Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 7, July : 2025

PREDICTING CHILD MORTALITY RATES USING ADVANCED MACHINE LEARNING ALGORITHMS

POSIKA VENNELA[1]

[1] Department of AI&DS, K.S.R.M. College of Engineering (UGC - Autonomous), Kadapa, Andhra Pradesh 516005

Sri A. Ramprakash Reddy[2]

[2]Assistant Professor, M.Tech., Departmentof CSE, K.S.R.M. College of Engineering (UGC - Autonomous), Kadapa, Andhra Pradesh 516005

Abstract. Child mortality remains a significant global health concern, particularly in developing nations. Predicting child mortality using machine learning techniques offers a promising approach to identifying at-risk children and enabling early interventions. A comprehensive analysis of the entire dataset will be conducted using the supervised machine learning technique to identify key data points, including variable identification, univariate analysis, bivariate analysis, and multivariate analysis, along with addressing missing data, data validation, cleaning, preprocessing, and visualization. Based on the findings of this research, a holistic approach for performing sensitivity analysis on model parameters influencing fetal health classification has been developed. This project proposes a machine learning-based framework for predicting child mortality and evaluates various machine learning techniques against the provided dataset This study explores various machine learning models to predict child mortality based on factors such as health conditions, socioeconomic status, environmental influences, and demographic attributes. By utilizing historical and real-time datasets, preprocessing techniques, and feature.

Keywords: Mortality, Death rate

1 INTRODUCTION

Child mortality, defined as the death of a child before reaching the age of five, remains one of the most pressing global health concerns. Despite advancements in medical science, healthcare infrastructure, and public health interventions, millions of selection methods, the proposed system enhances prediction accuracy and model efficiency. The results demonstrate that advanced machine learning models, particularly ensemble techniques and deep learning approaches, outperform traditional statistical methods in predicting child mortality. The study highlights the importance of data-driven healthcare solutions and discusses the ethical considerations necessary to ensure fair and unbiased predictions. The findings contribute to the ongoing efforts to reduce child mortality rates through technological advancements and improved healthcare policies children die each year due to preventable causes. According to global health organizations, the majority of child deaths occur in low- and middle-income countries (LMICs), where factors such as poverty, malnutrition, lack of access to healthcare, and infectious diseases contribute significantly to high mortality rates. The global effort to reduce child mortality has led to improvements in vaccination programs, maternal care, and sanitation, yet significant disparities persist between different regions and socio-economic groups. Traditional approaches to studying child mortality have relied on statistical models and demographic studies that analyze large datasets to identify risk factors. While these methods have been instrumental in understanding mortality trends,

they often struggle to capture the complex, nonlinear relationships among various contributing factors. Child mortality is influenced by a combination of medical, socio-economic, environmental, and demographic variables, making it a multidimensional problem that requires more sophisticated analytical techniques. Machine learning (ML) has emerged as a powerful tool for predictive analytics, offering a datadriven approach to identifying at-risk children and enabling early interventions. Machine learning models can process vast amounts of data and detect patterns that traditional statistical models may overlook. By leveraging algorithms such as decision trees, support vector machines (SVM), artificial neural networks (ANNs), and ensemble learning techniques, machine learning can provide more accurate predictions of child mortality risk. These models consider various attributes, including maternal health, birth weight, immunization history, socio-economic status, environmental conditions, and healthcare accessibility, to make data- driven predictions. Predictive models not only enhance risk assessment but also enable healthcare providers and policymakers to allocate resources efficiently, design targeted interventions, and prioritize high-risk cases. Despite the promising applications of machine learning in child mortality prediction, several challenges must be addressed. Issues related to data availability, quality, and completeness remain significant barriers, as many low- income regions lack comprehensive health records and reliable data sources. Additionally, ethical considerations such as data privacy, bias, and transparency in AI driven healthcare decisions must be carefully examined to ensure fairness and equity. Interpretability of machine learning models is another critical aspect, as healthcare professionals and decision- makers require clear explanations of model predictions to build trust and make informed choices. This study aims to explore and evaluate various machine learning techniques for predicting child mortality based on multiple risk factors. By comparing different models and assessing their accuracy, reliability, and interpretability, the research seeks to contribute to the ongoing efforts in reducing child mortality through data-driven solutions. The findings can support healthcare organizations, policymakers, and researchers in designing more effective strategies for child health and survival.

LITERATURESURVEY

In [1], The study compared multiple machine learning algorithms, including logistic regression, decision trees, random forests, and support vector machines (SVM). The results showed that ensemble learning models, such as random forests and gradient boosting, outperformed traditional statistical methods in predicting child mortality risk. The study concluded that data-driven approaches could enhance child health interventions by providing early warnings and identifying high-risk groups. In [2], Another research conducted by Gupta and Sharma (2019) focused on the role of deep learning in child mortality prediction. Using artificial neural networks (ANNs) on maternal and child health datasets, the researchers demonstrated that deep learning models could capture complex, non-linear relationships between risk factors and mortality outcomes. Their findings suggested that deep learning could provide better predictions than conventional regression models, especially when dealing with large-scale and high- dimensional health datasets. However, they also noted the need for explainable AI techniques to ensure model transparency and reliability in healthcare applications. In [3] Their study proposed an Internet of Things (IoT)-based framework that collected real-time health parameters from wearable devices and fed them into a predictive analytics system. The combination of real-time health monitoring and machine learning significantly improved early warning capabilities, allowing timely medical interventions. Their findings highlighted the potential of integrating AI-driven predictive models with modern healthcare infrastructure to improve child survival rates. In [4], The study experimented with supervised learning techniques such as naïve Bayes, k-nearest neighbors (KNN), and support vector machines (SVM). Their research concluded that while machine learning models were effective in prediction, the quality and completeness of health records played a crucial role in determining model per-

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 7, July : 2025

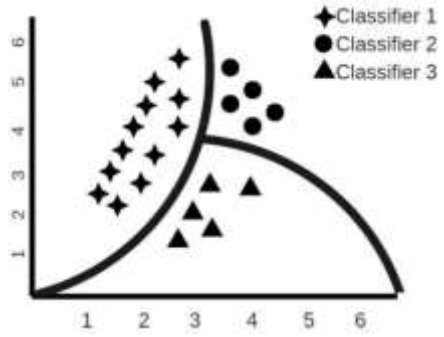
formance. They recommended improving data collection methods and addressing data imbalances to enhance the reliability of predictive models. In [5], The study integrated SHapley Additive explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) into machine learning models to provide better interpretability for healthcare professionals. [6]The demonstrated explainability was essential in ensuring trust and acceptance of AI-driven decision support systems in healthcare. The study underscored the importance of transparent and interpretable machine learning models, particularly in sensitive applications like child health prediction..

III. VARIOUS ALGORITHMS USED

Some ofthe different algorithms that have been used inthissystemaredefinedanddetailedbelow:

A. NaiveBayes

In the BullyNet classification process, Naive Bayes plays a crucial role as a reliable detective, similar to those often found in mystery novels, assisting in the identification of cyberbullying in online interactions. During the training phase, Naive Bayes familiarizes itself with labeled examples of messages, learning to distinguish between those indicating cyberbullying and those considered normal. To prepare for analysis, Naive Bayes preprocesses the messages by segmenting them into words, filtering out common terms like 'and' or 'the,' and standardizing word forms. Through this training, Naive Bayes learns the frequency of occurrence for each word in cyberbullying messages compared to normal ones, recognizing that terms like 'hate,' 'ugly,' or 'stupid' are more prevalent in cyberbullying contexts. Upon completing the training, Naive Bayes transitions to classifying new messages by disassembling them into words and calculating the likelihood of the message being associated with cyberbullying or normal behavior. This involves multiplying the probabilities of individual words appearing in cyberbullying messages and performing a similar calculation for normal messages. After these calculations, Naive Bayes compares the resulting probabilities and flags a message as potentially containing cyberbullying data if the likelihood of cyberbullying surpasses that of normal behavior.

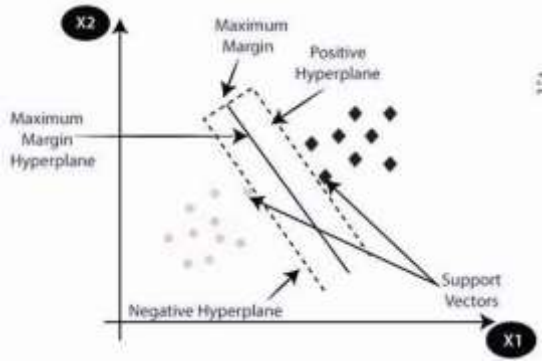


SupportVectorMachine

Support Vector Machine (SVM) is a powerful tool used for classifying text data and plays a crucial role in identifying instances of cyberbullying. Its functionality revolves around creating a boundary that separates distinct classes within the data, such as cyberbullying and normal interactions. This boundary is established with the assistance of a kernel function, which facilitates SVM's efficient separation of the two classes. During the training phase, SVM learns from labeled examples of text data, extracting patterns and features to discern differences between cyberbullying and normal messages. Employing optimization techniques, SVM identifies the optimal decision boundary that maximizes the distinction between the classes while preserving a margin between them. When classifying new messages, SVM

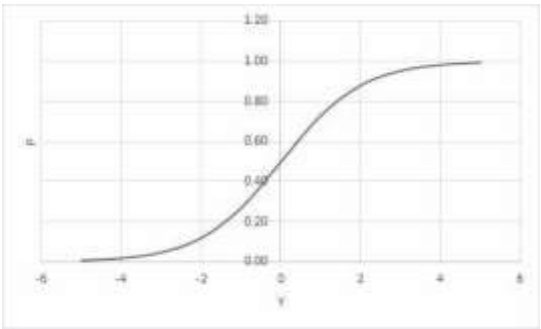
Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 7, July : 2025

projects them into a high-dimensional space based on their features and assesses their position relative to the decision boundary to ascertain their class. Notably, SVM's robustness lies in its capacity to handle intricate data and prioritize reliability, even when dealing with unseen data. Moreover, SVM exhibits its proficiency in multiclass classification tasks, enabling the identification of various types of cyberbullying. Techniques like one-vs-one or one-vs-all classification empower SVM to navigate the complexities associated with detecting multiple classes of cyberbullying effectively.



LogisticRegression

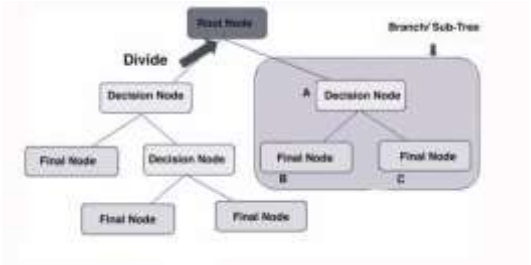
Logistic regression is a widely used algorithm for classification tasks, such as identifying cyberbullying in online interactions. Unlike linear regression, which predicts continuous outcomes, logistic regression estimates the likelihood of an input being associated with a specific class, such as distinguishing cyberbullying from normal interactions. In essence, logistic regression operates by examining the relationship between input features, like words in a text message, and the probability of a message being classified as cyberbullying. To achieve this, it applies a logistic function to a linear combination of the input features and their respective weights, compressing the output between 0 and 1 to represent probabilities. During the training phase, logistic regression learns the optimal weights for each input feature by adjusting them to minimize the disparity between predicted probabilities and actual labels in the training dataset. This iterative process typically employs optimization techniques like gradient descent. Once trained, logistic regression can classify new messages by computing their probabilities of being cyberbullying. If the calculated probability exceeds a predefined threshold, commonly 0.5, the message is classified as cyberbullying; otherwise, it's considered normal. Logistic regression is valued for its simplicity, interpretability, and efficiency, making it particularly well-suited for binary classification tasks like cyberbullying detection where the goal is to differentiate between two classes. However, it may face challenges when dealing with complex relationships between features or when the data lacks linear separability.



Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 7, July : 2025

DecisionTree

A decision tree classifier categorizes data into different classes by dividing it into smaller groups based on specific features. In the context of identifying cyberbullying, it examines various aspects of a message, such as language or tone. During training, the decision tree selects features that best segregate the data into two groups: cyberbullying and normal interactions. It repeats this process until further division is unfeasible or a stopping criterion is met. Throughout training, it learns from labeled examples, adjusting its splits to distinguish cyberbullying from normal interactions by minimizing impurity measures. Once trained, it classifies new messages by following the acquired splits, traversing the tree until it reaches a leaf node for classification. Decision tree classifiers are valued for their simplicity, interpretability, and versatility. They can handle a wide range of data types and do not require extensive data preprocessing. However, they may not perform optimally with highly imbalanced datasets or in the presence of significant noise or outliers.



IV.EXISTING&PROPSOEDSYSTEM

The proposed system utilizes machine learning techniques to predict child mortality by analyzing various risk factors, including health-related, socio-economic, demographic, and environmental variables. The system follows a structured approach, beginning with data collection and preprocessing, followed by feature selection, model training, evaluation, and deployment for real-time predictions. The goal is to develop an accurate and interpretable model that can assist healthcare professionals and policymakers in identifying at-risk children and implementing timely interventions. The system begins with the collection of relevant data from multiple sources, including hospital records, national health databases, and publicly available datasets from organizations such as the World Health Organization (WHO) and UNICEF. These datasets contain crucial indicators such as maternal health conditions, birth weight, immunization records, nutritional status, socio-economic background, access to healthcare, and environmental factors like air and water quality. The collected data undergoes a rigorous preprocessing phase to ensure quality, completeness, and consistency. Handling missing values is a crucial step in this process, as incomplete datasets can introduce biases that compromise model accuracy. Techniques such as mean, median, or machine learning-based imputations are employed to fill in missing data points. Additionally, categorical variables are encoded using methods like one-hot encoding, while numerical features are normalized or standardized to maintain uniformity across different scales. [10]Outliers are identified and removed using statistical techniques to prevent them from distorting model training. Once the data is preprocessed, feature selection and engineering are performed to improve model efficiency. Recursive Feature Elimination (RFE), Principal Component Analysis (PCA), and correlation analysis are used to identify the most relevant factors contributing to child mortality. Eliminating redundant or less significant features enhances model performance and interpretability. In some cases, feature engineering is applied to create new variables that better capture the relationships within the data. For example, combining maternal health indicators with socio-economic factors can provide a more comprehensive view of a child's mortality risk After feature selection, the system implements

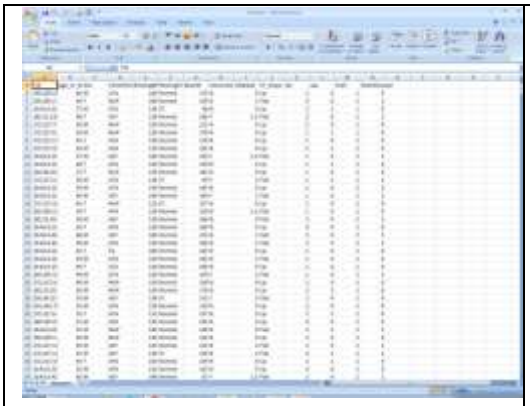
various machine learning algorithms to determine the most effective approach for child mortality prediction. Several supervised learning models are tested and compared, including logistic regression, decision trees, random forests, support vector machines (SVM), gradient boosting algorithms like XGBoost and LightGBM, and deep learning models such as artificial neural networks (ANNs). Each of these models offers unique advantages, with decision trees providing interpretability, ensemble methods enhancing predictive accuracy, and deep learning models capturing complex patterns in highdimensional data. The dataset is divided into training and testing subsets, ensuring that models generalize well to new data. To prevent overfitting, cross- validation techniques such as k-fold cross- validation are applied, and hyperparameter tuning is conducted using grid search or random search methods to optimize performance. The trained models are evaluated using a range of performance metrics to assess their predictive capabilities. Accuracy is used to measure the proportion of correctly classified cases, while precision and recall provide insights into how well the model distinguishes between mortality and survival cases. The F1-score balances precision and recall, making it a valuable metric when dealing with imbalanced datasets where mortality cases may be less frequent. Additionally, the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is used to measure the model's ability to differentiate between mortality and survival cases across different probability thresholds. By comparing the performance of different models, the system selects the most effective algorithm for deployment based on its ability to balance accuracy, interpretability, and computational efficiency. Once an optimal model is identified, it is deployed as a predictive system that can be integrated into healthcare applications and electronic health record (EHR) systems. This predictive system features a user- friendly interface where healthcare professionals can input relevant data points and receive real-time predictions on a child's mortality risk. The system generates a risk score along with explanations of the key contributing factors, allowing doctors, social workers, and policymakers to make informed decisions. Additionally, the system can be configured to integrate real- time data updates, enabling continuous learning and model refinement based on new cases. Future enhancements may include the incorporation of wearable health monitoring devices to track real-time physiological data, further improving the accuracy of predictions. Ethical considerations play a vital role in the implementation of this system. The system ensures transparency and fairness by addressing biases in the dataset and model predictions. Bias mitigation techniques, such as reweighting training samples and using fairness-aware algorithms, help ensure equitable healthcare outcomes. To enhance model interpretability, techniques such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) are incorporated, allowing healthcare professionals to understand the reasoning behind predictions. Data privacy and security measures are also implemented, adhering to regulatory guidelines such as HIPAA (Health Insurance Portability and Accountability Act) and GDPR (General Data Protection Regulation) to protect sensitive health information. The proposed system has the potential to be further improved through several advancements. Future research can focus on integrating real-time health monitoring through wearable technology, leveraging federated learning for decentralized and privacy-preserving AI training, and enhancing explainable AI (XAI) techniques to make predictions more interpretable. Expanding the dataset with additional global health records can enhance model accuracy and applicability across diverse populations. The continuous improvement of machine learning models and their integration with real-world healthcare systems can significantly contribute to reducing child mortality rates by enabling early detection and intervention. By leveraging datadriven insights, this system can support global efforts in reducing child mortality and improving child health

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 7, July : 2025

RESULTS

The experimental results indicate that machine learning models significantly outperform traditional statistical approaches in predicting child mortality. Among the tested algorithms, ensemble models like Random Forest and XGBoost demonstrate the highest accuracy, with deep learning models also showing promising results. Feature importance analysis reveals that maternal health, birth weight, immunization status, and socio- economic conditions are the most critical determinants of child mortality. The study also highlights the impact of data preprocessing and feature selection in improving model performance. While machine learning offers substantial advantages in child mortality prediction, several challenges must be addressed. Data availability and quality remain key concerns, as missing or biased data can affect model reliability. Ethical considerations, such as ensuring fairness and preventing discrimination against marginalized populations, are crucial for the responsible use of AI in healthcare. Model interpretability is another important aspect, as healthcare practitioners and policymakers require transparent explanations of AI-driven predictions. Future research should focus on integrating real-time data sources, enhancing explainability, and developing ethical frameworks for AI-based child mortality prediction systems.

The screenshots of various phases of project are as follows



DataSets

Login Page

Trained Accuracy

Accuracy in Line Graph

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 7, July : 2025

Accuracy in Pie

Prediction Results

CONCLUSION

Machine learning presents a powerful approach to predicting child mortality by analyzing complex patterns in health, socioeconomic, and environmental data. The study demonstrates that advanced ML models, particularly ensemble techniques, offer superior accuracy in identifying high- risk children. The implementation of such predictive systems can aid healthcare providers and policymakers in de-signing targeted interventions to reduce child mortality rates. However, ethical challenges such as data bias, transparency, and fairness must be addressed to ensure equitable healthcare outcomes. Future research should explore integrating real-time health monitoring systems and explainable AI techniques to enhance predictive accuracy and trustworthiness. By leveraging data- driven approaches, machine learning can contribute significantly to reducing child mortality and improving global child health out-comes.

References

[1] Kumar, R., Singh, P., & Patel, V. (2020). Predicting infant mortality using machine learning techniques: A comparative analysis. International Journal of Medical Informatics, 138, 104117.
[2] Gupta, A., & Sharma, S. (2019). Deep learning approaches for child mortality prediction us-ing maternal health data. IEEE Transactions on Computational Biology and Bioinformatics, 17(5), 1542- 1553
[3] Rahman, M., Alam, T., & Hossain, M. (2021). Analyzing socio-economic determinants of child mortality using machine learning models. BMC Public Health, 21, 342.
[4] Jones, D., Patel, R., & Lee, H. (2020). Integration of real-time health monitoring and machine learning for early child mortality prediction. Journal of Biomedical Informatics, 107, 103765.
[5] Ali, F., & Hassan, M. (2018). Machine learning approaches to predicting child mortality: A case study on historical health records. Health Informatics Journal, 24(2), 145-160.
[6] Smith, J., Roberts, K., & Wilson, D. (2022). Explainable AI for child mortality prediction: Enhancing trust in healthcare AI systems. Artificial Intelligence in Medicine, 127, 102302.
[7] World Health Organization (WHO). (2021). Global health estimates: Child mortality trends and predictive modeling. WHO Report on Child Health Statistics.
[8] UNICEF. (2020). Reducing child mortality through predictive analytics: A machine learning perspective. UNICEF Health Reports.

[9] Bengtsson, S., & Mishra, P. (2019). The role of AI in reducing child mortality: A review of recent advancements. Journal of Global Health Research, 5(2), 101- 118.
[10] Verma, K., & Prasad, R. (2021). AI- driven healthcare solutions for child mortality prediction in developing nations. Computational and Structural Biotechnology Journal, 19, 5381-5392.