Intelligent Systems
Dr. Octavio Loyola-Gonzalez

Nicolas Gonzalez Albo Mendez
A01730763

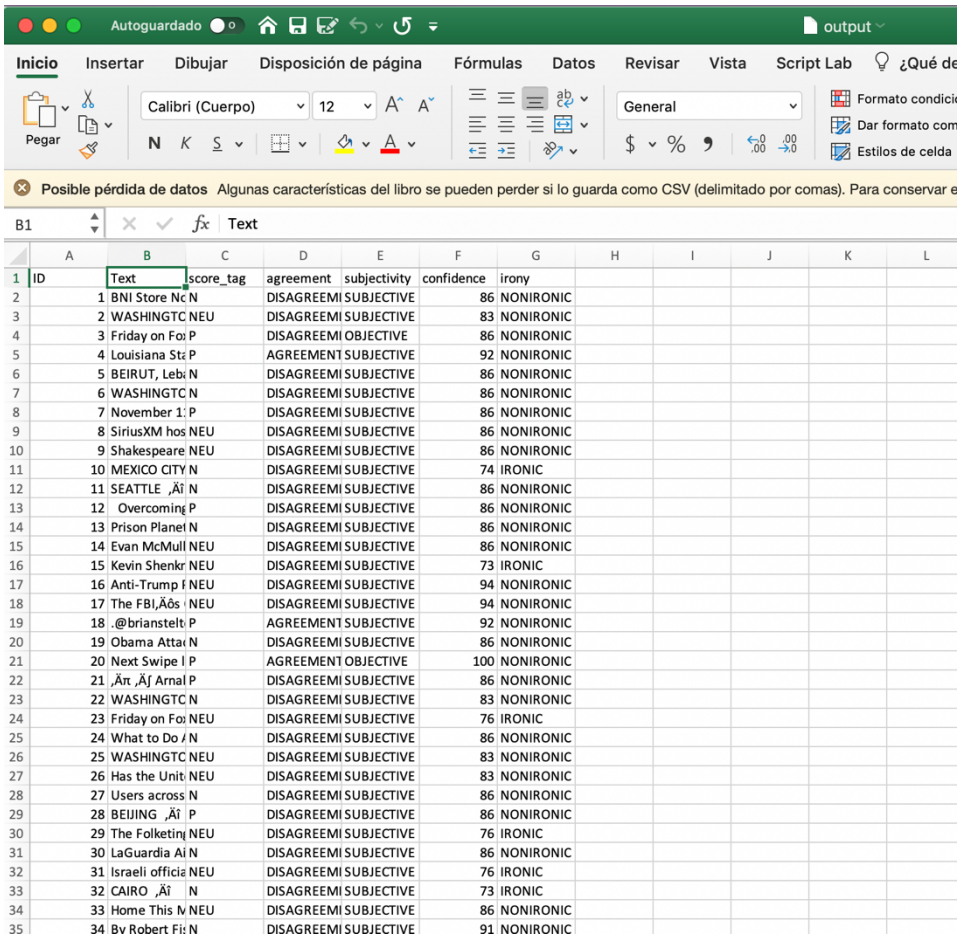
H4 – Testing two pattern-based classifiers

Introduction

Pattern based classifiers are tools that help us to determine the class of a certain object depending on the patterns that can be identified. Examples of pattern-based classifiers are J48, RandomForest, REPTree and FURIA.

Development

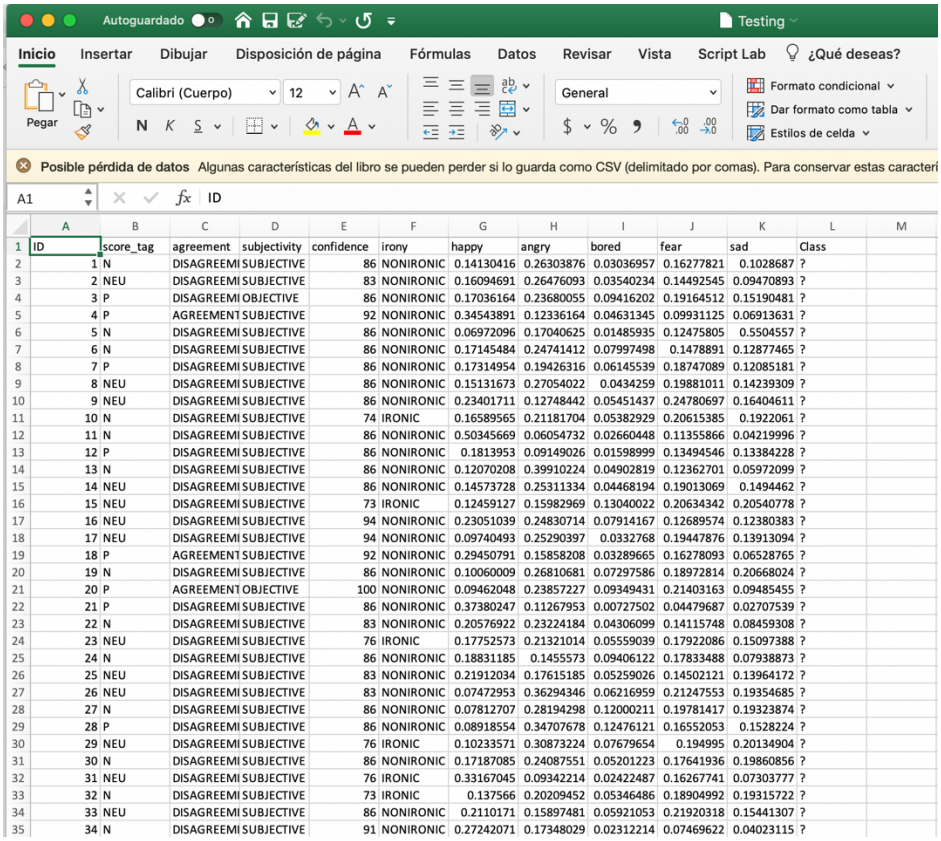
Using the python script from HW1, sentiments were extracted from the Testing.csv, emotions were not extracted because they were extracted in the previous homework. The columns from output were pasted into the Testing.csv of the previous homework. ID and Text columns were removed. Also, a new Column named “Class” was added and filled with the symbol “?”.



	A	B	C	D	E	F	G
1	ID	Text	score_tag	agreement	subjectivity	confidence	irony
2	1	BNI Store Nc	N	DISAGREEMI	SUBJECTIVE	86	NONIRONIC
3	2	WASHINGTC	NEU	DISAGREEMI	SUBJECTIVE	83	NONIRONIC
4	3	Friday on Fo	P	DISAGREEMI	OBJECTIVE	86	NONIRONIC
5	4	Louisiana St	P	AGREEMENT	SUBJECTIVE	92	NONIRONIC
6	5	BEIRUT, Leb	N	DISAGREEMI	SUBJECTIVE	86	NONIRONIC
7	6	WASHINGTC	N	DISAGREEMI	SUBJECTIVE	86	NONIRONIC
8	7	November 1	P	DISAGREEMI	SUBJECTIVE	86	NONIRONIC
9	8	SiriusXM hos	NEU	DISAGREEMI	SUBJECTIVE	86	NONIRONIC
10	9	Shakespeare	NEU	DISAGREEMI	SUBJECTIVE	86	NONIRONIC
11	10	MEXICO CITY	N	DISAGREEMI	SUBJECTIVE	74	IRONIC
12	11	SEATTLE ,Äî	N	DISAGREEMI	SUBJECTIVE	86	NONIRONIC
13	12	Overcoming	P	DISAGREEMI	SUBJECTIVE	86	NONIRONIC
14	13	Prison Planet	N	DISAGREEMI	SUBJECTIVE	86	NONIRONIC
15	14	Evan McMull	NEU	DISAGREEMI	SUBJECTIVE	86	NONIRONIC
16	15	Kevin Shenkr	NEU	DISAGREEMI	SUBJECTIVE	73	IRONIC
17	16	Anti-Trump f	NEU	DISAGREEMI	SUBJECTIVE	94	NONIRONIC
18	17	The FBI,Äôs	NEU	DISAGREEMI	SUBJECTIVE	94	NONIRONIC
19	18	.@brianstelt	P	AGREEMENT	SUBJECTIVE	92	NONIRONIC
20	19	Obama Atta	N	DISAGREEMI	SUBJECTIVE	86	NONIRONIC
21	20	Next Swipe I	P	AGREEMENT	OBJECTIVE	100	NONIRONIC
22	21	‚Äπ ‚Äî Arnal	P	DISAGREEMI	SUBJECTIVE	86	NONIRONIC
23	22	WASHINGTC	N	DISAGREEMI	SUBJECTIVE	83	NONIRONIC
24	23	Friday on Fo	NEU	DISAGREEMI	SUBJECTIVE	76	IRONIC
25	24	What to Do /	N	DISAGREEMI	SUBJECTIVE	86	NONIRONIC
26	25	WASHINGTC	NEU	DISAGREEMI	SUBJECTIVE	83	NONIRONIC
27	26	Has the Unit	NEU	DISAGREEMI	SUBJECTIVE	83	NONIRONIC
28	27	Users across	N	DISAGREEMI	SUBJECTIVE	86	NONIRONIC
29	28	BEIJING ,Äî	P	DISAGREEMI	SUBJECTIVE	86	NONIRONIC
30	29	The Folketin	NEU	DISAGREEMI	SUBJECTIVE	76	IRONIC
31	30	LaGuardia Ai	N	DISAGREEMI	SUBJECTIVE	86	NONIRONIC
32	31	Israeli officia	NEU	DISAGREEMI	SUBJECTIVE	76	IRONIC
33	32	CAIRO ,Äî	N	DISAGREEMI	SUBJECTIVE	73	IRONIC
34	33	Home This M	NEU	DISAGREEMI	SUBJECTIVE	86	NONIRONIC
35	34	By Robert Fi	N	DISAGREEMI	SUBJECTIVE	91	NONIRONIC

Figure 1. output.csv features are shown.

Nicolas Gonzalez Albo Mendez
A01730763



Autoguardado Testing

Inicio Insertar Dibujar Disposición de página Fórmulas Datos Revisar Vista Script Lab ¿Qué deseas?

Pegar Calibri (Cuerpo) 12 General Formato condicional Dar formato como tabla Estilos de celda

Posible pérdida de datos Algunas características del libro se pueden perder si lo guarda como CSV (delimitado por comas). Para conservar estas caracteri

A1 ID

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	ID	score_tag	agreement	subjectivity	confidence	irony	happy	angry	bored	fear	sad	Class	
2	1	N	DISAGREEMI	SUBJECTIVE	86	NONIRONIC	0.14130416	0.26303876	0.03036957	0.16277821	0.1028687	?	
3	2	NEU	DISAGREEMI	SUBJECTIVE	83	NONIRONIC	0.16094691	0.26476093	0.03540234	0.14492545	0.09470893	?	
4	3	P	DISAGREEMI	OBJECTIVE	86	NONIRONIC	0.17036164	0.23680055	0.09416202	0.19164512	0.15190481	?	
5	4	P	AGREEMENT	SUBJECTIVE	92	NONIRONIC	0.34543891	0.12336164	0.04631345	0.09931125	0.06913631	?	
6	5	N	DISAGREEMI	SUBJECTIVE	86	NONIRONIC	0.06972096	0.17040625	0.01485935	0.12475805	0.5504557	?	
7	6	N	DISAGREEMI	SUBJECTIVE	86	NONIRONIC	0.17145484	0.24741412	0.07997498	0.1478891	0.12877465	?	
8	7	P	DISAGREEMI	SUBJECTIVE	86	NONIRONIC	0.17314954	0.19426316	0.06145539	0.18747089	0.12085181	?	
9	8	NEU	DISAGREEMI	SUBJECTIVE	86	NONIRONIC	0.15131673	0.27054022	0.0434259	0.19881011	0.14239309	?	
10	9	NEU	DISAGREEMI	SUBJECTIVE	86	NONIRONIC	0.23401711	0.12748442	0.05451437	0.24780697	0.16404611	?	
11	10	N	DISAGREEMI	SUBJECTIVE	74	IRONIC	0.16589565	0.21181704	0.05382929	0.20615385	0.1922061	?	
12	11	N	DISAGREEMI	SUBJECTIVE	86	NONIRONIC	0.50345669	0.06054732	0.02660448	0.11355866	0.04219996	?	
13	12	P	DISAGREEMI	SUBJECTIVE	86	NONIRONIC	0.1813953	0.09149026	0.01598999	0.13494546	0.13384228	?	
14	13	N	DISAGREEMI	SUBJECTIVE	86	NONIRONIC	0.12070208	0.39910224	0.04902819	0.12362701	0.05972099	?	
15	14	NEU	DISAGREEMI	SUBJECTIVE	86	NONIRONIC	0.14573728	0.25311334	0.04468194	0.19013069	0.1494462	?	
16	15	NEU	DISAGREEMI	SUBJECTIVE	73	IRONIC	0.12459127	0.15982969	0.13040022	0.20634342	0.20540778	?	
17	16	NEU	DISAGREEMI	SUBJECTIVE	94	NONIRONIC	0.23051039	0.24830714	0.07914167	0.12689574	0.12380383	?	
18	17	NEU	DISAGREEMI	SUBJECTIVE	94	NONIRONIC	0.09740493	0.25290397	0.0332768	0.19447876	0.13913094	?	
19	18	P	AGREEMENT	SUBJECTIVE	92	NONIRONIC	0.29450791	0.15858208	0.03289665	0.16278093	0.06528765	?	
20	19	N	DISAGREEMI	SUBJECTIVE	86	NONIRONIC	0.10060009	0.26810681	0.07297586	0.18972814	0.20668024	?	
21	20	P	AGREEMENT	OBJECTIVE	100	NONIRONIC	0.09462048	0.23857227	0.09349431	0.21403163	0.09485455	?	
22	21	P	DISAGREEMI	SUBJECTIVE	86	NONIRONIC	0.37380247	0.11267953	0.00727502	0.04479687	0.02707539	?	
23	22	N	DISAGREEMI	SUBJECTIVE	83	NONIRONIC	0.20576922	0.23224184	0.04306099	0.14115748	0.08459308	?	
24	23	NEU	DISAGREEMI	SUBJECTIVE	76	IRONIC	0.17752573	0.21321014	0.05559039	0.17922086	0.15097388	?	
25	24	N	DISAGREEMI	SUBJECTIVE	86	NONIRONIC	0.18831185	0.1455573	0.09406122	0.17833488	0.07938873	?	
26	25	NEU	DISAGREEMI	SUBJECTIVE	83	NONIRONIC	0.21912034	0.17615185	0.05259026	0.14502121	0.13964172	?	
27	26	NEU	DISAGREEMI	SUBJECTIVE	83	NONIRONIC	0.07472953	0.36294346	0.06216959	0.21247553	0.19354685	?	
28	27	N	DISAGREEMI	SUBJECTIVE	86	NONIRONIC	0.07812707	0.28194298	0.12000211	0.19781417	0.19323874	?	
29	28	P	DISAGREEMI	SUBJECTIVE	86	NONIRONIC	0.08918554	0.34707678	0.12476121	0.16552053	0.1528224	?	
30	29	NEU	DISAGREEMI	SUBJECTIVE	76	IRONIC	0.10233571	0.30873224	0.07679654	0.194995	0.20134904	?	
31	30	N	DISAGREEMI	SUBJECTIVE	86	NONIRONIC	0.17187085	0.24087551	0.05201223	0.17641936	0.19860856	?	
32	31	NEU	DISAGREEMI	SUBJECTIVE	76	IRONIC	0.33167045	0.09342214	0.02422487	0.16267741	0.07303777	?	
33	32	N	DISAGREEMI	SUBJECTIVE	73	IRONIC	0.137566	0.20209452	0.05346486	0.18904992	0.19315722	?	
34	33	NEU	DISAGREEMI	SUBJECTIVE	86	NONIRONIC	0.2110171	0.15897481	0.05921053	0.21920318	0.15441307	?	
35	34	N	DISAGREEMI	SUBJECTIVE	91	NONIRONIC	0.27242071	0.17348029	0.02312214	0.07469622	0.04023115	?	

Figure 2. Modified Testing.csv showing the new sentiment and previous emotion's features

Then in Weka, both Training.csv and Testing.csv were changed to arff format, it is important to clarify that in both files, The class type was changed to nominal.

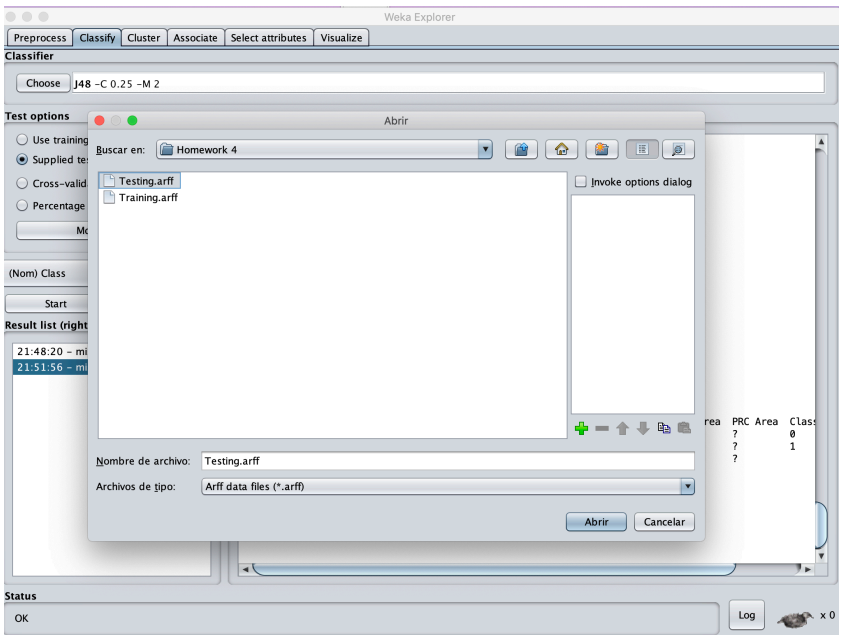


Figure 3. Testing.arff is chosen as the test set.

After all the above steps are made, we can pick the classifier that we want, for example J48, Random Forest, FURIA or REPTree.

Results

J48

In this case, J48 was chosen as the classifier. The output is generated in CSV format.

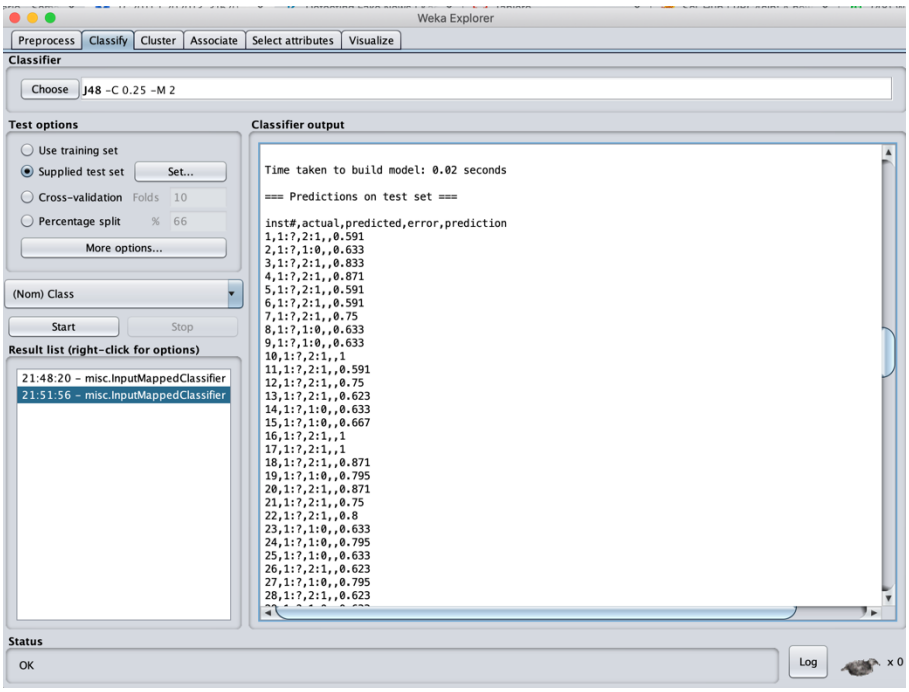


Figure 4. J48 output in CSV format.

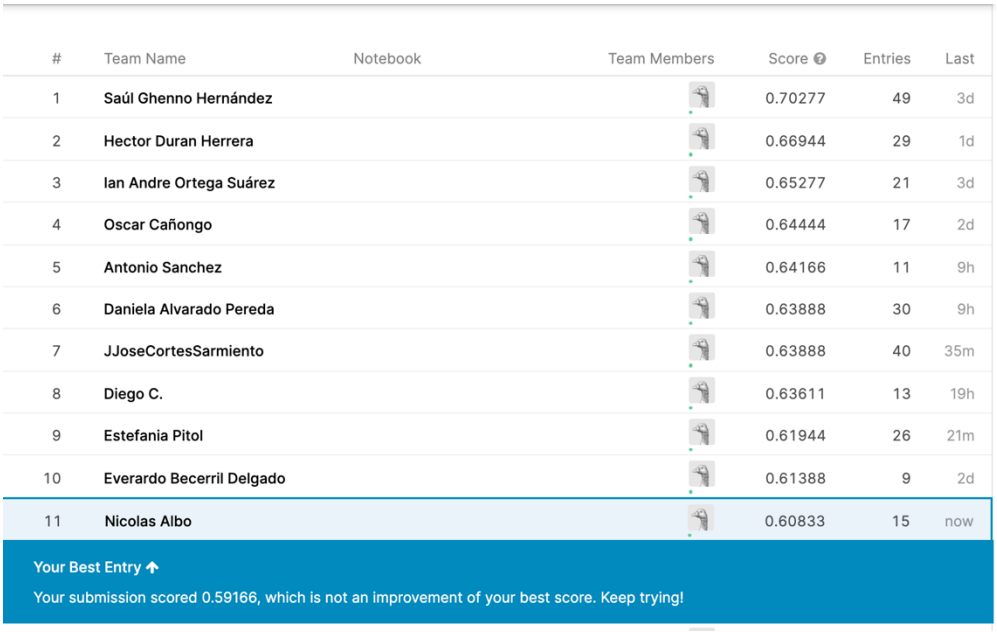
The ouput from Weka was copied into a new file named J48.csv and thanks to a Bash script that I developed; the file was given the format required for uploading to Kaggle. In fact, this Bash script is also used for formatting the following output CSVs that Weka generate for the other classifiers.

filter.sh

```
1  #!/bin/bash
2
3  cat $1 | tr "," ":" | cut -d: -f1,5 | tr ":" "," > output.csv
4
5
```

Figure 5. Simple Bash script that deletes the unnecessary columns of the original J48.csv

Then the new J48.csv is uploaded to Kaggle for its evaluation.



#	Team Name	Notebook	Team Members	Score	Entries	Last
1	Saúl Ghenno Hernández			0.70277	49	3d
2	Hector Duran Herrera			0.66944	29	1d
3	Ian Andre Ortega Suárez			0.65277	21	3d
4	Oscar Cañongo			0.64444	17	2d
5	Antonio Sanchez			0.64166	11	9h
6	Daniela Alvarado Pereda			0.63888	30	9h
7	JJoseCortesSarmiento			0.63888	40	35m
8	Diego C.			0.63611	13	19h
9	Estefania Pitol			0.61944	26	21m
10	Everardo Becerril Delgado			0.61388	9	2d
11	Nicolas Albo			0.60833	15	now

Your Best Entry ↑

Your submission scored 0.59166, which is not an improvement of your best score. Keep trying!

Figure 6. J48 result in Kaggle.

As it can be observed, the score obtained with this classifier couldn't beat the score of the last entry, that was a k-NN classifier from H3.

RandomForest

In this case, RandomForest classifier was chosen. The output is generated in CSV format.

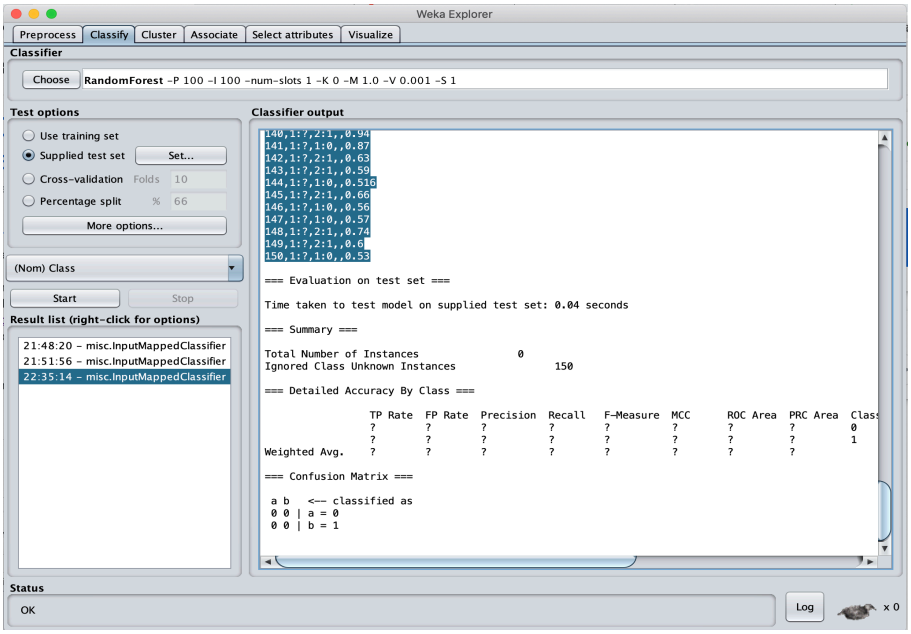
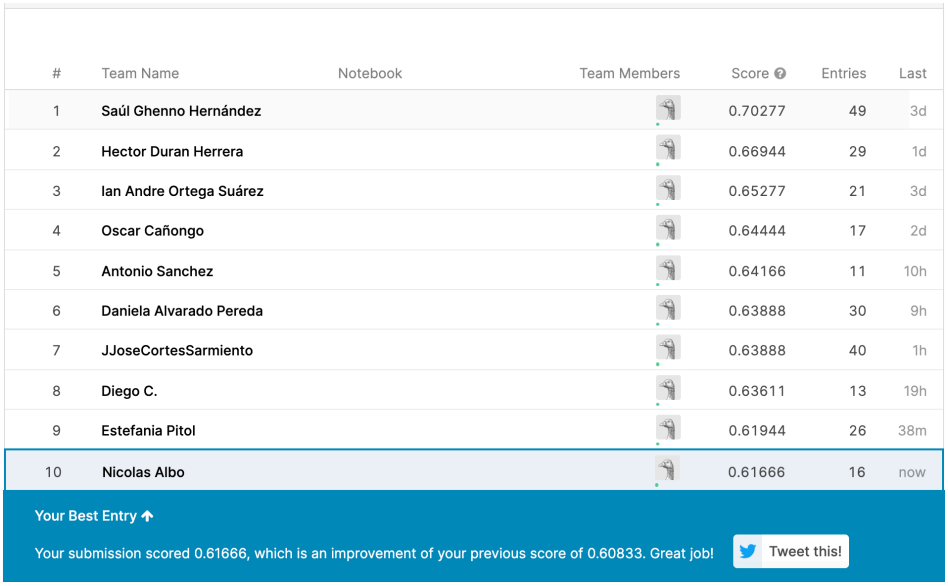


Figure 7. RandomForest output in CSV format.

Then RandomForest.csv is uploaded to Kaggle for its evaluation.



#	Team Name	Notebook	Team Members	Score	Entries	Last
1	Saúl Ghenno Hernández			0.70277	49	3d
2	Hector Duran Herrera			0.66944	29	1d
3	Ian Andre Ortega Suárez			0.65277	21	3d
4	Oscar Cañongo			0.64444	17	2d
5	Antonio Sanchez			0.64166	11	10h
6	Daniela Alvarado Pereda			0.63888	30	9h
7	JJoseCortesSarmiento			0.63888	40	1h
8	Diego C.			0.63611	13	19h
9	Estefania Pitol			0.61944	26	38m
10	Nicolas Albo			0.61666	16	now

Your Best Entry

Your submission scored 0.61666, which is an improvement of your previous score of 0.60833. Great job! Tweet this!

Figure 8. RandomForest result in Kaggle.

As it can be appreciated, the RandomForest classifier performed better than any other classifier that I have used in the past, therefore resulting in my best entry of all time.

REPTree

In this case, REPTree classifier was chosen. The output is generated in CSV format.

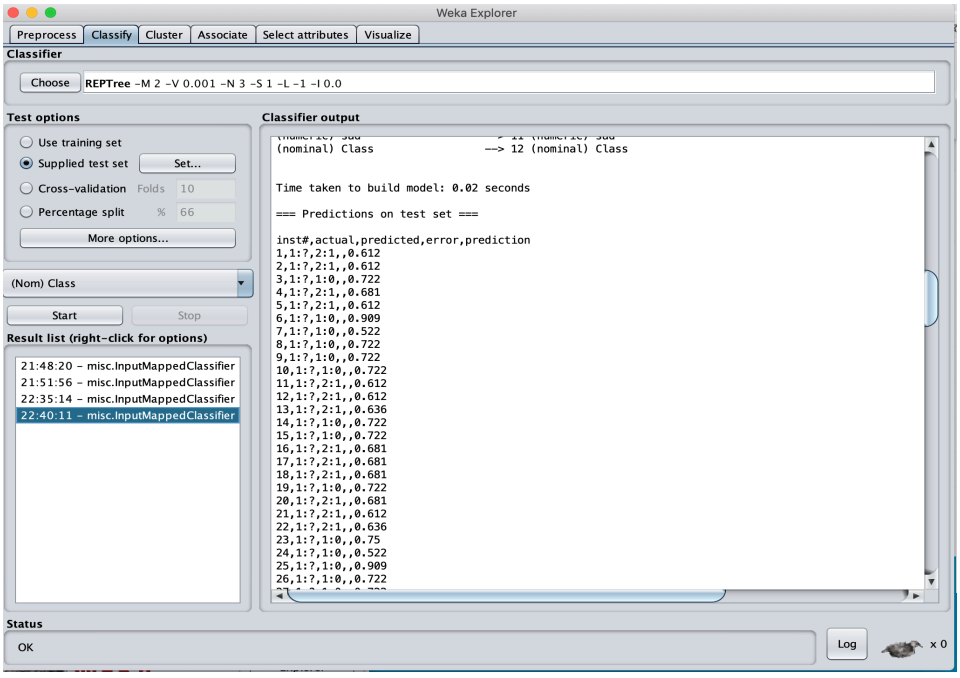
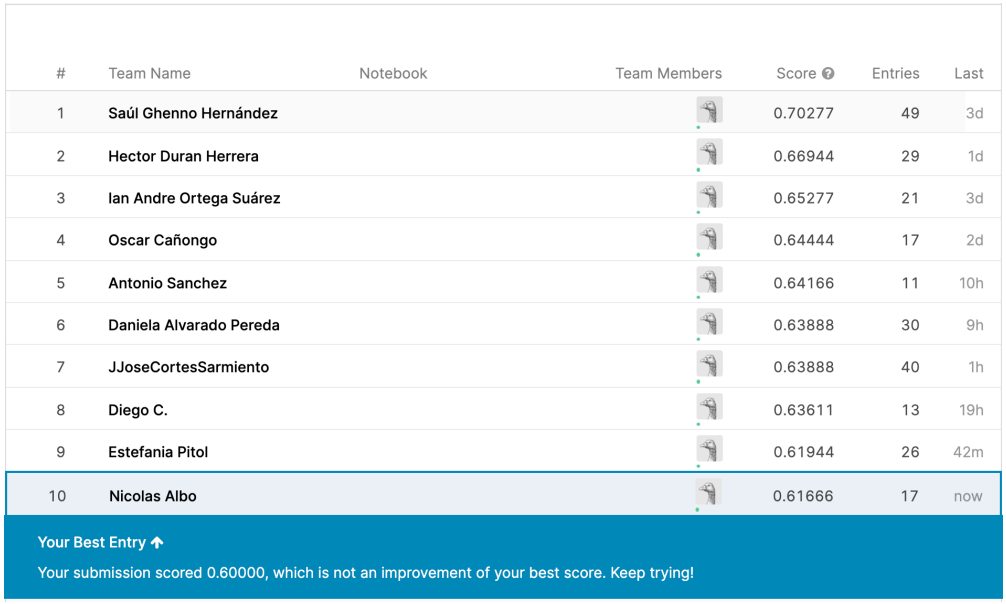


Figure 9. REPTree output in CSV format.

Then REPTree.csv is uploaded to Kaggle for its evaluation.

Nicolas Gonzalez Albo Mendez
A01730763



#	Team Name	Notebook	Team Members	Score	Entries	Last
1	Saúl Ghenno Hernández			0.70277	49	3d
2	Hector Duran Herrera			0.66944	29	1d
3	Ian Andre Ortega Suárez			0.65277	21	3d
4	Oscar Cañongo			0.64444	17	2d
5	Antonio Sanchez			0.64166	11	10h
6	Daniela Alvarado Pereda			0.63888	30	9h
7	JJoseCortesSarmiento			0.63888	40	1h
8	Diego C.			0.63611	13	19h
9	Estefania Pitol			0.61944	26	42m
10	Nicolas Albo			0.61666	17	now

As it can be appreciated, REPTree classifier did not had a results a good as the one generated by the RandomForest classifier, so my position in the Leaderboard remains equal.

Conclusion

Pattern-based classifiers are more useful than k-NN classifiers in certain situations. In this case, the RandomForest classifier performed better than the k-NN classifier that gave me my previous highest score in Kaggle, but, the other two Pattern-based classifiers that I used in this homework, could not beat the score produced by my k-NN classifier.