within the company and the customer base.

The challenges for the company include:

1. A culture of anxiety around big product launches.

2. A fear of those on the front lines with customers not being adequately able to answer questions.

3. A lack of communication of product changes, including the impact of those changes to customers

The challenges for customers, especially doctors who explain results to patients, include:

1. How do I explain to a patient that their results are now compared to a different threshold, age group, or other comparative standard?
2. How do I explain to a patient who has been tested under the previous normative version of data, that their results may now be different? Does this mean that previous results were inaccurate? Are the new results inaccurate?

If not adequately prepared for, and thoroughly considered, these issues lead to distrust, frustration and a lack of willingness for accepting future innovations. Furthermore, resources who have worked hard, and who understand the advances become frustrated and disheartened.

It is my job, as both Co-Founder and Chief Science Officer to bridge those gaps and ensure the launch is successful. Tools that communicate the process, such as DecisionsFirst, and the results, such as Cognos Analytics, can go a long way to graphically explain innovations and insights.

I am hopeful that through better cross-departmental communication, visual tools, and proactive thoughtfulness we will have a successful production launch of version two of the normative data. In preparation for the launch, I have developed a cross-departmental working group to break down barriers and build bridges of trust and understanding.