SMU MITB INTERNSHIP AT HP INC.

Reaching 4 Stars: Using Text Analytics To Analyse amazon Web Review

Company Overview

HP Inc. is an American IT company that develops printers and supplies, 3D printing solutions and personal computers. It has worldwide operations across many countries including Singapore. One of its core businesses is Imaging & Printing business. Within this business, there is the Home Printing Solution Quality (HPS-Q) department, which divides into Home Customer Solutions (HCS) and Home Business Solutions (HBS).

Internship Department and Team

The department where I interned is **Home Customer Solutions (HCS)** and the team is **Customer Insights and Experience (CIE)**. One of the KPIs of my team is to ensure the quality of printers by developing customer insights from various sources to help business meets its strategic objectives. My team works closely with other functional teams in R&D, Service and Support to meet the sentiment target measured by web review ratings.

Project Motivation, Scope and Objective

Business Question

How product support managers gain customer insights from Amazon web review data.

- The key motivation to develop a data science solution is currently there is no existing tool for them to gain a quick understanding of how customer perceived the printer products in the market.
- By manually annotating web reviews into different categories, the team can gain valuable insights and understanding of trends over time. However, this manual process is tedious and extremely time consuming.

Project Requirements

A solution which can help to automate the classification process and provide **early detection of a rising trend in negative sentiments**. Upon detection, send an **auto email trigger** to alert the product support managers. The key business objective of my internship project is to increase the star rating of printers on Amazon website which suggests better user experience and higher product quality.

Role and Responsibility

- Use text analytics to analyse web review data
- Build a R Shiny dashboard that reads in data, perform trend analysis and send auto email trigger to product managers
- Present the project and insights to director and team

By: Soh Hui Shan (22 Jul 2019)
Supervisor: Mr. Parammal, Sanil Kumar (HP) and
Dr. Andrew Koh (SMU)

Internship Project Overview

Automated Data Science Solution

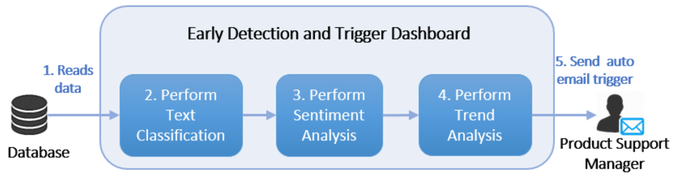

Figure 1: Process Overview of Automated Data Science Solution

The data science solution (Fig. 1) that I have created comprises of a dashboard, called Early Detection and Trigger Dashboard. It performs end-to-end automated analysis:

1) Dashboard is set up with seamless connection to the database and configured to read in data on a daily basis.
2) After data is read in, text classification is used to categorize web reviews into 4 predefined categories.
3) Next, web reviews are classified into positive or negative sentiment using sentiment analysis.
4) Perform trend analysis by analysing the trend changes of positive, negative and overall sentiment of each category over time.
5) Lastly, auto email trigger alert is sent to Product Support Manager when negative sentiment increase above the user-defined threshold.

Methodology of Data Analysis

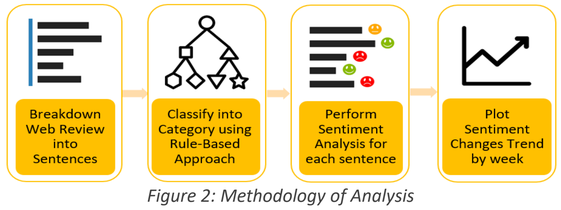

Figure 2: Methodology of Analysis

After data cleaning and data preparation, I performed the analysis (Fig. 2) to discover how negative review trends change across time.

1) Breakdown web review data into sentences
2) Classify sentences into 4 pre-defined categories using rule-based approach.
3) Perform sentiment analysis for each sentence by combining the sentiment value and star rating.
4) Plot the number of sentences for each sentiment type (positive, negative and overall) by week.

Accomplishments

- Insights generated are in-line with what was being observed in the market by the product team, suggesting that the model is useful
- Automated dashboard enables product team to gain timely insights on customer's experiences.
- Valuable time is saved as product team no longer have to classify manually.
- Seamless data connection remove the hassle of uploading data files.
- Early detection and auto email trigger capabilities empower product team to stay up-to-date when negative sentiment increases.

Dashboard Design

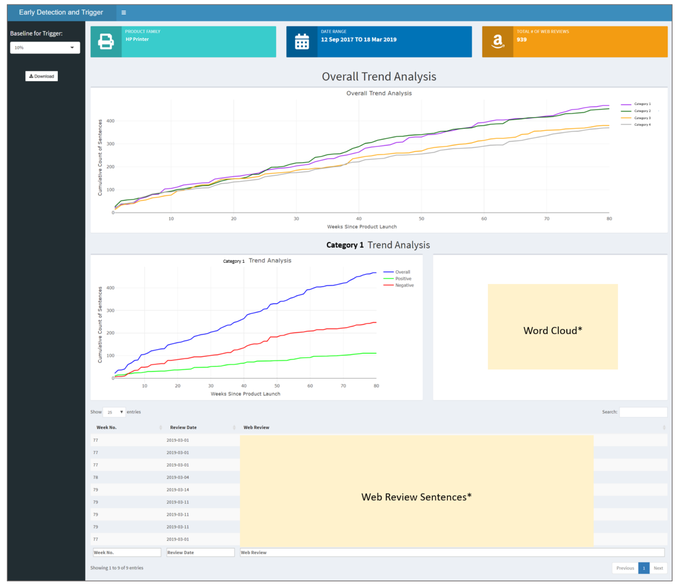

Figure 3: Early Detection and Trigger Dashboard

The views and functions which I have created at the dashboard include:

a) 3 Info boxes with product name, date range and number of reviews.
b) Overall trend graph for all 4 categories.
c) Individual trend graph for each category deep-dive.
d) Corresponding buzz sentiment word cloud for most popular issues.
e) List of reviews for each category with search function
f) Download file function to download data for further analysis.
g) Drop-down selection to customize trigger alert threshold level.

*Content is masked or replaced by generic terms due to data confidentially.

Reflections and Key Takeaways

One of my greatest takeaway from the internship experience is, I have observed data scientists should not work in silos to build models; in fact data scientists should work closely with the business team to understand their needs, and leverage upon the capabilities of data science to gain actionable insights that will help to achieve the business objectives. I have also gained very good exposure as I was given the opportunity to prepare executive slides and present my project and findings to the director and the team. I have honed by technical skills in R and Shiny and learnt PostgreSQL.

Acknowledgements

I sincerely thank my team, colleagues and fellow interns at HP for their guidance and support (and patience!) throughout the 6 months. The internship experience has been memorable and fruitful. I'm immensely grateful for the amazing opportunity to learn from and contribute to the team. My thanks extend to Andrew for his support and advices.