KamiLimu Research: Preeclampsia in Pregnant Women by Data Science Group 2

Members: Sebastian Chanzu, Eddah Chipchirchir, Denish Awajo

Project Name: Safe Mom

Project lead: Denish Awajo

Problem: Maternal Morbidity and Mortality caused by Preeclampsia

Table of Contents

Chapter One: Problem Background

Maternal mortality remains a critical global issue, with the World Health Organization (WHO) reporting an alarming rate of 342 deaths per 100,000 live births. This highlights the urgent need for innovative technologies, such as AI and predictive data models, to identify risks better early and provide timely care, ultimately preventing these tragic and avoidable deaths.

Kenya, in particular, faces a significant challenge. The country is ranked among the top 10 contributors to neonatal deaths globally (Malachi, 2017), with hypertensive disorders, including preeclampsia, accounting for 20% of maternal fatalities. According to a 2007 report by Aga Khan, preeclampsia affects between 5.6% and 6.5% of pregnant women, underscoring the need for more effective approaches to predict and manage this life-threatening condition.

Preeclampsia, a potentially life-threatening complication of pregnancy characterized by high blood pressure and organ damage, poses significant risks to maternal and fetal health if left untreated (American College of Obstetricians and Gynecologists, 2020). It is a leading cause of maternal mortality worldwide, with approximately 76,000 maternal deaths attributed to hypertensive disorders of pregnancy annually (World Health Organization, 2019). Early identification and management of preeclampsia are critical for reducing adverse outcomes for both mothers and babies.

In recent years, there has been a growing interest in leveraging machine learning and predictive analytics to improve healthcare outcomes. Predictive models have been successfully applied in various medical domains, including risk assessment, disease diagnosis, and treatment optimization (Rajkomar et al., 2019). By analyzing large datasets containing patient demographics, clinical variables, and health outcomes, these models can identify patterns and trends that may not be apparent to human clinicians alone.

The integration of predictive models into clinical workflows holds promise for enhancing prenatal care and reducing the burden of preeclampsia-related complications. With advances in technology and data analytics, healthcare providers can now leverage predictive algorithms to identify pregnant women at high risk of developing preeclampsia and initiate timely interventions to mitigate risks.

This project aimed to develop a machine learning-based predictive model for assessing the likelihood of pregnant women developing preeclampsia during antenatal clinic visits. By leveraging comprehensive datasets containing patient demographics, medical history, pregnancy details, and laboratory results, the predictive model seeks to enable healthcare providers to make informed decisions and intervene proactively. Ethical considerations, data privacy concerns, and regulatory requirements will be carefully addressed throughout the project lifecycle to ensure the responsible deployment of the predictive model in clinical practice. Through rigorous testing, validation, and integration with existing clinical workflows, this project seeks to contribute to improved maternal and fetal outcomes in managing preeclampsia.

Machine Learning and data analytics techniques provide the possibility to develop a preeclampsia prediction model and data analysis using the tracked records and hospital records (Jhee et al., 2019). This will allow increased surveillance of at-risk patients and reduce surveillance of patients who are less likely to develop preeclampsia by the doctor.

Chapter Two: Market Opportunity

Since 2009, Kenya has registered an improvement in demographic indicators with a population increasing by almost one million per year from 37.7 million in 2009 to 47.6 million in 2019 and it is projected to reach 57.6 million by 2030. (The Annual State of Kenya and State of World Populations reports 2023 launch, 2023)

This presents a challenge given that a 2024 study by BMC in Migori County revealed there is a shortage of clinicians, according to KNBS Kenya's patient-healthcare-practitioner ratio is 19 practitioners per 100,000 people which translates to 1 practitioner per 5000 patients. The low availability of healthcare workers impacts the quality of healthcare services provided and may even deny the opportunity for early detection and intervention for mothers at risk of preeclampsia.

Also, note that in Kenya over 90% of pregnant women attend at least one antenatal care visit during their pregnancy, despite this Kenya is currently among the top 10 countries contributing to neonatal deaths globally. (Malachi,2017)

We have identified two solutions that offer a similar approach. Fitbit is only limited to tracking using smartwatches. In contrast, flo only tracks the menstrual periods in women and therefore they do not identify the risks and likelihood of a pregnant woman developing preeclampsia.

Also, our solution seeks to help clinicians predict high-risk mothers during antenatal visits, hence more accurate predictions due to more data points than these solutions that rely on limited data points to make their predictions.

Chapter Three: Solution Idea

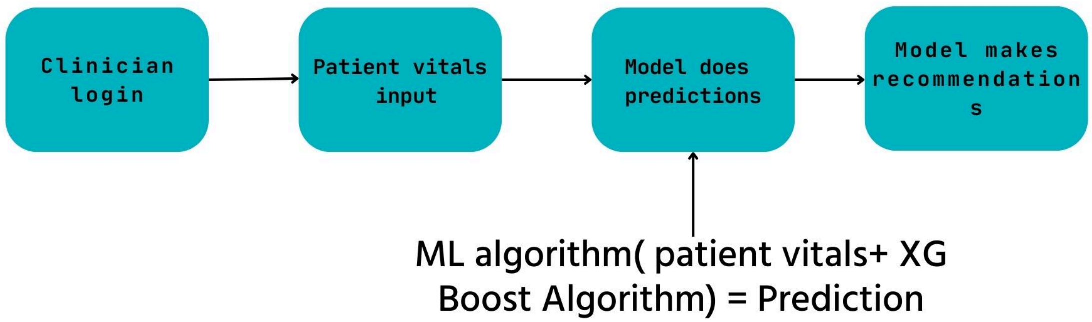
The target users for this project are clinicians, particularly those who work in maternal healthcare, including doctors, nurses, midwives, and community health workers. This group was identified because they play a critical role in monitoring and managing the health of expectant mothers. Given their front-line involvement in maternity care and experience their clinical judgment can be used to make judgments, these providers need tools to help them identify and manage high-risk pregnancies early, particularly those complicated by hypertensive disorders like preeclampsia.

The target users were identified through a comprehensive needs assessment that involved interviews with clinicians in Kenya to understand their specific challenges related to preeclampsia identification and management. This process was complemented by an in-depth literature review, which examined existing research on preeclampsia in Kenya and similar settings, revealing gaps in care and potential solutions. Additionally, expert consultations with professionals in maternal health, obstetrics, and healthcare technology provided further insights that helped inform the selection of the target users and shape the development of the solution. The choice of healthcare providers as the target users was made based on their direct involvement in the care process and their potential to positively influence patient outcomes. By equipping them with a predictive model for preeclampsia, they can intervene more effectively and reduce the risk of complications. The decision to focus on healthcare providers also stems from the existing gap in reliable, data-driven tools that can assist in predicting preeclampsia, especially in rural or resource-limited settings.

The target users were chosen based on their direct involvement in the care of expectant mothers and their need for a solution to improve preeclampsia identification and management. By focusing on these two groups, the solution can have a significant impact on reducing maternal morbidity and mortality in Kenya.

The solution being offered is a web-based platform that assists clinicians in predicting the likelihood of preeclampsia in expectant mothers by integrating patient data with a machine learning algorithm. This platform utilizes key patient vitals such as blood pressure, proteinuria, and fetal heart rate, which are processed by the XGBoost algorithm to generate accurate predictions. By enhancing early diagnosis and facilitating timely interventions, the system aims to significantly reduce maternal morbidity and mortality caused by preeclampsia.

<u>Proposed Solution</u>



Clinician Login: Healthcare providers log into the platform using their credentials.
Patient Vitals Input: Clinicians enter the relevant patient vitals data, such as blood pressure, proteinuria, and fetal heart rate, into the platform.
Model Predictions: The machine learning model, powered by XGBoost, will process this data to predict whether the patient is at risk of developing preeclampsia.
Recommendations: Based on the prediction, the platform provides recommendations to the clinician, such as increasing the frequency of monitoring, initiating preventive measures, or referring the patient to a specialist.

The proposed predictive platform directly addresses the problem of inadequate preeclampsia prediction in Kenya by providing healthcare providers with a reliable and data-driven tool. By identifying at-risk women early in their pregnancies that is during the eighth week of pregnancy mn;k, the platform enables timely interventions, reducing the risk of severe complications and maternal mortality. This solution has the potential to significantly improve maternal health outcomes in Kenya, particularly in rural areas with limited access to specialized care.

Chapter Four: Value Proposition

This project provides data-driven solutions to assess and mitigate health risks in pregnant women, empowering healthcare providers with predictive insights to improve maternal care outcomes and reduce complications during pregnancy.

Chapter Five: Designed Solution

XGBoost Algorithm: Chosen for its efficiency and accuracy in handling structured data, it helps predict the likelihood of preeclampsia based on patient vitals.

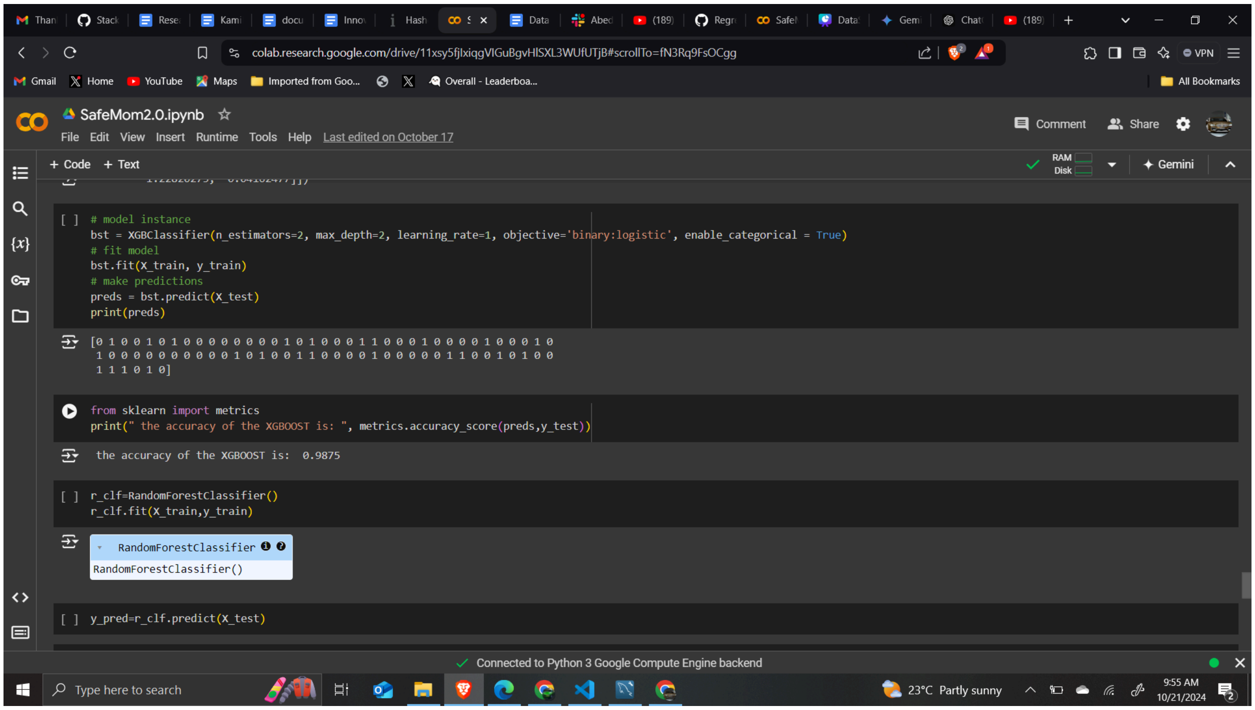
Flask Framework: Used to build the web-based platform due to its simplicity, flexibility, and compatibility with Python-based machine learning models.

Database: Employed to store patient vitals and login details, ensuring secure and organized data management in a MySQL database.

HTML/CSS: Used for creating the user interface, providing a seamless and accessible experience for healthcare providers.

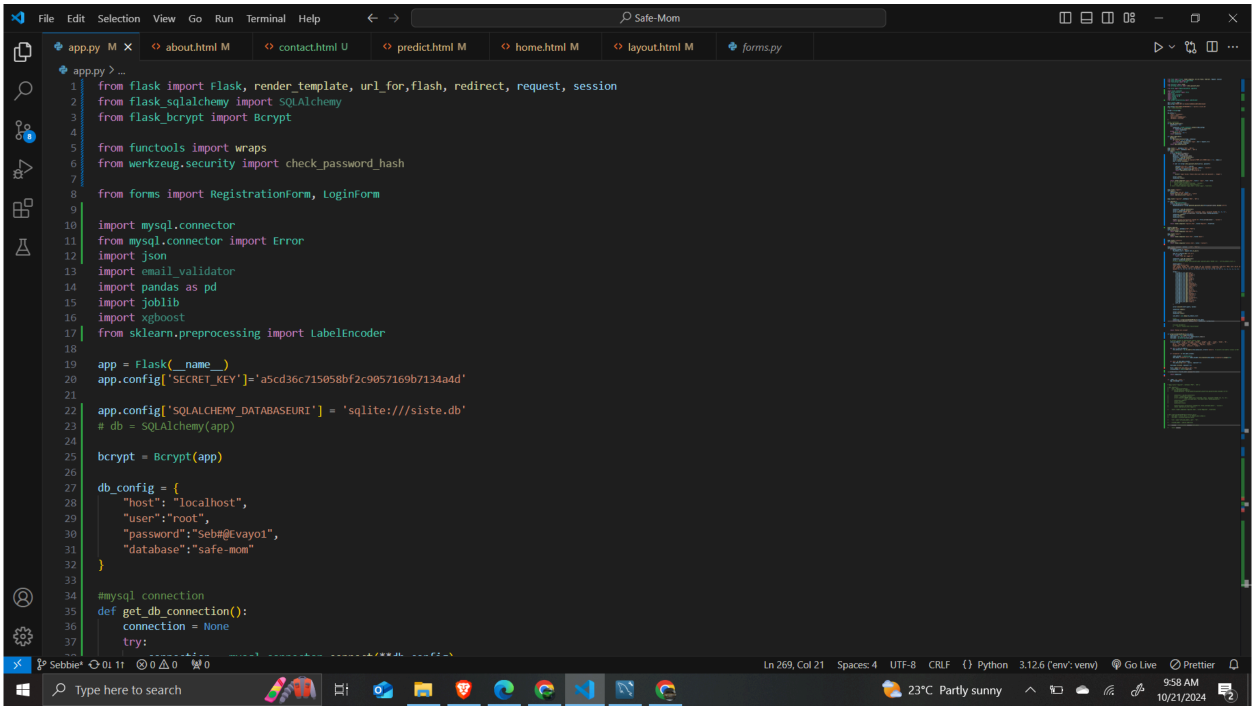
Fast API: We employed FastAPI to convert the machine learning model into a RESTful API, enabling seamless integration with the web platform.

MODEL SELECTION



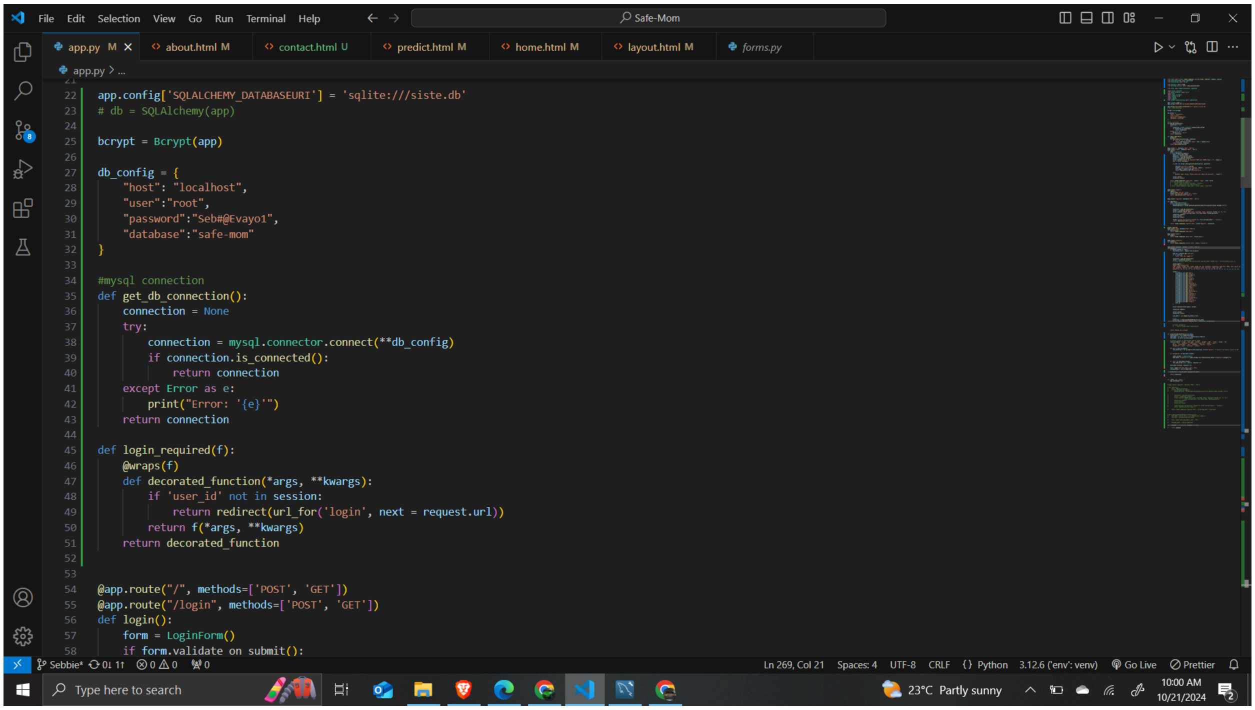
The XGBoost algorithm was used to predict the likelihood of a pregnant woman getting preeclampsia by analyzing patient vital signs. XGBoost was chosen for this because of its ability to handle tabular data efficiently and provide accurate predictions.

FLASK CONFIGURATION

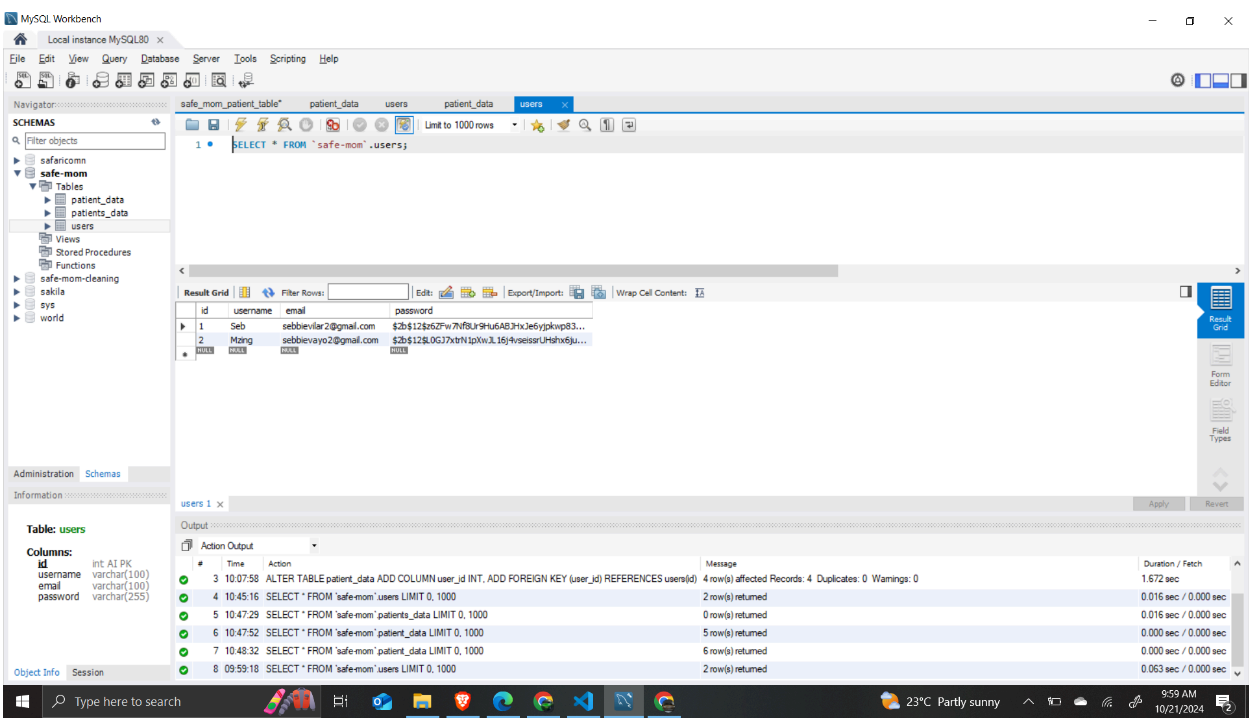


We used the flask framework to build the web-based platform, which the clinicians interact with. Flask's simplicity and compatibility with Python make it ideal for creating this platform.

DATABASE CONFIGURATION

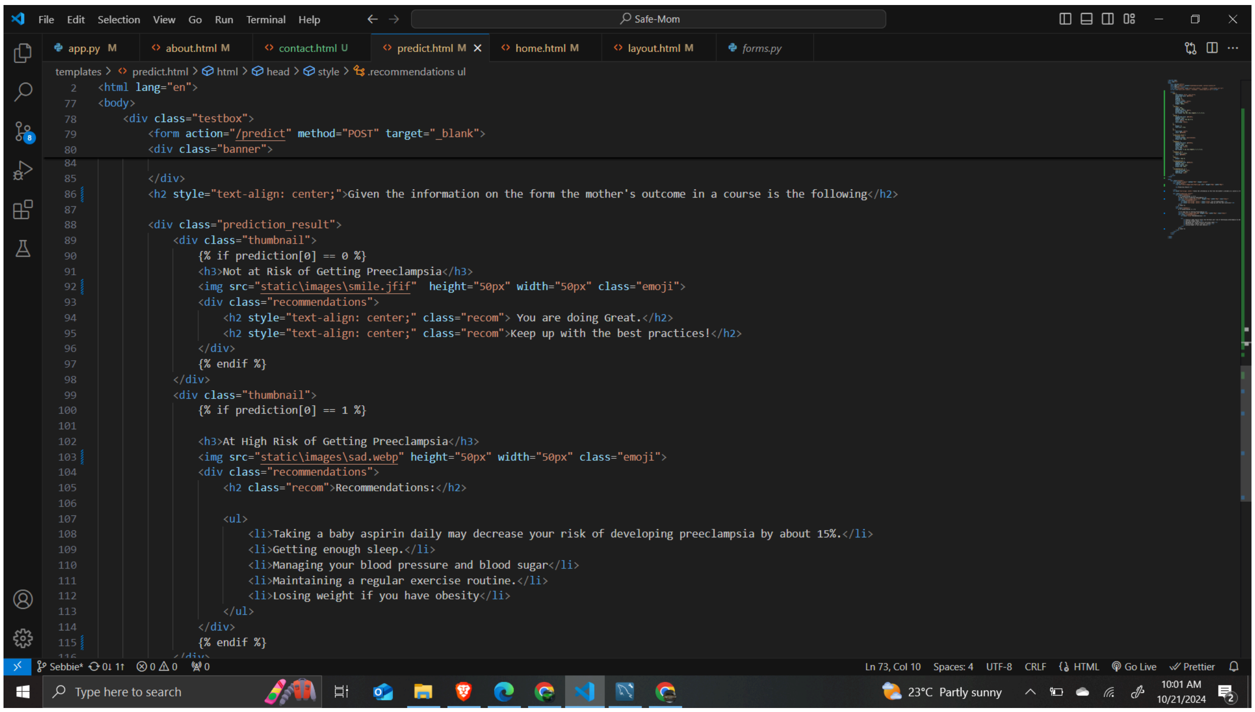


Integrating our Flask application with MySQL database.



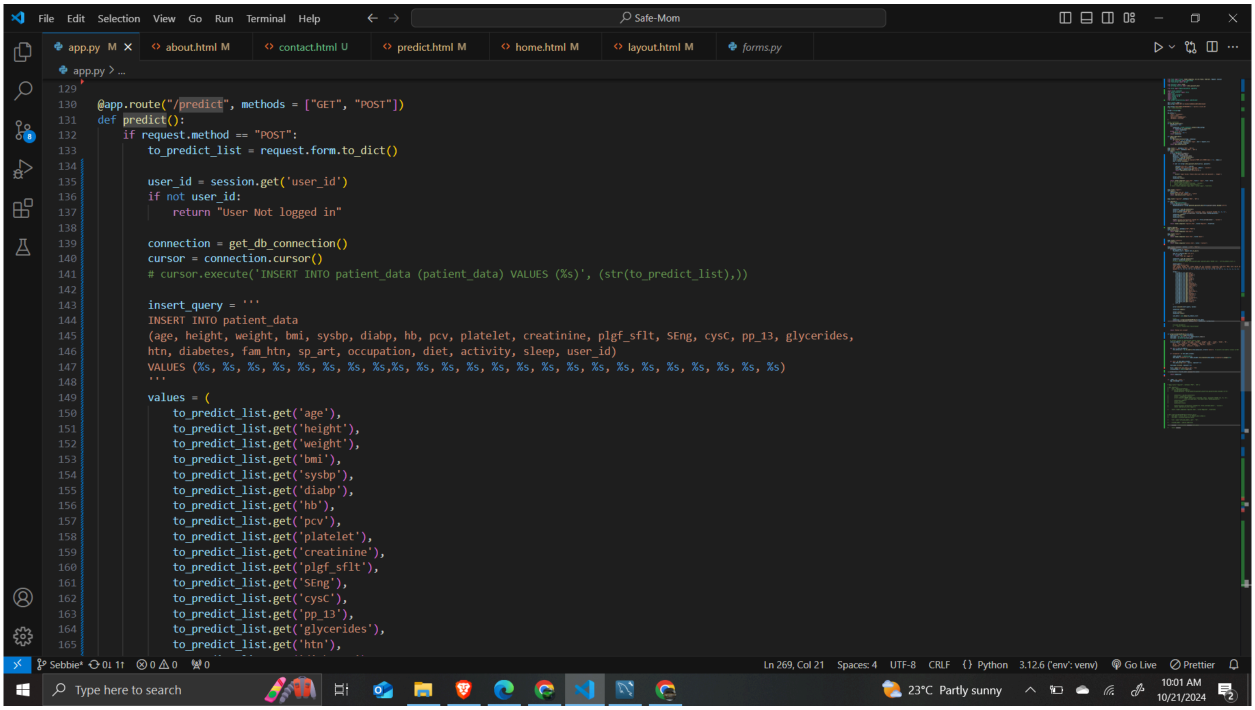
We used the MySQL database to store patient vital signs and login details securely. It ensures that the data is well-organized, easily retrievable, and protected from unauthorized access.

HTML/CSS INTEGRATION

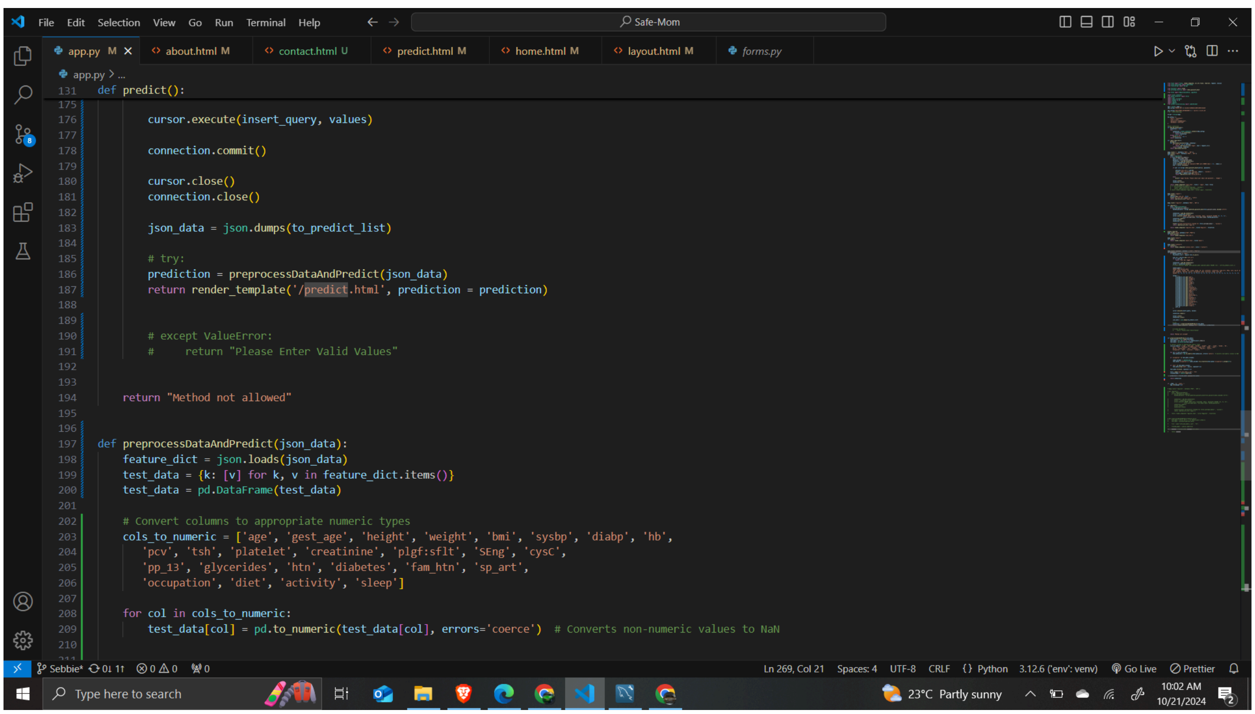


These technologies are used to create the user interface (UI) of the web platform. HTML structures the content (e.g., forms, tables), while CSS ensures the visual presentation (e.g., layout, colors, fonts) is user-friendly and visually appealing to healthcare providers.

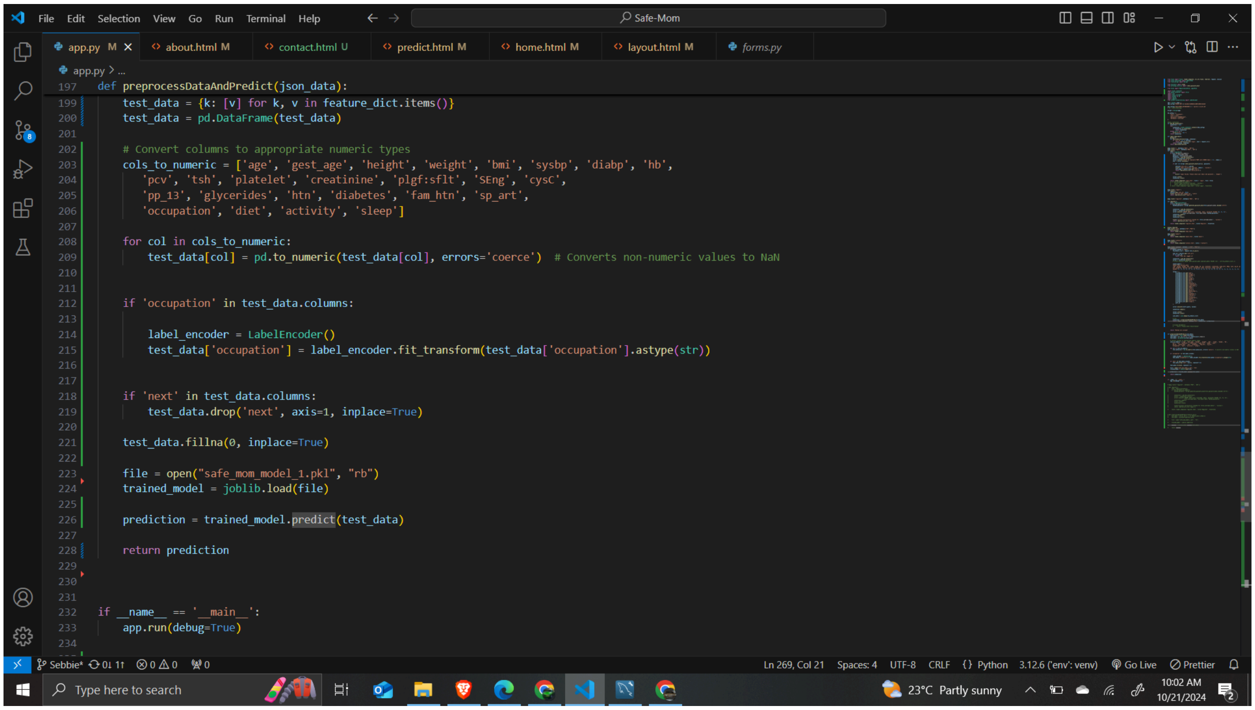
API INTEGRATION AND LINKING



```python
@app.route("/predict", methods = ["GET", "POST"])
def predict():
    if request.method == "POST":
        to_predict_list = request.form.to_dict()

        user_id = session.get('user_id')
        if not user_id:
            return "User Not logged in"

        connection = get_db_connection()
        cursor = connection.cursor()
        # cursor.execute('INSERT INTO patient_data (patient_data) VALUES (%s)', (str(to_predict_list),))

        insert_query = '''
        INSERT INTO patient_data
        (age, height, weight, bmi, sysbp, diabp, hb, pcv, platelet, creatinine, plgf_sflt, SEng, cysC, pp_13, glycerides,
        htn, diabetes, fam_htn, sp_art, occupation, diet, activity, sleep, user_id)
        VALUES (%s, %s, %s, %s, %s, %s, %s, %s,%s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s)
        '''

        values = (
            to_predict_list.get('age'),
            to_predict_list.get('height'),
            to_predict_list.get('weight'),
            to_predict_list.get('bmi'),
            to_predict_list.get('sysbp'),
            to_predict_list.get('diabp'),
            to_predict_list.get('hb'),
            to_predict_list.get('pcv'),
            to_predict_list.get('platelet'),
            to_predict_list.get('creatinine'),
            to_predict_list.get('plgf_sflt'),
            to_predict_list.get('SEng'),
            to_predict_list.get('cysC'),
            to_predict_list.get('pp_13'),
            to_predict_list.get('glycerides'),
            to_predict_list.get('htn'),
```



```python
def predict():

        cursor.execute(insert_query, values)

        connection.commit()

        cursor.close()
        connection.close()

        json_data = json.dumps(to_predict_list)

        # try:
        prediction = preprocessDataAndPredict(json_data)
        return render_template('/predict.html', prediction = prediction)


        # except ValueError:
        #     return "Please Enter Valid Values"


    return "Method not allowed"


def preprocessDataAndPredict(json_data):
    feature_dict = json.loads(json_data)
    test_data = {k: [v] for k, v in feature_dict.items()}
    test_data = pd.DataFrame(test_data)

    # Convert columns to appropriate numeric types
    cols_to_numeric = ['age', 'gest_age', 'height', 'weight', 'bmi', 'sysbp', 'diabp', 'hb',
      'pcv', 'tsh', 'platelet', 'creatinine', 'plgf:sflt', 'SEng', 'cysC',
      'pp_13', 'glycerides', 'htn', 'diabetes', 'fam_htn', 'sp_art',
      'occupation', 'diet', 'activity', 'sleep']

    for col in cols_to_numeric:
        test_data[col] = pd.to_numeric(test_data[col], errors='coerce')  # Converts non-numeric values to NaN
```



It is used to convert the machine learning model (likely the XGBoost model) into a RESTful API. This allows the machine learning model to be accessed and used by other components of the platform, such as the web application, by sending HTTP requests to the API.

Link to Github

https://github.com/SebbieMzingKe/Safe-Mom/tree/master/

Chapter Six: Business Model

Leasing Model: Hospitals can use the system free of charge for the first month, creating a trial period. They'll pay for continued access from the second month onwards, generating consistent revenue.

Subscription Model: Offer a flexible monthly or annual subscription, giving healthcare institutions access to predictive analytics services at a manageable cost.

Licensing & API Access: For a fee, allow other healthcare platforms to license your software or integrate your predictive models via API. Provide API access to individuals who would like to know their status for a fee.

Partnerships and Sponsorships: Collaborate with healthcare providers, insurance companies, or NGOs that may co-fund or subsidize the system for hospitals, increasing reach.

Grants & Donations: For non-profit sustainability, seek grants, launch crowdfunding campaigns, and develop partnerships with sponsors supporting maternal health. Host events and engage donors for continuous funding streams.

Chapter Seven: Responsible Computing

Inclusion: Ensuring broad accessibility by designing our solution to be inclusive by ensuring that healthcare providers, regardless of location or technological resources, can access the platform. By focusing on web-based delivery with minimal system requirements, we ensure that both urban and rural health facilities can benefit from this technology. Moreover, it aims to support a diverse range of users, including both experienced and less technically savvy healthcare professionals.

We will involve the community by collaborating with them and organizations working in maternal and child health to ensure that the solution meets their needs and concerns.

Accountability: Providing transparency in how the predictive model operates, detailing the factors contributing to its decisions. Clear documentation and explanations of model predictions will allow healthcare providers to make informed decisions, ensuring the platform supports rather than replaces professional judgment. Ethical oversight will be conducted through ethical review boards, and constant updates will ensure the system remains accountable to the latest medical standards.

Accessibility: The platform is designed to be user-friendly, using intuitive interfaces and making it easy for healthcare providers to navigate through patient data. We prioritize accessibility by optimizing the platform for mobile devices, ensuring that healthcare workers in regions with limited resources can still use the system.

Biasness: We address biases by incorporating diverse datasets that reflect different demographics, ensuring fairer predictions for all users. The model will be evaluated regularly to assess any emerging biases and update it to mitigate such risks.

Privacy and Security: The predictive platform for preeclampsia will prioritize data protection and privacy. It will adhere to the Kenya Data Protection Act 2019, collect only necessary data, and implement robust security measures to safeguard patient information. Informed consent will be obtained from patients before their data is collected and used, ensuring transparency and respect for their privacy rights.

Cultural Sensitivity: We recognize that maternal health challenges differ across regions, and the platform is designed with cultural sensitivity in mind. The user interface and educational materials provided to healthcare professionals are localized, taking into account cultural practices and local healthcare protocols. We aim to ensure that the platform does not promote a one-size-fits-all approach but rather adapts to the cultural contexts of the communities it serves.

Chapter Eight: Traction

Engagement with Healthcare Providers: We have spoken to several healthcare providers, including gynecologists and clinicians, to gather insights into how pre-eclampsia is currently diagnosed and managed. These discussions have revealed that clinical judgment primarily relies on three key indicators—protein in the urine, pregnancy status, and blood pressure. They confirmed that they typically diagnose pre-eclampsia around the 20th week of pregnancy, with approximately three cases per week being identified in clinics.

Understanding of Existing Challenges: Through interviews with doctors and clinicians, we have confirmed that while current methods are effective, they are time-consuming and rely heavily on clinical judgment. There is a lack of existing technological support to assist healthcare professionals in identifying the risk of pre-eclampsia early, which could greatly improve patient outcomes. Doctors have expressed that they would welcome a solution that could reduce their workload and enhance the accuracy of early risk identification.

Identification of Critical Risk Factors: Our research has led us to a deeper understanding of the indicators that healthcare providers rely on for pre-eclampsia diagnosis—high blood pressure above 140/90 mmHg, protein levels in urine, and pregnancy status. This has helped us focus our predictive model on the most relevant data points, allowing us to tailor the algorithm to the needs of Kenyan healthcare providers.

Validation of Solution Relevance: Through discussions with potential users, we validated the relevance of our solution. Healthcare providers have expressed interest in having a system that helps them assess the risk of pre-eclampsia early. They also confirmed that they would embrace an independent, tech-based solution integrated within the existing e-medical systems, provided it helps reduce their workload and does not complicate their workflows.

Chapter Nine: Funding/ Support Need

$550000 for 3 Years.

Platform Development: $100,000

- This includes building and refining the predictive model, creating the web-based platform, and integrating APIs for seamless data exchange. The costs will cover developer fees, cloud computing services, and testing across different devices and networks.

Staffing: $50,000

- Salaries for a team of developers, a project manager, and data scientists during the pilot phase, ensuring efficient development and initial deployment in a target healthcare facility.

Compliance and Data Security: $25,000

- This will cover legal fees for data protection compliance (e.g., Kenya Data Protection Act 2019), secure data storage, and encryption services, ensuring that patient information is safely managed

Hardware and Infrastructure: $30,000

- Purchase of essential hardware like servers, secure data storage solutions, and any additional networking tools for healthcare facilities that lack the required IT infrastructure.

Healthcare Provider Training: $25,000

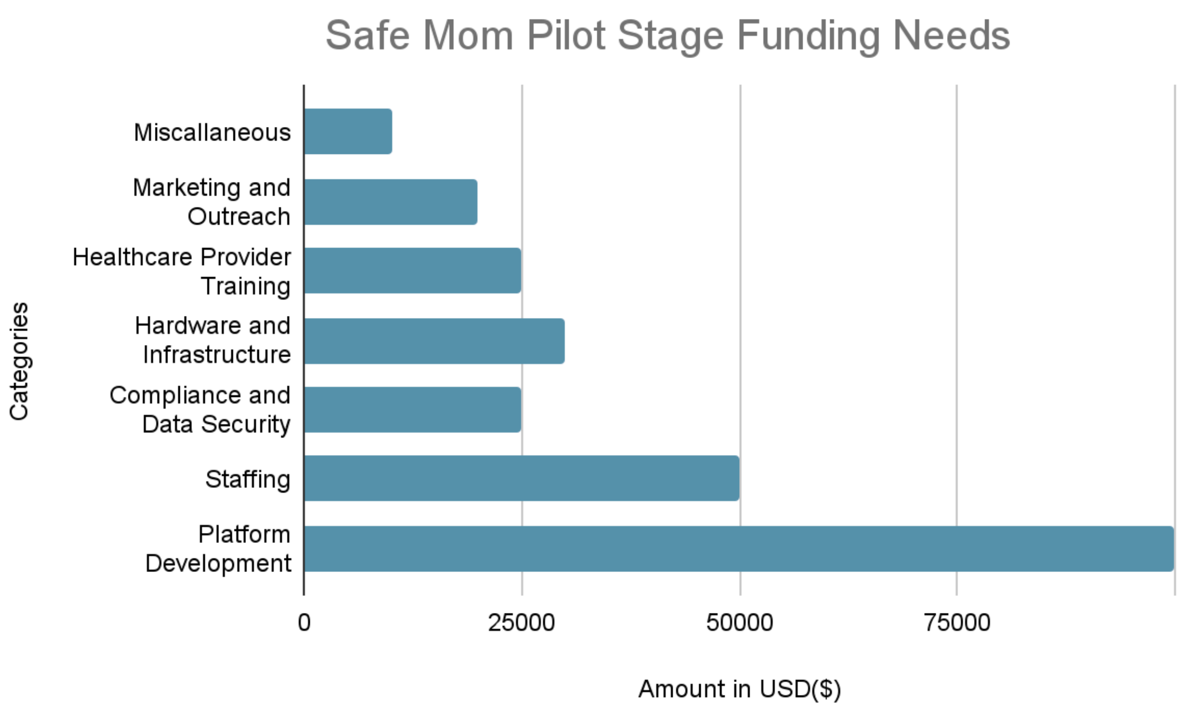
- Training programs for healthcare professionals using the system, including in-person workshops and creating comprehensive user manuals.

Marketing and Outreach: $20,000

- This will include digital marketing campaigns to raise awareness among healthcare providers, government agencies, and non-profits. This stage will also cover outreach efforts aimed at building partnerships with health ministries and potential donors.

Miscellaneous: $20,000

- Allowance for unexpected costs that may arise during the pilot phase, such as Hardware and Software upgrades, system upgrades or unforeseen legal requirements.



2. Scaling and Post-Pilot (Years 2–3)

Expansion to Additional Facilities: $120,000

- In the next two years, we plan to scale the platform to cover additional healthcare facilities. This will include infrastructure upgrades, deployment to rural and urban centers, and technical support for integrating the platform into new locations.

Further Platform Enhancements: $50,000

- Ongoing platform improvements include incorporating more advanced AI features, improving UI/UX design, and adapting to user feedback.

Operational Costs: $50,000 (per year)

- This will cover day-to-day expenses, including cloud services, server maintenance, and regular platform audits to ensure compliance and data integrity.

Ongoing Healthcare Provider Training: $20,000 (per year)

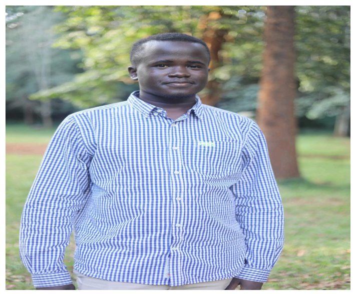
- Continued training programs to onboard new healthcare workers as the platform expands to more regions.

Compliance and Legal Updates: $20,000

- Ensuring that the platform complies with any updates in data protection laws and healthcare regulations.

Chapter Ten: The Team

To achieve our objective we have a team comprising Sebastian Chanzu (a data analyst), Denish Awajo (a software engineer), and Eddah Chepchirchir(an AI researcher).

		
Sebastian Chanzu **Data Analyst**	**Denish Awajo** **Software Engineer**	**Eddah Chepchirchir** **AI Researcher**