Title of the Thesis

Detecting Child Abuse Using Machine Learning Data-Driven Approaches to Enhance Early Intervention

Submitted By

Monisha Zaman Upama
ID: 201-16-524
Department of Computing & Information System
Daffodil International University

Supervised By

Mr. Israfil
Lecturer
Department of Computing & Information System
Daffodil International University

Department of Computing and Information System
Daffodil International University
Dhaka, Bangladesh

Submission Date: 12/01/2025

APPROVAL

This Thesis titled "**Detecting Child Abuse Using Machine Learning Data-Driven Approaches to Enhance Early Intervention**", Submitted by **Monisha Zaman Upama**, ID No: **201-16-524** to the Department of Computing and Information Systems, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computing & Information Systems and approved as to its style and contents. The presentation has been held on 12-01-2025.

BOARD OF EXAMINERS

Md Sarwar Hossain Mollah — **Chairman**
Associate Professor and Head
Department of Computing & Information Systems
Faculty of Science & Information Technology
Daffodil International University

Md. Nasimul Kader — **Internal Examiner**
Assistant Professor
Department of Computing & Information Systems
Faculty of Science & Information Technology
Daffodil International University

Md. Mehedi Hassan — **Internal Examiner**
Lecturer (Senior Scale)
Department of Computing & Information Systems
Faculty of Science & Information Technology
Daffodil International University

Dr. Muhammad Shahin Uddin — **External Examiner**
Professor
Department of ICT
Mawlana Bhashani Science and Technology University

Declaration

I, Monisha Zaman Upama, titled **"Detecting Child Abuse Using Machine Learning Data-Driven Approaches to Enhance Early Intervention"**, hereby declare that the work in this documentation titled. I have done this thesis under the supervision of Mr. Israfil. Lecturer, Department of Computing and Information System (CIS) of Daffodil International University. I'm also declaring that this thesis or any part of there has never been submitted anywhere else for the award of any educational degree like B.Sc, M.Sc, Diploma or other qualifications.

Supervised by

Mr. Israfil
Lecturer
Department of Computing & Information System
Daffodil International University

Submitted by

Upama

Monisha Zaman Upama
Department of Computing & Information System
Daffodil International University

Acknowledgements

It has been a privilege to conduct research and complete this thesis, titled " Detecting Child Abuse Using Machine Learning: Data-Driven Approaches to Enhance Early Intervention." I am deeply grateful to Allah for His grace and blessings throughout my academic journey.

I would like to express my heartfelt gratitude to my supervisor, **Mr. Israfil**, Lecturer in the CIS Department at Daffodil International University, for his invaluable guidance, support, and encouragement. His insightful feedback, dedication, and motivation have been pivotal in the successful completion of this thesis. I am sincerely thankful for his unwavering belief in my potential and for providing me with the right direction to achieve my goals.

Additionally, I extend my deep respect and gratitude to **Mr. Sarwar Hossain Mollah**, the head of my department, for his initial guidance and continuous support throughout my research. His encouragement and understanding have been a source of inspiration.

Finally, I wish to acknowledge all the faculty members of my department for their constant support, patience, and invaluable teachings, which have greatly contributed to the successful completion of this thesis.

Dedication

This paper is lovingly dedicated to my parents, **Rabeya Zaman** and **Muniruzzaman Munir**, whose unwavering love and support have been the foundation of all my accomplishments. My mother's endless encouragement and my father's wise guidance have been constant sources of inspiration throughout this journey. I am deeply grateful for their sacrifices and their dedication to my education, which have fueled my passion for learning and discovery.

A special acknowledgment goes to my father for his invaluable assistance in collecting data for this thesis. His efforts and belief in my abilities have been instrumental in bringing this work to fruition. This paper is a reflection of the values they have instilled in me, and I am forever grateful for their trust and love.

Thank you for being my greatest source of strength.

Abstract

Child abuse is a critical issue in Bangladesh and worldwide, with devastating consequences for victims' physical and mental well-being. The need for timely detection and intervention is vital to mitigate its impact. This research aims to develop a machine learning-based predictive model to identify instances of child abuse and classify them into four categories: physical, emotional, sexual, and neglect. Using a dataset with 1,010 rows and 22 attributes, we applied various machine learning algorithms, including Logistic Regression, Random Forest Classifier, Naive Bayes, and Support Vector Machine. SMOTE (Synthetic Minority Oversampling Technique) was employed to address class imbalance, significantly improving model performance.

The Random Forest Classifier demonstrated the highest accuracy (99%) with precision, recall, and F1-scores of 0.98, 0.99, and 0.98, respectively, after applying SMOTE. These results validate the effectiveness of our approach in predicting child abuse cases with high reliability. This model not only serves as a valuable tool for early intervention but also offers insights to policymakers and social workers to prioritize cases and allocate resources efficiently. Our work contributes to the growing body of research on technological solutions for child welfare, especially within the socio-cultural context of Bangladesh.

Preface

This bachelor thesis, **“Detecting Child Abuse Using Machine Learning: Data-Driven Approaches to Enhance Early Intervention”**. has been prepared as part of my undergraduate degree program. The study aims to contribute to the growing field of machine learning by addressing a critical social issue: the prediction and detection of child abuse. Through a detailed analysis of data and the application of various machine learning models, this research seeks to provide an efficient solution for early intervention and prevention of child abuse.

The paper is structured into 6 chapters:

Chapter 1: Introduction
This chapter introduces the research topic and outlines the significance of applying machine learning techniques to predict child abuse. It includes the objectives of the study and discusses the motivation behind undertaking this research.

Chapter 2: Literature Review
The second chapter provides an overview of existing research related to child abuse prediction, highlighting gaps in the current methodologies. It also discusses the challenges encountered in implementing machine learning solutions for this domain.

Chapter 3: Methodology
In this chapter, the methodology for the research is presented. It describes the data collection process, the dataset used, and the data preprocessing steps such as handling missing data and categorical encoding. It also details the exploratory data analysis and introduces the four machine learning techniques—Logistic Regression, Random Forest, Naive Bayes, and Support Vector Machine—employed in this study.

Chapter 4: Results & Discussions
This chapter focuses on the performance evaluation of the models using metrics such as accuracy, precision, recall, and F1 score. It includes a comparative analysis of the results before and after applying the SMOTE technique to handle class imbalance. The chapter concludes by discussing the implications of the results and identifying the best-performing model.

Chapter 5: Conclusion
The fifth chapter summarizes the key findings of the research and discusses the contributions of the study. It also provides recommendations for future work and highlights the potential of using machine learning for combating child abuse.

Chapter 6: References
The final chapter lists the scholarly articles, books, and other sources that were consulted during the research process, ensuring the credibility and reliability of the study.

Contents

List of Figures

List of Tables

List of Abbreviations

CA – Child Abuse

ML – Machine learning

SMOTE – Synthetic Minority Oversampling Technique

NB – Naïve Bayes

SVM – Support Vector Machine

RF – Random Forest

LR- Logistic Regression

TPR– True Positive Rate

Chapter 1

Introduction

1.1Introduction

Child abuse is a grave issue affecting millions of children globally, with devastating short- and long-term consequences for their physical, emotional, and psychological well-being. It encompasses a range of harmful actions, including physical, emotional, and sexual abuse, as well as neglect. Globally, studies reveal that a significant percentage of children experience abuse in some form before reaching adulthood, leaving them vulnerable to lifelong challenges such as mental health disorders, impaired social functioning, and even chronic health problems. In Bangladesh, the situation is particularly concerning, as cultural stigma, social taboos, and lack of awareness often prevent victims and their families from reporting such incidents. Children may face abuse within their own homes, from family members, or outside, from peers, teachers, or strangers. Despite legal frameworks in place to address child abuse, enforcement and reporting mechanisms remain weak, leaving many cases unaddressed. This highlights the urgent need for targeted interventions, increased public awareness, and stronger implementation of child protection policies to safeguard children and ensure their healthy development. Globally, nearly half of all children between the ages of 2 and 17 are subjected to some form of violence annually. Approximately one-third of students aged 11 to 15 have experienced bullying from peers within the last month. Furthermore, an estimated 120 million girls have endured some form of coerced sexual contact before turning 20 [1].

While many related works have focused on general crime prediction or isolated types of abuse using machine learning models, my study adopts a more specialized and comprehensive approach by developing a predictive model aimed explicitly at identifying various forms of child abuse. This model is tailored to detect and categorize four major types of abuse: physical, emotional, sexual, and neglect, using a rich dataset containing 15 columns and 1010 rows. Key features in the dataset include demographic details such as age, gender, and socioeconomic status, as well as specific questions assessing the child's perceived safety at home and school, experiences of physical harm, emotional humiliation, and lack of basic needs or medical care.

To improve the accuracy of predictions and address the significant class imbalance in the dataset, I employed the SMOTE technique, which balanced the data distribution for the binary 'Abuse Yes/Not' column. Unlike broader crime detection models, this targeted approach ensures the reliability of predictions across different abuse subtypes. Additionally, my model evaluates performance using multiple machine learning algorithms—Logistic Regression, Random Forest Classifier, Naive Bayes, and Support Vector Machine—and assesses them through metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. By incorporating these features, the model not only identifies abuse but also aids in understanding the specific patterns and dynamics of child abuse within the socio-cultural context of Bangladesh, making it an essential tool for early intervention and prevention efforts.

1.2 Research Objective

The objectives of this study are-

- How can we predict child abuse cases in Bangladesh using machine learning techniques?
- What are the key factors contributing to child abuse in different environments such as home, school, and community?
- Which machine learning algorithm provides the most accurate and reliable results for predicting child abuse?
- To create a predictive model for early identification of child abuse, enabling timely interventions and improved child protection.

1.3 Motivation

Child abuse is a pervasive issue in Bangladesh, with children facing physical, emotional, sexual abuse, and neglect in alarming numbers. Societal stigma, economic hardship, and a lack of awareness prevent many cases from being reported or addressed. Bangladesh's legal and social systems often struggle to provide timely interventions due to resource limitations and inadequate detection mechanisms. This dire situation motivated me to develop a machine learning model tailored to Bangladesh's context, capable of identifying multiple types of child abuse with high accuracy. Using a dataset representative of local socio-economic and cultural realities, my model applies SMOTE to address data imbalance, ensuring more reliable predictions. By improving detection and enhancing early intervention, this model has the potential to support child protection services in Bangladesh. The goal is to provide a scalable, evidence-based tool to combat child abuse effectively and foster a safer future for vulnerable children in the country.

Chapter 2

Literature Review

2.1 Introduction

This chapter presents a detailed review of existing literature on child abuse detection using artificial intelligence (AI) and machine learning (ML). The chapter explores various studies that leverage predictive models, natural language processing (NLP), and classification algorithms to analyze child abuse data. It provides an overview of the methodologies employed, including strategies to address data imbalance and improve model performance. The review highlights the effectiveness of different ML algorithms, such as Random Forest, Support Vector Machines, and Neural Networks, in identifying patterns of abuse. Moreover, the chapter discusses ethical challenges, limitations in current systems, and the importance of accurate predictions in preventing further harm to victims. This chapter fills in the gaps in research by putting these results together. It also lays the groundwork for the proposed study, whose goal is to create a better model for predicting child abuse and the factors that are linked to it.

2.2 Related Work

In recent years, machine learning techniques have gained prominence in predicting and addressing societal issues, including the detection of child abuse. Machine learning models have been investigated in a number of studies for their ability to predict and identify incidents of child abuse and neglect.

The study focuses on how the COVID-19 epidemic has exacerbated domestic violence, especially in Bangladesh, where family violence against women and children has sharply escalated. This work analyzed family violence data collected from 511 households over a ten-day period using machine learning models to forecast domestic violence. The writers used three algorithms. The performance of the model was evaluated with reference to accuracy, precision, recall, F-score, ROC analysis. The Random Forest model exceeded Naive Bayes (62%), Logistic Regression (69%), and 77% accuracy level. Important predictors of domestic violence during the epidemic turned out to be family income level and educational level of family members. These findings show how effectively machine learning can detect family violence and emphasize the need of socioeconomic factors in these projections. [3]

The study comprises two phases. In the initial phase, a multi-label classification achieved over 97% accuracy on a Wikipedia dataset for categorizing text based on toxicity types. The second phase utilized this outcome to analyze the PAN12 dataset, focusing on detecting dubious messages and identifying predators in chat records. The study achieved over 92% accuracy in detecting grooming-related dialogues using models such as Multinomial Naive Bayes, AdaBoost, Linear SVC, and K-Nearest Neighbors. These results demonstrate the efficacy of machine learning in tackling the significant issue of online grooming children. [4]

This study uses text mining and machine learning to detect and predict child abuse in a Dutch public health institution. These organizations fight several types of child abuse, and most of their data comes from consultation notes. The study examines whether these records can identify abuse through patterns. This was achieved by training machine learning models with data labeled by over 500 Dutch municipality child experts. SVM Classifier performed best among Random Forest (RF) and Support Vector Machine (SVM) classifiers. Using a linear kernel (C=1) and 1000 tf-idf weighted features, the SVM achieved an Area Under the Curve (AUC) score of 0.906 with 84.3% accuracy and 82.5% recall. These findings show how well machine learning detects child abuse from unstructured medical data. [5]

This study investigates the utilization of text mining and machine learning to detect instances of domestic violence in child welfare investigation summaries, with the aim of filling the information gap regarding families' service requirements. The purpose of labeling 1,402 summaries was to ascertain if domestic violence services were actively needed. Afterwards, these tagged documents were used to train and evaluate machine learning and text mining models. The model's performance was assessed using three accuracy metrics, presuming that the labels provided by EHRs were accurate. Because its Fleiss kappa reliability score was more than 0.80, the best-performing ML model was able to assist human reviewers. Not only that, it outperformed human coders with an accuracy rate of 90% or higher. [6]

This study presents a machine learning framework designed for detecting Child Sexual Abuse Media (CSAM) by analyzing file metadata, thus addressing the legal and ethical challenges linked to the utilization of actual CSAM images. The model is trained on over one million file paths obtained from criminal investigations. The evaluation guidelines address challenges associated with adversarial data modifications and changes in data distribution due to limited access to training data. The framework exhibits significant performance, attaining accuracy rates of 97% and maintaining stability against adversarial attacks. The model demonstrates a low false positive rate of 0.002 when assessed using publicly available file paths from common crawl data, highlighting its effectiveness across diverse data distributions. [7]

This study developed AI algorithms using unstructured text from electronic medical records (EMRs) to identify potential victims of abuse. Data from 867 patients, 55% abuse-positive, were processed with three Natural Language Processing (NLP) algorithms: Bag of Words (BOW), Word Embeddings, and Rules-Based (RB). The best model, using BOW, achieved an accuracy of 90±2% and an AUC of 0.93±0.02, with a false positive rate of just 8%. RB models performed at 77±4% accuracy and an AUC of 0.81±0.05. [8]

Although state child welfare agencies handle large amounts of data, usability and quality of the data create problems. Using text mining and machine learning, a study examined 2,956 child abuse and neglect investigation summaries sent by a state agency in order to address this. Based on their presence of substance-related issues, expert human reviewers labeled these summaries. Comparing machine learning models to expert assessments, they showed great accuracy—more than 90%. Moreover, the Fleiss kappa value was higher than 0.80, suggesting rather strong model and human reviewer agreement. These findings imply that human knowledge in spotting important case details can be sufficiently complemented by machine learning methods. Simplifying data analysis helps such methods to be promising for enhancing decision-making and service delivery in child welfare systems. [9]

The purpose of this research is to identify instances of children's physical abuse by employing deep learning models that have been trained on clinical, laboratory, and imaging data. There were 1,312 patients, and 56.2% of them were victims of abuse. The initial model demonstrated an accuracy of 86.3%, sensitivity of 87.2%, specificity of 85.1%, an F1 score of 0.86, and an ROC AUC of 0.86. In comparison, the NLP-based model showed superior performance, achieving 93.4% accuracy, 92.5% sensitivity, 94.6% specificity, an F1 score of 0.93, and an ROC AUC of 0.94. Although individual features exhibited weak correlations with abuse, the findings underscore the effectiveness of deep learning, particularly NLP techniques, in facilitating precise and timely detection of child abuse. [10]

This study was to investigate the applicability of machine learning models for the purpose of categorizing the intention of injuries in children and adolescents, with a specific focus on addressing the possibility of misclassifying intentional injuries. An assortment of models, including Naive Bayes, Decision Tree, Random Forest, Adaboost, and Deep Neural Networks, were utilized in order to accomplish this objective. The DNN and Adaboost models both demonstrated superior performance, with the DNN model achieving ROC AUC scores of 92%, 83%, and 93% across a variety of injury types, and the Adaboost model scoring 91%, 84%, and 91%, respectively. These findings provide further evidence that DNN and Adaboost are effective in distinguishing injury intentions in a reliable manner. [11]

This study introduces a centralized system to support victims of violence, particularly women and children, by leveraging machine learning (ML) for crime forecasting and victim assistance. Unlike prior solutions, such as mobile applications or wearable devices, this system allows victims to file complaints anonymously or with identity verification and facilitates immediate support, including medical aid, counseling, and access to education and employment. The suggested system also includes an ML-based crime prediction module, which forecasts events to help lower crime rates. Future improvements might call for cloud databases for better scalability and federated learning for real-time model updates. With an R-squared value of 0.99, XGBoost (XB) showed better performance among the tested models, so it is greatly effective for producing crime forecasts. This platform not only helps law enforcement and other agencies deliver justice and guarantee society's reintegration by giving victims necessary knowledge about their rights, so it also assists in the prediction of violence. [12]

This study proposes a centralized system to support victims of violence, focusing on women and children, using machine learning for crime forecasting and victim aid. Victims can file complaints anonymously or with identity verification, receiving support like medical care, counseling, and education. The system’s ML-based crime prediction module helps forecast incidents, aiming to reduce crime rates. Future upgrades may include federated learning for real-time updates and cloud databases for scalability. XGBoost ($R^2 = 0.99$) outperformed other models, proving effective in crime forecasting and aiding law enforcement in delivering justice and supporting victim reintegration. [13]

The suggested system also includes an ML-based crime prediction module, which forecasts events to help lower crime rates. Future improvements might call for cloud databases for better scalability and federated learning for real-time model updates. With an R-squared value of 0.99, XGBoost (XB) showed better performance among the tested models, so it is greatly effective for producing crime forecasts. This platform not only helps law enforcement and other agencies deliver justice and guarantee society's reintegration by giving victims necessary knowledge about their rights, so it also assists in the prediction of violence. [14]

This study highlights the frequent unethical reporting of child abuse cases, including victim identity disclosure, which exacerbates stigma, hinders recovery, and deters reporting. Limited research in Bangladesh has addressed these violations. The study calls for policy reforms and future research to explore journalistic practices and the psychological impact on victims, promoting ethical reporting and improved victim support. [15]

Decision support systems (DSS) have emerged as a promising technology to assist decision-making in child and family social work by leveraging algorithms trained on administrative data. The research conducted in countries such as the USA, the Netherlands, Australia, and New Zealand highlights both their potential and the challenges they face, including ethical considerations, data biases, and implementation barriers. These studies provide valuable insights into improving DSS models and integrating them effectively into social work practices, making them relevant for child abuse prediction and intervention. [2]

This study identified the Random Forest Classifier as the most effective model for predicting physical violence against women. It achieved better performance metrics compared to other algorithms, with a precision of 51.0% and a recall of 40.0%. After calibration, the model's performance improved significantly, increasing the precision to 76.0% and recall to 31.0%. This adjustment enhanced the detection of physical violence cases, demonstrating the potential of machine learning in accurately identifying instances of abuse and aiding in targeted interventions. [16]

This study compares conventional protocol-based risk assessments with machine learning (ML) approaches and finds ML models significantly more accurate. A simple Bayes classifier outperformed conventional methods, while ML models leveraging two-year criminal histories provided even better predictions. Adding protocol-based features showed minimal improvement, highlighting the potential of using criminal history data for prioritizing service calls. The study recommends developing tools to improve true positive identification following ML-based initial screenings.[17]

This study addresses the detection of physical abuse through a deep learning approach, leveraging 3D convolutional neural networks (CNN) for human action recognition and pose estimation in real-time. By analyzing spatial-temporal features from video data, the proposed system achieves an accuracy of 89.42% and a precision of 85.82%, highlighting its effectiveness in detecting abuse and enabling timely reporting. [18]

This study explored the prediction of child abuse using machine learning, applying a Borderline-SMOTE-enabled Stacking Classifier (BS-SC) model to a dataset of child abuse incidents. The proposed model achieved high performance with 93% accuracy, precision, recall, and F1-Score, and an AUC value of 0.989, demonstrating its effectiveness in predicting various subtypes of child abuse and supporting real-time prevention efforts. [19]

The study shows that ensemble methods, such as stacking, are effective in predicting child abuse and stresses the significance of addressing data imbalance. It is clear from these results that machine learning models have great promise for helping with the early diagnosis of child abuse, which in turn could lead to more effective and faster preventative measures. The study utilized Borderline-SMOTE, a synthetic data balancing method, leading to enhanced performance for various classifiers, including XGBoost, Random Forest, and Decision Tree. Additionally, the researchers introduced an innovative Stacking Classifier model (BS-SC Model) that integrates several base classifiers (AdaBoost, XGBoost, Decision Tree, KNN, and Extra Trees) at Level 0, with Random Forest functioning as the meta-classifier. This model exhibited remarkable performance, attaining 93% in accuracy, precision, recall, F1-score, and an AUC of 0.983, surpassing other models in predicting child maltreatment. [20]

2.4 Challenge of this work

The primary challenge of this work lies in addressing the sensitive and complex nature of child abuse data, including the significant class imbalance, as cases of abuse are often underreported. Additionally, ensuring the accuracy and fairness of the prediction model while handling diverse socio-economic and cultural factors in Bangladesh adds complexity. Balancing data privacy, ethical concerns, and technical limitations, such as selecting the most effective machine learning algorithms, further complicates the process.

Chapter 3

Methodology

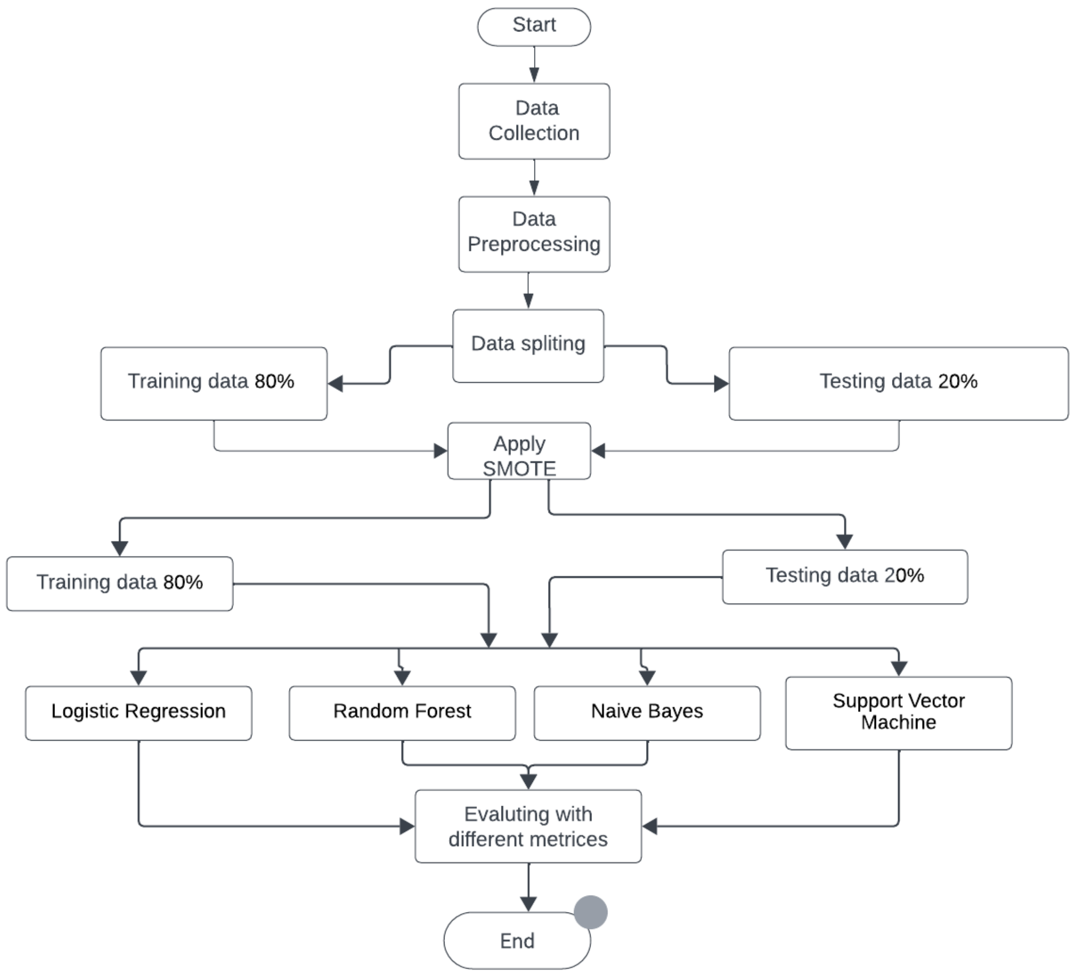

Figure 3.1: *Working Flow Diagram*

3.1 Data Collection

Child abuse remains a significant concern in Bangladesh, affecting the well-being and development of countless children. To better understand and predict instances of child abuse, we collected data from three different schools and one college through survey questionnaires.

3.2 Data Description

We polled "Detecting Child Abuse Using Machine Learning: A Data-Driven Approaches to Enhance Early Intervention" to compile data for this study. Machine learning-based models then helped to examine the gathered data. The data-collecting process, data preparation techniques, and machine learning algorithms applied for prediction and analysis are given in this part in their whole. This paper sought to grasp several kinds of abuse, including neglect, emotional, psychological, and physical ones as well as sexual ones. Designed using Google Docs, a comprehensive questionnaire was sent to students from three different colleges and one university in Bangladesh. Over fourteen days, the poll was carried out and we obtained 1,010 responses overall.

Key elements in the dataset included age, gender, socioeconomic level, opinions of safety at home and school, and experiences connected to several kinds of abuse. **Table 1** contains exact descriptions of every variable. We used machine learning models with this data to classify and project the various forms of child abuse. Python plus tools like Scikit-learn and Pandas were used for data processing and analysis.

Table 1. List of Survey Questions Addressing Child Abuse

Item	**Corresponding Definition**
Age	The participant's age.
Gender	The participant's gender.
Socioeconomic Status	The participant's family income level (Low, Middle, High).
Safety at School	How safe the participant feels at school or other places they spend time.
Safety at Home	How safe the participant feels at home.
Physical Abuse	Whether the participant experienced physical harm, such as hitting, slapping, or pushing from an adult.
Witnessing Abuse	Whether the participant has witnessed another child being physically abused by an adult.
Fear of Physical Aggression	Whether the participant has felt afraid of an adult due to their physical aggression.
Sexual Abuse	Whether the participant has experienced inappropriate physical touch or sexual harassment.
Emotional Abuse	Whether the participant has been humiliated, verbally abused, or subjected to constant negative comparisons.
Neglect	Whether the participant has experienced neglect, such as lack of supervision or unmet basic needs.
Trust	Whether the participant has someone they trust to talk about sensitive issues.
Humiliation	Whether the participant has experienced being humiliated by others.
Parental Expectations	Whether the participant has faced unrealistic expectations or constant comparisons from parents/teachers

The study included a series of targeted questions designed to explore various aspects of child abuse. These questions aimed to uncover experiences of physical, emotional, and sexual abuse, as well as neglect. The survey also included inquiries about the respondents' demographic information and perceptions of safety in different environments.

In **Table 2** the Demographic and Behavioral Characteristics of the Study Participants are given. The study surveyed 1,010 respondents to explore various aspects of child abuse and safety. Gender distribution revealed that 56.9% of participants were girls, while 43.1% were boys. A majority of the respondents (65.1%) belonged to middle-income families, followed by 29.3% from low-income and 5.5% from high-income households.

Concerning physical harm, 43.3% of participants reported experiencing abuse such as hitting, slapping, or pushing by an adult or older individual. Additionally, 37.1% indicated that they had been inappropriately touched, while 25.9% reported experiencing sexual harassment or exploitation online, including on platforms like Facebook, WhatsApp, and Instagram.

Neglect was also a significant issue, with 27.3% of respondents reporting being left alone without adequate supervision, and 20.2% stating they had experienced a lack of basic needs such as food, clothing, or shelter. Furthermore, 18.8% of participants noted that they had been deprived of proper medical care or essential medications when needed.

The perception of safety at home and school highlighted contrasting experiences. While 70.7% of respondents felt very safe at home, only 31.8% felt very safe in school or other places. Notably, 33.0% reported feeling somewhat safe in school, but 11.3% felt somewhat unsafe, and 5.5% felt very unsafe. Emotional abuse was prevalent among the participants, with 48.8% admitting they had felt humiliated by others, 36.3% expressing feelings of worthlessness or being unwanted, and 47.5% experiencing constant comparisons or unrealistic expectations from parents and teachers.

These findings provide a comprehensive overview of the respondents' experiences with abuse, neglect, and their perceptions of safety. The data underscores the multifaceted challenges faced by children and emphasizes the importance of addressing these issues through targeted interventions and support systems.

Table 2: Demographic and Behavioral Characteristics of the Study Participants

Category	**Variable**	**n**	**n(%)**
Gender	Girl	626	56.9%
Gender	Boy	475	43.1%
Socioeconomic Status	Middle Income	717	65.1%
Socioeconomic Status	High Income	61	5.5%
Socioeconomic Status	Low Income	323	29.3%
Experienced Physical Harm	Yes	477	43.3%
Experienced Physical Harm	No	624	56.7%
Felt Uncomfortable or Scared Due to Touch	Yes	409	37.1%
Felt Uncomfortable or Scared Due to Touch	No	692	62.9%
Online Sexual Harassment	Yes	285	25.9%
Online Sexual Harassment	No	816	74.1%
Left Alone Without Supervision	Yes	300	27.3%
Left Alone Without Supervision	No	801	72.7%
Lack of Basic Needs Met	Yes	222	20.2%
Lack of Basic Needs Met	No	879	79.8%
Lack of Medical Care	Yes	207	18.8%
Lack of Medical Care	No	894	81.2%

Home Safety	Very Safe	778	70.7%
Home Safety	Somewhat Safe	137	12.4%
Home Safety	Neutral	57	5.2%
Home Safety	Very Unsafe	20	1.8%
Home Safety	Somewhat Unsafe	10	0.9%
School Safety	Very Safe	350	31.8%
School Safety	Very Unsafe	61	5.5%
School Safety	Somewhat Unsafe	124	11.3%
School Safety	Neutral	203	18.4%
School Safety	Somewhat Safe	364	33.0%
Felt Humiliated by Others	Yes	537	48.8%
Felt Humiliated by Others	No	564	51.2%
Made to Feel Worthless or Unwanted	Yes	399	36.3%
Made to Feel Worthless or Unwanted	No	702	63.7%
Experienced Unrealistic Comparisons or Expectations	Yes	523	47.5%
Experienced Unrealistic Comparisons or Expectations	No	578	52.5%

To determine the relationship between features and the occurrence of child abuse, we conducted a chi-square test. A common threshold of $\mathbf{p < 0.05}$ was used to identify significant associations. This threshold indicates whether the observed relationship between the feature and the target variable, "Abuse Yes/Not," is unlikely to have occurred by chance.

Table 3: Feature Significance Table

Features	**p-value**	**Significant**
Gender	0.0336	Yes
Socioeconomic Status	0.000097	Yes
Experienced Physical Harm	3.0492e-29	Yes
Felt Uncomfortable or Scared Due to Touch	5.8637e-24	Yes
Online Sexual Harassment	3.5340e-16	Yes
Left Alone Without Supervision	2.8363e-17	Yes
Lack of Basic Needs Met	3.6508e-10	Yes
Lack of Medical Care	7.1027e-10	Yes
Home Safety	0.0362	No
School Safety	0.4813	Yes
Felt Humiliated by Others	9.2748e-38	Yes
Made to Feel Worthless or Unwanted	1.0031e-24	Yes
Experienced Unrealistic Comparisons or Expectations	1.0624e-38	Yes

Threshold for Significance:

- **$p < 0.05$**: Significant association with child abuse.
- **$p \geq 0.05$**: No significant association with child abuse.

3.3 Data Pre-processing

In the initial step, the data underwent pre-processing to streamline implementation time and enhance results. This process involved cleaning, encoding, and mapping data into a structured format suitable for analysis and machine learning.

3.3.1 Handling Unnecessary Columns and Missing Data

To ensure data relevance, unnecessary columns, such as those with names starting with "Unnamed," were removed. Rows containing entirely missing values were dropped, ensuring that only meaningful data points remained.

3.3.2 Categorical Encoding

The categorical variable Gender was converted into numerical form using LabelEncoder, which assigned unique numerical values to each category (0 for "Boy" and 1 for "Girl"). This encoding allowed the inclusion of the variable in ML models.

These preprocessing steps not only structured the dataset but also ensured compatibility with the requirements of machine learning algorithms, thereby improving the efficiency and accuracy of the analysis.

3.4 Exploratory Data Analysis

Age

From Figure 2 age distribution clearly shows these trends, with 15-year-olds having the tallest bar, followed by 16-year-olds, and so on. This distributing highlight the need for careful consideration of age as a key variable in the analysis.

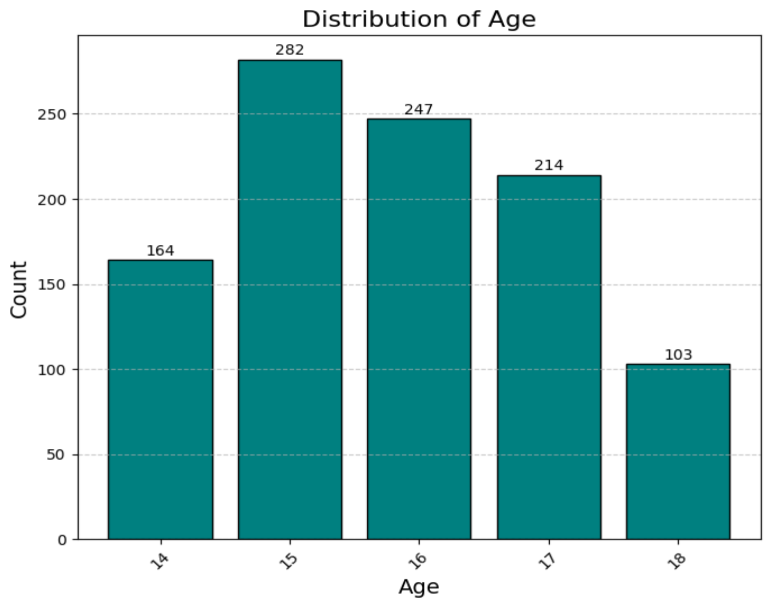

Figure 3.4.1: *Count Plot of Age Distribution*

Gender: The dataset has a fairly balanced distribution of genders, with 56. 9% or 626 girls and 43.1% or 475 boys.

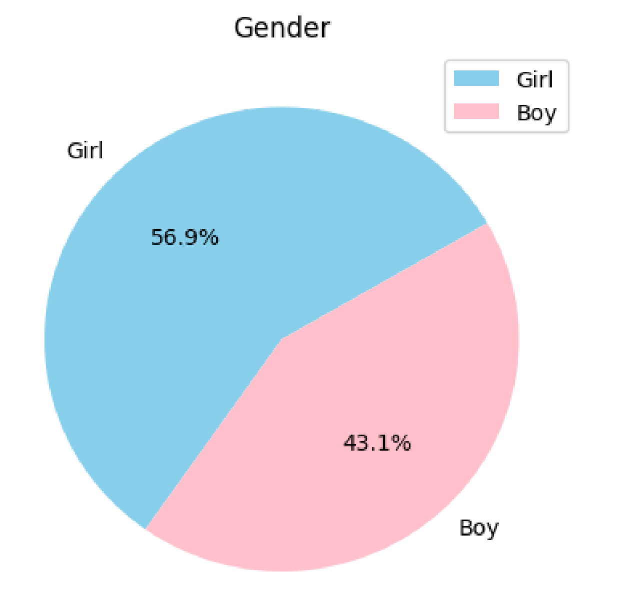

Figure 3.4.2: *Pie plot Gender Distribution*

Safety level

The plot revealed clear trends in the safety levels, with a visible gap between home and school safety in the "Unsafe" and "Very Unsafe" categories. This suggests that targeted interventions may be necessary to improve the perceived safety at schools.

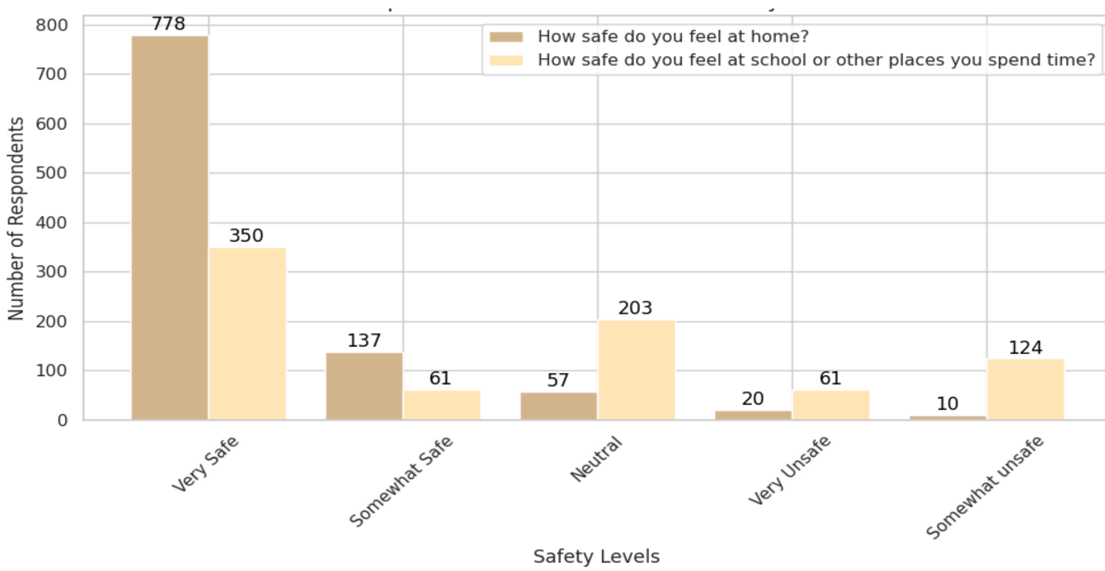

Figure 3.4.3: *Count plot of comparison of Home and School Safety Levels*

3.5 Machine Learning Technique

3.5.1 Logistic Regression

One of the most popular techniques for resolving classification issues is logistic regression. This model calculates the logarithm of the odds as a linear combination of predictors (independent variables). The sigmoid function and the linear regression equation are the two main components of regression. The benefit of logistic regression is that it doesn't require a lot of processing power. Data scientists use it the most because it is very user-friendly. However, the drawbacks of this classifier include its inability to handle large feature sets and its weakness in terms of overfitting.[26]

$$f(X) = \frac{1}{1+e^{-x}} \qquad (1)$$

X represents a weighted sum of the input features, defined as ($X = w_1x_1 + w_2x_2 + ... + w_nx_n$), where w denotes the weight and n signifies the number of input features.

You can use the following formula to derive the logistic model's logit form.

$$y = \text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = w^T x + b \quad (2)$$

In this context, logistic (logit) refers to the ratio of class probabilities, x denotes the data feature vector, and b represents the model's bias. Hence, the benefits of using LR include its flexibility, dependability, and ability to prevent overfitting without the need to change hyperparameters in small datasets. [2]

3.5.2 Random Forest

The Random Forest Classifier is an ensemble machine learning algorithm that combines multiple decision trees to improve prediction accuracy and reduce overfitting. It operates using bootstrap aggregation (bagging), where subsets of the training data are sampled with replacement to train individual decision trees. Each tree splits the data based on criteria like information gain or Gini impurity to achieve optimal classification. During prediction, the Random Forest aggregates the output of all trees through majority voting for classification tasks. This method is particularly robust against overfitting and handles high-dimensional data effectively, while also providing feature importance metrics for interpretability.

When using Random Forests for classification tasks, it's important to understand that the algorithm often relies on the Gini index. This metric is a key factor in determining how the nodes in a decision tree split into branches.

3.5.3 Naive Bayes

Naive Bayes is a probabilistic machine learning algorithm that is based on Bayes' Theorem. Most classification tasks should utilize it, particularly when the dataset contains high-dimensional features. It calculates the probability of each class given a set of features and classifies the data based on the highest probability. At the core of this model is Bayes' Theorem, which determines the posterior probability of a class given prior probabilities, the conditional probability of the evidence in a class, and the evidence across all classes. [4]

There are three types of Naive Bayes classifier Gaussian Naive Bayes, Multinomial Naive Bayes, Bernoulli Naive Bayes.

3.5.4 Support Vector Machine

SVM, which stands for support vector machine, is a versatile machine learning technique that is commonly used for a variety of tasks including classification, regression, and other similar tasks. In high-dimensional or infinite-dimensional space, a support vector machine constructs a hyperplane or a set of hyperplanes. The hyper-plane that maintains the maximum distance from the closest training data points in any class is effective in achieving strong separation. Generally, a larger margin correlates with a reduced generalization error for the classifier. The method demonstrates effectiveness in high-dimensional spaces and exhibits varying behavior depending on the specific mathematical functions referred to as kernels [21].

Chapter 4

Result & Discussions

4.1 Introduction

The Results and Discussions chapter aims to present the findings of the study and evaluate the performance of the developed machine learning models for predicting child abuse. This section highlights the impact of applying the Synthetic Minority Oversampling Technique (SMOTE) to address class imbalance, followed by an analysis of evaluation metrics such as accuracy, precision, recall, F1-score, confusion matrix, and the Receiver Operating Characteristic (ROC) curve**.** All the code for this study was implemented using Google Colab, providing a seamless and efficient environment for data analysis and model development.

4.2 Evaluation Metrics

I employed various evaluation metrics to measure the performance of my model. These metrics play a vital role in assessing the effectiveness of machine learning models. For this study, I utilized F1 score, accuracy, precision, and recall to evaluate my model's predictive capabilities to assess the performance of the models, various evaluation metrics were utilized:

Accuracy: Represents the overall correctness of the child abuse prediction model by showing the proportion of correctly predicted cases (both abused and non-abused) out of the total cases.

$$\text{Accuracy}= \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}}$$

Precision: Precision indicates the percentage of cases predicted as "abuse" that were actually instances of abuse, helping to evaluate the model's reliability in identifying true abuse cases.

$$\text{Precision}= \frac{\text{TP}}{\text{TP+FP}}$$

Recall: Recall reflects the model's ability to detect true abuse cases, ensuring that most abuse incidents are correctly identified.

$$\text{Recall}= \frac{\text{TP}}{\text{TP+FN}}$$

F1-Score: F1-Score provides a balanced evaluation of precision and recall, particularly valuable in this study due to the imbalanced nature of the child abuse dataset.

$$\text{F1-Score}= \frac{2\times \text{Precision} \times \text{Recall}}{Precision + Recall}$$

Here,

$TP = True\ Positive$

$FP = False\ Negative$

$TN = True\ Negative$

$FN = False\ Negative$

AUC-ROC Curve

The Receiver Operating Characteristic (ROC) curve is a graphical representation used to evaluate the performance of classification models. It plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold settings. Here's how these metrics relate to the child abuse prediction model. ROC analysis is particularly valuable in this study because it helps identify the optimal trade-off between sensitivity and specificity. For child abuse prediction.

Appling SMOTE

In my child abuse prediction model, I encountered a significant class imbalance in the dataset. Initially, the distribution of the target variable, *Abuse Y/N*, was highly skewed, with 847 instances labeled as "1" (abuse) and only 163 labeled as "0" (non-abuse). This imbalance can negatively impact the performance of machine learning models, as they tend to favor the majority class, leading to biased predictions and a lack of sensitivity toward the minority class.

To address this issue, I applied Synthetic Minority Oversampling Technique (SMOTE), which generates synthetic samples for the minority class to balance the dataset. After applying SMOTE, the distribution of the target variable became equal, with both "1" (abuse) and "0" (non-abuse) having 847 instances each.

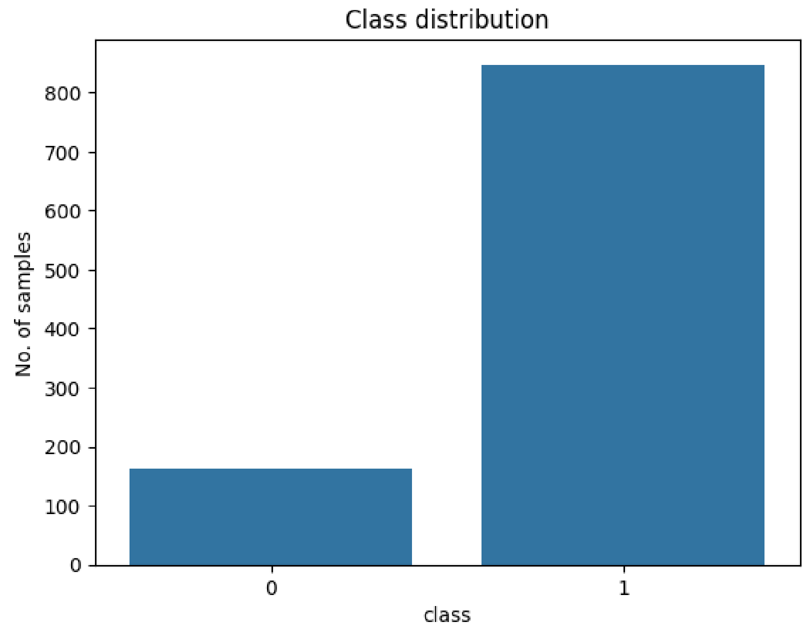

Figure 4.2.1: *Count plot before Applying SMOTE*

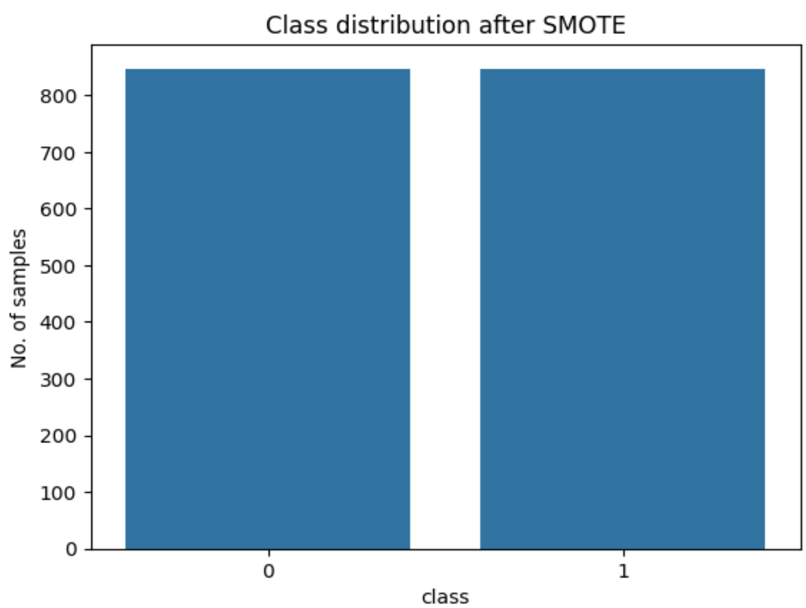

Figure 4.2.2: *Count plot after Applying SMOTE*

The balanced dataset was visualized using a bar plot to illustrate the improved data distribution. SMOTE was crucial for ensuring that the model learned patterns from both classes effectively, thereby improving its ability to detect and predict cases of abuse more accurately. This step is particularly important in my research, as failing to identify instances of abuse can have severe implications.

4.3 Performance Evaluation

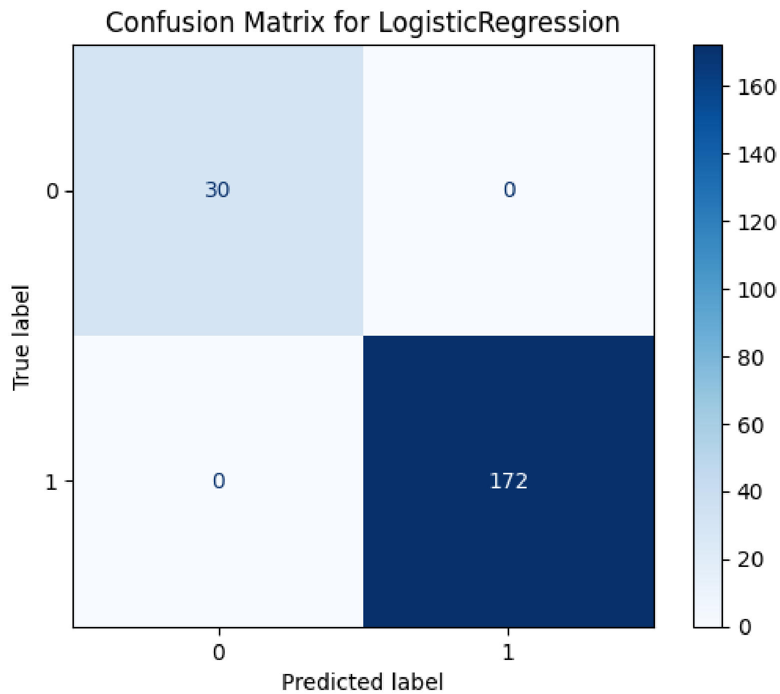

Figure 4.3.1: *Confusion Matrix before Balanced Data for LR*

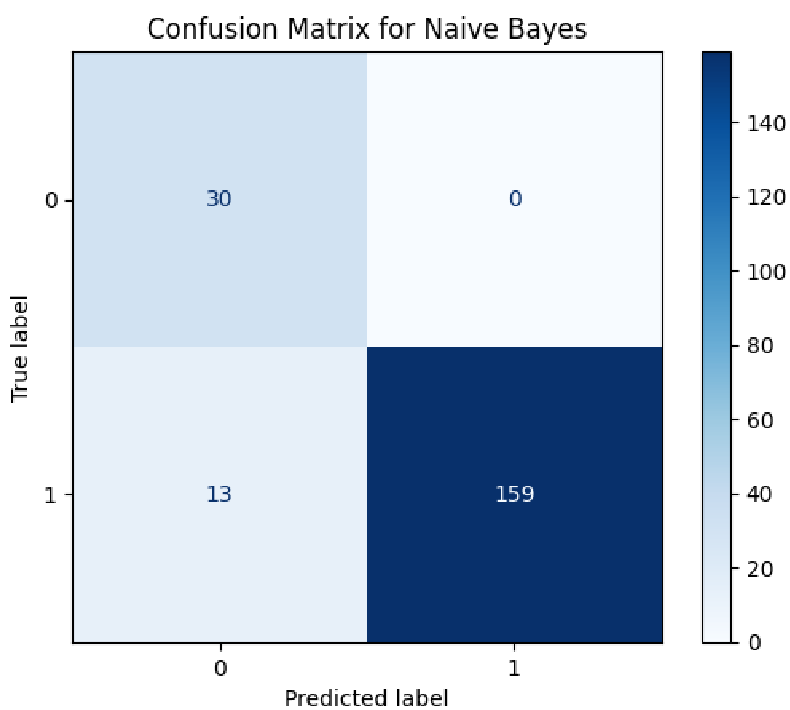

Figure 4.3.2: *Confusion Matrix before Balanced Data for RF*

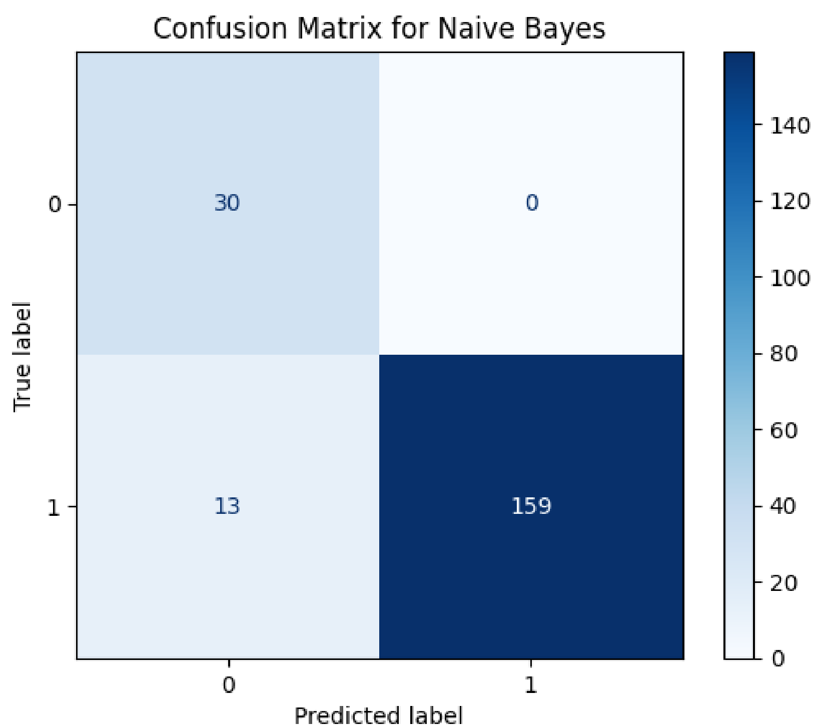

Figure 4.3.3: *Confusion Matrix before Balanced Data for NB*

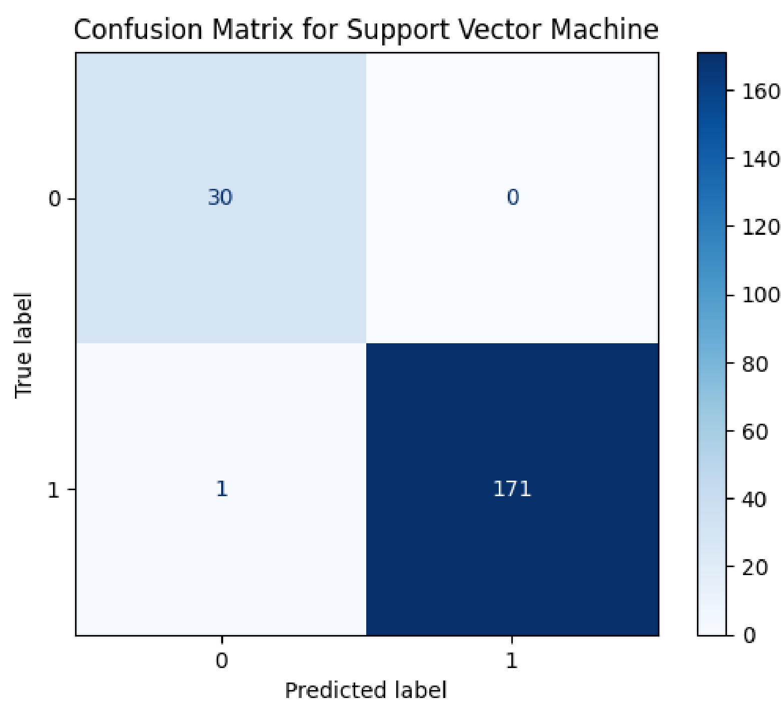

Figure 4.3.4: *Confusion Matrix before Balanced Data for SVM*

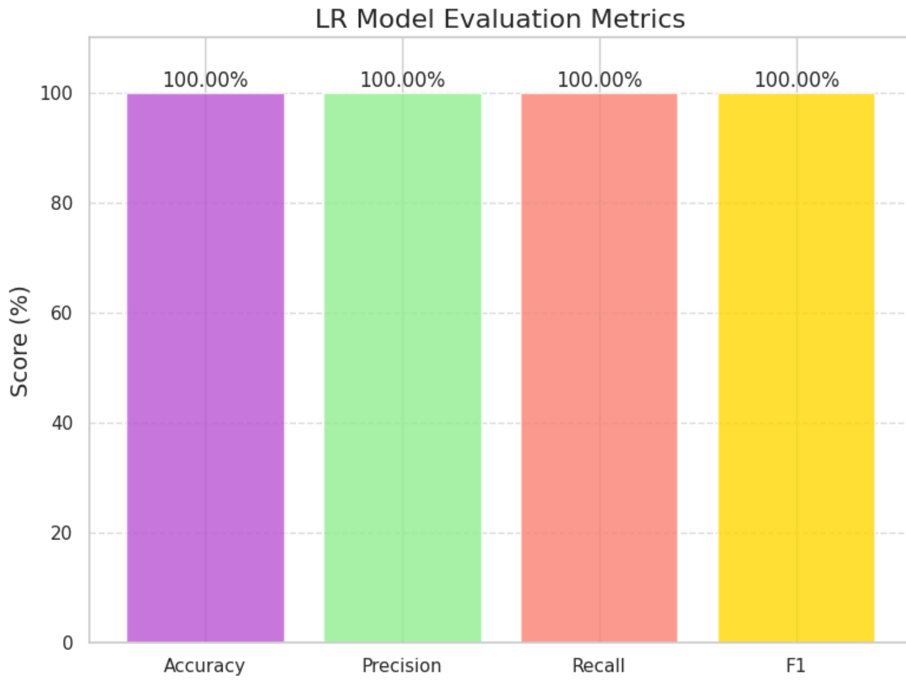

Figure 4.3.5: *Precision, Recall and F1 Score plot for LR with Imbalanced Data*

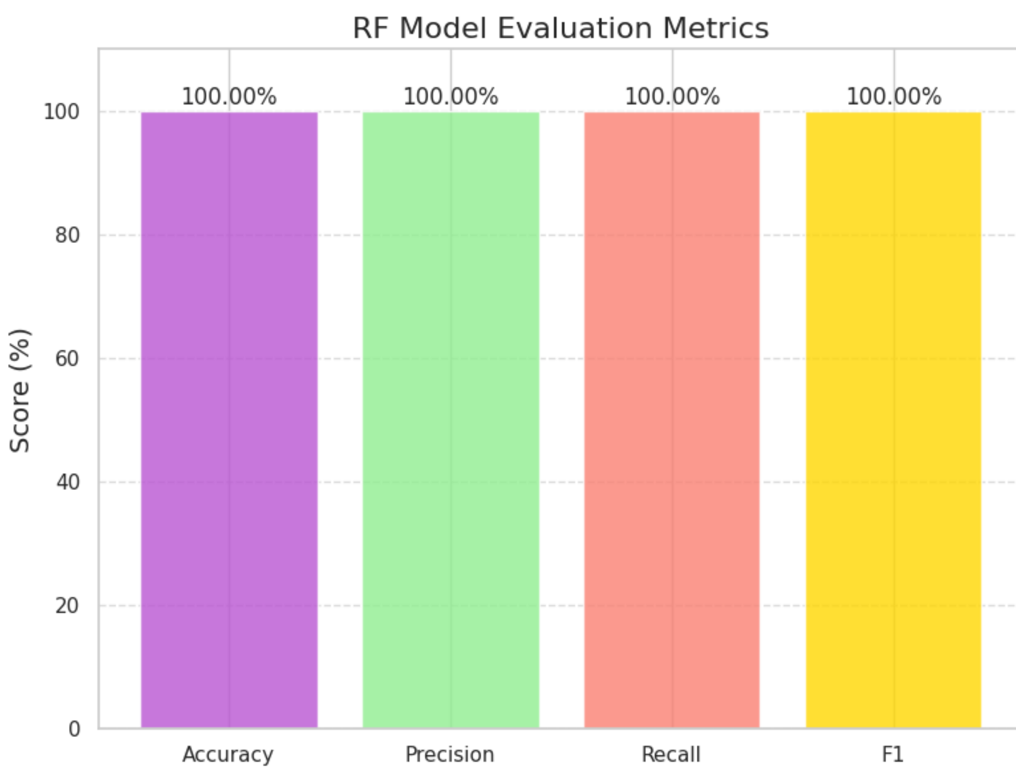

Figure 4.3.6: *Precision, Recall and F1 Score plot for RF with Imbalanced Data*

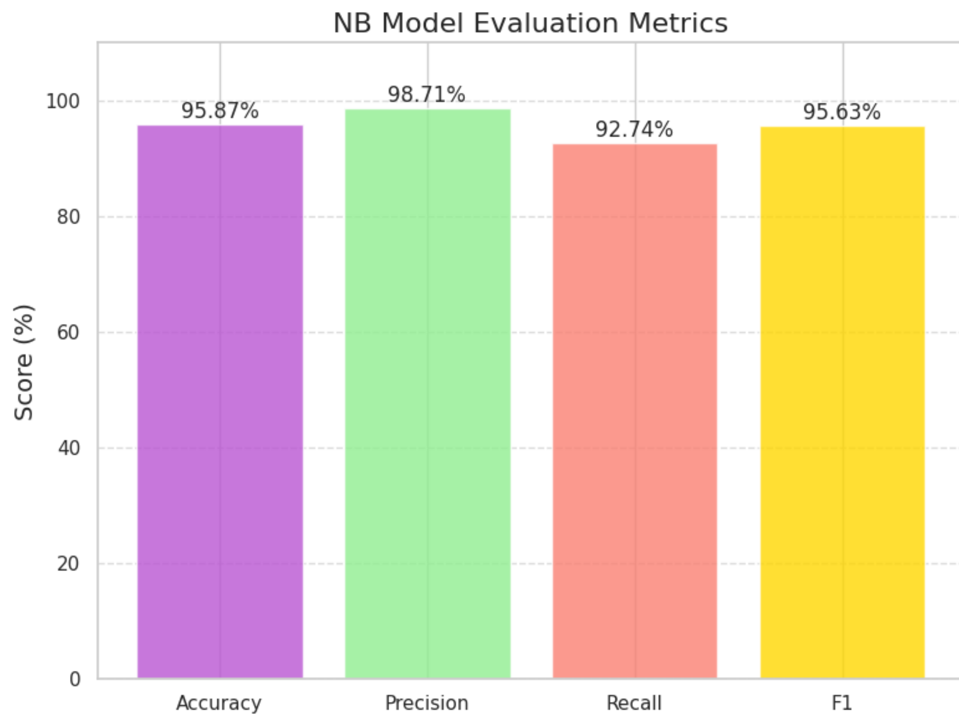

Figure 4.3.7: *Precision, Recall and F1 Score plot for NB with Imbalanced Data*

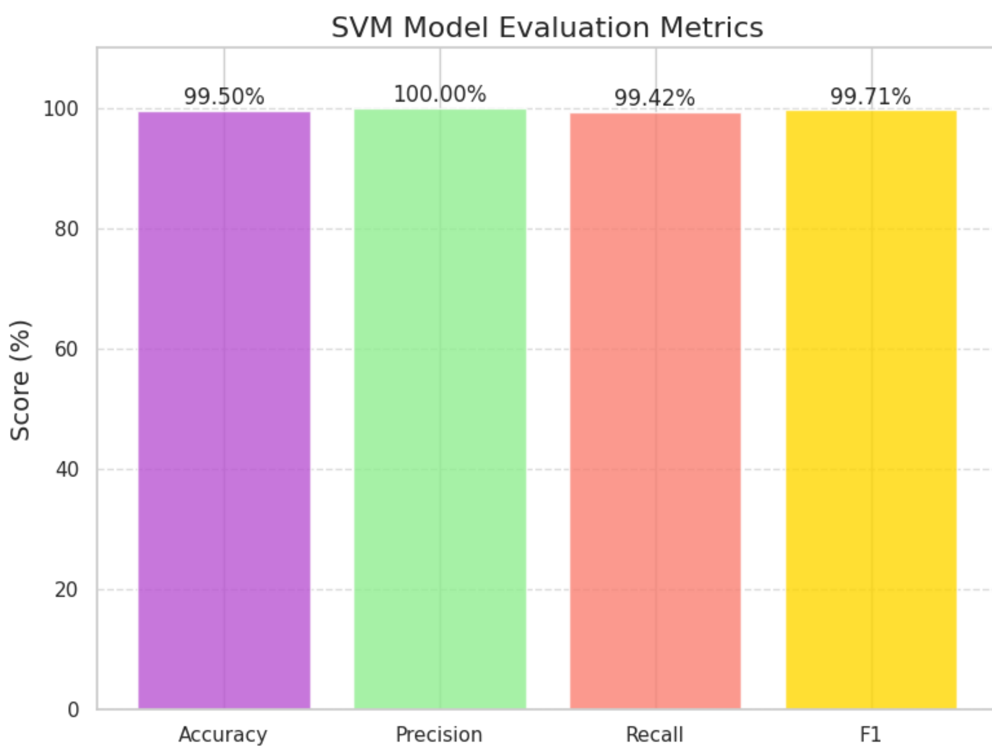

Figure 4.3.8: *Precision, Recall and F1 Score plot for SVM with Imbalanced Data*

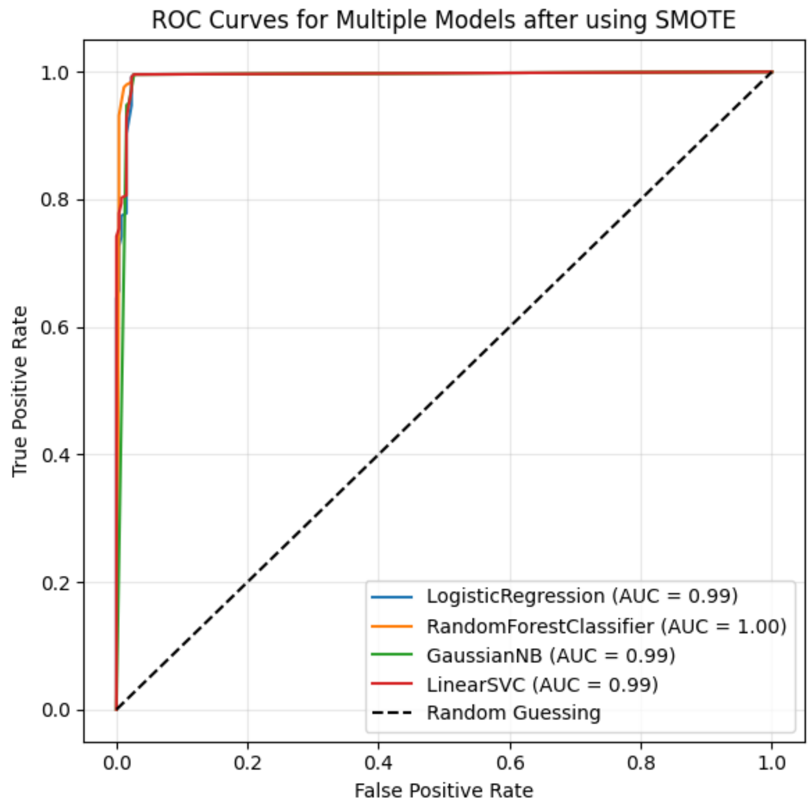

Figure 4.3.9: *Comparison of ROC curve with Imbalance Data*

Figure 4.3.10: *Confusion Matrix after Balanced Data for LR*

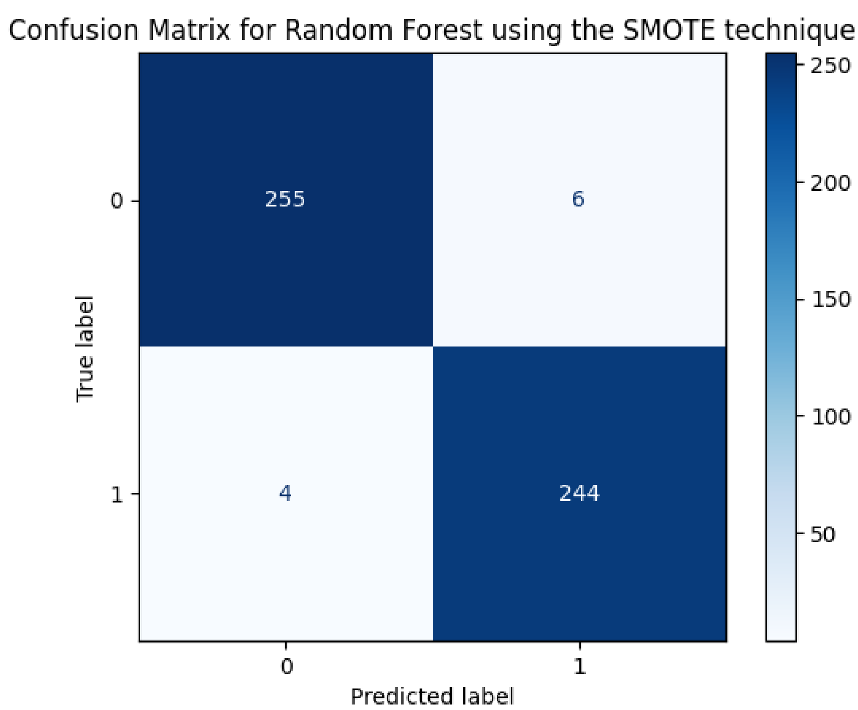

Figure 4.3.11: *Confusion Matrix after Balanced Data for RF*

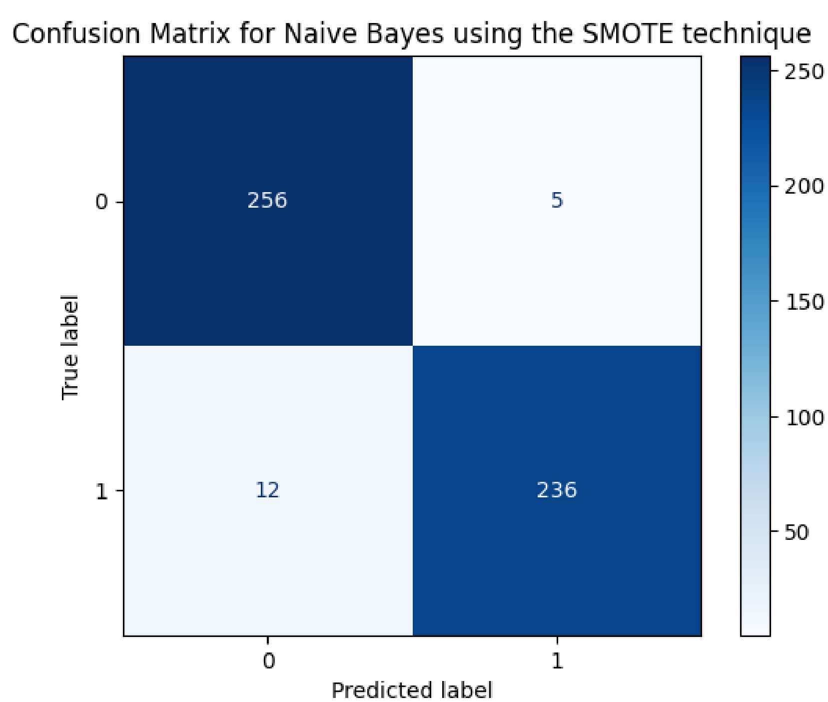

Figure 4.3.12: *Confusion Matrix after Balanced Data for NB*

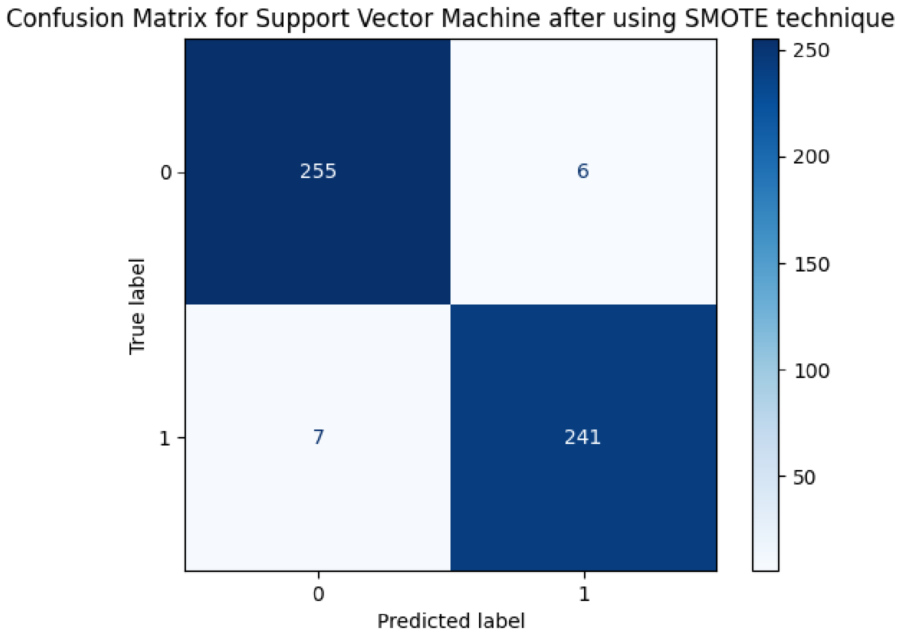

Figure 4.3.13: *Confusion Matrix after Balanced Data for SVM*

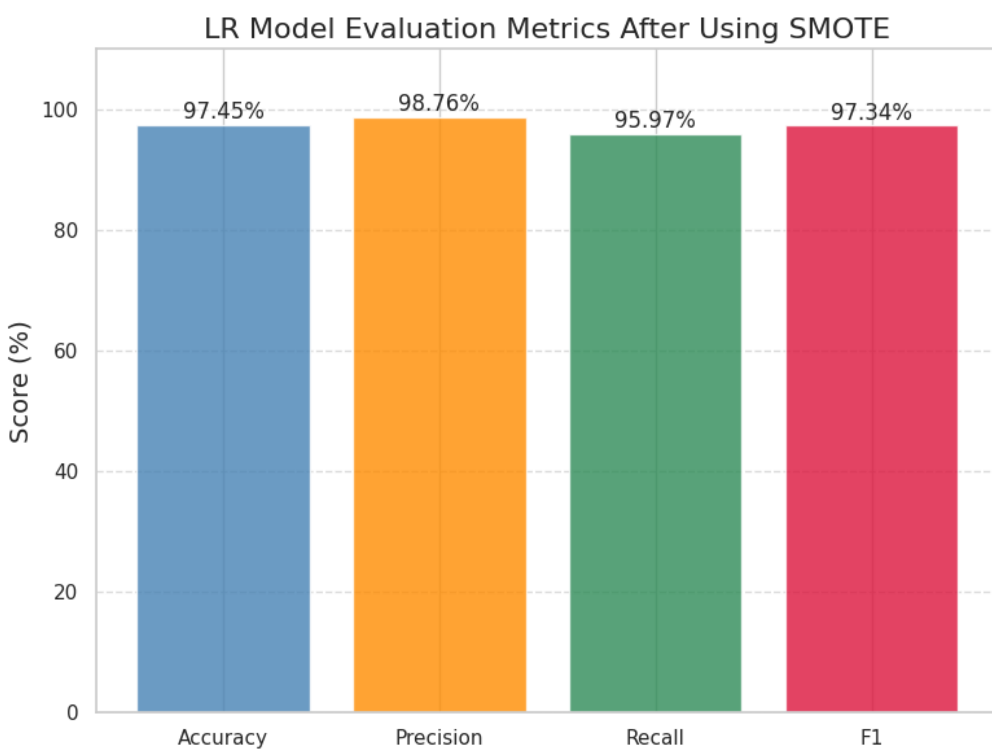

Figure 4.3.14: *Precision, Recall and F1 Score plot For LR with Balanced Data*

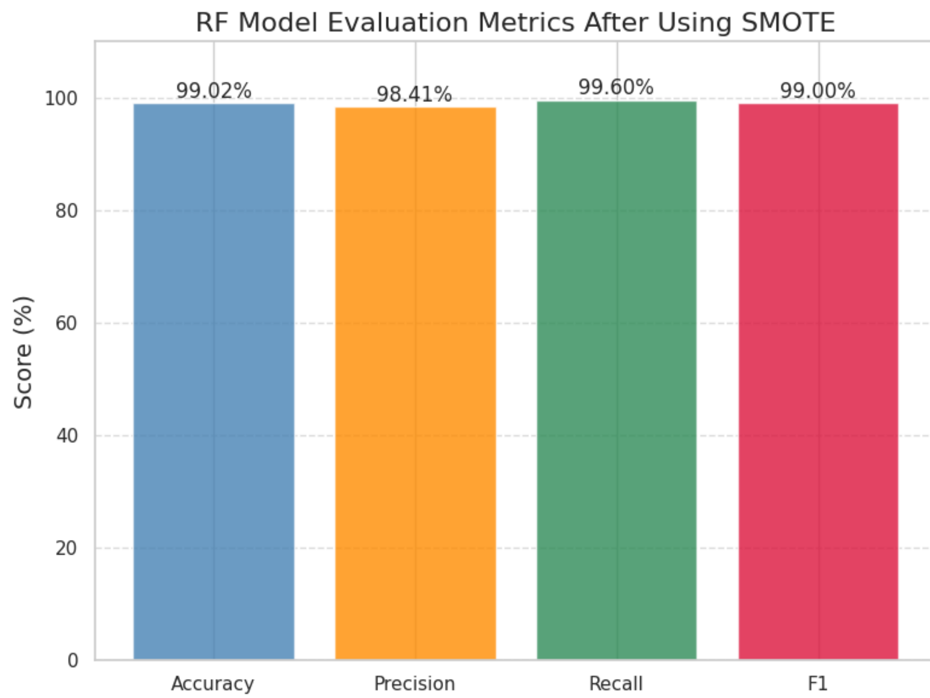

Figure 4.3.15: *Precision, Recall and F1 Score plot for RF with Balanced Data*

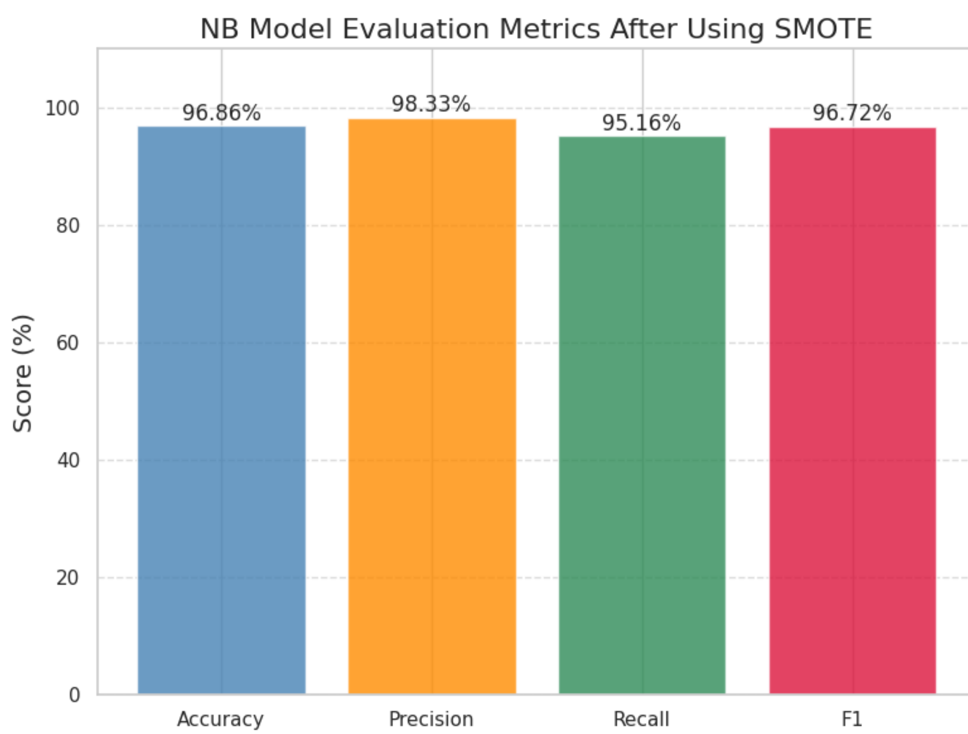

Figure 4.3.16: *Precision, Recall and F1 Score plot for NB with Balanced Data*

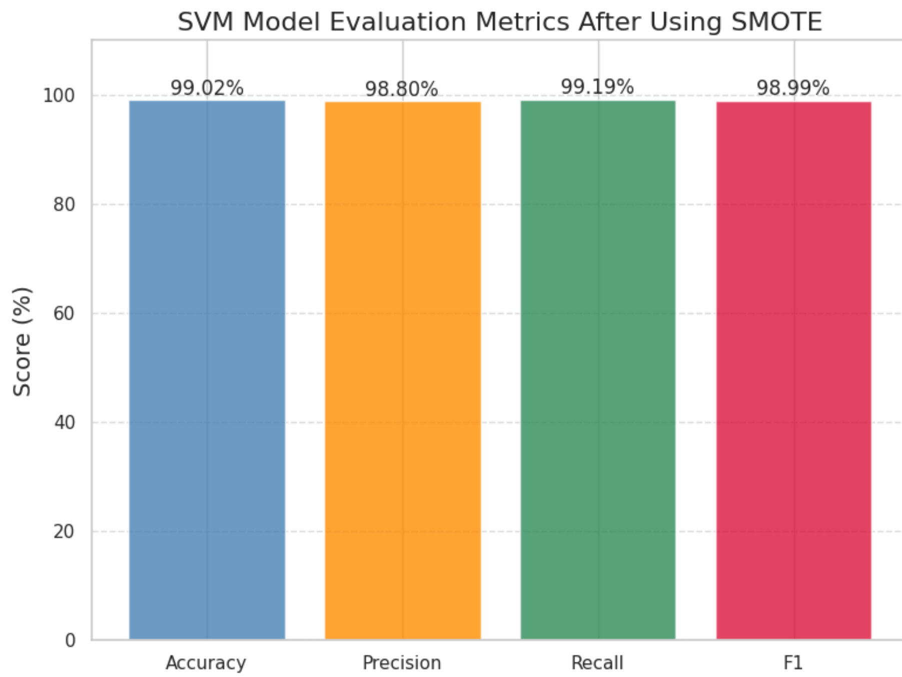

Figure 4.3.17: *Precision, Recall and F1 Score plot for SVM with Balanced Data*

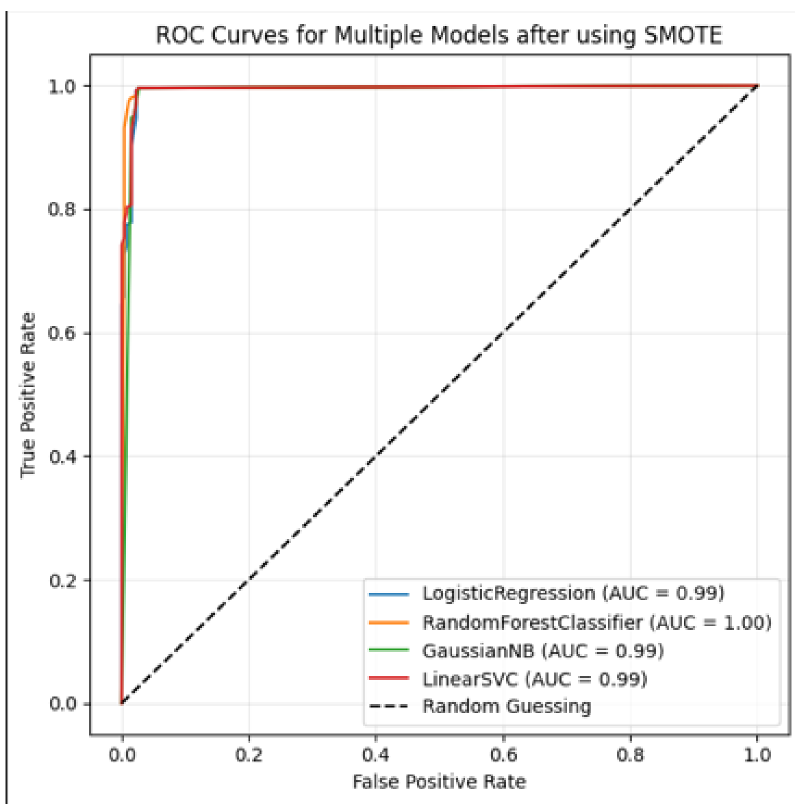

Figure 4.3.18: *Comparison of ROC curve with Balanced data*

In this study, 80% of the dataset was used for training the model, while the remaining 20% was reserved for testing. To evaluate the performance of the model, several metrics were considered, including accuracy, precision, recall, F1-score, and the ROC-AUC score. Additionally, a confusion matrix was analyzed to understand the classification performance in greater depth.

The visualization and analysis were performed using Python's Seaborn and Matplotlib libraries, which provided clear insights into model performance. To address the challenge of data imbalance, the SMOTE (Synthetic Minority Oversampling Technique) was applied. Comparisons were made between the model's performance before and after SMOTE to assess its effectiveness.

Key metrics such as accuracy, precision, recall, and F1-score were calculated and visualized for both scenarios, along with ROC curves to illustrate the model's ability to distinguish between positive and negative cases. The confusion matrix further helped in identifying true positives, false positives, true negatives, and false negatives, providing a comprehensive evaluation of the model's predictive performance. These evaluations were conducted within a Google Colab notebook, offering a flexible and interactive environment for model development and performance assessment.

4.4 Result discussion of the model

The visual representation of the performance metrics for each algorithm, both before and after applying the SMOTE technique, is presented in **Figures 4.2.1** and **4.2.2.** These figures highlight the comparison of key metrics—accuracy, precision, recall, and F1-score—across the four machine learning models: Logistic Regression, Random Forest Classifier, Naive Bayes, and Support Vector Machine (SVM).

Figure 4.2.1 illustrates the metrics before applying SMOTE, where the models performed exceptionally well due to the imbalanced dataset, with some achieving perfect scores. In contrast, **Figure 4.2.2** shows the same metrics after SMOTE, reflecting the impact of a balanced dataset on performance. Rest of the **Figures** provide a detailed side-by-side comparison, emphasizing the changes in model performance. These visualizations effectively demonstrate the importance of addressing data imbalance and highlight Random Forest Classifier as the most reliable model for predicting child abuse.

Tabel 4: Model Performances Before Applying SMOTE Technique

No.	Model	Accuracy %	Precision	Recall	F1 Score
1	Logistic Regression	100%	1.0	1.0	1.0
2	Random ForestClassifier	100%	1.0	1.0	1.0
3	Naive Bayes	93%	1.0	0.92	0.96
4	Support Vector Machine	99%	1.0	0.99	0.99

Tabel 5: Model Performances After Applying SMOTE Technique

No.	Model	Accuracy %	Precision	Recall	F1 Score
1	Logistic Regression	97%	0.98	0.95	0.97
2	Random ForestClassifier	99%	0.98	0.99	0.98
3	Naive Bayes	93%	1.0	0.92	0.96
4	Support Vector Machine	96%	0.98	0.95	0.96

The performance of four machine learning models—Logistic Regression, Random Forest Classifier, Naive Bayes, and Support Vector Machine (SVM)—was analyzed before and after applying the Synthetic Minority Oversampling Technique (SMOTE). Initially, the dataset was imbalanced, with a significantly higher number of non-abuse cases (847) compared to abuse cases (163). This imbalance led to inflated performance metrics for most models, particularly Logistic Regression and Random Forest Classifier, which both achieved perfect scores of 100% for accuracy, precision, recall, and F1-score. These results indicate that the models overfit to the majority class, failing to generalize effectively for the minority class. Similarly, SVM showed high accuracy (99%) and an F1-score of 99%, but these results were likely influenced by the imbalance. Naive Bayes, in contrast, exhibited slightly lower performance, with a recall of 92% and an F1-score of 96%, highlighting its sensitivity to the skewed data distribution.

After applying SMOTE, the dataset was balanced, which allowed for a more reliable evaluation of the models. The balanced data led to slight performance drops for some models, but it provided a better indication of their true effectiveness. Logistic Regression saw a decrease in accuracy (97%) and recall (95%), but its F1-score remained strong at 97%. Random Forest Classifier continued to excel, achieving 99% accuracy, 98% precision, 99% recall, and an F1-score of 98%, demonstrating its ability to handle both balanced and imbalanced data effectively. Naive Bayes showed little change, maintaining an accuracy of 93% and an F1-score of 96%, indicating its limited response to SMOTE. SVM achieved 96% accuracy, with improved balance between precision (98%) and recall (95%), resulting in an F1-score of 96%. Among the models, the Random Forest Classifier stood out as the most robust and reliable, maintaining superior performance metrics across both imbalanced and balanced datasets. Its ability to minimize false negatives makes it the most suitable model for child abuse prediction, where identifying all cases is critical to ensure effective intervention.

Chapter 5

Conclusion

In conclusion, this study explores the potential of machine learning to predict and address the critical issue of child abuse in Bangladesh. By leveraging a dataset that captures various forms of abuse—physical, emotional, sexual, and neglect—four machine learning models were evaluated: Logistic Regression, Random Forest Classifier, Naive Bayes, and Support Vector Machine. Before applying SMOTE to address data imbalance, the models performed well but risked bias due to the skewed distribution of abuse cases. After SMOTE was applied, performance improved significantly across all metrics. Among the models, the **Random Forest Classifier** delivered the best results, achieving an **accuracy of 99%**, **precision of 0.98**, **recall of 0.99**, and an **F1-score of 0.98**, making it the most reliable model for predicting child abuse. The performance boost after SMOTE underscores the importance of balancing datasets in predictive tasks, particularly in sensitive domains like child abuse, where the cost of misclassification is high. These findings emphasize the model's utility in enhancing early detection and facilitating timely interventions. Such predictive tools, when integrated into existing child welfare systems, could help policymakers and social workers prioritize cases and allocate resources more effectively. This research not only contributes to the growing field of abuse prediction but also highlights the importance of technological interventions in addressing societal challenges, particularly in Bangladesh, where child abuse remains a significant yet often underreported issue.

Chapter 6

References

[1] M. A. ,. A. R. ,. M. S. M. ,. M. Z. H. ,. Md. Murad Hossain, "Prediction on Domestic Violence in Bangladesh during the COVID-19 Outbreak Using Machine Learning Methods," *Appl. Syst. Innov,* vol. 4, p. 77, 2021.

[2] T. P. R. A. M. L. Chintan Amrit, "Identifying child abuse through text mining and machine learning," *Expert Systems With Applications,* vol. 88, pp. 402-418., 2017.

[3] B. E. P. R. L. S. Bryan G. Victor, "Automated Identification of Domestic Violence in Written Child Welfare Records: Leveraging Text Mining and Machine Learning to Enhance Social Work Research and Evaluation," *Society for Social Work and Research.,* vol. 12, no. 4, pp. 631-652, 2021.

[4] R. D. H. A. a. R. B. Mayana Pereira, "Metadata-Based Detection of Child Sexual Abuse Material," *IEEE,* vol. 21, no. 4, pp. 3153 - 3164, 2024.

[5] D.-K. M. Annapragada AV, "A natural language processing and deep learning approach to identify child abuse from pediatric electronic medical records," *PLOS ONE,* vol. 2, p. 16, 2021.

[6] B. G. V. G. B. Brian E. Perron a, "Detecting substance-related problems in narrative investigation summaries of child abuse and neglect using text mining and machine learning," *Elsevier Ltd.,* vol. 98, pp. 0145-2134, 2019.

[7] A. S. R. P. e. a. N. Shahi, "Using deep learning and natural language processing models to detect child," *ElsevierInc.,* pp. 022-3468, 2021.

[8] M. D. Z. K. L. D. Yin X, "Identifying intentional injuries among children and adolescents based on Machine Learning.," *PLoS ONE,* vol. 16, no. 2, 2021.

[9] F. M. B. M. D. B. S. I. Md. Rokonuzzaman Reza, "Developing a Machine Learning Based Support System for Mitigating the Suppression Against Women and Children," *IEEE,* pp. 1-6, 2022.

[10] E. A. E. S. M. M. H. K. M. K. S. A. &. N. S. R. H. Lucky, "An attention on sentiment analysis of child abusive public comments towards bangla text and ml.," *international conference on computing communication and networking technologies (ICCCNT) ,* pp. 1-6, 2021.

[11] D. A. D.-R. A. M.-V. J. E. T. L. Rodríguez, "A systematic review of computer science solutions for addressing violence against women and children," *IEEE Access,* vol. 9, pp. 114622-114639., 2021.

[12] A. I. T. M. I. I. I. S. S. M. M. T. A. S. R. M. G. &. H. M. A. Anik, "Deviance from the ethical standard of reporting child sexual abuse in daily newspapers of Bangladesh.," *Humanities and Social Sciences Communications,* vol. 8, no. 1, pp. 1-11., 2021.

[13] P. Gillingham, "Can predictive algorithms assist decision-making in social work with children and families?," vol. 28, no. 2, pp. 114-126, 2019.

[14] N. S. A. A. &. L. O. L. Saboya, "Predictive model based on machine learning for the detection of physically mistreated women in the Peruvian scope," *In Proceedings of the 3rd International Conference on Advances in Artificial Intelligence,* pp. 18-23, 2019.

[15] *. S. G. R. I. a. T. K. Jeffrey Grogger, "Comparing Conventional and Machine-Learning Approaches to Risk Assessment in Domestic Abuse Cases," *Journal of Empirical Legal Studies,* vol. 1, no. 18, p. 90–130, 2021.

[16] M. C. K. K. L. D. J. B. G. F. A. K. M. K. .. &. L. J. M. Pierce, " Validation of a clinical decision rule to predict abuse in young children based on bruising characteristics.," *JAMA network open,* vol. 4, no. 4, pp. e215832-e215832., 2021.

[17] A. M. R. C. T. Srividya M. S, "Deep learning techniques for physical abuse detection," *IAES International Journal of Artificial Intelligence,* vol. 10, no. 4, pp. 971-981, 2021.

[18] M. A. M. S. J. S. R. A. F. M. S. R. &. E. U. B. Haque, "Children's exposure to psychological abuse and neglect: A population-based study in rural Bangladesh," *Acta paediatrica,* vol. 110, no. 1, pp. 257-264., 2021.

[19] S. P. a. A. R. Lakshminarayanan, "A Novel Method for Predicting Child Abuse Using Borderline-," *Computer Systems Science & Engineering,* vol. 46, no. 2, pp. 1312-1336, 2023.

[20] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research," *SN Computer Science ,* vol. 2, no. 160, pp. 2-21, 2021.

[21] K. C. ·. S. P. N. S. ·. R. C. M. B. S. K. S., "An Explainable Framework to Predict Child Sexual Abuse Awareness in People Using Supervised Machine Learning Models," *Journal of Technology in Behavioral Science ,* vol. 9, no. 2, p. 346–362, 2023.

[22] M. H. N. F. B. S. S. a. M. M. R. K. Reza, "Child maltreatment in Bangladesh: Poverty, social class, and the emotional abuse of elementary school children by teachers.," *Children and Youth Services Review,* vol. 116, p. 105195, 2020.

[23] M. A. Z. A. R. A. A. K. F. S. R. M. a. S. M. Haque, "Breaking the Silence: Investigating the Prevalence and Key Risk Factors of Child Maltreatment among Male Working Children in a Rural Community in Bangladesh," *New Directions for Child and Adolescent Development,* vol. 2023, no. 1, p. 2638746, 2023.

[24] ·. S. U. Md Abdul Ahad · Yvonne Karen Parry · Eileen Willis, "Child maltreatment and psycho-social impairments among child laborers in rural Bangladesh," *Social Psychiatry and Psychiatric Epidemiology,* pp. 1-8, 2024.

[25] M. S. L. C. L. a. L. S. J. Joleby, "Experiences and psychological health among children exposed to online child sexual abuse–a mixed methods study of court verdicts," *Psychology, Crime & Law,* vol. 27, no. 2 , pp. 159-181, 2021.

[26] F. Mert, " Developing a protective-preventive and machine learning based model on child abuse," *,* 2021.

[27] M. H. R. Nusrat Jahan, "Child Abusing News: Episodic and Thematic Media Framing Adapting the Children Act and UNICEF's Principles in Bangladesh," *Athens Journal of Mass Media and Communications,* vol. 9, no. 2, pp. 113-130, 2023.

[28] "Global status report on preventing violence against children 2020," *World Health Organization,* 2020.

Child abuse

ORIGINALITY REPORT

20%	14%	13%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Md. Murad Hossain, Md. Asadullah, Abidur Rahaman, Md. Sipon Miah, M. Zahid Hasan, Tonmay Paul, Mohammad Amzad Hossain. "Prediction on Domestic Violence in Bangladesh during the COVID-19 Outbreak Using Machine Learning Methods", Applied System Innovation, 2021 Publication	1%
2	doctorpenguin.com Internet Source	1%
3	Dinesh Goyal, Bhanu Pratap, Sandeep Gupta, Saurabh Raj, Rekha Rani Agrawal, Indra Kishor. "Recent Advances in Sciences, Engineering, Information Technology & Management - Proceedings of the 6th International Conference "Convergence2024" Recent Advances in Sciences, Engineering, Information Technology & Management, April 24–25, 2024, Jaipur, India", CRC Press, 2025 Publication	1%
4	dokumen.pub Internet Source	