Custom chatbot predictor for diabetes stages

PROJECT MEMBER DETAILS:

Name : Abhiram Reddy Komma
ID: 700740172
Email : axk01720@ucmo.edu

Responsibilities and Contribution:

- I had myself contributed 100% to the project from proposal, implementation and to execution
- Using Diabetes symptoms dataset, initially pre processed the dataset and trained three classifiers namely, SVM (Support Vector Machine), K means Clustering, RF (Random Forest)
- Implemented python backend server using Flask
- Implemented chatbot endpoints to send and receive messages
- Did frontend chatbot page in HTML and integrated it with python backend using endpoints via API
- Performed Natural Language Processing and Named Entity Recognition features for the received input messages from the user using Spacy and SciSpacy modules
- Gathered classifier input features from user messages and predicted the diabetes stages

Motivation:

- Diabetes may not seem harmful initially, but as time goes on, it will affect vital organs in body from heart, vision, kidneys,etc.
- Early stage diabetes prediction is crucial, as proper diagnosis can be provided and in some cases, the diabetes can be reversed without taking any proper medications.
- Currently smart chatbots are the new way of interaction between users and websites
- Currently, there are no smart chat auto assistant which are integrated with machine learning techniques and provide automatic replies
- Chatbot can be easily integrated into any website as plugin which will act as customer support enabled with machine learning assistant

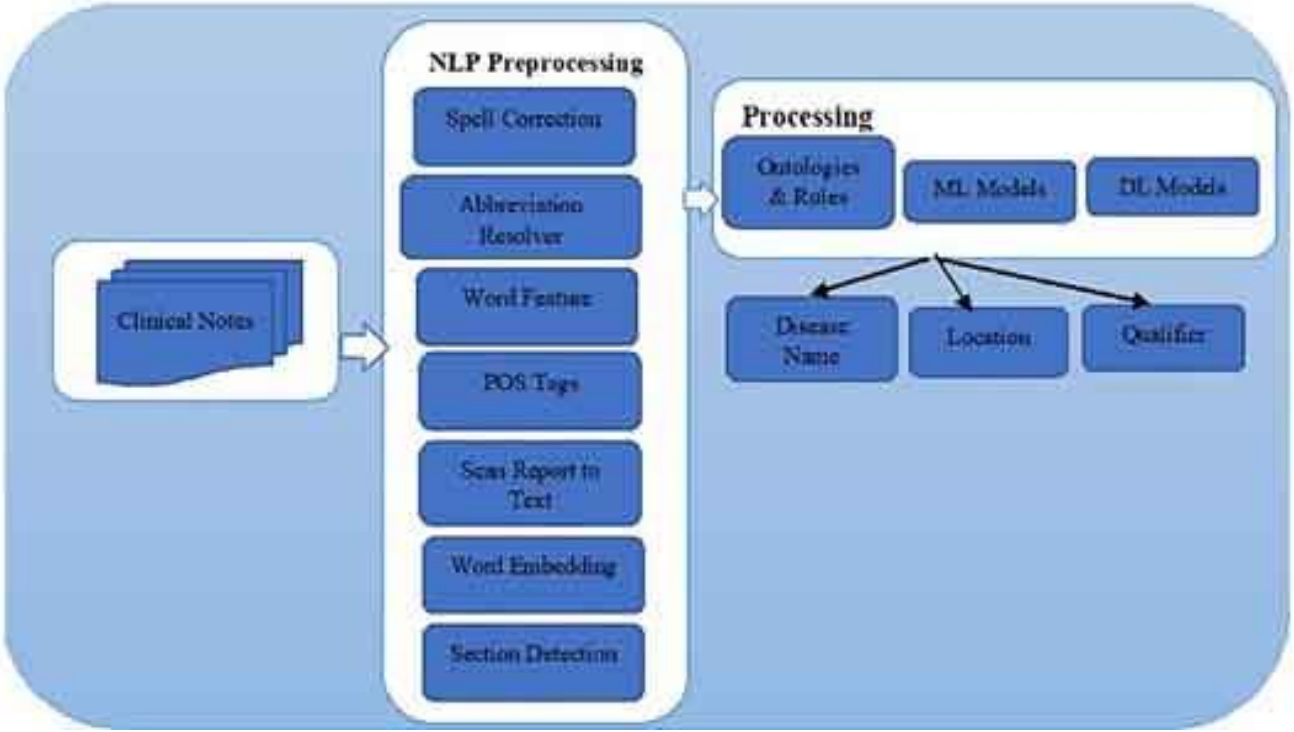
Objectives:

Chatbot should detect entities from user message and provide them to classifier as input features for the classification of diabetes stages

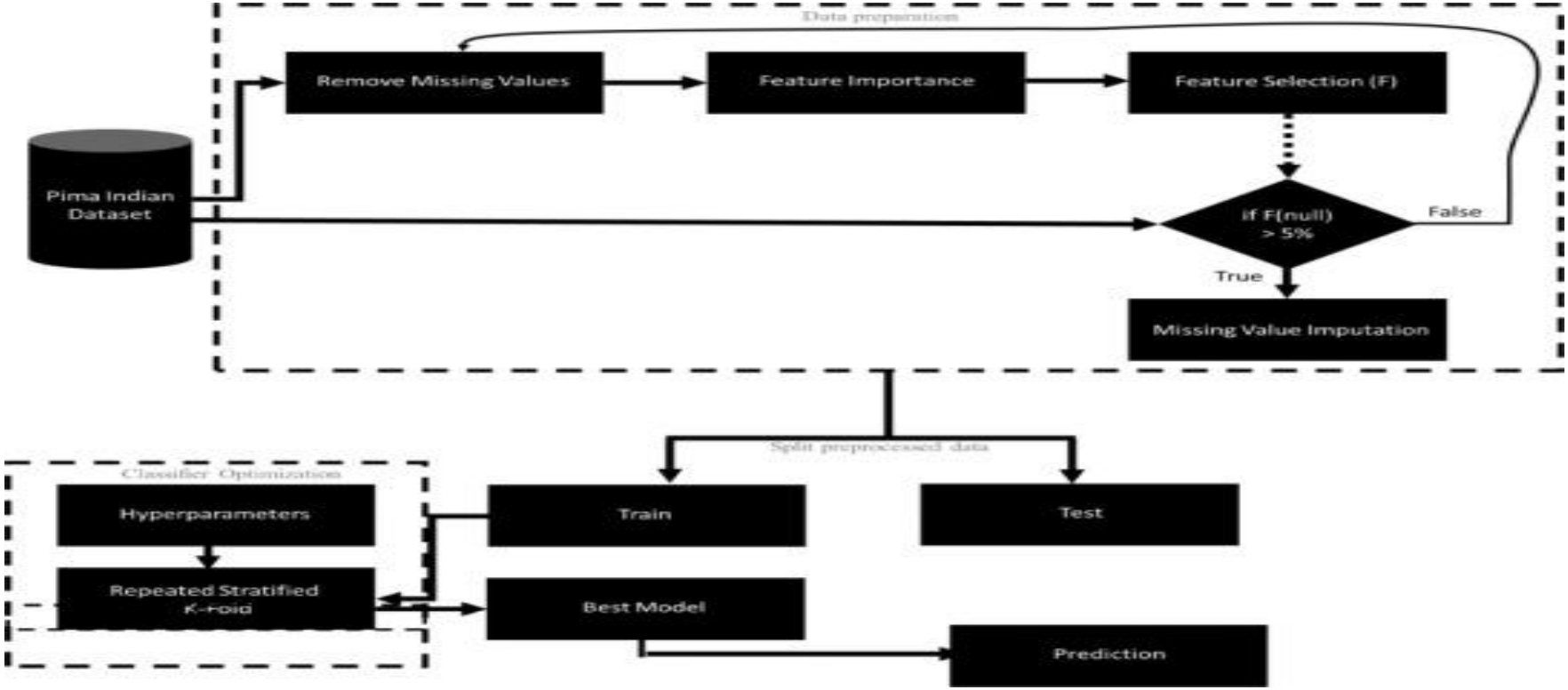
Input can be of two types,

- user message will consists of all symptoms related to diabetes and they will be given as input features to the classifier and the diabetes stage will be predicted.
- User message will be having only one or two symptoms, in this case algorithm will prompt questions related to input features using chatbot replies and gather respective inputs from user and predict diabetes stages.

Related Work: NLP for bio medical terms



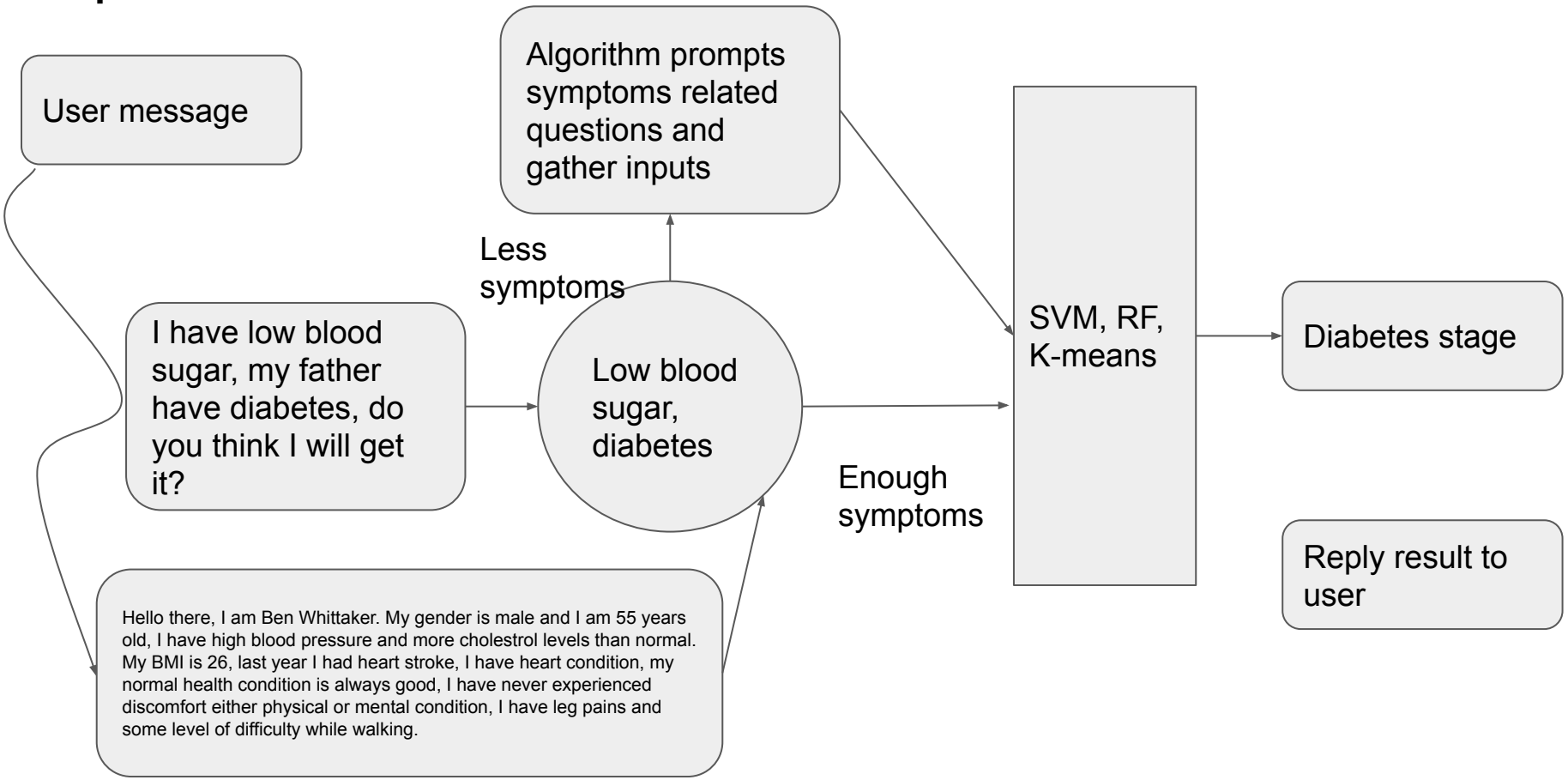
Related Work: Diabetes stages prediction from PIMA dataset

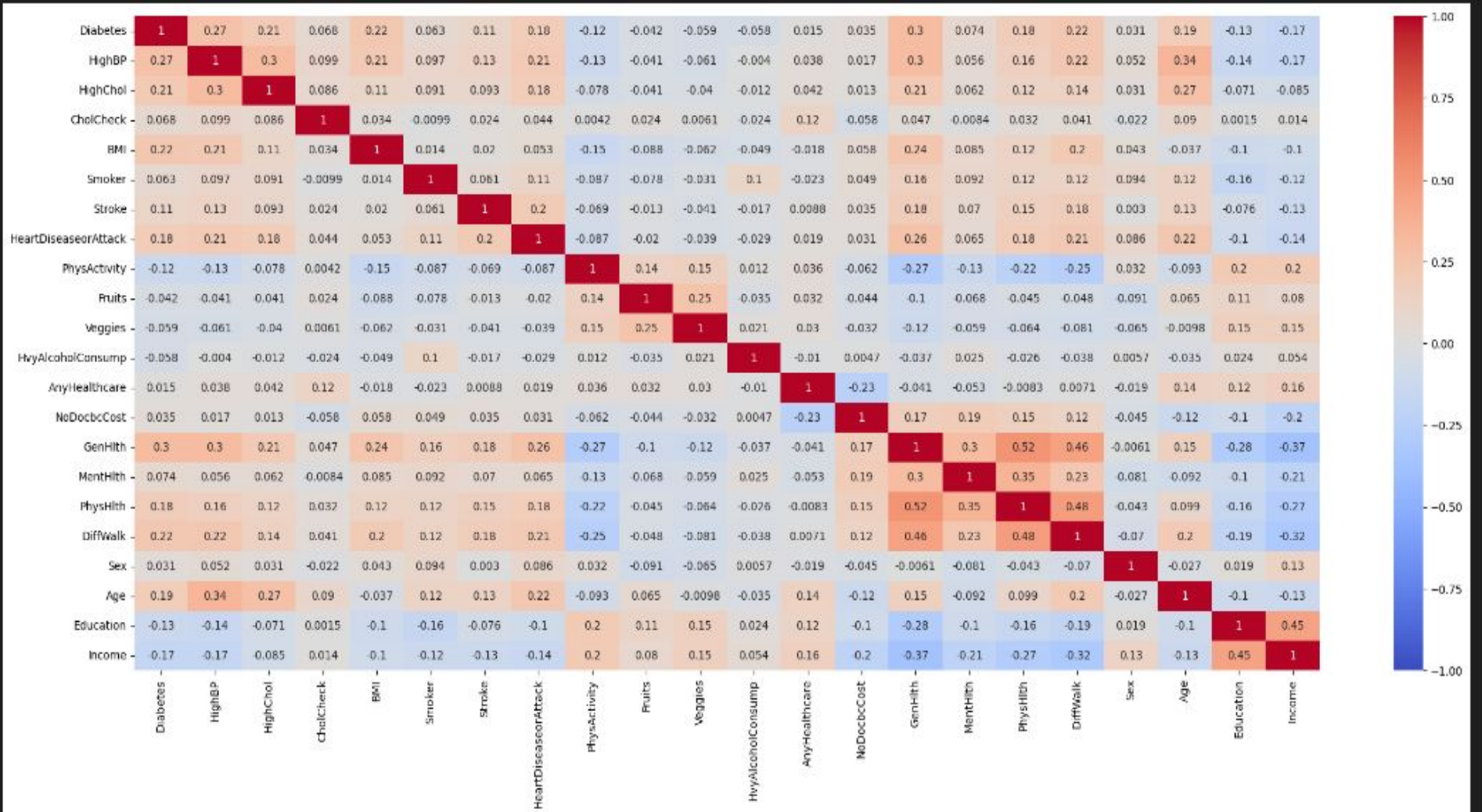


Problem Statement:

Currently there are no chatbot predictors which will gather user inputs and perform classification of diabetes stages

Proposed Solution:



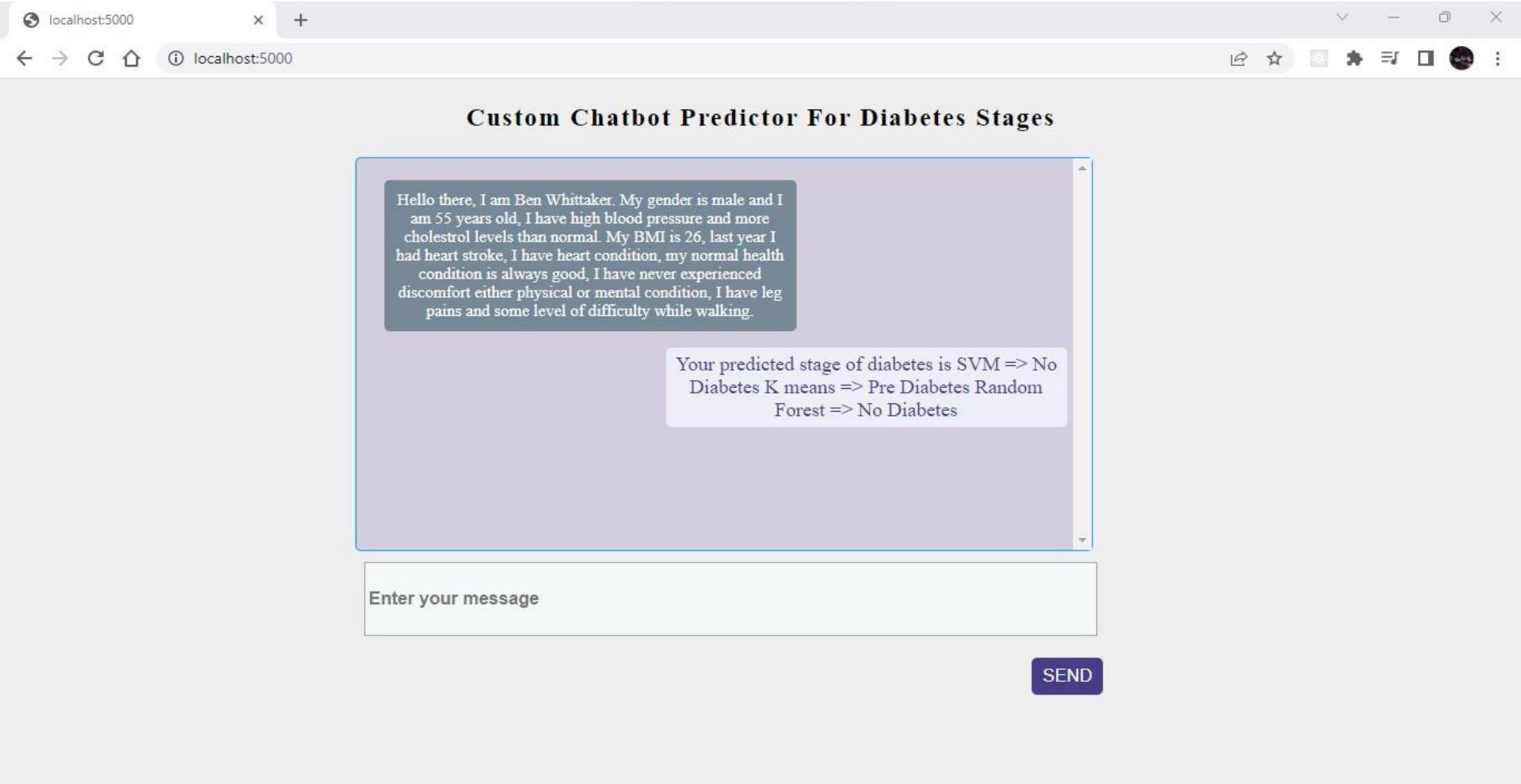

Diabetes 1 0.27 0.21 0.068 0.22 0.063 0.11 0.18 -0.12 -0.042 -0.059 -0.058 0.015 0.035 0.3 0.074 0.18 0.22 0.031 0.19 -0.13 -0.17
HighBP 0.27 1 0.3 0.099 0.21 0.097 0.13 0.21 -0.13 -0.041 -0.061 -0.004 0.038 0.017 0.3 0.056 0.16 0.22 0.052 0.34 -0.14 -0.17
HighChol 0.21 0.3 1 0.086 0.11 0.091 0.093 0.18 -0.078 -0.041 -0.04 -0.012 0.042 0.013 0.21 0.062 0.12 0.14 0.031 0.27 -0.071 -0.085
CholCheck 0.068 0.099 0.086 1 0.034 0.0099 0.024 0.044 0.0042 0.024 0.0061 0.024 0.12 -0.058 0.047 -0.0084 0.032 0.041 0.022 0.09 0.0015 0.014
BMI 0.22 0.21 0.11 0.034 1 0.014 0.02 0.053 -0.15 -0.088 -0.062 -0.049 -0.018 0.058 0.24 0.085 0.12 0.2 0.043 -0.037 -0.1 -0.1
Smoker 0.063 0.097 0.091 -0.0099 0.014 1 0.061 0.11 -0.087 -0.078 -0.031 0.1 -0.023 0.049 0.16 0.092 0.12 0.12 0.094 0.12 -0.16 -0.12
Stroke 0.11 0.13 0.093 0.024 0.02 0.061 1 0.2 -0.069 -0.013 -0.041 -0.017 0.0088 0.035 0.18 0.07 0.15 0.18 0.003 0.13 -0.076 -0.13
HeartDiseaseorAttack 0.18 0.21 0.18 0.044 0.053 0.11 0.2 1 -0.087 -0.02 -0.039 -0.029 0.019 0.031 0.26 0.065 0.18 0.21 0.086 0.22 -0.1 -0.14
PhysActivity -0.12 -0.13 -0.078 0.0042 -0.15 -0.087 -0.069 -0.087 1 0.14 0.15 0.012 0.036 -0.062 -0.27 -0.13 -0.22 -0.25 0.032 -0.093 0.2 0.2
Fruits -0.042 -0.041 -0.041 0.024 -0.088 -0.078 -0.013 -0.02 0.14 1 0.25 0.035 0.032 -0.044 0.1 -0.068 -0.045 0.048 0.091 0.065 0.11 0.08
Veggies -0.059 -0.061 -0.04 0.0061 -0.062 -0.031 -0.041 -0.039 0.15 0.25 1 0.021 0.03 -0.032 0.12 -0.059 -0.064 -0.081 -0.065 -0.0098 0.15 0.15
HvyAlcoholConsump -0.058 -0.004 -0.012 -0.024 -0.049 0.1 -0.017 -0.029 0.012 -0.035 0.021 1 -0.01 0.0047 -0.037 0.025 -0.026 -0.038 0.0057 -0.035 0.024 0.054
AnyHealthcare 0.015 0.038 0.042 0.12 -0.018 -0.023 0.0088 0.019 0.036 0.032 0.03 -0.01 1 -0.23 -0.041 -0.053 -0.0083 0.0071 -0.019 0.14 0.12 0.16
NoDocbcCost 0.035 0.017 0.013 -0.058 0.058 0.049 0.035 0.031 -0.062 -0.044 -0.032 0.0047 -0.23 1 0.17 0.19 0.15 0.12 -0.045 -0.12 -0.1 -0.2
GenHlth 0.3 0.3 0.21 0.047 0.24 0.16 0.18 0.26 -0.27 -0.1 -0.12 -0.037 -0.041 0.17 1 0.3 0.52 0.46 -0.0061 0.15 -0.28 -0.37
MentHlth 0.074 0.056 0.062 -0.0084 0.085 0.092 0.07 0.065 0.13 0.068 0.059 0.025 -0.053 0.19 0.3 1 0.35 0.23 0.081 0.092 -0.1 -0.21
PhysHlth 0.18 0.16 0.12 0.032 0.12 0.12 0.15 0.18 -0.22 -0.045 -0.064 -0.026 -0.0083 0.15 0.52 0.35 1 0.48 -0.043 0.099 -0.16 -0.27
DiffWalk 0.22 0.22 0.14 0.041 0.2 0.12 0.18 0.21 -0.25 -0.048 -0.081 -0.038 0.0071 0.12 0.46 0.23 0.48 1 -0.07 0.2 -0.19 -0.32
Sex 0.031 0.052 0.031 -0.022 0.043 0.094 0.003 0.086 0.032 -0.091 -0.065 0.0057 -0.019 -0.045 -0.0061 -0.081 -0.043 -0.07 1 -0.027 0.019 0.13
Age 0.19 0.34 0.27 0.09 -0.037 0.12 0.13 0.22 -0.093 0.065 -0.0098 -0.035 0.14 -0.12 0.15 -0.092 0.099 0.2 -0.027 1 -0.1 -0.13
Education -0.13 -0.14 -0.071 0.0015 -0.1 -0.16 -0.076 -0.1 0.2 0.11 0.15 0.024 0.12 -0.1 -0.28 -0.1 -0.16 -0.19 0.019 -0.1 1 0.45
Income -0.17 -0.17 -0.085 0.014 -0.1 -0.12 -0.13 -0.14 0.2 0.08 0.15 0.054 0.16 0.2 -0.37 -0.21 -0.27 -0.32 0.13 -0.13 0.45 1
Diabetes HighBP HighChol CholCheck BMI Smoker Stroke HeartDiseaseorAttack PhysActivity Fruits Veggies HvyAlcoholConsump AnyHealthcare NoDocbcCost GenHlth MentHlth PhysHlth DiffWalk Sex Age Education Income
1.00 0.75 0.50 0.25 0.00 −0.25 −0.50 −0.75 −1.00

```
22855          0          0    1    7
121922        30          1    1   10
232355         0          0    0    9
208496        10          0    0    7
232431         3          0    0   10

[50736 rows x 10 columns]
Fitting completed
SVC Classification Report:
              precision    recall  f1-sco

           0       0.84      1.00      0.9
           1       0.00      0.00      0.0
           2       0.62      0.02      0.0

    accuracy                           0.8
   macro avg       0.49      0.34      0.3
weighted avg       0.80      0.84      0.7

Score:  0.8442673841059603
```

```python
# Random Forest
from sklearn.ensemble import RandomForestClassifier
rf_model = RandomForestClassifier(n_estimators=100, max_features="auto", random_state=44)
rf_model.fit(X_train, y_train)
rf_pred = rf_model.predict(X_test)
print("RF predicted : \n", rf_pred)
from sklearn import metrics
print("Accuracy:",metrics.accuracy_score(y_test, rf_pred))

pickle.dump(rf_model, open('rf_model.pkl', 'wb'))
```

```
RF predicted :
 [0 0 0 ... 0 0 2]
Accuracy: 0.826631977294229
```

Results

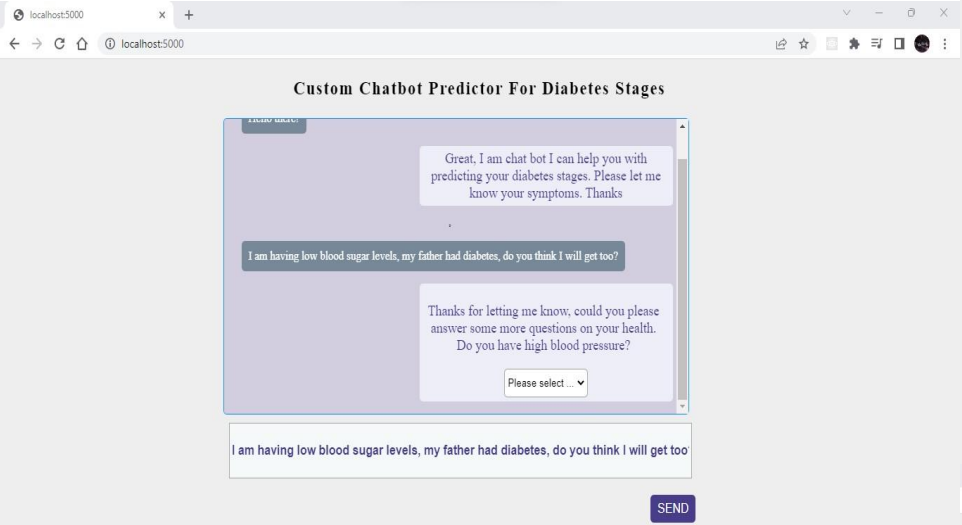

localhost:5000
localhost:5000
Custom Chatbot Predictor For Diabetes Stages
Hello there, I am Ben Whittaker. My gender is male and I am 55 years old, I have high blood pressure and more cholestrol levels than normal. My BMI is 26, last year I had heart stroke, I have heart condition, my normal health condition is always good, I have never experienced discomfort either physical or mental condition, I have leg pains and some level of difficulty while walking.
Your predicted stage of diabetes is SVM => No Diabetes K means => Pre Diabetes Random Forest => No Diabetes
Enter your message
SEND

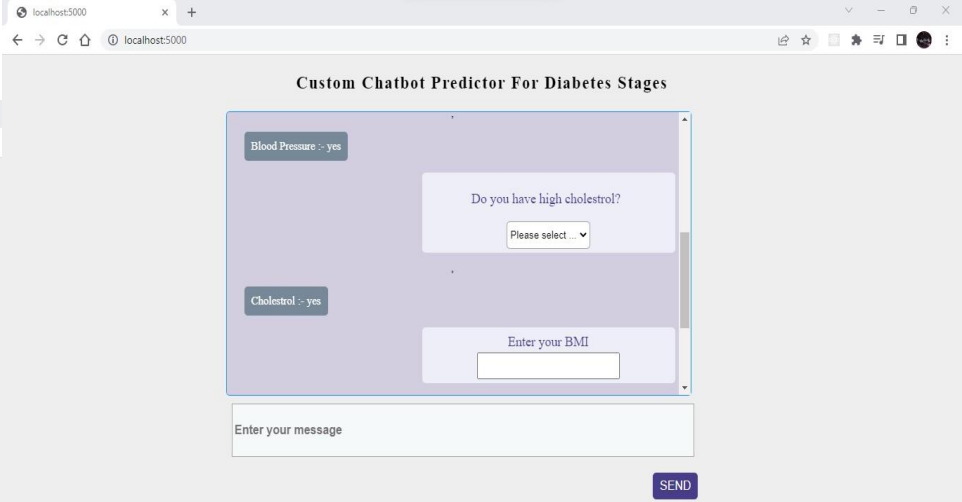
Results

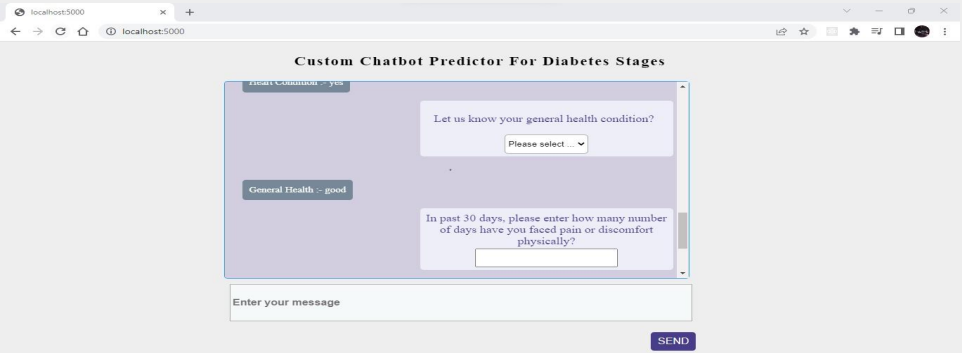
C:\WINDOWS\system32\cmd.exe - flask run

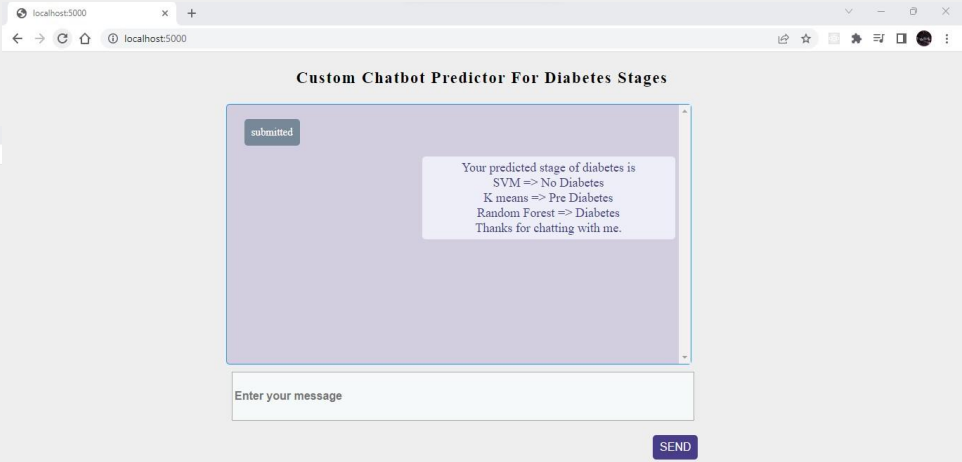
```
127.0.0.1 - - [05/Dec/2022 19:29:43] "GET / HTTP/1.1" 200 -
received message Hello there, I am Ben Whittaker. My gender is male and I am 55 years old, I have high blood pressure and more cholestrol levels than normal. My BMI is
26, last year I had heart stroke, I have heart condition, my normal health condition is always good, I have never experienced discomfort either physical or mental condi
tion, I have leg pains and some level of difficulty while walking.
General entities Ben Whittaker 18 31 PERSON
General entities 55 years old 60 72 DATE
General entities BMI 144 147 ORG
General entities 26 151 153 DATE
General entities last year 155 164 DATE
Hello 0 5 ENTITY
gender 36 42 ENTITY
male 46 50 ENTITY
I 55 56 ENTITY
years 63 68 ENTITY
I 74 75 ENTITY
blood pressure 86 100 ENTITY
cholestrol 110 120 ENTITY
levels 121 127 ENTITY
normal 133 139 ENTITY
BMI 144 147 ENTITY
year 160 164 ENTITY
heart stroke 171 183 ENTITY
I 185 186 ENTITY
heart condition 192 207 ENTITY
health condition 219 235 ENTITY
I 252 253 ENTITY
discomfort 277 287 ENTITY
physical 295 303 ENTITY
mental condition 307 323 ENTITY
I 325 326 ENTITY
leg pains 332 341 ENTITY
level 351 356 ENTITY
difficulty 360 370 ENTITY
walking 377 384 ENTITY
Entities: ['Ben Whittaker', '55 years old', 'BMI', '26', 'last year', 'Hello', 'gender', 'male', 'I', 'years', 'I', 'blood pressure', 'cholestrol', 'levels', 'normal',
'BMI', 'year', 'heart stroke', 'I', 'heart condition', 'health condition', 'I', 'discomfort', 'physical', 'mental condition', 'I', 'leg pains', 'level', 'difficulty', '
walking']
11 output
Array from entities: [[ 1  1 25  1  1  1 20  1  0 40]]
Inputs from entities :  [[ 1  1 25  1  1  1 20  1  0 40]]
got output : {'svm_predicted': {'output': 'No Diabetes'}, 'km_predicted': {'output': 'Pre Diabetes'}, 'rf_predicted': {'output': 'No Diabetes'}}
127.0.0.1 - - [05/Dec/2022 19:29:47] "POST /send_message HTTP/1.1" 200 -
```

Results


localhost:5000
Custom Chatbot Predictor For Diabetes Stages
Great, I am chat bot I can help you with predicting your diabetes stages. Please let me know your symptoms. Thanks
I am having low blood sugar levels, my father had diabetes, do you think I will get too?
Thanks for letting me know, could you please answer some more questions on your health. Do you have high blood pressure?
Please select ...
I am having low blood sugar levels, my father had diabetes, do you think I will get too
SEND


localhost:5000
Custom Chatbot Predictor For Diabetes Stages
Blood Pressure :- yes
Do you have high cholestrol?
Please select ...
Cholestrol :- yes
Enter your BMI
Enter your message
SEND


localhost:5000
Custom Chatbot Predictor For Diabetes Stages
Let us know your general health condition?
Please select ...
General Health :- good
In past 30 days, please enter how many number of days have you faced pain or discomfort physically?
Enter your message
SEND


localhost:5000
Custom Chatbot Predictor For Diabetes Stages
submitted
Your predicted stage of diabetes is
SVM => No Diabetes
K means => Pre Diabetes
Random Forest => Diabetes
Thanks for chatting with me.
Enter your message
SEND

REFERENCES:

- Cheheltani, R., King, N., Lee, S., North, B., Kovarik, D., Evans-Molina, C., . . . Dutta, S. (2022). Predicting misdiagnosed adult-onset type 1 diabetes using. ELSEVIER, 7.

- Chollette C, O., Smith, L., & Smith, M. (2022). Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective. ELSEVIER, 12.

- Houssein, E. H., Rehab E, M., & Abdelmgeid A, A. (2021). Machine Learning Techniques for Biomedical Natural Language Processing: A Comprehensive Review. IEEE, 26.

- Oyebode, O., Alqahtani, F., & Orji, R. (2020). Using Machine Learning and Thematic Analysis Methods to Evaluate Mental Health Apps Based on User Reviews. IEEE, 18.

THANK YOU