PRAXIS INDEPENDENT STUDY-SPRING 2021

HEMLATA SHARMA'22

FIELDSITE: KNOWCLICK

Faculty Advisor: Jennifer Spohrer

Field site Supervisor: Derek Burton

PRIMARY OBJECTIVES:

1.) GAIN A FUNDAMENTAL UNDERSTANDING OF PREDICTIVE ANALYTICS IN CONTEXT OF IMPROVING ONLINE BUSINESSES.

2.) LEARN TO BE ABLE TO NAVIGATE THROUGH RAW DATA TAKEN FROM A WEBSITE AND SYNTHESIZE AN EXPLANATORY HYPOTHESIS.

COURSE DESCRIPTION.

This course will help in gaining a fundamental understanding of the science of predictive analytics in the context of improving online businesses. In this course the hands-on experience with KnowClick will cover up the necessary key concepts of synthesizing an explanatory hypothesis while navigating through raw data.

Fieldsite Description:KnowClick LLC

KnowClick runs surveys at websites in order to understand what visitors think of the website experience:

- For many of these surveys, it is important to trigger them when the user is finished visiting the website and about to leave it.
- KnowClick has the technology to track intricate user behaviors on a website which can be analyzed to determine when a user is about to leave. This includes factors such as mouse movement, scrolling, time on the page, etc. The actual time when user exits is also recorded which helps in validating the prediction and further optimization.

BRIEF DESCRIPTION OF PROJECT

Goal:

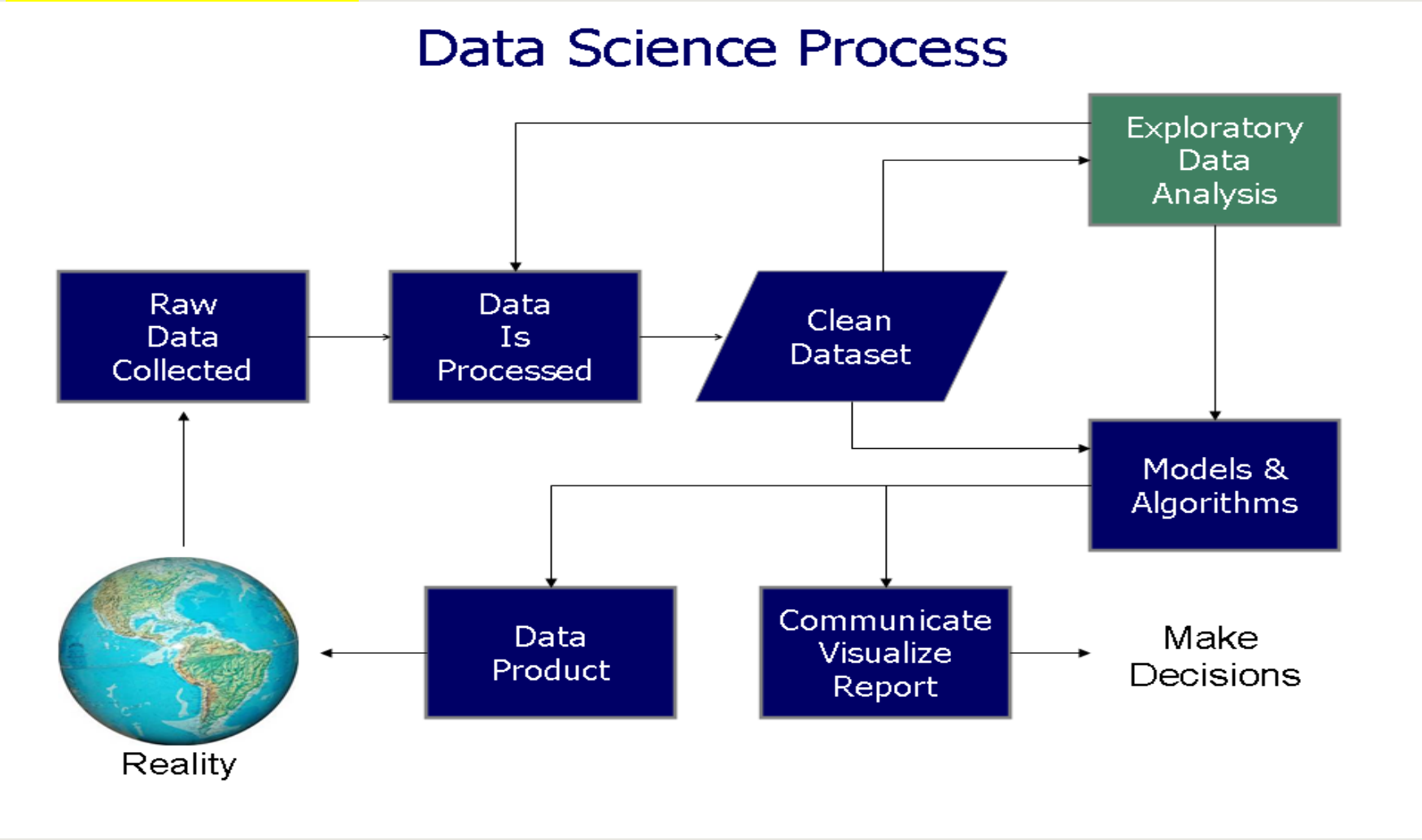
Analyse online user data to better predict when a user is about to leave a website. This analysis will inform how to design an algorithm that predicts a real-time score for each user's likelihood of leaving. This prediction will be based on their device type (computer or phone), scrolling, # of pages viewed, time on site, and many more metrics.

The variables used:

- Browser Name: Name of the browser used in the visit. Chrome, Edge, etc.
- Device Type: Desktop or Mobile.
- Event Name: Name of the event listener in KC Admin. Some relevant events are following:
    - "EI 10s timeout" fires after 10 seconds.
    - Max Scroll Depth on Page
- Exit Timing After Event: Main metric for predictivity. Compares Page of Visit Count to Pages Visited. If this is the final page view, classify the difference between Event Timestamp and Last Event Timestamp on Page. If the following page view is the last, this will say exited on next page. If there are 2 or more page views, it will say exited after 2+ pages.
- Pages Visited: Total number of page views in this visit.
- Path Coded: URL of this page view, split starting at the first /No query string.

- Referrer Type: Rough approximation of traffic source, e. g. visits from search, ad clicks, etc. To have this more accurate we'd need to manually code query strings of the entry page URL against Google Analytics data.
- Scroll Direction: Same as the value of the direction_scrolled event variable for Vertical Scroll events, but at event granularity.
- URL Coded: URL of this page view, minus query string.
- Max Scroll Depth on Page: Highest percent reported by the percent_scrolled event variable for the New Max Vertical Scroll % event on this page view
- Scroll Rate: Same as the value of the avg_scroll_speed event variable for Vertical Scroll events, but at event granularity. Measures in pixels/second
- User Type Based on Duration: I created this variable using the total duration of a user's stay. If the duration was less than 60 seconds, it was a "Random Visitor", if the duration was less than 120 seconds, it was a "Visitor Exploring" and the others were categorized as "Considerate Visitor".

REQUIRED STEPS IN A PREDICTIVE ANALYTICS PROJECT



The given steps are the ideal steps in a project done from scratch. I could not do all the steps because of the time constraint. I only focused on "Exploratory Data Analysis". I learnt to do basic interpretation by creating dataviews in Tableau.

Exploratory Analysis: Key Observations + Suggestions

KEY OBSERVATIONS

- For the largest number of "EI 10s Timeout" event, users left after 20s on both the devices.
- Considering the different web browsers, Firefox has the largest sample size. Most of the people using Firefox stayed for at least 20s.
- While comparing Max Scroll Depth on Page with the Exit Timing After the Event, I realized that for the "EI 10s Timeout" event, most of the people stayed on website at least for 20s.
- While scrolling downwards, a user is most likely to leave after 20 seconds . For scrolling upwards, the conclusion is not very certain because of the small sample size. Downward scrolling is more common.

SUGGESTIONS

- If "EI 10s Timeout" event is triggered in a visit, a user is likely to leave after 20s and thus we should trigger the survey after 20s.
- We can conclude that a user visiting the website through Chrome will most likely stay up until 20s and thus we can trigger the survey accordingly.
- Thus if "EI 10s Timeout" event is triggered then the user will at least stay for 20s regardless of the scroll depth on page except when scroll depth is 100, the user will view 2+ more pages which might extend the duration for more than 20s.
- Collecting more data specifically on people scrolling upwards might give a better comparison since scrolling upwards is not as common as scrolling downwards.

KEY OBSERVATIONS

- The sample size for random visitors was incomparably large. And most of those random visitors left the website after 20s.

SUGGESTIONS

- It can be concluded that random visitors are likely to leave after 20 sec. Collecting more data on other kinds of visitors might give a better picture of the comparison.

In general, the last suggestion would be to customize the first few pages of the website accurately especially the home page because those pages decide whether the user will stay or not. For the tools, I used Tableau using crosstabs to predict the exit intent of a user. Further, the analysis could be expanded by connecting Tableau to MySQL or any other coding platform to create machine learning algorithms.

TAKEWAYS FROM THE COURSE

- Initially I thought Data Science and predictive Analytics are used to refer to the same field but after talking to my mentors and doing some readings I realized predictive analytics is just one branch of data science. And data science is a very wide field.
- While interning with KnowClick, I realized that the contextual knowledge about the field or about the project I am working on is equally important as the technical skills. For example, in my project being aware of how a user behaves online was as useful as learning how to use Tableau.
- Though I was not being informed about the website that data was collected on and given to me, I could see the data structure applied to all kinds of businesses. Data is helping all the companies regardless of their different sizes or purposes build a better online presence.
- The last takeway is from the documentary called "Coded Bias". The documentary was a part of my discussion with my faculty advisor. For the first time, through the documentary I learnt about the impact of machine learning algorithms which are created using the data recorded in real life. The documentary emphasized how the use of machine algorithms in daily life are misclassifying people of color because it has been done in the past and that is the data machine algorithms are created from. The documentary helped me navigate the importance of ethics in data science.