

MIT School of Engineering
Department of Computer Science and Engineering

Project Synopsis

Group ID: 2

Project Title: MACHINE LEARNING PROJECT FOR DISEASE PREDICTION

Name of the Guide: Amol Dande

Group Members: Vince Francis Dsouza
Abhishek Tekavade
Aaditya Nair
Tanuj Sinha

Enrollment Number	Roll No.	Name of student	Email Id	Contact Number
MITU20BTCS0342	2203221	Vince Francis Dsouza	vincedsouza02@gmail.com	9767851524
MITU20BTCS0008	2203003	Abhishek Tekavade	abhishektekavade@gmail.com	9922839147
MITU20BTCS0314	2203046	Tanuj Sinha	tanujsinha2002@gmail.com	8484040035
MITU20BTCS0003	2203391	Aaditya Nair	aadityanair1803@gmail.com	7757984061

PROBLEM STATEMENT:

In day to day life, it is difficult for one person to go to a doctor and get a check of diseases as it takes quite a lot amount of money as well as time. In our developing and technology dependent life we totally rely on gadgets. So, there should be a way, with whose help a person can at least check whether he has a particular disease or not. Using tech like machine learning in predicting the diseases using symptoms or concerned medical data is a need of the hour

ABSTRACT:

Machine Learning and its approaches are generally helpful for healthcare and biomedical sectors for predicting the disease. For trivial symptoms, the difficulty is to meet the doctors at any time in the hospital. Big Data provides essential data regarding the diseases on the basis of the patient's symptoms. For several medical organizations, disease prediction is important for making the best feasible health care decisions. Conversely, the conventional medical care model offers input as structured that requires more accurate and consistent prediction. This project is planned to develop the multi-disease prediction using the various machine learning concept.

Machine Learning Approach for Identifying Disease Prediction Using Machine Learning is based on prediction modelling that predicts disease of the patients according to the symptoms provided by the users as an input / output to the system.

Here, the different datasets pertain to "Diabetes, lung cancer, liver and heart disease ", from the benchmark UCI repository is gathered for conducting the experiment also few datasets are taken from Kaggle.

Here we will use the concept of supervised Machine Learning in which implementation will be done by applying Decision Tree, Random Forest and few other algorithms which will help in early prediction of diseases accurately and better patients care. The results ensured that the system would be functional and user oriented for patients for timely diagnoses of diseases in a patient.

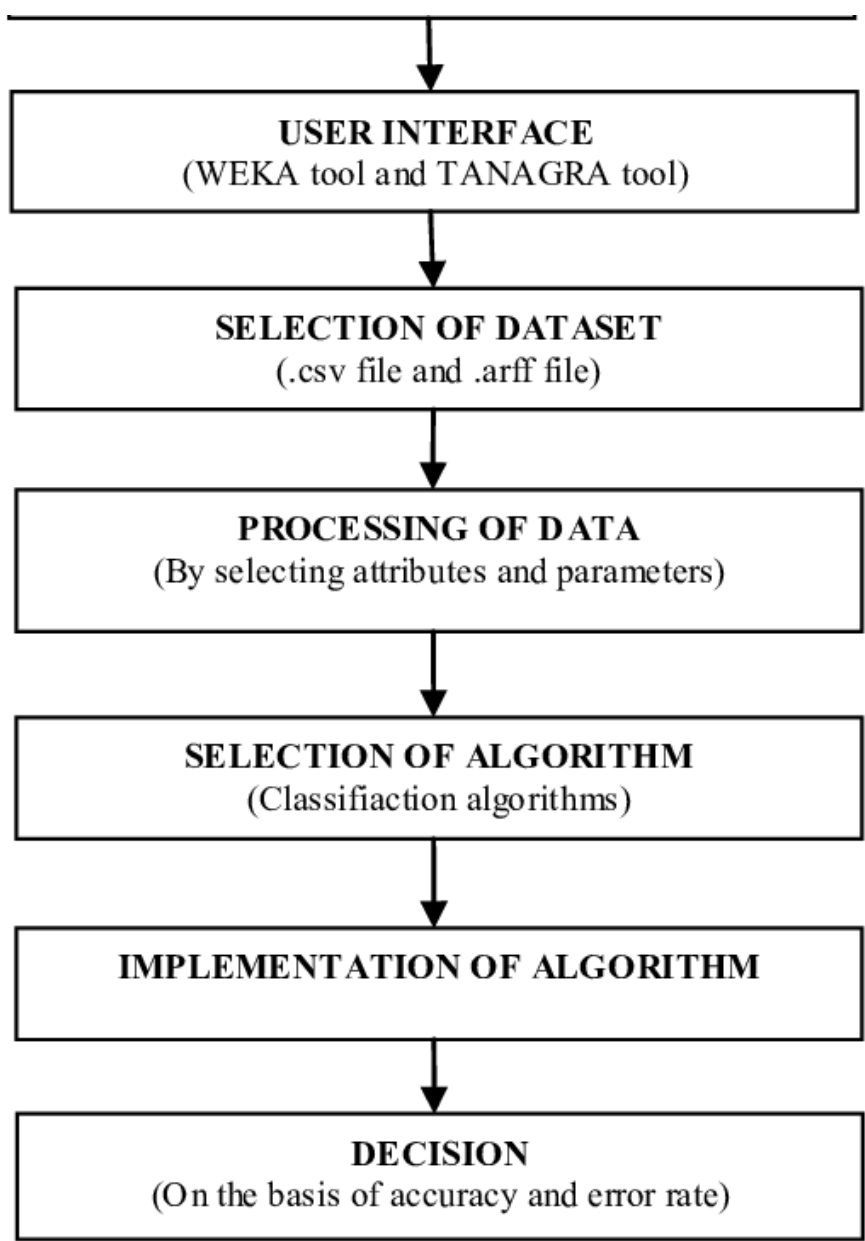
Literature Survey: Detail survey done

Sr No.	Paper Name	Author	Year of Publication	objective	Methodology	conclusion
1	Heart disease prediction using machine learning techniques : a survey	V.V. Ramalingam*, Ayantan Dandapath, M Karthik Raja	28/07/2018	To predict the presence or absence of heart related diseases accurately.	The following components have been used in this research paper <ul style="list-style-type: none">○ Naïve Bayes○ Support Vector Machine○ Decision Tree	There is a huge scope for machine learning algorithms in predicting cardiovascular diseases or heart related diseases. Each of the above-mentioned algorithms have performed extremely well in some cases but poorly in some other cases. Alternating decision trees when used with PCA, have performed extremely well but decision trees have performed very poorly in some other cases which could be due to overfitting. Ensemble models have performed very well because they solve the problem of overfitting by employing multiple algorithms (multiple Decision Trees in case of Random Forest). Systems based on machine learning algorithms and techniques have been very accurate in predicting the heart related diseases but still there is a lot scope of research to be done on how to handle high dimensional data and overfitting. A lot of research can also be done on the

						correct ensemble of algorithms to use for a particular type of data.
2	Heart Disease Prediction Using Machine learning and Data Mining Technique	Jaymin Patel, Prof. Teja IUpadhyay, Dr. Samir Patel	28/07/2018	The prediction system should not assume any prior knowledge about the patient records it is comparing. • The chosen system must be scalable to run against large database with thousands of data.	The following components have been used in this research paper <ul style="list-style-type: none"> Classification Tree Algorithms Used J48 algorithm: J48 with Reduced error Pruning: 	it is concluded that J48 tree technique turned out to be best classifier for heart disease prediction because it contains more accuracy and least total time to build. We can clearly see that highest accuracy belongs to J48 algorithm with reduced error pruning followed by LMT and Random Forest algorithm respectively. Also observed that applying reduced error pruning to J48 results in higher performance while without pruning, it results in lower Performance. The best algorithm J48 based on UCI data has the highest accuracy i.e. 56.76% and the total time to build model is 0.04 seconds while LMT algorithm has the lowest accuracy i.e. 55.77% and the total time to build model is 0.39seconds. In conclusion, as identified through the literature review, we believe only a marginal success is achieved in the creation of predictive model for heart disease patients and hence there is a need for combinational and more complex models to increase the accuracy of predicting the early onset of heart disease
3	Review of the State-of-the-Art of Brain-Controlled Vehicles	Amin Hekmatm Anesh, Pedro H J Nardelli	18/06/2022	It focuses on the most relevant topics on brain-controlled vehicles, with a special reference	The following components have been used in this research paper <ul style="list-style-type: none"> Bio-signal patterns 	They provide a systematic presentation of the most significant literature in the topic of BCV and BICAV from the past ten years

				to the terrestrial BCV and the aerial BCV, also called BCAV controlled by using bio-signals, such as electroencephalogram (EEG), Electrooculogram (EOG), and Electromyogram (EMG)	<ul style="list-style-type: none"> • machinelearning artificial intelligence simulator, • Aerial vehicle. 	
--	--	--	--	---	---	--

PROPOSED SYSTEM (BLOCK DIAGRAM):



CONCLUSION:

The project presented the technique of predicting the disease based on the symptoms, age, and gender of an individual patient. Different Machine learning algorithms were used to carry out the project such as the Random Forest, Decision Tree and Logistic Regression. Almost all the ML models gave good accuracy values. As some models were dependent on the parameters, they couldn't predict the disease and the accuracy percentage was quite low.

Once the disease is predicted, we could easily manage the medicine resources required for the treatment. This model would help in lowering the cost required in dealing with the disease and would also improve the recovery process.

We have also created a GUI for better interaction with the system by users which is very easy to operate. This paper shows that Machine Learning algorithm can be used to predict the disease easily with different parameters and models. To conclude, our system is helpful to those people who are always worrying about their health and they need to know what happens with their body. Our main motto to develop this system is to know them for their health. Especially, people who are suffering from mental illness like depression, anxiety.

They can come out of these problems and can live their daily lives easily. Besides, our system provides better accuracy of disease prediction according to symptoms of the user, and also it will provide motivational thoughts and images. In the end, we can say that our system has no boundary of the user because everyone can use this system.

ANNEXURE:

ANNEXURE I: FORM A-TITLE APPROVAL

MIT School of Engineering
Department of Computer Science and Engineering
Mini Project (I, II, III, IV) Topic Approval

Date: 05/02/2023

Class: TY CORE 2

Project Group ID: 02

Group Members:

Enrolment No	Name of Student	E-mail Address	Contact
MITU20BTCS0342	Vince Dsouza	vincedsouza02@gmail.com	9767851524
MITU20BTCS0003	Aditya Nair	aadityanair1803@gmail.com	7757984061
MITU20BTCS0008	Abhishek Tekavade	abhishektekavade@gmail.com	9922839147
MITU20BTCS0314	Tanuj Sinha	tanujsinha2002@gmail.com	8484040035

Project Title Evaluation Parameters:

Sr No.	Parameters	Topic 1	Topic 2	Topic 3
1	Title	Disease Prediction	Stock analysis and prediction	Sentimental analysis
2	Domain Expertise	Domain expertise can provide valuable insights into the development of disease prediction models and help to ensure that the models are accurate and effective.	Domain expertise in stock analysis and prediction helps to ensure that the predictive models are built on a solid understanding of the stock market and its underlying factors.	Sentiment analysis involves a deep understanding of natural language processing, machine learning, and statistical analysis techniques.
3	Technical Feasibility	This technique can be used to predict the likelihood of an individual developing	With the help of stock prediction and analysis, it is now possible to predict the behaviour of	Sentiment analysis is a field of natural language processing that involves the identification and

		certain diseases based on various risk factors such as genetic information, lifestyle factors, environmental factors, and medical history.	stock prices based on various market indicators, such as price trends, trading volumes, and economic indicators.	extraction of subjective information from text data, such as opinions, emotions, and attitudes. It has become an important tool for businesses, governments, and researchers to understand public opinion and sentiment.
4	Future Scope	<p>The future of disease prediction is promising, and there is a significant scope for its development and expansion. Some of the potential future developments in disease prediction include –</p> <ol style="list-style-type: none"> 1. Prevention and early intervention: Disease prediction models can be used to identify individuals at high risk of developing a particular disease and provide early intervention and prevention measures. This approach can help to reduce the incidence of diseases and improve overall health outcomes. 2. Global health monitoring: Disease prediction models can be used to monitor global health trends and detect emerging outbreaks of infectious diseases. This approach can help to prevent the spread of diseases and improve public health outcomes. 	<p>Some of the potential future developments in stock analysis and prediction include–</p> <ol style="list-style-type: none"> 1. Use of machine learning and artificial intelligence: Advancements in machine learning and artificial intelligence can help to develop more accurate and efficient predictive models. These models can analyse large volumes of data and identify complex patterns and relationships that may be missed by traditional statistical models. 2. Integration of alternative data sources: Alternative data sources such as social media sentiment, satellite imagery, and news articles can provide valuable insights into the stock market and its underlying factors. The integration of these data sources can help to improve the accuracy of predictive models. 	<p>The potential future scopes in sentiment analysis are --</p> <ol style="list-style-type: none"> 1. Analysis of audio and visual data: Sentiment analysis can be extended beyond text data to include audio and visual data such as speech and facial expressions. This can provide more comprehensive insights into the emotions and attitudes of individuals and groups. 2. Use in healthcare: Sentiment analysis can be used in healthcare to analyse patient feedback, identify trends and patterns, and improve patient satisfaction.
5	Applicability	Disease prediction has numerous applications in the healthcare industry, including-	Stock analysis and prediction has numerous applications across the finance industry, including-	Some of the most common applications of sentiment analysis include- 1. Customer service: Sentiment analysis can be

		<p>1.Early detection and prevention: Disease prediction can help healthcare providers identify patients who are at high risk of developing certain diseases or conditions. Early detection and prevention can lead to better health outcomes and reduce healthcare costs.</p> <p>2.Personalized treatment: Disease prediction can help healthcare providers develop personalized treatment plans based on an individual's risk factors and medical history.</p> <p>3.Population health management: Disease prediction can help healthcare providers identify and address health disparities and other population health issues.</p>	<p>1. Investment management: Stock analysis and prediction is used by investment managers to analyse market trends and predict future stock prices. This information is used to make informed investment decisions and manage investment portfolios.</p> <p>2.Trading: Stock analysis and prediction is used by traders to identify patterns and trends in stock prices, and to develop trading strategies that can maximize profits.</p>	<p>used to analyse customer feedback and complaints to identify patterns and trends in customer service issues. This information can be used to improve customer service processes and address issues more effectively.</p> <p>2. Healthcare: Sentiment analysis can be used to analyse patient feedback and improve patient satisfaction. It can also be used to monitor social media and other online sources to identify potential outbreaks of infectious diseases.</p>
Approved				
Remarks				

Sr. No. Name of Subject Expert

Signature

1. _____

2. _____

ANNEXURE II: FORM B - MARKET AND FINANCIAL FEASIBILITY

MIT School of Engineering

Department of Computer Science and Engineering

Viability Analysis Report (Filled by student and verify by guide)

Date: 20/03/2023

Class: TY CORE-2

Project Group ID: 02

Project Title: Disease Prediction

PROJECT TITLE EVALUATION PARAMETERS:

Sr. No.	Parameters	Description About Project	Marks (5)
1.	Business Ideas and Implementation from project	Data gathered from the project can be implemented to create more awareness in the health sector. Therefore, changing lives.	
2.	Market Survey (competitors, substitute products, potential market, etc.)	The Health Sector is an extensive and vast sector which will always be at the top of every list, both in an organization and on an individual level. The industry is estimated to generate a revenue of \$744.4 billion by 2030. Our potential market Includes – Well wise, Cure fit , Fit cloud and Health spot.	
3.	Market Acceptability of Product	The global Healthcare informatics market size is projected to reach USD 386820 million by 2028 , from USD 218500 million in 2021 as health related issues become more serious and fatal a need for Healthcare application software will also increase	
4.	Emerging Trends about Project and Product	Increase in the awareness of health and its related diseases in public on an individual level and increase in the same due to dependable software	
5.	Income Generation ideas through Project	This project can easily be turned into an income generation opportunity through – <ul style="list-style-type: none">• Algorithm can be patented and distributed providing profit with each installation• Data collected can be turned into profit	

6.	Project Profitability	We believe this project will be extremely profitable and viable. It can be turned into a connecting interface between the user and the doctor in emergencies .we plan on adopting this as a personal project and thus will take all measures to make this projectprofitable!	
7.	Cost Benefit Analysis	Serving community	
8.	Any Other Point	Our product will provide a fresh perspective to individualistic health and also act as an alarm with an interface that will be easy to use.	
Remark:			

Commercial Feasibility of project is evaluated based on the above parameters.

Project Approval Status: Approved / Not Approved

**(Name & Designation of Market
Expert) Signature with Date.**

ANNEXURE III: LITERATURE SURVEY PAPER OR LINKS

1. Shrestha, Ranjit , & Chatterjee, Jyotir. Moy. (2019).Heart Disease Prediction System Using Machine Learning. LBEF Research Journal of Science Technology and Management , 1(2), 115-132
2. Godse, Rudra A., Gunjal, Smita S., JagtapKaran A .,Mahamuni ,Neha S., &Wankhade, Prof. Suchita. (2019). Multiple Disease Prediction Using Different Machine Learning Algorithms Comparatively.International Journal of Advance Research in Computer and Communication Engineering, 8(12), 50-52
3. A. Gavhane, G. Kokkula, I. Pandya, and K. Devadkar, "Prediction of heart disease using machine learning," in 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2018, pp. 1275–1278.
4. Y. Hasija, N. Garg, and S. Sourav, "Automated detection of dermatological disorders through image-processing and machine learning," in 2017 International Conference on Intelligent Sustainable Systems (ICISS), 2017, pp. 1047–1051
5. S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction," BMC Medical Informatics and Decision Making, vol. 19, no. 1, pp. 1– 16, 2019