PROJECT REPORT

Project Title: - Addressing Domestic Violence in India

Under the Guidance of:
Consulting & Analytics Club
Indian Institute of Technology Guwahati

Date: May 2025 - June 2025

Table of Contents

Introduction & Project Overview

Introduction

Domestic violence is a serious and pervasive issue that affects approximately **1 in 3 women worldwide** during their lifetime, according to the World Health Organization. This form of violence includes physical, emotional, psychological, and economic abuse, often leading to profound and lasting consequences not only for the victims but also for their families and broader communities. The trauma inflicted by domestic violence can result in physical injuries, mental health disorders such as depression and anxiety, and intergenerational cycles of abuse. Despite its widespread prevalence, domestic violence remains significantly underreported in many regions due to social stigma, fear of retaliation, cultural norms, and inadequate access to support services.

Early identification of at-risk individuals and timely intervention are critical to preventing the escalation of violence and safeguarding the well-being of victims. However, traditional support systems typically depend on victims reaching out for help—something that is not always possible given the complexities and dangers surrounding abusive relationships. This gap highlights the urgent need for proactive, data-driven solutions that can predict potential risks and facilitate access to resources before violence occurs. Addressing this challenge, **Empower Sakhi** has been developed as an innovative application aimed at predicting domestic violence risk and providing users with immediate support and guidance, empowering them to take control of their safety and begin the journey toward healing.

Empower Sakhi: A Data-Driven Solution

To address this critical gap, **Empower Sakhi** has been developed as a proactive and accessible tool aimed at predicting the risk of domestic violence based on key socio-economic and behavioral factors. By leveraging machine learning and data analytics, the app provides personalized risk assessments, helping users recognize warning signs early.

Empower Sakhi not only estimates risk but also connects users to essential resources such as emergency contacts, counseling services, and shelters — fostering timely support while maintaining strict privacy and confidentiality. This innovative approach aims to empower vulnerable individuals and enhance preventive efforts.

Empower Sakhi serves as a crucial intervention by offering an intelligent and user-friendly platform to identify individuals who may be at risk of domestic violence before situations escalate. By integrating data-driven insights with real-world socio-economic and behavioral indicators, the app helps users understand their vulnerability and take informed steps toward safety.

Project Background, Problem Statement, and Objectives

Background

Domestic violence remains one of the most pervasive and deeply entrenched social issues in India, severely impacting the safety, dignity, and well-being of millions of women across all socio-economic groups. According to the National Family Health Survey (NFHS-5), nearly 30% of Indian women aged 15 to 49 have experienced domestic violence at least once in their lifetime. However, these figures likely underrepresent the true scale of the problem due to widespread underreporting driven by cultural stigma, fear of retaliation, and limited access to legal and social support.

The consequences of domestic violence are far-reaching, not only affecting victims but also their families, communities, and the nation's socio-economic development. It perpetuates cycles of gender inequality, hampers mental and physical health, and poses a serious public health challenge. The COVID-19 pandemic exacerbated these vulnerabilities, with reported cases of domestic abuse increasing significantly during lockdowns, highlighting the urgent need for innovative and proactive intervention strategies.

Problem Statement

Despite growing awareness and numerous government initiatives, domestic violence victims often face barriers to timely and effective help. Many support systems rely on victims self-reporting abuse, which may not always be feasible due to fear, shame, or lack of awareness. There is a critical gap in early risk identification and confidential access to resources, which results in delayed intervention and escalation of violence.

Traditional approaches have been reactive rather than preventive, lacking integration of modern technology that can enable proactive identification of high-risk individuals and personalized support. Furthermore, the diverse socio-economic and cultural fabric of India demands context-sensitive, data-driven solutions to address the root causes and facilitate effective prevention and response mechanisms.

About the Dataset

The dataset used in this project was carefully created to reflect real-world patterns of domestic violence. Since access to detailed personal data is limited due to privacy and ethical concerns, we combined publicly available information from trusted sources such as the National Family Health Survey (NFHS) and the National Crime Records Bureau (NCRB) with data I personally collected.

During my volunteering with a local NGO, I visited community camps and spoke directly with survivors and social workers. This experience gave me valuable insights into the common risk factors and personal experiences that victims face.

The first-hand exposure helped me understand the social and emotional contexts behind the numbers, allowing me to incorporate real, human-centered details into the dataset.

The data includes various personal and socio-economic factors such as age, education, income, marital status, number of children, employment status, and housing situation. It also captures sensitive elements like partner behavior, history of past violence, mental health issues, substance abuse, and previous reports of abuse. These details enable a comprehensive understanding of the factors contributing to domestic violence.

For example, the dataset contains entries like a 24-year-old unmarried individual with a technical diploma who has a history of substance abuse and past violence, and a 45-year-old married person living in a shelter with previous reports of abuse. This diversity in the data allows the Empower Sakhi app to identify at-risk individuals accurately and offer personalized support, aiming to intervene early and reduce the incidence of domestic violence.

Project Objectives

The **Empower Sakhi** project aims to fill this gap by leveraging machine learning and data analytics to develop a predictive mobile application designed to identify individuals at risk of domestic violence and provide timely, confidential support. The specific objectives include:

- **Risk Prediction:** Utilize socio-economic, behavioral, and historical data to build accurate predictive models that estimate the risk of domestic violence for individuals and households.
- **Early Intervention:** Enable early detection of risk factors to prevent escalation and promote timely help-seeking behaviors.
- **Resource Accessibility:** Offer users direct access to essential support services, including emergency helplines, counseling, legal aid, and shelters, while ensuring privacy and confidentiality.
- **Awareness and Education:** Increase user awareness about domestic violence indicators and empower users with knowledge and tools to protect themselves and others.
- **Data-Driven Policy Support:** Provide insights to policymakers, social workers, and law enforcement agencies for informed decision-making, improved resource allocation, and targeted interventions.

By addressing these objectives, Empower Sakhi seeks to transform domestic violence prevention and support systems in India, making them more proactive, accessible, and effective.

Exploratory Data Analysis (EDA) & Dashboard Insights

To build a meaningful and predictive solution for domestic violence risk, a comprehensive exploratory data analysis (EDA) was conducted. The goal was to identify key trends, correlations, and patterns that could provide insights into the root causes and risk factors associated with domestic abuse. This data-driven foundation is essential for designing effective interventions and personalized support within the **Empower Sakhi** application.

1. Data Cleaning and Preparation

The dataset, consisting of 150,000 records, was first cleaned and preprocessed to ensure reliability and consistency:

- **Missing values** were handled through imputation:
  - Numerical columns (like age, number of children) were filled with mean values.
  - Categorical columns (like education, marital status) were filled with the most frequent (mode) values.
- After preprocessing, the dataset had **zero missing values**, ensuring readiness for accurate analysis.

2. Statistical Summary

A statistical overview revealed that:

- The **average age** of individuals was around **33.5 years**, with most falling between **18 and 49** years.
- The **average number of children** was around **1**, and some had up to 3 children.
- Approximately **41% of the individuals had experienced domestic violence**, highlighting the seriousness of the issue in the data.
- Most individuals reported **previous reports of violence**, though a small fraction had made multiple reports—up to **six prior cases**.

3. Categorical Variable Exploration

We examined distributions across key social, demographic, and behavioral categories:

- **Education**: Most had completed upper or lower secondary school; fewer had diplomas or college degrees.
- **Income**: A majority reported earnings below ₹10,000/month, indicating economic vulnerability.
- **Marital Status**: The dataset included married, unmarried, divorced, and widowed individuals, with the majority being married.
- **Partner Alcoholism & Support System**: Around one-third reported alcoholic partners, and about half lacked a strong support system.

- **Mental Health & Employment**: A significant number of individuals reported mental health issues, and employment was split across full-time, part-time, homemakers, and unemployed.

All of this information was visualized using bar charts and count plots, helping us see the prevalence of each category clearly.

4. Correlation and Risk Analysis

Using a correlation matrix and feature-target analysis, we found that the strongest indicators of domestic violence were:

- **Past history of violence**
- **Mental health issues**
- **Substance abuse** (either self or partner)
- **Low income**
- **Low education levels**

Although some features like age or number of children had a very weak correlation, they were still considered for model training based on behavioral insights.

The **correlation matrix** revealed:

- A moderate correlation (0.16) between **previous reports** and **violence occurrence**
- Other numerical features had weaker but relevant relationships

5. Visual Analytics with Power BI

To enhance interpretability, we built a **Power BI dashboard** that visualizes core insights from the dataset. The dashboard includes:

- **Geographic Heatmaps** showing concentration of violence reports across demographic categories.
- **Risk Factor Charts** comparing features like income, education, and partner behavior with violence outcomes.
- **Trend Charts** visualizing how multiple features interact (e.g., unemployment + no support system + substance abuse = higher risk).

This dashboard served as a dynamic tool to:

- Present findings to stakeholders
- Identify high-risk individuals and communities
- Support future outreach planning by NGOs or government teams

Key Inferences from EDA

- **Socioeconomic status** (especially low income and unemployment) plays a vital role in increasing vulnerability.
- **Partner behaviors**, particularly alcoholism, are critical predictors.
- **Support systems** can act as a protective factor—those with a social or family support system were less likely to experience recurring violence.
- **Mental health concerns** and prior exposure to violence significantly increase risk.

Conclusion: Root Causes Identified

The EDA process confirmed that domestic violence is a complex issue influenced by intertwined social, psychological, and economic factors. The **primary risk indicators** included:

✔ **Past reports or exposure to violence**
People who've faced violence before are more likely to experience it again due to repeated patterns or vulnerable environments.

✔ **Presence of mental health issues**
Mental health struggles can increase emotional stress and conflict, raising the risk of domestic violence.

✔ **Alcohol or substance abuse**
Substance use—by either partner—can lead to loss of control, aggressive behavior, or unsafe home conditions.

✔ **Lower education and income levels**
Limited education and financial stress often make individuals more vulnerable to abuse and less able to leave unsafe situations.

✔ **Lack of a support system**
Without trusted family, friends, or community support, victims may feel isolated and unable to seek help or escape violence.

These insights shaped the logic behind the **risk prediction engine** in the Empower Sakhi app and validated the need for early intervention strategies that consider the full context of a person's life.

Predictive Modeling for Domestic Violence Risk Assessment

1. Overview

This section outlines the machine learning pipeline used to build a predictive model for assessing the risk of domestic violence. The model is trained on the preprocessed dataset using a Random Forest Classifier and aims to identify high-risk cases based on socio-economic, behavioral, and situational factors.

---

2. Data Preparation and Encoding

To begin, the dataset was imported and preprocessed:

- **Categorical Variables Encoding**: Label Encoding was applied to all categorical columns such as education level, income bracket, partner alcohol use, etc., to convert them into numerical format for model compatibility.
- **Handling Missing Values**: All missing values were dropped to ensure data quality.
- **Feature Scaling**: StandardScaler was used to normalize the feature values, which helps improve the performance of distance-based models and prevents bias due to scale differences.

---

3. Model Training

A **Random Forest Classifier** was chosen due to its ability to handle complex, non-linear relationships and its robustness against overfitting. The data was split into training and test sets in an 80:20 ratio, using stratified sampling to preserve the class distribution.

- **Model Configuration**:
    - Number of trees: 100
    - Maximum tree depth: 10
    - Random state: 42 (for reproducibility)

The model was then trained on the scaled training data.

To predict the risk of domestic violence, we needed a model that could work well with both numbers and categories (like income, education, marital status, etc.), handle complex patterns, and also tell us which factors are most important in the prediction. For these reasons, we chose the **Random Forest Classifier**.

A **Random Forest** is a machine learning model that combines the power of many decision trees — kind of like a group of experts each giving their opinion, and the final decision is based on the majority vote. This method helps improve accuracy and avoids common problems like overfitting, where a model learns the training data too well but performs poorly on new data.

How the Model Works

1. **Each decision tree** looks at a random part of the data and makes its own prediction.
2. The **forest** (i.e., all the trees together) then **votes** on whether a person is at risk or not.
3. The final result is based on the majority vote — which makes the prediction more stable and less biased by noise.

---

4. Model Evaluation

The model achieved an impressive performance on the test set:

- ✅ **Accuracy**: **93.71%**
- 📊 **Precision, Recall, F1-Score**:
  - Class 0 (No Violence): Precision = 0.94, Recall = 0.96
  - Class 1 (Violence): Precision = 0.94, Recall = 0.90
- 🧮 **Confusion Matrix**:
  - True Negatives: 17,043
  - False Positives: 722
  - False Negatives: 1,164
  - True Positives: 11,071

This performance indicates the model is fairly balanced and reliable in identifying both violence and non-violence cases.

---

5. Feature Importance

The model provided insight into which features most influence the risk of domestic violence:

Feature	Importance Score
Past Violence	0.248
Partner Alcoholic	0.244
Self-Substance Abuse	0.164
Previous Reports	0.076

Feature	Importance Score
Support System	0.05
Employment Status	0.051
Mental Health Issues	0.050

The top three predictors are **past violence**, **partner's alcohol abuse**, and **self-substance abuse**, confirming findings from the earlier EDA. These patterns reinforce the role of behavioral history and support systems in identifying at-risk individuals.

---

6. Model Saving

To ensure reusability and deployment-readiness, the trained model and scaler were saved using `joblib`. These can be loaded later for inference in web apps or dashboards without retraining the model.

```
joblib.dump(model, 'domestic_violence_model.pkl')
joblib.dump(scaler, 'scaler.pkl')
```

---

7. Model Performance

After training, the model was tested on unseen data and gave excellent results:

- ✅ **Accuracy:** 93.7% — which means the model correctly predicted risk in most cases.
- 📊 **Precision and Recall:** Both were strong, indicating the model is good at correctly identifying both high-risk and low-risk individuals.
- 🧮 **Confusion Matrix:** Showed that false positives and false negatives were quite low.

Empower Sakhi: A Technological Solution to Domestic Violence

Deployment, Risk Prediction, and Practical Impact

---

1. Introduction and Overview of Empower Sakhi Application

Domestic violence is a deeply entrenched social problem affecting millions, predominantly women, across the globe. Often hidden by stigma, fear, or lack of resources, victims suffer in silence, sometimes with tragic consequences. Addressing this requires innovative approaches that not only raise awareness but provide immediate and practical help.

Empower Sakhi is a web-based application designed to fill this critical gap. By integrating **machine learning risk prediction** with accessible emergency resources and emotional support, it empowers individuals at risk to evaluate their situation confidentially and take proactive steps toward safety.

The app targets survivors and potential victims of domestic violence but also supports counselors, NGOs, and law enforcement by providing a scalable digital tool for early identification and intervention. Accessible online without the need for installation, Empower Sakhi aims to break barriers of fear, accessibility, and awareness.

Why Empower Sakhi Matters?

- **Anonymity and Privacy:** Users can assess risk without revealing their identity, reducing fear of retaliation or stigma.
- **Immediate Access:** Connects users with helplines, safety tips, and police stations tailored to their location.
- **Empowerment through Knowledge:** By showing survivor stories and explaining risk factors, the app fosters hope and self-efficacy.
- **Wide Accessibility:** Designed for easy use on any device with internet and available in English and Hindi to reach diverse populations.

Empower Sakhi is not just an app but a digital ally that blends technology and social welfare to tackle domestic violence on a broad scale.

---

2. Deployment Architecture and Technology Stack

Deployment is a critical phase in making the app accessible and usable by the target audience. Empower Sakhi leverages modern cloud and web technologies for seamless distribution.

Technology Components

- **Frontend:** Built using **Streamlit**, a Python framework for rapid creation of interactive web apps. Streamlit enables a user-friendly, responsive interface with dynamic forms, buttons, and instant visual feedback.
- **Machine Learning Model:** The core prediction engine uses a **Random Forest Classifier**, a robust ensemble algorithm trained on datasets containing factors like age, income, partner's behavior, mental health status, substance abuse, and prior incidents of violence. The model outputs a binary risk assessment and a confidence probability.
- **Data Management:** Static datasets for emergency contacts and police stations are stored in CSV format and loaded via `pandas` for fast querying by user location.
- **UI Styling:** Enhanced with custom CSS and responsive design to ensure usability on both desktops and smartphones.

Deployment Methods

- **Local Deployment:** For development or offline environments, the app can be run locally with the command `streamlit run app.py`.
- **Cloud Deployment:** To maximize reach, the app is deployed on **Streamlit Cloud**:
    - Provides free and easy deployment linked to GitHub repositories.
    - Offers high uptime, scalability, and cross-platform access.
    - The app URL is publicly accessible at: https://empowersakhiapp.streamlit.app/

Alternative cloud platforms like Heroku, AWS, Google Cloud, or DigitalOcean can be used for enhanced control and scalability, but Streamlit Cloud was chosen for its simplicity and cost-effectiveness.

Benefits of this Deployment

- **Instant Global Access:** Anyone with internet can use the app without installation.
- **Cost-Effective:** No hosting costs for NGOs or social programs on Streamlit's free tier.
- **Scalable and Maintainable:** Easy to update and improve from the central code repository.

---

3. Features and Functional Modules of Empower Sakhi

The application is organized into intuitive modules that serve distinct functions to support users:

3.1 Risk Assessment Module

This is the app's cornerstone:

- Users input details about demographics, relationship dynamics, mental health, substance use, and history of abuse.
- Data preprocessing converts inputs into model-ready formats.
- The Random Forest model predicts the risk of domestic violence, classifying users into **High Risk** or **Low Risk** categories.
- The result screen provides empathetic explanations and tailored advice.

Example: A user reporting a partner's alcoholism and previous violent incidents is classified as high risk, prompting the app to recommend immediate contact with helplines and law enforcement.

3.2 Hear Her Stories

- A curated collection of anonymous survivor testimonials.
- Stories provide hope and a sense of community.
- Randomly displayed to offer new perspectives on each visit.
- Encourages users to know they are not alone and recovery is possible.

3.3 Emergency Help Section

- Lists critical helpline numbers such as **112, 1091, and 181**.
- Includes practical safety tips in both English and Hindi.
- Links to legal aid, online complaint portals, and women's shelters.
- Designed for easy access during moments of crisis.

3.4 Nearby Police Station Locator

- Accepts city and state input.
- Filters police station data to display local contacts.
- Enables prompt reporting and access to law enforcement.

3.5 Quick Exit Feature

- Provides urgent action steps to stay safe.
- Simple, clear instructions like calling emergency services or finding shelter.
- Available in both Hindi and English.

- Useful when users need to exit quickly and safely.

---

4. Real-World Impact and Use Cases

For Survivors and Potential Victims

Empower Sakhi provides a safe space for users to understand their risk without fear. Immediate access to resources and survivor stories helps victims:

- Make informed decisions.
- Prepare safety plans.
- Connect to emergency services.

For NGOs and Counselors

- Useful as a screening tool during outreach.
- Helps prioritize assistance based on risk.
- Raises awareness in communities through app dissemination.

For Government and Law Enforcement

- Supports data-driven approaches to violence prevention.
- Identifies high-risk regions for targeted interventions.
- Offers digital complement to traditional safety measures.

For Healthcare and Social Workers

- Assists in mental health and family support programs.
- Educates clients on recognizing abuse.
- Facilitates timely referrals.

For Community Centers and Shelters

- Provides a confidential platform accessible on local devices.
- Bridges communication gaps for survivors reluctant to seek face-to-face help.

---

5. Advantages and Future Enhancements

Key Advantages

- **Privacy-Centric:** No personal data stored, maintaining confidentiality.
- **User-Friendly:** Simple interface requires no prior technical knowledge.
- **Bilingual Access:** English and Hindi support broadens reach.
- **Data-Driven:** Uses validated machine learning for personalized predictions.
- **Portability:** Accessible from any device with internet.
- **Community-Oriented:** Survivor stories promote solidarity.

Future Scope

- **Real-Time Chat Support:** For immediate emotional assistance.
- **Legal Aid Automation:** Simplify understanding of rights and procedures.
- **Expanded Languages:** Adding regional Indian languages.
- **Mobile Apps:** Offline functionality for low-connectivity areas.
- **Anonymous Reporting:** Secure channels for abuse reporting.
- **Enhanced Data Integration:** Real-time location and social service info.

---

6. Conclusion

Empower Sakhi demonstrates the profound impact technology can have on social welfare. By combining machine learning risk prediction with accessible emergency resources and emotional support, it offers a scalable, cost-effective, and confidential tool to fight domestic violence.

The deployment on Streamlit Cloud ensures immediate global accessibility, making this life-saving resource available to anyone at risk, anywhere, anytime. Its modular design, thoughtful UI, and bilingual support ensure broad usability.

Moving forward, continued development, partnerships with NGOs and government agencies, and user feedback will enhance Empower Sakhi's capabilities and reach, transforming it from a pioneering application into a cornerstone of domestic violence prevention efforts.

Empower Sakhi is more than an app—it is a beacon of hope and a practical lifeline for survivors everywhere.

Summary

Overview of Domestic Violence and the Need for Innovative Solutions

Domestic violence is a critical social issue that affects millions globally, characterized by physical, emotional, psychological, and economic abuse within intimate relationships. Victims often remain trapped in abusive environments due to fear, stigma, and a lack of accessible support systems. Traditional methods of assistance can be limited by geographical, social, and technological barriers, leaving many survivors without timely help.

Addressing this challenge requires innovative approaches that leverage technology to provide discreet, immediate, and personalized support. Empower Sakhi was developed precisely to meet this need by offering an accessible, user-friendly platform that integrates machine learning-driven risk assessment with critical safety resources.

Empower Sakhi: A Technological Solution for Risk Prediction and Support

Empower Sakhi employs advanced machine learning algorithms to analyze key risk factors associated with domestic violence. By enabling users to confidentially input personal and relational data, the application delivers an informed risk prediction, helping individuals better understand their situation and take proactive safety measures.The platform further bridges the gap between victims and essential services by providing instant access to emergency helplines, local police contacts, safety guidelines, and survivor stories that offer emotional support. Its bilingual interface ensures inclusivity and broad usability.

Deployed on Streamlit Cloud, Empower Sakhi is readily accessible via web browsers across devices without requiring installation, ensuring ease of use while maintaining privacy and security. This deployment strategy facilitates rapid dissemination among vulnerable populations, NGOs, and authorities, enabling timely interventions.

Conclusion and Future Directions

Empower Sakhi exemplifies how technological innovation can be harnessed to empower vulnerable populations and address complex social challenges. By combining predictive analytics with practical support tools, the app fosters safer environments and informed decision-making. Continued enhancements, including multilingual support and real-time assistance features, will expand its impact and effectiveness, positioning Empower Sakhi as a vital resource in combating domestic violence.