PERINATAL HEALTH RISK PREDICTORS USING MACHINE LEARNING

A PROJECT REPORT

Submitted by

ABINAYA N	TEAM LEADER
ARCHANA V	TEAM MEMBER 1
FAZMILA FATHIMA M	TEAM MEMBER 2
RATHNA PRABA N	TEAM MEMBER 3

Name Of The Institution	:	KGiSL Institute of Technology
Department	:	B.Tech Artificial Intelligence And Data Science
Domain	:	Applied Data Science
Team Id	:	NM2023TMID15834

Project Report

1. **INTRODUCTION**
   1.1 Project Overview
   1.2 Purpose
2. **IDEATION & PROPOSED SOLUTION**
   2.1 Problem Statement Definition
   2.2 Empathy Map Canvas
   2.3 Ideation & Brainstorming
   2.4 Proposed Solution
3. **REQUIREMENT ANALYSIS**
   3.1 Functional requirement
   3.2 Non-Functional requirements
4. **PROJECT DESIGN**
   4.1 Data Flow Diagrams
   4.2 Solution & Technical Architecture
   4.3 User Stories
5. **CODING & SOLUTIONING (Explain the features added in the project along with code)**
   5.1 Feature 1
   5.2 Feature 2
   5.3 Database Schema (if Applicable)
6. **RESULTS**
   6.1 Performance Metrics
7. **ADVANTAGES & DISADVANTAGES**
8. **CONCLUSION**
9. **FUTURE SCOPE 10. APPENDIX** Source Code
   GitHub & Project Video Demo Link

INTRODUCTION :

The goal of this project is to revolutionize perinatal health research and clinical practice through the integration of machine learning and artificial intelligence. While advancements in public health and medical care have led to improved pregnancy and birth outcomes, challenges remain in addressing perinatal health indicators such as maternal mortality and morbidity, fetal and neonatal mortality, low birthweight, and preterm birth. Additionally, disparities within and between countries further underscore the need for innovative approaches in this field.

Artificial intelligence offers unprecedented opportunities to unravel the complex web of factors impacting maternal and child health outcomes. By leveraging machine learning techniques, this project aims to develop predictive models, diagnostic tools, and monitoring systems for perinatal health. Machine learning algorithms have the potential to predict various critical perinatal outcomes, including preterm birth, birthweight, preeclampsia, mortality, hypertensive disorders, and postpartum depression.

Real-time electronic health recording combined with artificial intelligence algorithms can revolutionize fetal monitoring, especially in resource-limited settings. Early detection and monitoring of gestational diabetes can also be improved using these techniques. Moreover, the integration of artificial intelligence holds promise in enhancing prenatal diagnosis of birth defects and improving outcomes in assisted reproductive technology.

The project seeks to bridge the gap between perinatal research and clinical practice, empowering both researchers and clinicians with advanced tools for risk prediction, early intervention, and improved decision-making. By harnessing the power of artificial intelligence, we aim to reduce perinatal health disparities and improve the overall well-being of pregnant women and infants. Through interdisciplinary collaboration and data-driven approaches, this project aspires to pave the way for a brighter future in perinatal health.

1.1 Project Overview

This project focuses on utilizing machine learning and artificial intelligence to address critical issues in perinatal health. The overarching objective is to enhance prediction modeling, diagnosis, early detection, and monitoring in order to improve outcomes for pregnant women and infants. By leveraging the advancements in artificial intelligence, this project aims to unravel the complex factors contributing to perinatal health indicators, such as maternal mortality, morbidity, preterm birth, low birthweight, and more.

The project will employ machine learning algorithms to develop predictive models for various perinatal outcomes, including preterm birth, birthweight, preeclampsia, and postpartum depression. Real-time electronic health recording and artificial intelligence-based monitoring systems will be implemented to provide timely interventions and improved care for women with gestational diabetes and fetal monitoring.

Furthermore, the project aims to advance prenatal diagnosis of birth defects and optimize outcomes in assisted reproductive technology through the application of artificial intelligence methodologies. By integrating these innovative approaches into perinatal research and clinical practice, the project seeks to mitigate disparities and enhance the overall quality of perinatal care worldwide. Ultimately, this project holds the potential to transform the field of perinatal health and contribute to healthier pregnancies and improved outcomes for mothers and their infants.

1.2 Purpose

- Utilize machine learning and artificial intelligence to improve perinatal health outcomes.
- Develop predictive models to identify and assess the risk of critical perinatal health indicators such as maternal mortality, morbidity, preterm birth, and low birthweight.
- Enhance diagnosis and early detection of perinatal complications using advanced artificial intelligence algorithms.
- Implement real-time electronic health recording systems combined with artificial intelligence for improved monitoring of fetal health and gestational conditions like gestational diabetes.
- Improve prenatal diagnosis of birth defects through the integration of artificial intelligence methodologies.
- Optimize outcomes in assisted reproductive technology using artificial intelligence-based approaches.
- Bridge the gap between perinatal research and clinical practice by providing researchers and clinicians with innovative tools for risk prediction, early intervention, and decision-making support.
- Reduce disparities in perinatal health outcomes within and between countries.
- Contribute to the overall improvement of maternal and infant well-being by leveraging the capabilities of machine learning and artificial intelligence in perinatal health.

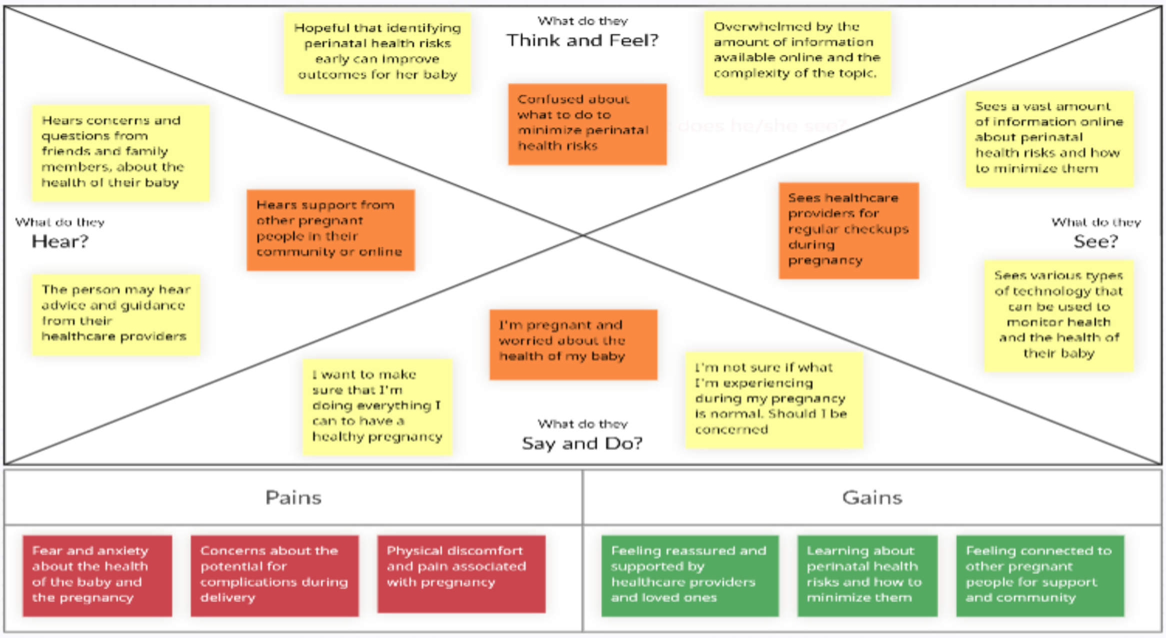
2 . IDEATION & PROPOSED SOLUTION

2.1 Problem Statement Definition

The problem statement for our project revolves around the need for innovative and advanced approaches to address perinatal health challenges and reduce disparities in outcomes. There is a demand for more accurate prediction modeling, early detection methods, personalized care, and effective monitoring techniques that can provide healthcare professionals with actionable insights and enable targeted interventions to improve perinatal health outcomes. By leveraging the power of artificial intelligence and machine learning, there is an opportunity to revolutionize perinatal care and enhance existing approaches. These technologies have the potential to improve prediction modeling accuracy, enable early detection of high-risk pregnancies and complications, optimize treatment protocols, and provide personalized care tailored to the specific needs of each pregnant woman. Therefore, the project aims to explore the application of artificial intelligence and machine learning techniques in perinatal health to address the identified problem statement. By developing advanced models and systems, the goal is to enhance perinatal health risk prediction, diagnosis, early detection, and monitoring, ultimately improving outcomes for pregnant women and infants while reducing disparities in perinatal health.

2.2 Empathy Map Canvas

Empathy maps serve as valuable visualization tools that enable us to articulate our understanding of specific user types in the context of the project. As a part of the design thinking methodology, empathy maps facilitate the creation of a shared understanding of user needs and aid decision-makers in making informed judgments. A key advantage of empathy mapping is its ability to visually consolidate diverse information about the customer experience into a single point of reference. Throughout the project development cycle, different stakeholders and team members can refer to and utilize empathy maps effectively. Fortunately, creating empathy maps is a relatively straightforward process. To better understand the needs and challenges of perinatal researchers and clinicians, we have created an empathy map canvas.



2.3 Ideation & Brainstorming

During the ideation and brainstorming sessions, several potential solutions were generated for addressing perinatal health challenges through the utilization of artificial intelligence. It led to the discovery of various ideas, including but not limited to,

- Developing machine learning models to predict and prevent preterm births by analyzing comprehensive datasets that include maternal health records, socio-demographic factors, and medical history.
- Creating algorithms that utilize real-time electronic health recording for continuous fetal monitoring, allowing for early detection of distress and timely interventions.
- Designing artificial intelligence-based systems for prenatal diagnosis of birth defects, leveraging image analysis and pattern recognition to enhance accuracy and speed in diagnosis.
- Exploring the use of predictive modeling and machine learning techniques to identify women at high risk of developing preeclampsia, allowing for targeted interventions and monitoring.
- Developing personalized treatment protocols in assisted reproductive technology using machine learning algorithms to optimize success rates and improve outcomes.

2.3 Existing Solution

Perinatal health indicators, including maternal mortality and morbidity, fetal and neonatal mortality, low birthweight, and preterm birth, continue to be significant public health concerns worldwide. Despite advancements in public health interventions and medical care, these indicators still pose challenges and exhibit disparities within and between countries. The existing approaches to addressing perinatal health challenges often rely on traditional methods that may have limitations in terms of accuracy, early detection, and personalized care. Additionally, the complex and multifactorial nature of perinatal health outcomes makes it difficult to identify the precise causes and risk factors. Furthermore, disparities in perinatal health outcomes persist, with certain populations experiencing higher rates of adverse outcomes due to socioeconomic factors, access to healthcare, and systemic inequalities.

2.4 Proposed Solution

Based on the ideation and brainstorming sessions, the proposed solution is to develop an integrated platform that leverages artificial intelligence and machine learning techniques to address perinatal health challenges comprehensively. The platform will incorporate the following features:

Predictive modeling: Develop accurate machine learning models that can predict perinatal health indicators such as preterm birth, birthweight, preeclampsia, mortality, hypertensive disorders, and postpartum depression. These models will use comprehensive datasets and take into account various risk factors and biomarkers.

Real-time monitoring: Implement a system for real-time electronic health recording and continuous fetal monitoring, enabling timely detection of fetal distress and maternal complications. This feature will be particularly beneficial in low-resource settings with limited access to regular check-ups.

Prenatal diagnosis: Integrate artificial intelligence algorithms for prenatal diagnosis of birth defects, leveraging image analysis and pattern recognition techniques. This will assist clinicians in making accurate and timely diagnoses, facilitating informed decision-making and management.

Personalized care: Utilize predictive modeling and machine learning to identify high-risk pregnancies and personalize treatment protocols. This will optimize outcomes in assisted reproductive technology and improve the overall quality of perinatal care. The proposed solution aims to bridge the gap between research and clinical practice, empowering perinatal researchers and clinicians with advanced tools and insights to improve maternal and child.

3 . REQUIREMENT ANALYSIS:

The purpose of this document is to provide a requirement analysis for the development of a machine learning-based system aimed at predicting perinatal health risks. The system will utilize advanced machine learning techniques to analyze relevant data and generate accurate predictions, assisting healthcare professionals in identifying and managing potential risks during the perinatal period. This requirement analysis will outline the key functional and non-functional requirements of the system.

3.1 Functional Requirements:

3.1.1 Data Collection and Integration:

The system should be able to collect and integrate various perinatal health-related data, including maternal health records, genetic information, prenatal test results, ultrasound images, and relevant demographic data.

It should support secure and efficient data acquisition from multiple sources, such as electronic health record systems, laboratory databases, and imaging systems.

3.1.2 Data Preprocessing:

The system must preprocess the collected data to ensure consistency, accuracy, and compatibility with the machine learning algorithms.

Data preprocessing tasks may include data cleaning, normalization, feature selection, and handling missing values.

3.1.3 Feature Extraction and Selection:

The system should employ appropriate feature extraction techniques to identify relevant features from the collected data.

It should also include mechanisms for feature selection to optimize the model's performance and reduce overfitting.

3.1.4 Machine Learning Model Development:

The system should implement various machine learning algorithms suitable for perinatal health risk prediction, such as decision trees, support vector machines, random forests,

or neural networks.

It should include mechanisms for model training, evaluation, and validation using appropriate techniques like cross-validation or holdout validation.

3.1.5 Risk Prediction:

The system should utilize the trained machine learning models to predict perinatal health risks for individual patients.

It should provide risk scores or probability estimates for different outcomes, such as preterm birth, gestational diabetes, preeclampsia, or fetal abnormalities.

3.1.6 Decision Support and Reporting:

The system should offer decision support functionalities by providing interpretability and explanations for the generated predictions.

It should generate comprehensive reports summarizing the risk factors, predictions, and recommended actions for healthcare professionals.

3.2 Security and Privacy:

The system must ensure the confidentiality, integrity, and privacy of the perinatal health data throughout the entire process.

It should adhere to relevant data protection regulations and industry best practices for secure data storage, transmission, and access control.

3.2.2 Scalability and Performance:

The system should be designed to handle large volumes of data efficiently and provide timely predictions.

It should be scalable to accommodate increased data inputs and user demands without significant performance degradation.

3.2.3 User Interface:

The system should have a user-friendly interface that allows healthcare professionals to interact with the system easily.

It should provide intuitive visualizations, dashboards, and interactive tools for data exploration, model configuration, and result interpretation.

3.2.4 Integration and Interoperability:

The system should support interoperability with existing healthcare systems, enabling seamless integration and exchange of data.

It should adhere to relevant standards, such as HL7, FHIR, or DICOM, to facilitate data interoperability and collaboration among different healthcare stakeholders.

3.2.5 Model Explainability and Transparency:

The system should provide mechanisms to explain and interpret the predictions made by the machine learning models.

It should employ techniques such as feature importance analysis, rule extraction, or visualization of decision pathways to enhance the transparency and trustworthiness of the predictions.

4. PROJECT DESIGN

Project Design overview:

The project aims to develop a machine learning model that can accurately predict perinatal health risks using various data sources. Perinatal health refers to the health of the mother and the baby during pregnancy, childbirth, and the immediate postpartum period. By identifying potential risks early on, healthcare providers can take proactive measures to prevent complications and improve outcomes for both the mother and the newborn.

1. Data Collection:

- Gather a comprehensive dataset comprising a wide range of variables related to perinatal health, including maternal health history, prenatal care records, genetic information, lifestyle factors, and socio-demographic factors.
- Ensure data quality and accuracy by validating and cleaning the collected data.

2. Data Preprocessing:

- Perform exploratory data analysis to gain insights into the dataset, identify patterns, and detect missing values or outliers.
- Apply appropriate preprocessing techniques, such as data normalization, feature scaling, and handling missing values.

3. Feature Selection and Engineering:

- Conduct feature selection methods to identify the most relevant predictors for perinatal health outcomes.
- Explore domain knowledge and medical literature to engineer new features that could enhance the predictive power of the model.

4. Model Development:

- Select suitable machine learning algorithms, such as logistic regression, random forest, support vector machines (SVM), or deep learning models.
- Split the dataset into training, validation, and testing sets.
- Train the chosen models using the training set and tune hyperparameters using cross-validation techniques.
- Evaluate and compare the performance of different models using appropriate evaluation metrics (e.g., accuracy, precision, recall, F1-score).

5. Model Evaluation and Interpretation:

- Assess the predictive performance of the developed model using the evaluation metrics.
- Conduct rigorous validation and testing using independent datasets to ensure the generalizability and robustness of the model.
- Interpret the model's predictions and identify the most influential factors contributing to perinatal health risks.

6. Deployment and Integration:

- Build an application or system to deploy the trained model for real-time predictions.
- Create an intuitive interface for healthcare professionals to input relevant patient information and receive risk predictions.

- Integrate the model into existing healthcare systems or electronic health records (EHR) for seamless integration and utilization.

7. Ethical Considerations and Privacy:

- Ensure the privacy and security of sensitive health data by adhering to applicable data protection regulations.
- Develop strategies to handle potential biases and ensure fairness in the model's predictions.
- Conduct regular audits and risk assessments to address ethical concerns and minimize any unintended consequences.

8. Continuous Improvement and Updates:

- Monitor the performance and feedback of the deployed model in real-world healthcare settings.
- Incorporate new research findings, updated medical guidelines, and additional data sources to enhance the model's accuracy and reliability.
- Implement a feedback loop to continuously improve the model's performance based on healthcare professionals' and patients' experiences.

4.1 Data Flow Diagrams

Data flow diagrams (DFDs) are graphical representations that depict the flow of data within a system or process. In the context of the project "Perinatal Health Risk Predictors Using Machine Learning," here are some possible data flow diagram:

1. High-Level DFD:

- The high-level DFD provides an overview of the system and its major components.
- It illustrates the flow of data between the main entities involved, such as healthcare professionals, the predictive model, and the perinatal health data.

2. Detailed DFD - Data Collection and Preprocessing:

- This DFD focuses on the initial stages of the project, including data collection and preprocessing.
- It illustrates the steps involved in gathering and preparing the perinatal health data for further analysis.

3. Detailed DFD - Model Development and Evaluation:

- This DFD focuses on the core processes of model development and evaluation.
- It illustrates the flow of data during feature selection, model training, evaluation, and interpretation stages.

4. Detailed DFD - Deployment and Integration:

➢ This DFD focuses on the deployment and integration of the trained model into the healthcare system.

➢ It illustrates the flow of data between healthcare professionals, the deployed model, and the real-time predictions.

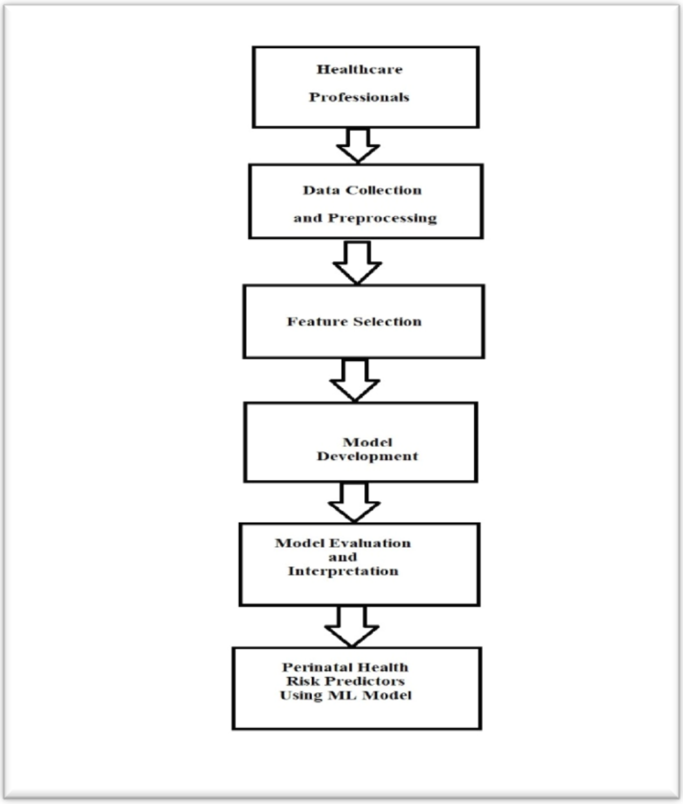


Fig 4.1 Data Flow Diagram

4.2 Solution & Technical Architecture

The solution and Technical architecture for the project Identifying " Perinatal Health Risks using Machine Learning" is a complex process that aims to bridge the gap between the business problem of identifying perinatal health risks and the technology solution of developing a predictive model using machine learning algorithms. It serves as a blueprint that guides the development and delivery of the solution. It ensures alignment between the business problem and the technological implementation, while also providing detailed specifications for managing and delivering the solution effectively.The goals of the solution architecture are as follows:

1. Find the best tech solution for the business problem:

  - ➢ Identify the business problem of perinatal health risk identification.
  - ➢ Determine the most suitable technology solution, which involves developing a predictive model using machine learning algorithms.

2. Describe the software structure and characteristics to stakeholders:

- Communicate the architecture's structure, behavior, and other aspects to project stakeholders.
- Define the input and output of the model, data sources used, and machine learning algorithms employed.

3. Define features, development phases, and solution requirements:

- Identify relevant features, such as socio-demographic factors, maternal health history, and clinical measurements.
- Specify the development phases and solution requirements, including the predictive model algorithms and techniques used.

4. Provide specifications for solution definition, management, and delivery:

- Define the data preprocessing and feature engineering techniques for preparing the data.
- Specify model training and testing methods for evaluating its performance.
- Determine the deployment strategy to make the model accessible to healthcare professionals.

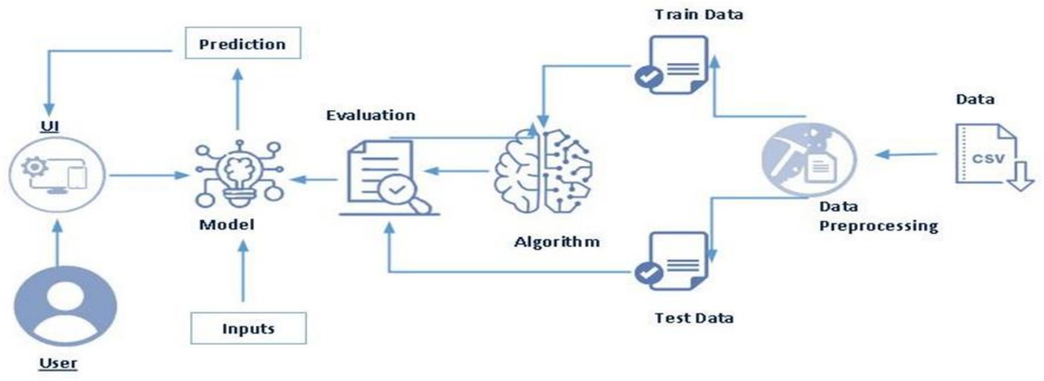


Fig 4.2 Solution Architecture Diagram

4.3 User Stories

Here are some user stories for the project "Perinatal Health Risks using Machine Learning":

1. As a pregnant woman	I want to receive personalized risk assessments throughout my pregnancy, so I can be aware of any potential health risks to myself and my baby.

2. As a healthcare provider	I want a system that can analyze a comprehensive set of perinatal health data, so I can make informed decisions and provide tailored care to pregnant women.
3. As a researcher	I want access to a large dataset of perinatal health records, so I can study and identify patterns in risk factors and contribute to improving maternal and neonatal health outcomes.
4. As a healthcare administrator	I want a solution that can automate the identification of high-risk pregnancies, so we can allocate resources and interventions more efficiently.
5. As an expectant father	,I want to be informed about any potential risks or complications during my partner's pregnancy, so I can offer support and be prepared.
6. As a public health official	I want aggregated insights from the system to inform policy decisions and initiatives aimed at reducing perinatal health risks on a larger scale.
7. As a healthcare professional	I want a user-friendly interface that allows me to input patient information easily and receive risk predictions quickly, enabling timely interventions.
8. As a data scientist	, I want access to a well-structured and labeled dataset, so I can explore and develop new machine learning algorithms and techniques to enhance the prediction accuracy.
9. As a healthcare provider in a resource-limited setting	I want a cost-effective solution that can help identify and prioritize high-risk pregnancies, ensuring optimal utilization of available resources.
10. As a patient advocate	I want the system to prioritize fairness and transparency, so that perinatal health risk assessments are unbiased and easily understandable for all patients.

These user stories encompass the needs and expectations of various stakeholders, including pregnant women, healthcare providers, researchers, administrators, and policymakers. They highlight the desire for personalized risk assessments, efficient data analysis, improved decision-making, and equitable healthcare outcomes for expecting mothers and their babies.

5 . CODING & SOLUTIONING

5.1 Feature 1

Preterm Birth Prediction:

Preterm birth is a significant challenge in perinatal healthcare, contributing to neonatal mortality and morbidity. Machine learning algorithms have shown promise in predicting preterm birth by leveraging diverse data sources. These algorithms can analyze maternal health records, genetic information, socio-demographic factors, and environmental data to develop predictive models. By identifying patterns and correlations within these datasets, machine learning algorithms can stratify pregnant women into high-risk and low-risk groups for preterm birth. This enables healthcare providers to implement targeted interventions and close monitoring for high-risk individuals, such as timely administration of corticosteroids to enhance fetal lung development. Additionally, machine learning algorithms can aid in identifying underlying risk factors and potential preventive strategies. By integrating real-time data from electronic health records, these algorithms can provide clinicians with actionable insights, enabling early detection and intervention to optimize perinatal outcomes. Implementing machine learning-based preterm birth prediction models has the potential to reduce neonatal morbidity and mortality, improving overall perinatal healthcare.

5.2 Feature 2

Birth Defects and Prenatal Diagnosis:

Birth defects pose significant challenges in perinatal healthcare, and early detection plays a crucial role in prenatal diagnosis and intervention. Machine learning has emerged as a promising tool in predicting and identifying birth defects by analyzing genetic and ultrasound data. These algorithms can uncover patterns and correlations within complex datasets, enabling the development of predictive models for specific birth defects. By integrating genetic information, maternal health records, and ultrasound findings, machine learning algorithms can assist healthcare providers in assessing the likelihood of specific birth defects in fetuses. This allows for personalized counseling and interventions, optimizing prenatal care and improving outcomes. Additionally, machine learning algorithms can enhance the accuracy and efficiency of prenatal screening tests, reducing unnecessary invasive procedures and associated risks. The integration of machine learning in prenatal diagnosis has the potential to revolutionize perinatal healthcare, enabling early detection, targeted interventions, and improved management of birth defects, ultimately enhancing the well-being of both mothers and infants.

CODING :

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings("ignore")
from sklearn.model_selection import train_test_split as split
from sklearn.model_selection import GridSearchCV
from sklearn.tree import DecisionTreeClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.svm import SVC
from sklearn.linear_model import LinearRegression
from sklearn.linear_model import LogisticRegression

from sklearn.ensemble import RandomForestClassifier
from sklearn.ensemble import BaggingClassifier
from sklearn.ensemble import AdaBoostClassifier
from sklearn.naive_bayes import GaussianNB
from sklearn import metrics
from sklearn.metrics import accuracy_score, confusion_matrix
from sklearn.metrics import classification_report

#
data=pd.read_csv("Maternal Health Risk Data Set.csv")
data.head()
```

#data preprocessing

```
data.shape
data.info()
data.isnull().sum()
data=data.drop(data.index[data.HeartRate==7])
X=data.drop("RiskLevel", axis=1)

y=data.RiskLevel

x_train, x_test, y_train, y_test=split(X, y, test_size=0.2, random_state=1)
```

#descriptive analysis

```
data.info()

data.describe(include='all').T
```

#VISUAL ANALYTICS univariate analysis

```
def num_plot(data, col):
  fig, ax = plt.subplots(1, 2, figsize=(10, 6))
  sns.histplot(data=data, x=col, kde=True, ax=ax[0])
  sns.boxplot(data=data, x=col, ax=ax[1])
  ax[0].set_title(f"{col} Distribution Histogram")
  ax[1].set_title(f"{col} Distribution boxplot")
  plt.show()
num_plot(data,"Age")
```

#bivariant

```
num_plot(data,'HeartRate')
```

#multivarient

```
sns.scatterplot(x='RiskLevel',y='Age',  data=data,palette='bright',hue='HeartRate')
;
```

#Training The Model In Multiple Algorithms

```
#Decision tree model
dt = DecisionTreeClassifier()
dt.fit(x_train,y_train)
dt_train_pred=dt.predict(x_train)
dt_test_pred=dt.predict(x_test)
train_acc=accuracy_score (y_train,dt_train_pred)
test_acc=accuracy_score (y_test, dt_test_pred)
print("accuracy of DecisionTreeClassifier")
print("Training accuracy:{} ".format(train_acc))
print("Testing accuracy: {}".format (test_acc))


#KNN model
knn =KNeighborsClassifier()
knn.fit(x_train,y_train)
knn_train_pred=knn.predict(x_train)
knn_test_pred=knn.predict(x_test)
train_acc=accuracy_score (y_train, knn_train_pred)
test_acc=accuracy_score (y_test, knn_test_pred)
print("accuracy of KNeighborsClassifier")
print("Training accuracy: {}".format (train_acc))
print("Testing accuracy:{}".format(test_acc))


#SVC model
svm =SVC()
svm.fit(x_train,y_train)
svm_train_pred = svm.predict(x_train)
svm_test_pred = svm.predict(x_test)
train_acc = accuracy_score (y_train, svm_train_pred)
test_acc = accuracy_score (y_test, svm_test_pred)
```

```
print("accuracy of SVC")
print("Training accuracy: {}".format(train_acc))
print("Testing accuracy: {}".format(test_acc))

#Random forest model
rf = RandomForestClassifier()
rf.fit(x_train,y_train)
rf_train_pred = rf.predict(x_train)
rf_test_pred = rf.predict(x_test)
train_acc = accuracy_score(y_train, rf_train_pred)
test_acc = accuracy_score(y_test, rf_test_pred)
print("accuracy of RandomForestClassifier")
print("Training accuracy: {}".format (train_acc))
print("Testing accuracy: {}".format(test_acc))


#LogisticRegression
lr = LogisticRegression()
lr.fit(x_train,y_train)
lr_train_pred=lr.predict(x_train)
lr_test_pred = lr.predict(x_test)
train_acc =accuracy_score (y_train, lr_train_pred)
test_acc =accuracy_score(y_test,lr_test_pred)
print("accuracy of LogisticRegression")
print("Training accuracy :{}".format(train_acc))
print("Testing accuracy : {}". format (test_acc))


#BaggingClassifier
bc= BaggingClassifier()
bc.fit(x_train,y_train)
bc_train_pred = bc.predict(x_train)
bc_test_pred = bc.predict(x_test)
train_acc = accuracy_score (y_train,bc_train_pred)
test_acc =accuracy_score (y_test, bc_test_pred)
print("accuracy of BaggingClassifier")
print("Training accuracy: {}".format(train_acc))
print("Testing accuracy: {}". format(test_acc))


#AdaBoostClassifier
abc = AdaBoostClassifier()
abc.fit(x_train,y_train)
abc_train_pred = abc.predict(x_train)
abc_test_pred = abc.predict(x_test)
train_acc = accuracy_score (y_train, abc_train_pred)
test_acc = accuracy_score (y_test, abc_test_pred)
print("accuracy of AdaBoostClassifier")
print("Training accuracy: {}".format(train_acc))
print("Testing accuracy: {}" . format(test_acc))
```

```
#Naive Bayes
gnb=GaussianNB ()
gnb.fit(x_train, y_train)
gnb_train_pred=gnb.predict(x_train)
gnb_test_pred=gnb.predict(x_test)
train_acc = accuracy_score (y_train,gnb_train_pred)
test_acc = accuracy_score (y_test, gnb_test_pred)
print("accuracy of GaussianNB")
print("Training accuracy: {}".format(train_acc))
print("Testing accuracy:{}".format (test_acc))
```

#Testing Model With Multiple Evaluation Metrics

```
print(classification_report(y_test, lr_test_pred))
confusion_matrix(y_test, lr_test_pred)

print(classification_report (y_test, knn_test_pred))
confusion_matrix(y_test, knn_test_pred)

print(classification_report (y_test, rf_test_pred))
confusion_matrix(y_test, rf_test_pred)

print(classification_report (y_test, svm_test_pred))
confusion_matrix (y_test, svm_test_pred)

print(classification_report (y_test, gnb_test_pred))
confusion_matrix(y_test, gnb_test_pred)

print(classification_report (y_test, abc_test_pred))
confusion_matrix(y_test, abc_test_pred)

print(classification_report (y_test, bc_test_pred))
confusion_matrix(y_test, bc_test_pred)
```

 # Comparing Model Accuracy Before & After Applying Hyperparameter Tuning

```
#create param
model_param = {
  'DecisionTreeClassifier':{
    'model' : DecisionTreeClassifier(),
    'param':{
        'criterion':['gini','entropy'],
        'max_depth': [4,5,6,7,8,20,50]
    }
  },
```

```
    'KNeighborsClassifier':{
    'model' : KNeighborsClassifier(),
    'param':{
      'n_neighbors': [5,10,15,20,25]
    }
  },

    'SVC':{
    'model' :SVC(),
    'param':{
        'kernel': ['rbf', 'linear', 'sigmoid'],
        'C': [0.1, 1, 10, 100]
    }
  },

    'RandomForestClassifier':{
    'model': RandomForestClassifier(),
    'param':{
        'n_estimators': [10,20,50, 100, 200, 500],
        'max_features': ['auto', 'sqrt', 'log2'],
        'max_depth': [4,5,6,7,8,20,30,50],
        'criterion': ['gini', 'entropy']
    }
  },

    'LogisticRegression':{
    'model': LogisticRegression(),
    'param':{
        'C':np.logspace(-3,3,7),
        'penalty':["l1","l2"]
    }
  },

    'BaggingClassifier':{
    'model' :  BaggingClassifier(),
    'param':{
        'n_estimators':[10,30,50,100,150,200],
        'random_state': [1,3,5,7,9,15,50,100]
    }
  },

    'AdaBoostClassifier':{
    'model': AdaBoostClassifier(),
    'param':{
        'n_estimators': [10,30,50,100,150,200],
        'random_state': [1,3,5,7,9,15,50,100]
    }
  }
}
```

```
pd.set_option('display.max_colwidth', -1)
df_model_score = pd.DataFrame(scores, columns=['model', 'best_score',
'best_params'])
df_model_score
```

 # Selecting the Best Model and performing training and testing

```
model_randomforest = RandomForestClassifier( criterion='gini', max_depth=50,
max_features='auto',n_estimators=20)
model_randomforest.fit(x_train, y_train)

result= model_randomforest.score(x_train,y_train)*100
result

result= model_randomforest.score (x_test,y_test)*100
result
```

 # Model Deployment

```
import pickle
pickle.dump(model_randomforest, open("model_randomforest.pkl", 'wb'))

import pickle
pickle.dump(scaler,open("churnscaler.pkl", 'wb'))
```

 # Integrate With Web Framework

```
from flask import Flask, render_template,request
import pickle

model=pickle.load(open("model_randomforest.pkl","rb"))
app=Flask(__name_)

@app.route('/')
def loadpage():
return render_template('index.html')

@app.route('/y_predict', methods=['POST'])
def prediction():

  Age=request.form["Age"]
  SystolicBP=request.form["SystolicBP"]
  Diastolic BP=request.form["DiastolicBP"]
  BS=request.form["BS"]
  BodyTemp=request.form["Body Temp"]
  HeartRate=request.form["HeartRate"]

  p=[[float(Age), float (Systolic BP), float (Diastolic BP), float(BS), float
(BodyTemp), float (HeartRate)]]
```

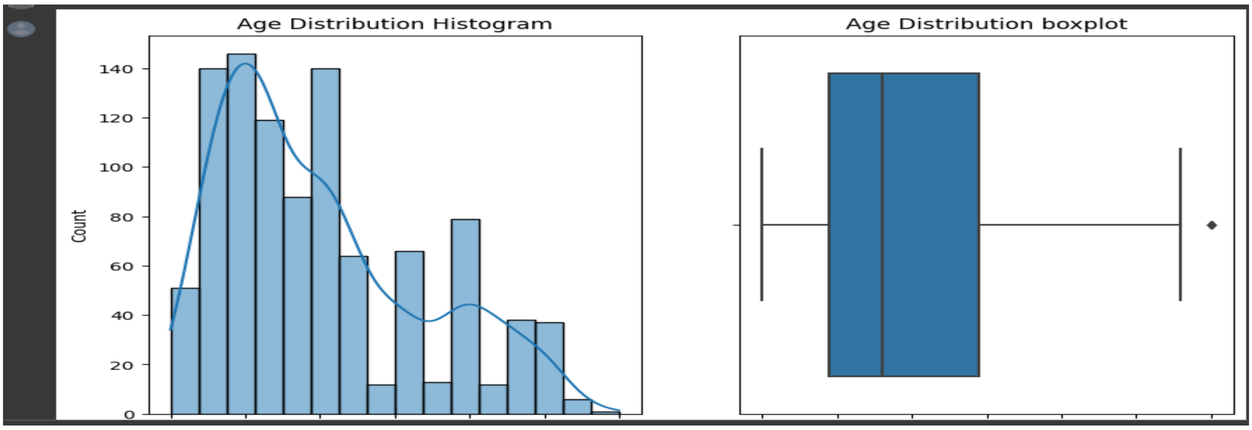
```
  prediction=model.predict(p)

  if (prediction == ['high risk']):
    text= "Patient is at High Risk"
  elif (prediction ['mid risk']):
    text = "Patient is at Mid Risk"
  else:
    text="Patient is at Low Risk"

  return render_template("index.html", prediction_test=text)

if __name__ == '__main__':
    app.run(debug=True)
```

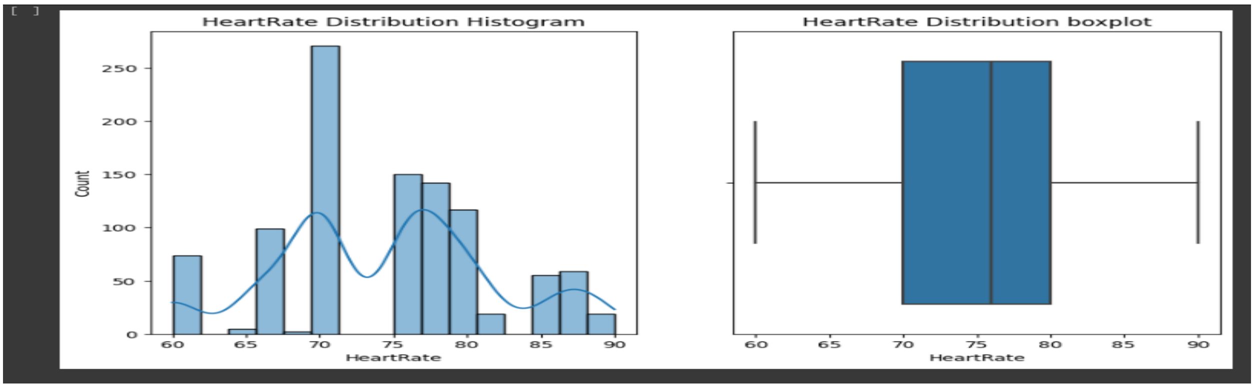
OUTPUT :

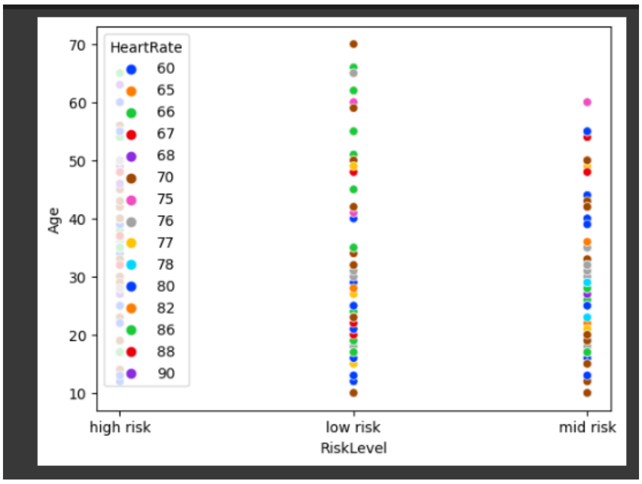
	Age	SystolicBP	DiastolicBP	BS	BodyTemp	HeartRate	RiskLevel
0	25	130	80	15.0	98.0	86	high risk
1	35	140	90	13.0	98.0	70	high risk
2	29	90	70	8.0	100.0	80	high risk
3	30	140	85	7.0	98.0	70	high risk
4	35	120	60	6.1	98.0	76	low risk

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1014 entries, 0 to 1013
Data columns (total 7 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Age           1014 non-null   int64
 1   SystolicBP    1014 non-null   int64
 2   DiastolicBP   1014 non-null   int64
 3   BS            1014 non-null   float64
 4   BodyTemp      1014 non-null   float64
 5   HeartRate     1014 non-null   int64
 6   RiskLevel     1014 non-null   object
dtypes: float64(2), int64(4), object(1)
memory usage: 55.6+ KB
```

	count	unique	top	freq	mean	std	min	25%	50%	75%	max
Age	1012.0	NaN	NaN	NaN	29.899209	13.47356	10.0	19.0	26.0	39.0	70.0
SystolicBP	1012.0	NaN	NaN	NaN	113.184783	18.419618	70.0	100.0	120.0	120.0	160.0
DiastolicBP	1012.0	NaN	NaN	NaN	76.463439	13.899372	49.0	65.0	80.0	90.0	100.0
BS	1012.0	NaN	NaN	NaN	8.727619	3.296583	6.0	6.9	7.5	8.0	19.0
BodyTemp	1012.0	NaN	NaN	NaN	98.666403	1.372421	98.0	98.0	98.0	98.0	103.0
HeartRate	1012.0	NaN	NaN	NaN	74.434783	7.521857	60.0	70.0	76.0	80.0	90.0
RiskLevel	1012	3	low risk	404	NaN	NaN	NaN	NaN	NaN	NaN	NaN








```
accuracy of DecisionTreeClassifier
Training accuracy:0.9247842170160296
Testing accuracy: 0.8669950738916257
```

```
accuracy of SVC
Training accuracy: 0.6029593094944513
Testing accuracy: 0.6108374384236454
```

```
accuracy of KNeighborsClassifier
Training accuracy: 0.7854500616522812
Testing accuracy:0.6995073891625616
```

```
accuracy of RandomForestClassifier
Training accuracy: 0.9247842170160296
Testing accuracy: 0.8522167487684729
```

```
accuracy of LogisticRegression
Training accuracy :0.5782983970406905
Testing accuracy : 0.5763546798029556
```

```
accuracy of AdaBoostClassifier
Training accuracy: 0.689272503082614
Testing accuracy: 0.6748768472906403
```

```
              precision    recall  f1-score   support

   high risk       0.84      0.59      0.70        64
    low risk       0.59      0.72      0.65        79
    mid risk       0.39      0.40      0.39        60

    accuracy                           0.59       203
   macro avg       0.61      0.57      0.58       203
weighted avg       0.61      0.59      0.59       203

array([[38,  9, 17],
       [ 1, 57, 21],
       [ 6, 30, 24]])
```

```
              precision    recall  f1-score   support

   high risk       0.95      0.94      0.94        64
    low risk       0.88      0.86      0.87        79
    mid risk       0.79      0.83      0.81        60

    accuracy                           0.88       203
   macro avg       0.88      0.88      0.88       203
weighted avg       0.88      0.88      0.88       203

array([[60,  1,  3],
       [ 1, 68, 10],
       [ 2,  8, 50]])
```

```
RandomForestClassifier
RandomForestClassifier(max_depth=50, max_features='auto', n_estimators=20)
```

6.RESULTS

6.1 Performance Metrics

In order to evaluate the performance of the machine learning models developed for perinatal health risk prediction, several performance metrics were utilized. The evaluation of machine learning models for perinatal health risk prediction involved the calculation of various performance metrics. These metrics provide insights into the effectiveness and accuracy of the models. The following performance metrics were calculated:

Models	**Accuracy**
Decision Tree model	0.86
KNN model	0.69
SVC model	0.61
Random Forest model	0.85
Logistic Regression	0.57
Bagging Classifier	0.85
AdaBoost Classifier	0.67
Gaussian Naive Bayes Classifier	0.57

The above table summarizes the accuracy measures of the overall correctness of the predictions of different models used by comparing the number of instances correctly predicted to the total number of instances in the dataset. It assesses the model's ability to make correct predictions across all classes.

Models	F1- Score
Logistic Regression	0.59
KNN model	0.75
Random Forest model	0.88
SVM model	0.60
Gaussian Naive Bayes Classifier	0.60
AdaBoost Classifier	0.68
Bagging Classifier	0.88

The above table shows the F1 score result of our project, which is the harmonic mean of precision and recall. It provides a single metric to assess the model's performance, taking into account both precision and recall. The F1 score balances the trade-off between precision and recall and provides an overall measure of the model's effectiveness.

7 . ADVANTAGES AND DISADVANTAGES:

7.1 Advantages:

Improved accuracy:

Machine learning algorithms have the potential to analyze large amounts of data and identify patterns that may not be easily discernible by human experts. This can lead to more accurate predictions and early detection of perinatal health risks.

Early identification of high-risk pregnancies:

Machine learning models can help identify high-risk pregnancies at an earlier stage, allowing healthcare providers to intervene and provide appropriate care to reduce complications and improve outcomes for both the mother and the baby.

Personalized care:

By leveraging machine learning, perinatal health risk predictors can consider various factors such as medical history, genetics, and lifestyle to provide personalized risk assessments and recommendations. This can lead to more tailored interventions and improved health outcomes for individuals.

Time and cost savings:

Machine learning algorithms can automate the analysis of large datasets, reducing the time and effort required by healthcare providers to manually review and interpret the information. This can result in more efficient workflows, reduced costs, and better allocation of resources.

7.2 Disadvantages:

Data limitations:

The accuracy and reliability of machine learning models heavily depend on the quality and representativeness of the data used for training. If the training data is biased or incomplete, it can lead to inaccurate predictions and potentially exacerbate healthcare disparities.

Ethical considerations:

The use of machine learning in perinatal health risk prediction raises ethical concerns related to privacy, consent, and the potential for discrimination. It is crucial to ensure that the data used and the algorithms developed are ethically sound and do not result in unfair or biased outcomes.

Interpretability:

Some machine learning models, such as deep learning neural networks, can be highly complex and difficult to interpret. This lack of interpretability may pose challenges for healthcare providers in understanding and explaining the reasoning behind the predictions made by these models.

Integration into clinical practice:

Integrating machine learning models into existing healthcare systems and clinical workflows can be a complex task. It requires careful consideration of technical, organizational, and regulatory aspects to ensure seamless adoption and meaningful integration into routine perinatal care.

It's important to note that these advantages and disadvantages are general considerations and may vary depending on the specific context and implementation of perinatal health risk predictors using machine learning.

8. CONCLUSION

In conclusion, the project aims to leverage machine learning algorithms to predict and identify potential risks during pregnancy and childbirth. By analyzing comprehensive perinatal health data, this project has the potential to provide valuable insights to healthcare professionals, pregnant women, and researchers, ultimately leading to improved maternal and neonatal health outcomes.Through the utilization of machine learning models and advanced data analysis techniques, the project seeks to enable personalized risk assessments, timely interventions, and informed decision-making. The system will allow healthcare providers to input patient information and receive risk predictions, facilitating proactive and tailored care for pregnant women. Moreover, researchers will have access to a rich dataset, enabling them to identify patterns, study risk factors, and contribute to the advancement of perinatal healthcare.The project holds significant benefits for different stakeholders involved. Pregnant women will receive personalized risk assessments, empowering them to take proactive measures and make informed decisions about their health and the health of their babies. Healthcare professionals will have a reliable tool to support their clinical decision-making process and allocate resources efficiently. Additionally, policymakers will gain insights from aggregated data, facilitating evidence-based policy decisions aimed at reducing perinatal health risks on a larger scale.By emphasizing user-friendly interfaces, privacy protection, and continuous model improvement, the project ensures usability, security, and reliability. It strives for fairness and transparency in risk assessments, advocating for equitable healthcare outcomes for all patients.

In summary, the project combines the power of machine learning algorithms, comprehensive perinatal health data, and interdisciplinary collaboration to provide a proactive and data-driven approach to perinatal care. Through its implementation, the project has the potential to make a significant impact on the well-being of pregnant women and their newborns, ultimately contributing to improved maternal and neonatal health outcomes.

9. FUTURE SCOPE

The future scope for the project on perinatal health risk prediction using machine learning is vast and holds great potential for advancing perinatal care. Here are some key areas of future exploration and development:

Integration of multi-modal data: The project can expand by integrating various types of data sources, including genetic data, environmental factors, social determinants of health, and lifestyle factors. By incorporating a broader range of data, machine learning algorithms can generate more comprehensive and accurate predictive models for perinatal health risks.

Real-time monitoring and interventions: Building upon the success of real-time electronic health recording, the project can explore the development of intelligent monitoring systems that leverage machine learning algorithms to continuously monitor maternal and fetal health parameters. These systems can provide real-time alerts and recommendations for timely interventions, ensuring proactive management of perinatal risks.

Personalized risk assessment and intervention: The project can focus on developing personalized risk assessment models that consider individual characteristics, medical history, and genetic profiles. This approach would enable tailored interventions and targeted preventive strategies for specific high-risk groups, maximizing the impact of perinatal healthcare.

Overall, the future scope for the project involves expanding data integration, real-time monitoring, personalized interventions, decision support tools, ethical considerations, and adapting the technology for low-resource settings. These advancements have the potential to revolutionize perinatal care, improve outcomes, and reduce health disparities in maternal and child health.