Assignment 1

Machine Learning

Question 2

Name: Keyur Kirti Mehta

Professor: Dr. Yuni Xia

1. Machine Learning for Prediction

Machine learning is a method to do data analysis. It's a process to by which different patterns can be learned from the data using without human intervention. There are different algorithms which when run on a data, creates a model. This model is used to predict the pattern observed in the data set. And then to give some concrete result on the future data elements which can be used for business purposes.

As a part of this assignment we used AWS Machine Learning tool which we calculate the prediction using the algorithms. We have to provide the dataset to the Machine learning tool of AWS, then it will create a model based on the data. And gives the evaluation result depending on algorithm. To get the evaluation result aka prediction, we need to provide 1 target value to the model. AWS machine learning uses 70 % of the data to create a model, and remaining 30 % of the data is used to create a test the created model or to give the prediction on that particular model. The prediction is given on the target value which we have specified to the AWS machine learning tool.

2. Dataset:

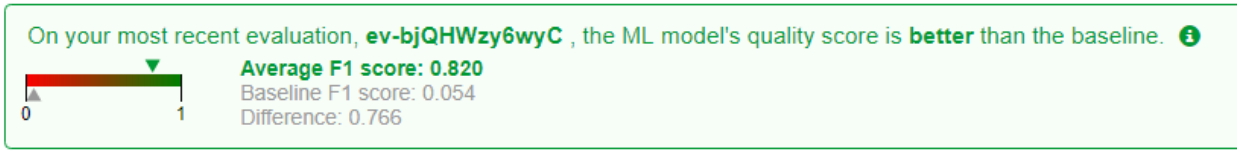
I have used data set of noble prize winners to test the machine learning tool. This dataset contains more than 700 records and more than 15 attributes. The attributes contain both types of data numeric and words.

	A	B	C	D	E	F	G	H	I	J	K	L	
1	Year	Category	Shared	Winner	Individual/ Organizatio	Gender	Place of Birth	Country of Birth	Year of Birth	Age	Country represented for Nobel Prize	Short Description	Description
2	1901	Chemistry		Jacobus Henricus van 't Hoff	Individual	Male	Rotterdam	Netherlands	1852	49	Netherlands	Chemist	Jacobus Henricus van 't H
3	1901	Literature		Sully Prudhomme	Individual	Male	Paris	France	1839	62	France	French Poet And Essayist	René François Armand (S
4	1901	Medicine / Physiolog		Emil Adolf von Behring	Individual	Male	?awice	Germany	1854	47	Germany	German Physiologist	Emil Adolf von Behring (
5	1901	Peace		Frédéric Passy	Individual	Male	Paris	France	1822	79	France	French Economist And Statesman	Frédéric Passy (May 20, 1
6	1901	Peace		Henry Dunant	Individual	Male	Geneva	Switzerland	1828	73	France, Switzerland	Inspiration For Red Cross	Jean Henri Dunant (8 Ma
7	1901	Physics		Wilhelm Röntgen	Individual	Male	Remscheid	Germany	1845	56	Germany		Wilhelm Conrad Röntger
8	1902	Chemistry		Hermann Emil Fischer	Individual	Male	Euskirchen	Germany	1852	50	Germany	German Chemist	Hermann Emil Louis Fisc
9	1902	Literature		Theodor Mommsen	Individual	Male	Garding	Germany	1817	85	Germany	German Classical Scholar, Jurist Ar	Christian Matthias Theoc
10	1902	Medicine / Physiolog		Ronald Ross	Individual	Male	Almora	India	1857	45	India, United Kingdom	British Doctor	Sir Ronald Ross, KCB, FRS
11	1902	Peace		Élie Ducommun	Individual	Male	Geneva	Switzerland	1833	69	Switzerland	Swiss Activist	Élie Ducommun (19 Febr
12	1902	Peace		Charles Albert Gobat	Individual	Male	Tramelan	Switzerland	1843	59	Switzerland	Swiss Politician And Lawyer	Charles Albert Gobat (M

3. Result:

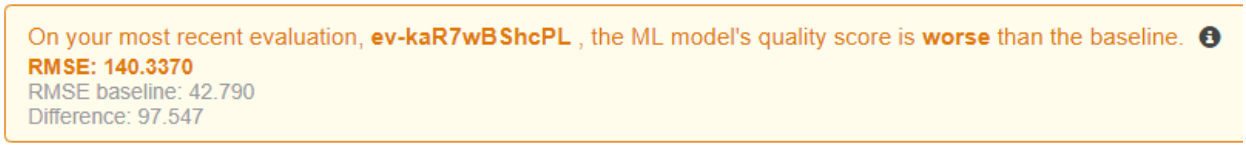
I test the prediction on this data twice for 2 different target attributes. One attribute was the category in which laureates have received the Nobel Prize and the second parameter was the year in which these laureates have born. The prediction result for both the case was different whose description is given as below:

a. Target: Category



In this model, I used target as “Category” field which has 5 different data value. AWS has used F1 parameter to calculate the prediction. This is statistical analysis measure, it uses precision and recall method to evaluate it. Precision is the number of correct positive results divided by the number of all positive results, and recall is the number of correct positive results divided by the number of all relevant samples. F1 scores lies between 0 and 1. Greater the score, model is good for prediction of the target field. In this case, we obtained F1 score of 0.820 which shows the very good prediction quality of ML model.

b. Target: Year of Birth



In this model, I used target as “Year of Birth” field which is numeric data field. AWS has used RMSE parameter to calculate the prediction. It is the sum, over all the data points, of the square of the difference between the predicted and actual target variables, divided by the number of data points. The smaller the value of RSME, better the model is. In this case, RMSE value we obtained was 140.33 which is far greater than the base line value. So, this model fails to predict the year of birth of laureates who will won the Nobel Prize or have already won the prize.

4. *Conclusion*

Using the Machine Learning, we can predict the pattern based on the data. However, we cannot guarantee that the model will always be correct or give right prediction. As seen, for same dataset, we obtained different results on different fields. So we require stronger algorithm to have error free prediction of the data.