

# My Approach to Preliminary Task

Name: Paritosh Yadav

Email: [y.paritosh@gmail.com](mailto:y.paritosh@gmail.com)

Contact Number: +91 9667483313

Dataset: <https://drive.google.com/file/d/1pZhzZnaPi74aKCQImSPzrTxWzVeE0qv/view?usp=sharing>

Github link: <https://github.com/Paritoshyadav/organic-chemica>

## Preprocessing:

### Steps in Data Preprocessing

Here are the steps I have followed;

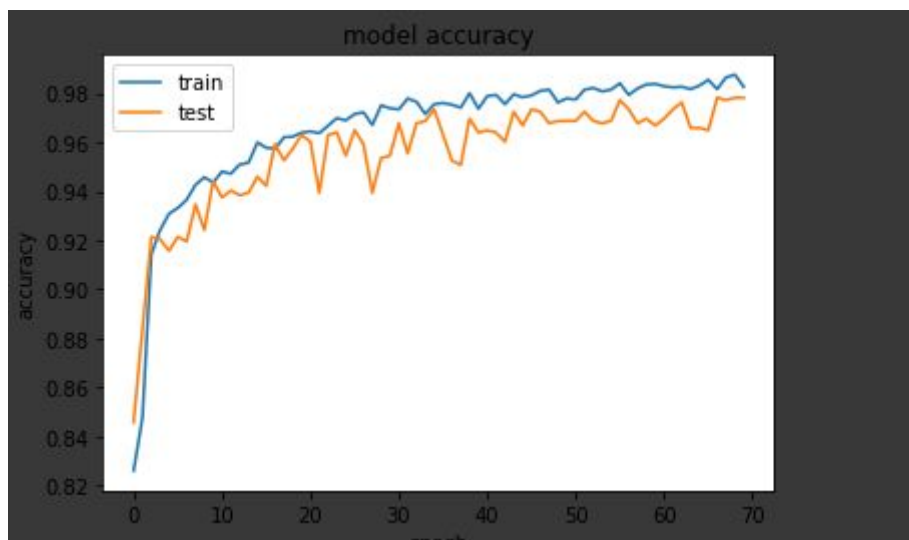
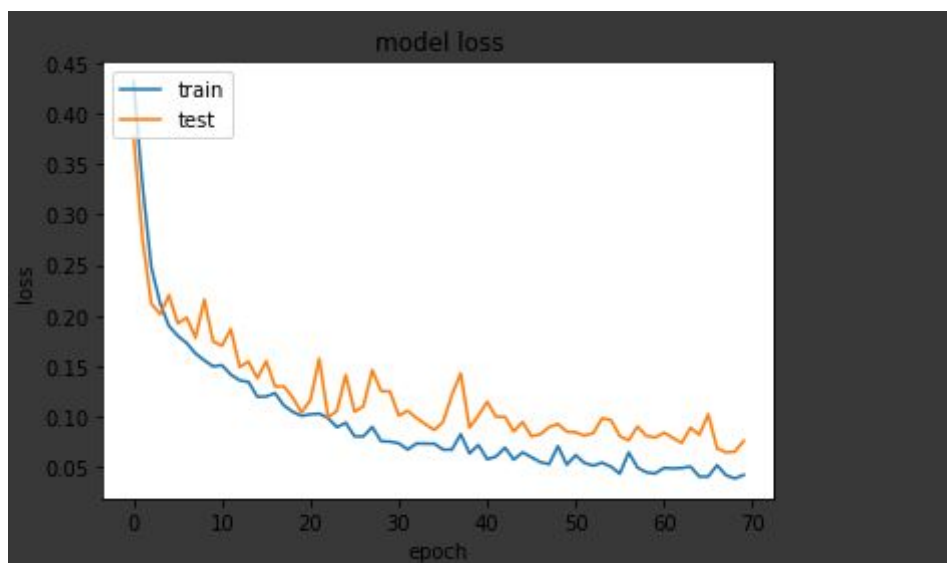
1. Import libraries
2. Read data
3. Checking for missing values
4. Checking for categorical data
5. Checking some Correlation in variables
6. Removing Unique or non -correlated columns
7. Standardize the data(Standardize doesn't give better result)
8. balancing the dataset(oversampling)#tried with both without sampling and with sampling
9. Data splitting

## Post Processing:

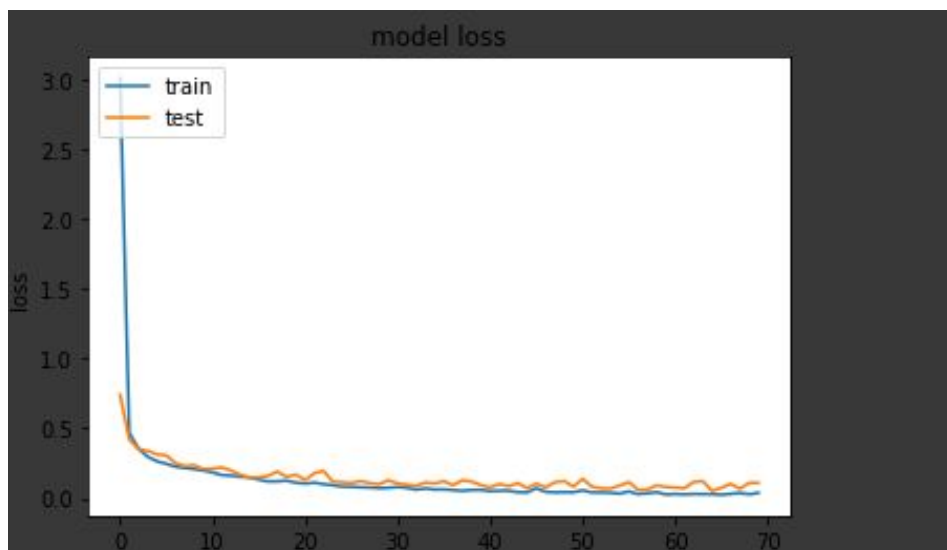
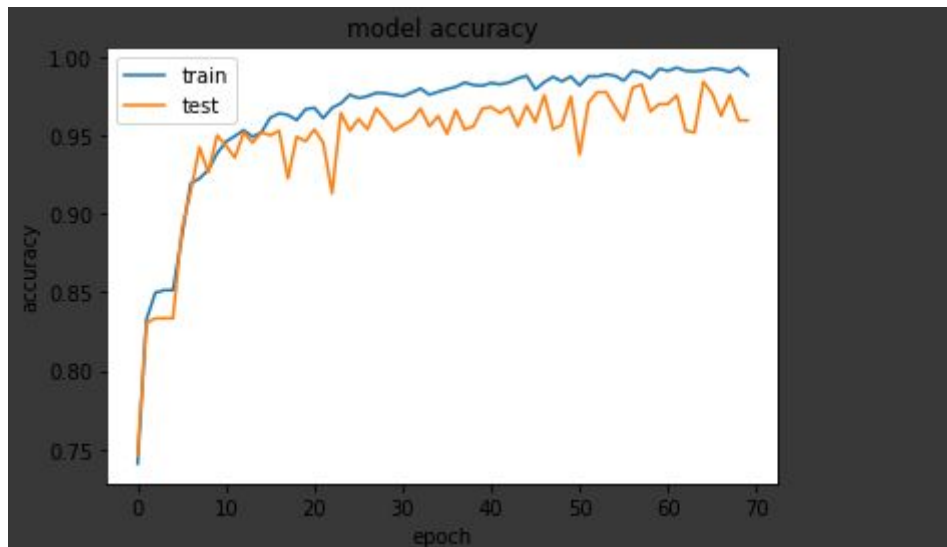
1. Data Visualization
2. Data summarization

## Graph

Result with Standardize data:



### Result without Standardize the data



# Final Report:

Without sampling:

	precision	recall	f1-score	support
0	1.00	0.96	0.98	1107
1	0.84	1.00	0.91	213
accuracy			0.97	1320
macro avg	0.92	0.98	0.95	1320
weighted avg	0.97	0.97	0.97	1320

With sampling(balanced):

	precision	recall	f1-score	support
0	1.00	0.99	0.99	1156
1	0.99	1.00	0.99	1077
accuracy			0.99	2233
macro avg	0.99	0.99	0.99	2233
weighted avg	0.99	0.99	0.99	2233