The social determinants of health are the economic and social conditions that influence individual and group differences in health status. Social Determinant of health are conditions in the environments in which people are born, live, learn, work, play, worship and age that affect a wide range of health, functionality and quality of life outcomes and risks.

**Unmet Social needs:** Inadequate or absence of any of the social determinant of health factors necessary to improve health

Insurance payers understands that, there are important factors beyond health care, although only 10% of the National Health Expenditure is spent on tackling these unmet social needs. With a desire to transform the health Industry from one of sick care to wellcare, US healthcare payers aims to address these unmet social needs affecting people with chronic condition(s) on Medicaid and Medicaid-Medicare Insurance plans.

**Solution:** Using 23 different SDOH, this application is predicting whether a person with given social determinants is prone to any chronic condition or not. It can help healthcare payers to understand and anticipate health conditions of their incoming customers based on their unmet social needs.



Step 0: Explain the problem you are going to solve using the selected data. [5 points]

**Step 1.1: Find a public dataset in a domain you like. This dataset must be have the following features. 1- there is a possibility of finding versions of the data. 2- possibility of change in the data. 3- possibility of receiving future updated on the data. 4- have at least two protected features.**

Source of Data- [Sources for Data on SDOH | Social Determinants of Health | CDC](https://www.cdc.gov/socialdeterminants/data/index.htm)

1. there is a possibility of finding versions of the data. - Yes
2. possibility of change in the data. - Yes
3. possibility of receiving future updated on the data. - Yes
4. have at least two protected features. – 2 protected features – Gender and Ethnicity

**Step 1.2: Define some ML metrics to evaluate your model. [5 points]**

**Classification Accuracy**

Classification Accuracy is what we usually mean, when we use the term accuracy. It is the ratio of number of correct predictions to the total number of input samples.

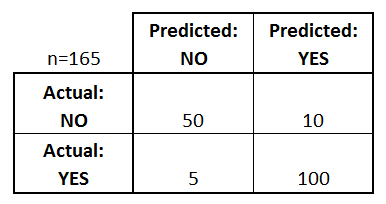


It works well only if there are equal number of samples belonging to each class.

# ****Confusion Matrix****

Confusion Matrix as the name suggests gives us a matrix as output and describes the complete performance of the model.

Lets assume we have a binary classification problem. We have some samples belonging to two classes : YES or NO. Also, we have our own classifier which predicts a class for a given input sample. On testing our model on 165 samples ,we get the following result.



Confusion Matrix

There are 4 important terms :

* **True Positives** : The cases in which we predicted YES and the actual output was also YES.
* **True Negatives** : The cases in which we predicted NO and the actual output was NO.
* **False Positives** : The cases in which we predicted YES and the actual output was NO.
* **False Negatives** : The cases in which we predicted NO and the actual output was YES.

Accuracy for the matrix can be calculated by taking average of the values lying across the**“main diagonal”**i.e



**F1 Score**

*F1 Score is used to measure a test’s accuracy*

F1 Score is the Harmonic Mean between precision and recall. The range for F1 Score is [0, 1]. It tells you how precise your classifier is (how many instances it classifies correctly), as well as how robust it is (it does not miss a significant number of instances).

High precision but lower recall, gives you an extremely accurate, but it then misses a large number of instances that are difficult to classify. The greater the F1 Score, the better is the performance of our model. Mathematically, it can be expressed as :



F1 Score

F1 Score tries to find the balance between precision and recall.

* **Precision :**It is the number of correct positive results divided by the number of positive results predicted by the classifier.



Precision

* **Recall :**It is the number of correct positive results divided by the number of ***all***relevant samples (all samples that should have been identified as positive).



Step 1.3: Define some business metrics to evaluate your model. [5 points]

Step 1.4: Define some software metrics to evaluate your model. [5 points]

**Step 2.1: Describe the dataset objective and the features made you decide on selecting it. [5 points]**

The social determinants of health are the economic and social conditions that influence individual and group differences in health status. Social Determinant of health are conditions in the environments in which people are born, live, learn, work, play, worship and age that affect a wide range of health, functionality and quality of life outcomes and risks.

**Unmet Social needs:** Inadequate or absence of any of the social determinant of health factors necessary to improve health

Insurance payers understands that, there are important factors beyond health care, although only 10% of the National Health Expenditure is spent on tackling these unmet social needs. With a desire to transform the health Industry from one of sick care to wellcare, US healthcare payers aims to address these unmet social needs affecting people with chronic condition(s) on Medicaid and Medicaid-Medicare Insurance plans.

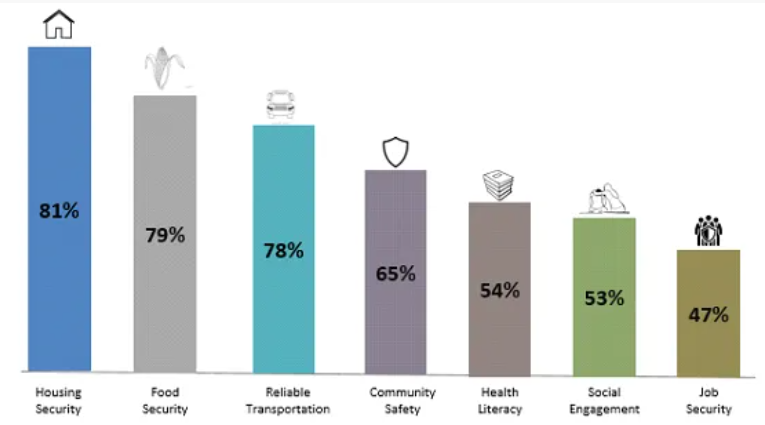
Step 2.1: Describe the quality of dataset using a radar chart with enough explanation. [5 points]

Step 2.2: upload your data in a public repo on https://git-lfs. github.com/. [5 points]

**Step 3: What are the features in your dataset? what is the target variable?. [10 points]**

Target variable – Diagnosed

**Step 4: Among the features what are the features with more predictive value? [5 points]**



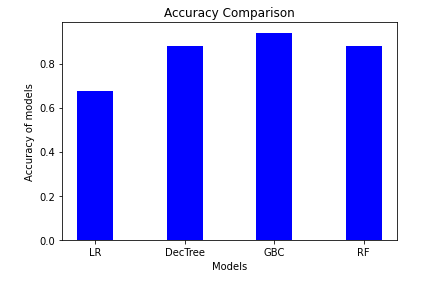
Majority of the respondents with chronic condition(s) consistently ranked Housing security or Food security Reliable transportation as their most pressing social needs.

**Step 5: Identify all protected features? (for example in some domains we can say Gender is a predictive feature. ) [5 points]**

Gender, Ethnicity

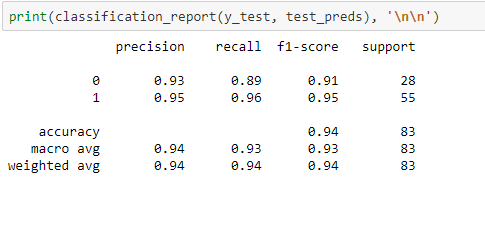
Step 6: Build a model to predict your target value. [5 points]

Step 7: Explain what model you utilized and the reason of choosing it. [5 points]



GBC is giving us best accuracy among all hence we will be going ahead with this model only for final results.

**Step 8: using the identified ML metrics evaluate your model. [5 points]**



Step 9: Perform error analysis on your dataset and try to improve the performance of your solution by investigating the samples in your dataset. [5 points]

Step 10: evaluate the fairness of your model. Use subsets of data to assess the fairness in regards to the protected features [5 points]