AMERICAN INTERNATIONAL
UNIVERSITY-BANGLADESH

Assignment-Mid Term

NAME: MD. AKHTARUZZAMAN EMON

ID: 18-36074-1

COURSE: DATA WAREHOUSING AND DATA MINING

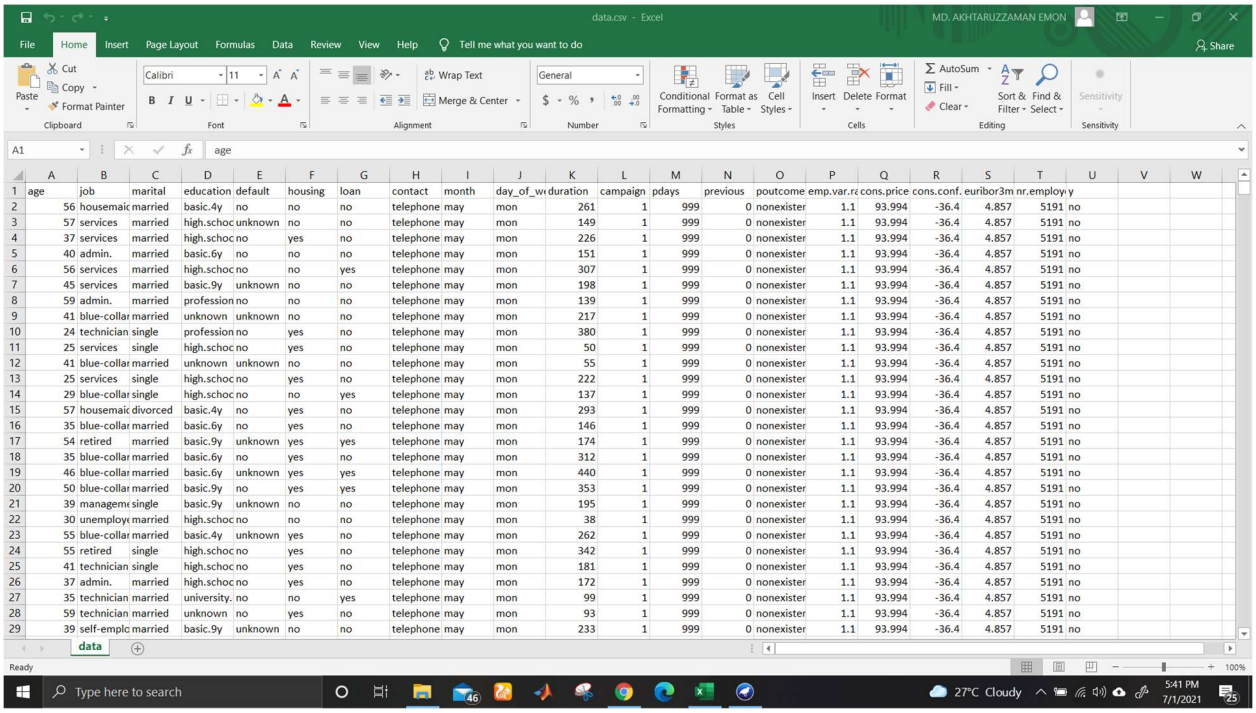
SECTION: A

SUBMITTED TO- DR. MD. MAHBUB CHOWDHURY MISHU

Introduction:

A data mining function called classification allocates objects in a collection to desired groups or classes. Classification's purpose is to correctly anticipate the target class for each case in the data. In this report I used K-Nearest Neighbor (KNN) procedure. The K-Nearest Neighbor algorithm is an example of a "lazy learner," meaning it does not build a model using the training set until the data set is queried. For this report I used a data set of a bank. There are many data of people in this data set. We use Weka software to find the result.

Data Set:



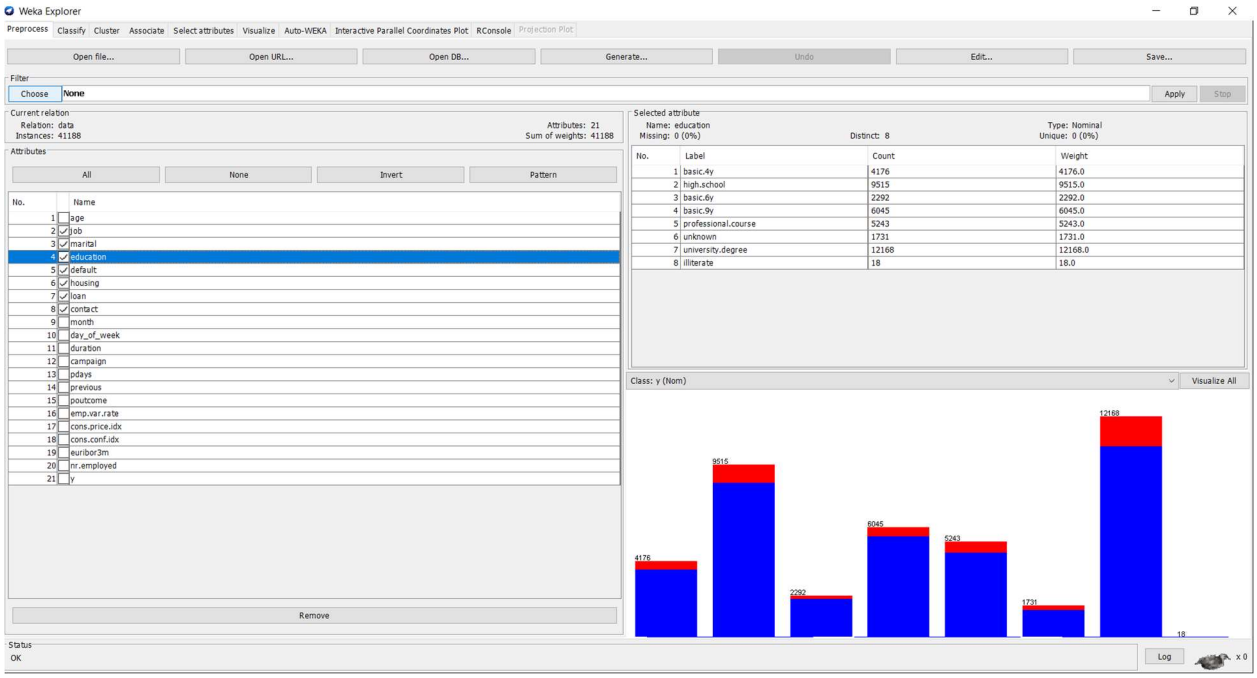
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	age	job	marital	education	default	housing	loan	contact	month	day_of_w	duration	campaign	pdays	previous	poutcome	emp.var.r	cons.price	cons.conf.	euribor3m	nr.employ	y
2	56	housemaic	married	basic.4y	no	no	no	telephone	may	mon	261	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
3	57	services	married	high.schoc	unknown	no	no	telephone	may	mon	149	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
4	37	services	married	high.schoc	no	yes	no	telephone	may	mon	226	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
5	40	admin.	married	basic.6y	no	no	no	telephone	may	mon	151	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
6	56	services	married	high.schoc	no	no	yes	telephone	may	mon	307	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
7	45	services	married	basic.9y	unknown	no	no	telephone	may	mon	198	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
8	59	admin.	married	profession	no	no	no	telephone	may	mon	139	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
9	41	blue-collar	married	unknown	unknown	no	no	telephone	may	mon	217	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
10	24	technician	single	profession	no	yes	no	telephone	may	mon	380	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
11	25	services	single	high.schoc	no	yes	no	telephone	may	mon	50	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
12	41	blue-collar	married	unknown	unknown	no	no	telephone	may	mon	55	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
13	25	services	single	high.schoc	no	yes	no	telephone	may	mon	222	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
14	29	blue-collar	single	high.schoc	no	no	yes	telephone	may	mon	137	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
15	57	housemaic	divorced	basic.4y	no	yes	no	telephone	may	mon	293	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
16	35	blue-collar	married	basic.6y	no	yes	no	telephone	may	mon	146	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
17	54	retired	married	basic.9y	unknown	yes	yes	telephone	may	mon	174	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
18	35	blue-collar	married	basic.6y	no	yes	no	telephone	may	mon	312	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
19	46	blue-collar	married	basic.6y	unknown	yes	yes	telephone	may	mon	440	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
20	50	blue-collar	married	basic.9y	no	yes	yes	telephone	may	mon	353	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
21	39	manageme	single	basic.9y	unknown	no	no	telephone	may	mon	195	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
22	30	unemploye	married	high.schoc	no	no	no	telephone	may	mon	38	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
23	55	blue-collar	married	basic.4y	unknown	yes	no	telephone	may	mon	262	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
24	55	retired	single	high.schoc	no	yes	no	telephone	may	mon	342	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
25	41	technician	single	high.schoc	no	yes	no	telephone	may	mon	181	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
26	37	admin.	married	high.schoc	no	yes	no	telephone	may	mon	172	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
27	35	technician	married	university.	no	no	yes	telephone	may	mon	99	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
28	59	technician	married	unknown	no	yes	no	telephone	may	mon	93	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
29	39	self-emplc	married	basic.9y	unknown	no	no	telephone	may	mon	233	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no

In this data set there are 21 columns and 41189 rows.

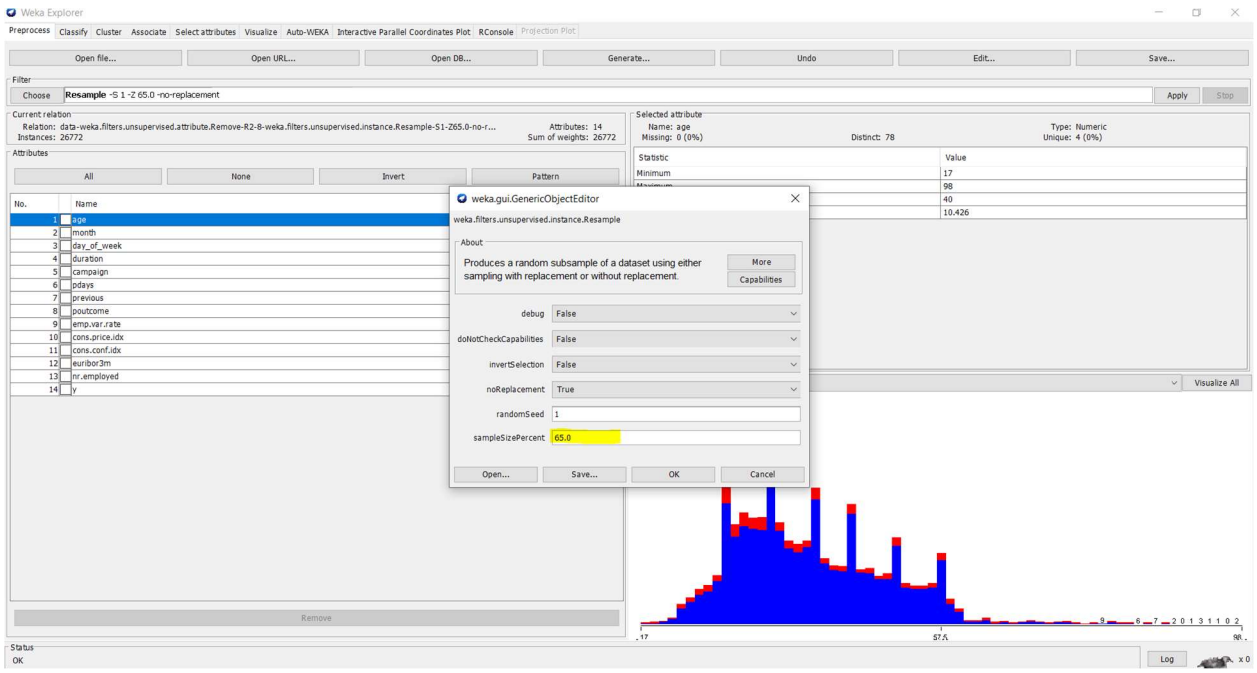
Dataset reference: https://www.kaggle.com/brijbhushannanda1979/bank-data

Procedure:

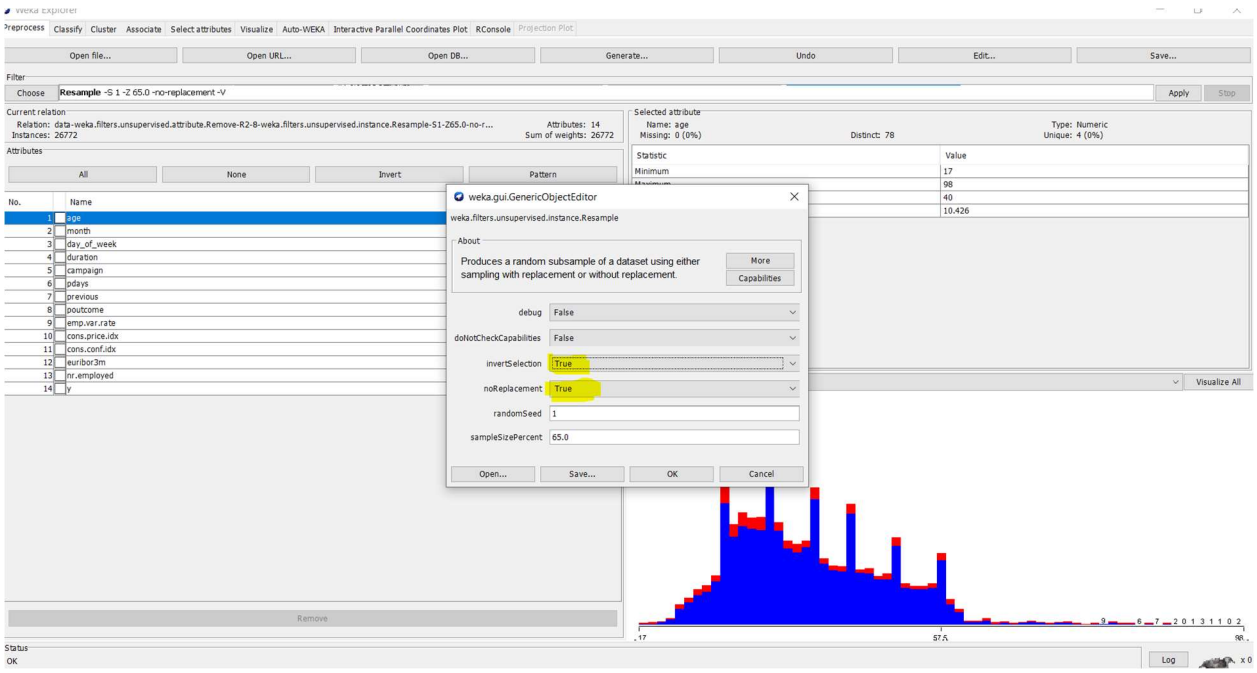
At first, we need to open the data set file in Weka software. Then we need delete unnecessary data from the data set. Then we have to follow the algorithms of K-Nearest Neighbor algorithm, Naïve Bayes.



Then we have to set the data in two set which are training set and test set.



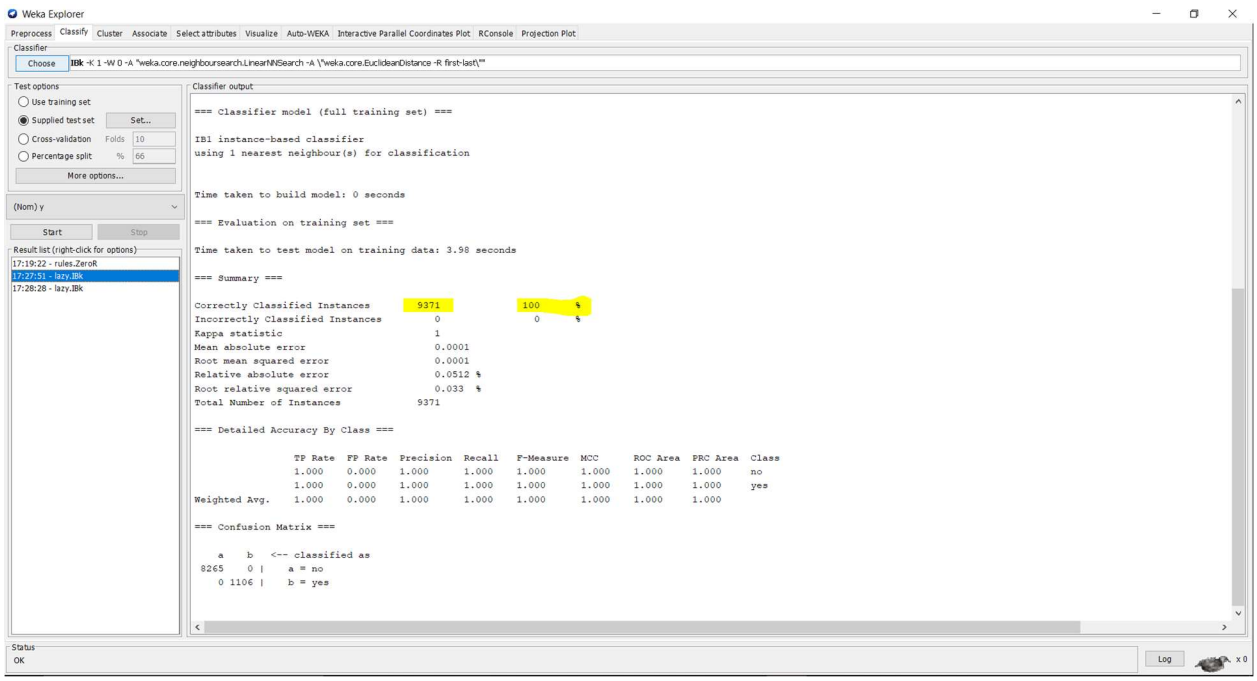
Here we set 65% of data for training set.



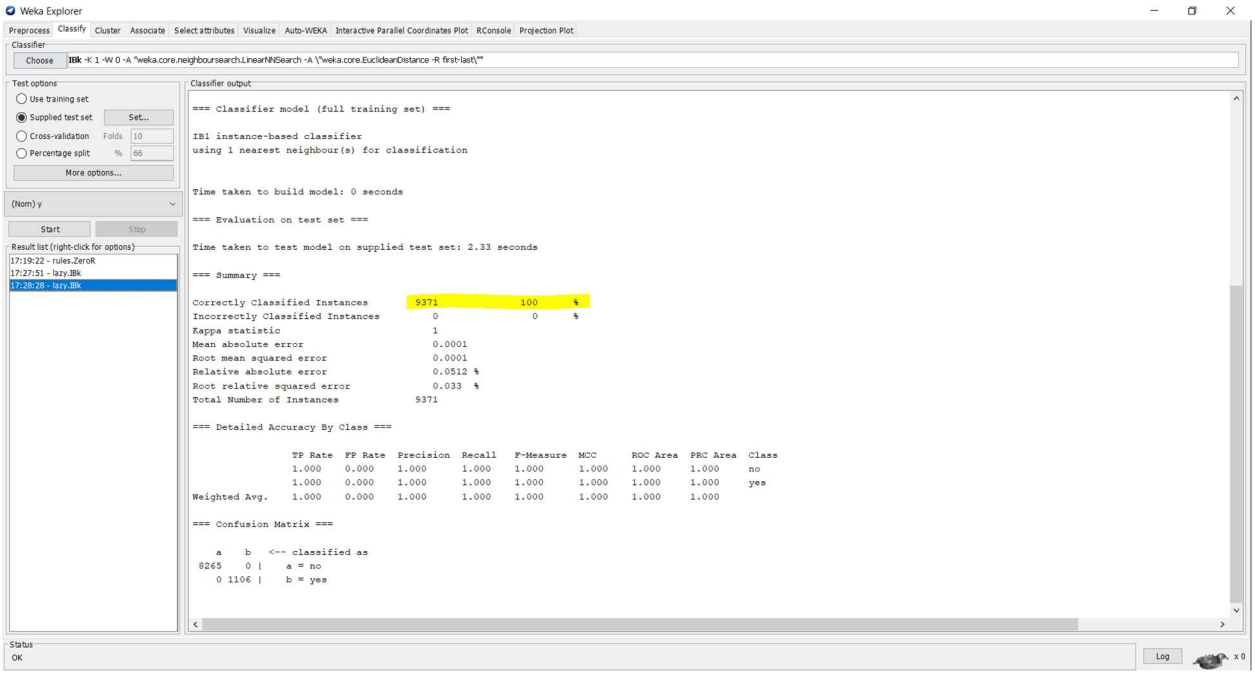
Here we set 55% of data for test set.

Result of K-Nearest Neighbor (KNN):

Training set:

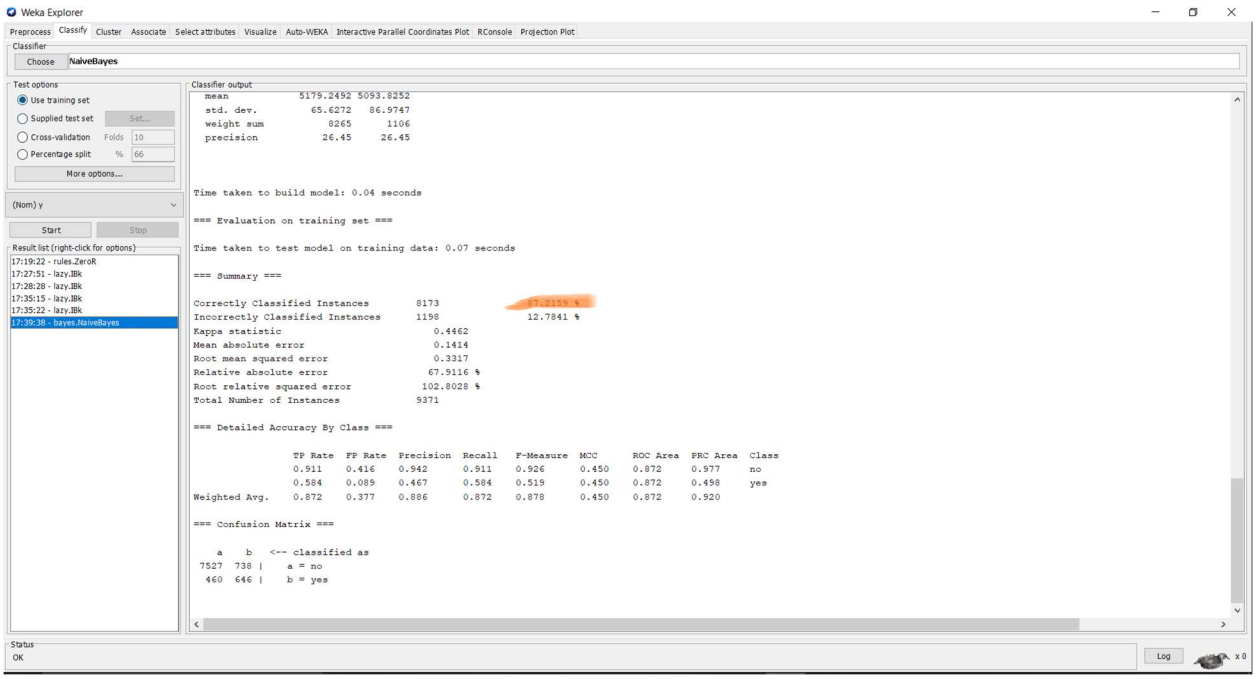


Test set:

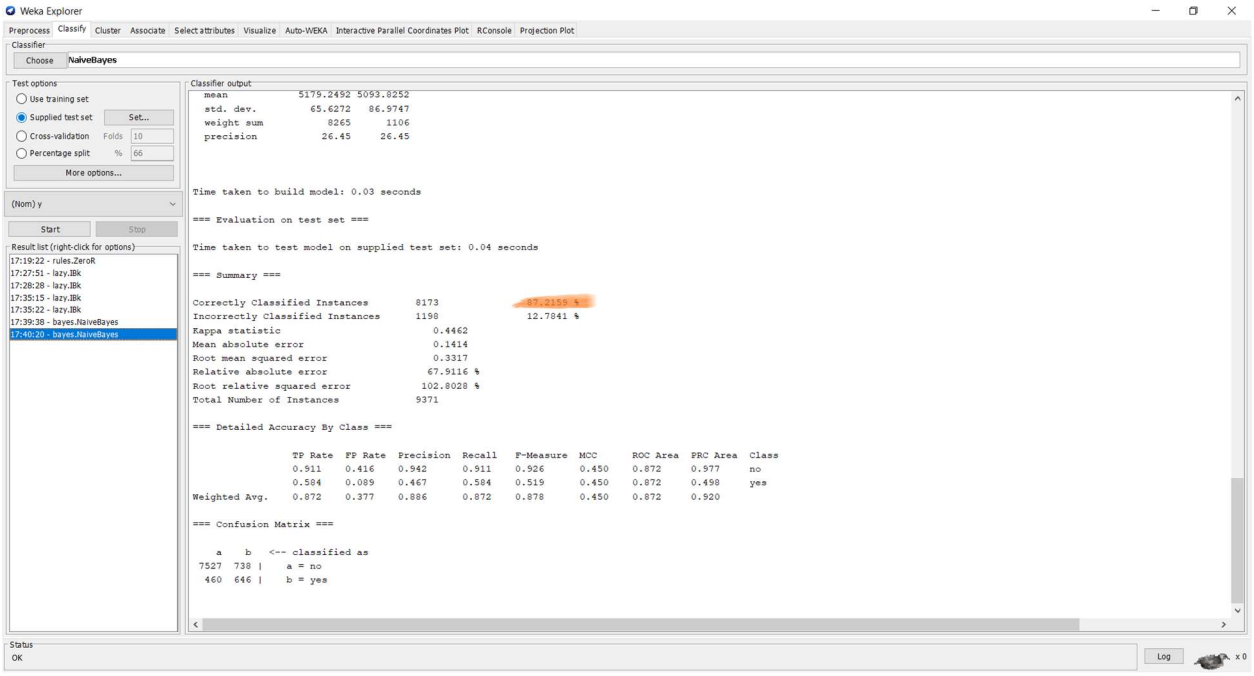


Result of Naïve Bayes:

Training set:



Test set:



Results:

By using these particular algorithms, we got some different results. We used K-Nearest Neighbor algorithm, Naïve Bayes procedure. From these K-Nearest Neighbor (KNN) procedure is the best algorithm for this problem. I got 100% accuracy of KNN in training set and also got 100% in test set. On the other hand, in Naïve Bayes procedure I got 87.22% accuracy in training set and test set. So, I prefer the K-Nearest Neighbor procedure.