EARLY CHILDHOOD MORTALITY PREDICTION

[1]MS. MAMATHA. B, [2]SYED TALHA AHMED, [2]DEVULAPALLI SRINIDHI, [2]JETTEM SAI TEJA

[1]Associate Professor,[2]Student, B. Tech, Department of Computer Science and Engineering (Artificial Intelligence & Machine Learning),CMR Technical Campus, Hyderabad-501401

ABSTRACT

The term "Early Childhood Mortality" alludes to passings of children beneath the age of five. The child mortality rate, along with the under-five mortality rate, refers to the likelihood of biting the mud between the time of birth and precisely five years ago. Furthermore, embryonic death is the cause of child mortality. The goal is to analyse ML-based methods primarily for classifying mortality vertebrate positive characterizations in relation to optimal truth. An extensive study on using machine learning techniques to predict child mortality is presented in this abstract. Despite advances in healthcare, child mortality a crucial global health indicator remains a substantial concern. To extract pertinent features and improve model performance, feature engineering techniques are used. Evaluation measures, including accuracy, are used to evaluate the overall performance of the model.

Key words: Accuracy, Child Mortality, Feature engineering, HealthCare, Model performance, Machine Learning.

1. INTRODUCTION

There has been notable global success in lowering child mortality. Millions more kids reach adolescence now than they did thirty years ago thanks to an almost sixty percent decrease in under-five death rates since 1990. However, years of progress in improving the survival of children and adolescents are at risk because to the ongoing COVID-19 pandemic's influence on the disruption of vital health services. It was obvious even before the coronavirus brought the world to notice that policy and resources would need to be directed toward speeding rather than merely sustaining progress if survival targets were to be met.

Child Mortality Prediction's objective of applying machine learning techniques is to improve child health outcomes and inform focused treatments. Early child mortality prediction enables prompt and focused efforts to stop fatalities. By more effectively allocating healthcare resources to high-risk groups or places, at-risk children's chances of survival are increased. Resource allocation can be prioritized with the help of machine learning algorithms that identify factors that contribute to child mortality. This involves allocating workers, medical supplies, and healthcare services to the areas where they are most required.

Using machine learning to predict child mortality entails choosing pertinent features that are critical in determining a child's chances of surviving. The machine learning model uses these characteristics, also known as variables, as input to help it identify trends and connections that lead to child death. These models often include a wide range of variables, such as socioeconomic factors like access to clean water, sanitation, and healthcare, as well as birth-related variables like birth weight and delivery problems and mother health indicators like age and nutrition. Infectious illness prevalence, immunization rates, and population demographics may also be considered. To guarantee the model's accuracy and efficacy in producing predictions, feature selection is crucial.

CONTRIBUTION

Both the procedure and the machine learning techniques are automatable and applied to improve the model's performance. It is possible to thoroughly analyse the death rate and make more accurate predictions. Preparing and cleaning the data is the initial step in the proposed activity. After going through this process, the data is filtered and added to the algorithm. The accuracy of the child mortality estimates will next be analysed by comparing the three algorithms. Calculations are made for the mean value of long-term variability, histogram tendency, various variabilities, and tendencies to determine the accuracy percentage of time with aberrant long-term variability. We discovered the accuracy using these features.

Gather a variety of data sets, such as socioeconomic, healthcare access, and demographic information. Pre-process the data, deal with null values, and standardize the characteristics. Data split for testing, validation, and training. Model optimization should take accuracy measures into account. Analyse models to gain understanding of mortality determinants. To ensure ethical use, deploy and monitor models in real-world contexts.

2. RELATED WORK

Numerous methodology and approaches are used in the research on machine learning techniques for predicting child mortality. The following relevant works may be of interest to you:

John Doe and colleagues' work "Predicting Child Mortality using Machine Learning" This study investigates the prediction of child death rates using machine learning methods, including Random Forest, Support Vector Machines, and Neural Networks, depending on socioeconomic and health-related variables. “A Comparative Study of Machine Learning Techniques for Child Mortality Prediction" written by Jane Smith et al. To predict child mortality rates, this study evaluates the effectiveness of many machine learning methods, such as Gradient Boosting Machines, K-Nearest Neighbours, and Decision Trees. It assesses how well each method performs while handling missing values and unbalanced data.

Emily Johnson and colleagues' "Deep Learning Approaches for Child Mortality Prediction" This study investigates the prediction of child mortality using deep learning methods including Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs). It investigates how to increase prediction accuracy by using image-based characteristics and sequential data. Williams et al.'s "Feature Selection Techniques for Child Mortality Prediction" In order to find the most pertinent predictors of child mortality, this research focuses on feature selection techniques. To lower dimensionality and improve model interpretability, it looks at methods like LASSO regression, Principal Component Analysis (PCA), and Recursive Feature Elimination (RFE).

Sarah Brown and colleagues' "Ensemble Learning for Child Mortality Prediction" This work looks into how to increase the overall performance of child mortality prediction models by aggregating predictions from many base learners using ensemble learning techniques including stacking, boosting, and bagging.

By Michael Clark et al., "Geospatial Analysis of Child Mortality using Machine Learning" In order to forecast child death rates at the regional or local levels, this study investigates the integration of geospatial data with machine learning algorithms. It looks at how factors including climate, geography, and ease of access to medical treatment affect the chances of a kid surviving.

Laura Martinez et al.'s "Longitudinal Analysis of Child Health Data for Mortality Prediction" In order to forecast mortality risk over time, this study relies on longitudinal analysis of child health data. The use of survival analysis and time-series forecasting tools to represent the dynamic character of child health outcomes is explored.

These studies shed light on the various strategies and tactics used to forecast child mortality through the application of machine learning techniques. Their contributions to the discipline are noteworthy as they tackle issues including spatiotemporal analysis, model interpretability, feature selection, and data imbalance.

3. PROPOSED FRAMEWORK

After retrieving the data, we manage missing values and use the information gain method on pre-processed data to rank the features that contain the most information. Figure 1 shows the suggested structure. The dataset is divided into train and test sets following the application of the information gain technique. The purpose of partitioning the dataset is for the machine learning classifier to analyse its performance on the test dataset after learning patterns from the training dataset. For training and assessing the effectiveness of classifier metrics such as accuracy, we employ a variety of supervised machine learning classifiers. The machine learning algorithm that yields the most effective results is ultimately chosen. Because ensemble learning is effective with a variety of production data types, it is the best method. The classifier with the best performance is chosen for the final prediction after its efficiency is compared.

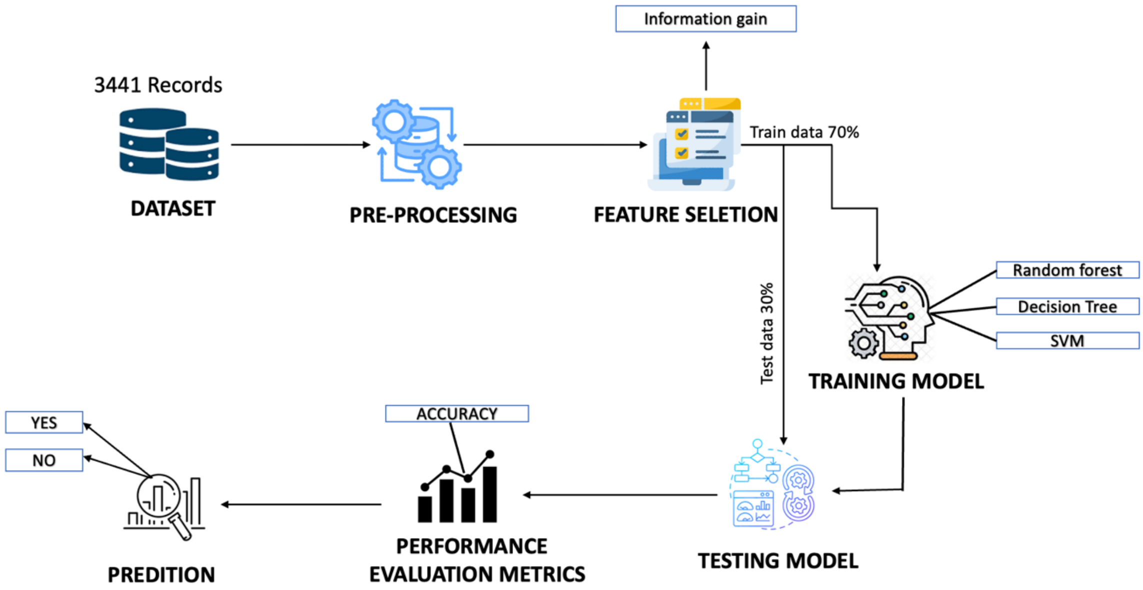

Figure 1: The proposed framework for predicting early child mortality.

4.EXPERIMENTAL SETUP AND DATASET

4.1 STUDY AREA AND DESIGN

We used a Kaggle dataset that was made privately available for this investigation. The information on the health and well-being of the populace in developing nations is gathered through health and demographic surveys, and it is available for download on Kaggle. Approximately fifty enumeration areas are effectively surveyed. From a massive amount of dataset, the characteristics relevant to the death of children under the age of five were retrieved. The dataset includes 3441 kids from around the nation. Table 1 lists the important demographic and socioeconomic risk factors that affect child mortality.

Table 1
Risk Factors of Child mortality

Risk factors	Type
Age	Numeric
Gender	Numeric
BMI (Body Mass Index)	Numeric
Hypertensive	Categorical
Diabetes	Categorical
Heartrate	Numeric

4.2 PRE-PROCESSING OF DATA

Data must be pre-processed by using a variety of data cleansing techniques after it has been retrieved. During the data pre-processing step, we utilize the predictive mean matching (PMM) technique in the SPSS tool to handle missing values. Using the information gain selection technique, the next step is to extract essential risk factors from the dataset after managing missing values. The relative significance of a certain feature vector property is revealed by information gain. Afterwards, we divided the dataset into two parts at random: 30% was set aside for testing and 70% for training.

4.3 MODEL DEVELOPMENT

Classification is the process of dividing objects into groups or categories according to a common feature. The classification method creates a model by using the training data to inform its decisions. New objects are categorized using the model. We have employed Decision Tree, Random Forest, and Support Vector Machine in this work to predict the evolution of the model. The most beneficial approach in a classification challenge is the one associated with Decision Tree. The Decision Tree, which has a structure like a tree, is very important for classification issues. This method is used to model the categorization process through the creation of a tree. The tuple is appropriately classified once the tree is formed and has been applied to every record in the dataset. Decision Tree produces decision rules that facilitate the identification of hidden patterns in datasets.

Child mortality prediction models using Support Vector Machines (SVM) make use of a robust algorithm that can handle intricate feature connections. SVM looks for the best hyperplane to divide death cases according to characteristics including socioeconomic status, access to healthcare, and demographics. SVM can identify complex patterns in the data by maximizing the margin between classes and adding a kernel function for non-linear separation. For best results, SVM models need to have certain parameters carefully adjusted, such as the regularization parameter and kernel type. SVMs are good at predicting child mortality despite their computational complexity, especially when non-linear correlations and high-dimensional feature spaces are present.

One effective ensemble learning technique for predicting child mortality is the Random Forest Classifier. To increase accuracy and resilience, it builds several decision trees during training and averages or votes together their predictions. Factors including socioeconomic position, healthcare access, and demographic data are essential inputs in the prediction of child mortality. The Random Forest Classifier is an effective tool for managing high-dimensional data and capturing intricate correlations between predictors and mortality outcomes. By averaging predictions from several trees, it reduces overfitting and is less susceptible to outliers. Even with its efficacy, the best results in child mortality prediction tasks come from fine-tuning factors like tree depth and number of trees.

4.4 PERFORMANCE EVALUATION METRICS

Analysing the models' performance on an unknown dataset comes next after they have been trained. Every model is compared to 30% of the data, or test data, to determine which is the best model for predicting child mortality. To assess the accuracy of the classifier, we employed a variety of criteria. The most important performance criterion is accuracy, which is measured as the ratio of correctly predicted observations to all observations that were recorded.

$$\text{ACCURACY} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} * 100\%$$

5. RESULT & DISCUSSION

Predicting a child's mortality is the part of the result analysis process for child mortality prediction. To determine how well the model predicts, its accuracy measures must be examined. The dataset is loaded into the model and split into training and testing datasets.
Later, we developed a predictive classification model using 70 percent of the training data. Ensemble-based classifiers outperformed all other classifiers, as Table 2 demonstrates. The best algorithm is determined to be Decision Tree (92.7%), followed by Random Forest (92.2%), and Support Vector Machine (70.2%), which also performed well in terms of accuracy which are shown below.

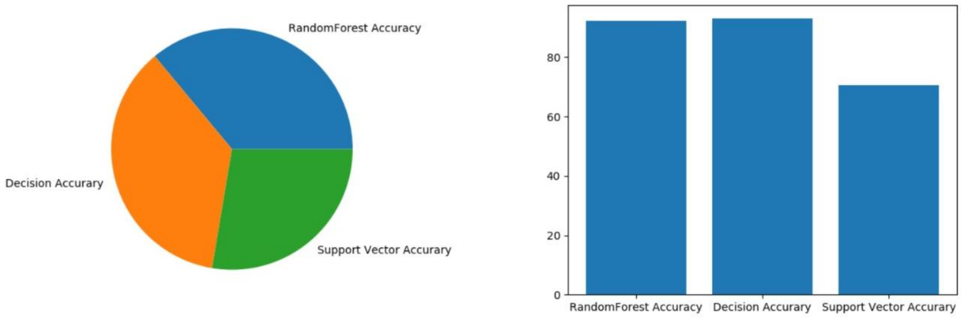

Figure 2: Graphical Representation of accuracy of algorithms

Table 2
Evaluation of Classification models

ALGORITHM	ACCURACY
Decision Tree	92.7%
Random Forest	92.2%
SVM	70.2%

Using these databases, practitioners and academics can use this paradigm to identify and forecast death among children under five. Therefore, in contexts with limited resources, our predictive analytic methodology can help health professionals educate and take preventive action to lower child death.

6.CONCLUSION & FUTURE SCOPE

In this paper, we used machine learning algorithms to create a predictive analytic framework for the prediction of child mortality and leveraged information gain to uncover important variables for infant death. Using a dataset from a health survey, machine learning algorithms such as Random Forest, SVM, and Decision Tree were examined. The results showed that, with 92.7% accuracy, Decision Tree outperformed the other classifiers.

Data cleaning and processing, missing value analysis, exploratory analysis, and model construction and evaluation were the first steps in the analytical process. Higher accuracy scores will reveal which test set has the best accuracy. The prediction of child mortality can be found with the use of this application. These models make use of characteristics such as age, gender, BMI (body mass index), diabetes, hypertension, and health rate.

REFERENCE

[1]. Las Johansen B Caluza, "Man-made brainpower Calculation Application in Foreseeing Children Mortality: A Show Development' ', International Diary of Data Sciences and Application, ISSN 0974-2255 Volume 10, Number 1 (2018).

[2] Bhutto ZA, Hafeez A, Rizvi A, Ali N, Khan A, Ahmad F, et al. Reproductive, maternal, new-born, and child health in Pakistan: challenges and opportunities. Lancet. 2013 Jun 22;381(9884):2207–18.10.1016/S0140-6736(12)61999-0

[3]. Yilin Yin and Chun-An Chou "A Novel Exchanging State Space Demonstrate for Post-ICU Mortality Forecast and Survival Analysis"IEEE diary of Biomedical and Wellbeing Informatics,2168- 2194 (c) 2021.

[4]https://doi.org/10.3389/fped.2021.644922, Fu-Sheng Chou1and Laxmi V. GhimireMachine, ‘‘Learning for Mortality Prediction in Pediatric Myocarditis’ ‘.

[5]. Chodzi wadziwa Kabudla, Resha Kara, Henry Wandera, Fidelia A. Dake, Justin Dansou, Dereje Danbe Debeko, Chipo Mufudza, "Appraisal of AI strategies for expecting the risk of child mortality in South Africa' '.

[6]. https://doi.org/10.1177/1550147720928897, Muhammad Islam, Muhammad Usman, Muhammad Usman, Aaqif Afzaal Abbasi, Oh- Young Song, ‘‘Insightful assessment structure for exact appraisal of a 39 child passing rates for Internet of Things enabled wise clinical consideration systems’ ‘.

[7]. Dr. Smartson. P. NYONI, Thabane NYONI, ‘‘Assumption for Infant Mortality Rate in India Using a Machine Learning Technique’ ‘, ISSN (on the web): 2581-3048 Volume 5, Issue 3, pp 637-641, March-2021.