

# **Data Science Intern at Data Glacier**

## **Week 5: Cloud and API deployment**

Name: Chenyu Wang

Batch code: LISUM 13:30

Submission date: Oct 03, 2022

Submitted to: Data Glacier

## Snapshot of each step of deployment:

### 1. Download Café data.

The café dataset contains 925 conversations of questions & answers regarding ordering food, of which 170 questions & 121 answers are unique.

|    | A                               | B   | C       | D          |
|----|---------------------------------|---|---------|------------|
| 1  | Question                        | Answer  | cluster | label      |
| 2  | what is price of Black Coffee   | Its our one of best, you can enjoy it at just CAD. 3.10 |         | 1 3.10 CAD |
| 3  | what is price of Black Coffee   | Its our one of best, you can enjoy it at just CAD. 3.10 |         | 1 3.10 CAD |
| 4  | what is price of Espresso       | Its our one of best, you can enjoy it at just CAD. 3.10 |         | 1 3.10 CAD |
| 5  | what is price of Espresso       | Its our one of best, you can enjoy it at just CAD. 3.10 |         | 1 3.10 CAD |
| 6  | what is price of Chans Espresso | Its our one of best, you can enjoy it at just CAD. 3.10 |         | 1 3.10 CAD |
| 7  | what is price of Chans Espresso | Its our one of best, you can enjoy it at just CAD. 3.10 |         | 1 3.10 CAD |
| 8  | what is price of Con Panna      | Its our one of best, you can enjoy it at just CAD. 3.10 |         | 1 3.10 CAD |
| 9  | what is price of Black Coffee   | Its our one of best, you can enjoy it at just CAD. 3.10 |         | 1 3.10 CAD |
| 10 | what is price of Espresso       | Its our one of best, you can enjoy it at just CAD. 3.10 |         | 1 3.10 CAD |
| 11 | what is price of Black Coffee   | Its our one of best, you can enjoy it at just CAD. 3.10 |         | 1 3.10 CAD |
| 12 | what is price of Black Coffee   | Its our one of best, you can enjoy it at just CAD. 3.10 |         | 1 3.10 CAD |
| 13 | what is price of Espresso       | Its our one of best, you can enjoy it at just CAD. 3.10 |         | 1 3.10 CAD |
| 14 | what is price of Espresso       | Its our one of best, you can enjoy it at just CAD. 3.10 |         | 1 3.10 CAD |
| 15 | what is price of Chans Espresso | Its our one of best, you can enjoy it at just CAD. 3.10 |         | 1 3.10 CAD |
| 16 | what is price of Chans Espresso | Its our one of best, you can enjoy it at just CAD. 3.10 |         | 1 3.10 CAD |
| 17 | what is price of Con Panna      | Its our one of best, you can enjoy it at just CAD. 3.10 |         | 1 3.10 CAD |
| 18 | what is price of Black Coffee   | Its our one of best, you can enjoy it at just CAD. 3.10 |         | 1 3.10 CAD |
| 19 | what is price of Espresso       | Its our one of best, you can enjoy it at just CAD. 3.10 |         | 1 3.10 CAD |
| 20 | how much is Black Coffee        | Its our one of best, you can enjoy it at just CAD. 3.10 |         | 1 3.10 CAD |
| 21 | how much is Black Coffee        | Its our one of best, you can enjoy it at just CAD. 3.10 |         | 1 3.10 CAD |
| 22 | how much is Espresso            | Its our one of best, you can enjoy it at just CAD. 3.10 |         | 1 3.10 CAD |
| 23 | how much is Espresso            | Its our one of best, you can enjoy it at just CAD. 3.10 |         | 1 3.10 CAD |
| 24 | how much is Chans Espresso      | Its our one of best, you can enjoy it at just CAD. 3.10 |         | 1 3.10 CAD |
| 25 | how much is Chans Espresso      | Its our one of best, you can enjoy it at just CAD. 3.10 |         | 1 3.10 CAD |

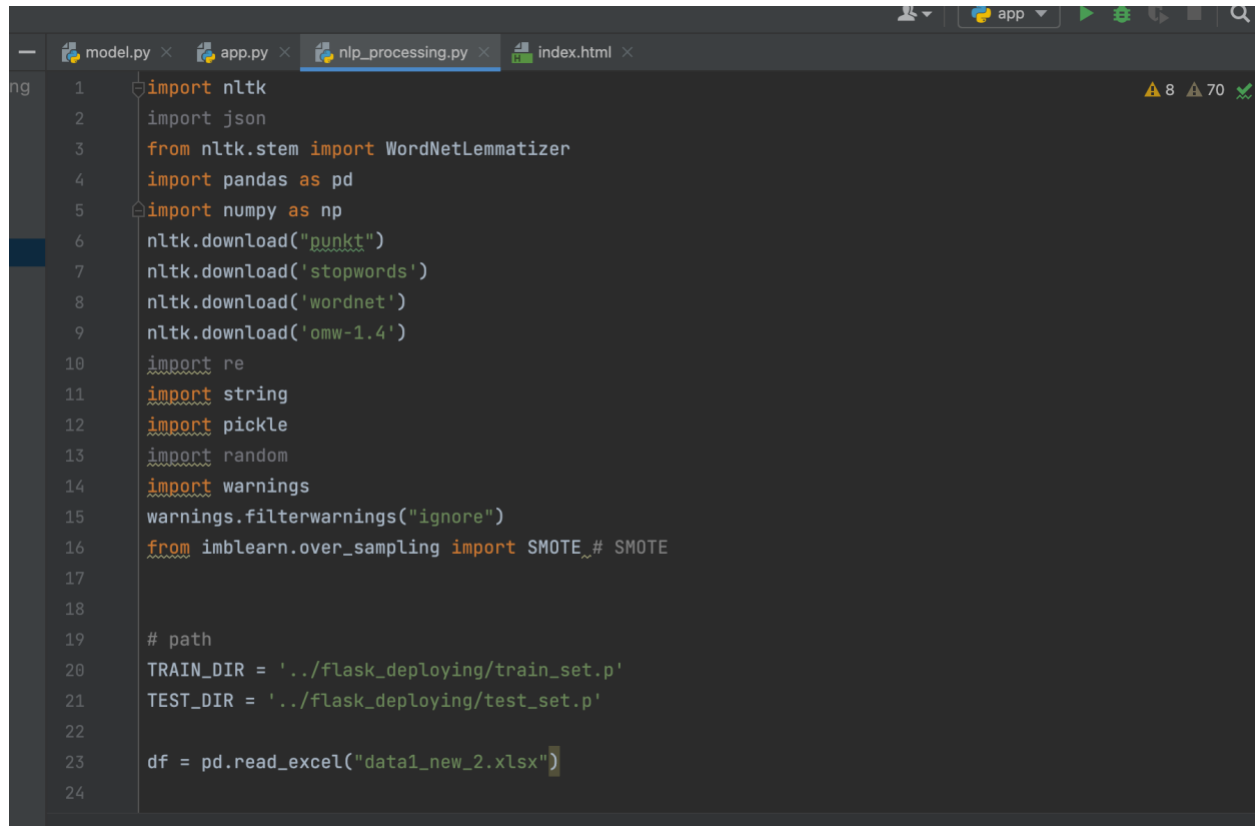
### 2. Create HTML index page which contains the input texts of question in the cafe data.

```
flask_deploying - index.html

1 <!DOCTYPE html>
2 <html>
3 <head>
4   <meta charset="UTF-8">
5   <title>ML API</title>
6   <link href='https://fonts.googleapis.com/css?family=Pacifico' rel='stylesheet' type='text/css'>
7   <link href='https://fonts.googleapis.com/css?family=Arimo' rel='stylesheet' type='text/css'>
8   <link href='https://fonts.googleapis.com/css?family=Hind:300' rel='stylesheet' type='text/css'>
9   <link href='https://fonts.googleapis.com/css?family=Open+Sans+Condensed:300' rel='stylesheet' type='text/css'>
10  <link rel="stylesheet" href="{{ url_for('static', filename='css/style.css') }}">
11
12 </head>
13
14 <body>
15   <div class="login">
16     <h1>Coffee ChatBot</h1>
17
18     <!-- Main Input For Receiving Query to our ML -->
19     <form action="{{ url_for('predict')}}" method="post">
20       <input type="text" name="Please write your question" placeholder="Please write your question" required
21
22       <button type="submit" class="btn btn-primary btn-block btn-large">Submit</button>
23     </form>
24   </div>
25 </body>
26 </html>
```

### 3. Create the nlp\_processing.py

In this file, the nltk package in python was used to clean the text dataset. The stopwords and punctuation from the texts were removed. The stemming was executed and making words lowercase for the final steps of the data cleaning process. The Bag of Words was performed for feature engineering steps to convert textual data to numeric. Finally, the training set and test set were saved as train\_set.p and test\_set.p



```
1 import nltk
2 import json
3 from nltk.stem import WordNetLemmatizer
4 import pandas as pd
5 import numpy as np
6 nltk.download("punkt")
7 nltk.download('stopwords')
8 nltk.download('wordnet')
9 nltk.download('omw-1.4')
10 import re
11 import string
12 import pickle
13 import random
14 import warnings
15 warnings.filterwarnings("ignore")
16 from imblearn.over_sampling import SMOTE # SMOTE
17
18
19 # path
20 TRAIN_DIR = '../flask_deploying/train_set.p'
21 TEST_DIR = '../flask_deploying/test_set.p'
22
23 df = pd.read_excel("data1_new_2.xlsx")
24
```

```

82
83 def main():
84     df = pd.read_excel("data1_new_2.xlsx")
85
86     lemmatizer = WordNetLemmatizer()
87
88     data = df['Question']
89     data_X = df['Question'].values.tolist()
90     y = df['cluster']
91     data_y = df['cluster'].values.tolist()
92     classes = df['cluster'].drop_duplicates().values.tolist()
93     classes = sorted(set(classes))
94
95     words = get_word(df)
96     train, test = get_train_test(data_X, words)
97
98     sm = SMOTE(random_state=777, k_neighbors=3)
99     train_sm, test_sm = sm.fit_resample(train, test)
100
101     pickle.dump(train_sm, open(TRAIN_DIR, 'wb'))
102     pickle.dump(test_sm, open(TEST_DIR, 'wb'))
103

```

#### 4. Create model.py

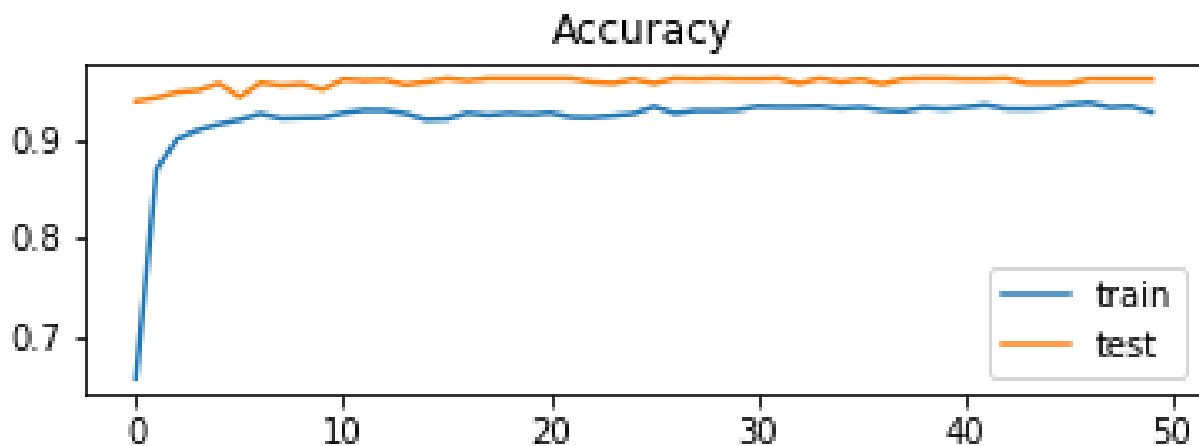
The Deep Neural Network (DNN) model was used for carrying out the accurate processing of the input user query to generate the most appropriate response. The first layer in our Neural Network consists of 128 neurons. The input shape of the first layer is 160 which is the value of the `train_X.shape[1]`. The second layer is the hidden layer which consists of 64 neurons. The relu activation function is used for these two layers. For preventing the overfitting problem, dropout, one of the regularization techniques, was used. Finally, the Softmax activation function is used for the output layer, which comprises 21 neurons that correspond to the 21 intent classes.

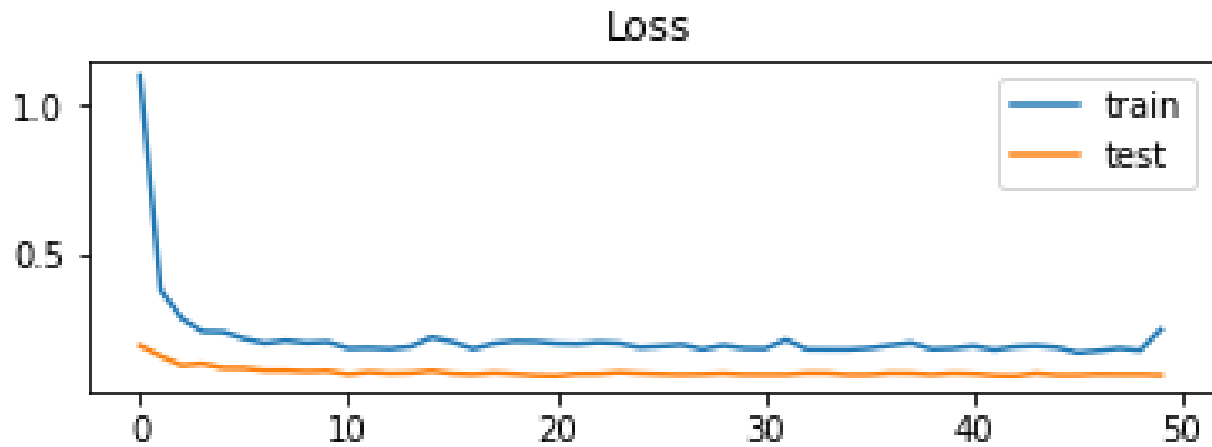
Model: "sequential\_2"

| Layer (type)        | Output Shape | Param # |
|---------------------|--------------|---------|
| dense_6 (Dense)     | (None, 128)  | 20608   |
| dropout_4 (Dropout) | (None, 128)  | 0       |
| dense_7 (Dense)     | (None, 64)   | 8256    |
| dropout_5 (Dropout) | (None, 64)   | 0       |
| dense_8 (Dense)     | (None, 21)   | 1365    |

=====  
Total params: 30,229  
Trainable params: 30,229  
Non-trainable params: 0

The loss and accuracy were used to evaluate the model when tuning hyperparameters during the training process. The accuracy of the Deep Neural Network (DNN) model on the training set and test sets are both 0.96.



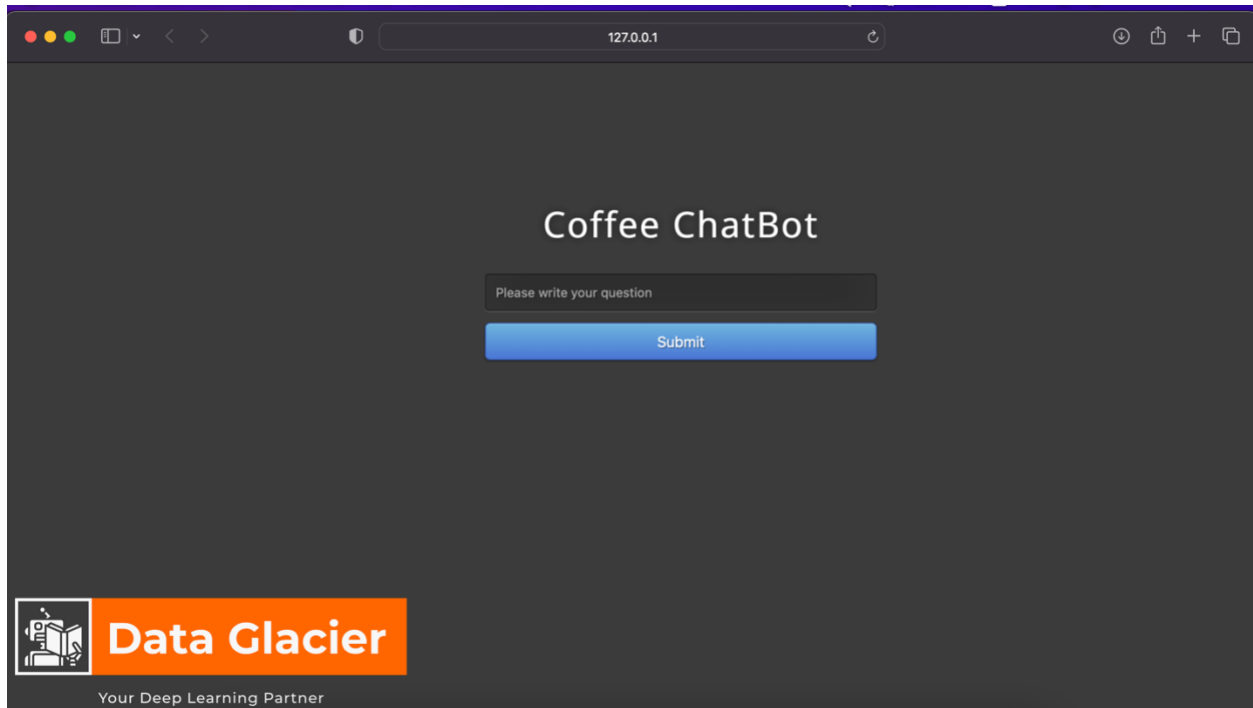


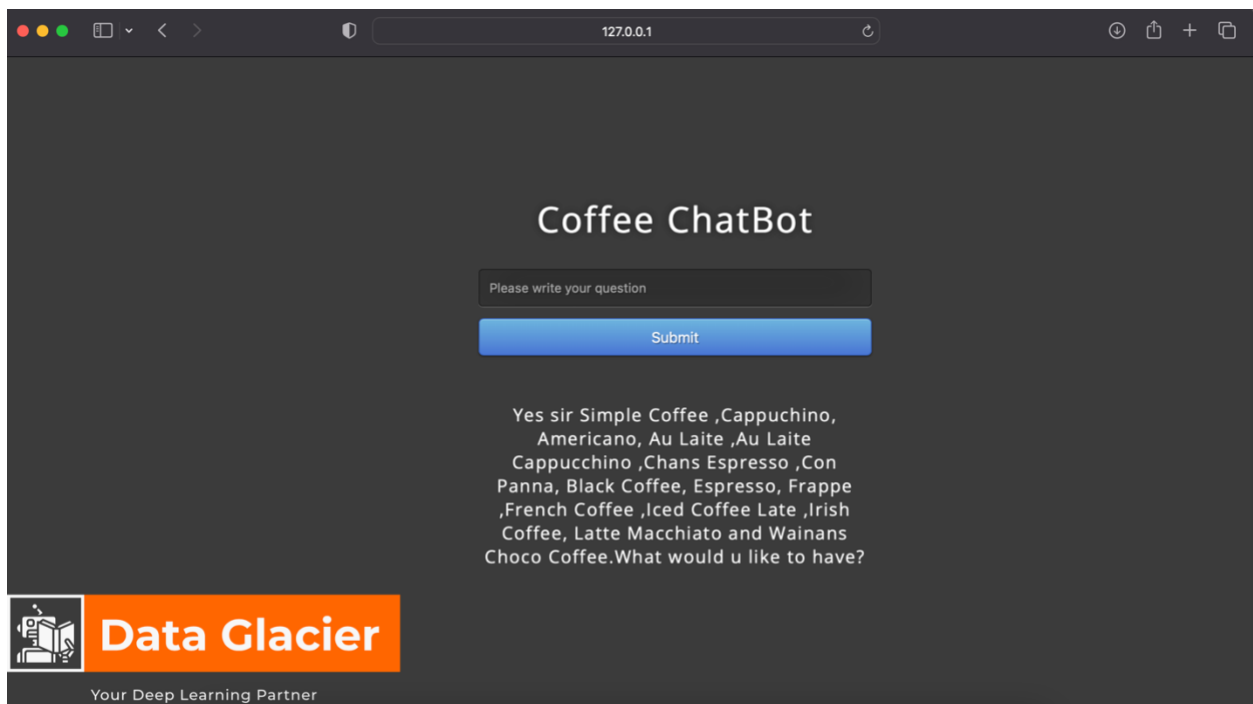
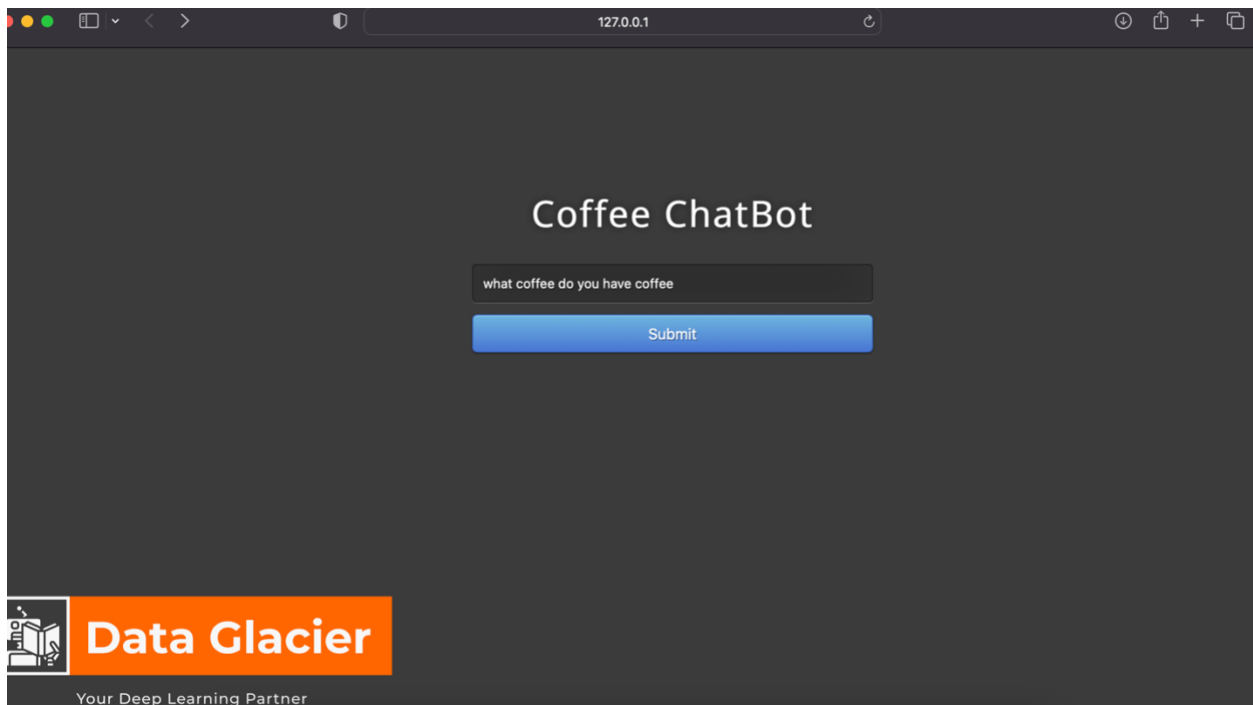
Finally, the trained model was saved as model.pkl

```
31 def create_model(train_X, train_Y):
32     model = Sequential()
33     model.add(Dense(128, input_shape=(train_X.shape[1],), activation='relu'))
34     model.add(Dropout(0.5))
35
36     model.add(Dense(64, activation='relu'))
37     model.add(Dropout(0.5))
38
39     model.add(Dense(train_Y.shape[1], activation='softmax'))
40     adam = Adam(learning_rate=0.01, decay=1e-6)
41     model.compile(loss='categorical_crossentropy', optimizer=adam, metrics=['accuracy'])
42
43     return model
44
45 def main():
46     train_sm = pickle.load(open(TRAIN_DIR, 'rb'))
47     test_sm = pickle.load(open(TEST_DIR, 'rb'))
48
49     train_X, test_X, train_Y, test_y = train_test_split(train_sm, test_sm, test_size=0.2, random_state=42)
50
51     model = create_model(train_X, train_Y)
52     model.fit(train_X, train_Y, validation_data=(test_X, test_y), batch_size=64, epochs=50, verbose=1)
53
54     pickle.dump(model, open(MODEL_DIR, 'wb'))
```

## 5. Create app.py

The flask app was created and saved pickle model was load. The question was received from the web and respond the answer to customer's question.







## 6. Deploy model on Heroku

I link my GitHub repository to my Heroku account for deploying the Coffee ChatBot. Second, requirements.txt, nltk.txt, and runtime.txt were created. These text files containing the python packages, nltk corpuses, and python version required to execute the application.


### Steps for Model Deployment on Heroku

#### 6.1 Create new app

Create New App

App name

Choose a region

 United States

⌵


Add to pipeline...

Create app

#### 6.2 Connect to GitHub repository where code is I uploaded

Personal ⌵ > coffee-chatbot-api

☆ Open app More ⌵

GitHub  chenyu-wang55/deploy\_heroku

Overview

Resources

Deploy

Metrics

Activity


Access

Settings

Add this app to a pipeline


Create a new pipeline or choose an existing one and add this app to a stage in it.

Add this app to a stage in a pipeline to enable additional features



Pipelines let you connect multiple apps together and **promote code** between them.

[Learn more](#)




Pipelines connected to GitHub can enable **review apps**, and create apps for new pull requests.


[Learn more](#)


Choose a pipeline

⌵

Deployment method

 Heroku Git  
Use Heroku CLI

 GitHub  
Connected

 Container Registry  
Use Heroku CLI

## 6.3 Deploy branch

## 6.4 view

### Manual deploy

Deploy the current state of a branch to this app.

### Deploy a GitHub branch

This will deploy the current state of the branch you specify below. [Learn more.](#)

Choose a branch to deploy

main

Deploy Branch

Receive code from GitHub



Build main 3cd4a176



Release phase



Deploy to Heroku

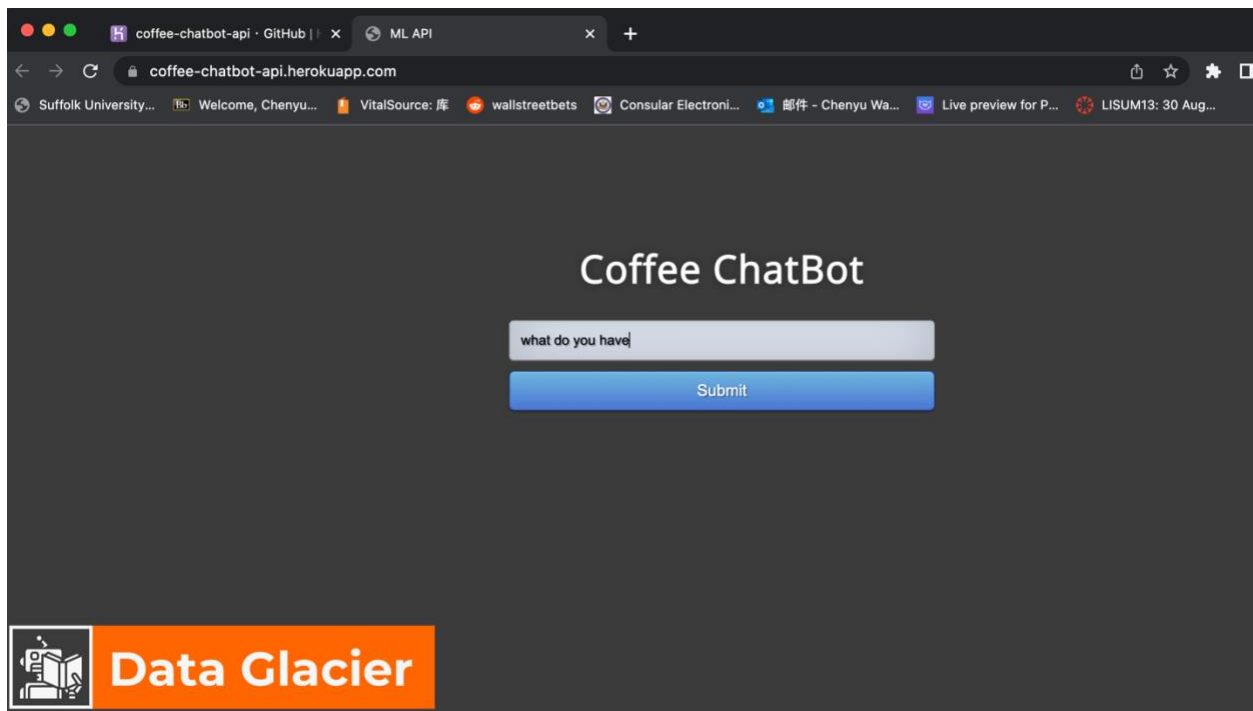


Your app was successfully deployed.

View

Link <https://coffee-chatbot-api.herokuapp.com/>

### Result display



coffee-chatbot-api · GitHub | x

ML API

Heroku Python Support | Heroku | x

+

← → ↻ coffee-chatbot-api.herokuapp.com/predict


Suffolk University... Welcome, Chenyu... VitalSource: 库 wallstreetbets Consular Electroni... 邮件 - Chenyu Wa... Live preview for P... LISUM13: 3

# Coffee ChatBot

Please write your question

Submit

Yes sir Simple Coffee ,Cappuchino,  
Americano, Au Laite ,Au Laite  
Cappuchino ,Chans Espresso ,Con  
Panna, Black Coffee, Espresso, Frappe  
,French Coffee ,Iced Coffee Late ,Irish  
Coffee, Latte Macchiato and Wainans  
Choco Coffee.What would u like to have?

 **Data Glacier**

Your Deep Learning Partner