INTRODUCTION

Toolbox for Computer Aided Diagnostic is an attempt to automate the medical diagnosis system . It is not to substitute the doctor instead it can assist doctors. If the performances is up to the mark then it can be made available to regions where there is dearth of hospitals and doctors .Also, People can get information that whether he has disease or not at home and at low cost. So, it saves lot of time to travel to doctor.

1.1 Motivation behind the choice of topic: In today's scenario accuracy is the main concern before prescribing or giving remedies to the patient. So the algorithm has to be accurate as much as it can as it's someone's life we are dealing with. According to the administrations report, every one in ten people suffer from Thyroid in India. Thyroid disorders impair normal functioning of the thyroid gland causing abnormal production of hormones leading to thyroidism. If left untreated, thyroid can cause elevated cholesterol levels, an increase in blood pressure, cardiovascular complications, decreased fertility, and depression.

The liver is essential for digesting food and eradicating toxic substances from our body. Liver disorders include hepatitis, cirrhosis, liver tumors, and liver abscess (collection of pus).

Also to propose a better algorithm for the detection of the Thyroid and Liver Disorder.

1.2 Purpose of Project: In rural areas of India, we can still find no doctors or any clinic for the support of the people. Even in the cities, it is difficult for person to take time from his tedious schedule for an appointment with the doctor. So our effort is to provide a better algorithm for detection of Liver Disorder and Thyroid disease.

1.3 Description of Project:

Our main aim is to automate the system with better efficiency, specially it will be useful in Rural Areas when everyone is not able to have good doctor to consult or the one whose schedule is too tedious that he has no time to consult doctor. This will be very useful for them.

Also, doctors can use our system to detect diseases very efficiently and easily.

BACKGROUND STUDY

We are using Android Studio as our front end, Fire-Base which is used as a Database for maintaining our records, Python for Socket Programming which is used as a connection between Android – Python and Machine Learning for implementing various algorithms to detect Hyperthyroid and Hypothyroid and Liver -Disorder Detection . We have studied 6 Research Papers for our Project, whose summary is mentioned in the later part.

The process of telling him to take test and then again to check his Test Report is so tedious for every Doctor and even for Patients also, as they have to visit the doctor again and again by keeping their other works aside. So, to alleviate this problem, we have developed an Application (using Android) which will help Doctors as well as Patients to detect Disease in very efficient way. Patient has to simply fill a form and to upload an image of his PDF. Our Application will give the desired result very easily and quickly It will save a lot of time. Also, we provide a simple interface that any non-skilled user can also operate to get the correct information.

The table shown below is the list of Research Papers that we have studied:-

2.1 Literature Survey:-

RESEARCH PAPER NAME	AUTHOR	SUMMARY
[1] "Thyroid Disease Diagnosis Based on Genetic Algorithms using PNN and SVM" IEEE , July 2007	Fatimah Saiti,Afsaneh Alavi Naini,Mahdi Aliyari shoorehdeli, Mohammad Teshnehlab	They have implemented a system which is able to provide the clinician with a rapid overview of a patient's medical problems, reporting a problem list, any medications taken and the procedures performed. They have used ICD-9 codes relating to recognized disorders and the tool is based on an NLP pipeline.
[2] "TDTD: Thyroid Disease Type Diagnostics" IEEE , May 2016	Ahmed , M. Abdul Rehman Soomrani	An universal medicine recommender system framework is implemented in which data mining technologies are applied . They have investigated various medicine recommendation algorithms of the SVM (Support Vector Machine), BP neural network algorithm and ID3 decision tree. They have selected SVM recommendation model for the medicine recommendation module to obtain good efficiency, high accuracy and model scalability.

[3] "Soft Computing Based Expert System for Hepatitis and Liver Disorders, IEEE March 2016"	Dr. R.R.Janghel,Dr. Anupam Shukla,Kshitij Verma	The authors has used many Artificial Neural Network Models like Back Propagation Algorithm, Probabilistic Neural Network ,Learning Vector quantization and Elman networks. using MATLAB. They predicted Liver Disorder and Hepatitis. Among all the above algorithms the best accuracy was obained by Backpropagation Algorithm.
[4] WEIGHTED KNEAREST NEIGHBOR CLASSIFICATION ON FEATURE PROJECTIONS, Research Gate October 2000	H. Altay Guvenir , Aynur	Author extends the KNN algorithm on Feature Projections. This algorithm has very low time complexity in comparison to former.It introduces weighted kNNFP and these algorithms were applied on various Datasets and their accuracy was recoorded, out of which the highest accuracy achieved was 62.7 % when k=10.
[5] Editing Training Data for kNN Classifiers with Neural Network Ensemble	Yuan Jiang, Zhi-Hua Zhou	Editing is used like Depuration algorithm after that KNN is applied. In editing all the columns are edited and compressed into smaller columns to increase accuracy. The experiments showed that after applying all the editing algorithms, the one having highest accuracy was about 69.33% using NNEE algorithm.
[6] "Thyroid disease diagnosis: A survey", IEEE October 2015	Shanu Shroff, Siddhi Pise, Pratiksha Chalekar,Suja S. Panicke	In [11], the efficiency summary like Fuzzy-KNN with 99.09\%,CAD based technique used for extracting features from 3D contrast-enhanced Ultra Sound (US) images after which knn, PNN, and DT algorithms were 95%.

REQUIREMENT ANALYSIS

a) Software Requirements

1. Windows 7 and above

2. Android Studio

3. Python Idle

b) Hardware

1. Processor- i3, i5, i7

2. Hard Disk- 1 TB

3. RAM-8 GB RAM, 1.83 GHz processor.

4. Android Phones of version 5.0(Lollipop) or above

c) Functional Requirements

1. User is required to fill all the details while Signing Up.

2. User is required to fill all his details in different forms provided in application.

3. After giving all his details, User is required to upload an image of his respective test's PDF .

5. User can edit as well as view all his details and after evaluation, he can check whether he has disease or not

d) Non-Functional Requirements

1. Validations should be there while creating an account like in email-Id, with proper syntax.

2. System will evaluate the disease and store the answer in database.

3. System will apply an appropriate algorithm of Machine Learning to detect the disease.

4. System will import content from Android to Python and again back from Python to Android.

e) User Requirements

The user requires simple software that tells him whether he has a particular disease or not. He simply wants an interface where he can feed all the details, upload an image of his PDF and when he clicks the "submit" button, the desired results are displayed. The user does not have to worry about algorithms and the process behind this application. The patient only knows about filling a basic form at the front-end. The retrieval should be fast and efficient.

f) UML Diagram

Here is the UML (Use Case) diagram of our Project. In this we have two actors namely Patient, Doctor and their corresponding functions.

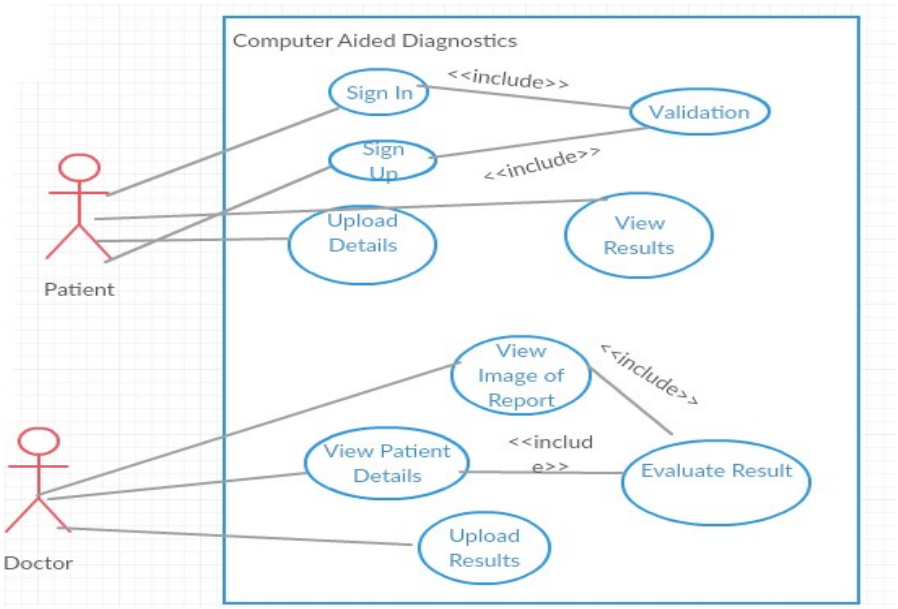

Figure-1(Use case diagram)

DETAILED DESIGN

4.1 PATIENT

Here is the detailed design of the Patient side: -

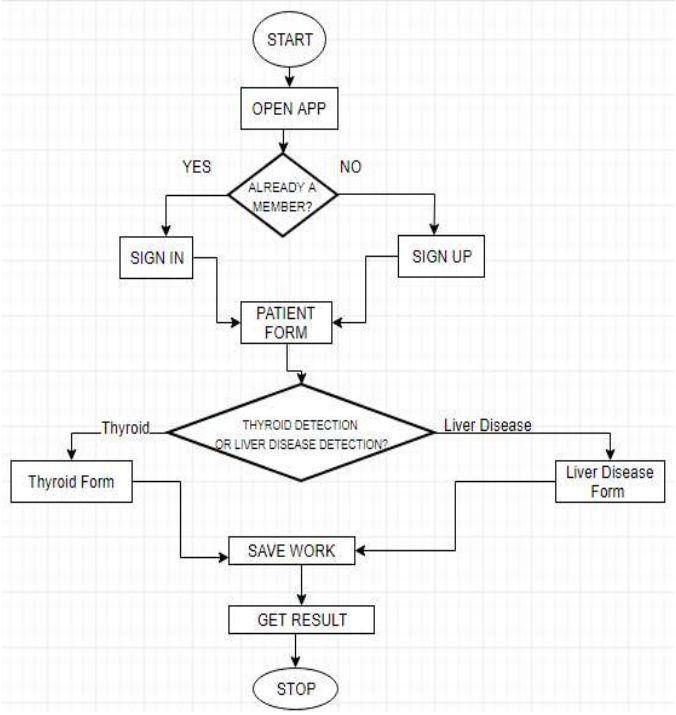

Figure-2(Flow Chart of Patient)

4.2 DOCTOR

Here is the detailed design of the Doctor's side:

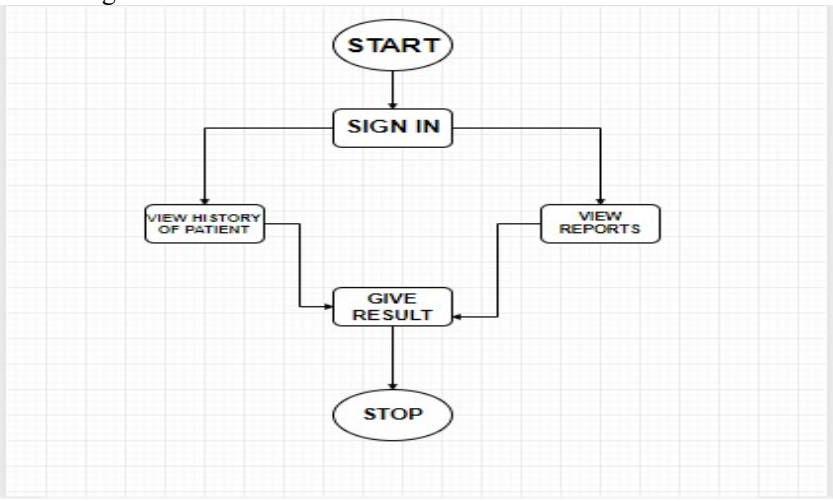

Figure-3(Flow Chart of Doctor)

IMPLEMENTATION

We have used Android Studio and Machine Learning for diagnosis of thyroid disease and liver disorder detection .At first, patient needs to fill the forms(Fig.1) and after submitting the form, Doctor will send a SMS(Fig.5) to the Patients. Then Patient will upload an image of his test report. After uploading an image, Doctor's assistant will fill all the details(Fig 7) in Thyroid or Liver disorder form and by Machine Learning Algorithm i.e. by using Random Forest Classifier, Decision Tree Classifier, Linear Discriminant Analysis, Ada-Boost Classifier, Bagging Classifier, Stochastic Gradient Descent, Logistic Regression, K-Nearest Neighbors.

Fig 4: Patient's Details

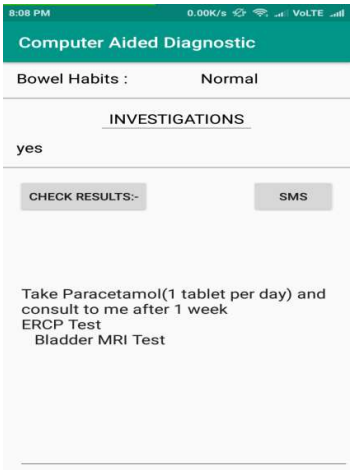

Fig 5: Tests

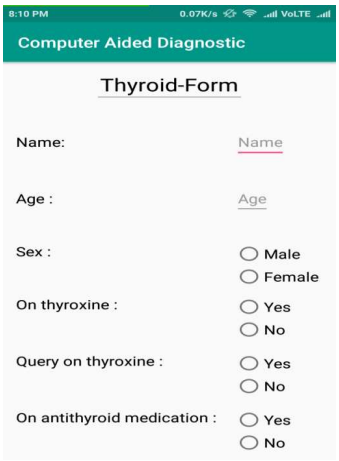

Fig 6: Thyroid Form

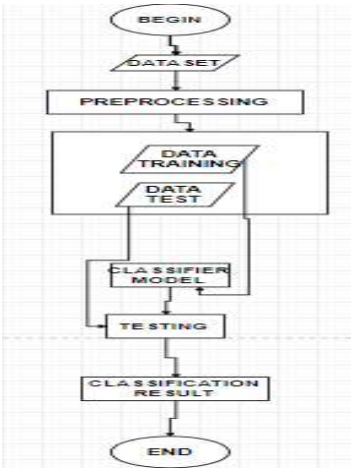

Fig 7: Machine Learning Flow chart

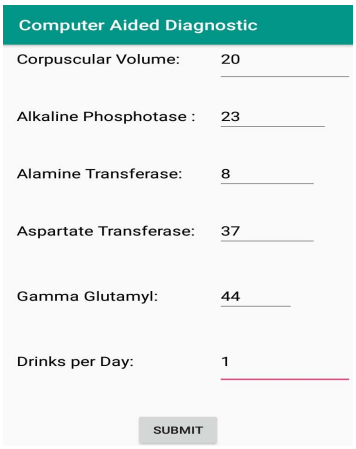

Fig 8: Liver Form

EXPERIMENTAL RESULTS AND ANALYSIS

We are using the real dataset of the patients. Doctor's assistant will feed the details of the patients and the result will be sent to the user. We have also trained our algorithm so that it can show the output with good accuracy.

In case of hypothyroid, combination of Bagging using LDA and Adaboost using decision tree is providing highest frequency which is 99.8%.

In case of hyperthyroid, combination of Bagging using decision tree and Adaboost using Random Forest and Bagging using LDA are providing highest frequency which is 99.1%.

In case of liver disorder, Bagging using decision tree is providing highest frequency which is 82.6%.

INFORMATION ON ACCURACY AND EXECUTION TIME OF HYPO-THYROID DATA-SET

S.No	CLASSIFIER	ACC.	EXEC. TIME
1	Adaboost using Random Forest	98.7%	0.72 sec
2	Bagging using Decision Tree	99.1%	0.85sec
3	Bagging using LDA	93.5%	1.19 sec
4	Adaboost using Decision Treet	99.5%	2.67 sec
5	Adaboost using Random Forest Bagging using Decision Tree	99.3%	1.03 sec
6	Adaboost using Random Forest Bagging using LDA	98.7%	1.55 sec
7	Adaboost using Random Forest Adaboost using Decision Tree	99.8%	2.87 sec
8	Bagging using Decision Tree Bagging using LDA	98.2%	2.24 sec
9	Bagging using Decision Tree Adaboost using Decision Tree	99.7%	2.87 sec
10	Bagging using LDA Adaboost using Decision Treet	99.8%	3.68 sec
11	Adaboost using Random Forest Bagging using Decision Tree Bagging using LDA	98.7%	2.65 sec
12	Adaboost using Random Forest Bagging using LDA Adaboost using Decision Tree	99.0%	4.73 sec
13	Adaboost using Random Forest Bagging using Decision Tree Adaboost using Decision Treet	99.5%	3.12 sec
14	Bagging using Decision Tree Bagging using LDA Adaboost using Decision Treet	99.6%	5.35 sec
15	Adaboost using Random Forest Bagging using Decision Tree Bagging using LDA Adaboost using Decision Treet	99.5%	5.86 sec

Table1: Accuracy & Exec. Time of Hypo-Thyroid Dataset

INFORMATION ON ACCURACY AND EXECUTION TIME OF HYPER-THYROID DATA-SET

S.No	CLASSIFIER	ACC.	EXEC. TIME
1	Adaboost using Random Forest	99.1%	0.52 sec
2	Bagging using Decision Tree	99.2%	0.66sec
3	Bagging using LDA	98.6%	1.22 sec
4	Adaboost using Decision Treet	98.2%	3.15 sec
5	Adaboost using Random Forest Bagging using Decision Tree	99.2%	0.81 sec
6	Adaboost using Random Forest Bagging using LDA	98.6%	0.98
7	Adaboost using Random Forest Adaboost using Decision Treet	98.1%	3.30 sec
8	Bagging using Decision Tree Bagging using LDA	98.2%	2.24 sec
9	Bagging using Decision Tree Adaboost using Decision Treet	98.6%	3.48 sec
10	Bagging using LDA Adaboost using Decision Treet	98.01%	4.49 sec
11	Adaboost using Random Forest Bagging using Decision Tree Bagging using LDA	98.8%	3.08 sec
12	Adaboost using Random Forest Bagging using LDA Adaboost using Decision Treet	98.7%	5.08 sec
13	Adaboost using Random Forest Bagging using Decision Tree Adaboost using Decision Treet	99.1%	4.07 sec
14	Bagging using Decision Tree Bagging using LDA Adaboost using Decision Treet	98.8%	5.35 sec
15	Adaboost using Random Forest Bagging using Decision Tree Bagging using LDA Adaboost using Decision Treet	98.7%	6.31 sec

Table2: Accuracy & Exec. Time of Hyper-Thyroid Dataset

INFORMATION ON ACCURACY AND EXECUTION TIME OF LIVER DISORDER

S.No	CLASSIFIER	ACC.	EXEC. TIME
1	Adaboost using Random Forest	81.1%	0.51 sec
2	Bagging using Decision Tree	82.6%	0.55sec
3	Bagging using LDA	62.3%	1.05sec
4	Adaboost using Decision Treet	60.8%	2.64 sec
5	Adaboost using Random Forest Bagging using Decision Tree	78.2%	0.91 sec
6	Adaboost using Random Forest Bagging using LDA	81.1%	1.74 sec
7	Adaboost using Random Forest Adaboost using Decision Tree	59.4%	3.22 sec
8	Bagging using Decision Tree Bagging using LDA	71.0%	1.90 sec
9	Bagging using Decision Tree Adaboost using Decision Tree	62.3%	3.30 sec
10	Bagging using LDA Adaboost using Decision Treet	63.7%	3.98 sec
11	Adaboost using Random Forest Bagging using Decision Tree Bagging using LDA	75.3%	2.61 sec
12	Adaboost using Random Forest Bagging using LDA Adaboost using Decision Tree	75.3%	4.59 sec
13	Adaboost using Random Forest Bagging using Decision Tree Adaboost using Decision Treet	76.8%	3.83 sec
14	Bagging using Decision Tree Bagging using LDA Adaboost using Decision Treet	78.2%	4.74 sec
15	Adaboost using Random Forest Bagging using Decision Tree Bagging using LDA Adaboost using Decision Treet	73.1%	5.25 sec

Table3: Accuracy & Exec. Time of Liver Disorder

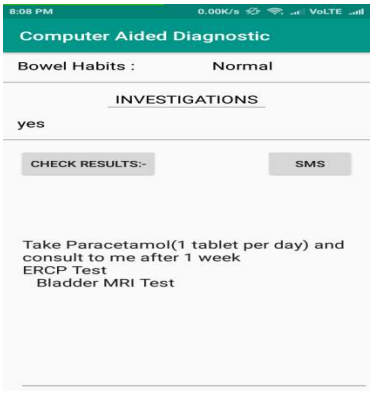

Fig 9:SMS of tests will be sent to the patients.

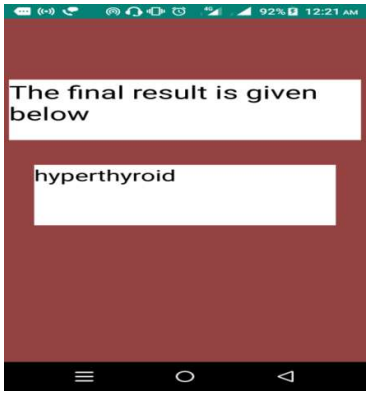

Fig 10:Final result which tells whether patient has thyroid disease or not.

6.1 TESTING

Test Id	Input	Expected Output	Actual Output	Pass or Fail
01.	'vidhi', '21', '00', '0', '0', '0', '0', '0', '0', '0', '0', '0', '0', '0', '1', '0', '0', '1', '1', '172.46', '1', '66.59', '1', '2.06', '1', '0', '1', '0', '1', '0', '2',	Primary hypothyroid	Primary hypothyroid	Pass
02.	'jyothi', '43', '00', '1', '0', '0', '0', '0', '0', '0', '0', '0', '0', '0', '0', '0', '0', '1', '1', '8.96', '1', '104', '1', '7.6', '1', '0', '1', '1.02', '1', '0', '2'	hyperthyroid	Secondary hypothyroid	Fail
03.	'laxmi', '40', '00', '1', '0', '0', '0', '0', '0', '0', '0', '0', '0', '0', '0', '0', '1', '1', '1.50', '1', '80', '1', '7', '1', '1', '1', '0', '1', '0', '5',	Primary hypothyroid	Primary hypothyroid	Pass
04.	'87', '61', '69', '56', '30', '2'	Liver disorder detected	Liver disorder detected	Pass
05.	'20', '23', '8', '37', '44', '1'	Liver disorder detected	Liver disorder not detected	Fail

Table4: Testing Table

CONCLUSION AND FUTURE SCOPE

As the nation is moving towards Digital India, we have also taken one step towards it by creating an application to detect Thyroid disease and Liver Disorder. In this tried to detect Thyroid and Liver Disorder by introducing Ensemble Learning. We have used many supervised ,Boosting and Bagging algorithm along with Voting classifiers. That best result in case of Hypo-Thyroid was obtained by using the Bagging using Linear Discriminant Analysis and Adaboost using Decision Tree with an accuracy of 99.8% and in case of Hyper-thyroid combination of Bagging using decision tree and Adaboost using Random Forest and Bagging using LDA are providing highest frequency which is 99.1%. The best accuracy in case of Liver Disorder is found by Bagging using Decision Tree algo with accuracy of 82.6%. Even if we get better accuracy in case of Thyroid Detection and Liver Disorder from our proposed models, further improvement is desired in both the cases. We will extend our project to other major diseases in future.

GANTT CHART

Here, is the Gantt chart of our project in which in the rows we have mentioned all the tasks done till current date and in the column we have mentioned the months.

		JAN	FEB	MAR	APR	MAY
Communication with mentor						
Planning						
App Design(Front end) & Database(Back-end)						
Discussion with mentor						
Applying Algorithms						
Mid -Term Evaluation						
Communication with mentor for further discussion						
Expanding our project to other diseases & applying various ML Algorithms.						
Documentation						
Final -Term Evaluation						

Gant Chart of our Minor Project

REFERENCES

RESEARCH PAPER:-

[1] Fatimah Saiti,Afsaneh Alavi Naini,Mahdi Aliyari shoorehdeli, Mohammad Teshnehlab "Thyroid Disease Diagnosis Based on Genetic Algorithms using PNN and SVM" IEEE , July 2009

[2] Jamil Ahmed , M. Abdul Rehman Soomrani , " TDTD: Thyroid Disease Type Diagnostics" IEEE , May 2016

[3] K.Pavya , B.Srinivasan," Feature Selection Algorithm to improve thyroid disease diagnosis" IEEE, November 2017

[4] Dr. R.R.Janghel,Dr. Anupam Shukla,Kshitij Verma Soft Computing Based Expert System for Hepatitis and Liver Disorders, IEEE March 2016

[5] H. Altay Guvenir , Aynur WEIGHTED KNEAREST NEIGHBOR CLASSIFICATION ON FEATURE PROJECTIONS, Research Gate October 2000

[6] Shanu Shroff, Siddhi Pise, Pratiksha Chalekar,Suja S. PanickerThyroid disease diagnosis: A survey, IEEE October 2015

ONLINE:-

[1] http://scikit-learn.org/

[2] www.youtube.com

[3] https://www.ieee.org/

[4] https://machinelearningmastery.com/category/python-machine-learning/

[5] https://pythonprogramming.net/machine-learning-tutorial-python-introduction/