**\*\*Proof of Concept: Predictive Modeling for Proactive Diabetes Management\*\***

**\*\*Problem Statement\*\***

**Frequent rehospitalization and emergency room (ER) visits among diabetic patients are a major challenge for healthcare systems and patient well-being. These events not only increase healthcare costs but also disrupt patient care and quality of life. To address this issue, we propose a proactive approach that leverages data analysis techniques to identify diabetic individuals at high risk of rehospitalization and ER visits.**

**\*\*Proposed Solution\*\***

**Our proposed solution encompasses a comprehensive framework for predictive modeling and risk stratification in diabetes management. By utilizing machine learning and graph-based methods, we aim to develop interpretable models that can accurately predict rehospitalization and ER visit outcomes, enabling early interventions and personalized care strategies.**

**\*\*Key Learnings and Adaptations\*\***

**Building upon the success of the Long COVID project, we have incorporated key learnings and adaptations into our diabetes management approach:**

**\* \*\*Data Relevance:\*\* Ensure that only relevant features and data are used for diabetes management.**

**\* \*\*Label Change:\*\* Adapt to any changes in data labels or create new labels if necessary.**

**\* \*\*Handling Multiple Records:\*\* Be prepared to handle cases where one diabetic patient may have multiple visits and records.**

**\* \*\*Encoding and Aggregation:\*\* Utilize appropriate encoding methods and aggregation strategies for diabetic data.**

**\* \*\*Model Performance:\*\* Experiment with various machine learning models, evaluate their performance, and consider both positive and negative labels.**

**\*\*Methodology\*\***

**Our methodology consists of several key steps:**

**1. \*\*Data Preprocessing:\*\* Employ MinMaxScaler normalization to ensure uniform feature scaling.**

**2. \*\*Sampling:\*\* Address class imbalance by dividing the dataset into two classes, balancing the sample dataset using under-sampling and stratified sampling.**

**3. \*\*Machine Learning Models:\*\* Leverage various classifiers, including Support Vector Classifier, Random Forest Classifier, Gradient Boosting Classifier, XGBoost Classifier, and LightGBM Classifier, to predict rehospitalization and ER visit outcomes.**

**4. \*\*Neural Network Configuration:\*\* Introduce a Sequential Model with Dense Layers and ReLU activation functions.**

**5. \*\*Loss Function:\*\* Employ Binary Cross-Entropy as the loss function, and the Adam optimizer to fine-tune the model.**

**\*\*Expected Outcomes\*\***

**The project is expected to yield several significant outcomes:**

**1. \*\*Comprehensive Literature Review:\*\* Assess the status of deep learning and GBDT models in tabular data analysis.**

**2. \*\*Model Interpretability:\*\* Investigate techniques for explaining gradient boosting models' predictions on raw tabular data using SHAP and LIME.**

**3. \*\*Feature Identification:\*\* Identify and leverage relevant features from a comprehensive dataset of diabetic patient records.**

**4. \*\*Predictive Model Development:\*\* Create interpretable machine learning algorithms capable of predicting rehospitalization and ER visit outcomes.**

**5. \*\*Predictive Model Evaluation:\*\* Evaluate the performance of predictive models using appropriate metrics.**

**6. \*\*Graph-Based Insights:\*\* Investigate the application of patient interaction graphs and symptom co-occurrence networks.**

**7. \*\*Model Selection and Custom Loss Functions:\*\* Select and fine-tune the most suitable machine learning models and introduce custom loss functions.**

**\*\*Conclusion\*\***

**Our proposed approach for predictive modeling in diabetes management has the potential to significantly improve patient care and reduce healthcare costs. By proactively identifying diabetic individuals at risk of rehospitalization and ER visits, we can enable early interventions and personalized care strategies, ultimately improving patient outcomes and quality of life**