Data Science Intern at Data Glacier
Week 4: Deployment on Flask

Name: Chenyu Wang

Batch code: LISUM 13:30

Submission date: Sep 28, 2022

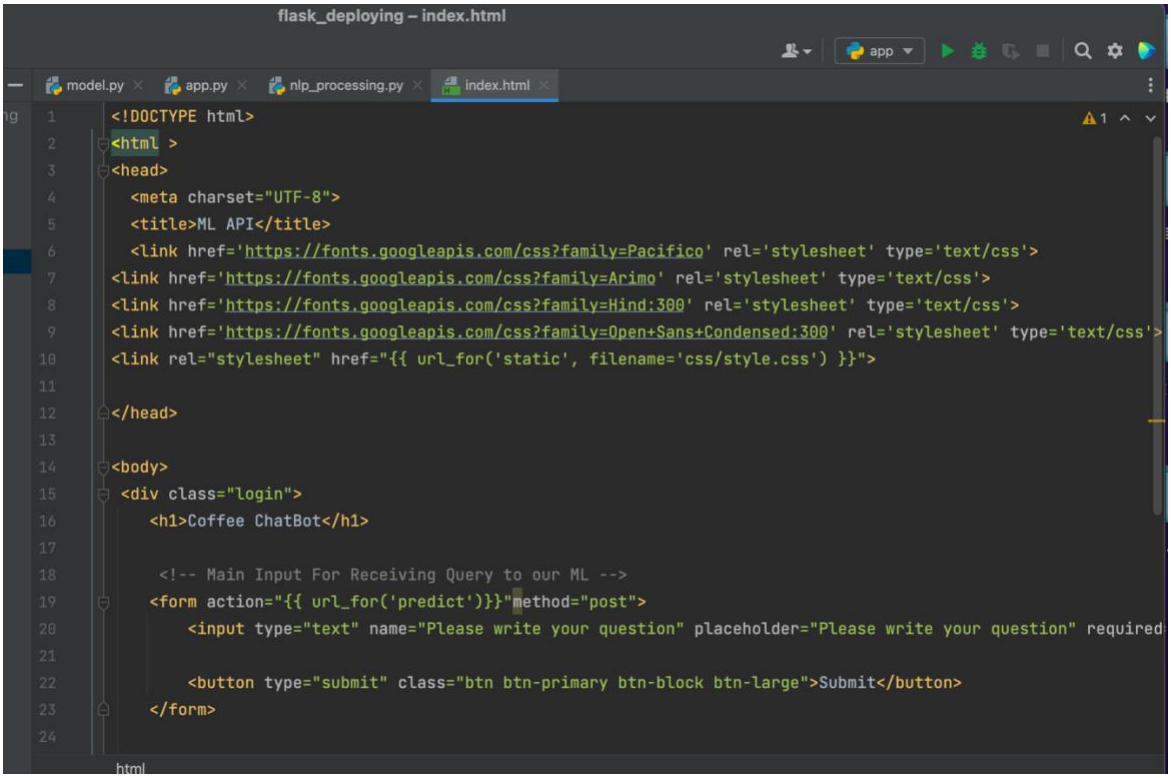
Submitted to: Data Glacier

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.

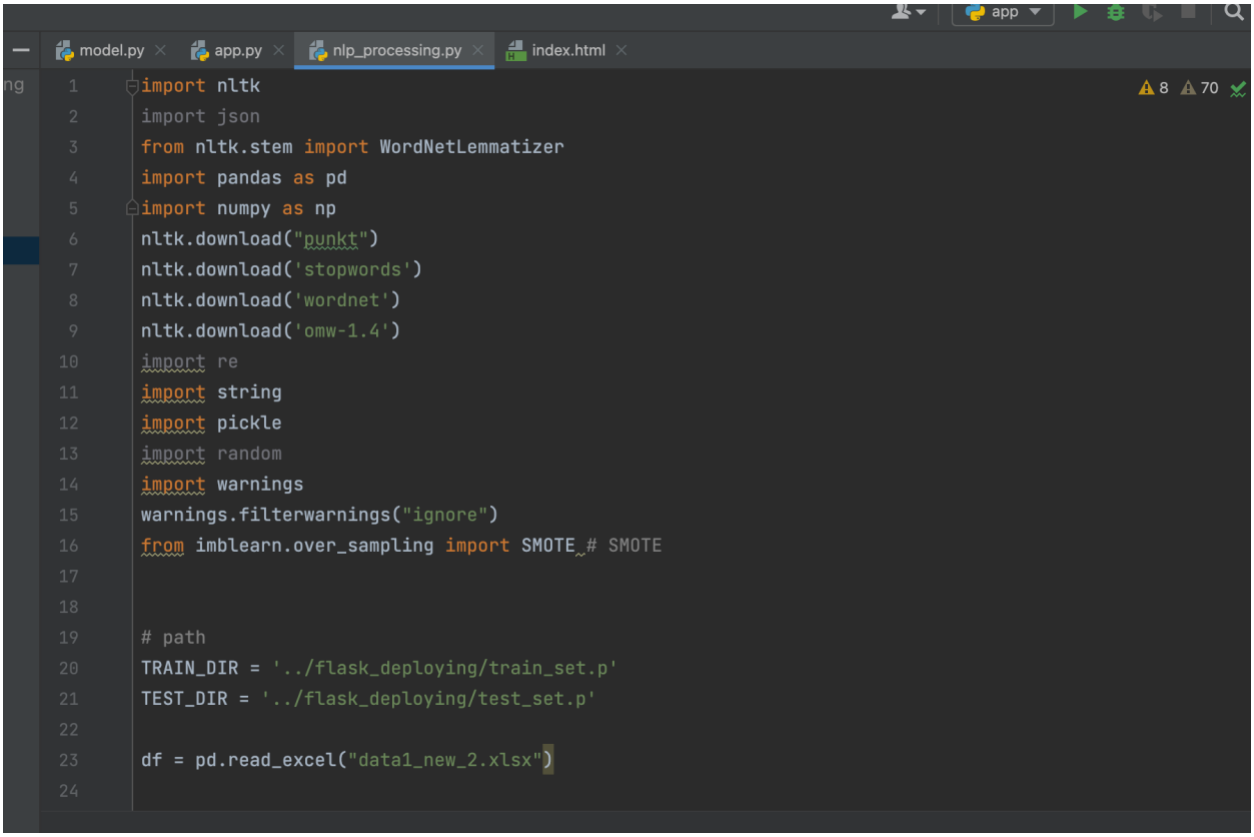
	A	B	C	D
1	Question	Answer	cluster	label
2	what is price of Black Coffee	Its our one of best, you can enjoy it at just CAD. 3.10	1	3.10 CAD
3	what is price of Black Coffee	Its our one of best, you can enjoy it at just CAD. 3.10	1	3.10 CAD
4	what is price of Espresso	Its our one of best, you can enjoy 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.**



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



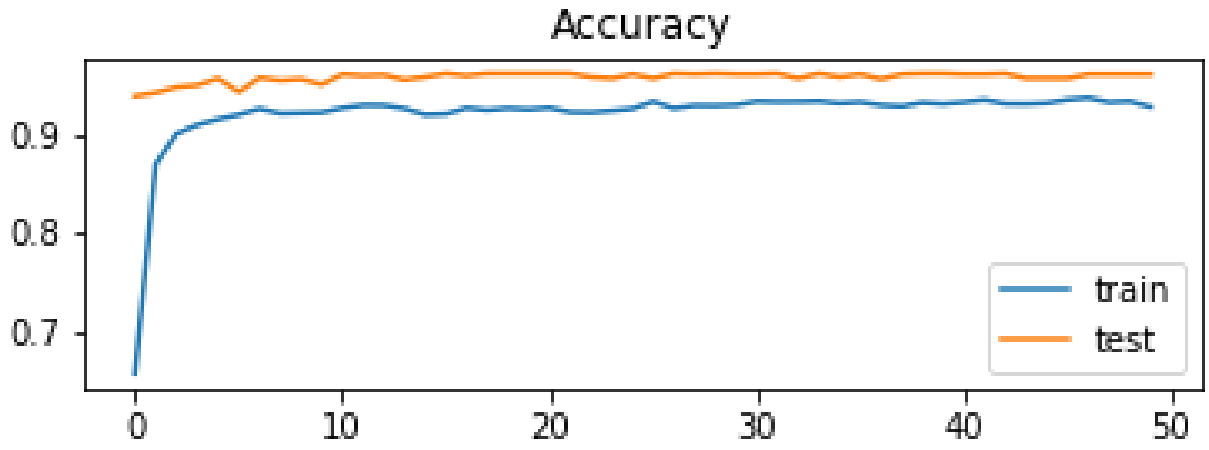
```
82
83  def main():
84      df = pd.read_excel("data1_new_2.xlsx")
85
86      lemmatizer = WordNetLemmatizer()
87
88      data = df['Question']
89      data_X = df['Question'].values.tolist()
90      y = df['cluster']
91      data_y = df['cluster'].values.tolist()
92      classes = df['cluster'].drop_duplicates().values.tolist()
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
```

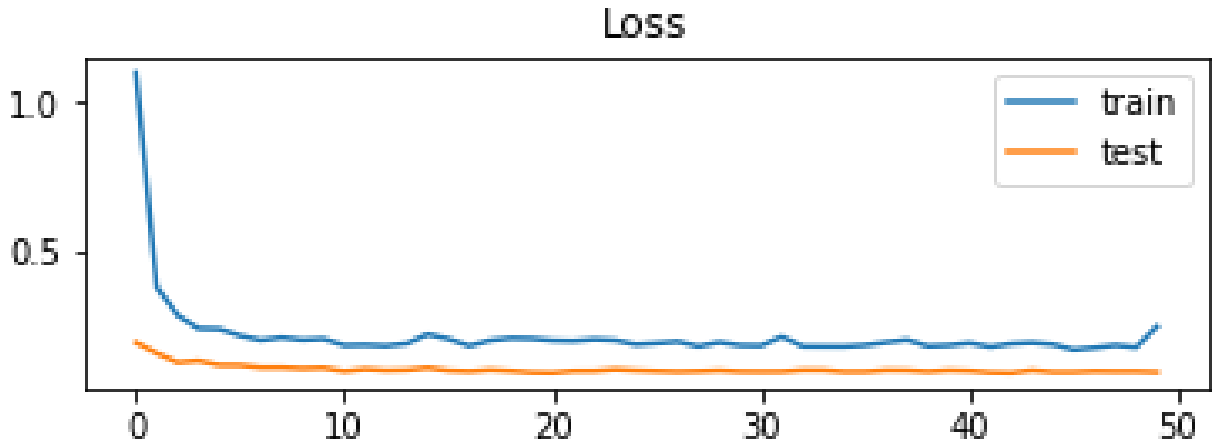
4. Create model.py

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.



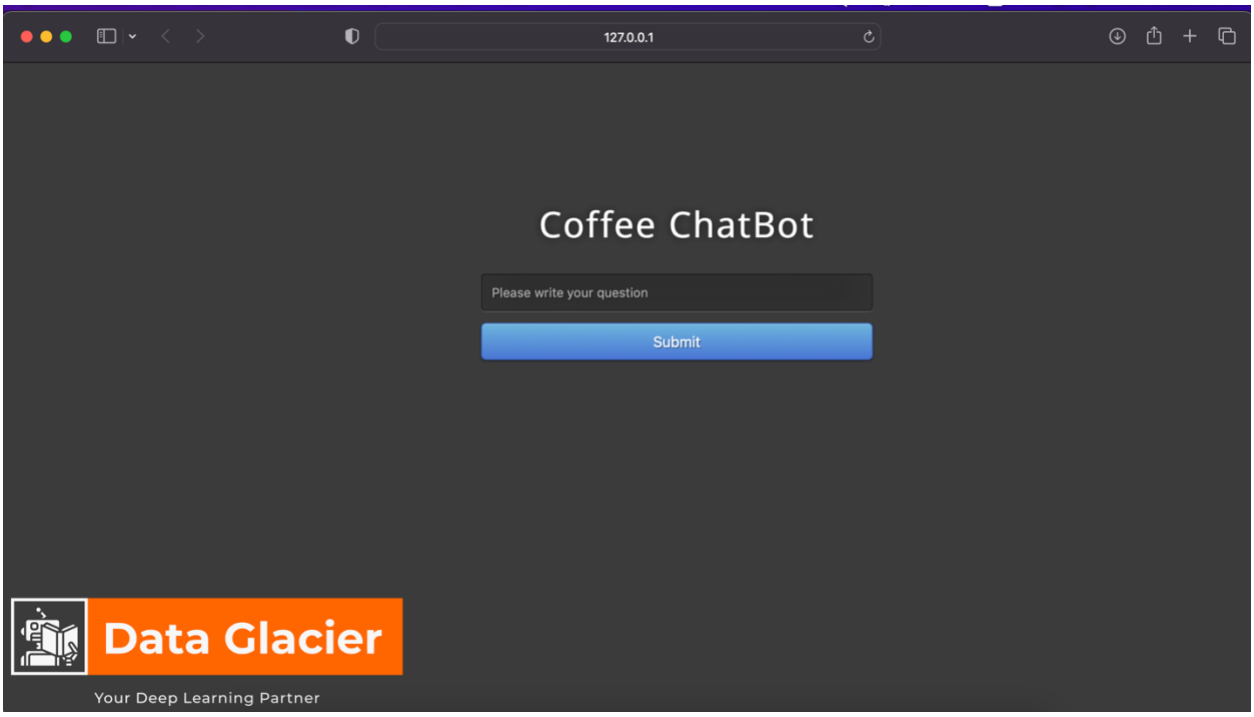


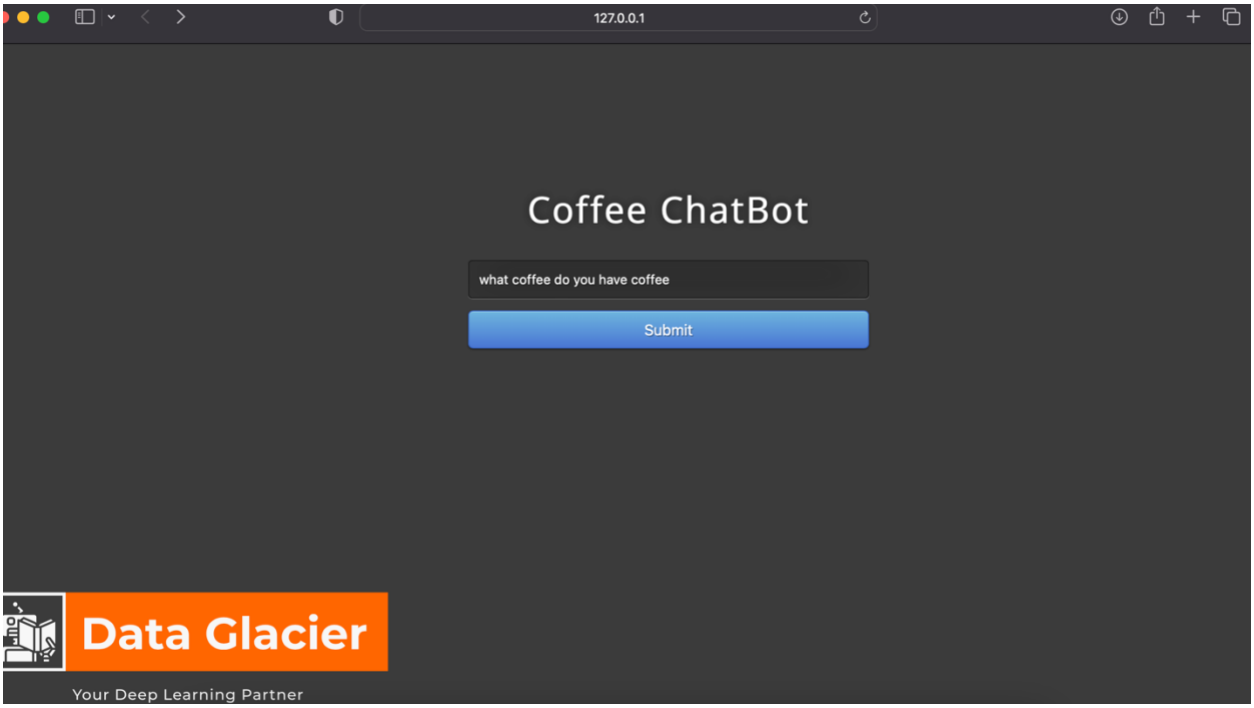
Finally, the trained model was saved as model.pkl

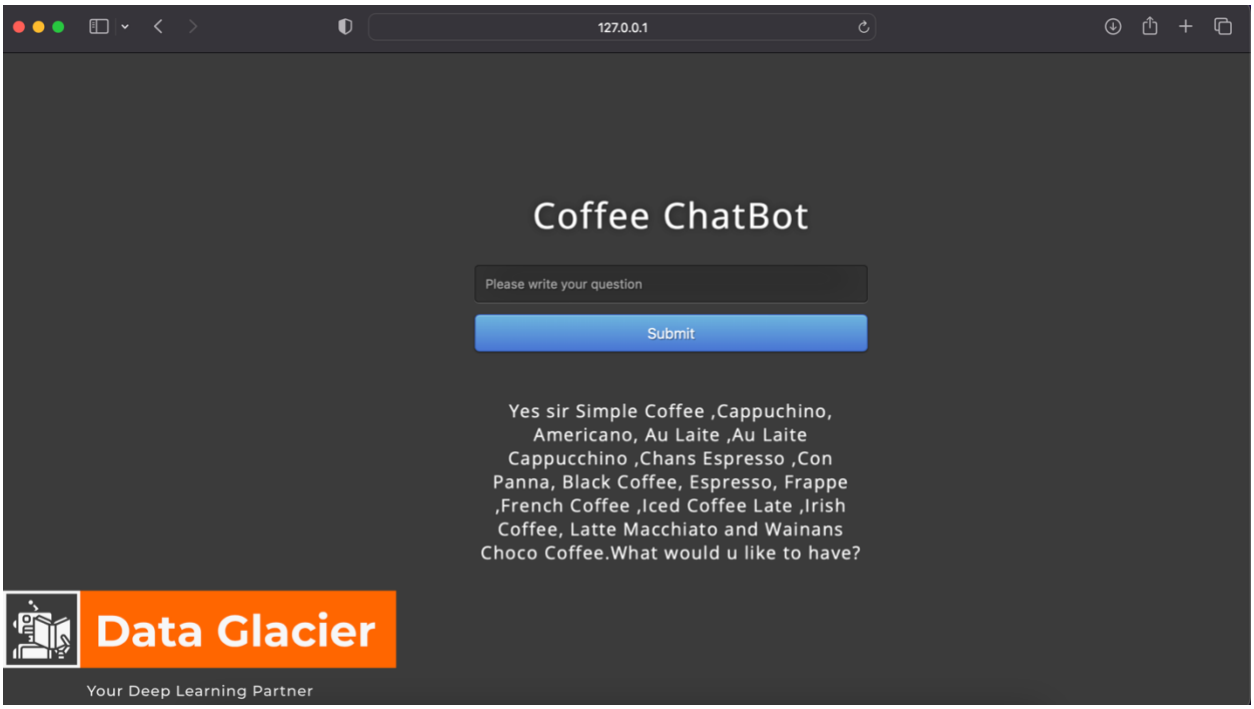
```python
def create_model(train_X,train_Y):
    model = Sequential()
    model.add(Dense(128, input_shape=(train_X.shape[1],), activation="relu"))
    model.add(Dropout(0.5))

    model.add(Dense(64,activation="relu"))
    model.add(Dropout(0.5))

    model.add(Dense(train_Y.shape[1],activation="softmax"))
    adam = Adam(learning_rate=0.01,decay=1e-6)
    model.compile(loss='categorical_crossentropy',optimizer=adam, metrics=["accuracy"])

    return model

def main():
    train_sm = pickle.load(open(TRAIN_DIR, 'rb'))
    test_sm = pickle.load(open(TEST_DIR, 'rb'))

    train_X, test_X, train_Y, test_y = train_test_split(train_sm, test_sm, test_size=0.2, random_state=42)

    model = create_model(train_X,train_Y)
    model.fit(train_X, train_Y, validation_data=(test_X, test_y), batch_size=64, epochs=50, verbose=1)

    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.




127.0.0.1
Coffee ChatBot
what coffee do you have coffee
Submit
Data Glacier
Your Deep Learning Partner


127.0.0.1
Coffee ChatBot
Please write your question
Submit
Yes sir Simple Coffee ,Cappuchino, Americano, Au Laite ,Au Laite Cappucchino ,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