Ain Shams University

Faculty of Computer & Information Sciences

Information Systems Department

Automatic Risk Assessment of School Violence

This documentation is submitted as required for the degree of bachelors in Computer and Information Sciences

By:

Mohammed Mostafa Refaat Abd Elhamed (20191700576)

Abdelrahman Hassan Mohamed Mahmoud (20191700360)

Abdelrahman Rezk Ibrahim Eldosoki (20191700337)

Raouf Fayek Attia Abdullah (2016170571)

Sherif Ahmed El-Sayed Ahmed (2016170219)

Supervised by:

Dr. Naglaa Fathy

June 2023

Acknowledgments

We extend our heartfelt gratitude to all those who have contributed to the success of our project. We would like to express our special appreciation to the Graduation Project Unit for their unwavering support and provision of valuable information, paving the way for students to apply their educational knowledge in real-world project design and analysis. We are sincerely grateful to our professors and industry experts for delivering engaging lectures that have greatly benefited us throughout our academic journey. Their insights and expertise have been invaluable. A special word of thanks goes to our project supervisor, Dr. Naglaa Fathy, for her enthusiastic encouragement and valuable feedback during this research. Her guidance has been instrumental in shaping our work. We also wish to acknowledge our parents and the members of Ain Shams University for their kind cooperation and unwavering support throughout this project.

Abstract

School violence defines violent activities that interfere with learning and are detrimental to students, schools, and the community. Youth violence and school violence exposure can result in a variety of unhealthy behaviors and outcomes, such as alcohol and drug abuse and suicide. School violence can lead to depression, anxiety, and many other psychological issues. The impact of school violence on students is significant and might their mental health and overall quality of life. Those who experience or witness violence at school. They are at a higher risk of engaging in substance abuse, including alcohol and drug misuse, as a means to cope with the traumatic experiences they have endured. To address this issue, we proposed a web-based application using machine learning techniques to help parents figure out whether their children are at risk of being violent to their colleagues or not. The web-based application utilizes machine learning algorithms to analyze various factors and indicators associated with potentially violent behavior. These factors may include a student's behavioral history, social interactions, academic performance, and emotional well-being. By considering these variables, the application generates an assessment of the child's risk of engaging in violent acts within the school environment.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ML	Machine Learning
NLP	Natural Language Processing
LMT	Logistic Model Trees
SVM	Support Vector Machine
CCHMC	Cincinnati Children’s Hospital Medical Center

Chapter 1 : Introduction

1.1 Introduction

This chapter will delve into several key components of the proposed project, encompassing the problem definition, motivation, objectives, and methodology, as well as the time plan and thesis outline. Each aspect will be thoroughly explored to provide a comprehensive understanding of the project's scope and direction.

1.2 Problem Definition

It is of utmost importance to establish a highly effective approach aimed at identifying children who are at risk for exhibiting violent behavior. The reason for this urgency lies in the fact that experiencing violence during childhood leaves a profound and enduring impact on an individual's overall health and well-being throughout their entire life. As children are particularly vulnerable, they face the highest risk of suffering from psychological distress or engaging in harmful behaviors that can adversely affect others.

1.3 Motivation

Youth violence and school violence exposure can result in a variety of unhealthy behaviors and outcomes, such as alcohol and drug abuse and suicide. School violence can lead to depression, anxiety, and many other psychological issues, including fear[1].

In Washington, there were a total of 93 violent deaths related to schools between June 2021 and June 2022. Students, teachers, and non-students were among those who perished in these school-related accidents. 12 homicide victims and 8 suicide victims (or 42%) of these fatalities were between the ages of 5 and 18. Students between the ages of 12 and 18 experienced more violent victimization in school in 2019 than outside of it [2]. The effects of school violence are extensive, affecting both students and staff as well as the entire student body. Youth who attend the most violent schools have worse academic achievement, lower school attendance, and higher dropout rates, according to studies [3]. Moreover, it demonstrates that cyberbullying in public schools increased to 16% in 2019–20 from 8% in 2009–2010[4].

Therefore, by implementing a robust system to identify at-risk children and provide appropriate interventions, we can mitigate the long-term consequences of childhood violence and promote healthier outcomes for both individuals and society as a whole.

1.4 Objectives

Developing machine learning techniques to automate the risk assessment process for school violence to:

- Reduce school violence.
- Maintain safety from school violence.
- Help parents figure out whether their children are at risk of being violent to their colleagues or not.

1.5 Methodology

To evaluate the risk of violence in adolescents, a thorough questionnaire was developed based on the research titled "Assessing the Risk of Violence in Adolescents in the Pediatric Emergency Department" [5]. The questionnaire encompassed various risk factors identified in the study. Data for this research was collected from two primary sources: social media platforms and direct surveys conducted with students from specific schools. To analyze the collected dataset, three widely employed machine learning algorithms, namely logistic regression, random forest, and support vector machine (SVM), were utilized. These algorithms were applied to the collected dataset to predict and assess the risk of violence in adolescents.

1.6 Time plan

To develop this project, a schedule was established utilizing an online Gantt chart tool called Office Timeline. The timeline, as depicted in Figure 1.1, was created to outline the key dates and milestones for the project's execution.

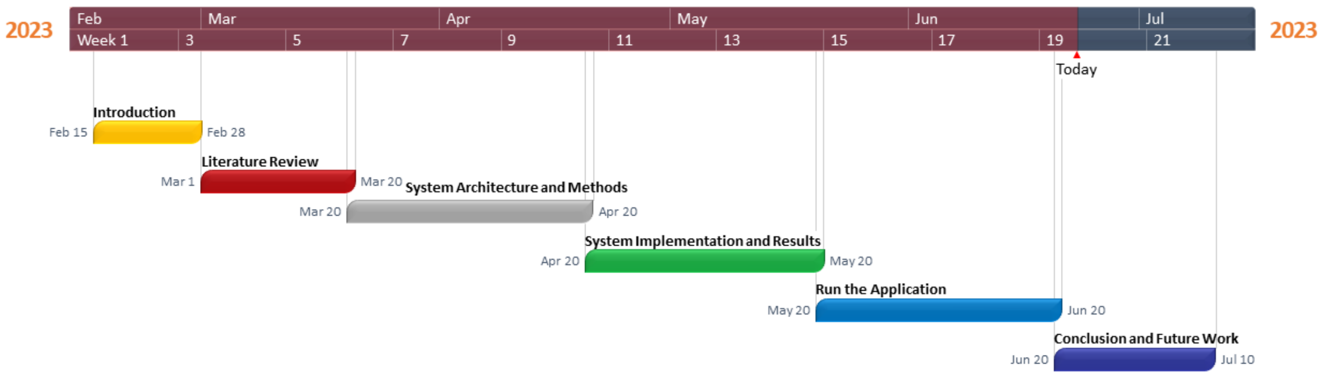


Figure 1:1 Time plan

1.7 Thesis Outline

This project consists of six chapters in addition to appendixes. These chapters are designed to represent the scientific steps we've taken toward our core goal. The following paragraphs provide a summary of the contents of each chapter:

Chapter 1: Introduction

This chapter provides an introduction to the project, encompassing the project overview, problem definition, motivation, objectives, methodology, and time plan.

Chapter 2: Literature Review

This chapter offers an exploration of the theoretical background, as well as a review of previous studies and works relevant to the project. It serves to establish a foundation of knowledge and understanding within the field, providing context for the current study and highlighting the existing research and contributions made by others.

Chapter 3: System Architecture and Methods

This chapter encompasses the system architecture, methods, and procedures employed in the project. It presents a detailed description and visualization of the system architecture.

Chapter 4: System Implementation and Results

This chapter includes the dataset used in the project, the software tools employed, the setup configuration, and a comprehensive account of the experiments conducted along with their corresponding results.

While chapter 5: Run the Application.

This chapter provides a step-by-step guide on how to run the web application.

Finally, in Chapter 6: Conclusion and Future Work.

The conclusion and future work section of this report serve as a reflection on the project's outcomes, offering insights into the project's significance and potential for further improvement and expansion.

Chapter 2 : Literature Review

2.1 Introduction

In this chapter, we will conduct a literature review to compare, analyze, explore, and comprehend the various efforts and directions taken to identify research gaps. Throughout this review, we aim to shed light on the future potential and scope of our project.

2.2 Theoretical Background

This section explains some of the important concepts such as school violence, Types of school violence, and the impact of school violence on students.

2.2.1 School Violence

School violence refers to any form of physical, verbal, or psychological aggression, harassment, or harm that occurs within a school setting. It involves acts of violence or aggressive behavior committed by students or directed toward students, teachers, or other school staff[6].

School violence can have severe consequences for the well-being, academic performance, and psychological development of the individuals involved. It creates an unsafe and disruptive learning environment, affecting not only the victims but also the overall school community. Preventing and addressing school violence requires a comprehensive approach involving education, awareness, support systems, and interventions to promote a safe and inclusive school environment for all students[6].

2.2.2 Types of school violence

School violence can manifest in various forms. Here are some common types of school violence:

1. Physical Violence: This includes physical acts of aggression, such as fights, assaults, or physical attacks on students, teachers, or staff members. It may involve the use of weapons or other objects to cause harm[7].

2. Bullying: Bullying is a persistent form of aggression that involves the repeated harassment, intimidation, or mistreatment of an individual by one or more students. It can be physical, verbal, or psychological and often occurs over an extended period[7].

3. Verbal Violence: Verbal violence refers to the use of abusive language, threats, insults, or derogatory remarks targeting individuals within the school environment. It can lead to emotional distress and can be part of a broader pattern of bullying[7].

4. Cyberbullying: With the proliferation of technology, cyberbullying has become a significant concern. It involves the use of electronic communication platforms, such as social media, online forums, or text messages, to harass, intimidate, or humiliate others[7].

5. Sexual Violence: This encompasses any form of unwanted sexual behavior or harassment within the school setting. It can include sexual assault, inappropriate touching, sexual coercion, or verbal sexual harassment[7].

6. Gang Violence: In some cases, schools may experience violence related to gang activities. Gang-related violence can involve conflicts between different gangs or individuals associated with gangs, leading to physical altercations or threats within the school environment[7].

7. Weapon-related Violence: This refers to incidents involving the possession or use of weapons within the school premises. It poses a significant threat to the safety and well-being of students, teachers, and staff members[7].

8. Hate Crimes: Hate crimes involve acts of violence or harassment motivated by bias or prejudice against a particular race, religion, ethnicity, gender, sexual orientation, or other characteristics. Hate crimes can occur within schools and have a detrimental impact on targeted individuals and the overall school climate[7].

2.2.3 The Impact of School Violence on Students

School violence can have a profound and long-lasting impact on students. Here are some key ways in which school violence can affect students:

1. **Physical Injuries:** Students who are victims of physical violence may suffer from physical injuries, ranging from minor bruises to more severe harm. These injuries can have both immediate and long-term consequences on their health and well-being[8].

2. Emotional and Psychological Effects: School violence often leads to emotional and psychological trauma for the victims. They may experience fear, anxiety, depression, and post-traumatic stress disorder (PTSD). These psychological effects can significantly impact their academic performance, social interactions, and overall mental health[8].

3. Academic Performance: The presence of violence within the school environment can disrupt students' concentration, focus, and ability to learn. Victims may experience difficulties in concentrating on their studies, leading to a decline in academic performance and achievement[8].

4. School Attendance and Engagement: Students who have experienced or witnessed school violence may become reluctant to attend school due to fear of further incidents. They may feel unsafe and avoid participating in school activities, leading to disengagement and a decline in their educational experience[8].

5. Social Relationships and Peer Interactions: School violence can damage students' social relationships and interactions. Victims may face social exclusion, isolation, or stigmatization, making it challenging for them to form trusting relationships with peers. This can lead to feelings of loneliness and alienation[8].

6. Long-term Effects: The impact of school violence can extend beyond the immediate period. Students who have experienced violence may carry the emotional and psychological scars into adulthood. They may have difficulties forming healthy relationships, trust issues, and an increased risk of experiencing violence later in life[8].

7. Mental Health Consequences: School violence is strongly associated with mental health problems among students. Victims may develop anxiety disorders, depression, low self-esteem, and other mental health issues. It is crucial to provide appropriate support and intervention to address these mental health consequences[8].

2.3 Related Work

In this section, we compare systems that are similar to our project and describe the tools, processes, and results of each system.

In [9], authors employed various machine learning algorithms to predict risk levels in an inpatient forensic psychiatry setting. Specifically, Seven machine learning methods Bagging, J48, Jrip, Logistic Model Trees (LMT), Logistic Regression, Linear Regression, and Support Vector Machine (SVM) were used to identify the combination of dictionaries and algorithms that best-predicted risk assessment scores. These algorithms were applied in conjunction with natural language processing techniques to analyze electronic mental health records. The dataset used in the study consisted of electronic mental health records from an inpatient forensic psychiatry setting. The records contained textual information, such as clinical notes and assessments. The model's ensemble approach can predict risk in an inpatient forensic psychiatry setting with an accuracy between 68% and 75%.

In[10], authors implemented a natural language processing (NLP) pipeline in previous research to extract information from clinical narratives. In this pipeline, the transcribed interviews underwent tokenization and lemmatization, with the removal of punctuation marks. Four machine learning methods, namely Logistic Model Trees (LMT), Logistic Regression, Linear Regression, and Support Vector Machine (SVM) were employed to identify the optimal combination of dictionaries and algorithms for accurately predicting risk assessment scores. The dataset was collected from a Psychiatric Response Center (PIRC) questionnaire that collects basic information including the subject's personality, school, and social and family dynamics. The model's ensemble approach can detect the risk of school violence with an accuracy between 90% and 94%.

In[11], the authors focus on conducting a pilot study to develop an automated risk assessment system for school violence. they used Logistic Model Trees (LMT), Logistic Regression, Linear Regression, and Support Vector Machine (SVM) was used to find the set of dictionaries and algorithms that most accurately predicted risk assessment scores using statistical analysis and four machine learning techniques. The dataset was collected from Cincinnati Children’s Hospital Medical Center (CCHMC) from psychiatry outpatient clinics, the inpatient units, and the emergency department. Participants (ages 12–18) were active students in 74 traditional schools (i.e. non-online education). The model's ensemble approach can detect the risk of school violence with an accuracy between 88% and 91%.

In our proposed system, we focused on creating a web-based application using machine learning techniques to help parents figure out whether their children are at risk of being violent to their colleagues or not. The web-based application utilizes machine learning algorithms to analyze various factors and indicators associated with potentially violent behavior. Three widely employed machine learning algorithms, namely logistic regression, random forest, and support vector machine (SVM), were utilized. These algorithms were applied to the collected dataset to predict and assess the risk of violence in adolescents. To evaluate the risk of violence in adolescents, a thorough questionnaire was developed based on the research titled "Assessing the Risk of Violence in Adolescents in the Pediatric Emergency Department". The questionnaire encompassed various risk factors identified in the study. Data for this research was collected from two primary sources: social media platforms and direct surveys conducted with students from specific schools. The accuracy results of different algorithms for automated risk assessment of school violence. The Random Forest algorithm achieved an accuracy of 93%, while Support Vector Machine (SVM) achieved 95% accuracy, and Logistic Regression achieved the highest accuracy of 97.5%. These accuracy measures indicate the performance of the respective algorithms in accurately predicting and assessing the risk of school violence.

Chapter 3 : System Architecture and Methods

3.1 Introduction

This chapter will mainly focus on system analysis and design, covering various aspects such as UML diagrams, system architecture, system users, functional and non-functional requirements, and database schema. Additionally, we will delve into the methods and procedures utilized in our project.

3.2 System Analysis and Design

3.2.1 System Architecture

Our proposed system comprises three modules, as depicted in Figure 3.1. The initial module encompasses the user interface, where children submit their questionnaires, and parents receive the risk assessment results for their children. The second module is the application layer, which takes the questionnaire results as input, processes them and employs machine learning techniques to assess the risk of violence for children. The assessed results are then presented as output to the parents. The third module involves collecting questionnaires to build a model that learns from them.

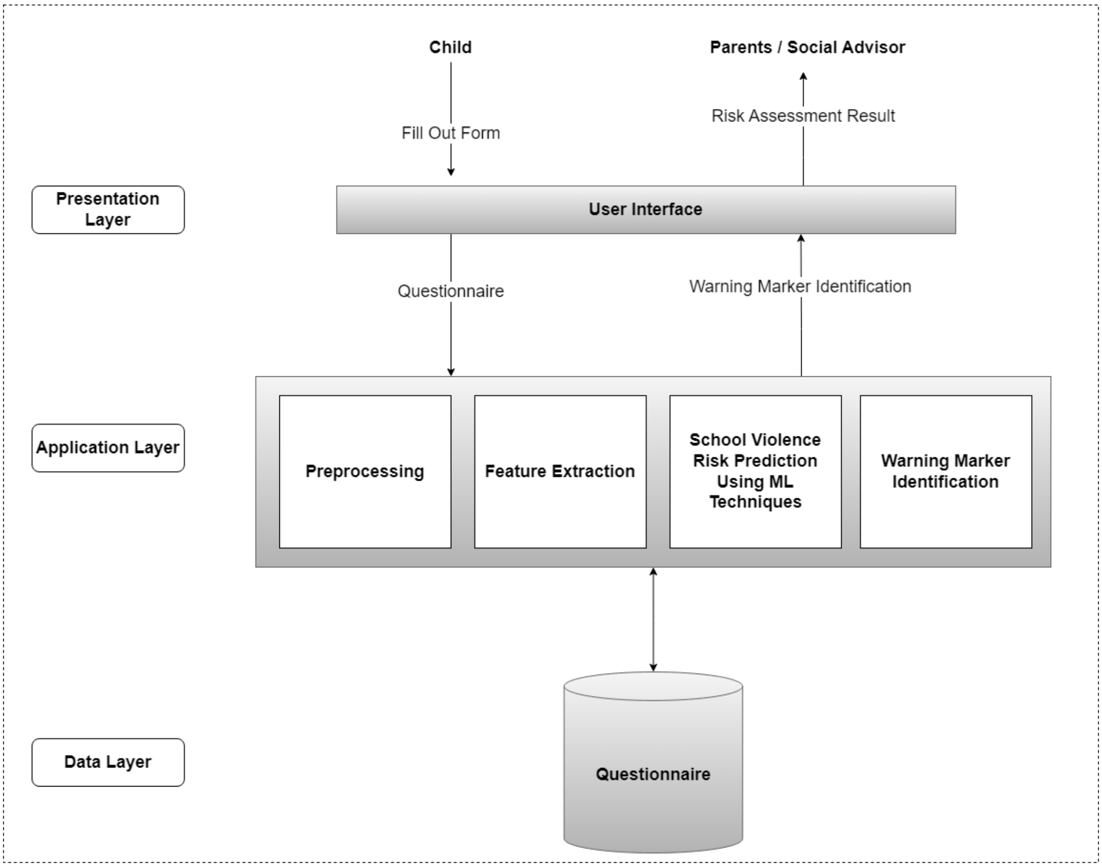


Figure 3:1 System Architecture

3.2.2 Functional Requirements

The proposed system provides functionalities mainly for two actors: children and parents, in addition to automated functions carried out by the AI model.

1) **Child Functionality:**
    - The child must be able to register on the website.
    - The child must be able to log in and log out of the website.
    - The child must be able to fill out the questionnaire.
2) **The parent or Social Advisor Functionality:**
    - The parent or social advisor must be able to register on the website.
    - The parent or social advisor must be able to log in and log out of the website.
    - The parent or social advisor must be able to receive notifications about the risk level of violence for their children.
    - The parent or social advisor must be able to view the dashboard of risk assessment of school violence
    - The parent or social advisor must be able to display tips to avoid the high risk of violence
3) **The AI Model Functionality:**
    - The system shall allow the AI system to detect the risk level of violence.

3.2.3 Non-Functional Requirements

Non-functional requirements are the characteristics and features that must be available in the system.

1) **Usability**: The destinations must be clear to the users and there is no ambiguity in them. There are explanatory texts for the operation of each of the system buttons.
2) **Efficiency**: The system must be efficient, work well, and serve the largest number of users
3) **Security**: The data in the system must have a high degree of protection against penetration or loss.
4) **Reliability**: The system must operate continuously and reliably and prevent incorrect data from being entered.

5) **Localization**: The software will be in Arabic and English so that the user can better understand the information found within the system.

3.2.4 System Users

A. Intended Users:

- **Child**: It is the child who will answer the questionnaire to find out whether he has violence towards his colleagues or not.
- **Parent**: They are parents who want to know the level of risk of violence for their children.
- **Social Advisor:** He is the one who wants to know the level of risk of violence for the child before deciding on their enrollment in school.

B. User Characteristics:

- To complete the questionnaire, children must fall within the age range of 11 to 14.

3.2.5 UML Diagrams

3.2.5.1 Use Case Diagram

The key requirement for a use case diagram is "how a system interacts with its environment involves a diagram and a description to illustrate the discrete activities that the users conduct," which can be created based on the functional requirements. as depicted in figure (3.2):

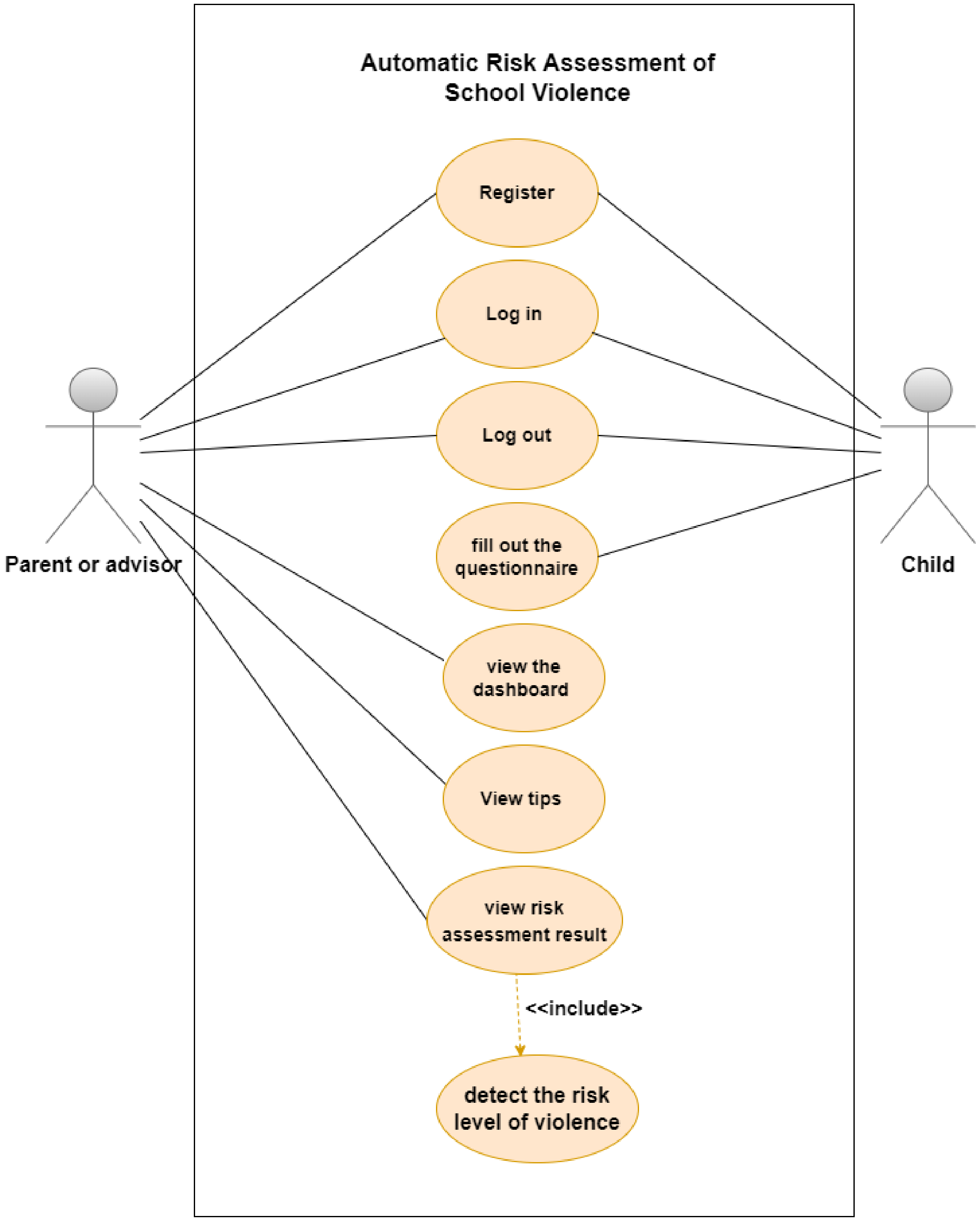


Figure 3:2 Use Case Diagram

3.2.5.2 Class Diagram

A class diagram in software engineering is a form of a static structure diagram that displays the classes, their characteristics, actions (or methods), and relationships between objects to illustrate the structure of a system. As depicted in Figure (3.3).

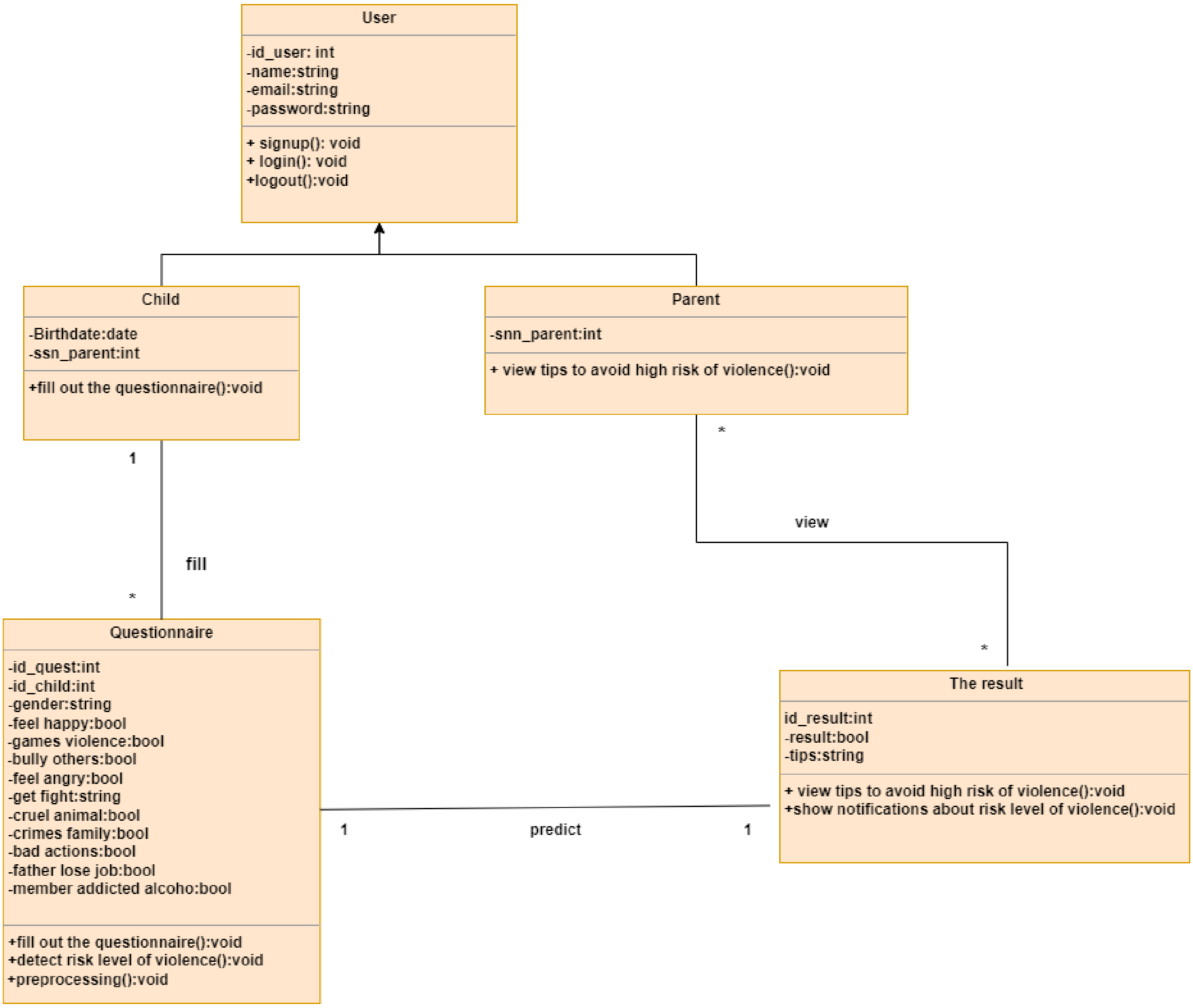


Figure 3:3 Class Diagram

3.2.5.3 Sequence Diagrams

Developers frequently use sequence diagrams to model the interactions between items in a single use case. They demonstrate the interactions that take place when a specific use case is executed and the order in which various system components interact with one another to perform a function. As depicted in Figure (3.4),(3,5).

1. **Sequence Diagram for Child**

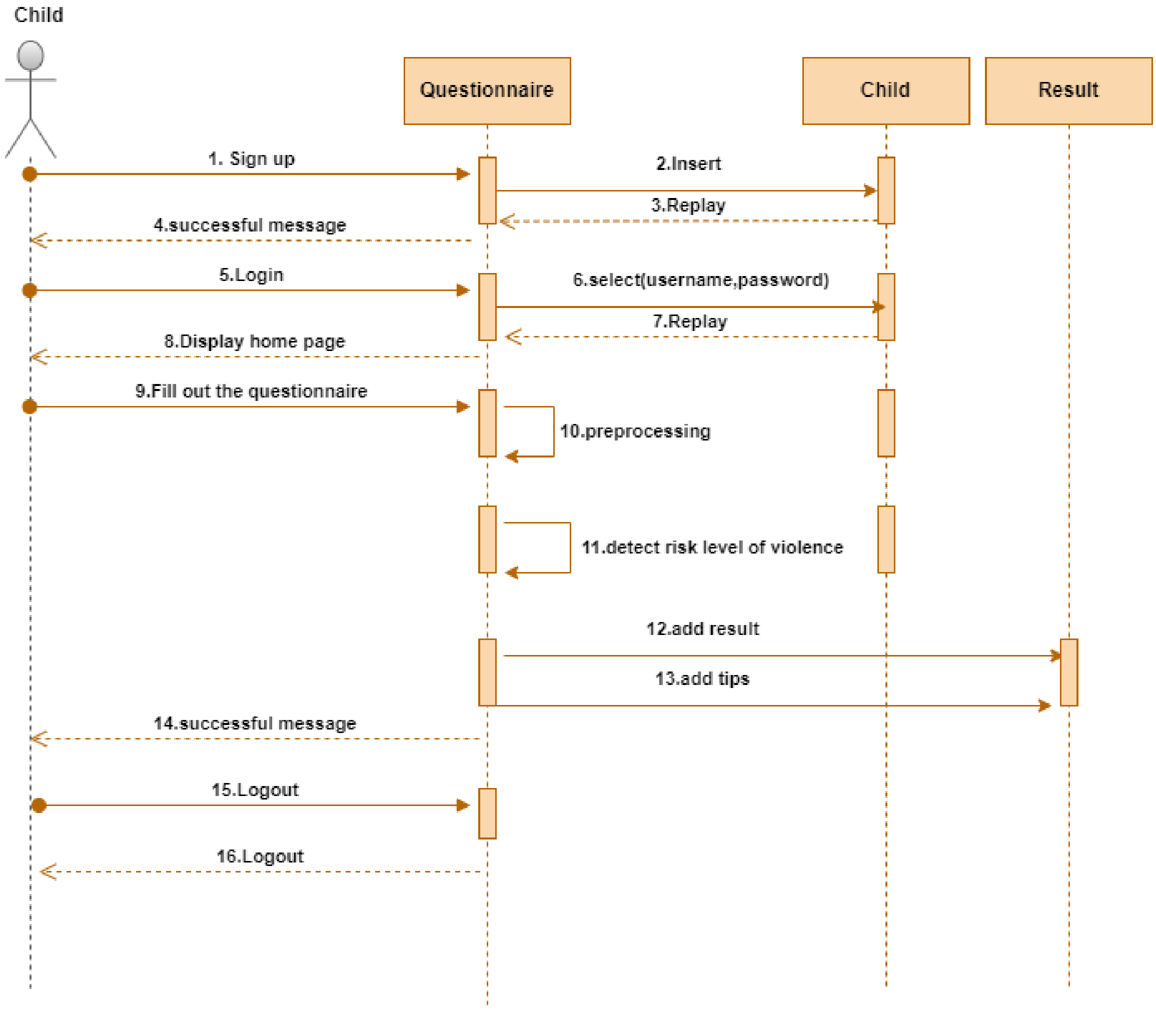


Figure 3:4 Sequence Diagram for Child

2. **Sequence Diagram for Parent or Social Advisor**

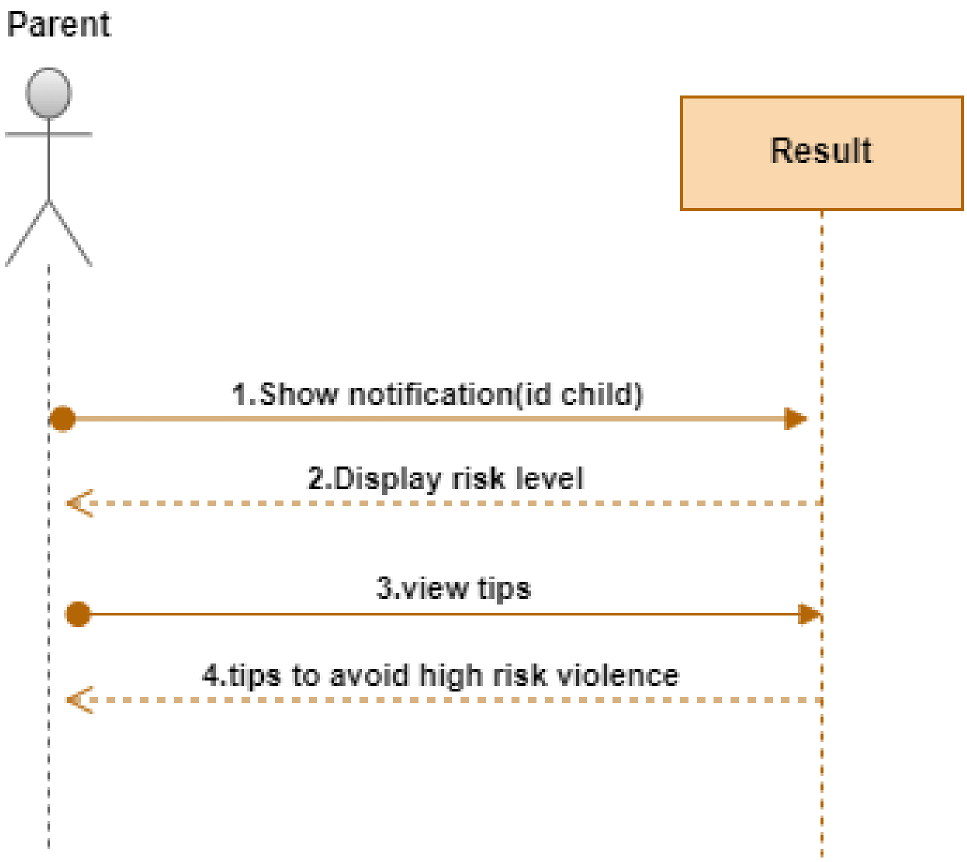


Figure 3:5 Sequence Diagram for Parent or Social Advisor

3.2.5.4 Database Diagram

Entity Relationship (ER) diagram show the connections between entity sets that are kept in databases. In other words, ER diagrams assist in describing how databases' log structures work as depicted in Figure (3.6).

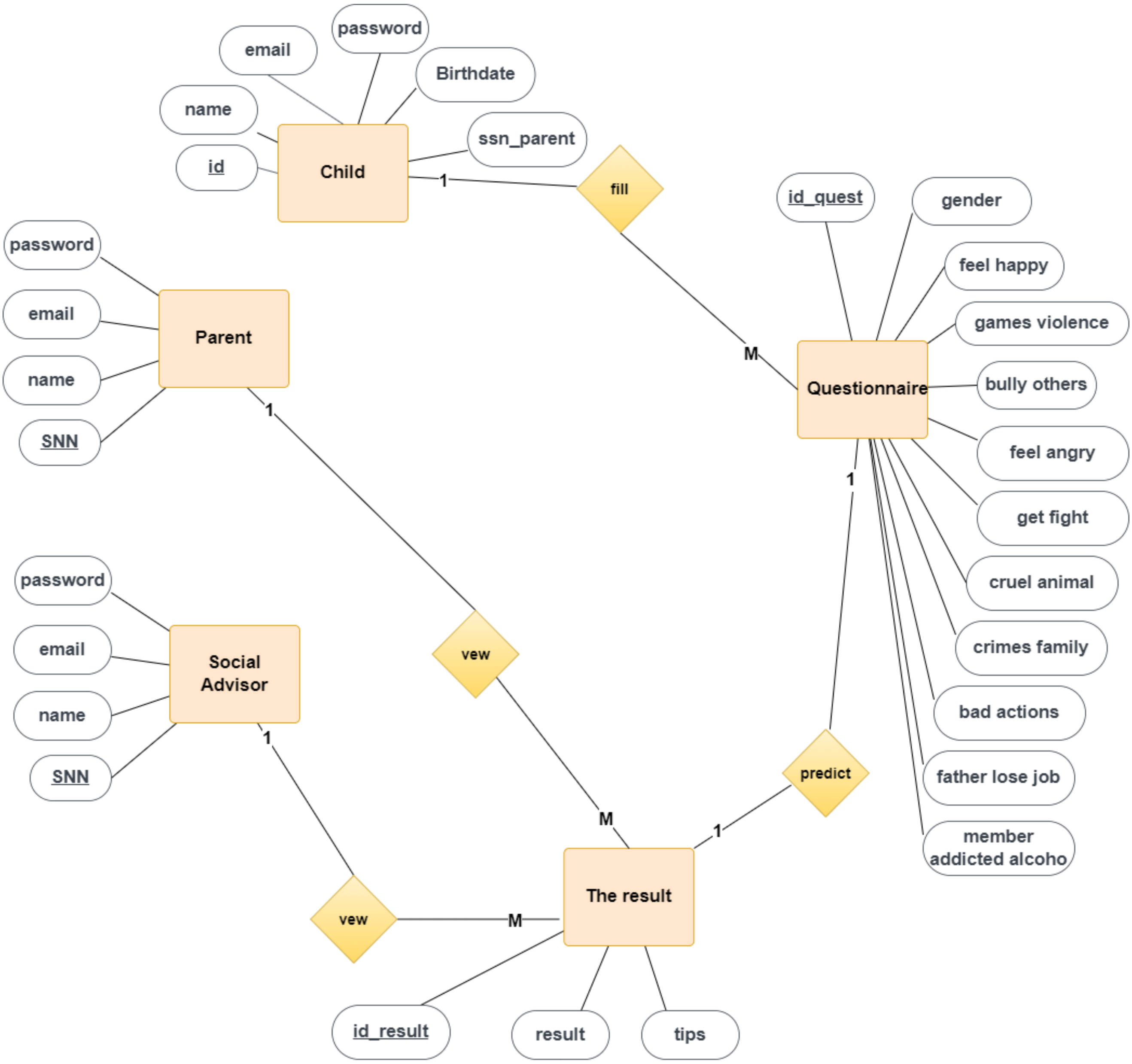


Figure 3:6 Entity Relationship (ER) diagram

3.3 Methods and Procedures Used

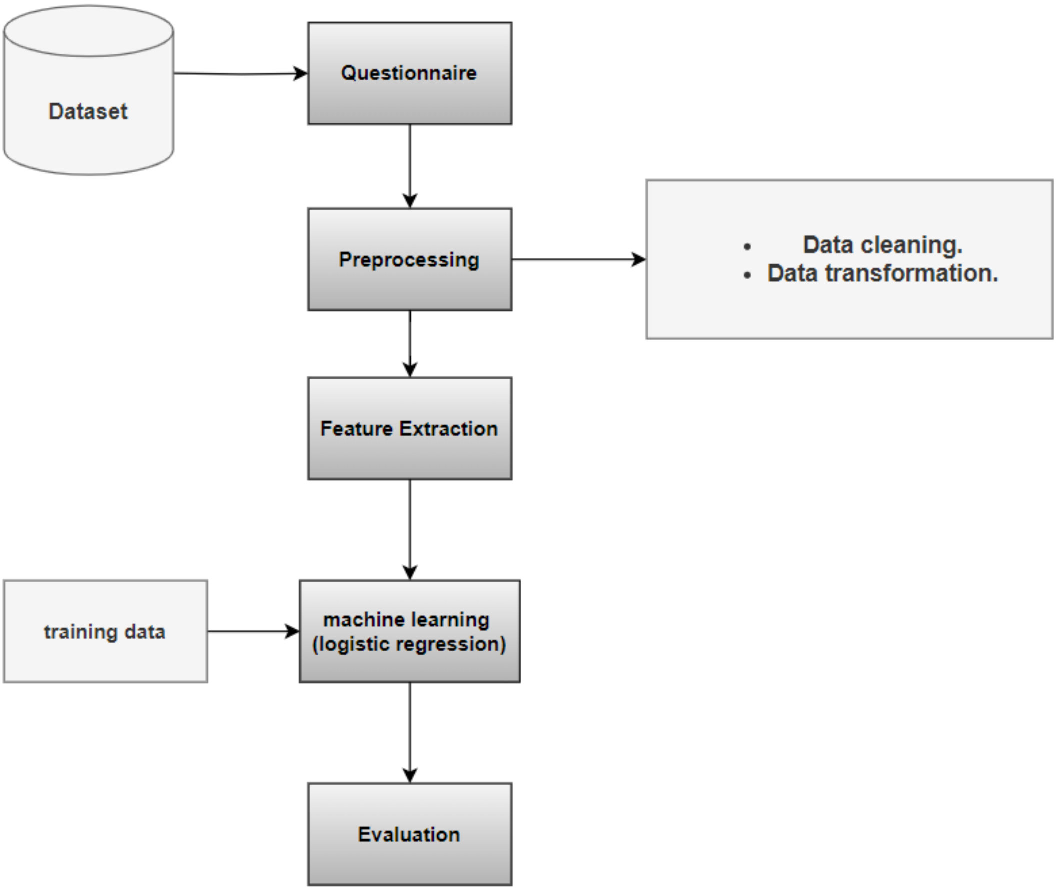


Figure 3:7 Methods and Procedures Used

1) Data Collection & Preparation

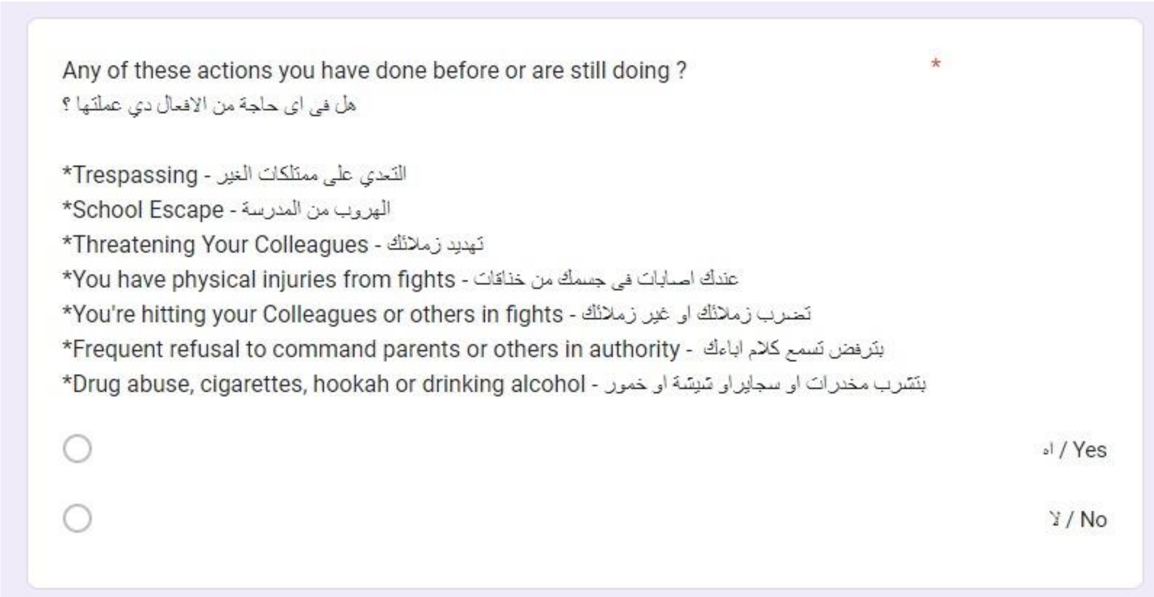
- **Creating Questionnaire**

To assess the risks of violence in adolescents, a comprehensive questionnaire was designed, incorporating the identified risk factors from the study titled "Assessing the Risks of Violence in Adolescents in the Pediatric Emergency Department" [5]. To ensure the validity and relevance of the included risk factors, an interview was conducted with Dr. Tarek at the Psychiatric Health Resort at Ain Shams University. During the interview, specific attention was given to assigning appropriate weights to each question in the questionnaire. The collaboration with Dr. Tarek not only confirmed the accuracy of the collected risk factors but also provided valuable insights into their significance. This collaborative approach enhanced the overall robustness and credibility of the research findings.

- **Collecting Data**

Data collection for this study involved two primary sources: social media and direct collection from students attending specific schools. Social media platforms were utilized to gather relevant data, while additional information was obtained through visits to Gamal Abdel Nasser Elementary School and Ahmed Shawky Preparatory School, both of which are under the administration of East Shubra El-Kheima. These schools were chosen as key locations for collecting data directly from the students.

- **Sample From Created Questionnaire**



Any of these actions you have done before or are still doing ? *
هل في اى حاجة من الافعال دي عملتها ؟

*Trespassing - التعدي على ممتلكات الغير
*School Escape - الهروب من المدرسة
*Threatening Your Colleagues - تهديد زملائك
*You have physical injuries from fights - عندك اصابات في جسمك من خناقات
*You're hitting your Colleagues or others in fights - تضرب زملائك او غير زملائك
*Frequent refusal to command parents or others in authority - بترفض تسمع كلام اباءك
*Drug abuse, cigarettes, hookah or drinking alcohol - بتشرب مخدرات او سجايراو شيشة او خمور

اه / Yes

لا / No

Figure 3:8 Sample 1 From Created Questionnaire

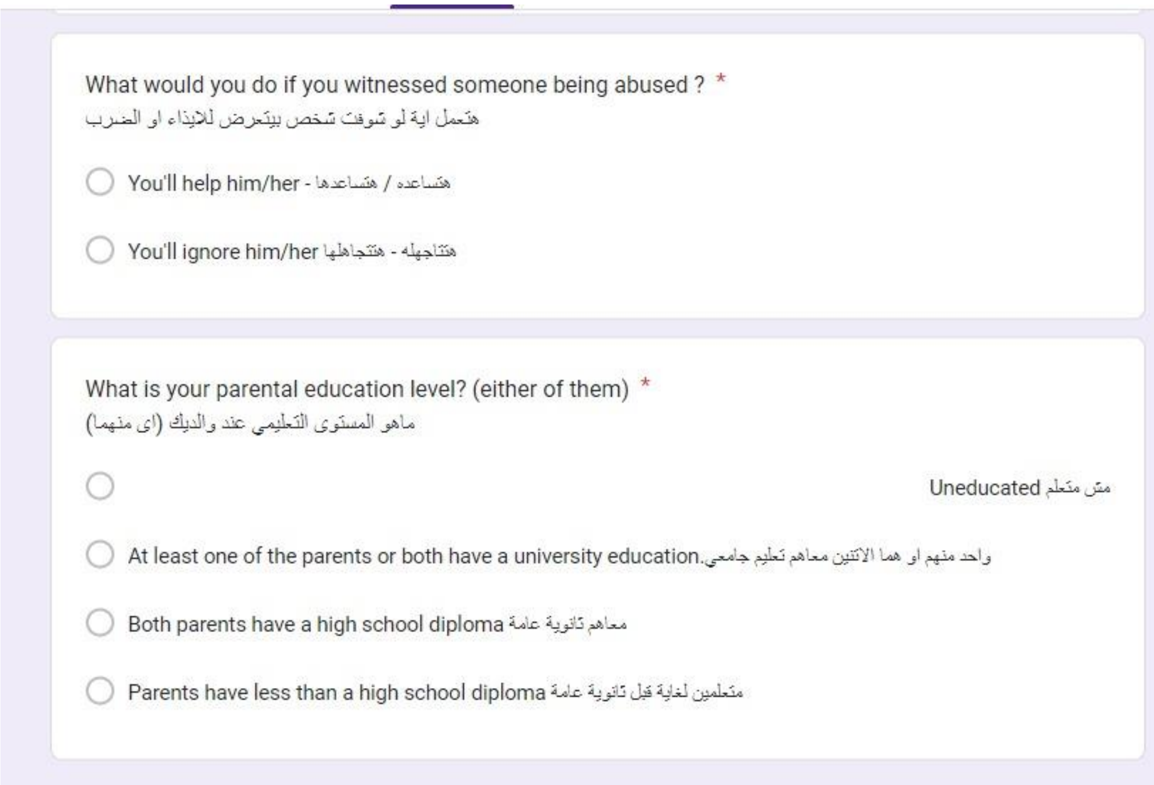


Figure 3:9 Sample 2 From Created Questionnaire

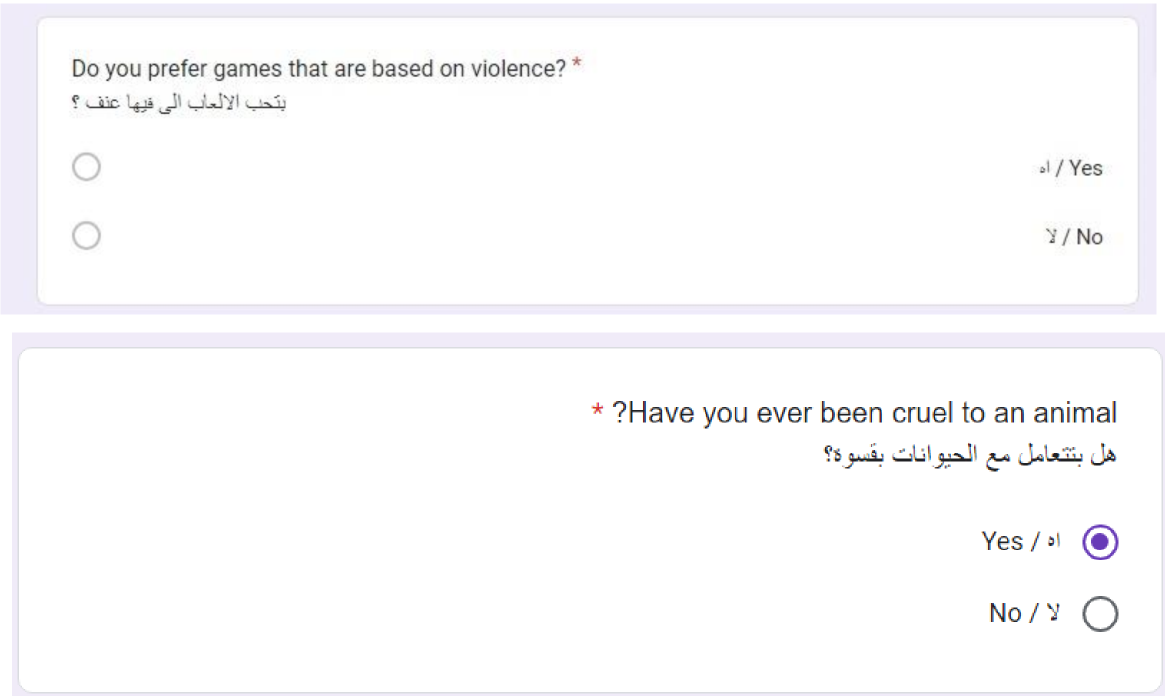


Figure 3:10 Sample 3 From Created Questionnaire

2) Preprocessing

Data Preprocessing includes the steps(Identifying and handling the missing values, Feature scaling, Splitting the dataset, and Encoding the categorical data) we need to follow to transform or encode data so that it may be easily parsed by machine learning.

- **Normal Dataset**

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Genderبنت	Ageكام سنة	Do you fee	Do you pre	Do you ha	Do you bul	Do you fee	When you	Have you e	Are your p	Any of the	What wou	What is yo	Did your fa	Is there a
2	Male / ولد	17-19 year	No / لا	Yes / اه	Yes / اه	No / لا	Yes / اه	Use your p	No / لا	No / لا	No / لا	You'll help	Parents ha	No / لا	Yes / اه
3	Male / ولد	11-13 year	No / لا	No / لا	Yes / اه	No / لا	Yes / اه	Use your p	No / لا	No / لا	Yes / اه	You'll help	At least on	No / لا	No / لا
4	Male / ولد	17-19 year	No / لا	Yes / اه	Yes / اه	Yes / اه	No / لا	Use your p	No / لا	No / لا	No / لا	You'll help	At least on	No / لا	Yes / اه
5	Female /	17-19 year	No / لا	No / لا	Yes / اه	No / لا	No / لا	Use your p	No / لا	No / لا	No / لا	You'll help	Both parer	No / لا	No / لا
6	Female /	17-19 year	Yes / اه	Yes / اه	No / لا	Yes / اه	Yes / اه	Use your p	No / لا	No / لا	Yes / اه	You'll igno	At least on	No / لا	No / لا
7	Male / ولد	17-19 year	No / لا	Yes / اه	Yes / اه	No / لا	No / لا	Use your p	No / لا	No / لا	No / لا	You'll help	Parents ha	No / لا	No / لا
8	Female /	17-19 year	No / لا	No / لا	Yes / اه	No / لا	Yes / اه	Brick or St	No / لا	No / لا	No / لا	You'll help	Parents ha	Yes / اه	Yes / اه
9	Male / ولد	11-13 year	No / لا	No / لا	Yes / اه	No / لا	Yes / اه	Use your p	No / لا	No / لا	No / لا	You'll igno	Both parer	No / لا	No / لا
10	Male / ولد	17-19 year	No / لا	Yes / اه	Yes / اه	No / لا	No / لا	Use your p	No / لا	No / لا	Yes / اه	You'll help	At least on	No / لا	Yes / اه
11	Female /	17-19 year	Yes / اه	Yes / اه	No / لا	Yes / اه	Yes / اه	Use your p	No / لا	No / لا	No / لا	You'll help	At least on	No / لا	No / لا
12	Male / ولد	17-19 year	No / لا	No / لا	Yes / اه	No / لا	No / لا	Use your p	No / لا	No / لا	No / لا	You'll help	At least on	No / لا	No / لا
13	Female /	17-19 year	No / لا	No / لا	Yes / اه	No / لا	Yes / اه	Use your p	No / لا	No / لا	Yes / اه	You'll help	At least on	Yes / اه	No / لا
14	Female /	17-19 year	No / لا	No / لا	Yes / اه	No / لا	No / لا	Use your p	No / لا	No / لا	No / لا	You'll help	At least on	No / لا	No / لا
15	Female /	17-19 year	No / لا	No / لا	Yes / اه	Yes / اه	Yes / اه	Use your p	No / لا	No / لا	Yes / اه	You'll help	At least on	No / لا	No / لا
16	Male / ولد	11-13 year	Yes / اه	Yes / اه	No / لا	Yes / اه	Yes / اه	Brick or St	Yes / اه	Yes / اه	Yes / اه	You'll igno	d مش متعلم	Yes / اه	Yes / اه
17	Female /	17-19 year	No / لا	No / لا	Yes / اه	No / لا	Yes / اه	Use your p	No / لا	No / لا	No / لا	You'll help	Both parer	No / لا	Yes / اه
18	Male / ولد	17-19 year	No / لا	Yes / اه	Yes / اه	No / لا	No / لا	Brick or St	No / لا	No / لا	Yes / اه	You'll help	At least on	No / لا	No / لا
19	Female /	17-19 year	No / لا	No / لا	Yes / اه	No / لا	No / لا	Use your p	No / لا	No / لا	No / لا	You'll help	Both parer	No / لا	No / لا
20	Male / ولد	17-19 year	No / لا	Yes / اه	Yes / اه	No / لا	No / لا	Use your p	No / لا	No / لا	No / لا	You'll help	At least on	No / لا	No / لا
21	Female /	17-19 year	No / لا	Yes / اه	No / لا	Yes / اه	Yes / اه	Use your p	Yes / اه	No / لا	Yes / اه	You'll igno	d مش متعلم	Yes / اه	Yes / اه
22	Male / ولد	14-16 year	No / لا	No / لا	Yes / اه	No / لا	Yes / اه	Use your p	No / لا	No / لا	No / لا	You'll help	At least on	No / لا	No / لا
23	Male / ولد	17-19 year	No / لا	No / لا	Yes / اه	No / لا	No / لا	Use your p	No / لا	No / لا	No / لا	You'll help	d مش متعلم	No / لا	No / لا
24	Female /	17-19 year	No / لا	No / لا	Yes / اه	Yes / اه	No / لا	Use your p	No / لا	No / لا	No / لا	You'll igno	At least on	No / لا	No / لا
25	Male / ولد	11-13 year	No / لا	No / لا	Yes / اه	No / لا	No / لا	Use your p	No / لا	No / لا	No / لا	You'll help	d مش متعلم	No / لا	No / لا
26	Male / ولد	17-19 year	Yes / اه	Yes / اه	No / لا	Yes / اه	Yes / اه	Brick or St	Yes / اه	Yes / اه	Yes / اه	You'll igno	d مش متعلم	Yes / اه	Yes / اه
27	Male / ولد	14-16 year	Yes / اه	Yes / اه	No / لا	Yes / اه	Yes / اه	Brick or St	Yes / اه	Yes / اه	Yes / اه	You'll igno	d مش متعلم	Yes / اه	Yes / اه
28	Female /	11-13 year	Yes / اه	Yes / اه	No / لا	Yes / اه	Yes / اه	Brick or St	Yes / اه	Yes / اه	Yes / اه	You'll igno	d مش متعلم	Yes / اه	Yes / اه
29	Female /	17-19 year	No / لا	No / لا	Yes / اه	No / لا	Yes / اه	Use your p	No / لا	No / لا	No / لا	You'll help	Both parer	No / لا	No / لا

Figure 3:11 Normal Dataset

- **Weighted Dataset**

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	gender	age	feel happy	games viol	emotional	bully other	feel angry	get fight	cruel anim	crimes fan	bad action	witnessed	parental e	father lose	member a	violence prediction	
2	Male	17-19 year	1	2	2	1	2	1	1	1	1	1	1.5	1	2	3	
3	Male	11-13 year	1	1	2	1	2	1	1	1	3	1	0.5	1	1	3	
4	Male	17-19 year	1	2	2	2.5	1	1	1	1	1	1	0.5	1	2	3	
5	Female	17-19 year	1	1	2	1	1	1	1	1	1	1	1	1	1	3	
6	Female	17-19 year	2.5	2	1	2.5	2	1	1	1	3	1.5	0.5	1	1	3	
7	Male	17-19 year	1	2	2	1	1	1	1	1	1	1	1.5	1	1	3	
8	Female	17-19 year	1	1	2	1	2	2	1	1	1	1	1.5	1.5	2	3	
9	Male	11-13 year	1	1	2	1	2	1	1	1	1	1.5	1	1	1	3	
10	Male	17-19 year	1	2	2	1	1	1	1	1	3	1	0.5	1	2	3	
11	Female	17-19 year	2.5	2	1	2.5	2	1	1	1	1	1	0.5	1	1	3	
12	Male	17-19 year	1	1	2	1	1	1	1	1	1	1	0.5	1	1	3	
13	Female	17-19 year	1	1	2	1	2	1	1	1	3	1	0.5	1.5	1	3	
14	Female	17-19 year	1	1	2	1	1	1	1	1	1	1	0.5	1	1	3	
15	Female	17-19 year	1	1	2	2.5	2	1	1	1	3	1	0.5	1	1	3	
16	Male	11-13 year	2.5	2	1	2.5	2	2	3	2	3	1.5	2	1.5	2	3	
17	Female	17-19 year	1	1	2	1	2	1	1	1	1	1	1	1	2	3	
18	Male	17-19 year	1	2	2	1	1	2	1	1	3	1	0.5	1	1	3	
19	Female	17-19 year	1	1	2	1	1	1	1	1	1	1	1	1	1	3	
20	Male	17-19 year	1	2	2	1	1	1	1	1	1	1	0.5	1	1	3	
21	Female	17-19 year	1	2	1	2.5	2	1	3	1	3	1.5	2	1.5	2	3	
22	Male	14-16 year	1	1	2	1	2	1	1	1	1	1	0.5	1	1	3	
23	Male	17-19 year	1	1	2	1	1	1	1	1	1	1	2	1	1	3	
24	Female	17-19 year	1	1	2	2.5	1	1	1	1	1	1.5	0.5	1	1	3	
25	Male	11-13 year	1	1	2	1	1	1	1	1	1	1	2	1	1	3	
26	Male	17-19 year	2.5	2	1	2.5	2	2	3	2	3	1.5	2	1.5	2	3	
27	Male	14-16 year	2.5	2	1	2.5	2	2	3	2	3	1.5	2	1.5	2	3	
28	Female	11-13 year	2.5	2	1	2.5	2	2	3	2	3	1.5	2	1.5	2	3	
29	Female	17-19 year	1	1	2	1	2	1	1	1	1	1	1	1	1	3	

Figure 3:12 Weighted Dataset

3) Feature Extraction

Feature extraction refers to the process of transforming raw data into numerical features that can be processed while preserving the information in the original data set. It yields better results than applying machine learning directly to the raw data.

4) Machine Learning Algorithms

In this study, three machine learning techniques logistic regression, random forest classifiers, and support vector machine(SVM) were employed to facilitate the task of the AI system, which involved predicting output values based on input data. The three techniques utilized were logistic regression is a statistical analysis method to predict a binary outcome, such as yes or no, based on prior observations of a data set. A logistic regression model predicts a dependent data variable by analyzing the relationship between one or more existing[12]. Random forest classifiers can handle thousands of input variables without variable deletion. And It gives estimates of what variables are important in the classification[13]. Support vector machine(SVM) can find the optimal hyperplane to correctly classify between data points of different classes[14]. These algorithms were selected due to their effectiveness in handling various types of data and their ability to generate accurate predictions. By employing these machine learning techniques, the study aimed to leverage their strengths and extract meaningful insights from the dataset.

5) Evaluation

The evaluation of a learning process involves assessing the extent to which knowledge or skills have been acquired and applied, using predefined criteria. One commonly used method for evaluation is the utilization of a confusion matrix. The confusion matrix provides a structured representation of the performance of a learning model by presenting the counts of true positive, true negative, false positive, and false negative predictions. This matrix helps in measuring the accuracy and effectiveness of the learning process by comparing the predicted and actual outcomes. By employing evaluation techniques such as the confusion matrix, the learning process can be objectively assessed and the level of knowledge or skill attained can be determined.

Chapter 4 : System Implementation and Results

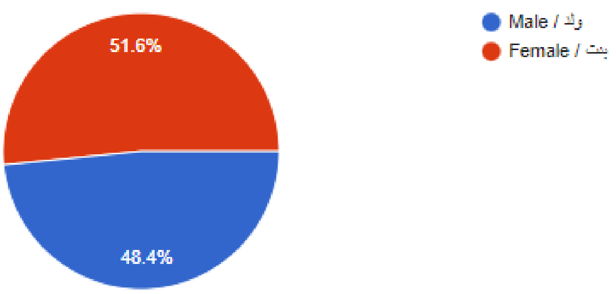
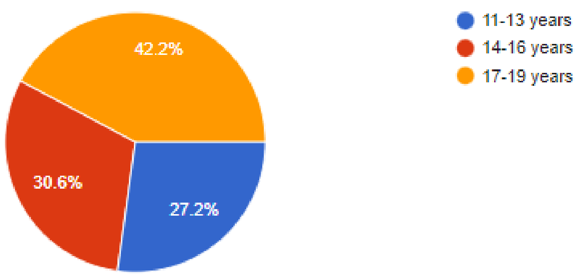
4.1 Introduction

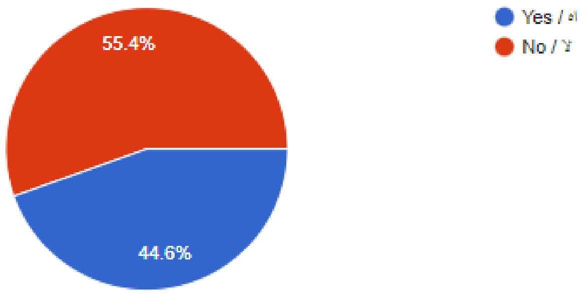
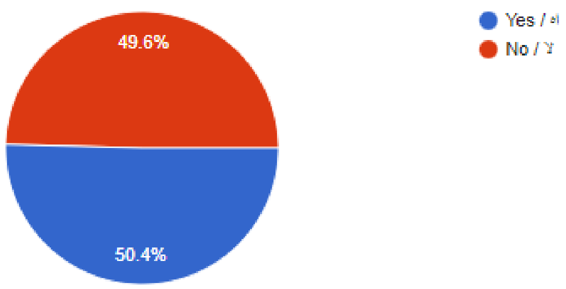
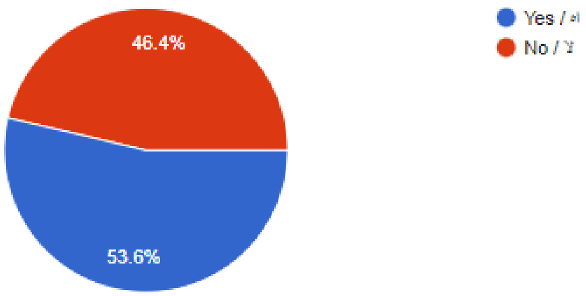
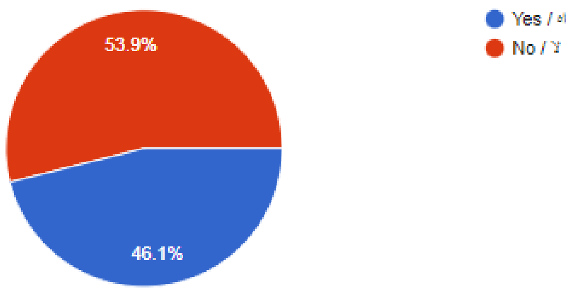
This chapter will cover various aspects of our study, including the dataset used, the software programs utilized, the configuration setup, as well as the experimental procedures, and the obtained results.

4.2 Materials Used

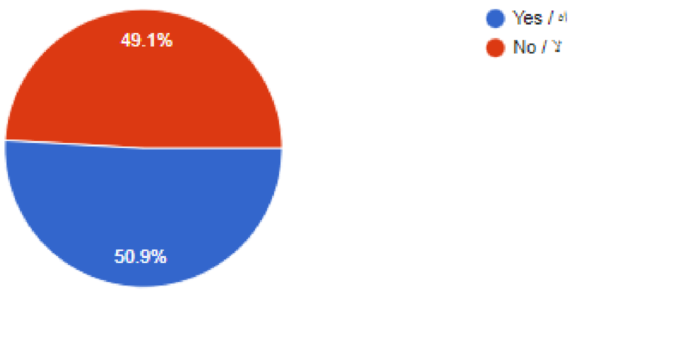
To gather the dataset, we created a questionnaire and distributed it through multiple channels, including social media platforms and primary and preparatory schools. We received a total of 1400 responses from students. The main objective is to collect data from the students to build the model. An online survey was created using a Google Form, and 1,400 people responded. Below are the survey questions and answers:

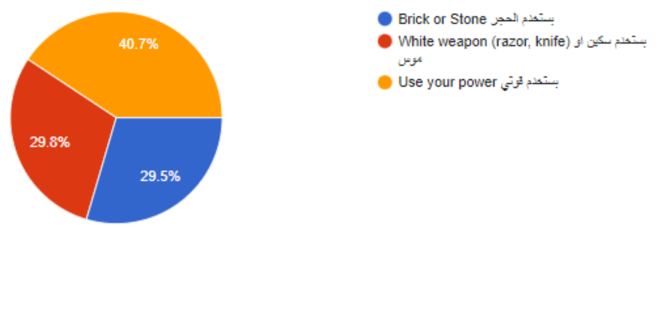
Table 4.1 Requirement gathering technique

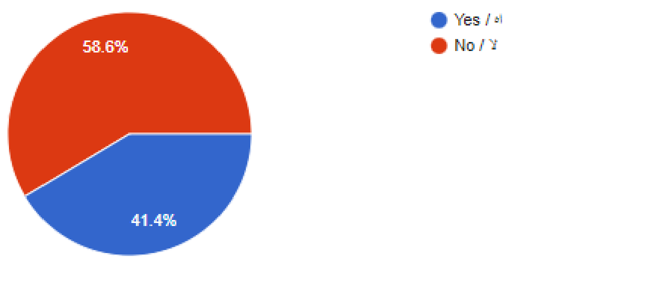
Questions	Percentage
1. Gender	
2. Age	

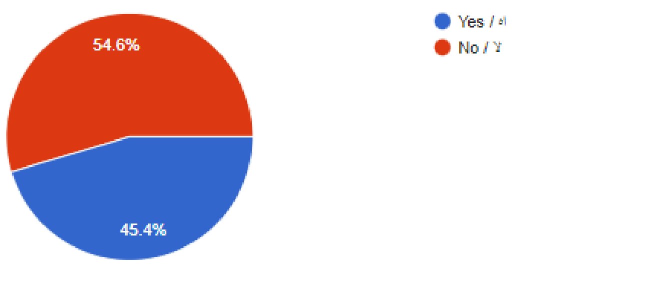
Do you feel happy when others suffer? هل تشعر بالسعادة لما الاخرين بيتألموا ؟	
Do you prefer games that are based on violence? بتحب الالعاب الى فيها عنف ؟	
Do you have severe emotional outbursts? هل بتتاثربالمواقف العاطفية؟	
Do you bully others physically or online? هل بتتنمر علي الناس؟	

<table>
<tr>
<td>Do you feel angry or lose your temper often?

بتبقي غضبان او بتفقد اعصابك باستمرار؟</td>
<td>
</td>
</tr>
<tr>
<td>When you get into a fight, what do you use?

لما بتتخانق بتستخدم ايه</td>
<td>
</td>
</tr>
<tr>
<td>Have you ever been cruel to an animal?

هل بتتعامل مع الحيوانات بقسوة؟</td>
<td>
</td>
</tr>
<tr>
<td>Are your parents wanted by the police or have they committed a crime before?

هل حد من بابا او ماما عمل خطأ غير قانوني ادى الى ان الشرطة تاخدهم القسم ؟</td>
<td>
</td>
</tr>
</table>

<table>
<tr>
<td>Any of these actions you have done before or are still doing?

هل فى اى حاجة من الافعال دي عملتها ؟

*Trespassing - التعدي على ممتلكات الغير

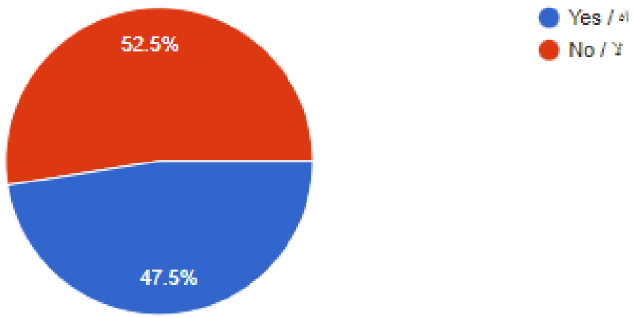
*School Escape - الهروب من المدرسة

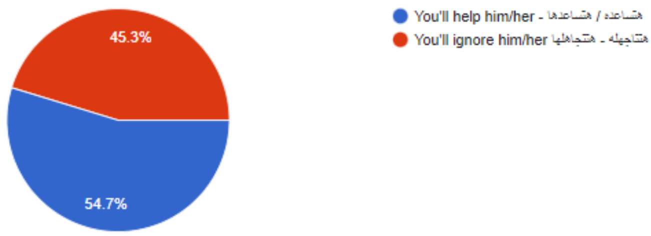
*Threatening Your Colleagues - تهديد زملائك

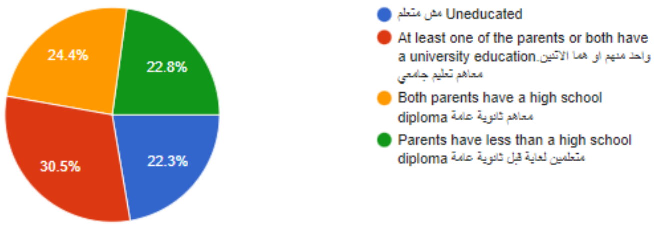
*You have physical injuries from fights - عندك اصابات فى جسمك من خناقات

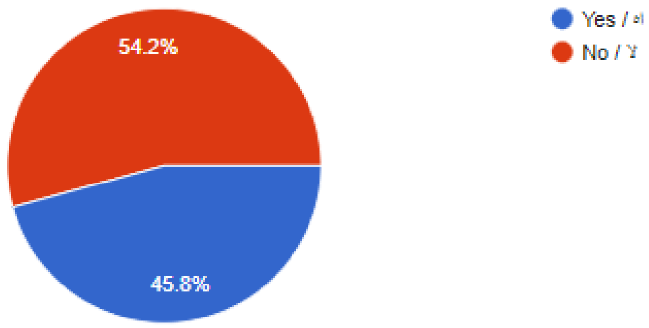
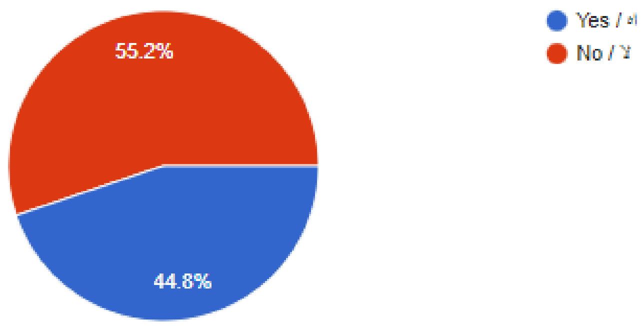
*You're hitting your Colleagues or others in fights - تضرب زملائك او غير زملائك

*Frequent refusal to command parents or others in authority - بترفض تسمع كلام اباءك

*Drug abuse, cigarettes, hookah or drinking alcohol - بتشرب مخدرات او سجايرو او شيشة او خمور</td>
<td>
</td>
</tr>
<tr>
<td>What would you do if you witnessed someone being abused?

هتعمل اية لو شوفت شخص بيتعرض للايذاء او الضرب؟</td>
<td>
</td>
</tr>
<tr>
<td>What is your parental education level? (either of them)

ماهو المستوى التعليمي عند والديك (اى منهما)؟</td>
<td>
</td>
</tr>
</table>

Did your father lose his job sometime for a long time? هل والدك فقد عمله لوقت طويل ؟	
Is there a family member addicted to alcohol or drugs? هل احد من عائلتك مدمن الخمور او المخدرات ؟	

4.3 Software Tools

Software tools used in our proposed system as depicted in Table 4.2

Table 4.2 Software Tools

Tool / Language	Reason for Using This Program
HTML	**HTML:** provides structure and semantic meaning to the content, ensuring proper interpretation and accessibility. The first version of HTML, HTML 1.0, was released in 1993[15].
CSS	**CSS**: allows developers to control the visual presentation and layout of web pages, resulting in consistent and appealing designs. CSS1 was introduced in 1996, and since then, CSS has evolved through different versions, with CSS3 being the latest major release[16].
JavaScript	**JavaScript:** adds interactivity and dynamic functionality to web pages, enhancing the user experience and enabling advanced features. It was initially introduced in 1995 and has undergone significant updates and enhancements over the years, with ECMAScript 6 (ES6) being a major version release[17].

Flask	**Flask:** is a popular Python web framework used for developing web applications. Company and Number of Issues: Flask is an open-source framework, maintained by a community of contributors rather than a specific company. Flask was initially released by Armin Ronacher in 2010[18].
Python	**Python:** is a versatile programming language that serves as a source language for building the backend of websites and developing AI models. By leveraging Python's capabilities in both backend web development and AI model development, developers can build integrated systems where web applications can interact with AI models to provide intelligent and dynamic functionality. Python is an open-source programming language developed and maintained by the Python Software Foundation, and its first release dates back to 1991[19].
Visual Studio Code	**Visual Studio Code (VS Code):** is an open-source source code editor developed by Microsoft. but it is an open-source project with contributions from a large community of developers worldwide. Visual Studio Code was initially released in April 2015. Since then, it has been continuously updated with new features, bug fixes, and performance improvements.VS Code is available for Windows, macOS, and Linux, making it a versatile choice for developers working on different operating systems[20].
Google Colab	**Google Colab:** is commonly used by data scientists, researchers, and developers for building AI models and performing machine learning tasks. It is available for free and provides access to a high-performance GPU and TPU (Tensor Processing Unit). This allows users to leverage powerful hardware resources for training deep learning models and executing computationally intensive tasks[21].
MYSQL	**MySQL:** is an open-source database system, which means it is free to use and has a large community of developers contributing to its development and support. This makes it a cost-effective solution for small to large-scale applications.MySQL was originally developed by MySQL AB, a Swedish company founded by Michael Widenius, David Axmark, and Allan Larsson. In 2008[22].

4.4 Hardware Specifications

Table (4.3) shows the system's hardware requirements.

Table 4.3 Hardware Specifications

Hardware	Specifications
Computer	Processor: Core i5 or higher. RAM: 8 GB. Hard Disk 256 GB
Third-party server	The web application will be tested on a local server before being hosted by a Third-party server.

4.5 Experimental and Results

After splitting the dataset into 70% for training and 30% for testing, Table (4.4) presents the accuracy results for logistic regression, random forest, and support vector machine. Both logistic regression and support vector machine achieved impressive performances, with their outcomes being relatively similar. Notably, they outperformed random forest in terms of accuracy. Logistic regression was chosen for classifying the risk assessment of violence.

Table 4.4 Experimental and Results 1 for accuracy for each algorithm

Model Name	Accuracy
Logistic Regression	0.9708
Random Forest	0.9250
Support Vector Machine	0.9667

Following the division of the dataset into 70% for training and 30% for testing, Table (4.5) presents the precision, recall, and f1-score results specifically for the low-risk category. This table provides insights into the performance of logistic regression, random forest, and support vector machine algorithms in accurately predicting low-risk outcomes. The precision, recall, and f1-score metrics were calculated to assess the effectiveness of each algorithm in this particular category.

Table 4.5 Experimental and results 1 for the precision, recall, and f1-score specifically for the low-risk category

Model Name	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.9708	0.97	1.00	0.98
Random Forest	0.9250	0.92	0.99	0.96
Support Vector Machine	0.9667	0.96	1.00	0.98

Following the division of the dataset into 70% for training and 30% for testing, Table (4.6) presents the precision, recall, and f1-score results specifically for the high-risk category. This table provides insights into the performance of logistic regression, random forest, and support vector machine algorithms in accurately predicting high-risk outcomes. The precision, recall, and f1-score metrics were calculated to assess the effectiveness of each algorithm in this particular category.

Table 4.6 Experimental and results 1 for the precision, recall, and f1-score specifically for the high-risk category

Model Name	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.9708	1.00	0.83	0.91
Random Forest	0.9250	0.93	0.61	0.74
Support Vector Machine	0.9667	1.00	0.80	0.89

After splitting the dataset into 50% for training and 50% for testing, Table (4.7) presents the accuracy results for logistic regression, random forest, and support vector machine. Both logistic regression and support vector machine achieved impressive performances, with their outcomes being relatively similar. Notably, they outperformed random forest in terms of accuracy. Logistic regression was chosen for classifying the risk assessment of violence.

Table 4.7 Experimental and Results 2 for accuracy for each algorithm

Model Name	**Accuracy**
Logistic Regression	0.975
Random Forest	0.9375
Support Vector Machine	0.9550

Following the division of the dataset into 50% for training and 50% for testing, Table (4.8) presents the precision, recall, and f1-score results specifically for the low-risk category. This table provides insights into the performance of logistic regression, random forest, and support vector machine algorithms in accurately predicting low-risk outcomes. The precision, recall, and f1-score metrics were calculated to assess the effectiveness of each algorithm in this particular category.

Table 4.8 Experimental and results 2 for the precision, recall, and f1-score specifically for the low-risk category

Model Name	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Logistic Regression	0.975	0.97	1.00	0.99
Random Forest	0.9375	0.90	0.63	0.74
Support Vector Machine	0.9550	0.96	0.99	0.97

Following the division of the dataset into 50% for training and 50% for testing, Table (4.9) presents the precision, recall, and f1-score results specifically for the high-risk category. This table provides insights into the performance of logistic regression, random forest, and support vector machine algorithms in accurately predicting high-risk outcomes. The precision, recall, and f1-score metrics were calculated to assess the effectiveness of each algorithm in this particular category.

Table 4.9 Experimental and results 2 for the precision, recall, and f1-score specifically for the high-risk category

Model Name	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Logistic Regression	0.975	0.98	0.84	0.91
Random Forest	0.9375	0.94	0.99	0.96
Support Vector Machine	0.9550	0.93	0.74	0.82

A comparison between the works reviewed in the previous section and our model is given in Table (4.10). It shows the characteristics of the models where our model has an advantage over other similar works.

Table 4.10 A comparison table showing the difference between our system and other systems

	Our Proposed System	**Risk prediction using natural language processing of electronic mental health records in an inpatient forensic psychiatry setting**	**Finding warning markers: Leveraging natural language processing and machine learning technologies to Detect the Risk of school violence**	**Automated Risk Assessment for School Violence: a Pilot Study**
Accuracy	(93.7% - 97.5%)	(68% - 75%)	(90% - 94%)	(88% - 91%)
Machine Learning	Logistic regression Random Forest Support Vector Machine (SVM)	Logistic Model Trees (LMT) Logistic Regression Linear Regression Support Vector Machine (SVM)	Logistic Model Trees (LMT) Linear Regression	Support Vector Machine(SVM) Logistic Model Tree(LMT)
Web-based Application	YES	NO	NO	NO
Give tips if children have school violence	YES	NO	NO	NO

Chapter 5 : Run the Application

5.1 Introduction

This chapter presents a detailed, step-by-step tutorial on how to execute the web application. The following instructions will guide you through the process.

5.2 Run the Application

1) After running the application, it will redirect you to the login page.

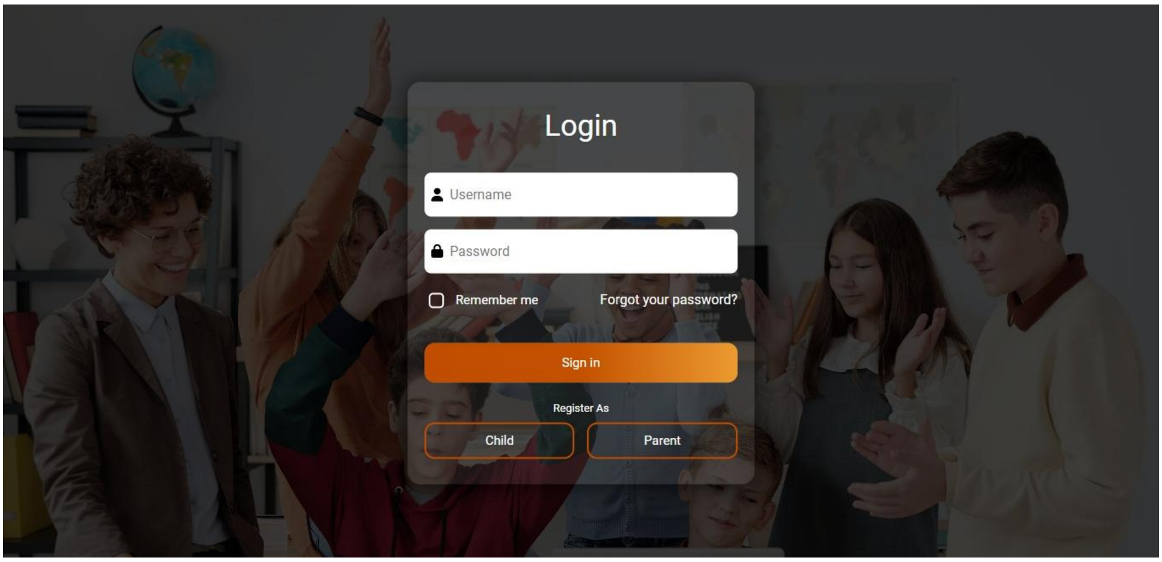


Figure 5:1 After running the application, it will redirect you to the login page.

2) Proceed by selecting the "Sign up as a parent" option and entering the required information.

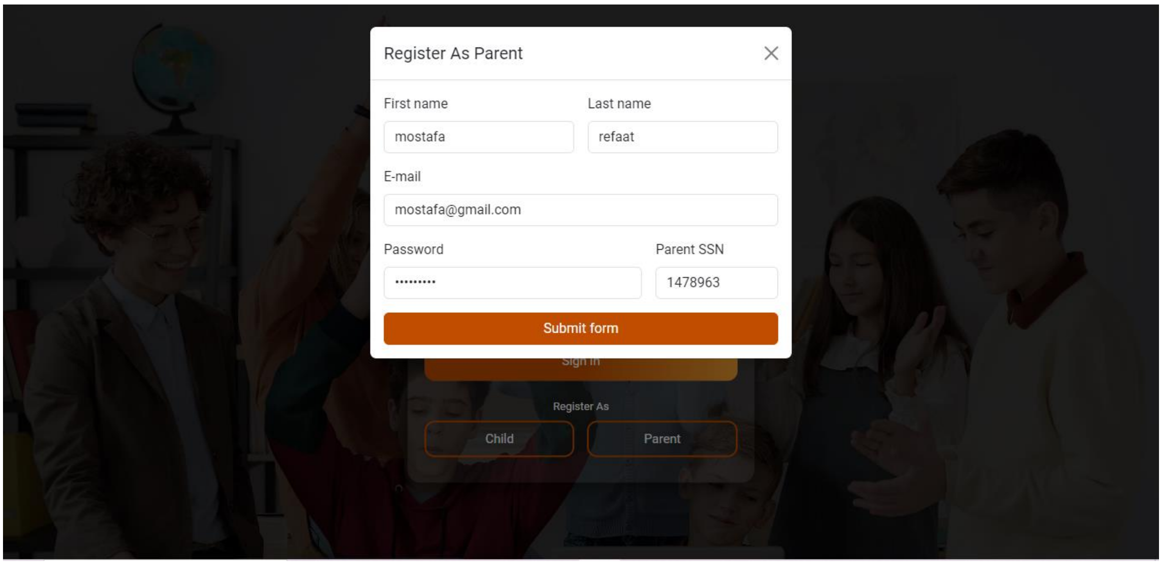


Figure 5:2 Sign up as a parent

3) Similarly, choose the "Sign up as a child" option and provide the necessary details.

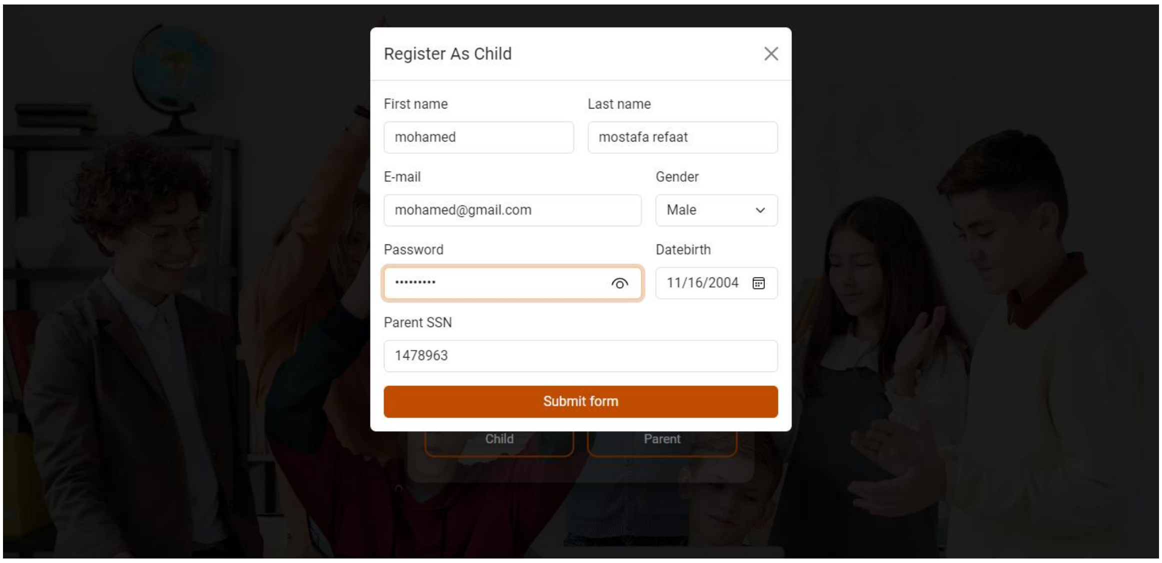
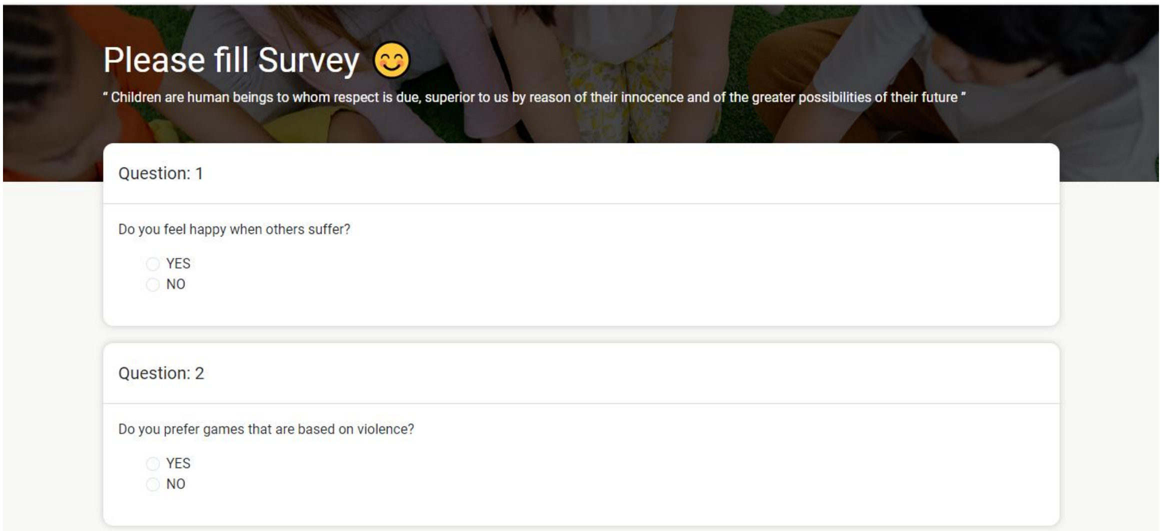


Figure 5:3 Sign up as a child

4) Once registered, log in as a child to complete the questionnaire.



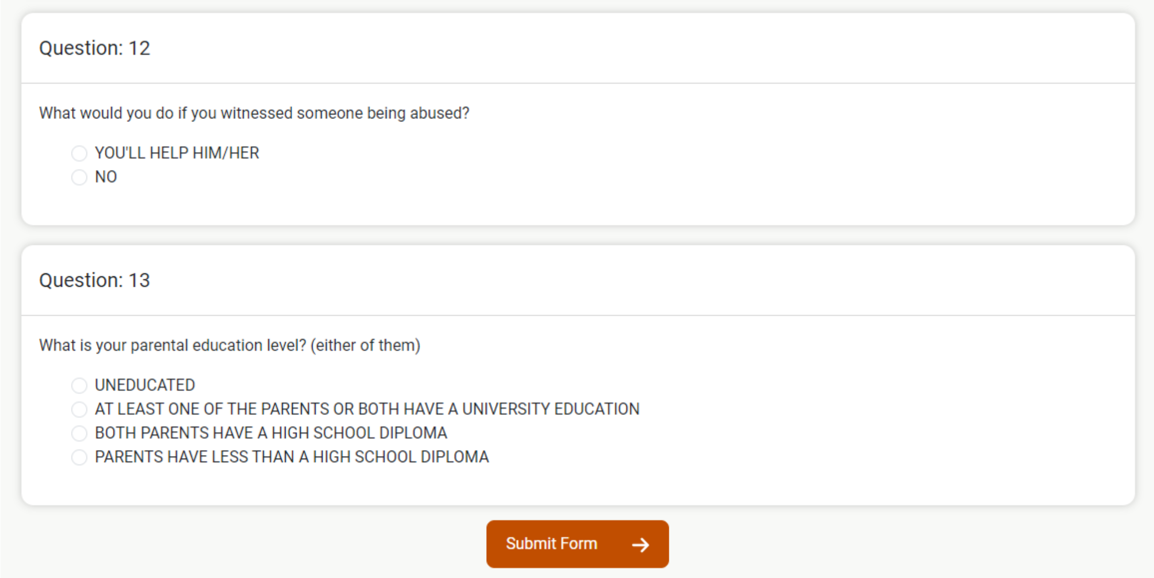


Figure 5:4 The questionnaire

5) Afterward, log in as a parent to view your child's results

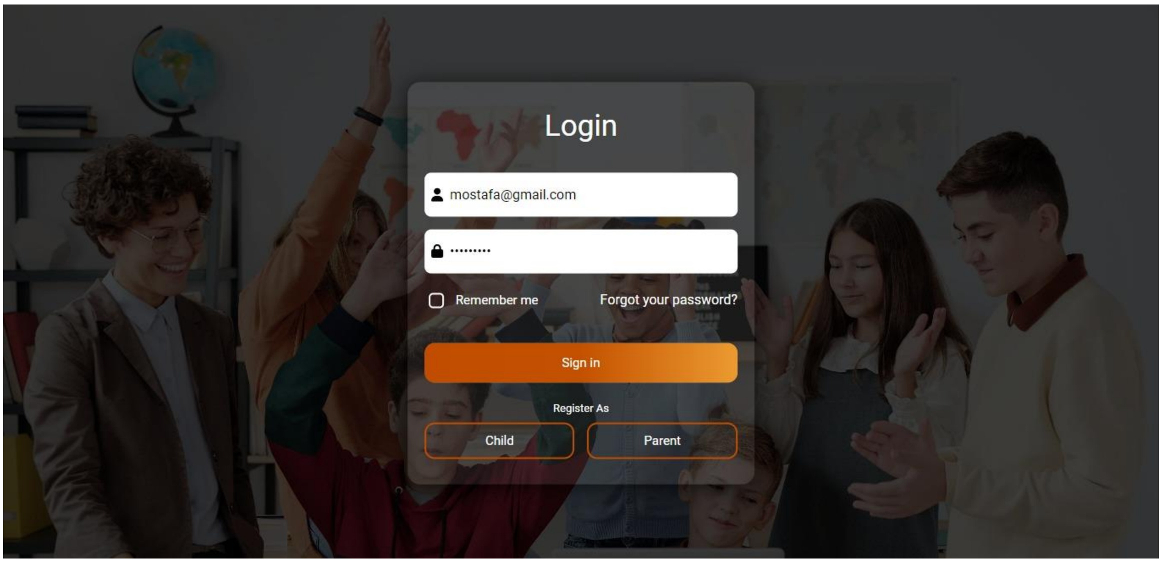


Figure 5:5 Log in as a parent

6) After logging in as a parent, you will be redirected to the home page.

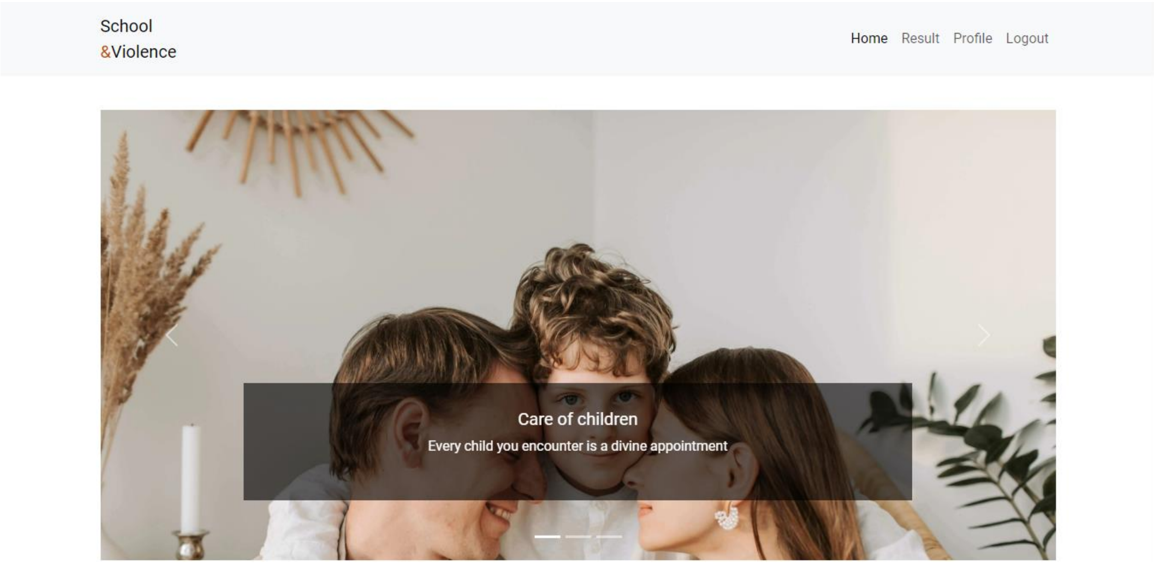


Figure 5:6 The home page

7) To view your child's results, simply click on the "Results" tab.

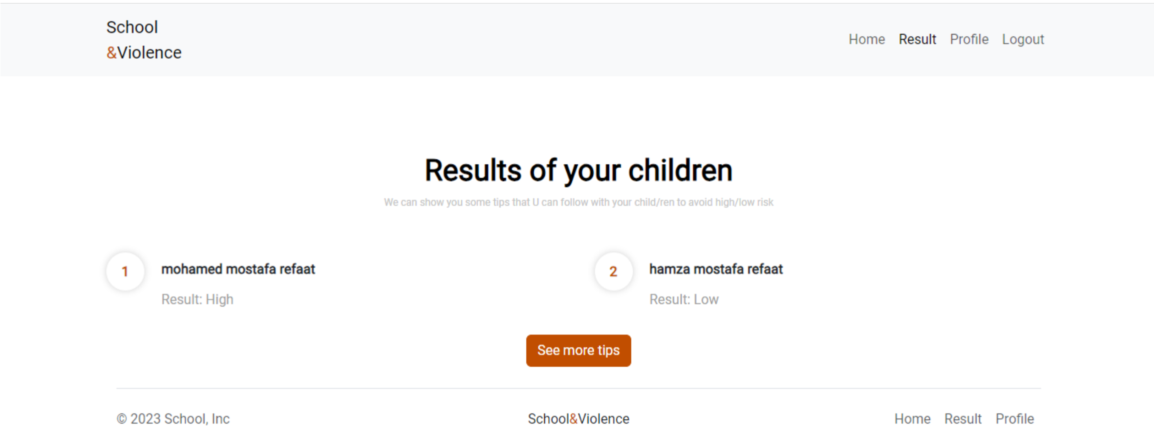


Figure 5:7 Child's results

8) To explore tips on how to avoid high-risk assessments of violence, simply click on the "See More Tips" button.

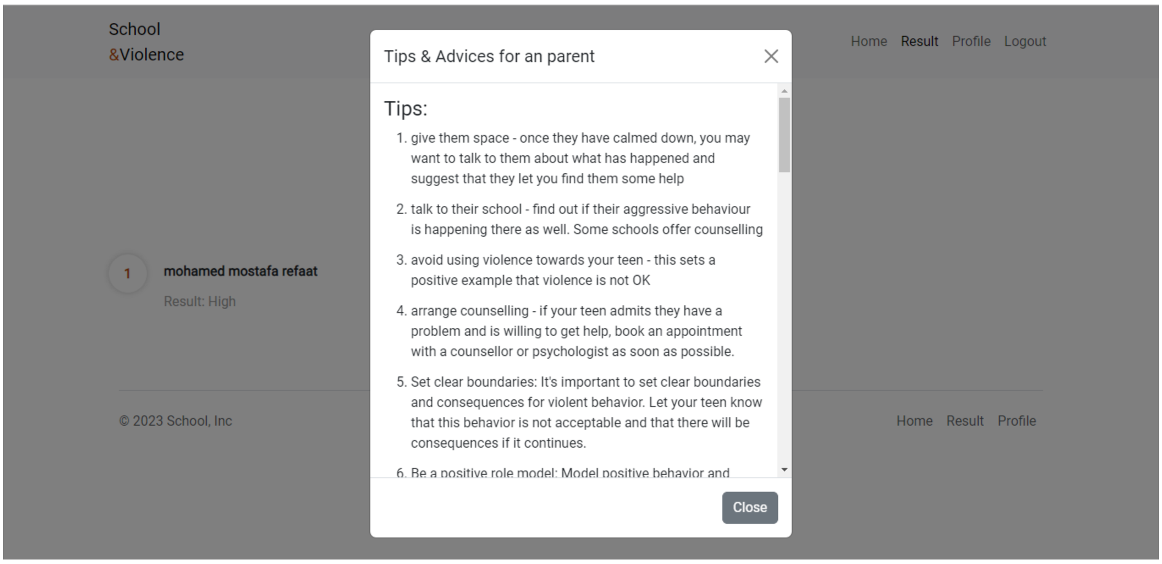
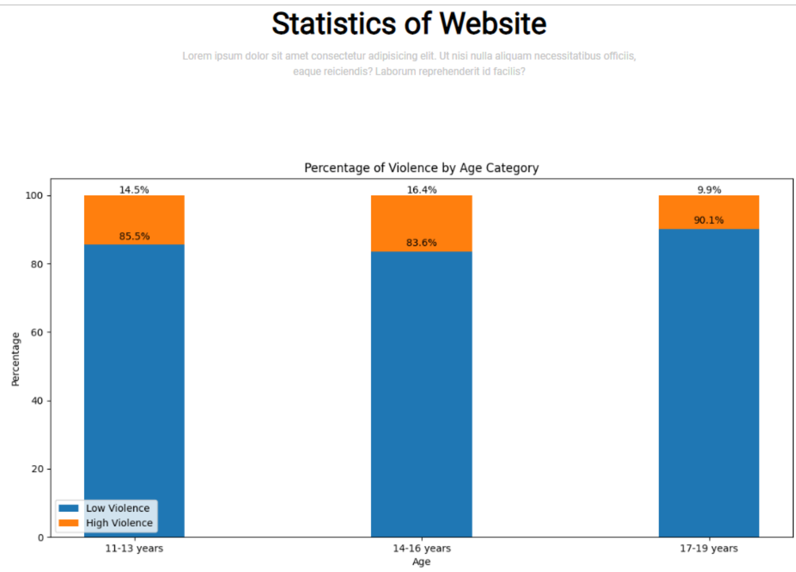


Figure 5:8 See More Tips

9) To access the dashboard containing information about the risk assessment of violence, click on the "Home" button. From there, you can view the comprehensive dashboard.



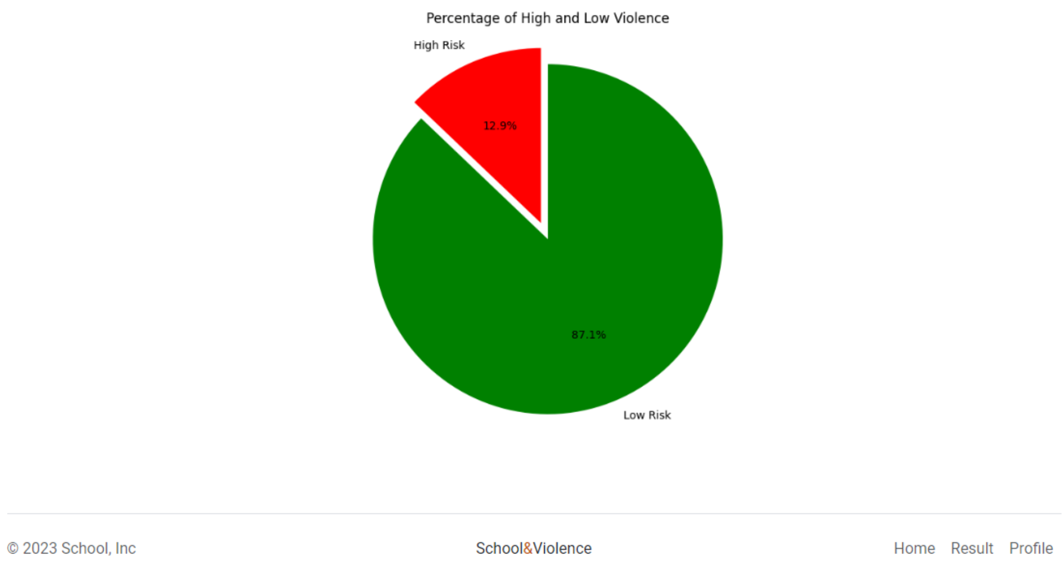


Figure 5:9 Dashboard

10) Additionally, if you want to view your personal information, click on the "Profile" button.

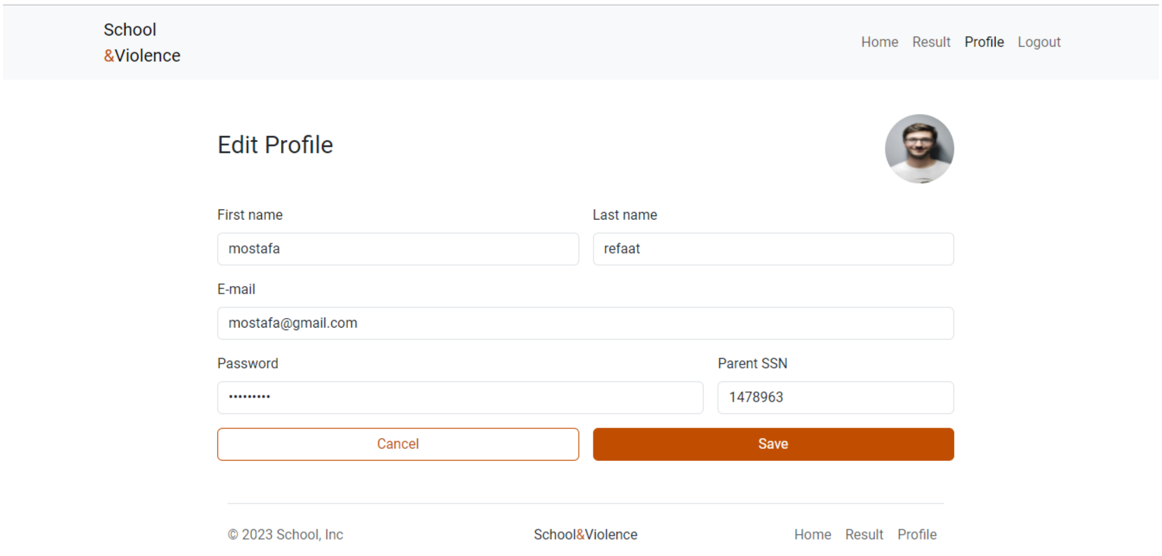


Figure 5:10 Profile Page

Chapter 6 : Conclusion and Future Work

6.1 Conclusion

It is of utmost importance to establish a highly effective approach aimed at identifying children who are at risk for exhibiting violent behavior. The reason for this urgency lies in the fact that experiencing violence during childhood leaves a profound and enduring impact on an individual's overall health and well-being throughout their entire life. As children are particularly vulnerable, they face the highest risk of suffering from psychological distress or engaging in harmful behaviors that can adversely affect others. So we decided to put a very effective approach through create a web-based application using machine learning techniques to help parents figure out whether their children are at risk of being violent to their colleagues or not. The web-based application utilizes machine learning algorithms to analyze various factors and indicators associated with potentially violent behavior. Three widely employed machine learning algorithms, namely logistic regression, random forest, and support vector machine (SVM), were utilized. These algorithms were applied to the collected dataset to predict and assess the risk of violence in adolescents. To evaluate the risk of violence in adolescents, a thorough questionnaire was developed based on the research titled "Assessing the Risk of Violence in Adolescents in the Pediatric Emergency Department". The questionnaire encompassed various risk factors identified in the study. Data for this research was collected from two primary sources: social media platforms and direct surveys conducted with students from specific schools. The accuracy results of different algorithms for automated risk assessment of school violence. The Random Forest algorithm achieved an accuracy of 93%, while Support Vector Machine (SVM) achieved 95% accuracy, and Logistic Regression achieved the highest accuracy of 97.5%. These accuracy measures indicate the performance of the respective algorithms in accurately predicting and assessing the risk of school violence.

6.2 Future Work

The current system has a few drawbacks that could potentially have fixes in the future:

1) Detecting the risks of violence among young people, not only for children.
2) Detecting the risks of violence against oneself.
3) To improve the interaction and engagement with the child, the system could consider using indirect questions that are more child-friendly. Instead of directly asking about risks of violence, the system can frame questions in a more subtle and playful manner.

References

[1] Fast Fact: Preventing School Violence |Violence Prevention|Injury Center|CDC. (n.d.).https://www.cdc.gov/violenceprevention/youthviolence/schoolviolence/fastfact.html

[2] Press Release - New NCES Data Show Increases in School Shootings and Cyberbullying in K–12 Schools Over the Last Decade - June 28, 2022. (n.d.). https://nces.ed.gov/whatsnew/press_releases/06_28_2022.asp.

[3] McCoy DC, Roy AL, Sirkman GM. Neighborhood crime and school climate as predictors of elementary school academic quality: a cross-lagged panel analysis. Am J Community Psychol. 2019 Sep;52(1–2):128–40.

[4] Burdick-Will J. School violent crime and academic achievement in Chicago. Sociol Educ. 2020 Oct;86(4):343–61. https://doi.org/10.1177/0038040713494225.

[5] Mroczkowski, M. M., Walkup, J. T., & Appelbaum, P. S. (2021). Assessing Violence Risk in Adolescents in the Pediatric Emergency Department: Systematic Review and Clinical Guidance. Western Journal of Emergency Medicine, 22(3). https://doi.org/10.5811/westjem.2021.1.49233

[6] Henry, Stuart. "What is school violence? An integrated definition." The annals of the American academy of political and social science 567.1 (2000): 16-29.

[7] Garaigordobil, Maite, and Vanesa Martínez-Valderrey. "Impact of Cyberprogram 2.0 on different types of school violence and aggressiveness." Frontiers in psychology 7 (2016): 428.

[8] Chen, Ji-Kang, and Hsi-Sheng Wei. "The impact of school violence on self-esteem and depression among Taiwanese junior high school students." Social indicators research 100 (2011): 479-498.

[9] Van Le, D., Montgomery, J., Kirkby, K. C., & Scanlan, J. (2018). Risk prediction using natural language processing of electronic mental health records in an inpatient forensic psychiatry setting. Journal of biomedical informatics, 86, 49-58.

[10] Ni, Y., Barzman, D., Bachtel, A., Griffey, M., Osborn, A., & Sorter, M. (2020). Finding warning markers: leveraging natural language processing and machine learning technologies to detect risk of school violence. International journal of medical informatics, 139, 104137.

[11] Barzman, D., Ni, Y., Griffey, M., Bachtel, A., Lin, K., Jackson, H., ... & DelBello, M. (2018). Automated risk assessment for school violence: A pilot study. Psychiatric quarterly, 89(4), 817-828.

[12] LaValley, Michael P. "Logistic regression." Circulation 117.18 (2008): 2395-2399.

[13] Pal, Mahesh. "Random forest classifier for remote sensing classification." International journal of remote sensing 26.1 (2005): 217-222.

[14] Suthaharan, Shan, and Shan Suthaharan. "Support vector machine." Machine learning models and algorithms for big data classification: thinking with examples for effective learning (2016): 207-235.

[15] Raggett, Dave, Arnaud Le Hors, and Ian Jacobs. "HTML 4.0." Last Modified: Monday, Nov 17 (1997).

[16] Abbasi, Sadegh, Farzin Mokhtarian, and Josef Kittler. "Enhancing CSS-based shape retrieval for objects with shallow concavities." Image and vision computing 18.3 (2000): 199-211.

[17] Guha, Arjun, Claudiu Saftoiu, and Shriram Krishnamurthi. "The essence of JavaScript." ECOOP 2010–Object-Oriented Programming: 24th European Conference, Maribor, Slovenia, June 21-25, 2010. Proceedings 24. Springer Berlin Heidelberg, 2010.

[18] Copperwaite, M., & Leifer, C. (2015). Learning flask framework. Packt Publishing Ltd.

[19] Millman, K. Jarrod, and Michael Aivazis. "Python for scientists and engineers." Computing in science & engineering 13.2 (2011): 9-12.

[20] Code, Visual Studio. "Visual studio code." línea]. Available: https://code. visualstudio. com (2019).

[21] Gunawan, Teddy Surya, et al. "Development of video-based emotion recognition using deep learning with Google Colab." TELKOMNIKA (Telecommunication Computing Electronics and Control) 18.5 (2020): 2463-2471.

[22] Kofler, Michael. "What Is MySQL?." The Definitive Guide to MySQL. Berkeley, CA: Apress, 2001. 3-19.