Cirrhosis: Liver transplant

NARVAEZ CAO ROMERO ALDO

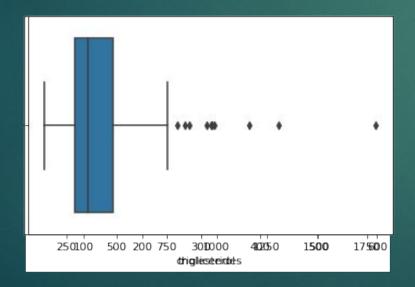
The goal of this project is taking as target the variable Status, and use it to predict if a person is candidate to receive a transplant or not

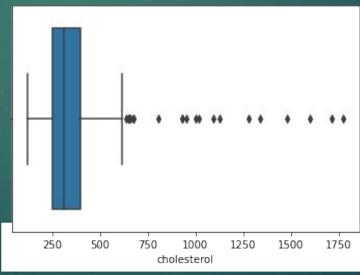
duration	status	drug	age	sex	ascites	hepatomology	spiders	edema	bilirubin	cholesterol	albumin	copper	phosphatase	SGOT	triglicerides	platelets	prothrombin	stage
400	2	1	21464	1	1	1	1	1.0	14.5	261.0	2.60	156.0	1718.0	137.95	172.0	190.0	12.2	4
4500	0	1	20617	1	0	1	1	0.0	1.1	302.0	4.14	54.0	7394.8	113.52	88.0	221.0	10.6	3
1012	2	1	25594	0	0	0	0	0.5	1.4	176.0	3.48	210.0	516.0	96.10	55.0	151.0	12.0	4
1925	2	1	19994	1	0	1	1	0.5	1.8	244.0	2.54	64.0	6121.8	60.63	92.0	183.0	10.3	4
1504	1	2	13918	1	0	1	1	0.0	3.4	279.0	3.53	143.0	671.0	113.15	72.0	136.0	10.9	3

There was a total of 312 rows and 20 columns

Cleaning and EDA

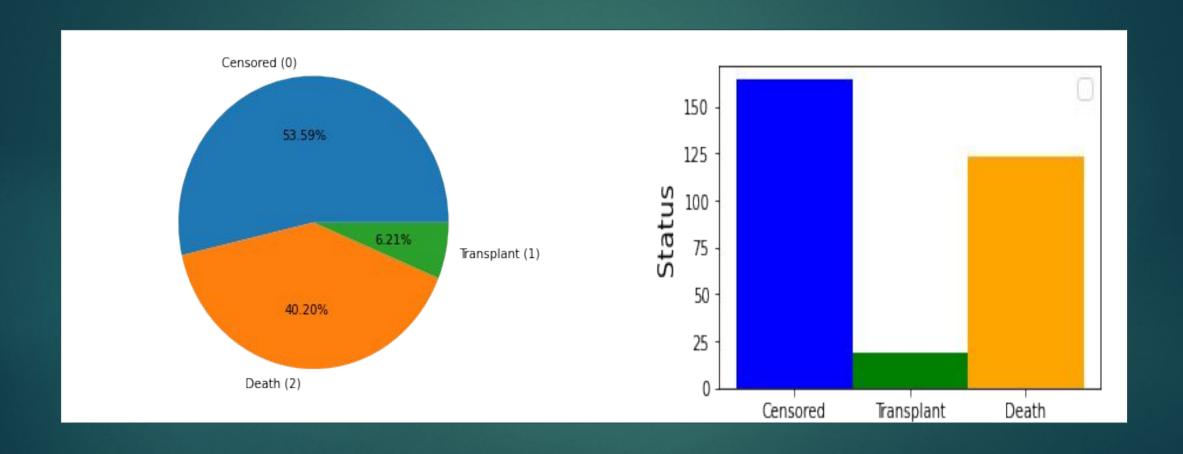
- Get rid off null values
- Replace the cholesterol and the triglycerides for their median and mean values
- Eliminate the rest of the Null



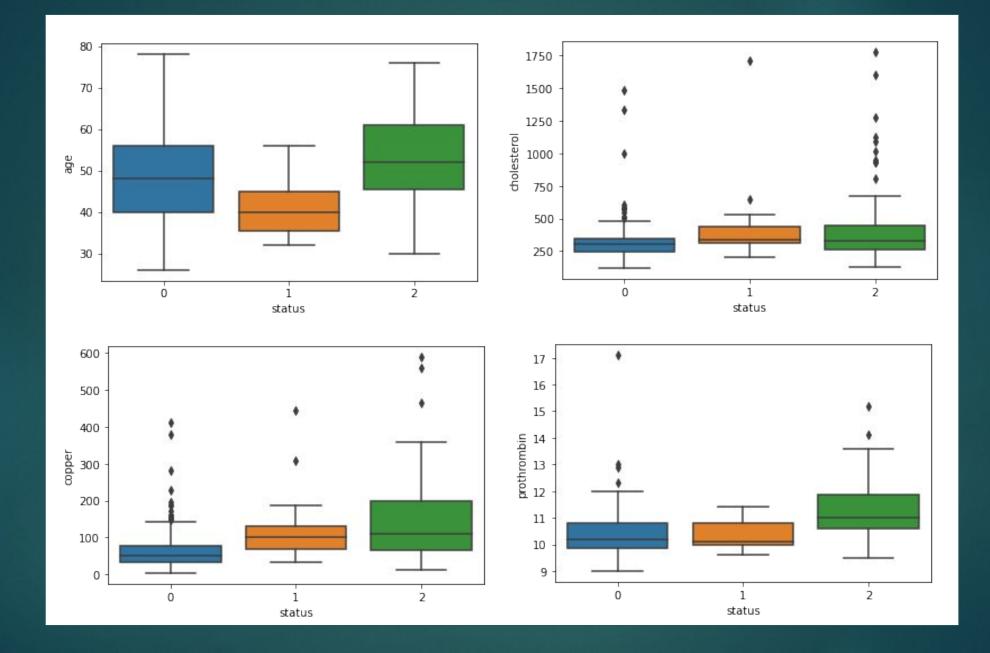


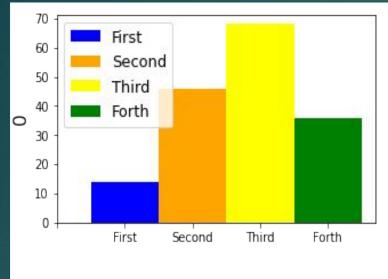
duration	0
status	0
drug	0
age	0
sex	0
ascites	0
hepatomology	0
spiders	0
edema	0
bilirubin	0
cholesterol	28
albumin	0
copper	2
phosphatase	0
SGOT	0
triglicerides	30
platelets	4
prothrombin	0
stage	0

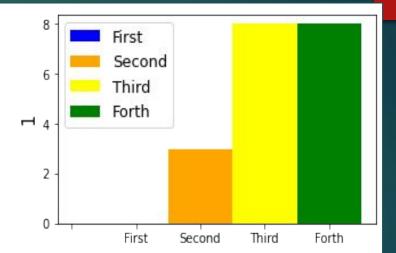
Taking a look to the target

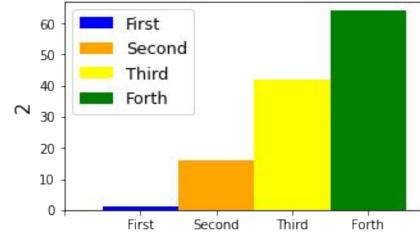


duration	1	-0.42	0.0045	-0.14	0.039	-0.3	-0.29	-0.26	-0.35	-0.44	-0.13	0.43	-0.36	0.14	-0.21	-0.15	0.19	-0.16	-0.36
status	-0.42	1	-0.026	0.2	-0.18	0.31	0.33	0.22	0.33	0.44	0.18	-0.31	0.39	0.23	0.29	0.21	-0.15	0.4	0.34
drug	-0.0045	-0.026	1	-0.14	0.047	-0.033	0.1	0.016	-0.018	0.088	0.027	0.0086	0.0005	-0.01	0.041	0.017	0.03	0.079	0.072
age	-0.14	0.2	-0.14	1	-0.23	0.24	0.068	-0.047	0.22	0.045	-0.15	-0.2	0.066	-0.04	-0.15	0.022	-0.14	0.2	0.17
sex	- 0.039	-0.18	0.047	-0.23	1	-0.014	-0.065	0.14	-0.041	0.029	0.002	-0.041	-0.24	-0.014	-0.0054	-0.05	0.099	-0.1	-0.041
ascites	-0.3	0.31	-0.033	0.24	-0.014	1	0.16	0.14	0.6	0.39	-0.055	-0.39	0.23	0.014	0.083	0.21	-0.21	0.32	0.24
hepatomology	-0.29	0.33	0.1	0.068	-0.065	0.16	1	0.3	0.18	0.31	0.14	-0.31	0.24	0.12	0.14	0.15	-0.19	0.19	0.47
spiders	-0.26	0.22	0.016	-0.047	0.14	0.14	0.3	1	0.28	0.29	0.059	-0.23	0.26	0.04	0.12	0.097	-0.16	0.24	0.28
edema	-0.35	0.33	-0.018	0.22	-0.041	0.6	0.18	0.28	1	0.41	-0.11	-0.4	0.26	0.026	0.14	0.081	-0.24	0.4	0.26
bilirubin	-0.44	0.44	0.088	0.045	0.029	0.39	0.31	0.29	0.41	1	0.38	-0.33	0.45	0.12	0.44	0.42	-0.088	0.36	0.24
cholesterol	-0.13	0.18	0.027	-0.15	0.002	-0.055	0.14	0.059	-0.11	0.38	1	-0.062	0.13	0.14	0.34	0.27	0.18	-0.033	0.0034
albumin	0.43	-0.31	0.0086	-0.2	-0.041	-0.39	-0.31	-0.23	-0.4	-0.33	-0.062	1	-0.26	-0.11	-0.21	-0.1	0.2	-0.23	-0.33
copper	-0.36	0.39	0.0005	0.066	-0.24	0.23	0.24	0.26	0.26	0.45	0.13	-0.26	1	0.19	0.29	0.28	-0.064	0.21	0.27
phosphatase	- 0.14	0.23				0.014		0.04	0.026	0.12	0.14	-0.11	0.19	1	0.12	0.17	0.14	0.087	0.053
SGOT	-0.21		- 1			0.083	0.14	0.12	0.14	0.44	0.34	-0.21	0.29	0.12	1	0.13	-0.12		0.16
triglicerides	A CALCADO	- KARROTO	0.017	200000		0.21	W00000	THE STATE OF	1400000	0.42	0.27	-0.1	0.28	0.17	0.13	1		0.024	144 (144)
platelets		-0.15	0.03	-0.14	0.099		-0.19		-0.24	-0.088	0.18	0.2	-0.064		-0.12	0.094	1	-0.22	
prothrombin	CONTRACTOR OF THE PARTY OF THE	0.4	0.079	0.2	-0.1	0.32	0.19	0.24	0.4		-0.033	The said			0.11		-0.22	1	0.25
stage		0.34	0.072	0.17	-0.041	T	0.47	0.28	0.26	7	0.0034		0.27	0.053	7	0.12	-0.25	0.25	1
	duration	status	drug	age	Sex	ascites	repatomology	spiders	edema	bilirubin	cholesterol	albumin	copper	phosphatase	SGOT	triglicerides	platelets	prothrombin	stage



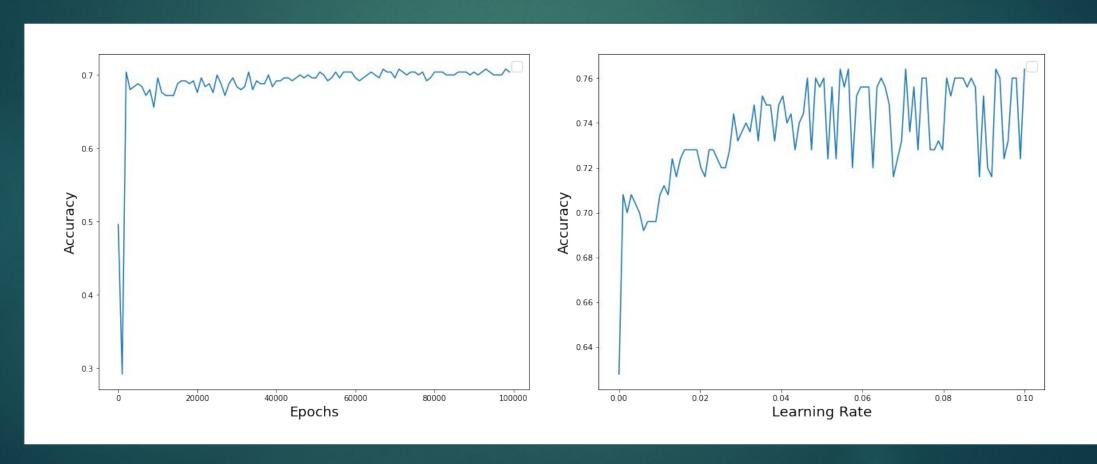




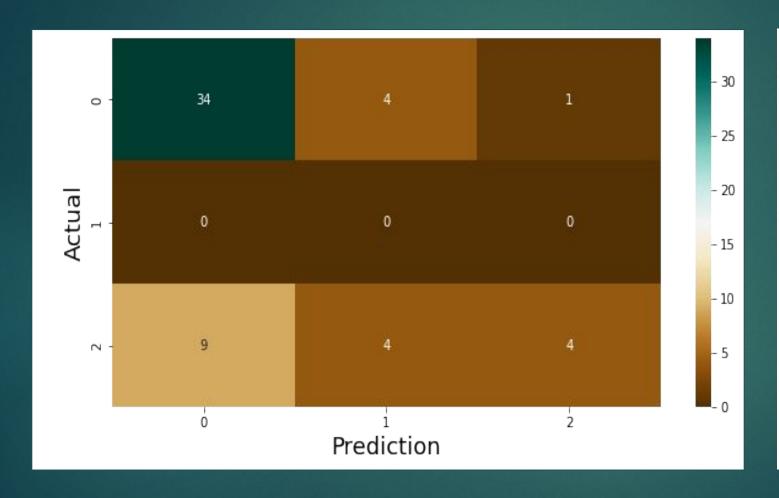


Training the model

It was used the Shallow ANN algorithm, and it was trained with 85% of the data. Along with the columns: hepatomology, copper, ascites, edema, bilirubin, albumin and prothrombin



It was obtained an accuracy of 72% for the Training data and 68% for the test data



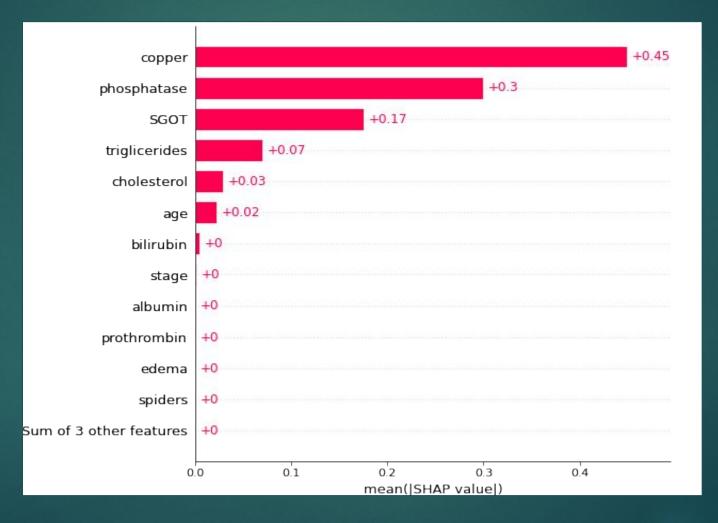
	Actual	prediction
0	0	0
1	0	0
2	0	0
3	0	2
4	0	2
5	0	0
6	1	0
7	1	2

We didn't get the best results. Let's take a look to the probabilities

	0	1	2
0	0.839933	0.033979	0.126088
1	0.715131	0.048004	0.236865
2	0.726844	0.040524	0.232632
3	0.089604	0.040857	0.869539
4	0.356984	0.057773	0.585243
5	0.666264	0.046141	0.287596
6	0.755494	0.039824	0.204682
7	0.155768	0.055069	0.789163
8	0.340955	0.054416	0.604628
9	0.809370	0.038302	0.152327
11 100	A		

Additional remarks

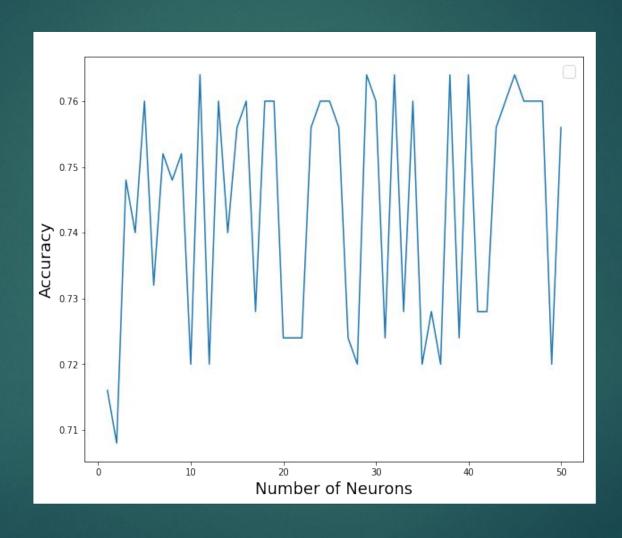
What is the effect of each variable in our model



Conclusions

- It was found really interesting relations between the status target and some other variables during the EDA
- It was understood, by observing the confusion matrix and the probabilities dataframe, why the accuracy of the model was not very precise. We can be able of getting a better one by increasing the amount of essays.
- We were able to observe which variables are the ones which has more impact in our model

Appendix



```
def predict(self, X):
   self.__forward__(X)#Printing the probabilities too
   return self.Z[2].argmax(axis=1),self.Z[2]
#Prediction and proabilities with the test data
y hat test, probs test = ann.predict(X test)
#Accuracy of the test data
accuracy(y_test,y_hat_test)
0.6785714285714286
#Creation of a data frame of the probabilites
result=pd.DataFrame(probs test)
result.head(10)
```