ShopBot

for earphones and headphones

Project Report



|  |  |
| --- | --- |
| **Team Members** | |
| Shashank Nigam | A0198469A |
| Lin Wenqi | A0198435R |
| Ng Mei Ying | A0198546L |
| Edmund Leow Kwong Wei  **TEAM MEMBERS**  Edmund Leow Kwong Wei  Ng Mei Ying  Wilson Lum Kok Keong A0198478A  M A S T E R O F TECH N O L O G Y  P R O J E C T R E P O RT    T E A M M E M B E R S  Edmund Leow Kwong Wei - 1  Ng Mei Ying - 2  Wilson Lum Kok Keong - 3  M A S T E R O F TECH N O L O G Y  P R O J E C T R E P O RT    T E A M M E M B E R S  Edmund Leow Kwong Wei - 1  Ng Mei Ying - 2  Wilson Lum Kok Keong - 3 | A0198458H |

Contents

[1 Executive Summary 1](#_Toc35955037)

[2 Business Problem 1](#_Toc35955038)

[3 Solution Approach 2](#_Toc35955039)

[3.1 Approach and Objective 2](#_Toc35955040)

[3.2 Datasets 2](#_Toc35955041)

[3.3 System Architecture 3](#_Toc35955042)

[4 Chatbot Design 4](#_Toc35955043)

[4.1 Chatbot Personality 4](#_Toc35955044)

[4.2 Intent Classification and Entity Recognition 4](#_Toc35955045)

[4.3 Intents 4](#_Toc35955046)

[4.3.1 Product recommendation 4](#_Toc35955047)

[4.3.2 Explanation of terminology 4](#_Toc35955048)

[4.3.3 Product pricing 4](#_Toc35955049)

[5 Challenges and Limitations 4](#_Toc35955050)

[5.1 Heroku Deployment 4](#_Toc35955051)

[6 Test Results 5](#_Toc35955052)

[7 Conclusions 5](#_Toc35955053)

[8 References 5](#_Toc35955054)

# Executive Summary

TODO

# Business Problem

In today’s world, traditional shopping is declining year-on-year, while e-commerce has been rising and is expected to reach 2.05 billion buyers globally by 2020 [1]. This has led to a much greater variety of products at lower prices, as multiple sellers across different countries compete for a slice of the market [2].

However, in a bid to differentiate their products, retailers often advertise their features with marketing jargon that adds to consumer confusion [3]. In addition, with the plethora of online marketplaces [4], even after deciding on the “best item” item, the consumer needs to decide which platform has the “best” price. All these factors result in consumer confusion and decision paralysis, where consumers experience greater anxiety and stress, and sometimes “buyers’ regret” after purchase, fearing that they have not chosen the “best” possible product [5, 6]. In 2016, a survey of 1000 Australians found 86 per cent believe too many products was making buying decisions harder [7]. Overwhelmed with choices and decisions, they may even delay or stop shopping all together [8].

In particular, there has been a surge in demand for headphones in the last few years. More than one million headphones were sold per day in 2018 and sales are expected to increase by two-fold in 2024 [9]. Consumers are no longer content with standard earphones that are bundled with mobile phones or MP3 players, and headphones and earphones have become lifestyle accessories for the trendy. Confronted with a large variety of product models, consumers may not always understand the different specifications or find it tedious browsing through countless items to find the product that best suits them.

# Solution Approach

## Approach and Objective

The project objective was to build a product recommendation chatbot for headphones and earphones to give users better “decision simplicity” in their shopping. The main intents are summarised in *Table 1*, and will be discussed in detail in *Chatbot Design*.

Table 1. Chatbot intents and descriptions

|  |  |
| --- | --- |
| Chatbot Intent | Description |
| Product recommendation | Select suitable products according to rating and sentiment in product reviews, as well as, product features deemed important to user (eg long battery life, suitable for sports, etc) |
| Explanation of terminology | Shortlist typical terminology and store the terms into a lookup table; if exact term is not found, word similarity between query and keys would be performed and closest match will be returned. |
| Product price and information | Webscrape prices from Amazon periodically (eg weekly/daily), and display price, features and website link to purchase in chatbot |

We assumed that we were an e-commerce website that sold headphones (eg Treoo). Because of a very wide selection of products, our customers might face the “Paradox of Choice”, and be overwhelmed with choices. In order to increase the conversion rate of web traffic to actual sales, a virtual shop assistant in the form of a chatbot, would be used to identify their needs and facilitate their choice-making decisions. In the event that our store did not hold stock, we would even recommend products from Amazon, so that we would become the one-stop website for headphones.

## Datasets

Several datasets would be utilised for this project. For “Product recommendation” intent, the Amazon Review Data (2018) which contained reviews from 1996 to 2018 **Invalid source specified.** as well as reviews of headphones and earphones from Amazon extracted from webscraping were used for sentiment analysis and also to obtain the list of available products to recommend.

For “Explanation of terminology” intent, several websites were web-scraped to obtain a glossary of terminologies. For “Product Pricing” intent, Amazon website was web-scraped for price and delivery costs based on the list of products that were shortlisted in “Product Recommendation”.

The list of dataset sources is summarised in *Table* 2.

Table 2. Dataset sources and usage

|  |  |  |
| --- | --- | --- |
| Purpose | Dataset source | Dataset type |
| Sentiment Analysis | Amazon Reviews (2018) dataset from  <https://nijianmo.github.io/amazon/index.html>  Headphone and earphone reviews from Amazon | Text |
| Information extraction | Product review websites | Text |
| Aspect extraction | Headphone and earphone reviews from Amazon | Text |
| Topic modelling | Headphone and earphone reviews from Amazon | Text |
| Short-listing of products | Shopping websites (Amazon, Treoo) | Text,  product IDs |
| Explanation of terminology | Websites specialising in headphones ([Headphones.com](https://www.headphones.com/pages/glossary), [CrutchField.com](https://www.crutchfield.com/S-P3Fy2Oh1DMm/learn/headphones-glossary.html), [Krisp.ai](https://krisp.ai/blog/glossary-headphone-terms/)) | Text |
| Product Pricing | Amazon Website and Treoo Website | Prices |

## System Architecture

The chatbot was hosted on [DialogFlow](https://dialogflow.com/) due to its ease of setup and wide integration with multiple platforms including Google Assistant, Slack and Facebook Messenger. However, intent classification and entity recognition were performed by [Rasa NLU](https://rasa.com/docs/rasa/nlu/about/), as it allowed greater customisation in terms of processing and training.

In order to easily embed our chatbot into our website as a pop-up, we leveraged [Kommunicate](https://www.kommunicate.io/), as it also allowed easy integration with DialogFlow.

When a user enters a message into the chatbot platform, its API sends the message to DialogFlow. Instead of using DialogFlow’s NLP engine to process all intents and entities, we utilised it for only basic categorisation and sent the raw input to our flask webhook app for preprocessing and then forward it to Rasa NLU model to determine intents and entities. The appropriate response is formulated in the flask app and then returned to the user. The detailed system architecture is shown in *Figure 1*.

|  |
| --- |
|  |

Figure 1. Overall System Architecture

Ideally, for deployment, the Flask app and Rasa NLU server should be hosted in Heroku server. However, there were several problems in doing so[[1]](#footnote-1), and finally for this prototype, it was assumed that the Flask app and Rasa NLU server would be ran locally, and exposed to the internet with a public url provided through ngrok.

# Chatbot Design

This section discusses the different aspects of our chatbot design, including chatbot persona, entities, intents and conversation flows.

## Chatbot Personality

Our target audience will be mainly young people within the age range 18-34., as these were deemed to be the people that had the highest chance of intending to purchase earbuds within the next 12 months [10].

There are 12 archetypes commonly used to define a brand persona, which are functionally identical to the 12 Jungian archetypes [11]. As our target audience are young people, we will stereotype them as Explorers, who are ambitious and always seek out new things. In order for them to identify well with our bot, we will mirror the personality of our audience [12] and also give it the same archetype Explorer, with a friendly, energetic and enthusiastic personality. This will be projected in its interactions with our users. For example, when returning the price for a specific item, it might respond with something like “Woohoo! It’s your lucky day! 🤩 JAYS A ONE is available at S$58.00!”

In addition, in order to not be repetitive, there would be multiple variations for each response of our chatbot. This was catered for by having choosing randomly from lists at different parts of the response, thereby having many multiple permutations.

In coming up with a name, it was important that we had a memorable name that was linked to our core business of earphones and music, so that our users could easily remember it. We came out with AudioPhil, as a play on the word “audiophile”, who is a person who is enthusiastic about high-fidelity sound reproduction. Phil is also uncommon enough while being an easy name to remember.

In coming up with an avatar picture, we wanted the users to have reasonable expectations of our chatbot. It was found that consumers expected chatbots to behave like human agents but wanted it to be clear that they are bots [13]. In addition, if the chatbot was too convincing as a human, people were more likely to speak quickly and less clearly, and have higher expectations for the system [14]. Hence, we adopted a cartoon-like avatar so that they will know immediately that it was a bot. At the same time, we wanted it to appear friendly, hence we opted for a human avatar. Since our chatbot is supposed to be like a virtual shop assistant, and since we recommend headphones, it would be nice if he had headphones on him.

With all these in mind, our final chatbot persona is given in *Figure 2*.

|  |  |  |
| --- | --- | --- |
| **AudioPhil** | **Interests and Passions**   * Listening to music * Audio history and latest technology in headphones * Trying out the newest headphones on the market * Helping people find their ‘headphone soulmate’   **Quotes**     |  | | --- | | Hello there! I am Phil. Let’s work together to discover the perfect headphones for you today! | |
| **Friendly • Energetic • Attentive** |
| **Age**: 20  **Occupation**: Earphone Purchase Specialist  **Location**: Singapore  **Archetype**: Explorer (ENTP) |

Figure 2. Chatbot Persona

## Intent Classification and Entity Recognition

Basic natural language processing (NLP) was first carried out using DialogFlow. However, it had several major limitations such as (i) unable to perform customised preprocessing and processing of inputs, and (ii) unable to easily input a large list of training and test phrases. Hence, we adopted a hybrid approach, where intents and entities were defined within DialogFlow, and a first-round basic intent classification and entity recognition was done. If DialogFlow fails to recognise the intent, second-round processing for intent classification and entity recognition was done by Rasa.

This allowed us to adopt the following custom preprocessing pipeline for user inputs as shown in *Figure 3*. By performing such preprocessing, we simplified the training examples that were required (eg. No need to consider Singlish; no need to consider both singular and plural tenses, etc).



Figure 3. Custom pre-processing pipeline

Training example datasets were generated using a third-party tool [Chatito](https://github.com/rodrigopivi/Chatito), by defining examples in a custom domain specific language (DSL), which simplified the generation of samples. A snippet of the chatito file used to generate training samples for “Explanation of terminology” is given in *Figure 3*. From the simple definitions below, training and testing datasets could be easily generated by iterating and permutating through the different combinations that were defined.

In order to streamline our workflow, a script (\_generate\_lookup\_files.py) was also created to generate lookup tables for the entities of the intents, as well as to add these to the chatito file that would be used to generate training and testing examples.

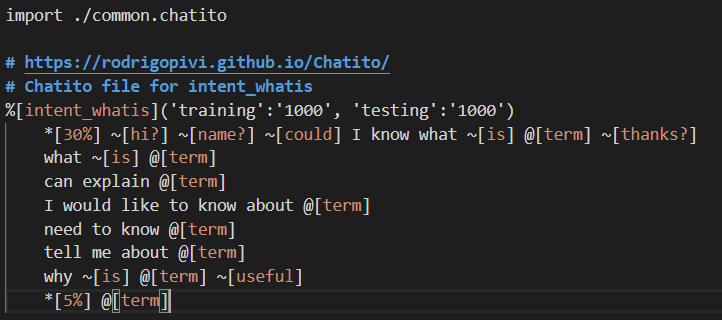


Figure 4. Example snippet of chatito file

In Rasa NLU, incoming messages are processed by a sequence of components that will executed one after another in a so-called processing pipeline. The three main parts are (i) tokenization, (ii) featurization and (iii) entity recognition/intent classification/response selectors[[2]](#footnote-2). The following pipeline, supervised embeddings in *Figure 5* was used. This was so that word vectors could be customised for our specific domain.



Figure 5. Rasa pipeline using supervised\_embeddings

## Intent Details

An overview of intents is shown below. There are three main intents for our system

* Product Recommendation
* Explanation of terminology (FAQ)
* Product Information

In general, each of the intents rely on a database that was scraped from websites as described previously.



### Product Recommendation

TODO

### Explanation of Terminology

In order to get a list of domain-specific terminologies, we searched for websites that had ready-made glossary terms for headphones, and finalised on glossary terms from three websites (See *Table 3*). In cases where there were repeated glossary terms, we just took the description that was shorter. Interestingly, the number of overlapping terms was small, which helped to expand our range of glossary terms, giving us a total of 116 terms (See *Figure 3*). All web-scraping was performed in the notebook Glossary.ipynb.

Table 3. Websites for glossary terms

|  |  |  |
| --- | --- | --- |
| **Website** | **Website URL** | **Total terms** |
| Headphones.com | <https://www.headphones.com/pages/glossary> | 63 |
| Crutchfield.com | <https://www.crutchfield.com/S-P3Fy2Oh1DMm/learn/headphones-glossary.html> | 36 |
| Krisp.ai | <https://krisp.ai/blog/glossary-headphone-terms/> | 28 |

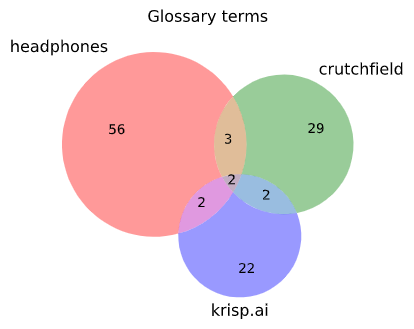


Figure 6. Venn diagram showing overlaps of glossary terms among the webscraped sites

If the search query was found in our list, we would return the description and source of information to the user.

In the event that it was not found, we would perform similarity matching of query with terms in the database, based on word vectors using [spacy model 'en\_core\_web\_md'](https://spacy.io/models/en#en_core_web_md) (English multi-task CNN trained on OntoNotes, with GloVe vectors trained on Common Crawl), and return the closest matched term.

Otherwise, we would just tell the user that we do not know what that term is.

See **User Guide** for test scenarios for the intent.

### Product Information

We had two lists of products from Treoo and Amazon Singapore. For Amazon, we realised that the international site had a lot of products that did not ship to Singapore but had more sellers, in different conditions (new or used), different delivery cost (free or not, and if not, price differed greatly depending on type of delivery).

In order to simplify these considerations, we instead took products from Amazon Singapore, which all had free delivery and had only new items. There were numerous items that did not state a brand (*See Figure 7 and Figure 8*). Filtering for only products that had a brand, model and price, we had a total of 87 brands consisting of a total of 230 models.

A picture containing food, room

Description automatically generated

Figure 7. Packed bubble chart showing Amazon products before filtering

A screenshot of a social media post

Description automatically generated

Figure 8. Amazon product frequency by brand before filtering

For Treoo, there was a much greater number of total models (*see Figure 9 and Figure 10*), but product name was very inconsistent. Some inconsistencies, using AKG models as examples, are shown in *Table 3*. Therefore, various pre-processing methods had to be done to standardise them. In total, there were 144 brands and 1643 different models.

Table 4. Example inconsistencies in Treoo product names for AKG brand

|  |  |
| --- | --- |
| **Example** | **Comment** |
| AKG K52 | Ideal format |
| AKG K 545 | Extra space between K and 545 |
| AKG K845BT Wireless Bluetooth Over-the-Ear Headphone | Features are also included in the description |

A picture containing ball, room

Description automatically generated

Figure 9. Packed bubble chart showing Treoo products before filtering

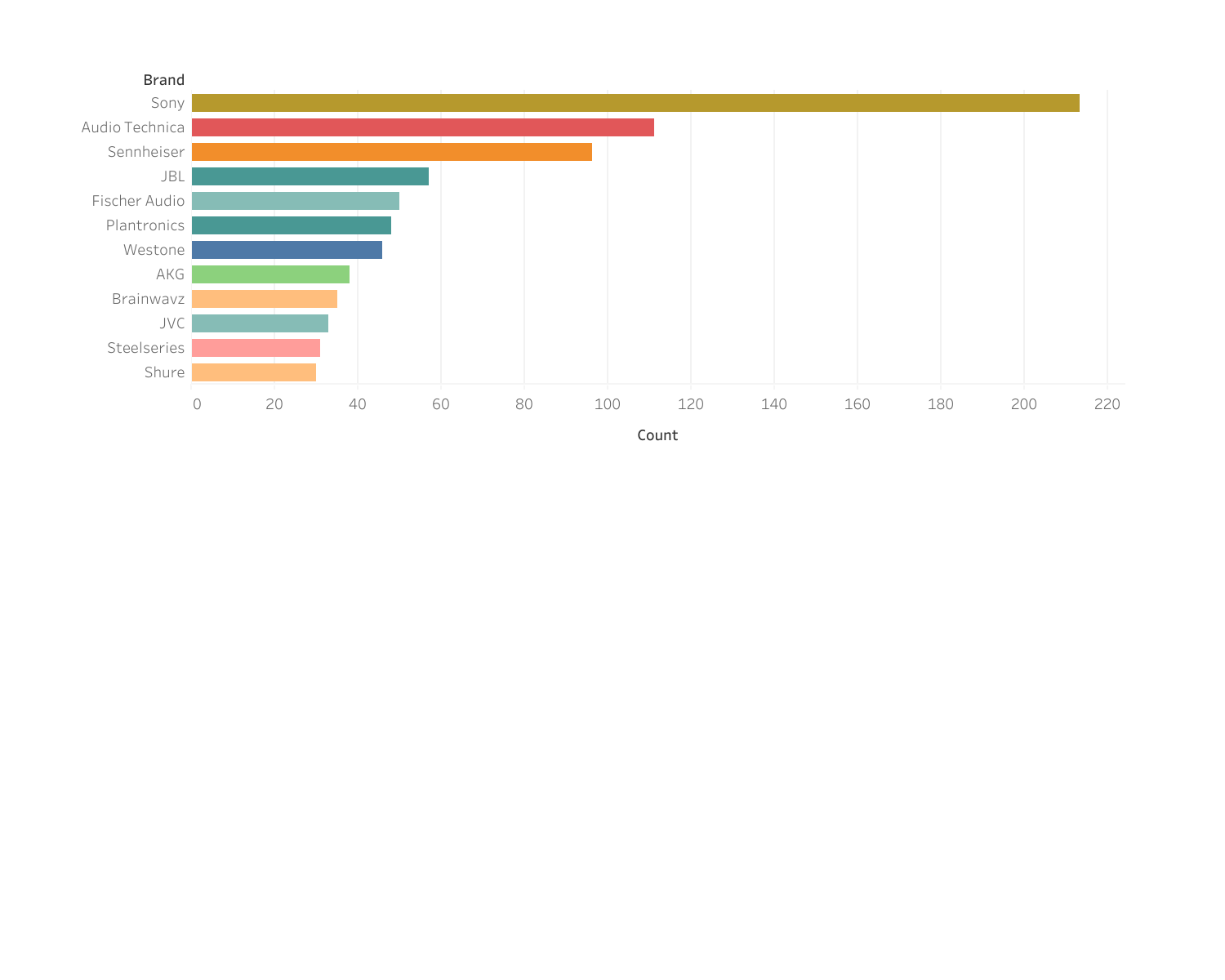


Figure 10. Treoo product frequency by brand before filtering

We combined both lists of brands and models and found that there were 29 brands that overlapped. However, there were only 24 models that existed in both Treoo and Amazon (See *Figure 3*). This small overlap suggested that there were possibly similar models that should be regarded as the same model. However, to capture more variation in user queries, and in the interest of time, we decided not to further process the model names.

For the same brand and model, we took the item with the lower price. It was found that when the same item existed in both Treoo and Amazon, Amazon was almost always the cheaper one.

All pre-processing was performed in the script file \_generate\_lookup\_files.py.

|  |
| --- |
|  |

Figure 11. Venn diagrams showing Treoo and Amazon brands and models

In order to match the user’s query, we would first check if a product model was given. If not, we would ask them for it. If the model is unique (ie. Only 1 brand has this specific model), we could then retrieve the results. Otherwise, we would ask them for the brand name as well. If a suitable brand-model pair was found, then results would be returned.

See **User Guide** for test scenarios for the intent.

# Test Results

The intents for “Explanation of Terminology” and “Product Information” were trained on 1000 samples and tested on another 1000 samples each in Rasa NLU. Intent classification evaluation was performed via [rasa’s in-built functionality](https://rasa.com/docs/rasa/1.0.9/user-guide/evaluating-models/#intent-classification). The intent prediction confidence distribution showed that most samples were predicted with a very high confidence (*Figure 12*). The confusion matrix also showed that most intents were correctly classified (*Table 5*).

Table 5. Confusion table for intents

|  |  |  |  |
| --- | --- | --- | --- |
|  | **greet** | **intent\_price** | **intent\_whatis** |
| **greet** | 100 | 0 | 0 |
| **intent\_price** | 1 | 998 | 1 |
| **intent\_whatis** | 0 | 0 | 1000 |

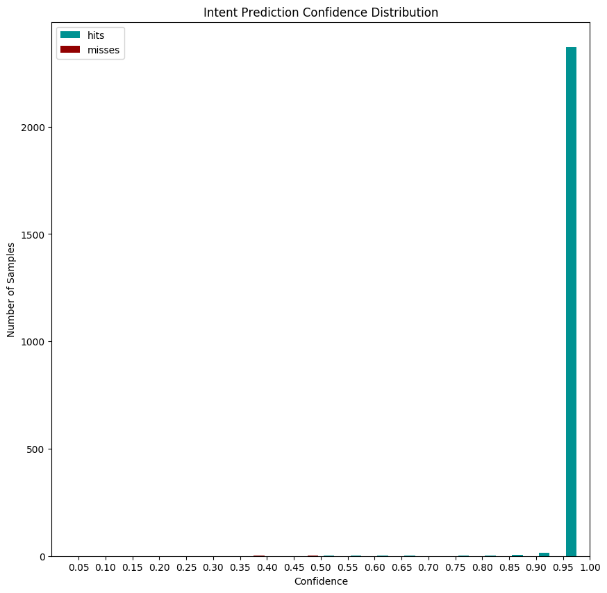


Figure 12. Intent Prediction Confidence Distribution

For “Product Recommendation”, intent classification and entity recognition were done mainly in Google DialogFlow, which did not require a large training set, and had no easy way to evaluate performance.

TODO

# Challenges and Limitations

We encountered several issues with DialogFlow, Rasa NLP, and Heroku deployment while working on the chatbot and will elaborate on them in the following subsections.

## Heroku Deployment

Initially, we deployed our system automatically to Heroku. However, later on in our development, we realised the limitations of the free tier of Heroku.

Firstly, we exceeded the maximum [slug size](https://devcenter.heroku.com/articles/slug-compiler#slug-size) limit of 500MB was exceeded, leading to a failed deployment. This was because we were using many different libraries especially in terms of NLP (eg spacy model and rasa). To circumvent this, the flask webhook was hosted on one app, and rasa NLU server was hosted on another.

Next, we exceeded the memory quota ([errors R14 and R15](https://devcenter.heroku.com/articles/error-codes)) where our dyno in Heroku required too much memory. This was due to loading of too many large variables into memory, such as the spacy 'en\_core\_web\_md' model (91MB), and various lookup tables for our querying of headphone models and brands. We tried to switch to the spacy ‘en\_core\_web\_sm’ model and managed to deploy, but faced performance degradation for similarity matching.

After lots of time trying to fix the issue, we finally gave up in the interest of time, and decided to only cater for a local deployment. However, these errors would be non-existent if using higher tiers of Heroku such as the [Performance M](https://www.heroku.com/pricing) which has 2.5GB RAM.

## DialogFlow

DialogFlow has an unconfigurable timeout of five seconds for webhook responses. This feature was to ensure that the chatbot responds within an acceptable time limit for natural conversation. However, this meant that we could not fetch real-time price information for the user via web-scraping, as it would exceed the time constraints. Hence, we resorted to using cached data for our results.

In an actual deployment, we would periodically web-scrape (eg weekly) and update our data cache to ensure that prices were not too different from the current price. An alternative was to subscribe to APIs that allowed the retrieval of product price and information from multiple sellers, which would be typically much faster than web-scraping. Unfortunately, there were no free sources of APIs for this, and we could not choose this option.

# Conclusions

In this project, we utilised Google DialogFlow, and rasa NLU for intent classification and entity extraction and performed various text mining techniques for classification and information extraction. By using sentiment mining of Amazon reviews through a deep neural network, we were also able to recommend a suitable product to the user based on the features that were important to him. This could greatly help a visitor to the website narrow down his choices or get relevant information that he required, thereby increasing sales conversion and revenue for the website.

# References

|  |  |
| --- | --- |
| [1] | Oberlo, “Find out How Many People Shop Online in 2020,” [Online]. Available: https://www.oberlo.com/statistics/how-many-people-shop-online. |
| [2] | “The growth of e-commerce means more choice and better deals,” The National, 2 Jan 2018. [Online]. Available: https://www.thenational.ae/opinion/editorial/the-growth-of-e-commerce-means-more-choice-and-better-deals-1.692313. |
| [3] | R. Carter, “Lost in translation: The dangers of marketing jargon,” Fabrik Brands, 1 Aug 2019. [Online]. Available: https://fabrikbrands.com/the-dangers-of-marketing-jargon/. |
| [4] | “The biggest online marketplaces that you should know,” IONOS, 6 8 2019. [Online]. Available: https://www.ionos.com/digitalguide/online-marketing/online-sales/the-biggest-online-marketplaces-that-you-should-know/. |
| [5] | B. Schwartz, The paradox of choice : why more is less, New York: Ecco, 2004. |
| [6] | L. M. Lyengar S, “When choice is demotivating: Can one desire too much of a good thing,” *Journal of Personality and Social Psychology,* vol. 79, pp. 995-1006, 2000. |
| [7] | Knowledge@Wharton, “Online Shopping Choices: Less Is Sometimes Better Than More,” Wharton University of Pennsylvania, 17 Dec 2013. [Online]. Available: https://knowledge.wharton.upenn.edu/article/online-shopping-choices-less-sometimes-better/. |
| [8] | A. A. &. Intelligence, “Global Earphones and Headphones Market to Reach Values of $36 Billion During the Period 2018−2024,” PRNewswire, 14 Feb 2019. [Online]. Available: https://www.prnewswire.com/news-releases/global-earphones-and-headphones-market-to-reach-values-of-36-billion-during-the-period-20182024--market-research-by-arizton-300795642.html. |
| [9] | T. O. Team, “The State of Play 2019: What’s next for audio tech,” Qualcomm, 5 Sept 2019. [Online]. Available: https://www.qualcomm.com/news/onq/2019/09/05/state-play-2019-whats-next-audio-tech. |
| [10] | N. Smith, “Does Your Brand Have Multiple Personality Disorder? A Look at Brand Archetypes,” nvision designs, 10 November 2015. [Online]. Available: http://www.nvision-that.com/design-from-all-angles/what-is-your-brand-personality-a-look-at-brand-archetypes. |
| [11] | T. Manifest, “How to Give Your Chatbot a Personality,” 11 Dec 2018. [Online]. Available: https://chatbotslife.com/how-to-give-your-chatbot-a-personality-5e0fb239b28c. |
| [12] | Leah, “What Do Your Customers Actually Think About Chatbots?,” Userlike, 30 Jan 2019. [Online]. Available: https://www.userlike.com/en/blog/consumer-chatbot-perceptions. |
| [13] | K. Waddell, “Chatbots Have Entered the Uncanny Valley,” The Atlantic, 21 Apr 2017. [Online]. Available: https://www.theatlantic.com/technology/archive/2017/04/uncanny-valley-digital-assistants/523806/. |
| [14] | V. MacMillan, “How to Develop a Chatbot Persona That Fits Your Brand,” Medium, 26 April 2018. [Online]. Available: https://chatbotslife.com/how-to-develop-a-chatbot-persona-that-fits-your-brand-c48055970372. |
| [15] | T. F. Tay, “Tourist arrivals, spending in Singapore at record highs,” Straits Times, 14 February 2019. [Online]. Available: https://www.straitstimes.com/singapore/tourist-arrivals-spending-at-record-highs. [Accessed 28 August 2019]. |

1. See *Challenges and Limitations* [↑](#footnote-ref-1)
2. See <https://rasa.com/docs/rasa/nlu/choosing-a-pipeline/#understanding-the-rasa-nlu-pipeline> for details [↑](#footnote-ref-2)