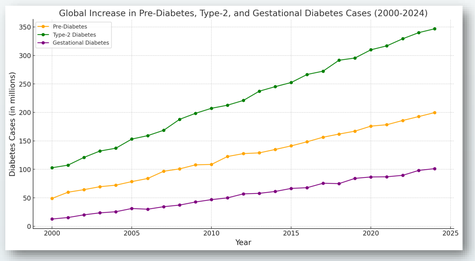
REAL TIME AI POWERED HEALTH COACH FOR DIABETES

School of Science and Engineering-University of Missouri Kansas City
DATA DIVERS (Ashna Ali, Sai Kiran Basetty, Sreevardhan Reddy Soma, Venkata Sai Veeramalla)

Introduction

- Diabetes is a critical global health challenge, with Type-2 and gestational diabetes on the rise, often leading to severe complications like cardiovascular disease and kidney failure if undetected.
- Our project harnesses AI to build a predictive model for early diabetes diagnosis, targeting pre-diabetes, Type-2, and gestational diabetes.
- By integrating demographic, lifestyle, and biometric data, and leveraging machine learning techniques such as Gradient Boosting and LSTM deep learning models.



Objectives

- **Early detection and Intervention**:
  By identifying high-risk individuals early, this project enables timely intervention, reducing the likelihood of diabetes progression and severe complications.
- **Personalized Health Guidance**:
  With AI-driven insights tailored to individual risk factors, patients receive customized recommendations, improving lifestyle management and adherence.
- **Reduced Healthcare Costs**:
  Preventing or delaying diabetes onset lowers the need for expensive treatments, reducing the financial burden on both patients and healthcare systems.
- **Accessible and Scalable Solution**:
  Once deployed, the model can be accessed widely, providing a cost-effective and scalable approach to diabetes screening and prevention, especially in underserved areas.

Methodology

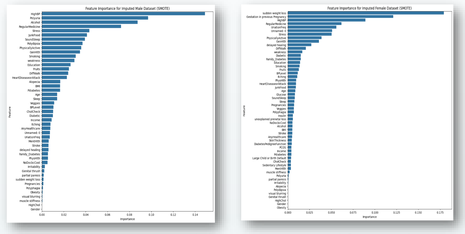
Data Collection & Preprocessing:

- Sources: Collected five datasets from sources like GitHub, Kaggle, and CDC, including specialized datasets like the PIMA Indians dataset and the BRFSS survey from CDC.

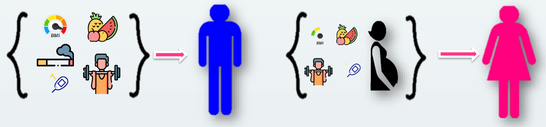
kaggle GitHub + CDC U.S. CENTERS FOR DISEASE CONTROL AND PREVENTION BRFSS Behavioral Risk Factor Surveillance System →

Methodology

- Feature Selection: Filtered features relevant to diabetes diagnosis (e.g., physical activity, diet, age, BMI).



- Gender-Based Split: Separated data into male and female datasets to handle gender-specific diagnoses (e.g., gestational diabetes for females).



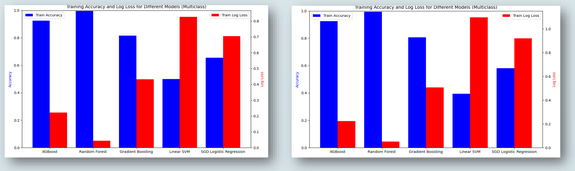
- Encoding & Scaling: Label-encoded categorical features and normalized numerical values; added a 'Patient_ID' column to organize data for time series analysis.

CGM Data Simulation:

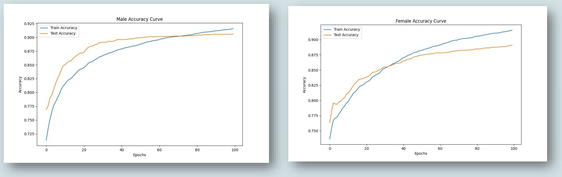
- Data Generation: Simulated Continuous Glucose Monitoring (CGM) data for 24 timestamps per day across all four diabetes classes.
- Preprocessing: Converted timestamps to seconds after midnight and normalized glucose readings and time using MinMax scaling.

Model Selection (Structured Data):

- Model Testing: Evaluated XGBoost, Random Forest, Gradient Boosting, Linear SVM, and SGD Logistic Regression.

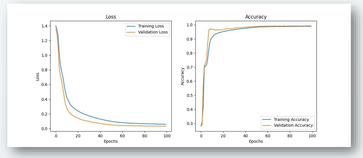


- Results: XGBoost and Random Forest outperformed others, with XGBoost chosen as the final model due to reliable accuracy (around 90-85%).



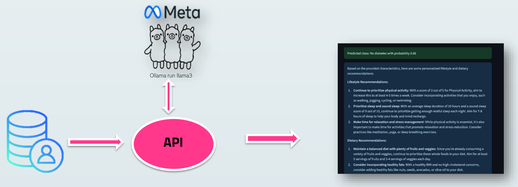
CGM Model (Time-Series Data):

- LSTM Model: Developed an LSTM model for classifying diabetes status based on CGM data, optimized for temporal patterns.
- Preprocessing: Mapped condition and gender to numeric codes, converted timestamps, applied MinMax scaling, and reshaped data into 24-hour sequences per patient.
- Training: Trained over 100 epochs with dropout, regularization, early stopping, and learning rate reduction to prevent overfitting.
- Outcome: Achieved effective classification accuracy, enabling reliable predictions of diabetes status from CGM data.



LLaMA (LLM)FOR RECOMMENDATION:

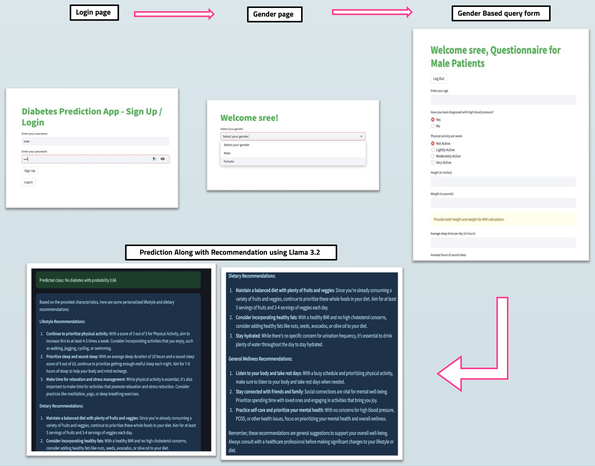
- Personalized Input Collection: Users provide health details like age, BMI, and lifestyle factors through a simple questionnaire.
- LLaMA 3.2 Processing: The model analyzes formatted data via an API, generating personalized health recommendations.



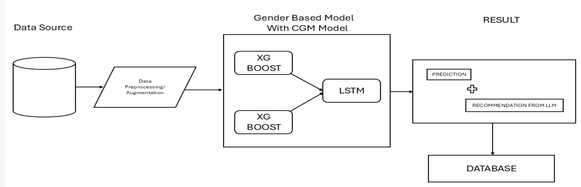
- Immediate, Actionable Guidance: Users instantly receive tailored advice to manage diabetes risk and enhance well-being.

RESULTS:

At the End the user is able to know the predicted result along with recommendations.



Workflow



Future Work

- **Patient Compliance and Data Quality**: Ensuring accurate data is challenging due to inconsistent patient logging. Future improvements will leverage wearable devices for automated, high-quality data collection.
- **Scalability for Real-time Data Handling**: Real-time data from multiple users requires a robust backend. Future enhancements will focus on scalable infrastructure to manage simultaneous user data streams efficiently.

Conclusion

- Our project harnesses machine learning, structured health data, and continuous glucose monitoring to transform early diabetes detection and management.
- With separate models tailored for male and female health profiles, we enhance prediction accuracy and provide personalized, actionable insights.
- Empowering users in the fight against diabetes and promoting better health outcomes.

Reference

1. **2024 Fall CS5588 Data Science Capstone** (Instructor: Dr. Yugi Lee)
2. Hasan, Md. K., Alam, Md. A., Das, D., Hossain, E., & Hasan, M. (2020). Diabetes Prediction Using Ensembling of Different Machine Learning Classifiers.IEEEAccess,8,7651676531. https://doi.org/10.1109/access.2020.2989857.
3. Wu, Y.-T., Zhang, C., Ben W.J. Mol, Kawai, A., Li, C., Chen, L., Wang, Y., Sheng, J.-Z., Fan,J., Shi, Y., & Huang, H.-F. (2021). Early Prediction of Gestational Diabetes Mellitus in the Chinese Population via Advanced Machine Learning. The Journal of Clinical Endocrinology and Metabolism 106(3),e1191e1205. https://doi.org/10.1210/clinem/dgaa899.
4. Papers with Code - Diabetes Prediction. (2022). Paperswithcode.com. https://paperswithcode.com/task/diabetes-prediction.
5. Diabetes Prediction using Machine Learning. (n.d.). Kaggle.com. https://www.kaggle.com/code/ahmetcankaraolan/diabetes-prediction-using-machine-learning.

Scan This QR code for more project details :