***‘Vira’ – a model Chatbot for Alaskan airlines to help customers with Covid19 related questions***

***Team Members***

*Bhargavi Sankula (qv4668)*

*Kavitha Perumal (mv9432)*

*Pooja Bhansali (gs6262)*

*Ponmalar Rajan (zy5227)*

*Sowmya Chintha (jf8997)*

### **CONTENTS**

### NARRATIVE DESCRIPTION ABOUT THE PROJECT BACKGROUND. . . . . . . . . . . . . . . . .1

### RESEARCH QUESTION. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2

### WHY IS THE PROJECT USEFUL/IMPORTANT?. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2-3

### DATA DESCRIPTION . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. . . . . . . . . . . 3-6

### Data pre-processing . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5-6

### SUMMARY STATISTICS . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .6-11

### 5.1 Word cloud . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .6-7

### 5.2 Lexical Dispersion Plots. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .7-8

### 5.3 Accuracy. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .8-9

### 5.4 Bar plot to show distribution of topics/questions . . . . . . . . . . . . . . . . . . . . . . . . . 10

### 5.5 Bar plot to show average length of questions in topics/questions. . . . . . . . . . . . 11

### METHODOLOGIES REQUIRED TO COMPLETE THE ANALYSIS. . . . . . . . . . . . . . . . . . . . 11-13

### A brief step by step process of our chatbot building . . . . . . . . . . . . . . . . . . . . . . . 12-13

### ANALYSIS RESULT REPORT AND EXPLANATION . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .13-17

1. CONCLUSION. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .17

### REFERENCES. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 18

### **1.Narrative description about the project background**

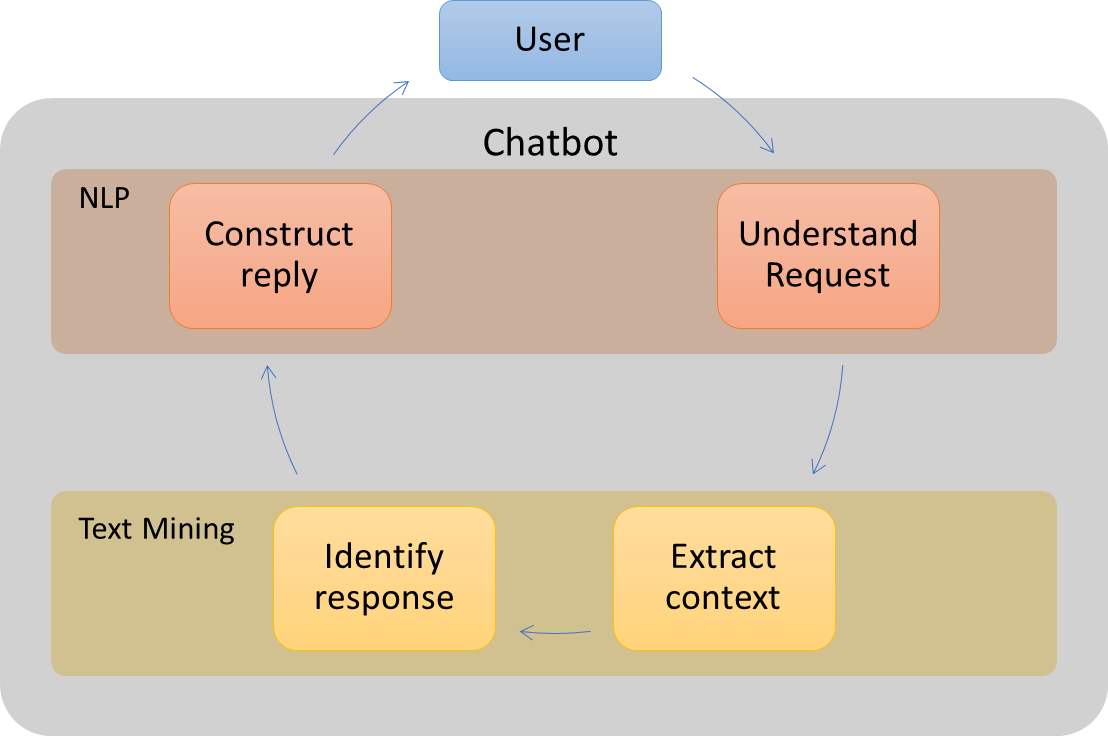
In the current COVID-19 pandemic situation, many Airline companies are struggling to keep up their customer care service as high volumes of travel related enquiries and requests come in the form of calls, chat requests and emails. Some notable airlines like Singapore Airlines have already come up with beta chatbot ‘Kris’ which is helping with so many covid19 related travel questions.

So, when we decided to develop a model chatbot for our project, Alaska airlines was an easy pick since they do not have an existing chatbot. Headquartered in Seattle, ‘Alaska Airlines’ is one of the most popular domestic airlines in the US. They are fifth largest airlines based on traffic and have services to many domestic and International destinations. They are one of the highly rated airlines in customer satisfaction and were recently ranked highest for a 12th consecutive year in the J.D. Power North America Airline Satisfaction Study.

Alaska Airlines, known for their uncompromised customer experience we noticed had to discontinue their chat services temporarily due to practical difficulty to support the huge volume of customer service requests under current situation. Some of these requests can be very trivial like Travel Advisories, Cancellations, Baggage/refund policies and other services.

A quick solution would be to introduce a Chatbot which can help answer these basic customer questions. Chatbots understand human natural language and provide intuitive conversation friendly responses. It can be programmed in this situation to learn the various frequently asked questions and respond based on the context of the questions asked.

**Our model Chatbot ‘Vira’ has been developed using textual Dataset taken from Alaska Airlines website’s Covid19 Pandemic Breakdown related content displayed as bulletins and hyperlinks in their main page. We believe ‘Vira’ may be a quick solution for Alaska Airlines to serve their customers efficiently by answering Covid19 related customer questions.**



***Figure 1-1. Chatbot working cycle***

### **2. Research question**

Under the current corona pandemic outbreak situation, traditional customer call/chat service of airlines that involves human resources may not work because of shelter in place orders and also many airlines are experiencing high call volumes with questions related to cancellations and changing flights. Alaskan airlines have discontinued their chat services, so what practical technique/tool Alaskan airlines can rely on to address this issue and assist their customers in questions related to cancellations, flight Status and travel advisories ?

### **3. Why is the project useful/important? What are the implications of the research?**

* Our model Chatbot ‘Vira’ would be highly useful given the current pandemic situation which unfortunately prompted Alaska airlines to discontinue their chat services due to high call volumes. So, Vira at 84% accuracy rate can help panic stricken customers to get answers in just one click away. Also, it can help Alaska airlines to take that extra load off from customer care representatives but at the same provide customer service without any quality and efficiency compromise. So, ‘Vira’ would be an absolute win-win.

**A screenshot of a social media post

Description automatically generated**

***Figure 3-1 snapshot taken from Alaska Airlines website main page showing their chat is currently unavailable***

* We believe our chatbot Vira may have a great impact in future in the areas of  ‘Dialog systems’ and ‘Automated customer service’ for Alaskan airlines if adapted . ‘Vira’ at the development model phase has good accuracy, so with more data generated and collected from live conversations in future will help to further improvise it as a complete functional chatbot. We also believe it might help Alaska airlines not only with their customer service but also highly useful from a business perspective in increasing cost, resources, and time efficiency .

### **4. Data Description**

Vira is a retrieval-based model i.e. it uses the predefined rules to train on and respond from the provided responses. Our model uses the message and context of the conversation to select the best response from a list of bot messages.

The predefined intents are the expected responses defined in the python programming behind the scenes by considering what user has provided. For this to happen data collection in a predefined intent is key for our chatbot development. Once the user intent has been determined, Vira will respond to the query with the appropriate response .

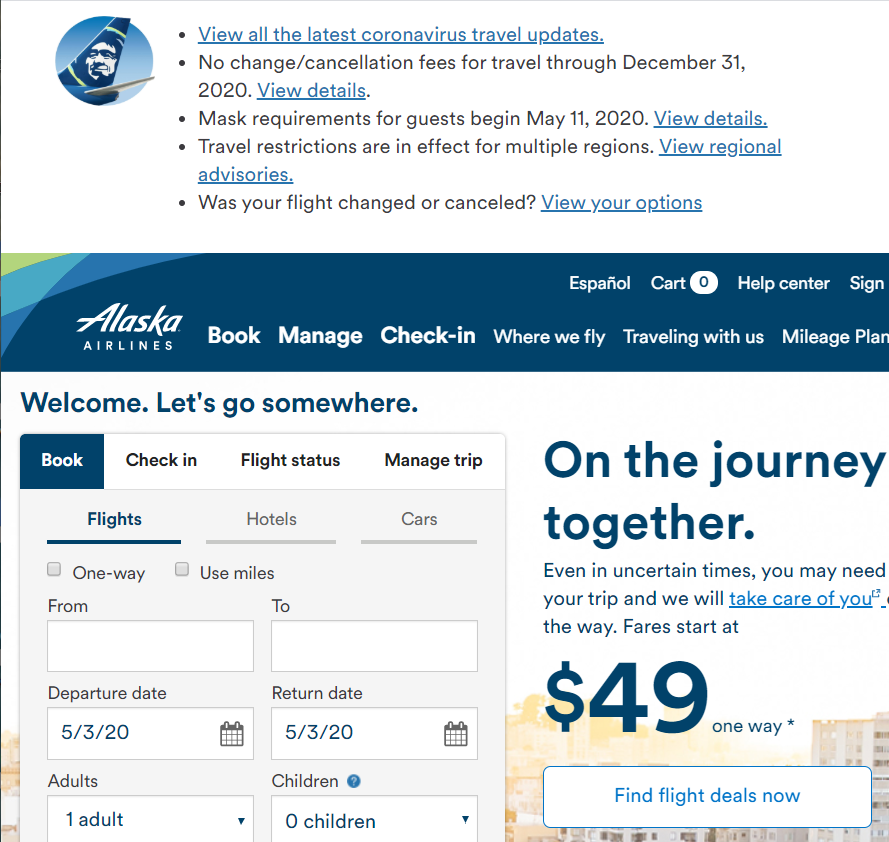
For our data set, we directly referred Alaska Airlines which has bulletined important links for questions related to Covid19. This raw textual data collected from their website has been converted into a cleaned data set in the format of subject/topic line and appropriate response. The predefined intents that we used for subject ending with full stop (.) and response starting with a hashtag (#) , ending with a full stop (.) respectively .

The raw text data which was carefully cleaned to become our dataset can be found at this link,

[https://www.alaskaair.com/content/advisories/travel-advisories?int=AS\_HomePage\_AdvisoryBR\_L2||2020\_CV\_AW||-prodID:Awareness&lid=HomePage\_AdvisoryBR\_CancelFees#flex](https://www.alaskaair.com/content/advisories/travel-advisories?int=AS_HomePage_AdvisoryBR_L2%7C%7C2020_CV_AW%7C%7C-prodID:Awareness&lid=HomePage_AdvisoryBR_CancelFees#flex)

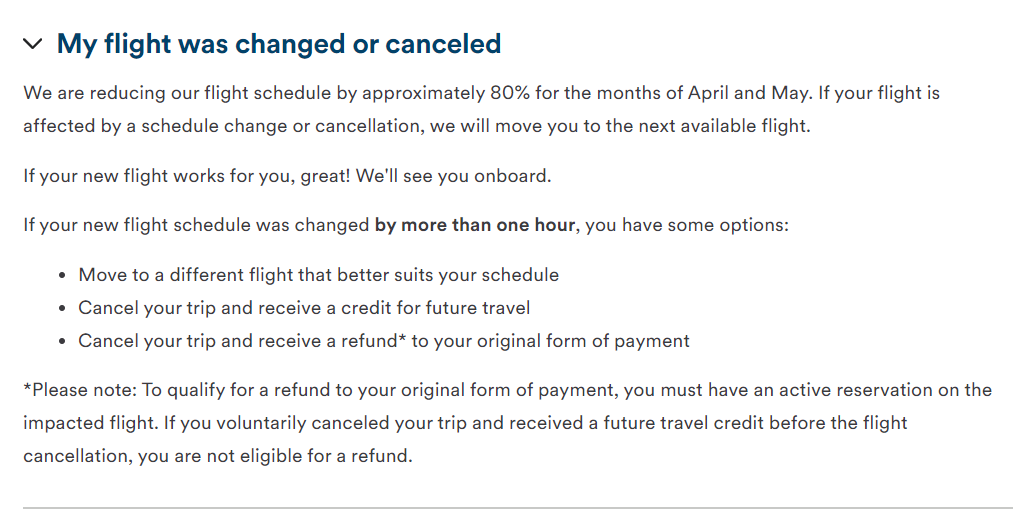
In above link, Alaska Airlines tried their best to provide answers to five important topics all related to Covid19 and this raw textual data have been taken and preprocessed to form five different topics of our dataset that are listed below,

1. **Cancellation**
2. **Travel and Regional Advisories for covid19**
3. **Seating and Safety measures for covid19**
4. **Baggage**
5. **Onboard services and Lounge services for covid19**



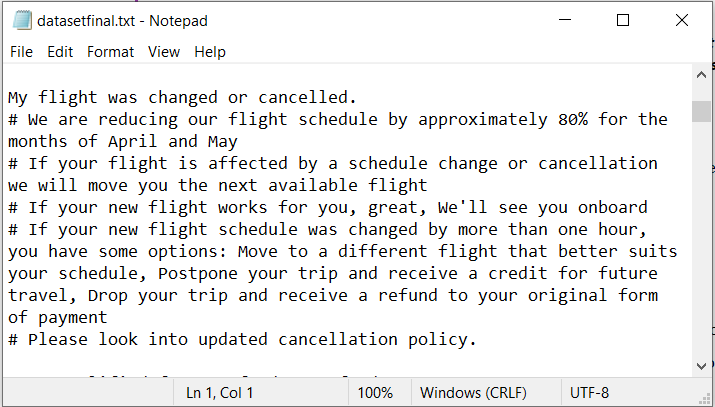
***Figure 4-1 showing Alaska Airlines website main page with covid19 related bulletins***

***and hyperlinks at top from where textual raw data has been taken for our dataset***



***Figure 4-2 snapshot of raw textual data related to cancellation policies***

***in Alaska Airlines website***

******

***Figure 4-3 snippet of cleaned Cancellation section from our Dataset that is***

***in accordance with Figure 4-2 Raw Data taken from Alaska Airlines website***

### **4.1 Data pre-processing**

Since we have the data set in text format i.e. strings , we need to preprocess the text using the NLTK package . We used basic text preprocessing modules as follows :

* Entire text conversion from uppercase into lower case, so that the model does not treat the same words in upper or lower cases as different .
* Tokenization**:** To convert normal text strings into a list of tokenized words. Sentence tokenization is used to find the list of sentences and word tokenization to find the list of words, respectively .
* Removing noise (punctuations) and stop words for English language, since we built chatbot for Alaska airlines for English speaking users only .
* Stemming of words  to result in a single word.
* Lemmatization of words to get the actual words .

### **5. Summary Statistics**

Summary statistics gives a brief and simple description of our data. There are literally dozens of ways

to display summary data using graphs or charts.

For our project we have used following formats to quickly summarize our data:

1. **Word Cloud**
2. **Lexical Dispersion Plots**
3. **Accuracy**
4. **Bar plot to show distribution of topics/questions**
5. **Bar plot to show average length of questions in topics/questions**

### **5.1 Word cloud**

* Word clouds work in a simple way: the more a specific word appears in our dataset which is textual data, the bigger and bolder the word appears, the more often it is mentioned within our dataset and the more important it is.
* Since our dataset is text based, we wanted to make use of word cloud as visualization format to highlight important data points and immediately convey crucial information

A screenshot of a cell phone

Description automatically generated

***Figure 5-1 Word cloud depicting most important text data points***

### **5.2 Lexical dispersion plots**

* A word’s importance can be weighed by its dispersion in a corpus. Lexical dispersion is a measure of a word’s homogeneity across the parts of a corpus.
* For an airline chatbot project, what would be the most repeated and important words? Yes, Travel and Flight. Below lexical dispersion plots display occurrences of words ‘Travel’ and ‘Flight’.

A screenshot of a cell phone

Description automatically generated ***Figure 5-2 Lexical Dispersion plot for word ‘Flight’***

A screenshot of a social media post

Description automatically generated

***Figure 5-3 Lexical Dispersion plot for word ‘Travel’***

### **5.3 Accuracy**

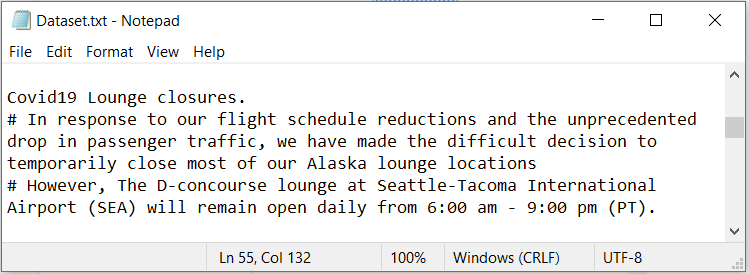
* Since our chatbot Vira is still on development phase, we calculated accuracy by randomly asking 50 different questions and then calculated the accuracy as below:

***Accuracy = No of questions our chatbot Vira replied accurately/Total No of questions asked***

***= 42/50 => 0.84 or 84%* accuracy is pretty good for a Covid19 related model chatbot**

* **And how did we know that the answers are accurate?**

We manually compared the answer extracted to a random question with the dataset of that  topic.  For example, we asked ‘Vira’ a question about Lounge policy and it answered accurately in accordance with the Lounge Closures section from the dataset. We manually compared ‘Vira’ answers for randomly drawn questions with our dataset like above example and validated.



***Figure 5-4 snapshot of actual ‘Lounge closures’ topic from our Dataset***

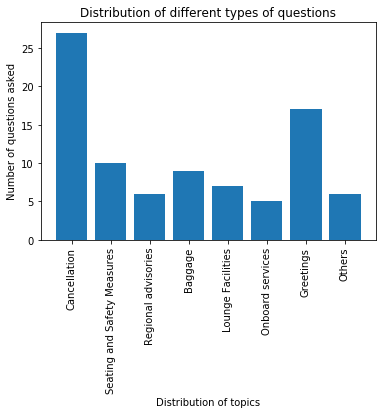
A screenshot of a social media post

Description automatically generated

***Figure 5-5 snapshot of Vira chatbot window with lounge related question answered***

### **5.4 Bar plot to show distribution of topics/questions**

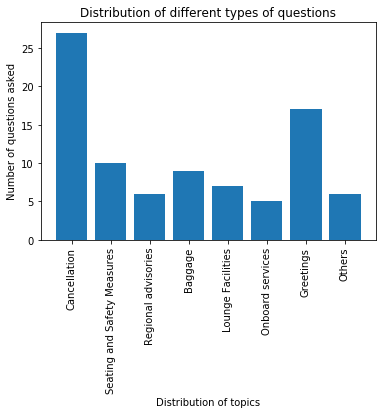
We used the bar plot to give an approximate representation of the [distribution](https://en.wikipedia.org/wiki/Frequency_distribution) of topics or questions.



***Figure 5-6 Bar plot of topics distribution***

### **5.5 Bar plot to show average length of questions in topics/questions**

We again used bar plot to display the approximate average length of questions in each selected topic.



***Figure 5-7 Bar plot of average length of questions for each topic***

**6. Methodologies required to complete the analysis**

Natural language processing (NLP) is one of the most promising fields of artificial intelligence that uses natural languages to enable human interactions with machines. There are several exciting Python libraries for NLP, such as Natural Language Toolkit (NLTK), sklearn, etc. Using the above concept from python’s NLTK library, we have built our chatbot Vira. Following are the methodologies used:

1. **Natural Language Toolkit (NLTK)**NLTK has been used to build ‘Vira’ since it is known to be a leading platform for working with human language data. NLTK provides easy-to-use interfaces to resources such as WordNet and text processing libraries for classification, tokenization, stemming, tagging, parsing, etc. In ‘Vira’ we have extensively used tokenization and stemming to pre-process our dataset.
2. **TF-IDF**For evaluating the significance of words in text documents, we have used the statistical method Term Frequency-Inverse Document Frequency (TF-IDF). This gives us the ability to break down the text into a matrix of words and compute their frequencies along with their weightages. This helps Vira understand the context of the text accurately and in turn provide relevant answers to user queries.
3. **Cosine Similarity**

### We have used Cosine Similarity to numerically quantify similarity between two words in our chatbot. Since we are using TF-IDF vectorizer, calculating the dot product provides us the similarity scores. These calculated scores are used by Vira to understand the context appropriately.

### **Tkinter Library for GUI**

We have created an intuitive environment for users to interact with Vira using the tkinter package. Tkinter provides various UI elements, such as labels, buttons, text boxes etc. for building GUI applications. tkinter is one of most efficient ways to create GUI applications in Python.

### **6.1 A brief step by step process of our chatbot building**

**Step 1**: Importing necessary packages explained above, that are needed to build our bot.

**Step 2**: Preprocessing our dataset by Word tokenizing, Removing ASCII values, Removing tags of any kind, Part-of-speech tagging, and Lemmatization

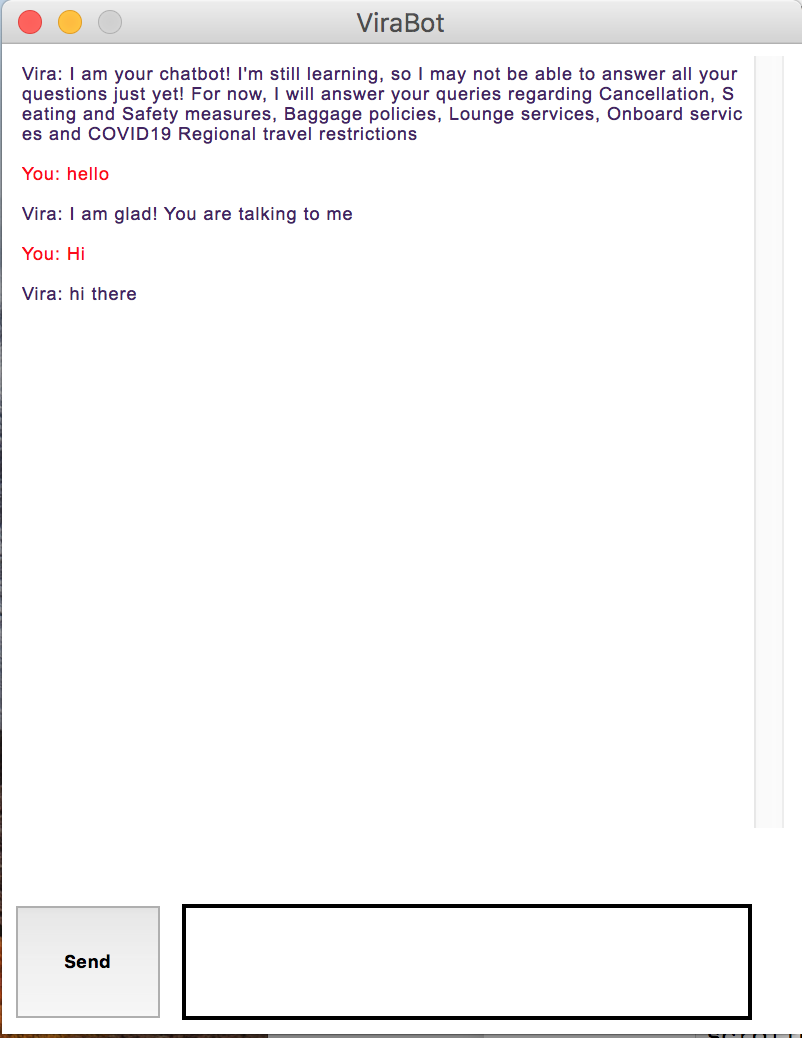
**Step 3** : Defining a function called 'Response' using TF IDF and Cosine Similarity concept to match the words in the user's question to the words in our dataset and give an appropriate response. In case there is no match, then Vira would give something like," I'm sorry I don’t understand you".

**Step 4**: Creating a GUI using Tkinter library that is available in python, and GUI created are like existing chatbots. Initially when GUI opens there will be an introduction for our chatbot. Then the user will ask questions based on pre instructions listed by chatbot. Once the user enters his/her question and enters the 'Send' button, the 'Response' function would be invoked. This function extracts relevant output that would be printed in the GUI chat window.

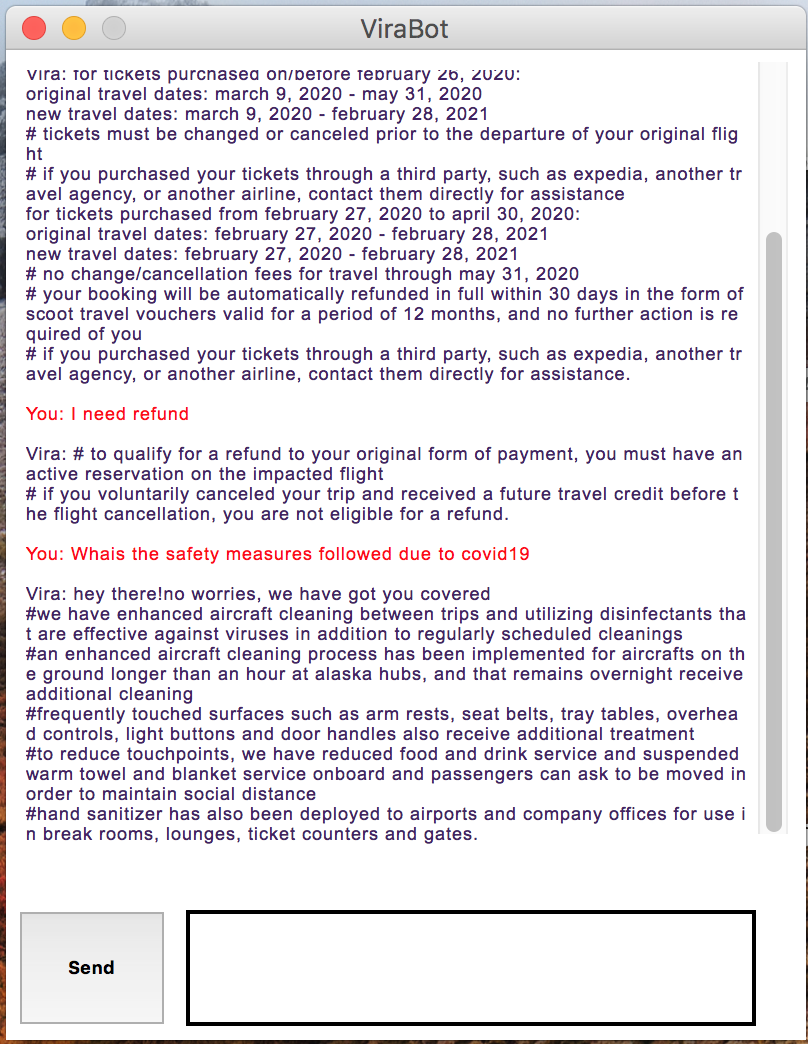
**Step 5** : Process will continue until the user enters the "Bye! take care." line or closes the chat window.

### **7. Analysis result report and explanation**

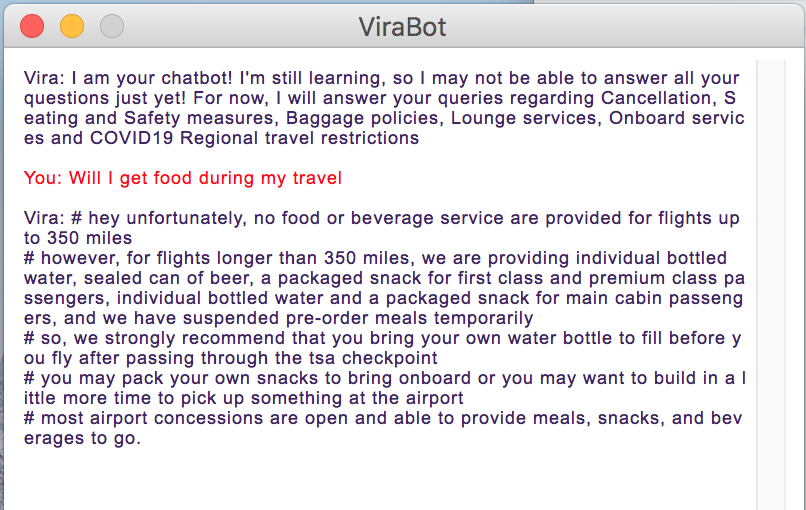
Our chatbot Vira is using tf-idf and cosine similarity concepts to match the words in user’s questions with the words in the dataset to give appropriate answers as explained previously. We have given a few working demo shots of Vira below which would give you a better idea on its working efficiency.

******

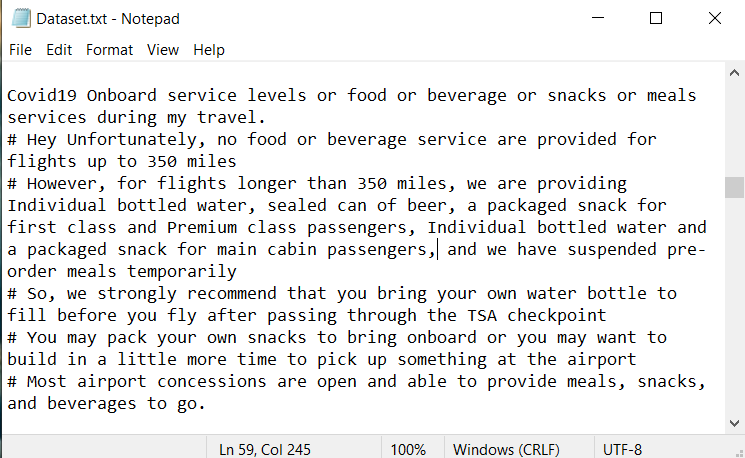
***Figure 7-1 snapshot of Vira’s GUI chat window popping out successfully***



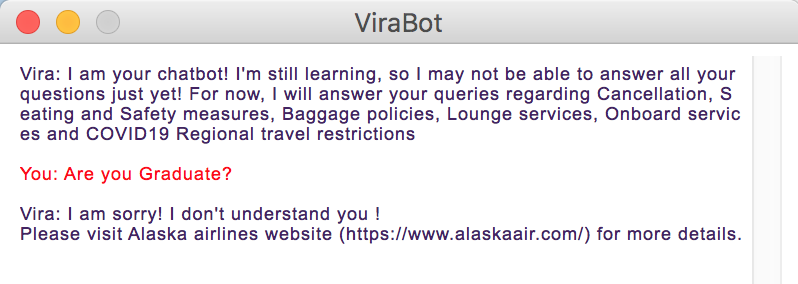
***Figure 7-2 snapshot of Vira answering to questions related to Refund and Covid19 safety measures taken by Alaska Airlines and the answers are is in accordance with our dataset***



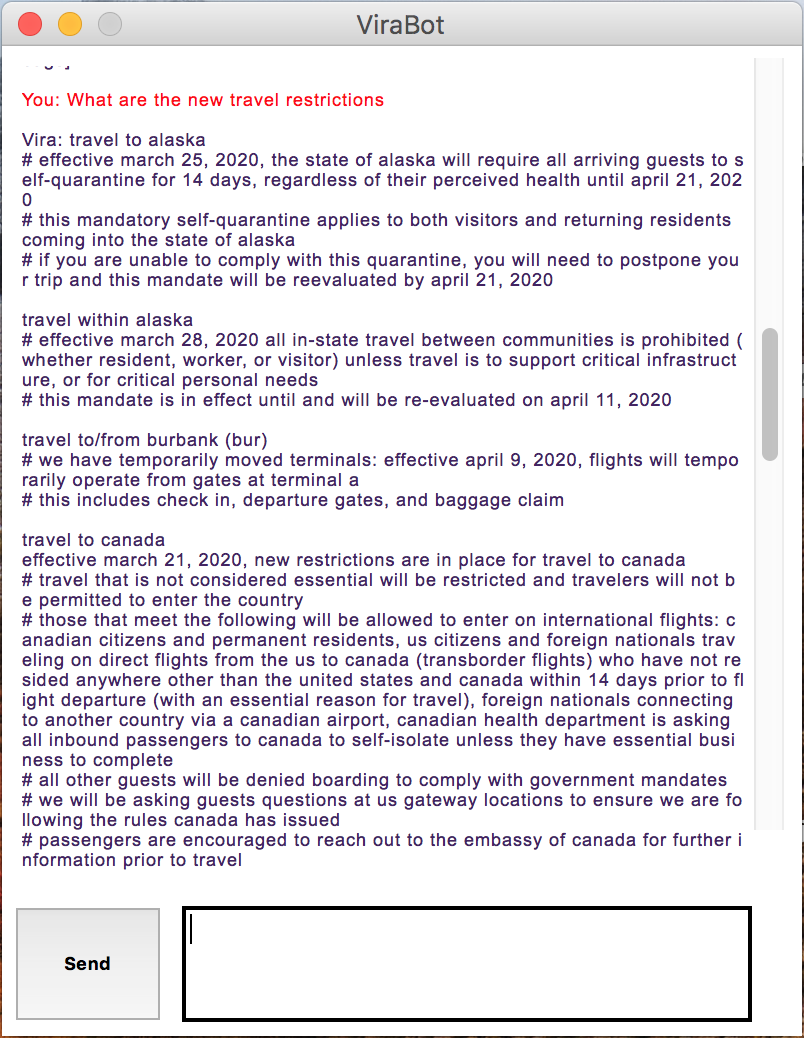
***Figure 7-3 snapshot of Vira answering to questions related to Alaska airlines onboard food service and the answer extracted by Vira is in accordance with our dataset***



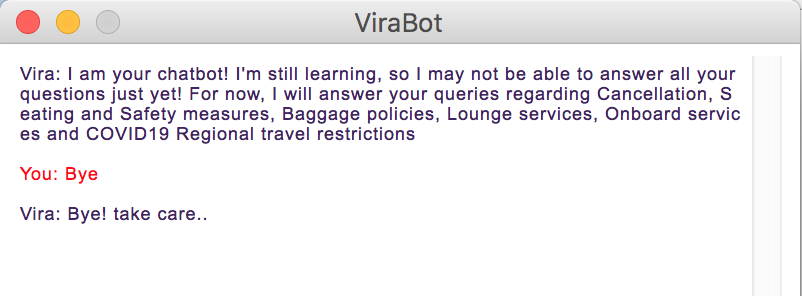
***Figure 7-4 showing a snapshot of the onboard service section from our dataset which was extracted correctly and displayed by Vira to questions related to onboard food service as shown in figure 7-4.***



***Figure 7-5 snapshot of Vira answering cleverly to a totally irrelevant question that was not related to Alaska Airlines and Covid19.***



***Figure 7-6 snapshot of Vira answering to questions related to Covid19 Travel restrictions and again answers are is in accordance with our dataset which was taken from airlines website.***

******

***Figure 7-7 Finally defining the chatbot user conversation handler.***

***Note: The program will exit if you type Bye, shutdown, exit or quit.***

### 

### **8. Conclusion**

### Chatbots are becoming an increasingly important part of digital customer service mix. We strongly believe 'Vira' has potential to become an important part of digital customer service strategy.

### Our model chatbot ‘Vira’ has an accuracy of 84% , so with further training and testing the accuracy can be increased. Also, ‘Vira’ answers now only to covid19 related questions but more sections can be added in future to answer other service-related questions as well.

### On business perspective cost to pilot would be relatively low so ‘Vira’ is cost efficient. Also, time and resources efficient.

### **REFERENCES**

* Natural Language Processing with Python - by Steven Bird, Ewan Klein, and Edward Loper - Published by O’Reilly Media
* Mining Text Data - by Charu Aggarwal
* Newsroom - J.D. Power 2019 North America Airline Satisfaction Study:

<https://newsroom.alaskaair.com/2019-05-29-Alaska-Airlines-captures-number-one-ranking-for-12th-consecutive-year-among-Traditional-Carriers-in-the-J-D-Power-2019-North-America-Airline-Satisfaction-Study>

* Alaska Airlines latest coronavirus travel updates: <https://www.alaskaair.com/>
* Singapore Airlines chatbot: <https://www.singaporeair.com/en_UK/contact-us/>
* Google.com for all day to day references and internet searches
* Dr. Peng Xie for valuable inputs and feedback throughout the project execution