Predicting Risk of Comorbidities in Type 2 Diabetes

Ferhat Yilmaz

Associate Director – Ops Excellence

Speaker Disclaimer

- I am an employee of Novo Nordisk Inc. and my business address is 800 Scudders Mill Rd, Plainsboro, NJ 08536

- This presentation is made at the request of Corinium Global Intelligence / RE•WORK. The presenter, although an employee of Novo Nordisk Inc., is presenting on his own behalf. The information provided and any views expressed during the presentation are those of the presenter and may not be attributed to Novo Nordisk Inc.

Agenda
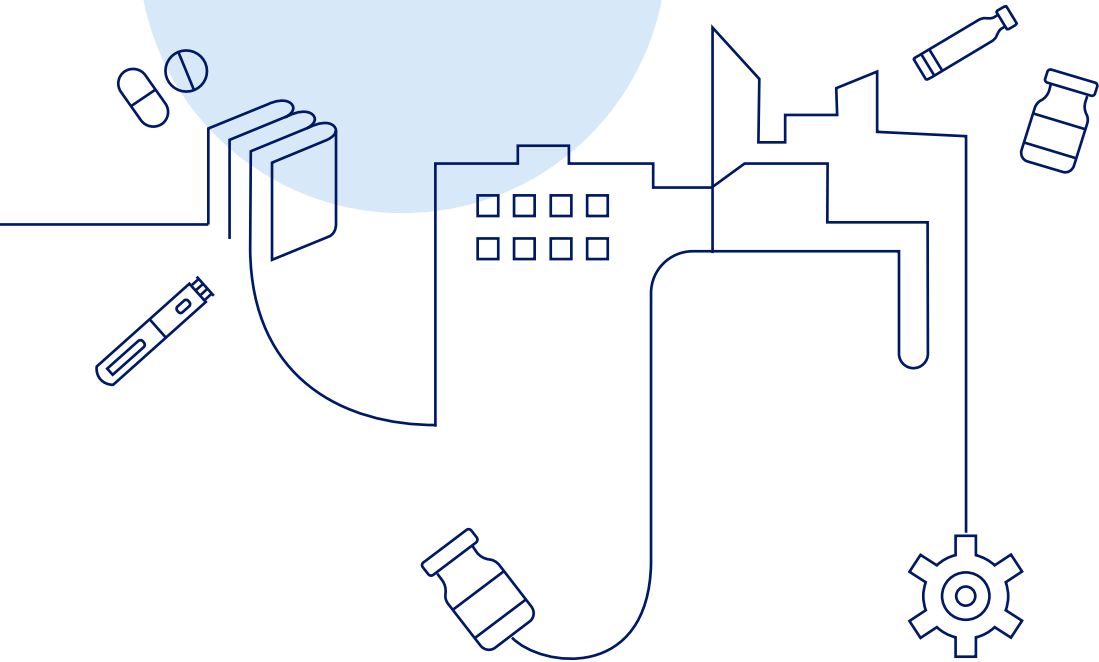
Background & Design

Dataset

Machine Learning Model

Q&A

Background

Why

- Ensure whole-patient insight for comprehensive understanding of health and wellness pain points.
- Identify root causes that limit our ability to develop transformative health innovations.
- Drive improved health outcomes through innovative, patient-centered approaches.

How

- Develop a tool that enhances our clinical insights to understand patients better.
- Deliver more personalized and innovative tools to improve long-term health outcomes.

Design

Outline

End user requirements

Compliance & Regulations

Data ingestion

Model Development

Feature Engineering
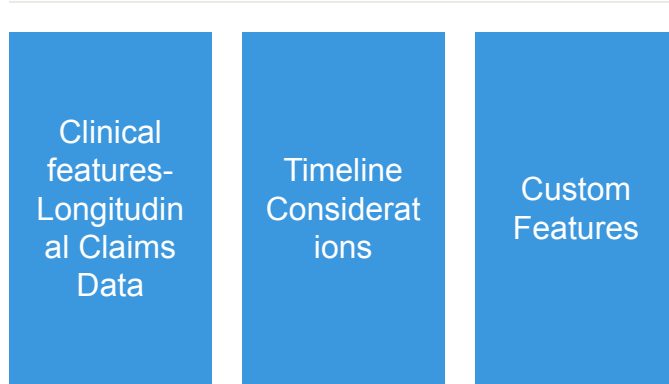
Programmatic Labeling

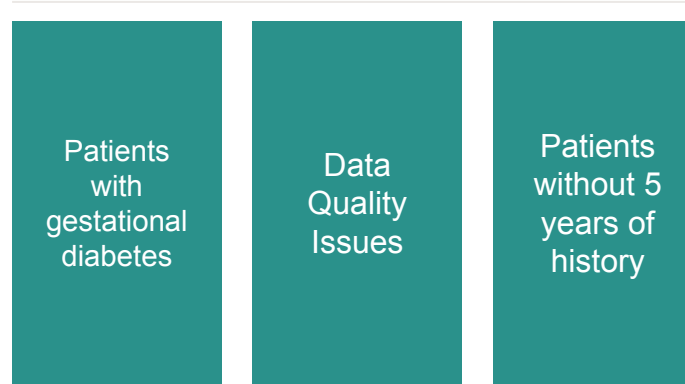Predictive Model

Model in Production

Automation process

Front End

Dataset

Feature Store

| Clinical features- Longitudinal Claims Data | Timeline Considerations | Custom Features |

Exclusion Criteria

| Patients with gestational diabetes | Data Quality Issues | Patients without 5 years of history |

Defining the cohort

| ICD codes | Dx and RX information | Data Availability |

Employ machine learning techniques to predict individuals* with high risk of developing type 2 diabetes comorbidities within 5 years horizon

Focus areas: T2D Comorbidities

Individuals that don't have 5 years of data

Machine Learning Model

Baselines

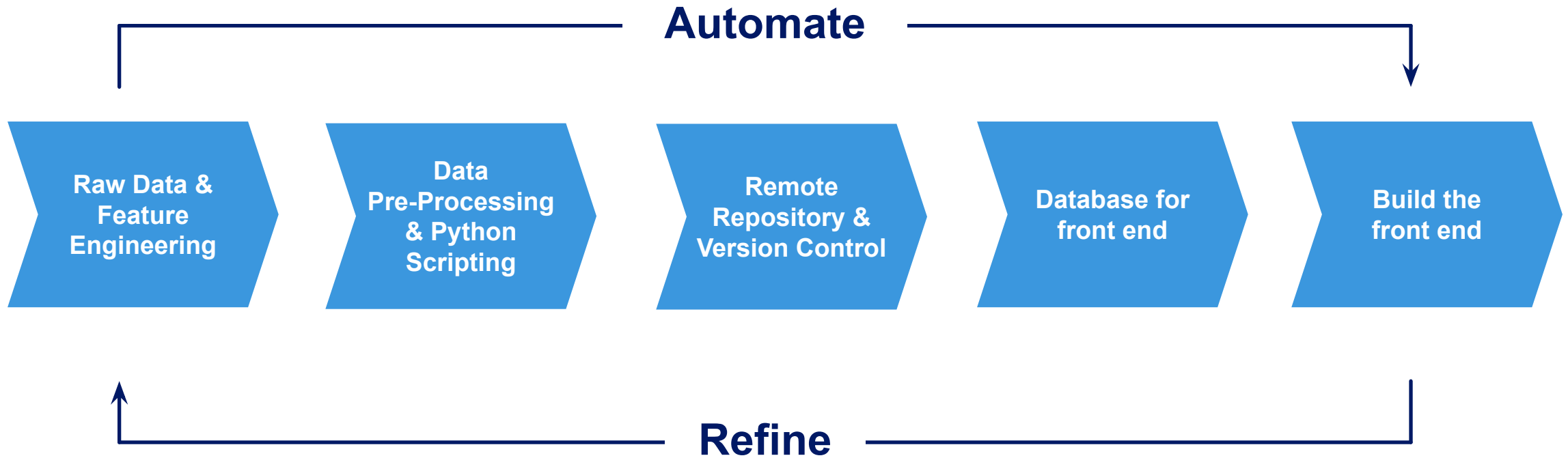- Logistic Regression
- Random Forest
- XGBoost
- LightGBM

Validation

- 5-fold cross validation

Tackling the data imbalance

- Under-sampling
- Threshold setting

Machine Learning Model

Automate

Raw Data & Feature Engineering	Data Pre-Processing & Python Scripting	Remote Repository & Version Control	Database for front end	Build the front end

Refine

Challenges

Time Window:

- How to find the ideal time window?
- Quantitative vs Qualitative approach?

Processing Large Data:

- Computing resources?
- Framework?
- Skillset?
- Cost?

Domain Knowledge:

- DS Skillset?
- Importance of having TA SME in the product team

Q&A