SYNOPSIS REPORT

PROJECT TITLE

MEDSYNTH: PREDICTIVE ANALYTICS FOR SYNTHETIC PATIENT DATA

SUBMITTED BY:

- **FIDA TANEEM**
- **ISHA SHRIVASTAVA**
- **MANTHAN M Y**
- **NISCHITHA KN**
- **PADAKANDLA VENKATA NAGA SAI HASINI**

SUBMITTED TO:

GLOBAL NEXT CONSULTING INDIA PRIVATE LIMITED

Multi-Morbidity Risk Prediction System - Project Synopsis

Project Overview

This project develops a comprehensive **Medical Risk Prediction Dashboard** designed to assess multi-morbidity risk in patients through an intelligent healthcare analytics system. The platform determines whether a patient has more than two concurrent medical conditions using advanced machine learning techniques applied to key health and demographic features. The system is specifically engineered to support healthcare professionals in proactive patient care, early intervention strategies, and optimal resource allocation.

Objectives

Primary Goal: Create a robust, end-to-end machine learning pipeline capable of accurately predicting multi-morbidity risk (>2 medical conditions) from patient health data with real-time probability scoring.

Secondary Goals:

- Implement comprehensive data preprocessing and quality assurance protocols for healthcare data
- Apply advanced statistical analysis and dimensionality reduction techniques optimized for medical datasets
- Integrate synthetic data generation to enhance model training and address healthcare data limitations
- Deploy a dual-mode solution supporting both individual patient assessment and batch processing
- Ensure model interpretability with clear risk probability visualization
- Provide scalable architecture for clinical workflow integration

Methodology

Data Processing Pipeline

The project employs a systematic approach beginning with comprehensive data collection from patient records, focusing on key health indicators and demographic variables. Rigorous preprocessing handles missing values, inconsistencies, and data quality issues common in healthcare datasets. Advanced outlier detection algorithms identify and treat anomalous data points to ensure statistical integrity while preserving clinically relevant edge cases.

Statistical Analysis and Feature Engineering

Exploratory data analysis reveals underlying patterns and relationships within multi-morbidity risk factors. Statistical hypothesis testing validates key assumptions and identifies significant predictors across demographic groups. Principal Component Analysis reduces dimensionality while preserving essential medical information, optimizing computational efficiency and model performance for real-time predictions.

Machine Learning Framework

The system integrates both supervised and unsupervised learning approaches to capture diverse patterns in multi-morbidity development. CTGAN (Conditional Tabular Generative Adversarial Network) generates synthetic patient data to address class imbalance and enhance training dataset diversity, particularly crucial for rare condition combinations and underrepresented demographic groups.

Model Selection and Validation

Rigorous evaluation protocols compare multiple algorithms to identify the optimal predictor for multi-morbidity classification. Cross-validation techniques ensure robust performance estimates across diverse patient populations, while feature importance analysis provides clinical interpretability essential for healthcare decision-making.

Technical Innovation

The project incorporates cutting-edge synthetic data generation through CTGAN, addressing critical healthcare analytics challenges including data scarcity, privacy concerns, and demographic bias. The dual-mode architecture supports both

individual patient assessments and batch processing capabilities, enabling scalable deployment across various clinical workflows while maintaining high standards of data security and model performance.

System Features

Single Patient Prediction Mode

The platform provides individual patient risk assessment through an intuitive interface accepting demographic and health parameters. Real-time probability scoring delivers immediate multi-morbidity risk percentages with visual probability indicators, enabling point-of-care decision support.

Batch Processing Capabilities

Healthcare institutions can upload CSV files containing multiple patient records for simultaneous risk assessment. This feature supports population health management, screening programs, and large-scale clinical research initiatives with efficient batch processing (supporting files up to 200MB).

Risk Visualization

The system presents risk probabilities through clear visual indicators:

- **Low Risk**: Visual confirmation with specific probability percentages
- **Progressive Risk Scale**: 0-100% probability visualization
- **Color-coded Risk Categories**: Immediate visual risk stratification

Expected Outcomes

The deliverable system provides healthcare professionals with:

- Accurate multi-morbidity risk predictions with specific probability scores
- Real-time individual patient assessment capabilities
- Batch processing for population health management
- Interpretable insights into demographic and health-based risk factors

- User-friendly interface requiring minimal technical expertise
- Scalable architecture supporting institutional deployment across clinical settings

Clinical Impact

This system enables proactive healthcare management by identifying patients at risk of developing multiple concurrent conditions before they manifest simultaneously. The probability-based scoring system supports evidence-based clinical decision-making, allowing healthcare providers to implement targeted early intervention strategies. This approach potentially reduces healthcare costs through preventive care optimization and improves patient outcomes through personalized risk-stratified care planning.

Implementation Framework

The solution deploys through a sophisticated Streamlit-based **Medical Risk Prediction Dashboard** ensuring accessibility across diverse healthcare settings. The interface seamlessly handles both individual patient data entry and bulk CSV uploads, providing instant risk assessments with confidence intervals and clear probability visualizations. The system's dual operational modes support various clinical workflows from individual consultations to population health screening programs.

Quality Assurance and Validation

Comprehensive testing protocols validate model performance across diverse patient populations and demographic groups. The system demonstrates reliable risk stratification with clear probability thresholds. Artifact management systems ensure reproducibility and version control, supporting continuous improvement and regulatory compliance requirements essential for healthcare applications.

User Experience Design

The dashboard features an intuitive design optimized for healthcare professionals:

- **Clean Interface**: Professional medical dashboard aesthetic
- **Dual Functionality**: Single patient and batch processing modes
- **Visual Feedback**: Clear risk categorization with probability indicators
- **File Upload Integration**: Seamless CSV processing with drag-and-drop functionality
- **Real-time Results**: Immediate probability scoring and risk categorization

This project represents a significant advancement in predictive healthcare analytics, combining methodological rigor with practical clinical applicability specifically tailored for multi-morbidity risk assessment. The system addresses the growing challenge of complex patient care in healthcare systems while providing actionable insights through both individual and population-level analysis capabilities.

Project workflow for predictive Analytics