**RAMYA R**

**AU711021104037**

# **Phase 1: Problem Definition and Design Thinking**

**AI Based Diabetes Prediction System**

**AI Based Diabetes Prediction System**

# **Phase 1: Problem Definition and Design Thinking**

## Problem Definition:

## The problem is to build an AI-powered diabetes prediction system that uses machine learning algorithms to analyze medical data and predict the likelihood of an individual developing diabetes. The system aims to provide early risk assessment and personalized preventive measures, allowing individuals to take proactive actions to manage their health.

## Design Thinking:

* Design thinking methodology for developing an AI-powered diabetes prediction system is impressive. It covers each of the essential steps, such as:

**Data Collection:**

* Ensure that the dataset is representative and diverse, covering a broad range of demographics and medical conditions.
* Consider collecting longitudinal data to track changes in health over time.
* Pay attention to data quality and accuracy to avoid introducing bias into the model.

**Data Preprocessing:**

* Handle missing data appropriately (imputation or removal) and consider using techniques like mean imputation or regression imputation.
* Normalize numerical features to have a consistent scale (e.g., using Min-Max scaling or Z-score normalization).
* Address class imbalance if present by oversampling the minority class or using techniques like Synthetic Minority Over-sampling Technique (SMOTE).

**Feature Selection:**

* Use domain knowledge and exploratory data analysis to identify which features are most relevant for diabetes prediction.
* Consider techniques like recursive feature elimination (RFE) or feature importance scores from tree-based models to aid in feature selection.

**Model Selection:**

* Experiment with various machine learning algorithms, but also consider deep learning models (e.g., neural networks) for complex patterns in the data.
* Use cross-validation to assess model performance and ensure it generalizes well to unseen data.
* Hyperparameter tuning is crucial. Utilize techniques like grid search or random search to optimize model parameters.

**Evaluation:**

* Choose evaluation metrics that align with the problem and its implications. For medical diagnosis, metrics like sensitivity (recall) and specificity are often more informative than accuracy.
* Consider using a confusion matrix to visualize true positives, true negatives, false positives, and false negatives.
* ROC-AUC is useful for assessing the model's ability to discriminate between classes.

**Iterative Improvement:**

* Continuously monitor the model's performance and retrain it with updated data as new information becomes available.
* Explore advanced techniques like ensemble methods (e.g., stacking) and model interpretability techniques (e.g., SHAP values) to enhance both prediction accuracy and interpretability.
* Conduct rigorous A/B testing when introducing significant changes to the model to ensure improvements translate into real-world benefits