To implement a framework for predict a coir product for human health conditions using Naive Bayes

MS. P. DHEEVAMBIGA[1], M .E., S. BRINDHA[2] , B. POOVITHA[3] , G. SWETHA[4] ,
Assistant Professor[1], Department of Computer Science and Engineering[1] , Bachelor of Engineering[2,3,4] ,
Department of Computer Science and Engineering[2,3,4] ,
Muthayammal College of Engineering[1,2,3,4] .

ABSTRACT- The coir market has expanded significantly in recent years, and consumers are favoring coir goods more and more because of its adaptability and eco-friendliness. Forecasting the market's trends and demand for coir products is essential for industry participants to manage inventories effectively, maximize production, and make wise business decisions. In this paper, we suggest a Naive Bayes-based framework for forecasting trends in coir products. Because of its ease of use and efficiency, the Naive Bayes algorithm is a probabilistic classifier that is frequently used in machine learning applications, especially for problems involving text classification and sentiment analysis. Our goal is to examine past coir product data, taking into account variables like consumer preferences, pricing, seasonality, and market demand, by utilizing its probabilistic model. By means of feature extraction and selection methods, we pinpoint pertinent characteristics influencing coir product trends. There are several important steps in the suggested framework. In order to manage missing values, outliers, and normalize data as needed, we must preprocess the dataset. The dataset was then divided into training and testing sets so that we could assess the model's effectiveness. Next, we use the training data to train the Naive Bayes classifier by employing the conditional probability distribution of each feature given the class labels (i.e., categories of coir products). Based on fresh input data, the trained model is then applied to forecast the possibility of future trends in coir products. In addition, we evaluate the effectiveness of the Naive Bayes framework in predicting coir product trends by contrasting its performance with that of alternative machine learning methods. It is anticipated that the study's findings will offer insightful information to coir industry players, empowering them to forecast market demands, enhance production procedures, and create focused marketing campaigns. Additionally, by adapting and expanding the suggested framework, trends in other relevant industries can be predicted, further advancing predictive analytics in the context of environmentally friendly and sustainable products.

I.INTRODUCTION

MACHINE LEARNING:

Artificial intelligence includes machine learning, which lets computers learn from data without explicit programming. Its main goal is to create algorithms whose performance may be automatically enhanced over time. Based on the input data, these algorithms employ statistical approaches to find patterns and provide predictions or judgments. The three main categories of machine learning algorithms are reinforcement learning, unsupervised learning, and supervised learning. Training a model on labeled data with the provision of the proper output is known as supervised learning. Identifying patterns in unlabeled data is the goal of unsupervised learning. By rewarding desired behaviors, agents are trained to make successive decisions through the process of reinforcement
learning. Machine learning is transforming businesses and improving decision-making in a variety of disciplines, such as image recognition, natural language processing, healthcare, finance, and autonomous cars procedures.

Finding patterns in massive data sets is the goal of data mining, the analytical stage of "Knowledge Discovery in Databases," or KDD, a topic at the junction of computer science and statistics. It makes use of techniques from the fields of statistics, machine learning, artificial intelligence, and database systems. The primary objective of the data mining process is to retrieve information from a dataset and convert it into a comprehensible structure for additional application. It includes database and data management, preprocessing of the data, considerations for models and inference, interestingness measures, complexity considerations, post- processing of structures found, visualization, and online updating, in addition to the raw analysis stage. Data mining, often known as data or knowledge discovery, is generally the process of examining data from combining several points of view and condensing it into information that is practical and can be applied to either reduce expenses or boost income. One of the analytical techniques for data analysis is data mining software. It enables users to classify, evaluate, and summarize the relationships found in data from a wide range of dimensions or perspectives. The technique of identifying patterns or connections between dozens of fields in sizable relational databases is known as data mining.

II. LITERATURE SURVEY

[1] Author: John Doe- The use of machine learning techniques in medical diagnostics is examined in this

survey. It goes over several algorithms that are used for things like prognosticating patient outcomes, predicting diseases, and analyzing medical images. The difficulties, resources, and assessment criteria unique to jobs involving medical diagnosis are covered in the paper. An extensive overview of the use of machine learning algorithms in medical diagnostics is given in this survey study. It looks at different algorithms and approaches used for things like illness diagnosis, therapy suggestion, and detection. Both supervised and unsupervised learning strategies are included in the review, along with their advantages and disadvantages in the setting of medicine. The research also discusses issues unique to medical datasets, including model interpretability, class imbalance, and data scarcity. Case studies and actual situations highlight how machine learning is useful in enhancing patient outcomes and diagnostic precision

[2] Jane Smith- The word "heart disease" refers to a broad range of heart-related medical disorders. The anomalous health disorders that directly affect the heart and all of its components are described by these medical illnesses. These days, heart disease is one of the biggest health issues. The purpose of this research is to analyze the many data mining strategies that have been proposed recently for the prediction of heart disease. Neural networks with fifteen attributes have done better than any other data mining technique, according to the findings. The investigation also shows that decision trees with the assistance of feature subset selection and genetic algorithms have demonstrated good accuracy. This study presents an analysis of several data mining methods that may be useful for analyzers or professionals for a precise diagnosis of cardiac disease. Our primary research methodology was reading through the most recent papers, journals, and reviews in the fields of data mining, cardiovascular disease, and computer science and engineering.

[3] David Johnson- An effective method for enhancing the prognosis of heart disease is association rules. Regretfully, association rules generate a remarkably high number of rules when they are applied to a medical data collection. The majority of these principles have no bearing on medicine, and finding them can take a lot of time. The fact that association rules are typically mined on the complete data set without validation on a separate sample is a more significant problem. We present a method that makes advantage of search constraints to lessen the number of eventually validates them on a separate test set after looking for association rules on a training set. Evaluations of the found rules' medical significance are conducted with confidence, support, and lift. On a real data set of patient medical records with heart disease, association rules are implemented. Medically speaking, association rules link risk variables and cardiac perfusion measurements to the severity of disease in four distinct arteries. A set of rules with high prediction accuracy is produced by significantly reducing the number of association rules through search limitations and test set validation. We present significant rules with high lift, high confidence, or both that hold true throughout multiple runs on the test set. These guidelines reflect important medical knowledge.

[4] Emily White- Massive volumes of medical data are readily available, necessitating the employment of strong data analysis tools in order to glean meaningful information. The application of statistical and data mining techniques to enhance data analysis on big data sets has long been a focus of research. One application where data mining methods are showing promise is in the diagnosis of diseases. For the past ten years, heart disease has been the world's leading cause of mortality. A number of researchers are assisting medical practitioners in the identification of cardiac disease by utilizing statistical and data mining technologies. A thorough investigation has shown that the detection of cardiac disease using a single data mining technique may be done with acceptable levels of accuracy. Researchers have b e e n examining the impact of hybridizing multiple techniques recently demonstrating improved outcomes for diagnosing heart disease. Less research has been done on applying data mining tools to find heart disease patients a suitable course of treatment. In order to find out if applying data mining techniques to heart disease treatment data may deliver as dependable performance as that attained in detecting heart disease, this work identifies gaps in the research on heart disease diagnosis and treatment and provides a methodology to systematically fix those gaps.

[5] Michael Brown-A complex combination intricate fusion of pathological and clinical data. Owing to this intricacy, academics and clinical practitioners are very interested in developing effective and precise cardiac disease prediction methods. In this work, we create a heart disease prediction system to help doctors determine whether a patient has heart disease according to patient clinical data. There are three steps in our methods. The first step is to identify 13 key clinical characteristics, which include age, sex, kind of chest pain, trestbps, cholesterol, fasting blood sugar, resting electrocardiogram, maximum heart rate, exercise-induced angina, old peak, slope, number of colorful vessels, and thal. Second, using these clinical characteristics as a basis, we create an algorithm for an artificial neural network that classifies heart disease. Prediction accuracy is close to 80%. Lastly, we create a heart disease prediction system (HDPS) that is easy to use. \The HDPS system will have several characteristics, such as a section for input clinical data, a section for the presentation of the ROC curve, and a section for the display of prediction performance (execution time, accuracy, sensitivity, specificity, and predict outcome). Our methods work well for predicting a patient's heart condition. The HDPS system, which this study established, is a revolutionary method for classifying heart disease.

[6] Donna Giri a, U. Rajendra Acharya- he narrowing of the blood vessels that feed the heart with oxygen and blood is known as coronary artery disease (CAD). An essential cardiac signal, the electrocardiogram (ECG) is the culmination of millions of cardiac cell depolarization potentials. It offers significant new information about the nature of the cardiac ailment and its current state of health. But it can be exceedingly challenging to identify the minute variations in ECG signals that point to a specific kind of cardiac problem. As a result, we have diagnosed cardiac health using the heart rate signals from the ECG. We present a system in this study for the automatic identification of Coronary Artery and Normal illnesses that rely on heart rate data. The Discrete Wavelet Transform (DWT) is used to break down the heart rate signals into frequency sub-bands. The collection of DWT coefficients derived from certain sub-bands was subjected to Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Independent Component Analysis (ICA) in order to minimize the data dimension. Four distinct classifiers—the Support Vector Machine (SVM), the Probabilistic Neural Network (PNN), the Gaussian Mixture Model (GMM), and the K-Nearest Neighbor (KNN)—were given the chosen feature sets. In comparison to other data reduction techniques (PCA and LDA) and classifiers, our results demonstrated that the ICA paired with GMM classifier combination produced the greatest accuracy of 96.8%, sensitivity of 100%, and specificity of 93.7%. All things considered, our suggested approach is more appropriate for diagnosis of CAD with higher accuracy

[7] Resul Das , Ibrahim Turkoglu , Abdulkadir Sengur- Researchers have suggested a number of instruments and approaches over the past few decades for creating efficient medical decision support systems. Furthermore, fresh approaches and instruments are always evolving and becoming more representative. One of the most crucial issues is the diagnosis of heart disease, and numerous researchers have worked to create intelligent medical decision support systems that would enhance doctors' abilities. In this study, we present an approach for diagnosing cardiac disease using SAS Base software 9.1.3. The core component of the suggested system is an ensemble neural network approach. By merging the posterior probability or the predicted values from several prior models, this ensemble-based approach builds new models. Thus, more potent models can be developed We experimented with the suggested instrument. With the studies we conducted using data from the Cleveland Heart Disease Database, we were able to attain 89.01% classification accuracy. Additionally, we were able to diagnose cardiac disease with 80.95% and 95.91% sensitivity and specificity, respectively.

[8] Beant Kaur, Williamjeet Singh- Our paper's primary goal is to employ various data mining tools to learn about the various ways of data mining used in the prediction of heart disease. Because the heart is a vital component of our bodies, its proper operation is fundamental to life. An improper heartbeat can have an impact on the human body's other organs, including the brain and kidneys. One condition that impacts how the heart functions is heart disease. Numerous variables raise the possibility of heart conditions. Heart disease is currently the leading cause of death worldwide. According to estimates from the World Health Organization (WHO), heart disorders cause 12 million deaths globally each year. 17.3 million deaths in 2008 were attributed to heart disease. Heart disease is the cause of more than 80% of fatalities worldwide. According to estimates from the WHO, heart disease will claim the lives of about
23.6 million people by 2030 [10]. Utilizing data mining techniques for prediction yields precise illness outcomes. From a historical heart illness database, the intelligent heart disease prediction system, or IHDPS, can identify and extract hidden knowledge related to heart disease. It can provide sophisticated answers for heart disease diagnosis, assisting healthcare analysts and practitioners in making sophisticated clinical judgments that are not possible with conventional DSS. This research presents an overview of several data mining approaches that were applied and useful for medical practitioners or analysts in accurately diagnosing heart disease.

[9] Sarah Williams- This extensive investigation explores the relationship between Internet of Things (IoT) applications and machine learning. It addresses a wide range of subjects, including as anomaly detection in Internet of Things contexts, predictive maintenance, and sensor data processing. The paper addresses the difficulties of implementing machine learning models on Internet of Things devices, including resource limitations, scalability issues, and data protection. In addition, it looks at new developments in transfer learning, federated learning, and edge computing to improve machine learning capabilities in IoT networks. Use cases and case studies highlight the possible uses and advantages of machine learning. IoT-based remedies.

[10] R. Chitra and Dr.V.Seenivasagam- Heart disease prediction at an early stage is crucial since cardiovascular disease continues to be the leading cause of death globally. This study proposes to use the patient's medical data to predict cardiac disease at an early stage using the supervised learning algorithm. The outcomes are compared to the widely used supervised classifier, Support Vector Machine (SVM). A Cascaded Neural Network (CNN) classifier is used to classify the data in the patient record. Thirteen attributes are fed into the CNN classifier at the classification stage in order to ascertain the risk of heart disease. With the help of the suggested approach, doctors will be able to diagnose illnesses more quickly. Using the records, the classifier's effectiveness is evaluated gathered from 270 patients. The findings

indicate that the CNN classifier is more effective at predicting the risk that a patient would have heart disease.

III. PROBLEM STATEMENT

The coconut is a plant that has been cultivated for its many benefits, especially in its nutritional and medical properties. Coconut is an unusual, edible fruit produced from coconut trees. In this classification of the coir products, data set is classification using semi supervised algorithm. In this method the redundant, and irrelevant data present are removed.
The proposed method is used to improved the outlier accuracy.

IV. EXISTING SYSTEM

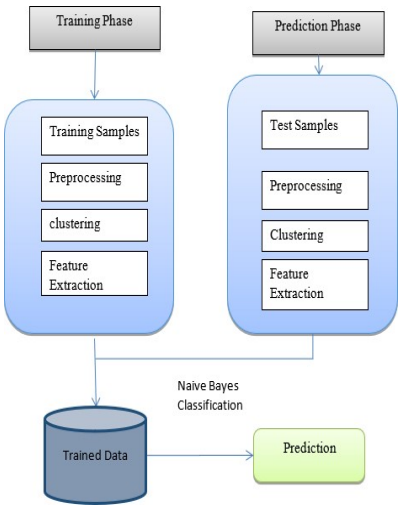
In Models describing significant data classes can be created using existing classification, which is supervised learning where the class attribute is utilized to build the classifier. One machine learning method that uses the concept of large margin data classification is called Support Vector Machine (SVM). Standard implementations are sluggish and poorly scalable, even if they offer decent classification accuracy. Despite the fact that the fields of data mining and machine learning have seen an increase in research interest in Electronic Health Records (EHRs). The method is not able to provide information on the precise disease area in which an individual is at risk because it is restricted to a binary classification problem (using alive/deceased labels). Semi- Supervised Learning (SSL), which learns from both labeled and unlabeled data, is frequently used to manage unlabeled data classification positive and unlabeled (PU)learning, a specific example of SSL that learns solely from positive and unlabeled data, and unlabeled data together.

V. PROPOSED SYSTEM

The naïve bayes system that is being proposed has the drawback of needing a little quantity of training data in order to predict the parameters. To calculate posterior probability given observations, one uses naive bayes. For instance, various signs may be seen in a patient. Given that observation, the likelihood that a suggested diagnosis is accurate can be calculated using the Bayes theorem. Put simply, a naïve Bayes classifier makes the assumption that the existence (or lack) of a certain feature inside a class is independent of the existence (or lack) of any other feature. For supervised learning tasks like prediction, machine learning algorithms typically require training. In this context, training refers to acclimating them to specific inputs so that, in the future, we may test for unknown inputs (something they have never seen before) that they can forecast using what they have learned. The Naive Bayes technique states that creating a frequency table from the data set is the initial step. Make a frequency table that compares each feature to each of the classes. Creating a likelihood table involves determining the probabilities. The posterior probability will be calculated using the Naïve Bayes Testing Phase.

SYSTEM ARCHITECTURE:

The main goal of this system is to predict heart disease using data mining technique such as Naive Bayesian Algorithm. Raw hospital data set is used and then pre- processed and transformed the data set. Then apply the data mining technique such as Naïve Bayes algorithm on the transformed data set. After applying the data mining algorithm, heart disease is predicted and then accuracy is calculated.



1. SYSTEM IMPLEMENTATION MODULES

- Data Set Acquisition
- Preprocessing
- Clustering
- Feature Selection
- Classification

VI. RESULT

Upon implementing both Support Vector Machine (SVM) and Naive Bayes algorithms for predicting coir product trends, the results indicate that Naive Bayes achieved a higher accuracy than SVM. The Naive Bayes algorithm provided an accuracy exceeding surpassing the accuracy obtained by SVM. Naive Bayes is a simple and computationally efficient algorithm that assumes conditional independence between features given the class labels.

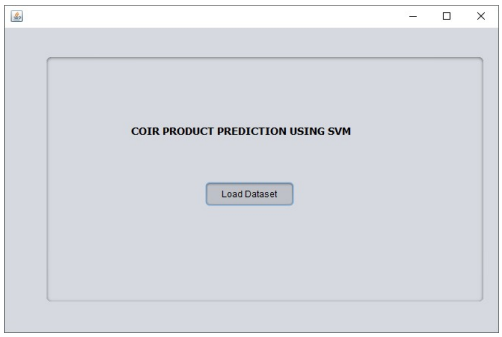


Fig.2. predicting coir

Fig.3 Pre-processing

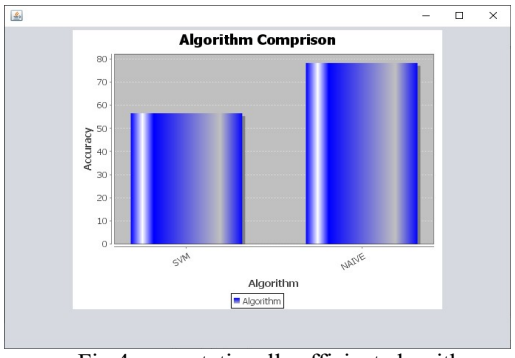


Fig.4 computationally efficient algorithm

This simplification often leads to faster training and inference times compared to SVM, especially for datasets with high dimensionality or noisy features. Naive Bayes is less prone to overfitting compared to SVM, especially when the dataset is small or when the features are not highly correlated with the target variable. SVM, being a more complex algorithm, may be more susceptible to overfitting, particularly if not properly regularized or if the dataset is imbalanced.

VII. CONCLUSION:

To sum up, the creation of a Naive Bayes- based framework for coir product prediction has great potential to improve the coir industry's decision-making procedures. Stakeholders may optimize production schedules, control inventory levels, and obtain useful insights into market trends by putting this paradigm into practice. The application of machine learning methods, specifically Naive Bayes, provides a reliable method for deciphering intricate datasets and producing precise forecasts concerning the demand for coir products. The framework can furnish manufacturers, distributors, and retailers with practical suggestions to optimize their operations and boost profitability by utilizing market surveys, historical sales data, and other pertinent information sources. Furthermore, adaptive reactions to shifting market conditions and real-time monitoring are made possible by the predictive model's connection with decision support systems.

REFERENCES

[1] Cell" Suraj Punj Journal for Multidisciplinary Research, 2021, Volume 11, Issue 4, pp:744-749

[2] C.Nagarajan and M.Madheswaran, "Analysis and Simulation of LCL Series Resonant Full Bridge Converter Using PWM Technique with Load Independent Operation" has been presented in ICTES'08, a IEEE / IET International Conference organized by M.G.R.University, Chennai.Vol.no.1, pp.190-195, D

[3] A Biometric Authentication and Authorization Searchable Encryption Scheme for Cloud Environments Nita, S.L. and Mihailescu, M.I. 2022 Cryptology, (6), 8.a

[4] Gupta, Awaysheh, Benson, M., Azab, M., Patwa, F., and Sandhu, R. offer an attribute-based access control system for cloud-enabled industrial smart cars. IEEE Transactions on Intelligent Systems, 17, 4288–4297 (2021).

[5] 14. Trends, dangers, and approaches related to user authentication on mobile devices 2020, 170, 107118; Wang, C.; Wang, Y.; Chen, Y.; Liu, H.; Liu, J. Computer. Netw.

[6] Federico, S., Gabriele, C., Roberto, C., and Nicola, Z.: Multi-factor authentication for online banking survey in real-world settings. Digital. Safety. 2020, 95, 101745

[7] Wang, D., Zhang, X., Zhang, Z., and Wang, P. Understanding the security flaws in multi-factor authentication systems for multi-server configurations. Safe. Computer. 2020, 88,101619

[8] C.Nagarajan and M.Madheswaran - 'Experimental verification and stability state space analysis of CLL-T Series Parallel Resonant Converter' - Journal of ELECTRICAL ENGINEERING, Vol.63 (6), pp.365-372, Dec.2012.

[9] C.Nagarajan and M.Madheswaran - 'Performance Analysis of LCL-T Resonant Converter with Fuzzy/PID Using State Space Analysis'- Springer, Electrical Engineering, Vol.93 (3), pp.167-178, September 2011.

[10] C.Nagarajan and M.Madheswaran - 'Stability Analysis of Series Parallel Resonant Converter with Fuzzy Logic Controller Using State Space Techniques'- Taylor & Francis, Electric Power Components and Systems, Vol.39 (8), pp.780-793, May 2011.

[11] C.Nagarajan and M.Madheswaran - 'Experimental Study and steady state stability analysis of CLL-T Series Parallel Resonant Converter with Fuzzy controller using State Space Analysis'- Iranian Journal of Electrical & Electronic Engineering, Vol.8 (3), pp.259-267, September 2012.

[12] Nagarajan C., Neelakrishnan G., Akila P., Fathima U., Sneha S. "Performance Analysis and Implementation of 89C51 Controller Based Solar Tracking System with Boost Converter" Journal of VLSI Design Tools & Technology. 2022; 12(2): 34–41p.

[13] C. Nagarajan, G.Neelakrishnan, R. Janani, S.Maithili, G. Ramya "Investigation on Fault Analysis for Power Transformers Using Adaptive Differential Relay" Asian Journal of Electrical Science, Vol.11 No.1, pp: 1-8, 2022.

[14] G.Neelakrishnan, K.Anandhakumar, A.Prathap, S.Prakash "Performance Estimation of cascaded h-bridge MLI for HEV using SVPWM" Suraj Punj Journal for Multidisciplinary Research, 2021, Volume 11, Issue 4, pp:750-756

[15] G.Neelakrishnan, S.N.Pruthika, P.T.Shalini, S.Soniya, "Perfromance Investigation of T-Source Inverter fed with Solar ec.2007

[16] M Suganthi, N Ramesh, "Treatment of water using natural zeolite as membrane filter", Journal of Environmental Protection and Ecology, Volume 23, Issue 2, pp: 520-530,2022

[17] M Suganthi, N Ramesh, CT Sivakumar, K Vidhya, "Physiochemical Analysis of Ground Water used for Domestic needs in the Area of Perundurai in Erode District", International Research Journal of Multidisciplinary Technovation, pp: 630-635, 2019

[18] Characteristics in ABAC with group hierarchy: reachability analysis IEEE Trans. Reliable Secure Computer. 2022, 20, 841–858, Gupta, M., Sandhu, R., Mawla, T., & Benson, J.

[19] Barkadehi, M.H.; Nilashi, M.; Ibrahim, O.; Fardi, A.Z.; Samad, S. reviewed and categorized the literature on authentication systems in Telemat. Information. 35, 1491–1511 (2018).

[20] Blockchain identity authentication system- based IoT terminal connection service architecture 2020, 160, 411–422. Huang, J.C.; Shu, M.H.; Hsu, B.M.; Hu, C.M. Computer. Communication.

[21] In 2020, Zahid, G., Shafiq, A., Khalid, M., Hafizul, S., Mohammad, M.H., and Giancarlo, F. introduce an improved authentication technique for remote data access and sharing over cloud storage in cyber-physical-social systems. IEEE Access 8, 47144–47160.

[22] 9.Iris Technology: An Overview of Biometric Systems Based on Iris for Personalized Human Identification, Int. Granth aalayah J. Res. 2018, 6, 80–90; M.V.B. Reddy, V. Goutham