StatisticalMachine Learning Approaches to Liver DiseasePrediction

TEAM ID: PNT2022TMID14570

1. INTRODUCTION

Project Overview

It is essential to diagnose liver disease early in order to save lives and take the necessary steps to control the condition. The ensemble approach has been successfully used in a number of sectors, particularly in the field of medical science. This study examines the early diagnosis of liver illness using a variety of ensemblemethodologies.

Purpose

Our project's objective is to analyse data from liver patients with a focus on the correlations between a vital list of liver enzymes, proteins, age and gender using them to try and predict the likeliness of liver disease. We are creating a model in this instance by utilizing different machine learning algorithms to identify the most accurate model. and incorporate into web applications built with flask. By entering certain parameters into the web application, users can predict the disease.

2. LITERATURE SURVEY

a. Existing problem

The liver related diseases are identified by analysing liver function blood test reports and scan reports. This paper while employing different data mining algorithms to ease this process, it is possible to reduce the time for diagnosing the liver disease. But it doesn't assisting making effective decisions and some accuracy Issues.

b. References

1. **TITLE:** "A Comparative Analysis of Unsupervised Machine Techniques forLiver Disease Prediction"

AUTHOR: Varun Vats

YEAR: 2018

He considered three different ML (Machine Learning) algorithms. A comparison of these algorithms had been carried out for evaluating their forecasting accuracy and computing intricacy. These algorithms included AP(Affinity Propagation), K means and DBSCAN. This work was dedicated to the

medical dataset based on lever disorders. This work made use of the Silhouette coefficient to measure the comparative efficiency of the considered algorithmic approaches.

2. TITLE: "Prognosis of Liver Disease using Machine LearningAlgorithms"

AUTHOR: Vyshali J Gogi

YEAR: 2018

He stated that the healthcare sector had a lot of data but this data was of no use [17]. This ample data required a leading analytic tool so that the hidden relationship and the valuable knowledge could be determined. The liver disease referred to the medical condition of the human liver-related to the human liver. The liver diseases led to sudden changes in health conditions that governed the functioning of the liver affecting other internal body organs. This work made use of several classification algorithms based on datamining. These algorithms included DT (Decision Tree), LD (Linear Discriminant), SVM Fine Gaussian, and LR (Logistic Regression). This work made use of Lab-based metrics of patients in the form of a liver dataset.

3. **TITLE:** "Accuracy Prediction Using Machine Learning Techniques for IndianPatient Liver Disease"

AUTHOR: Auxilia
YEAR: 2018

She stated that the use of medical datasets had attracted the medical experts globally [18]. The use of ML (Machine Learning) algorithms was quite common as a branch of making selection expressively helpful networks for the prediction of diseases by arranging therapy-based datasets. Grouping schemes had been generally employed as a segment of the curative domain for extracting order more efficiently as compared to a signal classification model. The disorders of

the Liver malady could be described as liver damage or sickness. Liver disorder can be categorized into several categories. This work made use of standard Indian liver illness patient records as a database for providing support to the researcher

4. **TITLE:** "New Fuzzy-ANWKNN algorithm for the successful prediction of liverdisorder" **AUTHOR**: Pushpendra Kumar

YEAR: 2019

He stated that it was a very difficult task for the doctors to detect the consequences of liver disorders on a person. In general, researchers used datasets based on LFT (Liver Function Test) for implementing classification algorithms so that the predictions about liver disorders could be generated. The dataset based on ground truth had several problems such as a class imbalance in the liver disorder data. With regard to the majority classes, the classic algorithms of classification generated influenced outcomes. This work presented a new Fuzzy-ANWKNN algorithm for the successful prediction of liver disorder.

5. TITLE: "To accurately predict liver disorder by means of several data mining algorithms"

AUTHOR: Sanjay Kumar

YEAR: 2018

He described different classification approaches by implementing them on the dataset of patients suffering from liver diseases . The main objective herewas to accurately predict liver disorder by means of several data mining algorithms. This work performed the analysis using the dataset of real-time patients to build classification paradigms for the prediction of liver diseases. This work implemented five classification algorithms on the used dataset. This work analysed different metrics such as precision, recall, and accuracyfor

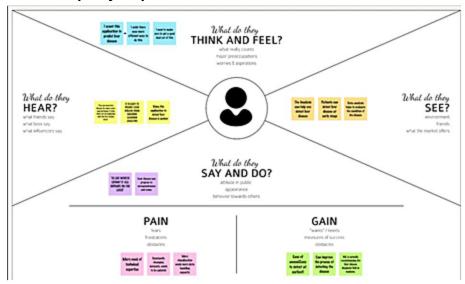
determining the efficiency of the implemented classification models.

c. :ProblemStatement Definition

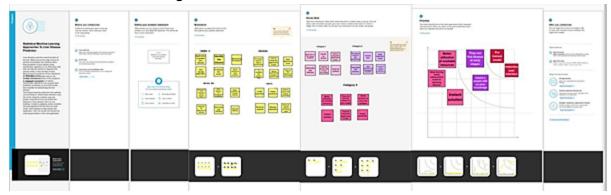
The main objective of this project is to analyse the parameters of various classification algorithms and compare their predictive accuracies so as to find out the best classifier for determining the liver disease. Here we are building a modelby applying various machine learning algorithms to find the best accurate model and integrate it to a flask-based web application. User can predict the disease by entering parameters in the web application. ML algorithms are new techniques to handle many hidden problems in medical data sets. This approach can help healthcare management and professionals to explore better results in numerous clinical applications, such as medical image processing, language processing, and tumor or cancer cell detection, by finding appropriate features.

3. IDEATION & PROPOSEDSOLUTION

Empathy Map Canvas



Ideation & Brainstorming



a. Proposed Solution

S.No.	Parameter	Description
1.	Problem Statement	Statistical Machine Learning Approaches to Liver Disease Prediction
2.	Solution description	Analyze the parameters of various classification algorithms and compare their predictive accuracies so as to find out the best classifier for determining the liver disease.
3.	Performance	Effective development of application
4.	Social Impact / Customer Satisfaction	With the help of this analysis patients can detect liver disease an early stage and evaluate the condition of the disease
5.	Availability	24/7 service
6.	Precision of Disease	Based on accuracy and better performance of implemented classifiers, the development of application to predict liver disease of the patient.

b. Problem Solution fit

The Problem-Solution Fit simply means that you have found a problem with your customer and that the solutionyou have realized for it actually solves the customer's problem. It helps entrepreneurs, marketers and corporate innovators identify behavioral patterns

Purpose:

Solve complex problems in a way that fits the state of your customers. Succeed faster and increase your solution adoption by tapping into existing mediums and channels of behavior. Sharpen your communication and marketing strategy with the right triggers and messaging. Increase touch-points with your company by finding the right problem-behavior fit and buildingtrust by solving frequentannoyances, or urgent orcostly problems.

4. REQUIREMENT ANALYSIS

Functional requirement

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement	Sub Requirement		
FR-1	User Registration	Registration through Form		
FR-2	User Confirmation	Confirmation via Email		
FR-3	User Profile	Filling the profile page after logging in		
FR-4	Uploading dataset	Patient data needed to upload		
FR-5	Requesting solution	Patient data is compared with pre-defined data and solution is generated		
FR-6	Solution display	The predicted solution is displayed on the screen		

Non-Functional requirement

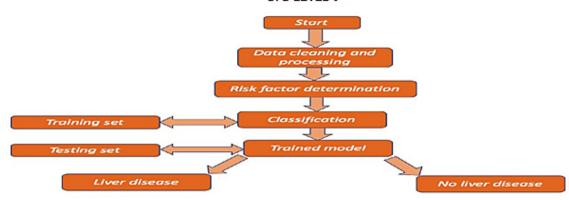
Following are the non-functional requirements of the proposed solution.

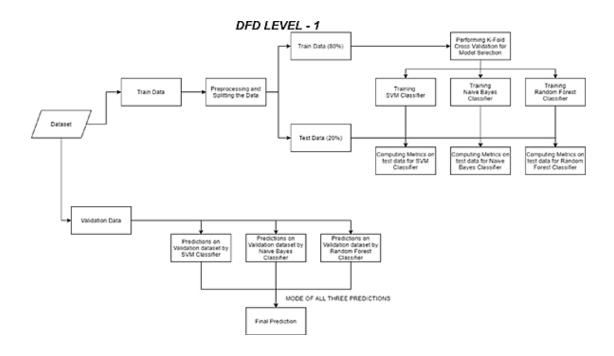
FR No.	Non-Functional Requirement	Description			
NFR-1	Usability	The system allows the user to perform the tasks easily, effectively and efficiently.			
NFR-2	Security	Assuring all data inside the system or its parts to be protected against malware attacks or unauthorized access.			
NFR-3	Reliability	The website does not recover from failure quickly as it is running on a single server.			
NFR-4	Performance	Response time and net processing time is fast.			
NFR-5	Availability	The system will be available upto 95% of time			
NFR-6	Scalability	The website is scalable			

5. PROJECT DESIGN

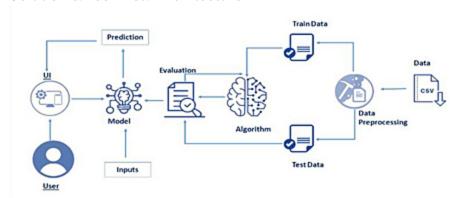
5.1: Data Flow Diagrams

DFD LEVEL-0





Solution & Technical Architecture



User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard	High	Sprint-1
	Login	USN-2	As a user, I can log into the application by entering email & password	I can login using my Email ID accounts or user credentials	High	Sprint-1
	Dashboard	USN-3	As a user, I can view the page of the application where I can upload my data and the result is predicted	I can access my account/ dashboard	High	Sprint-2
Customer (Web user)	Registration	USN-4	As a user, I can login to web dashboard just Like website dashboard	I can register using my username and password	Medium	Sprint-3
	Login	USN-5	As a user, I can login to my web dashboard with the login credentials	I can login using my User credentials	High	Sprint-3
	Dashboard	USN-6	As a user, I can view the web application where i can upload my data and the result is predicted	I can access my account/ dashboard	High	Sprint-4
		USN-7	As a user, the result predicted to me should be of higher accuracy	I can access my account/ dashboard	High	Sprint-4
Care Executive	Login	USN-8	As a care executive, I can login to the website using my login credentials	I can login to the website using my login credentials	High	Sprint-5
	Dashboard	USN-9	As a care executive, I can view the dashboard of application	I can access my dashboard	High	Sprint-5

6. PROJECT PLANNING & SCHEDULING

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	10	5 Days	22 Oct 2022	26 Oct 2022	10	30 Oct 2022
Sprint-2	15	10 Days	26 Oct 2022	04 Nov 2022	15	06 Nov 2022
Sprint-3	15	10 Days	04 Nov 2022	13 Nov 2022	15	13 Nov 2022
Sprint-4	10	7 Days	13 Nov 2022	19 Nov 2022	10	20 Nov 2022

7. Coding and Solutioning

7.1 Featuring

```
a =float(request.form['ar'])
agr =float(request.form['agr'])

X=[[age,gender,tb,db,ap,aal,aa2,tp,a,agr]]

# NOTE: manually define and pass the array(s) of values to be scored in the next line

payload_scoring = ("input_data": [{"field": [['age','gender','tb','db','ap','aa1','aa2','tp','a','agr']], "values": X}]}

response_scoring = requests.post('https://us-south.ml.cloud.ibm.com/ml/v4/deployments/58fa0711-e9d9-4847-adca-ce45c5a43f93/predictions?version=2022-11-18', json=paylo-headers=('Authorization': 'Bearer' + mltoken))
predictions = response_scoring_ison()

predict = predictions = response_scoring_ison()

predict = predictions['predictions'][0]['values'][0][0]

| return render_template('Output.html',prediction = 'You have a liver disease and you must take treatment')

| return render_template('Output.html',prediction = 'You have dont have liver')

| app.run(debug-True)
```

7.2 Feature 2

```
| Teacher | Teac
```

```
| deader class="jumbotron" | deader class="jumbotron" | div class="row row-hodder" style="padding-top: 60px;" | div class="col-12 col-1m-8" | div class="col-12 col-1m-8" | div class="col-12 col-1m-8" | div class="col-12 col-1m-8" | div class="auto" | div class
```

```
| discpanding in the color of your urine or stool.
| discpanding easily.
|
```

```
| clock two state | clock | cl
```

8. TESTING

8.1 TestCases

- The home page and the result page is tested .It is working well without issues.
- · The app was tested for functionality.
- The scoring end point application is slower than the normal flask app.

8.2 User AcceptanceTesting

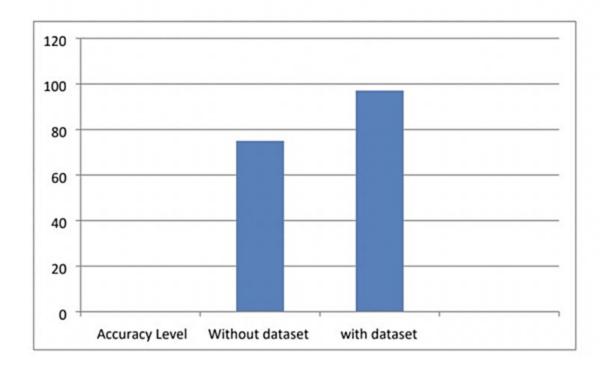
This report shows the number of test cases that have passed, failed, and untested

Section	Total Cases	Not Tested	Fail	Pass
Print Engine	8	0	0	8
Client Application	48	0	0	48
Security	2	0	0	2

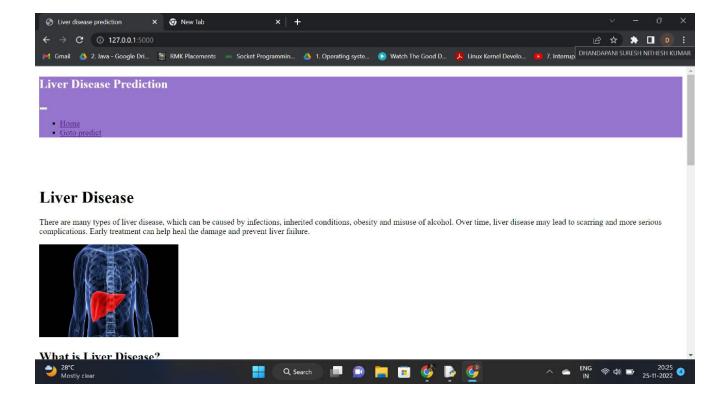
Outsource Shipping	3	0	0	3
Exception Reporting	7	0	0	7
Final Report Output	4	0	0	4
Version Control	3	0	0	3

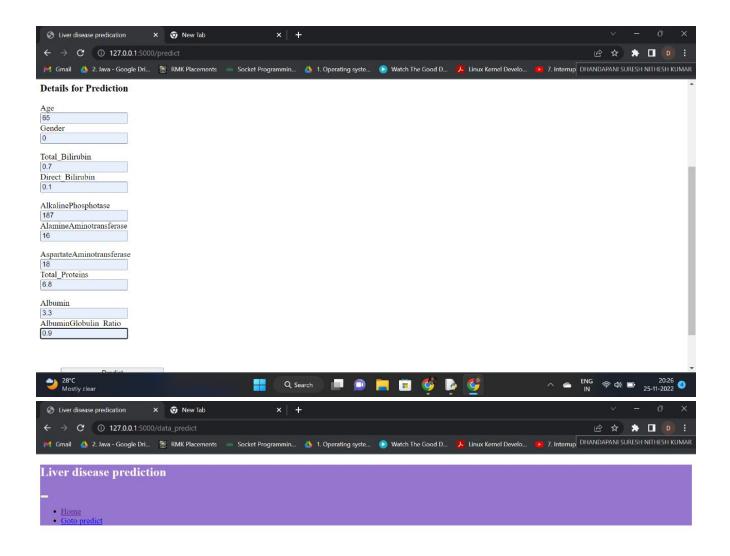
9. RESULTS

9.1 Performance Metrics



9.2 OUTPUT





PREDICTION RESULT

You have a liver disease and you must take treatment



10.Conclusion

This projectthis will,

- 1. Help healthcare management and professionals to explore better results in numerous clinical applications, such as medical image processing, language processing, and tumor orcancer cell detection, by finding appropriate features.
- Several statistical and machine learning approaches (e.g., simulation modeling, classification, and inference) have been used by researchers and lab technicians for betterprediction. The clinical results are more data-driven than model-dependent

1. ADVANTAGES AND DISADVANTAGES

Advantages:

- 1. This helps in early diagnosis of liver disease.
- 2. It makes the process simple and easier.
- 3. We are able to monitor the patient in an effective manner.
- 4. We can also help patients who are not in a condition to directly consulta doctor.

Disadvantages:

- 5. There are inaccuracies which cannot be avoided.
- 6. For people who have difficulty in accessing internetor those who do not have electronic

gadgets, this is not a feasible solution.

2. FUTURE SCOPE

- 1. In biomedical science, accuracy and speed are two important factors that should beconsidered chiefly in dealing with any disease.
- In this regard, Machine Learning techniques can be of great help to physicians. Withadvances, several machines have enteredin our lives.

2. One of the most famous areas where computers as the mostlyused machines can be helpful is knowledge extraction with the help of a machine (machinelearning).

13.Appendix

13.1 Github and Project Demo Link https://github.com/IBM-EPBL/IBM-Project-31056-1660195526

Project Demo Link

https://github.com/IBM-EPBL/IBM-Project-31056-1660195526/blob/main/Output%20video.mp4