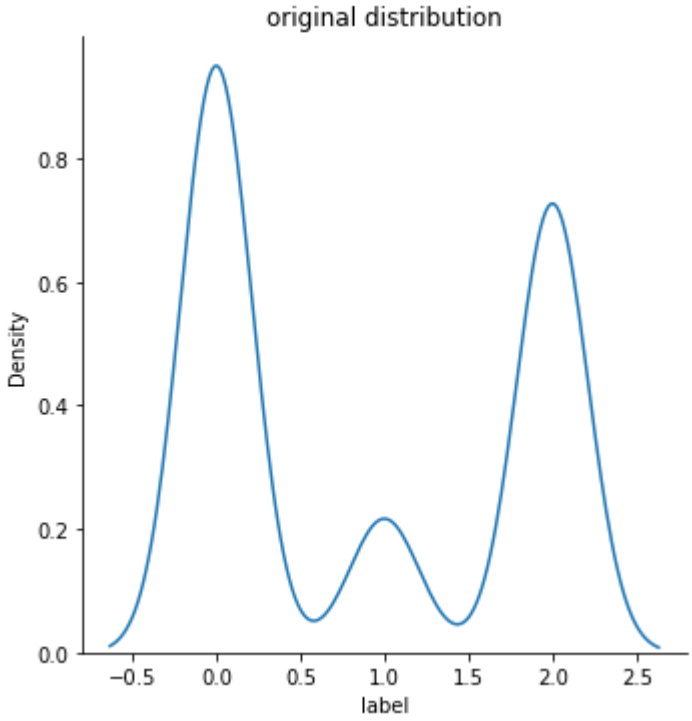
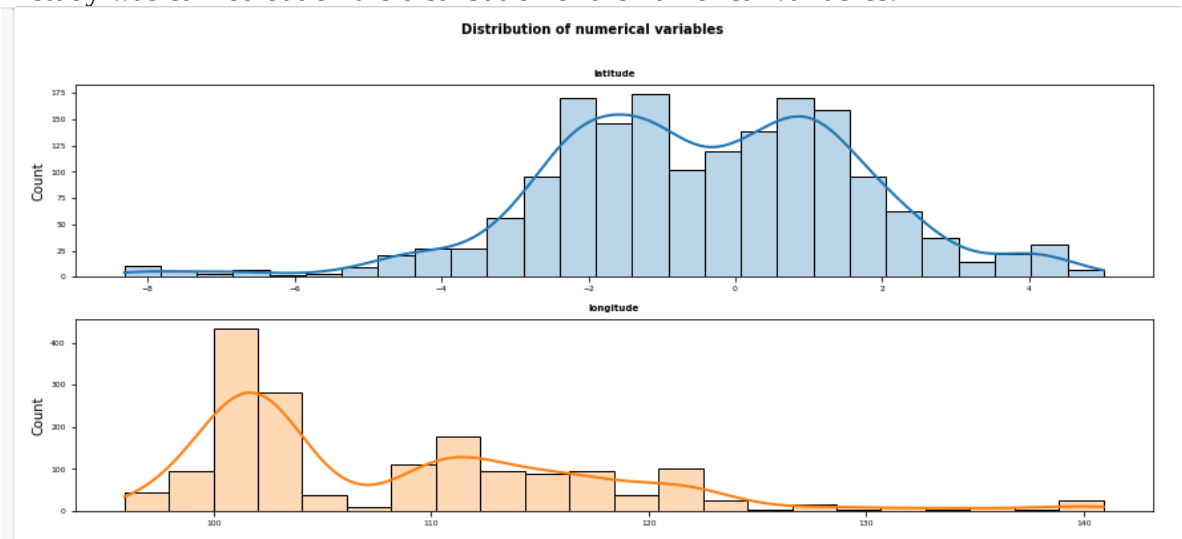
Name: Juan David Moreno

Challenge: Zero deforestation mission

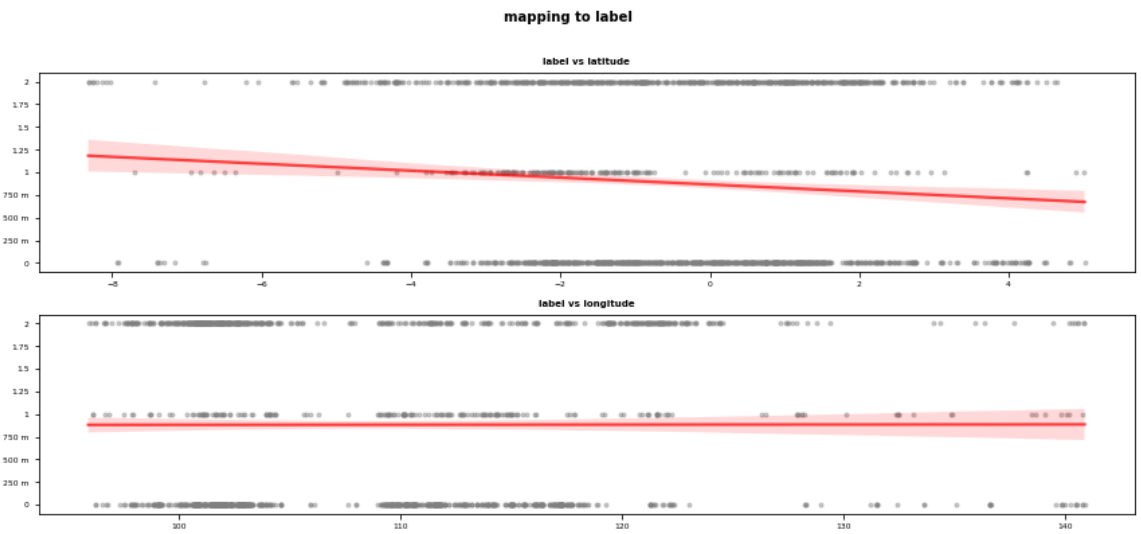
The dataset was read and the information was verified according to the data type of the variable and which had null data. Then a graph of the distribution of the response variable was made.



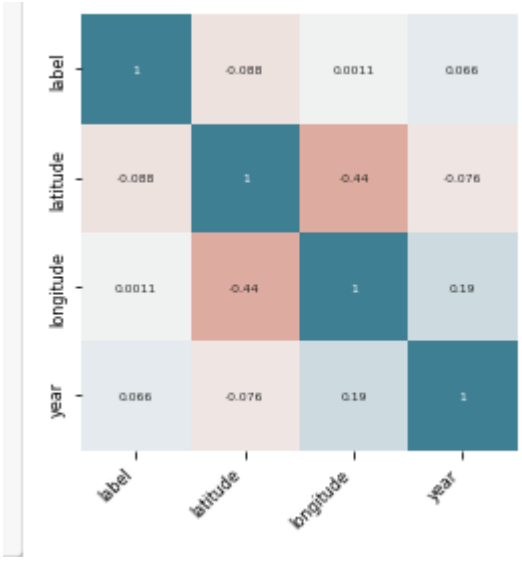
A study was carried out on the distribution of the numerical variables.



One more distribution was made from each variable.



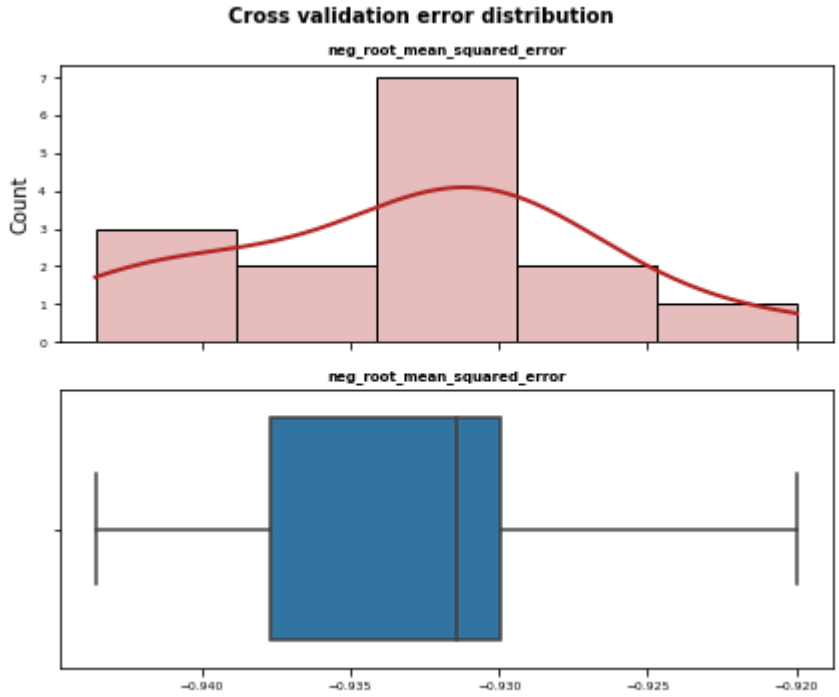
A correlation was made between the variables and the main one that is label, in this case the length variable is the most related



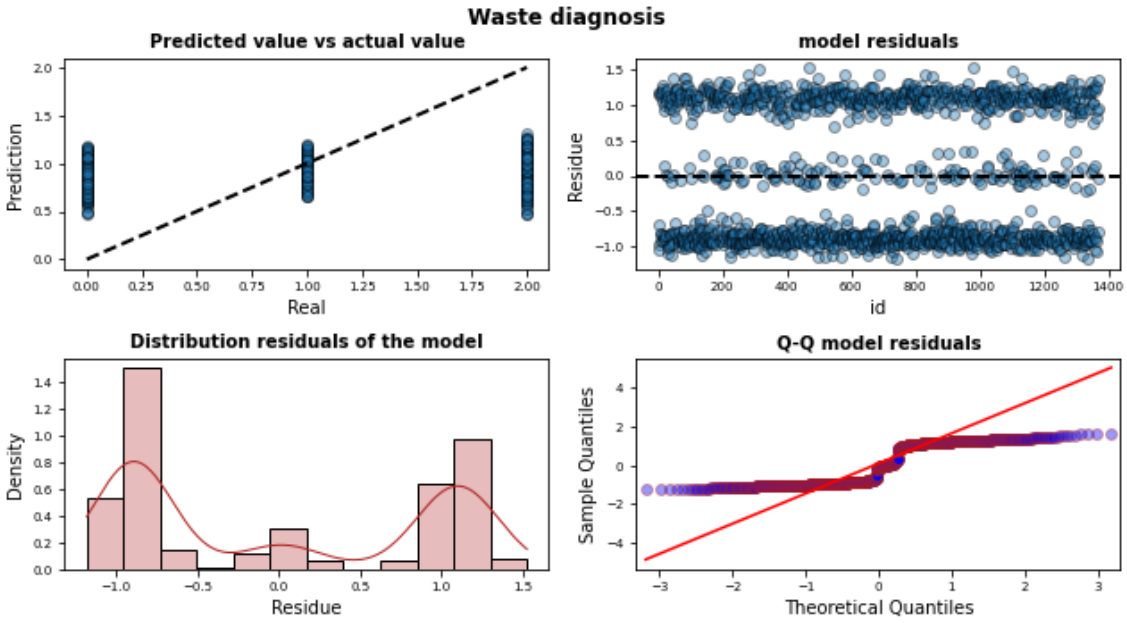
A repeated cross validation was performed to evaluate the trained dataset.

	fit_time	score_time	test_r2	train_r2	test_neg_root_mean_squared_error	train_neg_root_mean_squared_error
0	0.031000	0.012996	0.010383	0.753224	-0.919973	-0.468552
1	0.019999	0.010008	0.014655	0.752982	-0.941200	-0.462747
2	0.025003	0.015994	-0.010230	0.755747	-0.942061	-0.463030
3	0.029000	0.010999	0.011594	0.753745	-0.931235	-0.465160
4	0.024009	0.010001	0.018087	0.753021	-0.931392	-0.465026
5	0.024993	0.012001	0.006839	0.754305	-0.931381	-0.465097
6	0.020007	0.011000	-0.010687	0.755039	-0.933839	-0.465535
7	0.020000	0.011000	0.010816	0.753895	-0.943564	-0.462011
8	0.019001	0.011000	0.012130	0.753032	-0.926562	-0.466724
9	0.020000	0.011000	-0.000853	0.754809	-0.926925	-0.466329
10	0.019004	0.011001	0.016760	0.752431	-0.936738	-0.463974
11	0.020001	0.009999	0.012853	0.753749	-0.932043	-0.464813
12	0.020000	0.011000	0.005811	0.753769	-0.938582	-0.463521
13	0.020004	0.010000	0.013154	0.753548	-0.930378	-0.465318
14	0.020000	0.011000	0.004597	0.754014	-0.929478	-0.465991

In the following graph it can be seen that the boxes and whiskers are asymmetric to the left, which has a normal distribution, where 50% of the data is represented.



A diagnosis of the predictions of the cross-validation was made, where the noise that can be presented in each graph by the residuals that date in the comparison between the real and predictive data can be observed.



At the end, the tested data set is passed to be able to make predictions according to the label variable that represents the categories of the images, and in which this prediction information is passed to a file called "predictions.json" and is attached along with the code in the repository.

	label	prediction
1016	2	0.806150
1541	2	0.735384
454	0	0.866263
889	0	0.937558
222	0	0.760271