Name: Xin Peng, Yuchen Zhao
Date: 12/22/2022
NYU NetID: xp2083, yz8759
Section: 3
Total in points: Professor's Comments:

Project 4
Affirmation of my Independent Effort: Xin Peng, Yuchen Zhao

1. Case analysis

Nowadays, insurance companies are facing more and more opportunities as well as risks. Diseases, disasters and global emergencies are happening more and more frequently and can easily cost companies millions of dollars. How to leverage data has become a key part to control risks and gain profits.   Our program can simulate and provide solutions in cases showing below:

a. Deal with unstructured data
    Market research is one of the biggest sources of data in daily activities to enhance the performance of our model. Unstructured data from market research needs to be uploaded to our database and the model has to be updated according to these new data. Associates are able to upload unstructured data using the program interface, and the updated model will provide more precise quote suggestions.
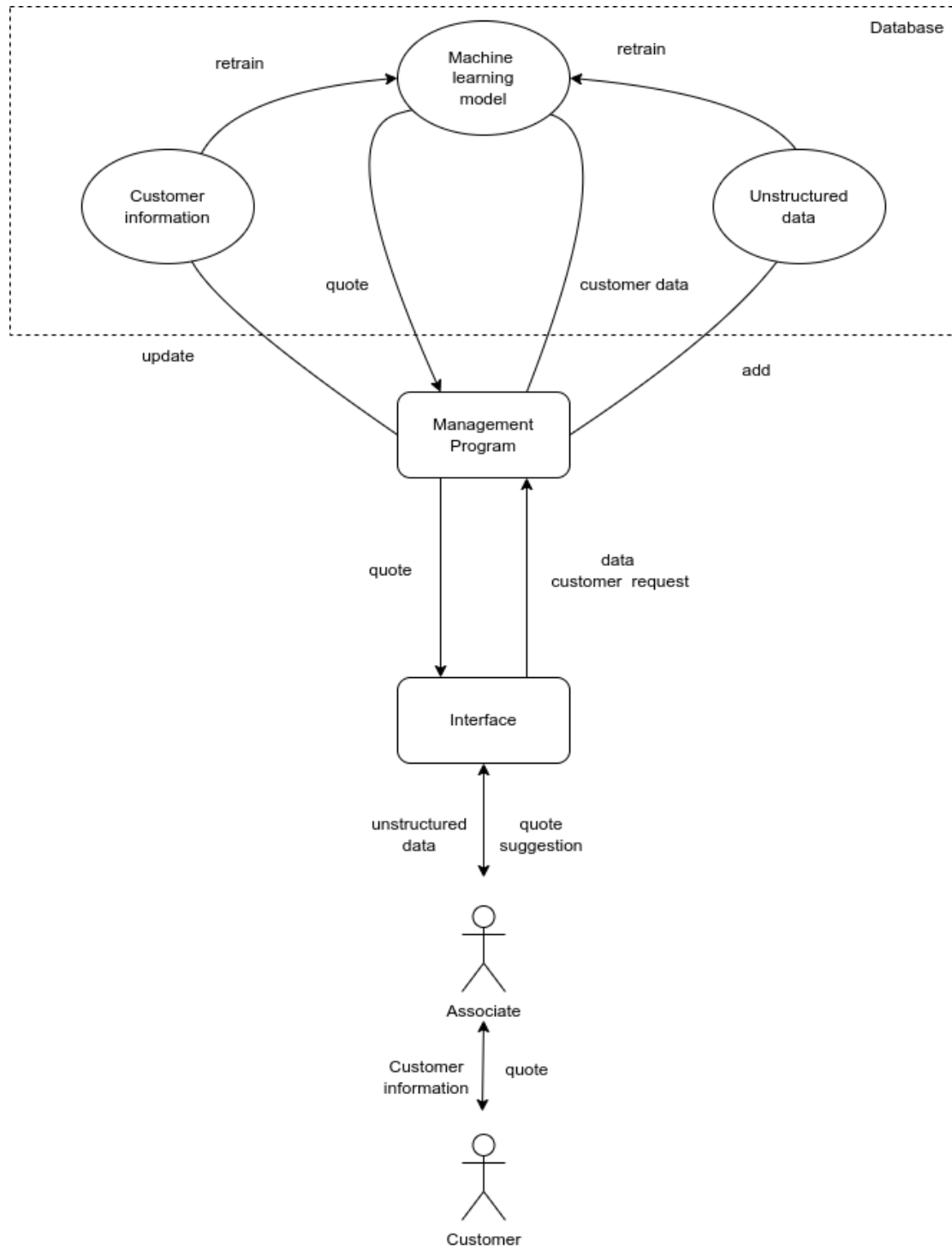
b. Give quote suggestions
    Associates need quote suggestions to help their decisions on premiums and policies. Our database stores huge amounts of personal data and claims, which can be a good source of information. When there comes a new customer asking for a quote or a customer hoping to renew the plan, our associates are able to upload related information and get a quote suggestion from our system. The quote suggestion usually includes the policy and price.

2. Logistic design

There will be three parts in our modern Enterprise Data Architecture (EDA): the management program, the database and the interface. The interface provides interactions with associates to deliver information between customers and the management program. The management program will act as the core part of the whole system, which can interact with the cloud database, manage the machine learning model and pass results back to the interface. The cloud database is a place where stores customers data and unstructured data gained from market researching.
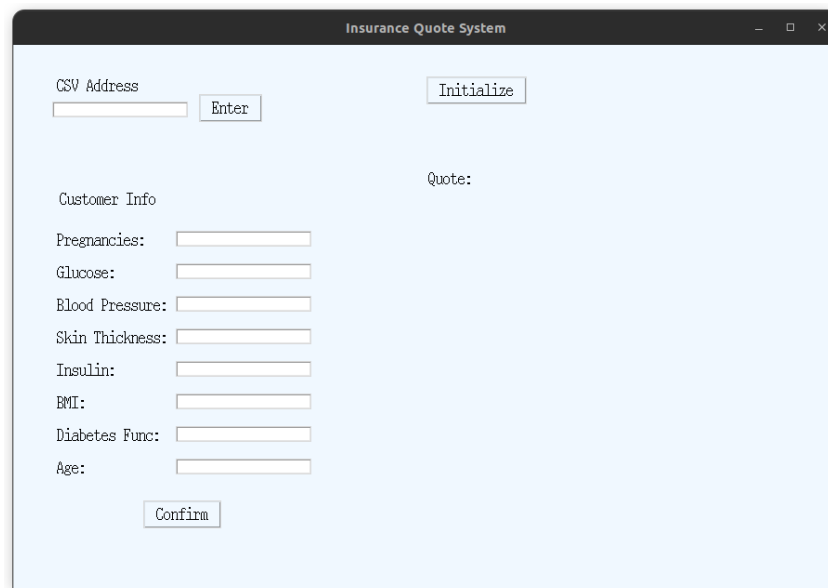
Due to the fund limitation, we design an offline data structure to mock the cloud database. This data structure stores customer information and the unstructured data as well as the machine learning model.

The whole design is shown in the graph below:

3. Implementations

a.   Interface

The user interface is implemented with the python programming language using the tkinter package. It has two modes: the quote suggestion mode and the unstructured data input mode. It is designed to provide support for associates.

In the quote suggestion mode, associates input customers' examination data, the interface will pass these data to the management program. Then, the management program will return the quote suggestion, including diabetes policy with the suggested premium.

In the unstructured data input mode, a csv file storing plenty of patients examination records is the expected input, and this file will be handed to the management program for advanced processing.

b.   Management program

The management program is the core of the whole architecture. It is implemented with the Python programming language. The management program consists of three parts: the unstructured data processing part, the model updating part, and the quote retrieval part.

For the unstructured data processing part, it will transfer unstructured data into structured data and store it into the mocked database. It will also trigger the model updating part, which retrains the machine learning model using the updated data in the mocked database.

For the quote retrieval part, when receiving the quote request from the user interface, the management program will give a prediction result using the machine learning

model stored in the mocked database and return it to the user interface. The customer data will also be stored into the mocked database.

After a certain amount of time, the management program will automatically retrain the machine learning model using both unstructured data and the customer data to gain the most real-time insight into the current situation.

c. Mocked database

Due to the fund limitation, we gave up the attempt to connect the cloud database and our management program. Instead, we use a data structure in csv file format to simulate the cloud environment. The data will be uploaded to the cloud manually.

The mocked database consists of three parts: unstructured data, customer information and the machine learning model. They are mainly manipulated by the management program and provide support for quote application.

Github link:
https://github.com/yuchenzhao726/DMS2022FALL_project.git