Name: Chenhao Jin

NetID: cj2234

Section: 001

Database System Project Part 4

(Github: https://github.com/jch0625/DatabaseSystem2433)

This application can enable customers to get a quote on their basic health condition information and physical check result. This application used 6 Random–forest regression models to predict the probability of getting 5 kinds of chronic disease, and the probability of getting at least one chronic disease. The initial data is from the UCI machine Learning Repository (https://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease). The quote is based on the probability prediction result. The basic information needed for the quote are below :

age - age
al - albumin
pc - pus cell
bgr - blood glucose random
sc - serum creatinine
pcv - packed cell volume
rc - red blood cell count
bp - blood pressure
su - sugar
pcc - pus cell clumps
bu - blood urea
pot - potassium
wc - white blood cell count
sg - specific gravity
rbc - red blood cells
ba - bacteria
sod - sodium
hemo - hemoglobin

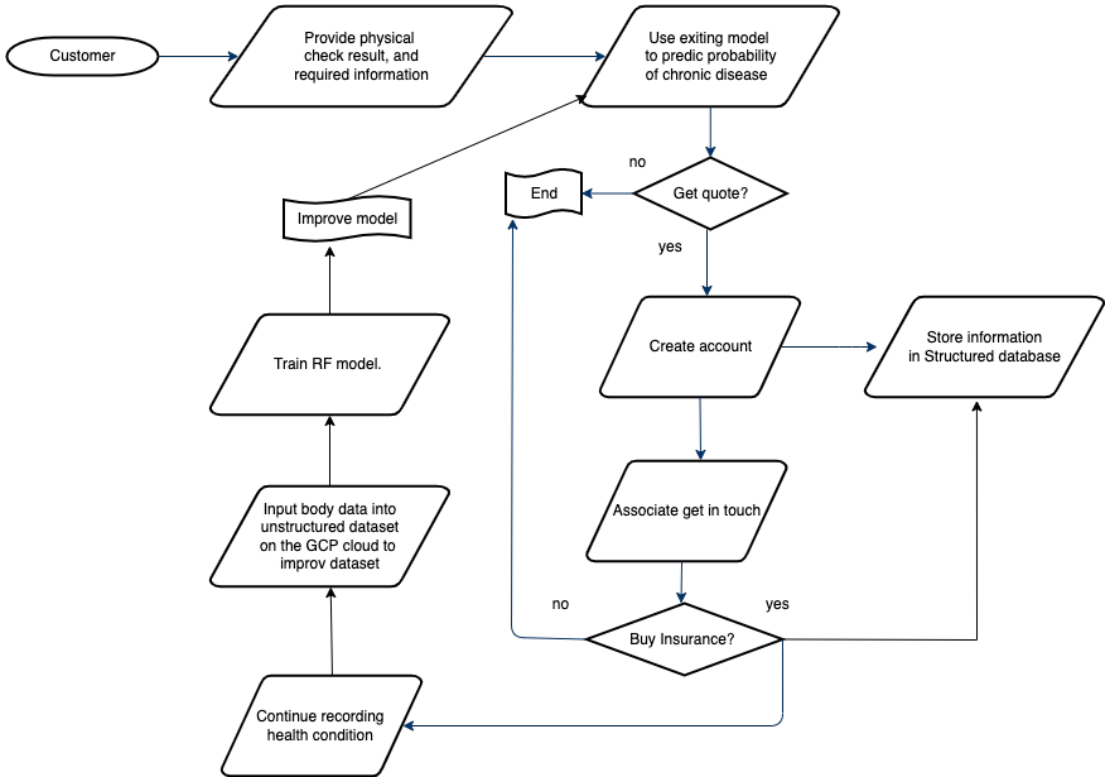
Each model will predict the probability of respective chronic diseases. For now, the quoted amount is calculated by the sum of probability multiply a given amount for each disease. In the future, the company can use machine learning or deep learning to get an efficient function. Customers can choose “Basic”, “Standard” and “Premium” for their insurance. They can check the price for each level.

Tkinter is used for the user interface. Customers can get a quote first and decide whether to create an account and buy the product. Using entry.get() can get the user’s input on the interface. Use SQL to insert value to respective column and table. As the figure showed, the insertvalue_cus function will get the parameter, and insert those values into the customer table.

The initial data from UCI for model training is insufficient. This application use pymongo to connect the MongoDB atlas. If the customer decides to create an account and buy insurance, the health information will be stored as unstructured data in MongoDB Atlas cloud

anonymously. The company can track users' health situations and update the data in Mongo DB. This dataset can be used to train and improve the Random Forest model used above for the prediction of chronic disease. By increasing the number of customers and dataset, more data will be fed to the model and the model will be more curated.

```
client = MongoClient( "mongodb+srv://
dbms:jchjchjch@dbns.swthwi3.mongodb.net/?
retryWrites=true&w=majority", tlsCAFile = certifi.where())
db = client["dbmsfinal"]
collection = db["dbms"]
collection.insert_one({"age":age,'bp':bp,'sg':sg,'al':al,'su':su,'
rbc':rbc,'pc':pc,'pcc':pcc,'ba':ba,'bgr':bgr,'bu':bu,'sc':sc,'sod'
:sod,'pot':pot,'hemo':hemo,'pcv':pcv,'wc':wc,'rc':rc})
```



The customer will fill in some information for account creation.

First name, Last name, Address, Phone number, SSN, account name, and password.

Application use pymysql to connect the local MySQL database, and save First name, Last name, SSN, and phone number into the Customer table. The SSN will be the primary key of

the table. First name and Last name are indexed for increasing query speed when searching by name.

Account name, Address details and Password will be stored in the Account table. The account ID is auto-generated by increasing order, the start data will also be generated. Account ID is the primary key of this table. The account name is indexed to make login faster.

```
def get_connection():
    usr = "root"
    pw = "jchjchjch"
    h = "localhost"

    conn = pymysql.connect(
        user=usr,
        password=pw,
        host=h,
        cursorclass=pymysql.cursors.DictCursor,
        autocommit=True
    )
    return conn

def insertvalue_cus(Customer_ssn, Cus_Fname, Cus_Lname, Cus_phone):
    sql = f'''INSERT INTO dbms.Customer(Customer_ssn, Cus_Fname, Cus_Lname, Cus_phone) VALUES (%s,%s,%s,%s)'''
    conn = get_connection()
    cur = conn.cursor()
    res = cur.execute(sql,args=(Customer_ssn,Cus_Fname, Cus_Lname, Cus_phone))
    conn.commit()
    result = cur.lastrowid
```

The account id will be queried and store in new associate table Account_customer. This table only contain SSN and account id to connect two table. This is many to many relation.

The quote amount predicted by model will be saved in Contract table as contract amount. The contract state date, account id for this contract, and contract status will be stored in this table. The primary key contract id is auto generated by increase order.

LogIn

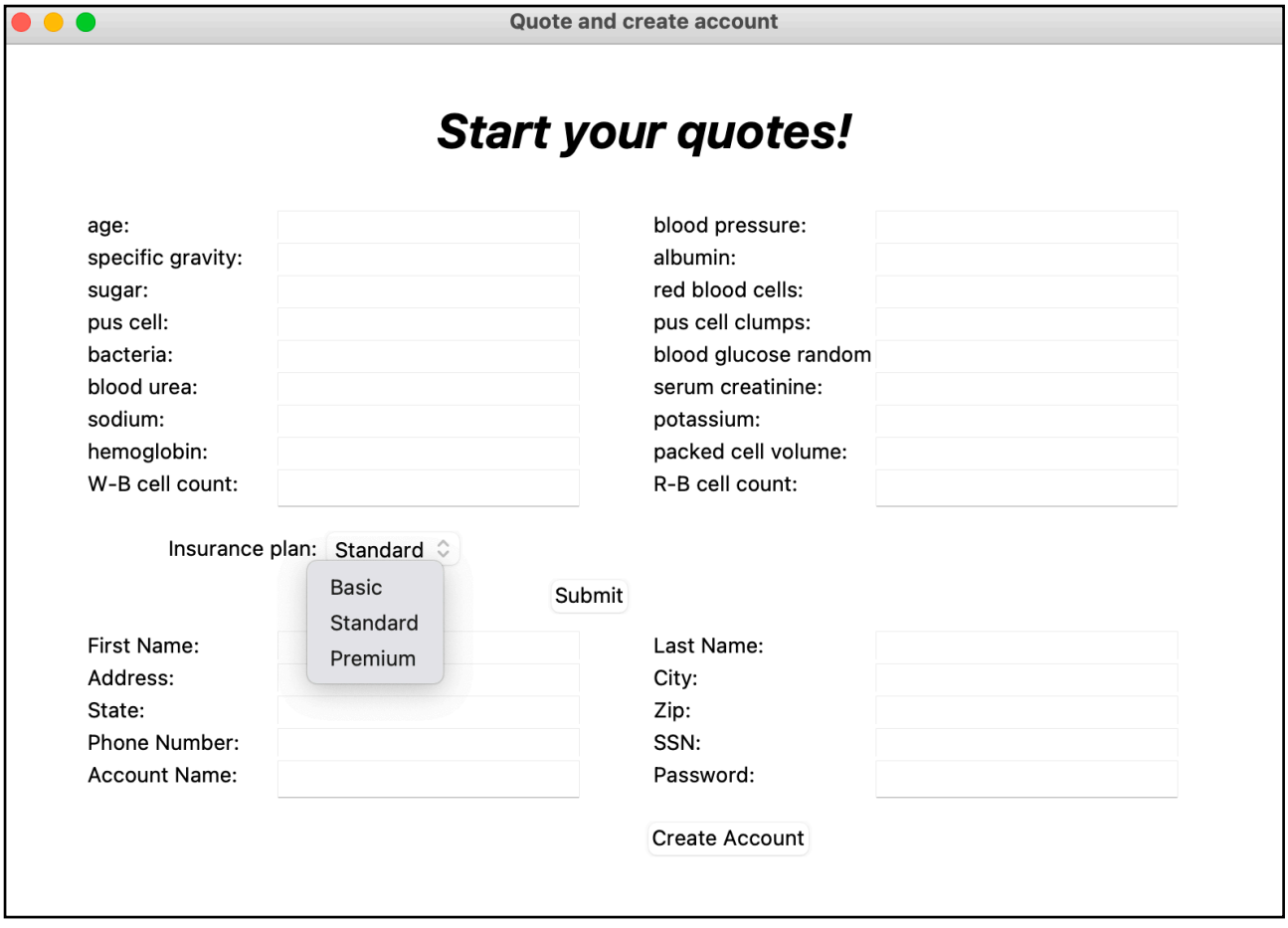
Users can use their account name and password to log in to their accounts. If the account and password don't match the record in the database, the application will notify you. After login, user can see their account details and contract details. The query will first match the account_name and password. Then, the query will select the needed column of this account. The account number(PK) will be used to query the Contract detail in the Contract table as a

foreign key. The account number will also be used to first get the customer SSN in Customer_Account. And SSN will be used as the primary key to query information in the Customer table.

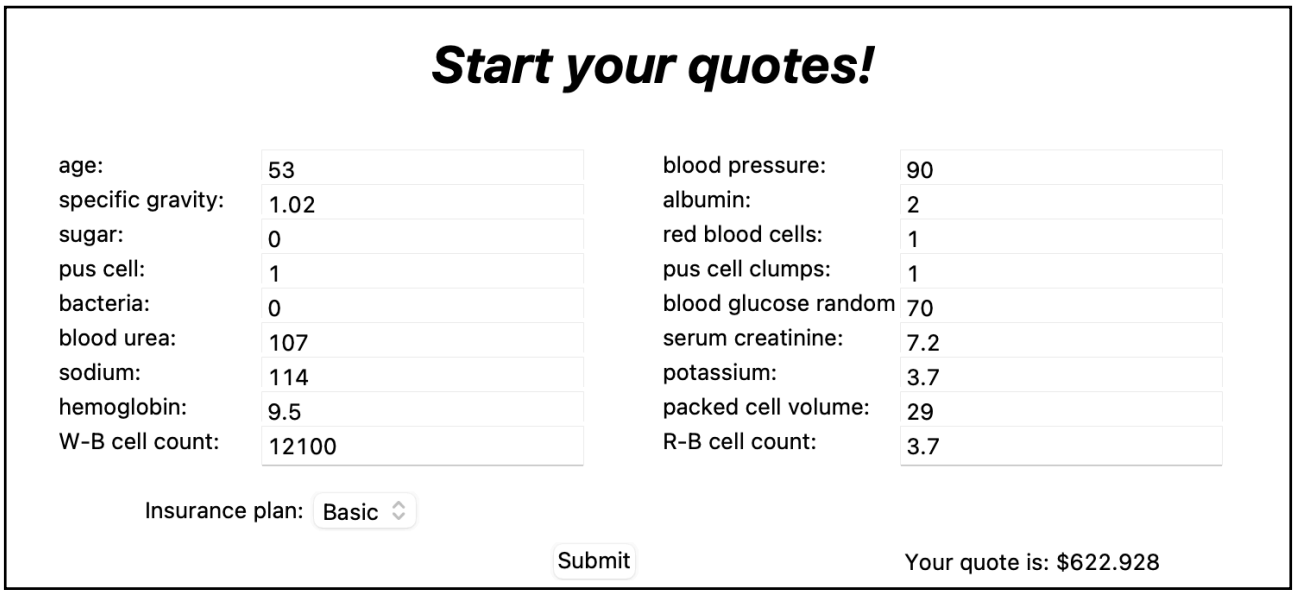
The health condition data will be stored in JSON file without any user's information. Those data are stored in the cloud which makes sure personal privacy. Those data are only used to train Random forest models. Even if the account information is leaked, other people can't get customers' privacy information. Customers' health data will not be stored if they don't create an account or buy insurance. Further health data collection also needs customers' consent. Customers may enter the data with some mistakes. It will influence the model training. The data should be clean before feeding the model. The requirement of health data for the quote is high. This may make the customer do a physical check-in purpose. In the future, the company can reduce the field required, and find out several top important features.

Example

Starting page

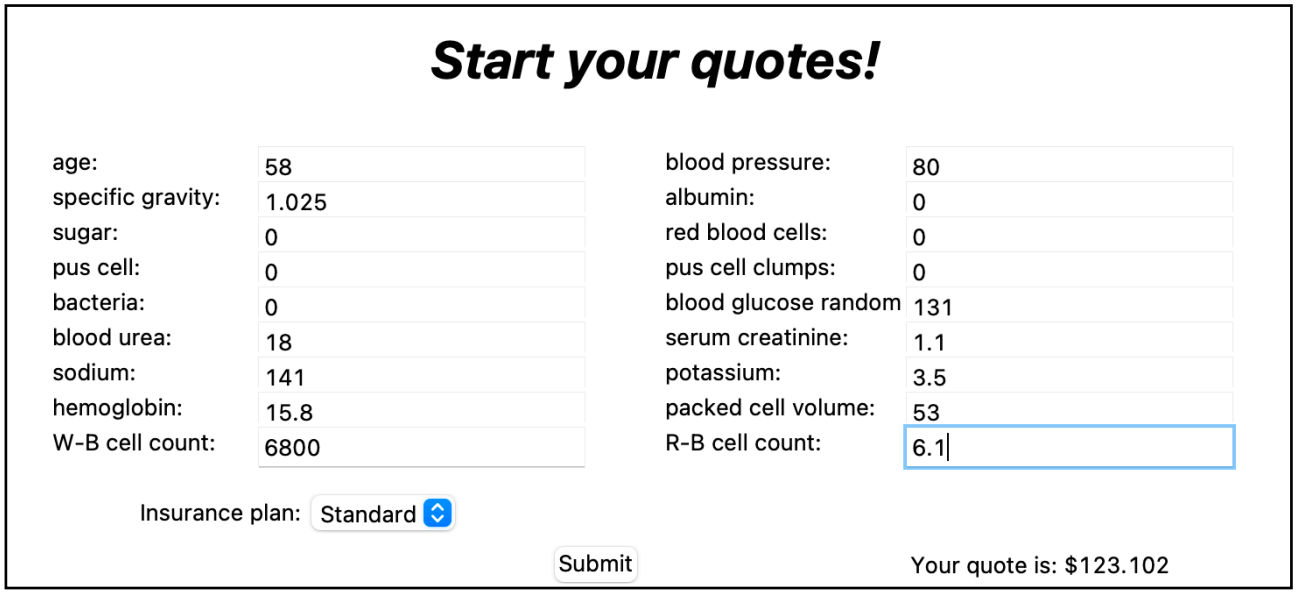


This is a person who has three kinds of chronic disease

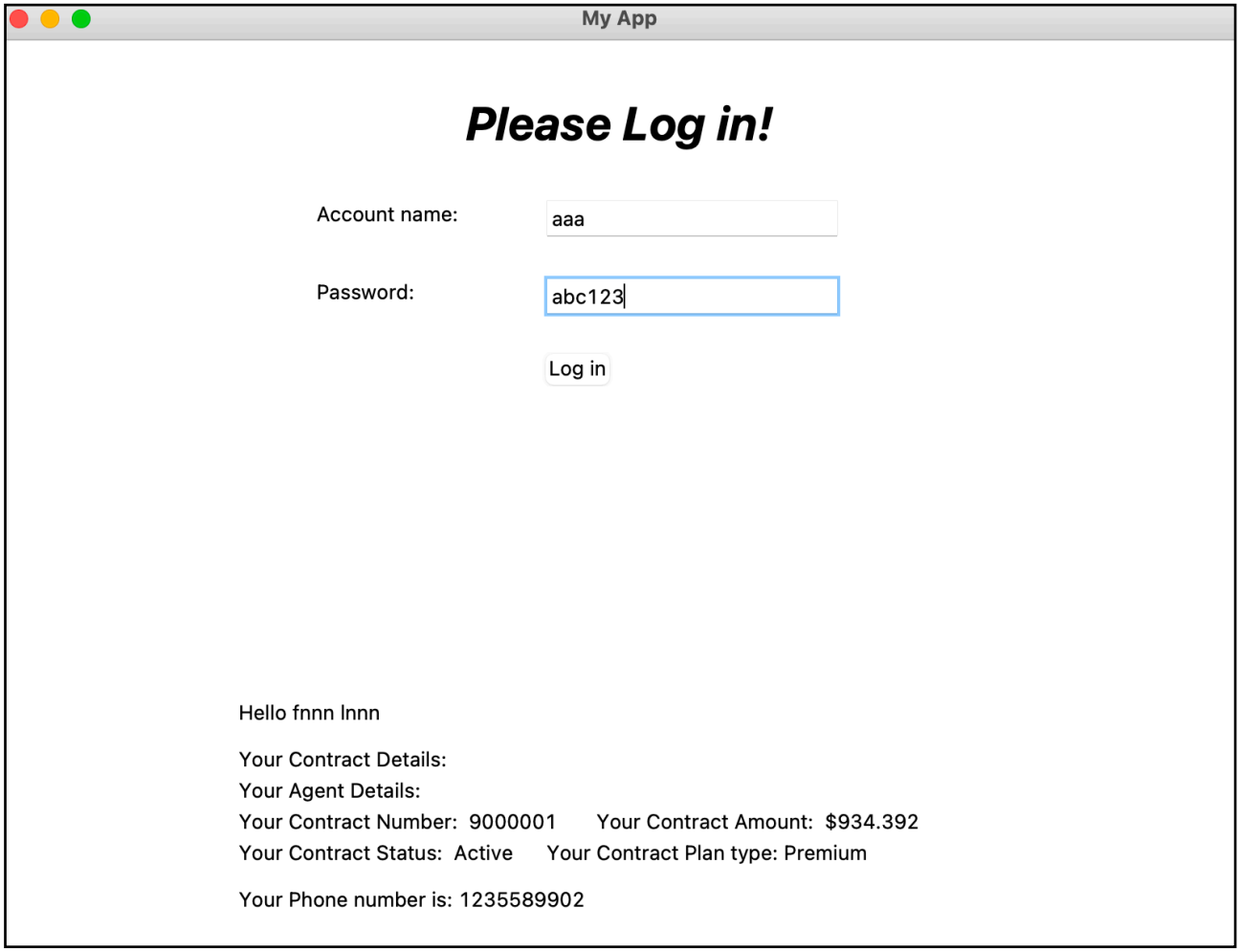


Costumer can change plan in selection box to check price

This is a quote of healthy person who doesn't have any chronic disease.



Log in page


My App
Please Log in!
Account name:
aaa
Password:
abc123
Log in
Hello fnnn lnnn
Your Contract Details:
Your Agent Details:
Your Contract Number: 9000001
Your Contract Amount: $934.392
Your Contract Status: Active
Your Contract Plan type: Premium
Your Phone number is: 1235589902