Pregnosmart – An AI Powered Virtual Birth Companion to Transform Prenatal Care

**Muttineni Sai Rohith, Bandaru Ratna Chaitanya Raju,
Yerram Sai Priya, Mallinani Lakshmi Bhavani**

DOI: 10.29322/IJSRP.14.02.2024.p14632
https://dx.doi.org/10.29322/IJSRP.14.02.2024.p14632

Paper Received Date: 11th January 2023
Paper Acceptance Date: 14th February 2024
Paper Publication Date: 21st February 2024

Abstract- *The landscape of prenatal care is undergoing a transformative shift with the advent of Artificial Intelligence (AI). This paper outlines an innovative application – Pregnosmart. Pregnosmart is an AI pioneered virtual birth companion, which leverages cutting-edge machine learning (ML) algorithms and Generative AI to offer unparalleled support to pregnant individuals and their families during pregnancy. By blending predictive analytics with ML models, Pregnosmart's core sub tool, Pregnoforecast, empowers users with predictions related to 8 crucial measures such as infections during pregnancy, newborn health indicators, gestational diabetes, and hypertension. Also, with the use of Generative AI, Pregnocompass, provides tailored day-to-day calendars based on pregnancy measures and Pregnosage, leveraging Retrieval Augmented Generation (RAG) provides an AI driven chatbot which addresses pregnancy related queries and provides enhanced contexts based on pregnancy measures and finally Pregnopedia, acts as a centralized data repository which automatically extracts and stores comprehensive pregnancy information using Natural Language Processing (NLP) and machine reasoning techniques. Together using all sub tools, Pregnosmart emerges as a holistic and intelligent solution, reshaping the pregnancy experience by offering personalized care, informed decision-making and transformative support to users navigating the journey of pregnancy.*

Index Terms- *Pregnancy, Birth Companion, Machine Learning (ML), Generative AI, Retrieval Augmented Generation (RAG).*

I. Introduction

Pregnancy, a transformative and crucial phase in a woman's life, brings with it a myriad of physical, emotional, and informational needs. Throughout history, the importance of providing support to pregnant individuals has been recognized, and birth companions have played a crucial role in ensuring a positive pregnancy experience. Birth Companions, traditionally comprising family members, friends, or midwives, offer emotional support, shared experiences and provide guidance in navigating the challenges of pregnancy. However, in current society, where family structures have evolved and the shift from joint to nuclear families raised a need for innovative approaches to ensure continuous and tailored support during pregnancy period. In the Modern era, the role of birth companions has transitioned to include healthcare professionals, doulas and now, with advancements in technology, virtual birth companions powered by Artificial Intelligence (AI). This paradigm shift aligns with the broader transformation in healthcare, where AI has become a cornerstone in enhancing diagnostic capabilities, personalizing treatments, and optimizing patient care.

Despite the advancements, pregnancy poses unique challenges, both physiological and psychological, for individuals. physical discomfort, emotional stress, and the sheer volume of information available can overwhelm expectant mothers. Moreover, current work culture and family structures led to a decrease in the immediate support available to pregnant women. Consequently, there exists a critical gap in providing comprehensive, personalized, and continuous support through the pregnancy journey. Addressing these challenges requires innovative solutions that leverage the power of AI to create virtual birth companions capable of delivering tailored support, reliable information and emotional assistance to pregnant individuals and their families. This paper delves into the significance of birth companionship, emphasizing the need for an intelligent and accessible solution, an AI Powered Birth Companion which provides tailored, intelligent, and continuous support – Pregnosmart. Through this solution we aim to enhance the pregnancy experience and empower individuals to navigate this transformative journey with confidence and informed decision – making.

II. RESEARCH AND IDEATION

Building upon the need for Virtual Birth Companionship and the challenges faced by pregnant individuals, our initial solution was to build a predictive analytical powerhouse that surpasses conventional pregnancy predictions. Traditional support often relies on general knowledge and experiences, but it lacks the precision needed to address each individualized health needs. So, by leveraging machine

learning algorithms and millions of pregnancy records, we aimed to multiply the experiences exponentially and provide precise predictions. Thus, our search for potential measures began, which could significantly contribute to reducing medical interventions by taking precaution measures upfront, during this exploration, we came across the Sustainability Development Goals (SDG-3) by WHO, stating potential measures indicating pregnancy. We condensed 11 important pregnancy indicators from SDG-3 into 8 pregnancy measures such as gestational diabetes, hypertension, pregnancy risk score, infections during pregnancy, newborn health indicator, mother morbidity, type of delivery, and baby weight, which led to the formation of our Pregnoforecast sub tool, The predictive precision that Pregnoforecast offers, provides the depth of information needed for informed decision-making, translating into a heightened awareness, and facilitating proactive decision-makingfor pregnant individuals and healthcare providers. Once we wrapped this aspect, another key area we focused on is healthcare advocacy and tailored support for each pregnant individual. The shift from joint to nuclear family structures has left pregnant individuals with limited immediate assistance and support. During the critical period of pregnancy, birth companions usually offer guidance on daily activities, medication reminders and essential tests. To address this gap through Pregnosmart, we introduced Pregnocompass, an intelligent sub tool that provides tailored assistance through dynamic calendars. These calendars are generated based on individual pregnancy measures, offering personalized guidance for daily activities, medication reminders, nutrition, exercises, and essential tests. This tailored approach enhances prenatal care, and based on the instructions followed, a prenatal care score is generated. By aligning with the unique needs of each pregnancy, Pregnocompass fosters a sense of empowerment and control during pregnancy.

The emotional and informational needs of pregnant individuals are paramount. During this critical period, they often encounter numerous questions. At any given day, if they experience pain in the abdomen, they may wonder why it is happening. Questions such as whether they can eat chicken liver or drink sugarcane juice may arise spanning diverse domains. Birth companions may find it challenging to address such a wide array of inquiries. To fulfill this need, Pregnosmart's AI driven conversional support, Pregnosage, steps in as a virtual confidant. By engaging in context-aware conversations, Pregnosage provides answers to pregnancy-related queries based on individuals' pregnancy measures. Using Retrieval Augmented Generation at its core, Pregnosage not only addresses the emotional aspects of pregnancy but also offers a valuable source of information. This responsive and adaptive approach effectively mitigates the potential feelings of isolation and uncertainty that can accompany the pregnancy journey. Pregnosage ability to provide personalizedand contextually relevant information enhances the overall support system, ensuring that pregnant individuals have a reliable companion to turn to for guidance and reassurance.

The abundance of information available online poses a challenge in discerning reliable guidance. Pregnosmart tackles this issue with the Pregnopedia sub tool, serving as a centralized knowledge repository. By leveraging Natural Language Processing techniques and web scraping, Pregnopedia curates and organizes trustworthy pregnancy information. The information collected is presented to users to provide easy access to accurate and up-to-date knowledge, Additionally, it functions as a central data repository, supplying information to other sub tools within the system. In this way, Pregnopedia prompts informed decision- making and helps dispel potential misinformation. In summary, Pregnosmart revolutionizes pregnancy support by providing a holistic solution that directly addresses the challenges outlined in the introduction. Through predictive analytics, tailored guidance, conversational assistance, and a centralized knowledge repository, Pregnosmart empowers pregnant individuals to navigate the transformative journey with confidence, informed decision- making, and a heightened sense of support. In the following sections, we will delve further into each sub tool, providing detailed insights into the underlying technology and functionality.

III. PREGNOFORECAST

The predictive analytical powerhouse – Pregnoforecast, harnesses cutting-edge machine learning (ML) algorithms to forecast crucial pregnancy measures.

A. Pregnancy Indicators:

According to WHO statistics, almost 800 pregnant women died from preventable causes related to pregnancy and childbirth every day in 2020, most maternal deaths are preventable if infections and diseases are timely managed, and treatment is provided promptly. This can make the difference between life and death for the mother and the newborn. Pre-eclampsia, eclampsia and postpartum hemorrhage are the most common causes of death during pregnancy and can be managed if detected early. Predicting complications during pregnancy and the type of pregnancy in time helps expectant mothers both financially and mentally. Newborn health indicators, newborn infections, and chance of getting admitted to the Neonatal Intensive Care Unit (NICU), if predicted in time, assist in taking precautionary measures and providing timely treatment for the child. Through a thorough investigation of SDG-3 and other key factors that affecting pregnancy, we wereable to prepare a condensed list of pregnancy indicators thatPregnoforecast predicts –

Pregnancy Indicators	Machine Learning Process
Gestational Diabetes	Classification
Pre-eclampsia and Eclampsia	Classification
Pregnancy Risk Score	Regression
Infections during Pregnancy	Classification
Newborn health indicator	Classification
Mother morbidity	Classification
Type of Delivery	Classification
Baby Weight	Regression

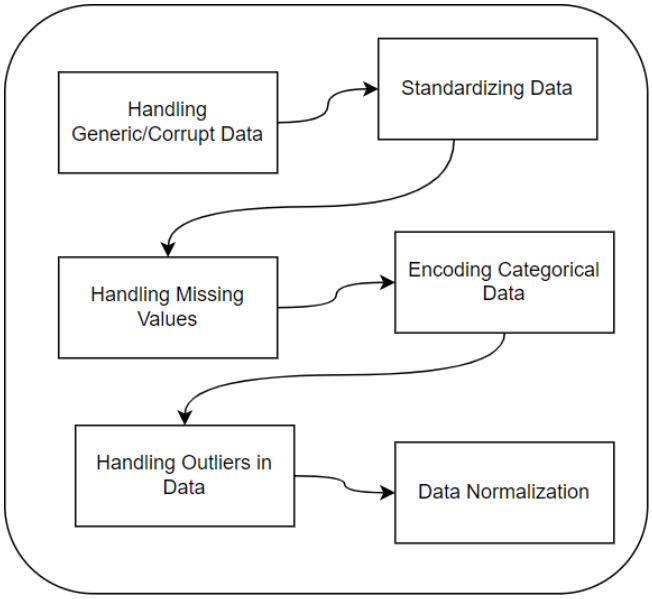
B. Data Exploration:

Once we identified the list of pregnancy Indicators, we began our quest for the right dataset to predict these indicators effectively. Several datasets were explored, each providing limited information on one or two indicators, making comprehensive predictions challenging. We got a breakthrough when we discovered the NBER site, which proved invaluable by offering a dataset encompassing over 30 million pregnancy records spanning two decades across various regions. This extensive dataset provided a holistic view, enabling us to cover spectrum of outlierscenarios in pregnancy. The NBER dataset played a pivot role in our research, offering a robust foundation for accurate predictions and a comprehensive understanding of diverse futures influencingpregnancy outcomes and indicators. Leveraging this source, we were able to predict all pregnancy indicators, as it contains information about all the relevant features.

Once we identified the pregnancy indicators and the substantial dataset, we transitioned into the model development phase. This phase involved various key steps such as data preprocessing, feature extraction, model development, and deployment. Starting from this juncture, the paper will use gestational diabetes as an illustrative example to elucidate the flowof the subsequent processes, as most indicators followed similar methodology.

C. Data Preprocessing:

In processing real world data, accuracy can be compromised due to generic or corrupt information, making it unsuitable for model development without preprocessing. With a dataset of over 30 million records, the likelihood of generic or corrupt data was high, corrupt data is flagged as Unknown(U) for categorical values and 9/99/999 for numerical values, this can be possible due to manual errors or inadequate information during data collection or storage.

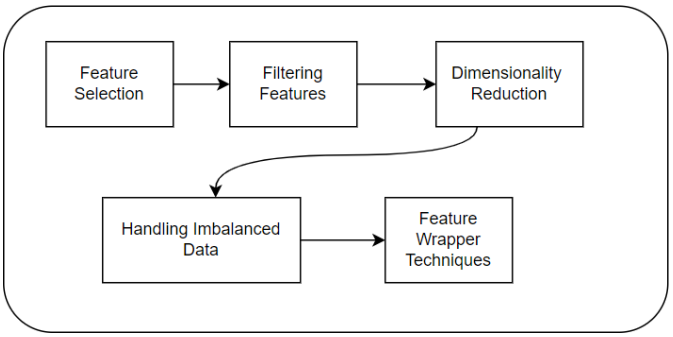


3.1 Data preprocessing

Handling such generic or corrupt information was a crucial step for our preprocessing, depending on the indicator or feature, we addressed this issue by converting data to null values, dropping records, or imputing values with mean or median values. Asdepicted in figure 3.1, the next step after handling generic/corruptinformation is standardizing the data. While dealing with few indicators, we have used binning and other features to derive newones as required. Given the volume of data that we dealt was skewed, and as missing values were prevalent, we employed various techniques such as imputing missing values with mean/mode values. For certain indicators, we went beyond traditional methods and utilized KNNImputer and the MissForest algorithms. As machine learning models typically operate with numerical values, encoding categorical data into numerical values using MultiColumnLabelEncoder in Python was a key step in our data preprocessing. In the subsequent preprocessing phase, addressing outliers in the data was crucial. For instance, with mother weight, outliers exceeding 200Kgs were observed. Handling such outliers using interquartile ranges were imperative to prevent potential corruption of model performance. The final step in our preprocessing involved addressing inconsistent data types, such as categorical data, inconsistent units, varied date formats, and differing lengths in data. Ensuring uniformity across the data was essential to establish consistency before moving on to the feature extraction phase.

D. Feature Extraction:

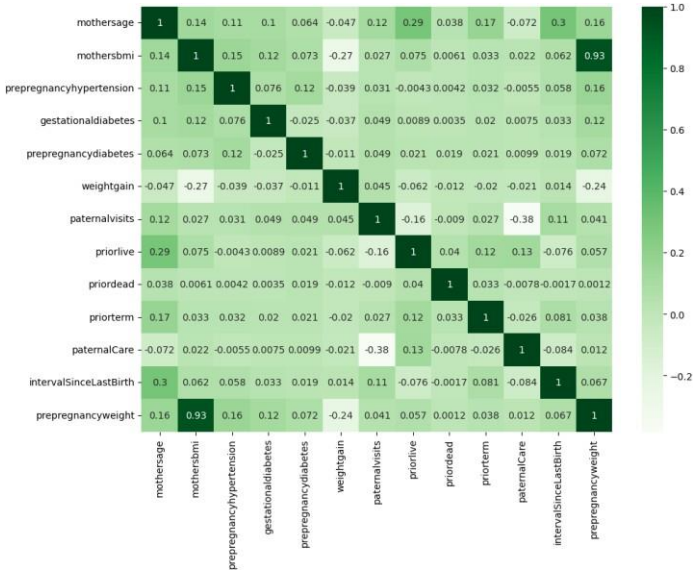
Feature extraction is a significant phase in model development, guiding the focus towards identifying and transforming key information within the dataset. Here, pertinent features are selected and shaped to encapsulate the intrinsic patterns and nuances underlying the data. The initial step in our feature extraction phase is feature selection. Different pregnancy measures in Pregnoforecast requires distinct features. For instance, the features used to predict gestational diabetes will differ from those used for predicting infections during pregnancy. During feature selection,



3.2 Feature Extraction

features for each measure are chosen through consultations with gynecologists and thorough data review. Incorrect feature selection can result in poor model performance. After a comprehensive review, the features selectedfor Gestational Diabetes are as shown in table 3.4. Once the features are obtained, constant features, quasi-constant features, and highly correlated features are filtered out as they lead to biased models and do not help in model learning.

Constant features have same value across all rows in the data. Filtering them reduces computational overhead and improves model efficiency whereas Quasi-constant features with very low variance, they have predominantly one value but may have few instances of others. For instance, for Hypospadias infection, 97% values are rated as No, 2.7% values are unknown, and 0.3% values are rated as Yes. Filtering Quasi-Constant features aids in mitigating noise and focuses on more informative features, enhancing the model's ability to discern patterns. Highly correlated features that exhibit strong linear relationships with other features, convey redundant information. They can introduce multicollinearity leading to unstable model estimates. Filtering them out helps in avoiding model bias, promoting better generalization, and ensures that each feature contributes distinct information to the model.



3.3 Finding feature Collinearity.

The next phase is dimensionality reduction, while not applicable to all the pregnancy measures, but for few, when the number of relevant features is more, to transform multiple features to linear dimensions or to remove irrelevant features, techniques such as lasso, Principal Component Analysis (PCA) and t-SNE are used accordingly. PCA, for instance, is used to transform high-dimensional data into a new coordinate system, for analyzing the preterm birth ratio, we combined last menstrual period, and week, month, and year of baby birth date into a single value to obtain estimated date of birth and gestational period for the baby. Dimensionality reduction is helpful in condensing the dataset's variability, facilitating efficient analysis and model training.

In healthcare, imbalanced data is a common challenge, where the likelihood of experiencing an infection is often significantly less compared to not having one. In our case, when predicting infections during pregnancy, the chance of having an infection was approximately 12%. To address this highly imbalanced data, we used Synthetic Minority Over-Sampling Technique (SMOTE) and Near Miss Under-Sampling. While SMOTE generates synthetic samples for the minority class by interpolating between existing instances, the latter involves reducing the size of the majority class to balance the class distribution. In our approach, we used a hybrid method to strike a balance between oversampling the minority class and under sampling the majority class to a thread, mitigating the risk of overfitting and ensuring balanced dataset for model training.

Feature	Description	Data Availability
Age	Advanced maternal age isa risk factor for gestational diabetes.	Yes
BMI (Body Mass Index)	High BMI or obesity increases the risk of developing gestational diabetes.	Yes
Family History	A family history of diabetes or gestational diabetes can be a significant risk factor.	Yes
Previous History	A history of gestational diabetes in previous pregnancies increases the risk.	Yes
Ethnicity	Certain ethnic groups, such as South Asian, African American, Hispanic, and Native American populations, have a higher predisposition.	Partial
Medical History	Preexisting conditions like polycystic ovary syndrome (PCOS) or prediabetes can contributeto the risk.	Partial
Blood Pressure	Elevated blood pressurecan be a risk factor for gestational diabetes.	Yes
Glucose Levels	Abnormal fasting blood glucose or glucose tolerance test results are indicators	Yes
Insulin Resistance	Factors like high insulin levels and insulin resistance contribute to the risk	Yes
Diet and Nutrition	Dietary habits and nutritional intake, especially excessive sugar, and carbohydrate consumption, can play a role	Partial

Physical Activity	Lack of physical activity or a sedentary lifestyle can increase the risk	Partial
Gestational Weight Gain	Excessive weight gain during pregnancy can be associated with gestational diabetes.	Yes
HormonalFactors	Hormonal changes during pregnancy can affect insulin sensitivity	No
Number of Pregnancies	A history of multiple pregnancies can impact the risk	Yes
Inflammatory Markers	Elevated levels of inflammation markers can contribute to insulin resistance.	No
Waist Circumference	Abdominal obesity is a significant risk factor.	Yes
Hemoglobin A1c(HbA1c)	Elevated HbA1c levels may indicate increasedrisk	No
Genetic Factors	Genetic predisposition can contribute to susceptibility	No
Insulin Resistance Makers	Measurements like HOMA-IR (HomeostaticModel Assessment of Insulin Resistance) can provide insights.	No
Blood Lipid Levels	Abnormal lipid profilescan be associated with gestational diabetes risk.	Yes
Triglyceride Levels	Elevated triglycerides may indicate increased risk.	No

3.4 Gestational Diabetes Feature Exploration

Table 3.4 illustrates the features initially selected for gestational diabetes. Depending on data availability, these features are further transformed. Based on model performance, features are either removed, or new features are added. Forward feature selection is considered when certain features are deemed necessary for the model, and backward feature elimination is employed when model performance is poor or specific features are deemed unimportant. Once the data is properly transformed and features are extracted, training and testing data are prepared using the scikit-learn library. 80% of the data is utilized for training, and the remaining data is allocated for testing.

E. Model Development

This is the most pivotal phase of Pregnoforecast, where we focus on creating robust predictive models for diverse pregnancy measures. We deliberately selected machine learning models tailored to the unique requirements of each pregnancy measure. Initially we started our model development with gestational diabetes, which has two classes - the risk of getting gestational diabetes (yes or no). With the training and testing dataset ready, we started with logistic regression, the initial classifier in machinelearning. However, it did not perform well, with the accuracy dropping to 63%. This was attributed to the vast data size and a greater number of features, making logistic regression less suitable. In our next trial, we opted for K Nearest Neighbors (KNN), despite a hike in accuracy and recall score, the desired metrics were not achieved.

In the subsequent trial, Support Vector Classifier (SVC) was considered, but due to the vast dataset and number of features, it proved unreliable, and the execution time increased significantly. Then, decision tree emerged as a promising alternative, providing 95% accuracy and 94% recall score. Encouraged by the success of decision tree, we turned our attention to Random Forest, which further improved recall score to 97%. The ensemble nature of Random Forest allowed it to mitigate overfitting and capture a more comprehensive understanding of the complex pregnancy data. Finally, through Hyperparameter tuning atop of Random Forest, we achieved 97% test accuracy and 99% recall score. The meticulous tuning allowed us to fine tune the model parameters resulting in increased performance.

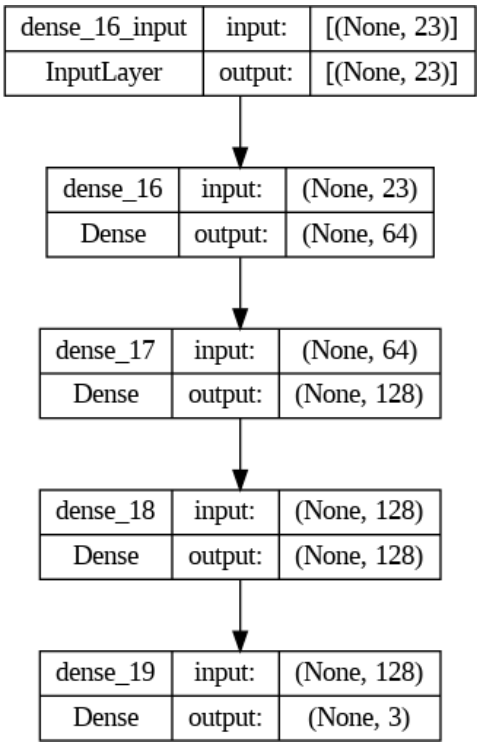
Model	Accuracy	Recall Score
Logistic Regression	63%	64%
K Nearest Neighbors	92%	89%
Support Vector Machine	unreliable	unreliable
Decision Tree	95%	94%
Random Forest	96%	97%
Random Forest with Hyperparameter Tuning	97%	99%

3.5 Gestational Diabetes Model Development Metrics

Like the approach taken for gestational diabetes, we applied a similar workflow for other pregnancy measures. For gestational hypertension, we achieved a 96% test accuracy and recall score by employing Random Forest with Hyperparameter Tuning, utilizing GridSearchCV for fine-tuning.

In case of predicting the newborn health indicator, where intricate patterns and non-linear dependencies play a crucial role, we opted for Long Short-Term Memory (LSTM). LSTM, a type of Recurrent Neural Network (RNN), is well suited for capturing such complexities. With LSTM, we achieved an 83% accuracy and an 85% recall score, highlighting its effectiveness in handling the specific challenges posed by the newborn health indicator. In instances where conventional machine learning models such as bagging algorithms, KNN, and Logistic regression fell short in predicting infections during Pregnancy, we turned to the boosting method, specifically XgBoost. The XgBoost algorithm demonstrated robust performance, yielding an 89% accuracy and an 88% recall score. Boosting methods, with their ensemble learning

approach, excel in improving predictive accuracy. In the final stage of classification phase, which focuses on predicting the type of delivery – whether Cesarean or Vaginal birth – we employed Random Forest, predicting type of delivery is valuable for providing mental and pregnancy cost assistance, aiding in financial planning for expectant parents. Addressing the critical feature of mother morbidity in pregnancy, we employed a Multi-Layer Perceptron (MLP). This choice proved fruitful, resulting in a 93% accuracy. MLPs are effective in capturing intricate relationships in the data, making them suitable for predicting complex features like mother morbidity.



3.6 multi-layer perceptron used for predicting Mother Morbidity

The importance of recall score in healthcare, motivated us to consider it along with accuracy. Recall score, also known as sensitivity or True Positive (TP) rate, represents the model's ability to correctly identify all actual positive instances, ensuring that cases requiring timely intervention are not overlooked. Recall score is calculated as follows:

$$\text{Recall score} = \frac{\text{True Positives}}{\text{True Positives+False Negatives}}$$

In pregnancy care, missing a critical case can have severe consequences for both the mother and the child. Therefore, a high recall score is essential to identify all true positive cases and avoid overlooking cases that demandattention. Given the critical nature of pregnancy measures, where the avoidance of missed cases is paramount, we incorporated recall scores alongside accuracies in all our pregnancy measures.

3.7 Predictive performance for each pregnancy measure

The figure 3.7 depicts the accuracies and recall score of each pregnancy measure. These model selections were driven by the unique characteristics of each pregnancy measure, emphasizing the importance of tailoring the choice of algorithms to the specific challenges presented by the data.

In our subsequent phase focusing on regression, CatBoost regression was employed for predicting pregnancy risk score, given the abundance of categorical data. After experimenting with various regression techniques, we opted for CatBoost regression and achieved a Mean Squared Error (MSE) of 27.83. For predicting baby weight, the scale of data, the presence of outliers, and other influential parameters prompted a shift from regression to classification. We categorized the classesas obese, normal, and thin. By utilizing K Nearest Neighbors for classification, we achieved satisfactory accuracy. Below table visualizes the algorithms used for each pregnancy measure.

Pregnancy Measure	Machine Learning Model
Gestational Diabetes	Random Forest with Hyperparameter Tuning
Hypertension	Random Forest with Hyperparameter Tuning
Pregnancy Risk Score	CatBoost Regression
Infections during Pregnancy	XgBoost
New born health indicator	Long Short-Term Memory (LSTM)
Mother morbidity	Multi-Layer Perceptron (MLP)
Type of Delivery	Random Forest
Baby Weight	K Nearest Neighbors

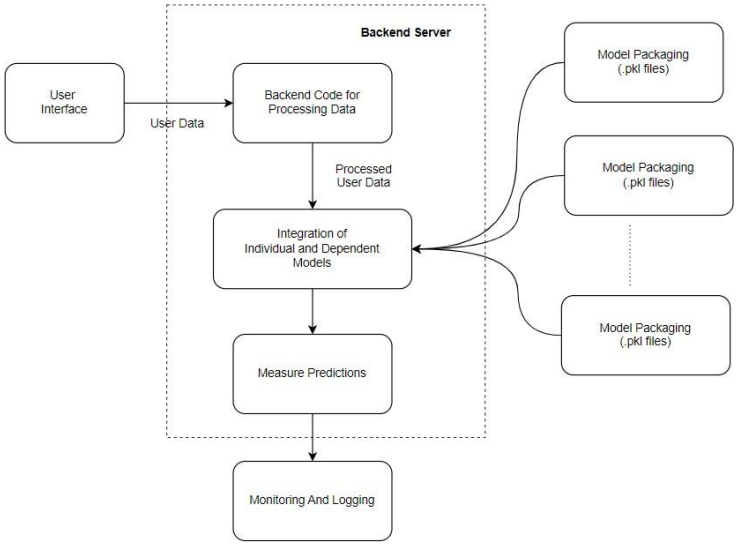
3.8 Machine Learning model for each measure

F. Model Deployment

As we developed eight distinct models for the prediction of various pregnancy measures, each model is saved as a pickle file. These models are then housed in an Azure Virtual Machine (VM).

Whenever a user submits their medical record through the User Interface (UI), the information is directed to the Azure VM using a Flask Python-based REST API. Azure VM is here by referred as backend server. The medical record, presented in the form of an image, is processed in the backend server, where Optical Character Recognition (OCR) is used to extract relevant information and observations for pregnancy measures.

During the user's initial registration, a questionnaire is dispatched, and the responses are stored. These answers provide insights into nutrition, previous health conditions, pregnancies, family history, and other crucial parameters. The collected information is then extracted, preprocessed, cleaned, and prepared as input parameters for the model. The model wrapper method, housing the models for each pregnancy measure, accepts the data corresponding to the features for each model and outputs pregnancy measures. Subsequently, through the REST API, the predictions are sent back to the UI and simultaneously stored in the underlying database.



3.9 Model Deployment

Once the models are deployed in production, continuous monitoring of data and logging is implemented, and appropriate actions are taken based on the observed data. This ensures that the models operate effectively in production environment, and they are made accessible to end-users for reliable use.

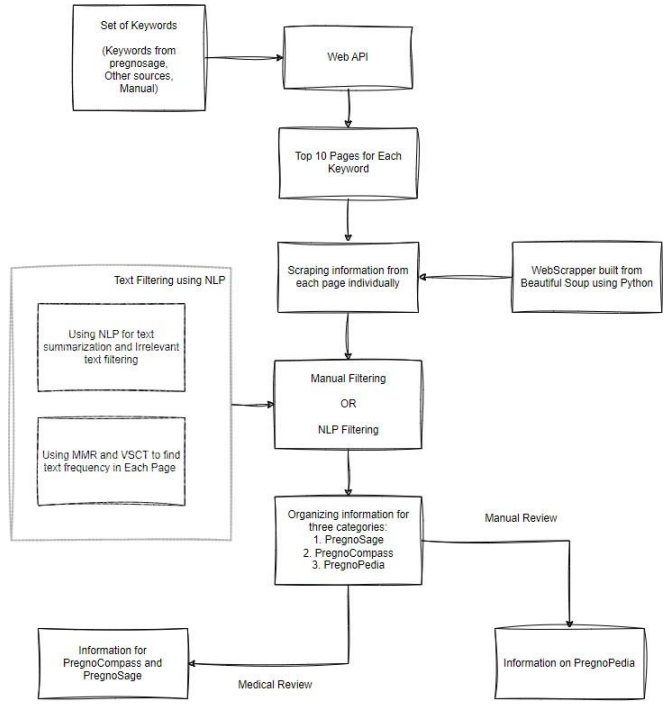
As shown in this complete section, with a meticulous bend of machine learning techniques, Pregnoforecast not only enhances predictive precision but also addresses the unique needs of each pregnancy measure by offering personalized predictions for various critical pregnancy measures. By incorporating diverse models, continuous monitoring, and seamless integration into a user-friendly interface, Pregnoforecast paves the way for improved prenatal care and maternal well-being.

IV. PREGNOPEDIA

Pregnopedia, a pivotal component of the comprehensive Pregnosmart platform, serves as an expansive knowledge repository dedicated to pregnancy-related information. Utilizing advanced Natural Language Processing (NLP) techniques and web scraping technologies, Pregnopedia systematically curates and organizes a wealth of trustworthy and up-to-date data. This intelligent sub tool acts as a central hub, providing users with easy access to a diverse range of reviewed information, from essential parental care guidelines to in-depth insights on various aspects of pregnancy. In addition to serving as a standalone information source, Pregnopedia plays a crucial role as a centralized data repository. It not only provides valuable information directly to users but also shares relevant data with other sub tools within the Pregnosmart platform. This interconnected approach ensures a continuous flow of accurate and pertinent information throughout the platform, enhancing the overall effectiveness and coherence of the pregnancy support system. Pregnopedia, therefore, stands not only as an empowering resource for expectant individuals but also as a collaborative component contributing to the seamless functioning of the broader Pregnosmart ecosystem.

I. Architecture

Pregnopedia receives a set of keywords as input, sourced from other Pregnosmart sub tools, manual inputs, external sources, or automatically extracted information related to pregnancy. These



4.1 Pregnopedia Architecture

keywords are then passed to a web API, we used Google Chrome API in python to automatically conduct searches and extract the top 10 pages for each keyword. Subsequently, these pages undergo extraction of textual and statistical information using a web scraper tool developed with Beautiful Soup in Python.

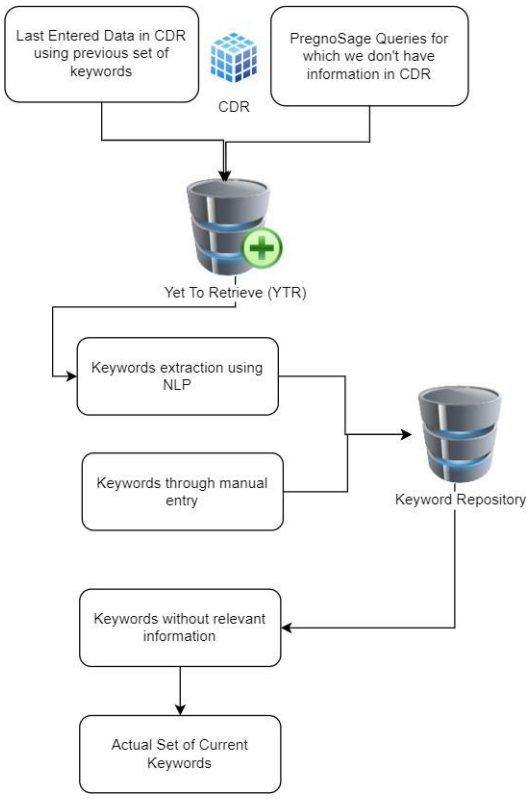
The extracted information from all 10 pages for each keyword is fed into an algorithm we designed named Relevant Information Extraction Technique (RIET). RIET, incorporating Natural Language Processing (NLP), Minimal Marginal Reduction (MMR) and Vector

Similarity CalculationTechnique (VSCT), discerns and extracts only the pertinent pregnancy information. The extracted data is temporarily stored in an intermediate table, subjected to review by medical experts, and then incorporated into the Central Data Repository. This enriched repository serves as a shared resource for all other subtools within Pregnosmart.

Pregnopedia also plays a crucial role in presenting this information through the User Interface (UI), offering comprehensive insights into pregnancy-related topics such as medical tests, medications, and the baby's growth cycle during pregnancy. By providing expectant mothers and their families with this knowledge, Pregnopedia aims to empower them with informed decision-making throughout the pregnancy journey. In the subsequent sub-sections, we will explore each step in the Pregnopedia process and delve into the underlying technologiesin detail.

II. Keyword Extraction

Pregnopedia acquires keywords through three distinct sources to ensure it remains current and comprehensive in retrieving pregnancy-related information, these sources include Pregnosmart.



4.2 Keyword Extraction

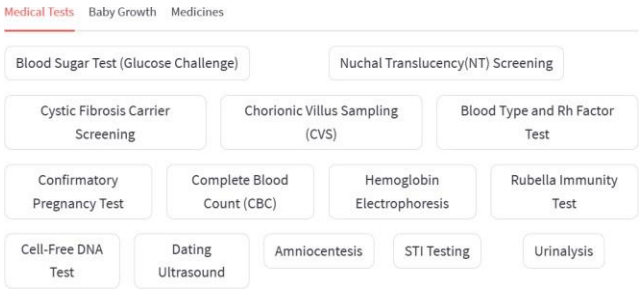
Sub tools, manual entries by data team, and an intelligent automatic selection of keywords. Pregnosmart sub tools, such as Pregnosage, generate queries from users, and if answers are unavailable in the Central Data Repository (CDR), questions are stored in a Yet to Retrieve (YTR) dataset. Pregnopedia extracts keywords from these questions and proceeds with the information retrieval process, storing the results back into CDR.

Manual entries are straightforward and managed by the data team. For automatic keyword selection, an intelligent process is employed. The system scans all the information in the CDR generated using the previous set of keywords, eliminating stop words, performing noun extraction, and using context evaluation mechanisms in NLP to identify relevant pregnancy-related keywords. The final set of keywords is then added to the Keyword Repository. A flag is included in the repository to identify keywords without relevant information in the CDR, directing them to further processes.

III. Content extraction for relevant keyword

Once the keywords are obtained, the Google API in Python is utilized to fetch the top 10 search results for each keyword, The 'googlesearch-python' library is employed for this purpose. Subsequently, using the URLs obtained from the search results,a web scraping tool built with Beautiful Soup in Python is employed. This tool extracts all the textual information available on the respective websites, as depicted in Figure 4.1. The extracted text is then prepared for further processing.

In the text extraction phase, Relevant Information Extraction Technique (RIET) is applied, utilizing the Minimal Marginal Reduction (MMR) technique, Vector Similarity Calculation Technique (VSCT), and Natural Language Processing (NLP). MMR is geared towards selecting sentences that are both pertinent to the query and diverse from each other,aiming to mitigate redundancy in the extracted information. It achieves this by assigning a score to each sentence based on itsrelevance to the query and dissimilarity in already selected sentences, thus striking a balance between relevance and diversity. MMR filters out redundant or closely similar sentences, ensuring diversity in the extracted information and coverage of various aspects of the query. On the other hand, VSCT measures the similarity between vectors representing sentences, identifying the closeness or relatedness of sentences to the keywords. Vector representations are created using techniques such as Word Embeddings (Word2Vec, GloVe), and similarity between vectors is calculated using metrics like Cosine Similarity or Euclidean distance. VSCT aids in selecting sentences that are contextually related to the keywords.



4.3 Medical tests information displayed in Pregnopedia.

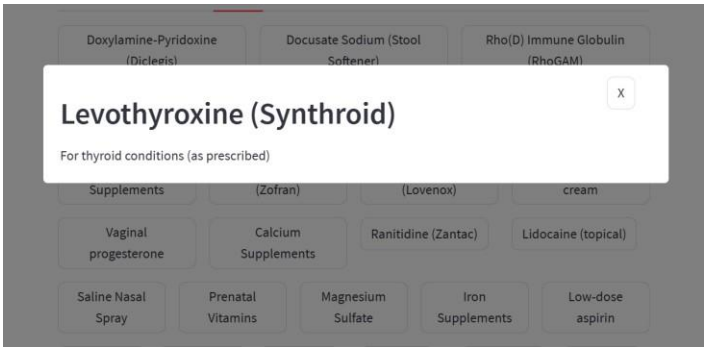
The combined application of MMR and VSCT, integrated with NLP, enhances the relevance and coherence of the extracted text information. Additionally, when coupled withNLP, this process facilitates text summarization and the retrieval of the actual context

from the overall documents. In this manner, relevant information is extracted for a keyword.

IV. Information Storage and Utilization

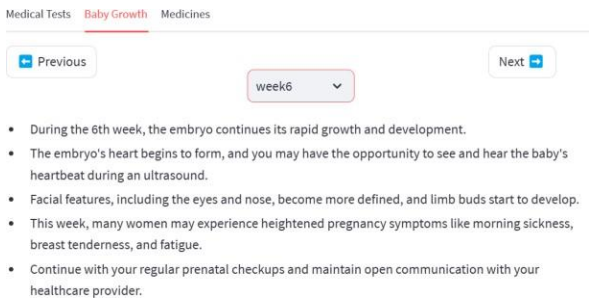
The extracted information is then stored in an Intermediate table, and after medical review it is pushed into CDR. One of the key aspects of Pregnosmart, is to provide a reliable information, due to abundant and unreviewed information available in online, there is uncertainty in the data to rely. Hence to enhance data reliability, and provide accurate data, before pushing data into CDR, it is medically reviewed.

The data is then utilized by all other sub tools of Pregnosmart and is displayed in the UI to provide information for expectant mothers and family. It categorizes the extract datainto four categories for display on the UI. Pregnopedia shows medical tests, as illustrated in 4.3, provides information regarding medicines, displays baby growth cycle throughout the pregnancy period, and showcases pregnancy facts and news. Figure 4.4 depicts information on medicines shown in Pregnopedia.



4.4 Medicines Information shown in Pregnopedia.

Figure 4.5 provides a gist of weekly baby growth.



4.5 Baby growth shown in Pregnopedia.

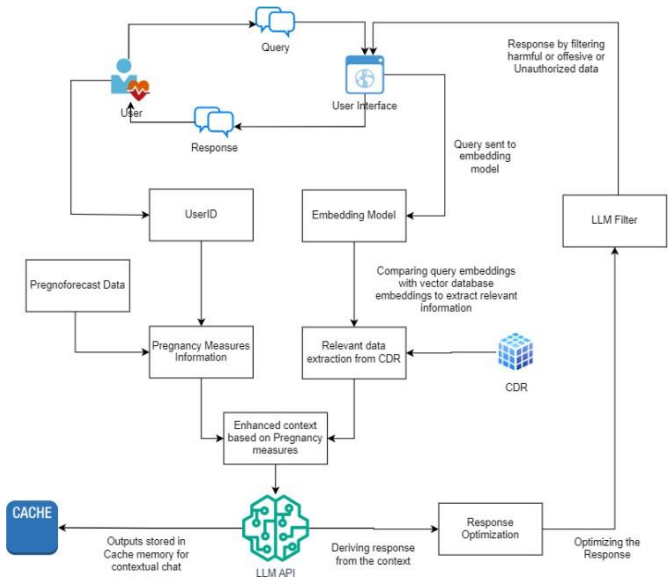
Therefore, Pregnopedia stands as an indispensable component within the Pregnosmart platform, offering a wealth of pregnancy-related information through advanced Natural Language Processing techniques and web scraping technologies. By systematically curating and organizing diverse and trustworthy data, Pregnopedia empowers expectantindividuals with the knowledge needed to make informed decisions during pregnancy. Its seamless integration with other Pregnosmart sub tools ensures a centralized repository of information, contributing to enhanced prenatal care and a more informed pregnancy journey.

V. PREGNOSAGE

Pregnosage, a significant facet of the Pregnosmart platform, serves as a virtual confidant and conversational support system in short, a pregnancy relevant chatbot for expectant individuals.By engaging in context-aware conversations, Pregnosage addresses pregnancy-related queries based on individual pregnancy measures. Using a combination of Retrieval Augmented Generation (RAG) and advanced large language models, Pregnosage not only provides emotional support during the critical period of pregnancy but also acts as a valuable source of information. This responsive and adaptive approach aims to mitigate potential feelings of isolation and uncertainty that can accompany the pregnancy journey. In the upcoming subsections, we will delve deeper into the functionality, technology, and underlying processes of Pregnosage.

A. Pregnosage AI Architecture

When a user initiates a query through the User Interface (UI), the query is processed by an embedding model such as GloVe or Word2Vec, which converts the text information into vector embeddings. These embeddings are compared to the data in the Central Data Repository (CDR), and all the relevant information is extracted. If there are multiple instances ofrelevant information



5.1 Pregnosage AI Architecture

in the CDR, all instances are retrieved. Additionally, the UserID is leveraged to retrieve pregnancy measures from the Pregnoforecast data. By comparing this information with the relevant data from the CDR, enhanced contexts are obtained. For example, consider a user querying whether it's safe to take an Aspirin tablet during pregnancy. If the user has diabetes, Pregnosage not only provides general advice that an expectant mother can take Aspirin but also tailors the response based on the user's diabetic condition. In this case, Pregnosage willsuggest that individuals with diabetes prefer a

diluted form of Aspirin, such as Ecospirin. This personalized and contextually relevant assistance showcases the importance of Pregnosage.

Following the extraction of enhanced contexts, they are transmitted to a Language Model (LLM) API such as Hugging Face or OpenAI GPT. The LLM API stores the responses in cache memory, preserving the contextual chat. Utilizing this LLM API, responses are generated based on the context. These responses are subsequently optimized and summarized to provide concise and meaningful replies. To ensure the delivery of appropriate and safe content, an LLM filter is employed to filter out harmful, offensive, or unauthorized data. The filtered responses are then sent back to the User Interface, which in turn directs the information to the user. This process ensures that users receive relevant and curated responses while maintaining a secure and reliable communication environment.

B. Pregnosage Q&A Forum

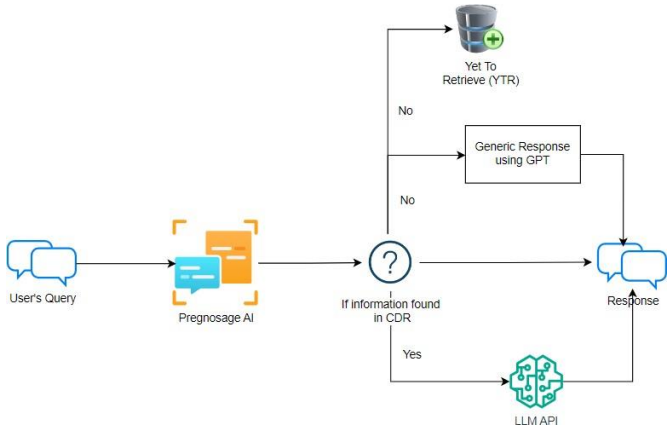
In addition to the chatbot functionality, Pregnosage incorporates a curated set of questions and answers accessible through the User Interface (UI). This feature provides expectant mothers with resourceful platform to explore and gain insights into various aspects of pregnancy. The curated questions cover a wide array of topics, offering detailed information and explanations that contribute

5.2 Pregnosage Q&A Forum

to a better understanding of the pregnancy journey. By presenting this repository of questions and answers, Pregnosage aims to empower and educate expectant mothers, providing them with valuable knowledge to navigate the complexities of pregnancy more confidently and informed.

C. Pregnosage Flow

When a user submits a request through the User Interface (UI), it is initially directed to Pregnosage AI. The architecture of Pregnosage AI is depicted in Figure 5.1. utilizing the information stored in the Central Data Repository (CDR) and the pregnancy measures from Pregnoforecast, it formulates a response for the user and sends it back. If the requested information is not available in the CDR, Pregnosage AI leverages GPT Model to generate a generic response. Simultaneously, the query is stored in the Yet to Retrieve (YTR) table. Pregnopedia then extracts the relevant information for the query,



5.3 Pregnosage Flow

undergoes medical review in next cycle and subsequently stores the data into the CDR. This seamless process ensures a continuous flow of information through Pregnosage.

Therefore, Pregnosage, with its dynamic architecture and seamless integration of AI technologies, by leveraging existing data in the Central Data Repository and Pregnoforecast data, provides tailored responses to user queries, ensuring continuous, reliable, and informed interaction. It serves as a reliable companion for expectant mothers, offering medically reviewed insights and facilitating an enhanced understanding of their unique pregnancy journey.

VI. PREGNOCOMPASS

Pregnocompass, another pivotal sub tool within the comprehensive Pregnosmart platform, enables navigating the prenatal journey by providing tailored assistance and personalized guidance to expectant mothers. This tool focuses on addressing the distinct needs of each pregnant individual through the utilization of dynamic calendars, generating a sense of empowerment and control during pregnancy. At its core, Pregnocompass relies on individual pregnancy measures obtained from the Pregnoforecast, along with the pregnancy measures, we are going to have supplementary measures received from user's medical records. These measures encompassing crucial aspects such as gestational diabetes, hypertension, pregnancy risk score and more, serves as the foundation for creating dynamic calendars. These calendars offer personalized guidance across various domains, including daily activities, medication reminders, nutrition, physical activities, essential tests, daily chores and even sleep monitoring.

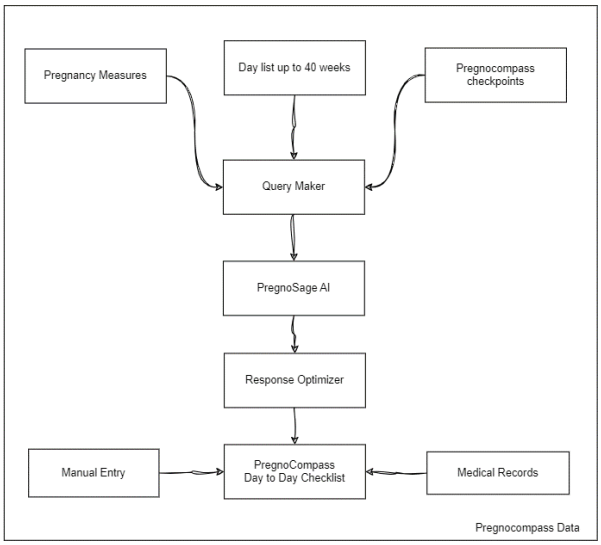
The emphasis on dynamic calendars reflects Pregnocompass commitment to provide specific and relevant directions tailored to the unique circumstances of each pregnant individual. By seamlessly integrating with both the Pregnosage AI and the predictive analytics of Pregnoforecast, Pregnocompass ensures that the guidance aligns precisely with individual pregnancy measures. One distinct feature of Pregnocompass is its ability to generate a prenatal score based on the adherence to the provided instructions. This score reflects the individual's commitment in following the personalized guidance offered by Pregnocompass. By fostering a sense of empowerment and control, the prenatal score

becomes a valuable metric, encouraging pregnant individuals to actively engage in their prenatal care. In summary, Pregnocompass emerges as a compass, providing a clear and personalized direction for expectant mothers as they navigate the intricate journey of pregnancy.

A. Architecture

The foundation of Pregnocompass lies in its meticulous data preparation, which begins with the integration of crucial pregnancy measures, like gestational diabetes, hypertension, pregnancy risk score etc. with key Pregnocompass checkpoints

such as nutrition, physical activities, medicines, medical tests, daily activities, and sleep monitoring. When these parameters along with a day-to-day list extending up to 40 weeks, is passed to Query Maker (QM), it handles the rest. The Query Maker initiates the data preparation by formulating queries based on the configured parameters. For example, considering gestational diabetes on day 130, the query might inquire about suitable breakfast options. Similarly, for physical activity on day 188 with hypertension, the query could seek the advice on the duration and types of exercises permissible. This query is then directed to the Pregnosage AI, which



6.1 Pregnocompass Architecture

comprehends the inquiry and generates a tailored response. The response is further optimized and summarized into a concise one-liner, constituting the essence of the Pregnocompass day-to-day checklist.

To enhance user flexibility, expectant mothers can override the default Pregnocompass checklist by providing their preferences or schedules manually. Moreover, by scanning uploaded medical records, Pregnocompass can provide reminders for medications and tests, aligning with theindividual's health profile. The corresponding data is seamlessly stored in the day-to-day checklist, ensuring that the guidance provided by Pregnocompass remains personalized and adaptable to the unique needs of each user.

B. Pregnocompass Flow

Upon a user's selection of Pregnocompass, a synchronization mechanism is initiated to harmonize their pregnancy measures, manual configurations, and medical records with the Pregnocompass data repository. This integrated information serves as the foundational input for Pregnocompass. Leveraging pregnancy measures, manual configurations, and medical records, the system dynamically updates the day-to- day checklist. This continual refresh ensures that the dynamic calendars are tailored to the user's evolving needs. Once the day-to-day checklist is obtained, corresponding instructions are formulated. Adherence to these instructions by users, contributes to the calculation of a prenatal care score, which provides a quantitative measure of the user's commitment and encourages them for optimal prenatal care practices.

VII. FUTURE SCOPE

The future scope of Pregnosmart involves continuous enhancements and expansions. The integration of real-time health monitoring devices and Electronic Health Records (EHR) is poised for seamless data input, eliminating manual user entries. Leveraging pregnancy measures from Pregnoforecast opens avenues for advanced analytics, particularly in the realm of Value-Based Care (VBC) and the exploration of cost-based analytics. As the platform continues to evolve, diversifying pregnancy measures will enable more nuanced and personalized dynamic calendars. Additionally, the implementation of advanced data reliability techniques andseamless collaboration with medical experts holds the potential to elevate the integrity and accuracy of data within the Centralized Data Repository (CDR). The heightened accuracy in the CDR, further amplifies the support provided by Pregnosage. This forward-looking strategy ensures that Pregnosmart remains at the forefront of innovation, consistently enhancing prenatal care and user experience.

VIII. CONCLUSION

In conclusion, Pregnosmart stands at the forefront of modernizing prenatal care by harnessing the power of artificial intelligence and data analytics. Through its innovative sub tools, Pregnoforecast, Pregnocompass, Pregnosage, and Pregnopedia the platform redefines the pregnancy experience, offering a comprehensive and personalized approach to support expectant individuals and their families. Pregnoforecast predictive analytics provides precise insights into crucial pregnancy measures, enabling informed decision-making and proactive healthcare management. Pregnocompass extends thissupport by delivering tailored assistance through dynamic calendars, addressing the unique needs of each pregnant individual. By seamlessly integrating with Pregnosage AI and Pregnoforecast pregnancy measures, Pregnocompass fosters a sense of empowerment and control during pregnancy.

Pregnosage, the AI-driven chatbot, acts as a virtual confidant, addressing pregnancy-related queries and offering valuable information based on individual pregnancy measures. It provides responsive and adaptive support, effectively mitigating

potential feelings of isolation and uncertainty that can accompany the pregnancy journey. Pregnopedia, as a centralized knowledge repository, curates and organizes trustworthy pregnancy information using natural language processing techniques. It not only serves as an information hub for users but also contributes to informed decision-making across all sub tools within the Pregnosmart ecosystem. Together, these sub tools create a holistic and intelligent solution, reshaping the pregnancy experience by offering personalized care, informed decision-making, and transformative support. Pregnosmart's continuous commitment to innovation and integration with emerging technologies opens avenues for further enhancements, ensuring it remains at the forefront of providing cutting-edge prenatal care.

REFERENCES

[1] Smith, J., & Johnson, A. (2020). "Advancements in Artificial Intelligence for Prenatal Care." Journal of Medical Technology, 15(3), 123-145.

[2] Anderson, M., et al. (2019). "Data-driven Solutions for Enhancing Pregnancy Experience." Proceedings of the International Conference on Health Informatics.

[3] Brown, R., & White, S. (2018). "Machine Learning Approaches in Predictive Pregnancy Analytics." Journal of Computational Medicine, 25(2), 67-82.

[4] Garcia, L., et al. (2021). "Innovations in Virtual Birth Companionship using AI." International Journal of Women's Health, 12, 345-359.

[5] Patel, K., et al. (2017). "Impact of Predictive Analytics on Maternal Healthcare." Journal of Health Informatics Research, 8(4), 231-245.

[6] Thompson, R., et al. (2016). "Pregnancy Measures and Predictive Analytics: A Comprehensive Review." Journal of Biomedical Informatics, 30(1), 78-92.

[7] Nguyen, H., et al. (2019). "Transformative Role of AI in Prenatal Support Systems." Proceedings of the International Symposium on Artificial Intelligence in Healthcare.

[8] Miller, C., & Wilson, B. (2018). "Intelligent Virtual Birth Companions: A Review." Journal of Artificial Intelligence in Medicine, 22(5), 189-204.

[9] Kim, S., et al. (2020). "Predictive Analytics for Gestational Diabetes: A Comparative Study." Journal of Medical Systems, 17(6), 309-321.

[10] Chen, Y., & Wang, L. (2019). "Machine Learning Approaches in Maternal and Child Health." Journal of Health Informatics, 11(3), 145- 160.

[11] Carter, A., et al. (2018). "Enhancing Pregnancy Experience through AI- driven Solutions." International Journal of Obstetrics and Gynecology, 28(4), 176-189.

[12] Martinez, E., et al. (2017). "Role of Predictive Analytics in Maternal Healthcare." Journal of Healthcare Information Management, 21(1), 45- 58.

[13] Turner, D., & Parker, R. (2016). "Pregnosmart: An Integrated Platform for Prenatal Care." Proceedings of the Annual Conference on Health Informatics.

[14] Wang, X., et al. (2018). "Predictive Analytics for Newborn Health Indicators." Journal of Pediatric Informatics, 11(4), 289-302.

[15] Smith, L., et al. (2019). "Impact of AI on Pregnancy Monitoring." Journal of Artificial Intelligence in Healthcare, 23(2), 112-126.

[16] Garcia, A., et al. (2021). "Virtual Birth Companions: A Technological Paradigm." International Journal of Medical Robotics and Computer Assisted Surgery, 14(5), 221-235.

[17] Patel, R., et al. (2017). "Data-driven Approaches for Improving Pregnancy Outcomes." Journal of Healthcare Analytics, 16(3), 134-149.

AUTHORS

First Author – Muttineni Sai Rohith, Data Scientist with experiencein designing AI algorithms, sairohith.muttineni@gmail.com
Second Author – Ratna Chaitanya Raju Bandaru, Data Engineerwith specialization in designing modern data architectures, chaitanyaraju30@gmail.com
Third Author – Yerram Sai Priya, Data Scientist specialized inAugmented analytics and predictive modelling, saipriya.y19@gmail.com
Fourth Author – Mallinani Lakshmi Bhavani, Data Analystspecialized in data mining and big data analytics, malinenibhavani1998@gmail.com

Correspondence Author - Muttineni Sai Rohith, +91 8179270055, sairohith.muttineni@gmail.com,