Table of Content

● Introduction

We are team Doctor Cloud. We focused on the healthcare industry and devoted to provide cloud-based services to predict and improve personal health status. Cardiovascular diseases and stroke are the largest causes of deaths in the US. A person dies of a heart stroke every 4 minutes in the USA ("Stroke Facts | cdc.gov", 2019).

We want hospitals to create value at not only the business level, but also create value at the patient experience level. By improving the patient experience the hospital can generate lot of revenue which can be made possible with an effective digital transformation strategy.

Mayuresh's parents are doctors and they have their own hospital back in India. So this is a very interesting topic of stroke prediction using ML model, which will help us gain meaningful insights from such a implementation in a traditional healthcare setup which is not yet fully digitized. Such an idea will pave a path towards digital transformation of hospitals in developing countries such as India and China. This project would provide us some insights on which kind of people are more likely to suffer from stroke. In practice, we assume that all personal data in this project can be obtained through wearable devices and integrated into the IoT platform to make predictions later. In our discussion, the cloud platform should be able to provide a rich healthcare analytics library and organizations should be able to deploy these advanced analytics including this stroke predictive model into their business module.

The business problems hospitals are facing can be summarized as follows:

- How can they use the vast amount of patient data, collected from IoT devices, to improve healthcare and leverage business revenue.
- How to get real-time predictive analysis for patients thereby reducing their time spent in hospital admissions and making healthcare affordable for them.
- How to increase the level of patient engagement and loyalty by reducing patient leaking (patients not returning back to the primary hospital and seeking other rival healthcare providers).
- How to protect patients privacy and ensure data security

● Impact

We are planning to use hybrid cloud strategy and take advantages of both private cloud and Microsoft Azure public cloud. We'd use some Azure services including IoT service, development service, machine learning service and database service.

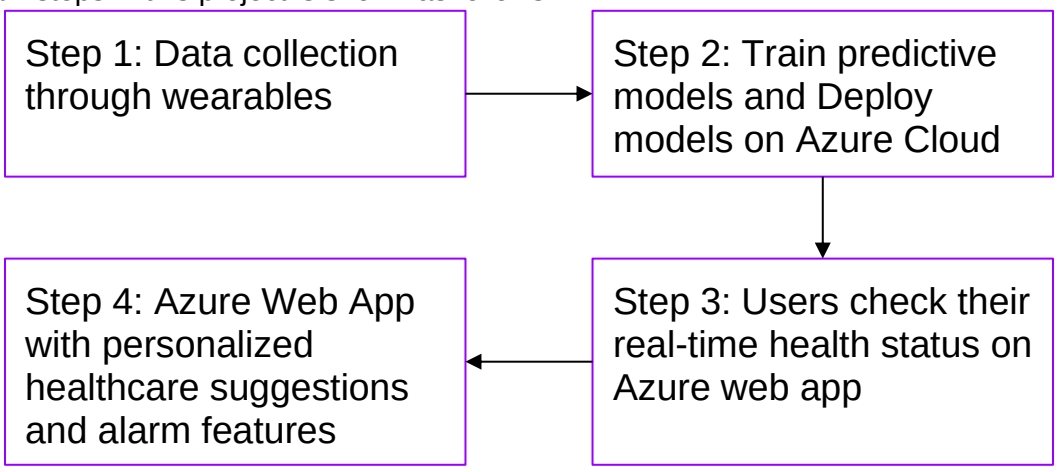
The Internet of things (IoT) service can help us to capture, gather, monitor and analyze data (vitals and other relevant body measurements) from wearable devices. Development service would help us to test our predictive model and track potential issues. We would use AI and machine learning service to infuse our model into applications and data sets. The database service is also critical for this project, since we want to store the data in Azure database and ensure cyber security.

● Change

This project would change our business model by saving costs. With Azure public cloud, we can decrease our local server numbers. The data storage and computational costs are scalable to real-time needs. Another change is that some patients could simply use the web app to check their health status, and don't need to go to the hospital. Overall, the project would deliver significant business benefits, without any upfront costs and provisioning time.

● Transformation

The main steps in this project is shown as follows.



There are three phases in our digital transformation plan.

- Phase 1: Digital Frontend Transformation

We'll make some projects focusing on customer-facing areas. These projects are driven by individual technical managers. We would also train IT staff and improve processes to focus on customer experience. We will run an initial pilot test to get feedback on the effectiveness before an organization wide roll out.

- Phase 2: Digital Integrated Enterprises
  We'll integrate these separate digital initiatives and to build up a corresponding digital-end. This phase aims to implement cross-functional digital intelligence for the whole company.

- Phase 3: Digital Ecosystems
  There's supposed to be a digital ecosystems emerging trend in different industries. Different players are pursuing collaborative business models to provide innovative services to customers. We'll collaborate with external companies and build this ecosystem to continue to add values to our business models.

● Methods

The developments in the field of machine learning have given way for growth of tools that can assist physicians in disease diagnosis. Early diagnosis and prognosis of stroke are crucial for timely prevention and cure. We will try to design a model with high accuracy(>~90%) that will help in stroke prediction based on various physiological parameters that are used as risk factors. This project is aimed to predict stroke based on multiple physical and lifestyle characteristics such as age, hypertension, heart diseases, smoking status and gender. These parameters are chosen based on our initial research on this topicand has been collected from visiting various web resources, journals such as Centers for Disease Control and Prevention(CDC) and 'Heart', which is an international peer reviewed journal that keeps cardiologists up to date with important research advances in cardiovascular disease.

- Data Collection

  We collected more than 50,000 observations from patients and healthy people. The data source is Kaggle.com. We hope to learn the pattern behind the messy data and get some insights on the potential factors that would lead to stroke. This project would be helpful for doctors to pay more attention to those who are more likely to suffer from stroke, and for public to know more about stroke. Finally, we'd provide some recommendations for stakeholders including doctors, people who care about health, health organizations and so on.

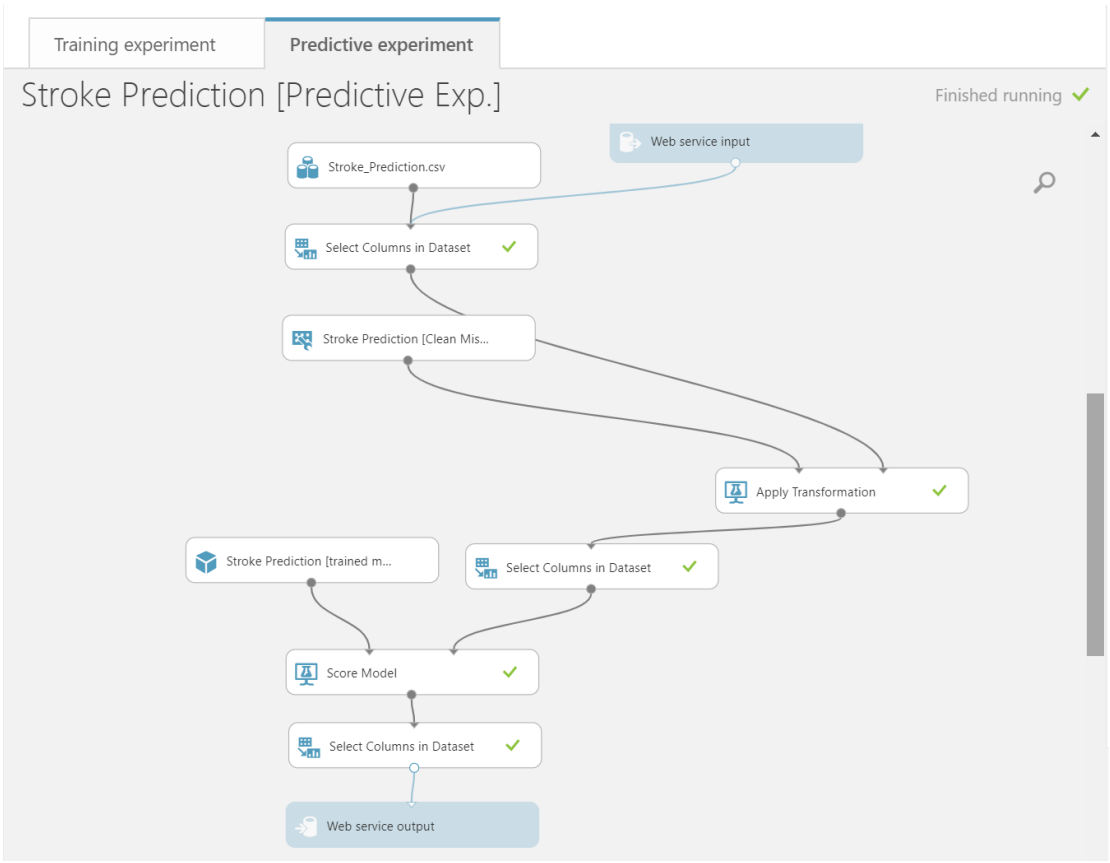
- Data Model description

  We have eleven explanatory variables as follows. Based on these values, we'll make prediction on the probability that the patient would suffer from stroke.

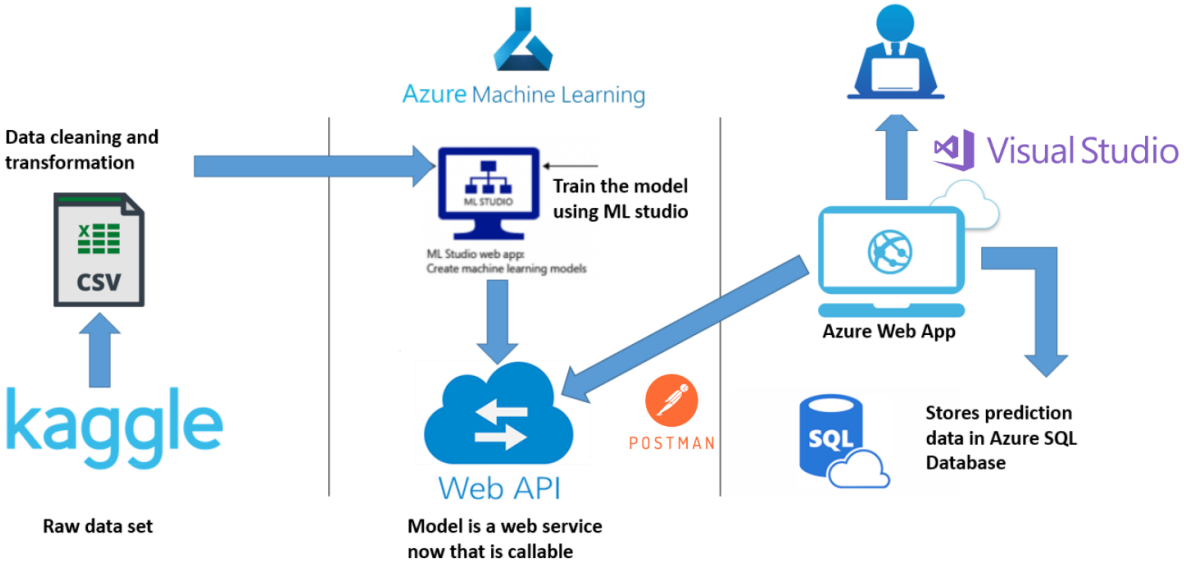
Variable Name	Data Type	Description
ID	int	Patient ID
Gender	string	Gender of Patient
age	int	Age of Patient
hypertension	binary	0 indicates no hypertension; 1 indicates suffering from hypertension
Heart disease	binary	0 indicates no heart disease; 1 indicates suffering from heart disease
Ever married	string	Yes or no
Work type	string	Type of occupation (five classifications would be given for users to choose from)
Average glucose level	float	Average Glucose level measured after meal
Residence type	string	Area type of residence (Urban/Rural)
bmi	float	Body mass index
Smoking status	string	Patient's smoking status (three types of smoking status would be given for users to choose from)

- Data cleaning
  We used Python and Microsoft Azure platform to perform data cleaning task. For the data cleaning part, we dropped the rows with missing bmi variable, around 10% of the total observations.

- ML model training
  We used Azure ML studio to train the model and wrapped a predictive web service. All 10 features were selected to build the logistic regression model. The web service output is the predicted probability that the patient would suffer from stroke. Our final predictive model is shown as follows.


Training experiment
Predictive experiment
Stroke Prediction [Predictive Exp.]
Finished running
Web service input
Stroke_Prediction.csv
Select Columns in Dataset
Stroke Prediction [Clean Mis...
Apply Transformation
Stroke Prediction [trained m...
Select Columns in Dataset
Score Model
Select Columns in Dataset
Web service output

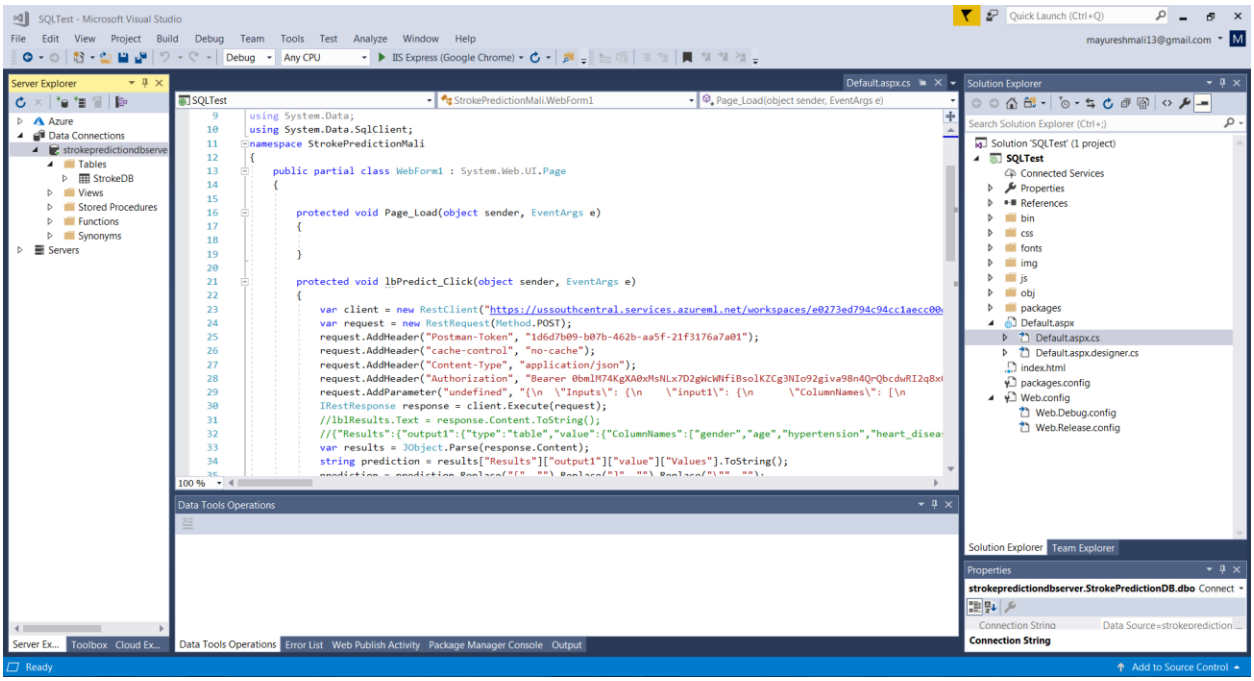
- Design



Above is a high level overview of our approach to Database operations. We have obtained a dataset containing patient information from Kaggle.com . We performed data cleaning and transformation operations on the data set to feed as input to Azure Machine Learning Studio for modeling. We used logistic regression model for predictive analysis and splitted the data set into training and test data(75:25).After we are satisfied with our model, we deployed our predictive model built in Azure Machine Learning studio as a scalable, fault tolerant Web service. Using this model we would call the Web API in our Azure Webapp. The web application would accept test parameters from the user as input and provide prediction result as output. These input parameters would be run by the Web API against the model and return a result in the form of a prediction to the user. The input features and prediction result would then be stored in the Azure SQL database for later access if a user/admin wishes to see the past prediction results.

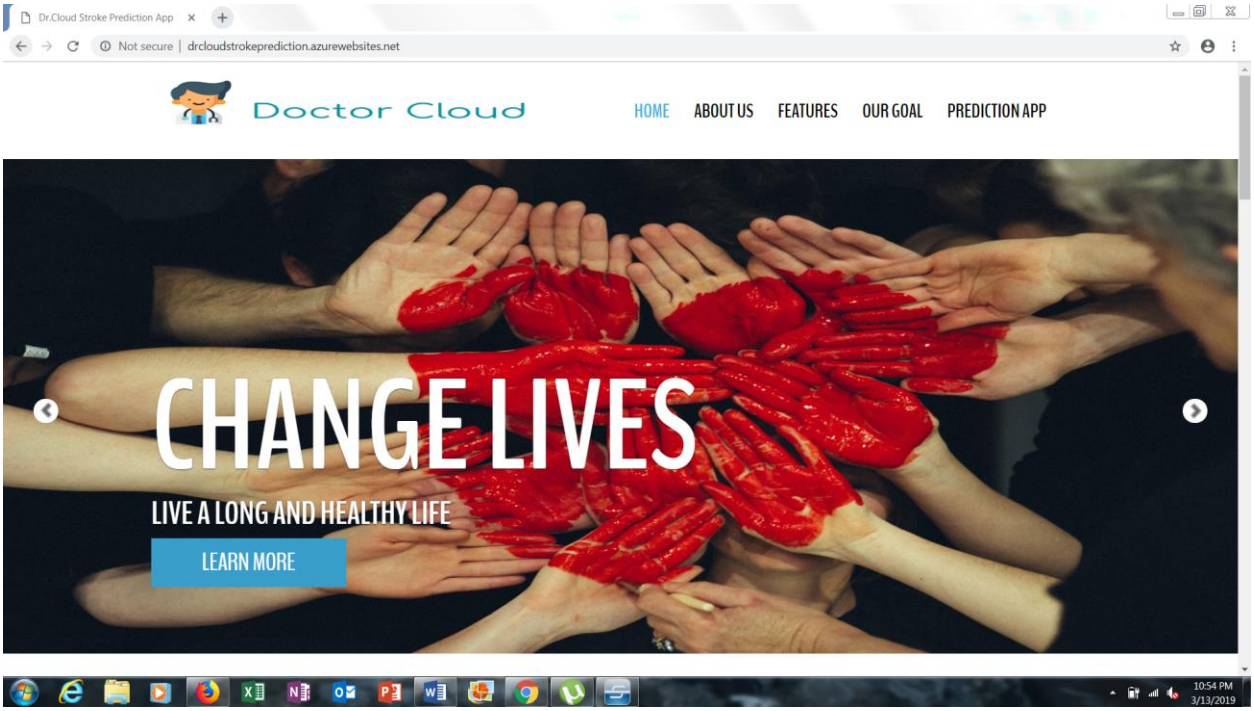
We used ROC curve to evaluate our model. The AUC for our model is 0.859. The threshold we set for our model is 0.5. The overall accuracy rate is 98.3%, which is pretty good.

- Results



C# code for developing the front end of the Azure web application using ASP.Net

We developed a web app using C#, HTML, ASP.NET and deployed it on the Azure platform. This web application made use of RestSharp framework to make use of the Web API to communicate with our Predictive model, which was deployed as a web service. Furthermore, we used the Newtonsoft frameworks to parse, query and modify JSON using JObject. We also added a functionality to save the predictions results and data back to the Azure SQL Database. Finally to complete the web app experience we developed an intuitive user interface which talked about our goals and the work we are doing and redirecting the users to the prediction app.



The web application's opening page

The prediction page of the web application

The deployment of this web app ties in strongly to the strategic goals of the hospital for which we are appointed. We want to create value for business as well as patients. This strategy has a chain reaction where one event leads to another to achieve the final goal of leveraging business. Our efforts will bear fruit in three primary areas -

Prioritize the customer experience - a customer experience-oriented company as an enterprise that focuses its strategy, energy and budget on processes that enhance the knowledge of and engagement with customers in this case the patients. This helps to tackle the business problem of patient leaking and loyalty.

ROI of Digital Transformation - With a better customer engagement, patients will be more likely to visit the same doctors and hospitals more often to stay healthy. Making the patients aware about the health benefits of such regular checkups will lead to repeat business and increased revenue for the healthcare organization.

Managing information - With properly trained IT staff in place, hospitals will be better equipped to handle data more effectively and comply with healthcare regulations and laws (HIPAA, FDA) without paying any fines for non-compliance.

● Discussion

What were the most interesting sub-problems related to your work? What design decisions were particularly difficult? What has the audience learned from your work?

The most interesting sub-problem for us was to choose the suitable methods for ETL, machine learning algorithm as well as model deployment on Azure. We first used Python Jupyter Notebook to perform data cleaning and transformation process. Then we loaded the cleaned data into ML studio to train the model. For the machine learning algorithm, we used logistic regression since it’s faster and more accurate than others. For the model deployment, we used C# to build a web app on Azure, which enables patients/doctors to input the parameters, and the app can output the prediction on the webpage.

The challenging part was the connection to Azure SQL database. We tried many times to store the data obtained from the front-end and the prediction results into our database and eventually succeeded to store our results in master database. This project is helpful for people to understand the whole design and implementation process for an Azure web app with machine learning algorithms deployed in the back-end. It gives them an end to end solution for deploying ML based web applications on Microsoft Azure successfully.

● Future Work

We would like to focus our efforts on the following things going forward -

Data Cleaning Processes - First of all, more work is needed on cleaning the data. Due to time and resource constraints we were unable to devote much time on clean data collection and transformation. We followed basic ETL strategy to cleanse the data so that our input feed for the ML model gave us the expected results that we were seeking for. Normalizing the data is a critical first step. We can use Amazon Mechanical Turk's (MTurk) human labeling service to have real-life people annotate or clean your labels to make sure your models are being trained on high-quality data.

Hybrid Cloud strategy - Hybrid cloud is a cloud computing environment that uses a mix of private cloud and public cloud services with orchestration between the platforms allowing data and applications to be shared between them. Digital business transformation agendas are focused on investments to make money. The primary benefit of a hybrid cloud is agility. The need to adapt and change direction quickly is a core principle of a digital business and this can be achieved by combining public clouds, private clouds, and on-premises resources.

BI analytics - The beauty of Cloud BI applications is that they are accessible on multiple devices and web browsers. This is circumventing traditional software barriers such as the requirement to access the application on-site. Using targeted dashboards customized for doctors to view individual high risk patients, doctors will be able to identify opportunities for improvement. The hospital then can develop and deploy highly targeted, specific interventions to promote those improvements in care, such as proactive measures to be taken by patients to reduce stroke chances.

Using Complex and highly accurate ML algorithms - Since we explored Azure ML Studio more than AutoML, it will be interesting to see what level of customizability we get. Selecting the best algorithms and an efficient hyper-parameter tuning we can get models with a very high accuracy.

Appendix

Each team member must contribute to the project. As a part of the index provide information on who did which sections in the project.

Task	Team Member

Data collection, methodology framework design	together
Machine learning model in ML studio	Yufei Li/ Mayuresh Mali
Final web app	Mayuresh Mali
Final slides and report	together

Web references

Stroke Facts | cdc.gov. (2019). Retrieved from https://www.cdc.gov/stroke/facts.htm

Image references

"https://www.freepik.com/free-photos-vectors/banner" Banner vector created by starline - www.freepik.com