PROJECT STATEMENT OF WORK

ORGANIZATION		
CLIENT	NAME	XX
	PHONE	123456
	EMAIL	example@gmail.com
	MAILING ADDRESS	xxx,xxx,xxx
PROVIDER	NAME	AI algorithm
	PHONE	xxx
	EMAIL	verazhang3333@gmail.com
	MAILING ADDRESS	Guangzhou,China
DATE: Oct.28 2020		AUTHOR: Juan Zhang

PROJECT	
PROJECT NAME	SYMTOM CHECKER
CLIENT	Some Company
BRAND	xx Tech
PRODUCT	Easy Symtom Checker

DESCRIPTION | BACKGROUND

One of the universal problem now is that when people are wondering if they might be sick, they ask the question through web Doc, or from other people on the internet who has similar symptoms. The consequence of this is the web Doc. might not know enough information about the person, so sometimes this person receives inaccurate estimation from the web source, Or get no answer at all. To tackle this problem, the new application , named Easy Symptom Checker, will give better advice and estimation for users. Easy Symptom Checker is a phone application, where people can input their symptoms, and the software will lead them to the next question that will be helpful to find the potential illness they might have contracted. Or in another case, people think they have some kind of disease, they want to know the likelihood of assumed disease, by going through some question on the application, the user will get a percentage of the likelihood of the specific disease.

DATA SOLUTION

Data Source:

This Easy symptom checker is only to estimate the basic surgical disease. Inside of this Application, there are huge amount of the data sets that are been used as training set for our estimator, the data sets are from the domestic hospitals and disease control center, also hospital records from other countries and regions.
the records will be updated as we collecting the latest records from the hospitals.

Data Assumptions:

Question	Why model accuracy from user end dropped?		
Assumptions	Lacking of tranning data sets	Accracy between different kind of disease has large variance.	The questions are too many which cause user give up on checking.
How to fix	Check the how many records have been used for tranning the model, especially some rare disease.	Analysis the average amount of data set is needed for each kind of desease, then find out how much of the symtoms(features) it need to determin a kind of disease.	Check the model accracy while trainning model, it might select too many less relevant symtoms .
Data required	Collect more data from hospitals and disease center.	Require more records from hospital, or suggest coperative hospital collect more symtoms from patients.	Try to use different model to compare and contrast while selecting features , or reduce the complexity of model.

Data limitations:

1. The age of the records is not averagely distributed, some times we need more records for rare diseases.

2. Some hospitals are not willing to share patients' records.

3. Some diseases records are hard to collect in domestic, this can result from some illness are related to geograpical reason.

4. The probability of wrong records, which means the patient is been wrongly diagnosed. This will affect model accuracy.

5. Missing values of the records, also some records that are too old.

Other Constrains:

The worker who do data cleanning job lacking of experieces.

PROJECT TEST PROCESS

1. Using Precision and Recall mothod to test the accuracy of the model, as well as RMSE.

2. Using PR and ROC curve to test the stability of the performance of the model.

3. Using Cross Valivation to determine the devision of the trainning data and test data.

4. Adjust the hyperparameter of the model, then find out if the model is overfiting or underfitting, Analysis the reason then take corespondent measure to fix the problems.

DELIVERABLE MATERIALS

STAKEHOLDERS	
IT AFFECTED TEAMS	xx
NON-IT AFFECTED TEAMS	xx
STEERING COMMITTEE	xx
CUSTOMERS	xx
POTENTIAL / OTHER	xx

TIMELINE	
Oct.28 2020	SOW V1
NOV.23 2020	Data Aquisiton and Understanding; SOW V2
NOV.23 2020	Modelling
NOV. 23 2020	*Prototyping*
DEC.28 2020	*Deployement*