# IVF Trigger Day Prediction System

End-to-End Machine Learning Pipeline
Automated Clinical Decision Support using Airflow, MLflow & FastAPI

360 DigiTMG®
Digital Transformation | Management | Governance

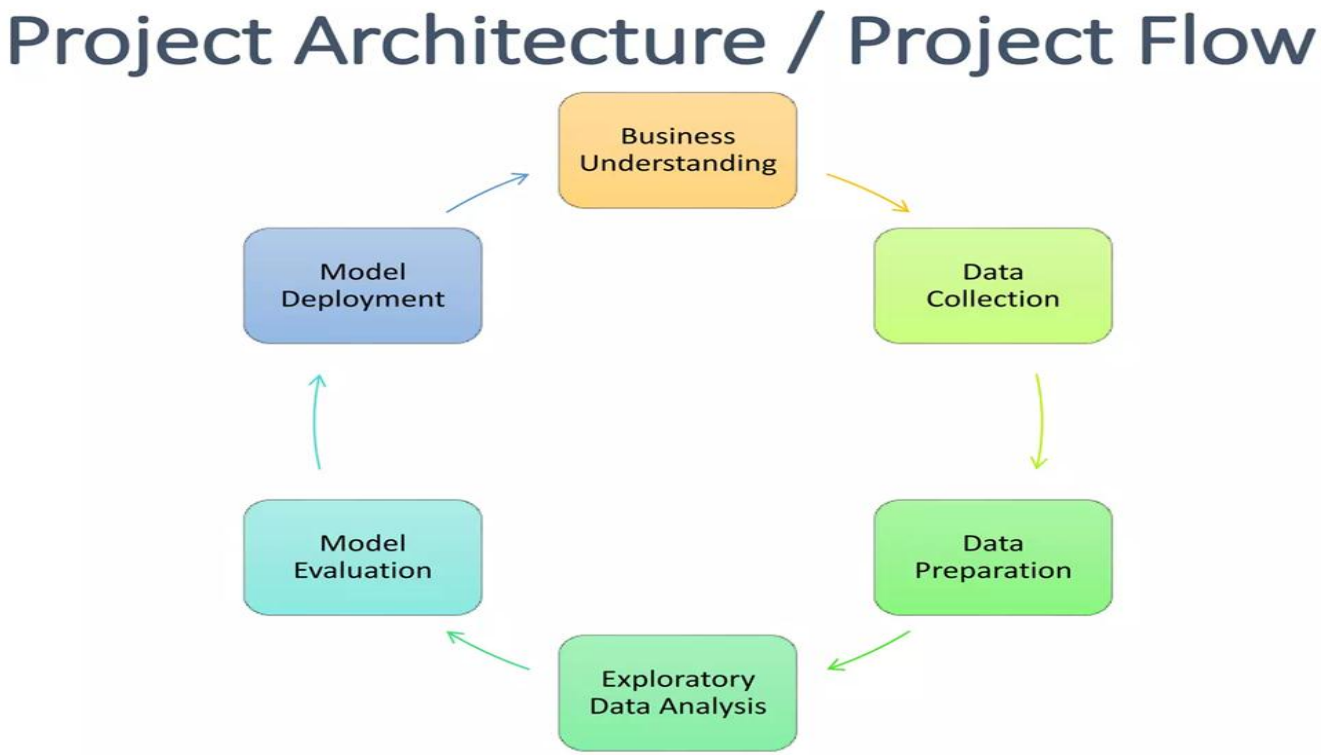
# Team Members

**Name:** Jeshwanth
**Role:** Data Science & ML Pipeline Developer
**LinkedIn:** (Jeshwanth B. - Bengaluru, Karnataka, India | Professional Profile | LinkedIn)

# Contents

- Business objective
- Business Constraints
- Project Architecture
- Data collection and details
- Exploratory Data Analysis
- Visualization
- Modeling
- Evaluation
- Deployment

# Project Overview and Scope



# Business Problem

- IVF trigger day selection is critical for treatment success
- Manual decisions depend heavily on experience
- Risk of inconsistency and sub-optimal outcomes

360 DigiTMG®
Digital Transformation | Management | Governance

# Business Objective

## Objective

- Predict accurate trigger day using patient data

## Constraints

•Limited medical datasets
•Data privacy and quality issues

360 DigiTMG®
Digital Transformation | Management | Governance

# CRISP-ML(Q) Methodology

**There are six stages of CRISP-ML(Q) Methodology**

**1.Business and data understanding**

**2.Data preparation**

**3.model building**

**4.Model evaluation**

**5.Model deployment**

**6.Monitoring and maintenance**

360 DigiTMG®
Digital Transformation | Management | Governance

# Technical Stacks

- Python, Pandas, NumPy
- Scikit-learn
- Apache Airflow
- PostgreSQL / TimescaleDB
- FastAPI & Uvicorn
- MLflow, Great Expectations, Evidently

360 DigiTMG®
Digital Transformation | Management | Governance

# Data Collection and Understanding

- Historical IVF clinical datasets
- Hormonal values and follicle measurements
- Structured patient-cycle records

# Data Information

- Each row represents one IVF cycle
- Features include hormone levels, age, follicle size
- Target variable: Trigger Day

360 DigiTMG®
Digital Transformation | Management | Governance

# Data Dictionary

- **Age:** Patient age
- **FSH, LH, Estradiol:** Hormone levels
- **Follicle Size:** Average follicle diameter
- **Trigger Day:** Predicted output

360 DigiTMG®
Digital Transformation | Management | Governance

# System Requirements

- Python 3.9+
- PostgreSQL database
- 8 GB RAM minimum
- Linux / Windows OS

# Exploratory Data Analysis [EDA]

- Distribution analysis of hormone levels
- Correlation between features and trigger day
- Detection of outliers

# Missing Values Observation

- Missing hormone values observed
- Median imputation applied
- Data validated using Great Expectations

360 DigiTMG®
Digital Transformation | Management | Governance

# Data Preprocessing

- Handling missing values
- Outlier treatment
- Feature scaling

# Data Preprocessing

- Encoding categorical variables
- Train-test split
- Feature engineering

360 DigiTMG®
Digital Transformation | Management | Governance

# Data Visualization

- Hormone trends vs trigger day
- Boxplots and histograms
- Correlation heatmaps

360 DigiTMG®
Digital Transformation | Management | Governance

# Model Building

- Unsupervised clustering applied to identify patient patterns
- Clusters used to understand hormonal and follicle behavior
- Logistic Regression,
- Random Forest Classifier,
- Gradient Boosting,
- Hyperparameter tuning using cross-validation applied
- Random Forest model trained for trigger day prediction
- Feature importance used for interpretability

360 DigiTMG®
Digital Transformation | Management | Governance

# Model Accuracy Comparison

- Models evaluated using accuracy and validation metrics
- Random Forest outperformed other models
- Balanced precision and recall achieved

360 DigiTMG®
Digital Transformation | Management | Governance

# Best Model –

**Random Forest Classifier**

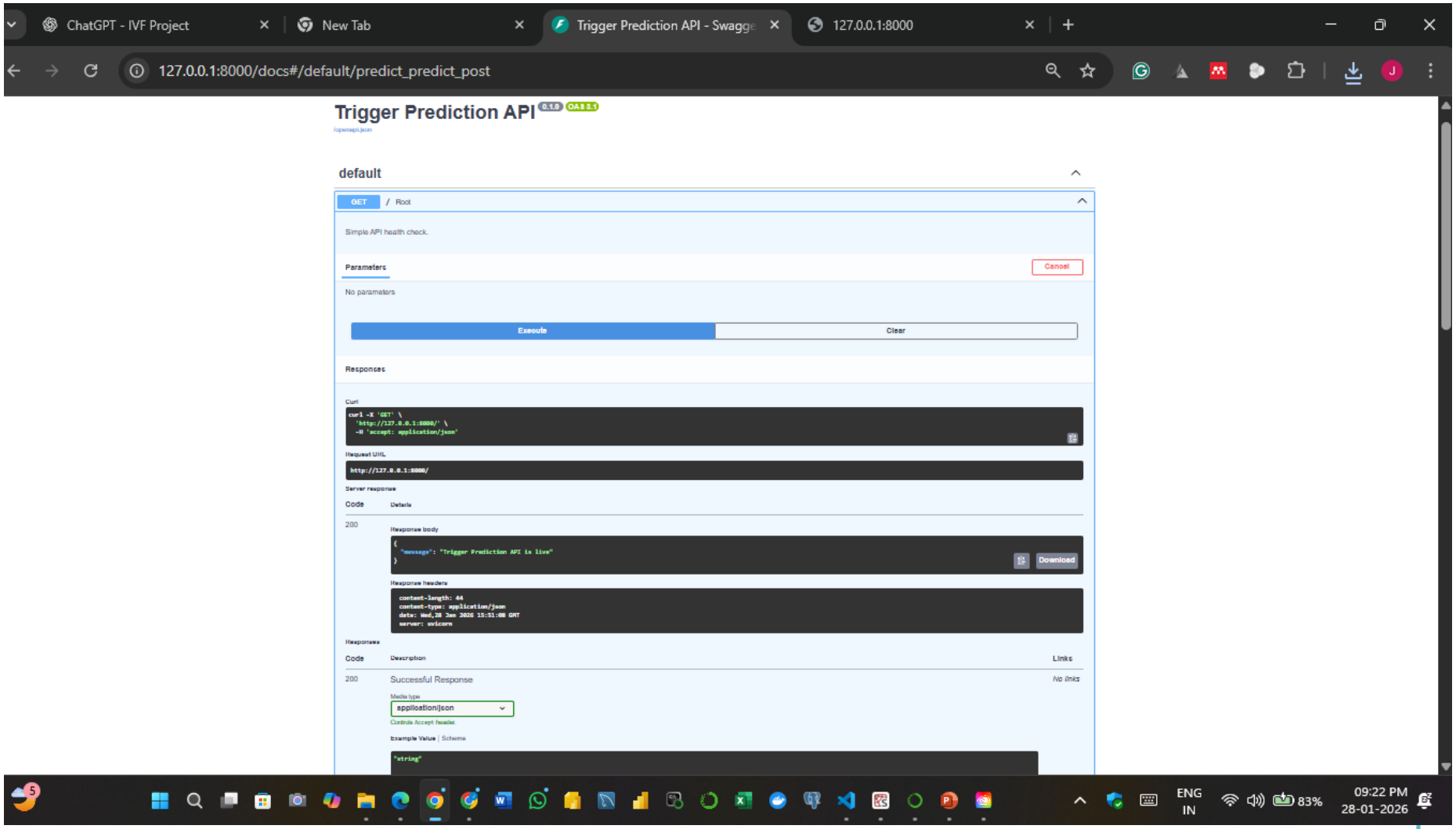
- High prediction accuracy
- Robust to noisy medical data
- Provides feature importance insights

# Model Deployment - Strategy

- Model deployed as REST API using FastAPI
- Automated pipelines using Apache Airflow
- Experiment tracking and model versioning with MLflow

360 DigiTMG®
Digital Transformation | Management | Governance

# Screen shot of output




360 DigiTMG®
Digital Transformation | Management | Governance

# Video of output

https://drive.google.com/file/d/1fvSS60VD8hdaSFXxTvGCo0fG42cTTBDm/view?usp=sharing

# Challenges

- Limited availability of high-quality medical data
- Handling missing and inconsistent hormone values
- Ensuring data privacy and confidentiality
- Balancing model accuracy with interpretability

360 DigiTMG®
Digital Transformation | Management | Governance

# Future Scopes

- Use larger and more diverse clinical datasets
- Improve clustering techniques for patient grouping
- Integrate real-time hospital data systems
- Explore advanced ML models if more data is available

360 DigiTMG®
Digital Transformation | Management | Governance

# Queries ?


360 DigiTMG®
Digital Transformation | Management | Governance


Thanks