ML Model for detecting possible SARS-COV-19 patients

-Software Projects Management (MPS)-

Team:

Nicolae, Andreea-Cristina, 343C4 - Team Leader/Project Manager

Coman, Diana-Stefania, 343C5 – Developer

Petcu, Monica-Alexandra, 341C2 – Developer

Sendroiu, Ioan, 343C5 - Tester/Developer

Musatescu, Costin-Teodor, 343C5 – Developer

Constanda, George- Adrian, 343C5 – Developer

Document's purpose

The purpose of this document is to create an inference system of the chance of infection with COVID. In order to realize this project, we use a data set with medical and demographic information about patients tested for COVID and a machine learning model using the data set.

Our team composed of six people have different roles from data processing to developers that create the machine learning model, to be tested by our members.

Document's content

The document consists of three essential sections:

1. Data modeling
2. Program flow
3. Testing components

General description

In spite of our implementation, we start with researching statistics of corona virus infection necessary for the machine learning model. Another first step was to establish training, testing and inference data sets. The second step was to develop the machine learning model. The final step was to train this model, test it using data sets and calculate the metrics in order to obtain the inference.

Technologies

Jupyter Notebooks - Python

Version control

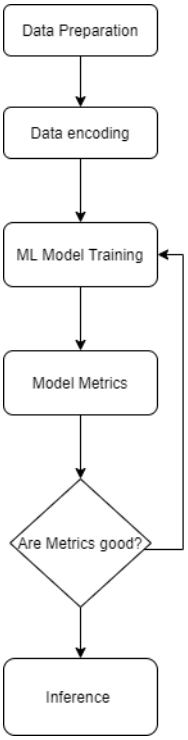
Github: https://github.com/andreeaNic/Error404

Data modeling

System architection description

- Data preparation
    - prelucrareDate - extract and model the original data set received as input
- Encoding date input
    - encodingTime - auxiliary function for data encoding (week)
    - encodingAge - auxiliary function for data encoding (age)
    - encodingSimptome - auxiliary function for data encoding (symptoms)
    - encodingDiagnostic - auxiliary function for data encoding (covid suspect)
    - encodingData - main function for data encoding
- Performance metrix
    - perfomanceMetrix - calculate Accuracy, Precisions, Recall, F1, CM
- Machine learning model
    - Data preparation
    - Encoding date input
    - LogisticRegression - train the logistic regression model
    - performanceMetrix

Program flow

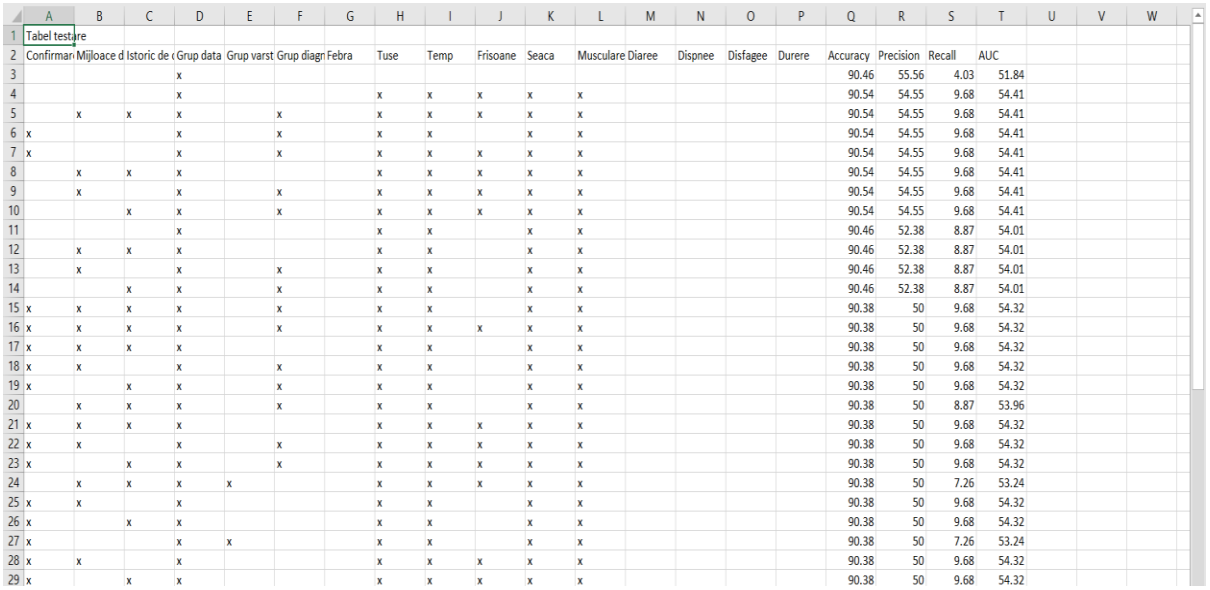


Testing components

We tested the machine learning model considering more combinations of simptoms apart from some fixed ones: result date, dry and cough, high temperature, muscular pains.

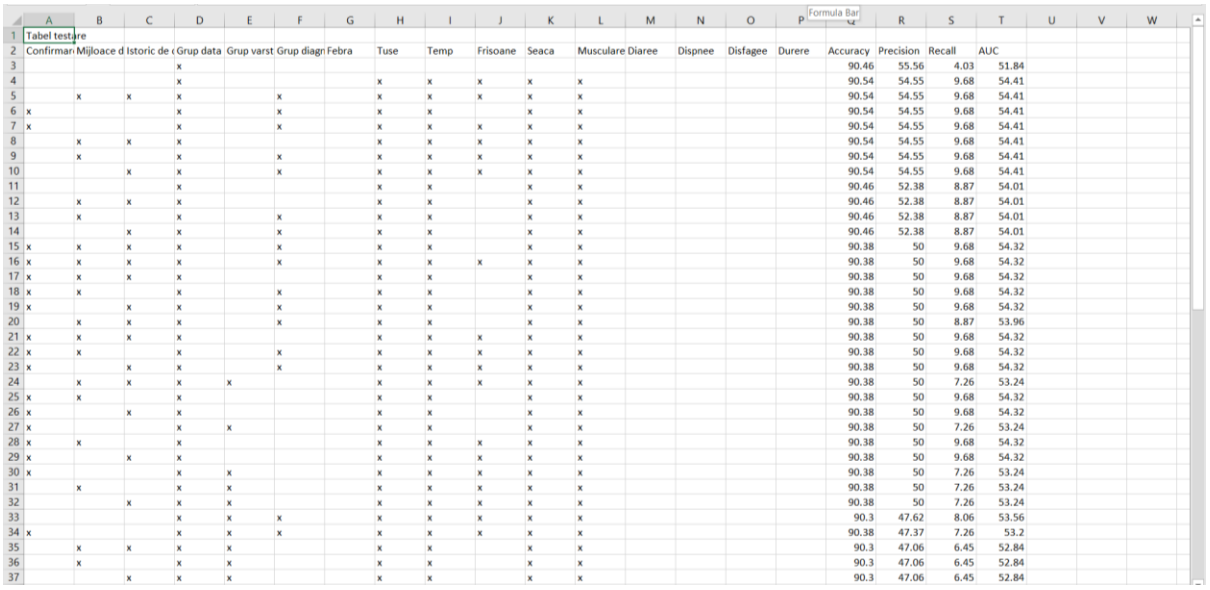
The results are stored in three tables:

- Positives – considering inconclusive/uncertain people



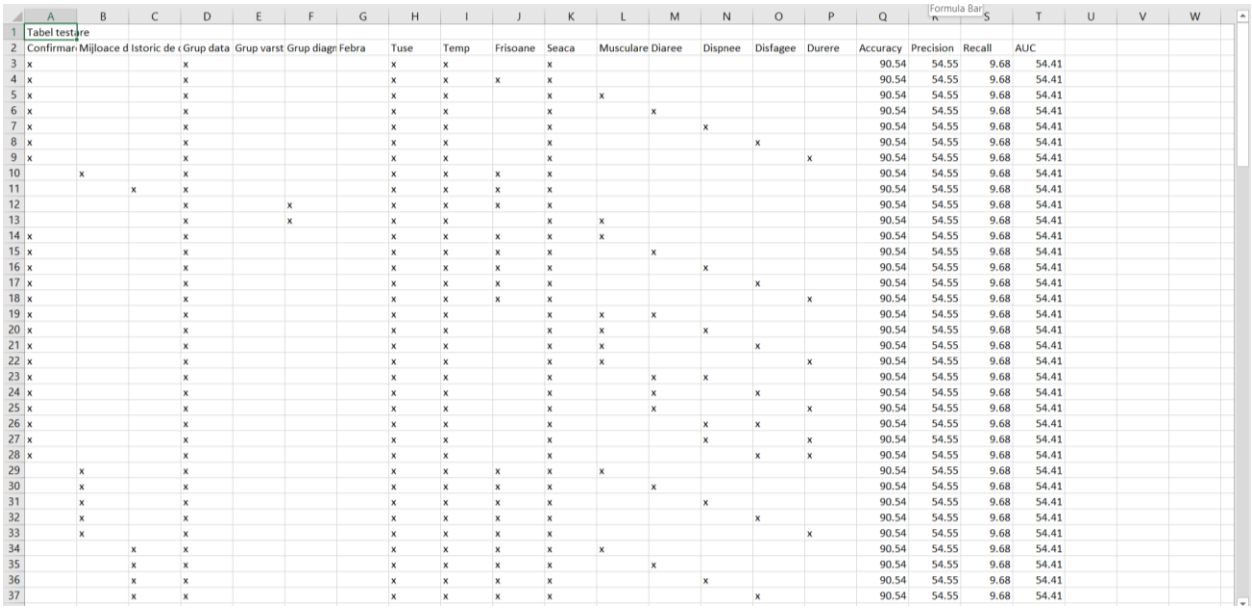
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	Tabel testare																						
2	Confirmare	Mijloace d	Istoric de	Grup data	Grup varst	Grup diagn	Febra	Tuse	Temp	Frisoane	Seaca	Musculare	Diaree	Dispnee	Disfagee	Durere	Accuracy	Precision	Recall	AUC			
3				x													90.46	55.56	4.03	51.84			
4				x				x	x	x	x	x					90.54	54.55	9.68	54.41			
5		x	x	x		x		x	x	x	x	x					90.54	54.55	9.68	54.41			
6	x			x		x		x	x		x	x					90.54	54.55	9.68	54.41			
7	x			x		x		x	x	x	x	x					90.54	54.55	9.68	54.41			
8		x	x	x				x	x	x	x	x					90.54	54.55	9.68	54.41			
9		x		x		x		x	x	x	x	x					90.54	54.55	9.68	54.41			
10			x	x		x		x	x	x	x	x					90.54	54.55	9.68	54.41			
11				x				x	x		x	x					90.46	52.38	8.87	54.01			
12		x	x	x				x	x		x	x					90.46	52.38	8.87	54.01			
13		x		x		x		x	x		x	x					90.46	52.38	8.87	54.01			
14			x	x		x		x	x		x	x					90.46	52.38	8.87	54.01			
15	x	x	x	x		x		x	x		x	x					90.38	50	9.68	54.32			
16	x	x	x	x		x		x	x	x	x	x					90.38	50	9.68	54.32			
17	x	x	x	x				x	x		x	x					90.38	50	9.68	54.32			
18	x	x		x		x		x	x		x	x					90.38	50	9.68	54.32			
19	x		x	x		x		x	x		x	x					90.38	50	9.68	54.32			
20		x	x	x		x		x	x		x	x					90.38	50	8.87	53.96			
21	x	x	x	x				x	x	x	x	x					90.38	50	9.68	54.32			
22	x	x		x		x		x	x	x	x	x					90.38	50	9.68	54.32			
23	x		x	x		x		x	x	x	x	x					90.38	50	9.68	54.32			
24		x	x	x	x			x	x	x	x	x					90.38	50	7.26	53.24			
25	x	x		x				x	x		x	x					90.38	50	9.68	54.32			
26	x		x	x				x	x		x	x					90.38	50	9.68	54.32			
27	x			x	x			x	x		x	x					90.38	50	7.26	53.24			
28	x	x		x				x	x	x	x	x					90.38	50	9.68	54.32			
29	x		x	x				x	x	x	x	x					90.38	50	9.68	54.32			

- Negatives - considering negative people



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	Tabel testare																						
2	Confirmare	Mijloace d	Istoric de	Grup data	Grup varst	Grup diagn	Febra	Tuse	Temp	Frisoane	Seaca	Musculare	Diaree	Dispnee	Disfagee	Durere	Accuracy	Precision	Recall	AUC			
3				x													90.46	55.56	4.03	51.84			
4				x				x	x	x	x	x					90.54	54.55	9.68	54.41			
5		x	x	x		x		x	x	x	x	x					90.54	54.55	9.68	54.41			
6	x			x		x		x	x		x	x					90.54	54.55	9.68	54.41			
7	x			x		x		x	x	x	x	x					90.54	54.55	9.68	54.41			
8		x	x	x				x	x	x	x	x					90.54	54.55	9.68	54.41			
9		x		x		x		x	x	x	x	x					90.54	54.55	9.68	54.41			
10			x	x		x		x	x	x	x	x					90.54	54.55	9.68	54.41			
11				x				x	x		x	x					90.46	52.38	8.87	54.01			
12		x	x	x				x	x		x	x					90.46	52.38	8.87	54.01			
13		x		x		x		x	x		x	x					90.46	52.38	8.87	54.01			
14			x	x		x		x	x		x	x					90.46	52.38	8.87	54.01			
15	x	x	x	x		x		x	x		x	x					90.38	50	9.68	54.32			
16	x	x	x	x		x		x	x	x	x	x					90.38	50	9.68	54.32			
17	x	x	x	x				x	x		x	x					90.38	50	9.68	54.32			
18	x	x		x		x		x	x		x	x					90.38	50	9.68	54.32			
19	x		x	x		x		x	x		x	x					90.38	50	9.68	54.32			
20		x	x	x		x		x	x		x	x					90.38	50	8.87	53.96			
21	x	x	x	x				x	x	x	x	x					90.38	50	9.68	54.32			
22	x	x		x		x		x	x	x	x	x					90.38	50	9.68	54.32			
23	x		x	x		x		x	x	x	x	x					90.38	50	9.68	54.32			
24		x	x	x	x			x	x	x	x	x					90.38	50	7.26	53.24			
25	x	x		x				x	x		x	x					90.38	50	9.68	54.32			
26	x		x	x				x	x		x	x					90.38	50	9.68	54.32			
27	x			x	x			x	x		x	x					90.38	50	7.26	53.24			
28	x	x		x				x	x	x	x	x					90.38	50	9.68	54.32			
29	x		x	x				x	x	x	x	x					90.38	50	9.68	54.32			
30	x			x	x			x	x	x	x	x					90.38	50	7.26	53.24			
31		x		x	x			x	x	x	x	x					90.38	50	7.26	53.24			
32			x	x	x			x	x	x	x	x					90.38	50	7.26	53.24			
33				x	x	x		x	x	x	x	x					90.3	47.62	8.06	53.56			
34	x			x	x	x		x	x	x	x	x					90.38	47.37	7.26	53.2			
35		x	x	x	x			x	x		x	x					90.3	47.06	6.45	52.84			
36		x		x	x			x	x		x	x					90.3	47.06	6.45	52.84			
37			x	x	x			x	x		x	x					90.3	47.06	6.45	52.84			

- More negatives



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	Tabel testare																						
2	Confirmare	Mijloace d	Istoric de c	Grup data	Grup varst	Grup diagn	Febra	Tuse	Temp	Frisoane	Seaca	Musculare	Diaree	Dispnee	Disfagee	Durere	Accuracy	Precision	Recall	AUC			
3	x			x				x	x		x						90.54	54.55	9.68	54.41			
4	x			x				x	x	x	x						90.54	54.55	9.68	54.41			
5	x			x				x	x		x	x					90.54	54.55	9.68	54.41			
6	x			x				x	x		x		x				90.54	54.55	9.68	54.41			
7	x			x				x	x		x			x			90.54	54.55	9.68	54.41			
8	x			x				x	x		x				x		90.54	54.55	9.68	54.41			
9	x			x				x	x		x					x	90.54	54.55	9.68	54.41			
10		x		x				x	x	x	x						90.54	54.55	9.68	54.41			
11			x	x				x	x	x	x						90.54	54.55	9.68	54.41			
12				x		x		x	x	x	x						90.54	54.55	9.68	54.41			
13				x		x		x	x		x	x					90.54	54.55	9.68	54.41			
14	x			x				x	x	x	x	x					90.54	54.55	9.68	54.41			
15	x			x				x	x	x	x		x				90.54	54.55	9.68	54.41			
16	x			x				x	x	x	x			x			90.54	54.55	9.68	54.41			
17	x			x				x	x	x	x				x		90.54	54.55	9.68	54.41			
18	x			x				x	x	x	x					x	90.54	54.55	9.68	54.41			
19	x			x				x	x		x	x	x				90.54	54.55	9.68	54.41			
20	x			x				x	x		x	x		x			90.54	54.55	9.68	54.41			
21	x			x				x	x		x	x			x		90.54	54.55	9.68	54.41			
22	x			x				x	x		x	x				x	90.54	54.55	9.68	54.41			
23	x			x				x	x		x		x	x			90.54	54.55	9.68	54.41			
24	x			x				x	x		x		x		x		90.54	54.55	9.68	54.41			
25	x			x				x	x		x		x			x	90.54	54.55	9.68	54.41			
26	x			x				x	x		x			x	x		90.54	54.55	9.68	54.41			
27	x			x				x	x		x			x		x	90.54	54.55	9.68	54.41			
28	x			x				x	x		x				x	x	90.54	54.55	9.68	54.41			
29		x		x				x	x	x	x	x					90.54	54.55	9.68	54.41			
30		x		x				x	x	x	x		x				90.54	54.55	9.68	54.41			
31		x		x				x	x	x	x			x			90.54	54.55	9.68	54.41			
32		x		x				x	x	x	x				x		90.54	54.55	9.68	54.41			
33		x		x				x	x	x	x					x	90.54	54.55	9.68	54.41			
34			x	x				x	x	x	x	x					90.54	54.55	9.68	54.41			
35			x	x				x	x	x	x		x				90.54	54.55	9.68	54.41			
36			x	x				x	x	x	x			x			90.54	54.55	9.68	54.41			
37			x	x				x	x	x	x				x		90.54	54.55	9.68	54.41			