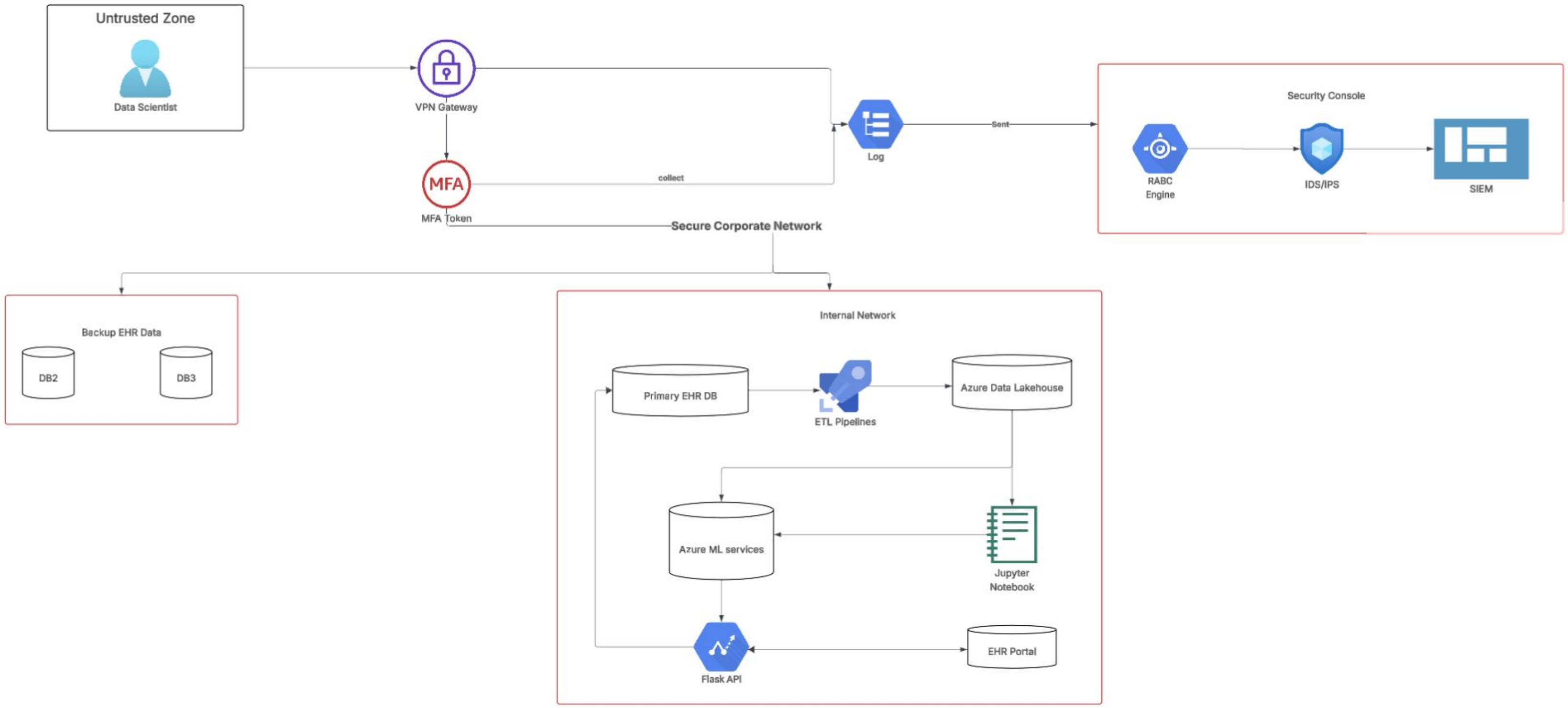
Diabetes Readmission

Group 1: Yicheng Pan, Prabhakar Pandey,
Miaoxuan Zhang, Danny Chen, Vanessa Chen

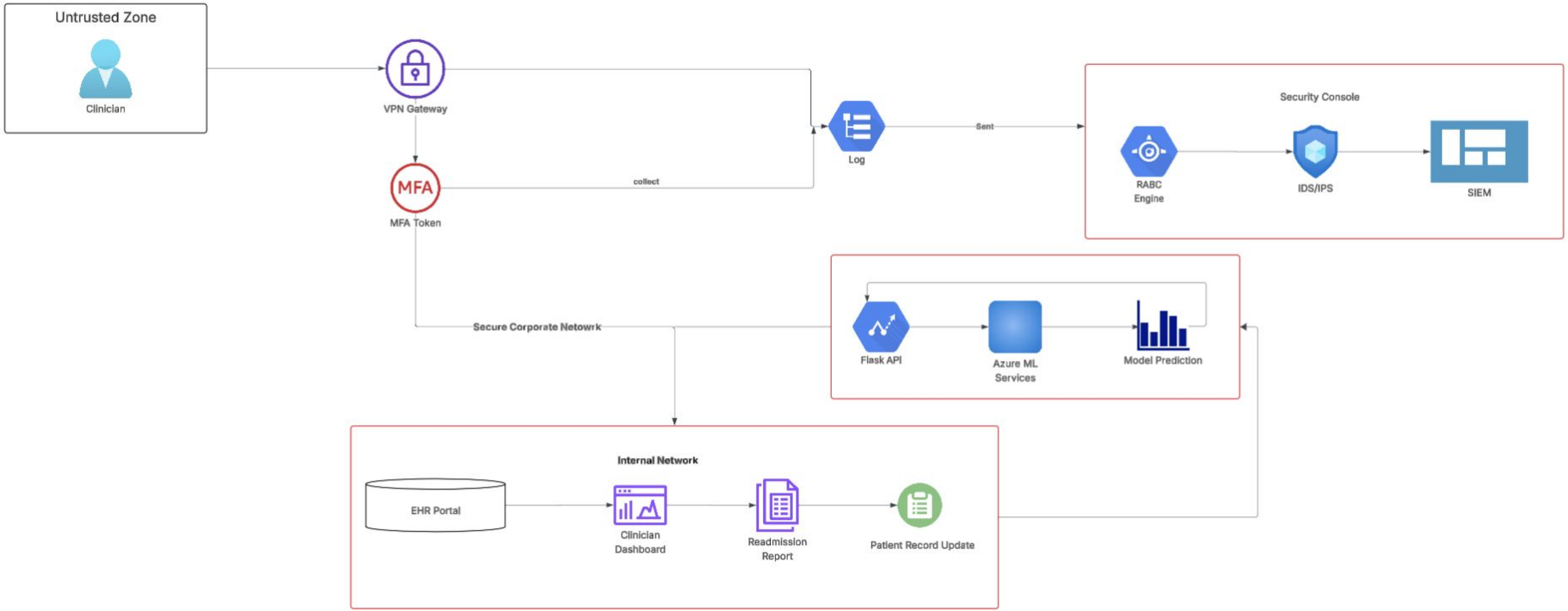
Tools and Technologies

- Lucidchart
- Excel
- Jupyter Notebook
- Python
- Tableau
- Canva
- ChatGPT

Network Diagram (Data Scientist)



Network Diagram (Clinician)



🔒 Secure Entry: Identity Verification and Role Control

🛡️ VPN + MFA

Encrypted tunnel plus identity check — like a security gate with both badge and fingerprint

👥 RBAC (Role-Based Access Control)

Only lets you into rooms you're authorized for — like a hotel key card that opens only your room

📉 Principle of Least Privilege

Just enough access to do your job — no access to unnecessary areas, reducing blast radius

Surveillance Watch and Respond

📊 SIEM (Security Information and Event Management)

Centralizes logs from all systems for unified threat visibility

🛡️ IDS/IPS (Intrusion Detection/Prevention Systems)

Blocks malicious traffic like a digital gatekeeper

📝 Audit Logging

Tracks all access and activity in the environment

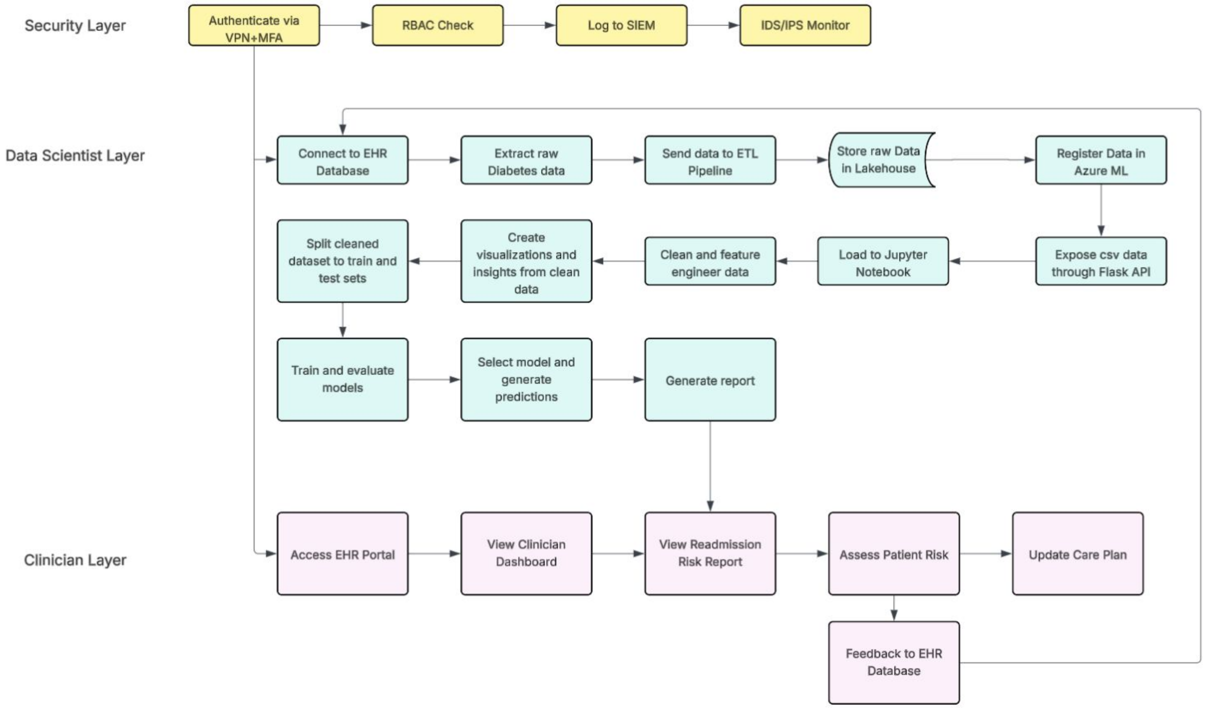
Following The Trails: A Three Layered Approach

- **Security Layer**
  - Protect sensitive health data

- **Data Scientist Layer**
  - Process and analyze data

- **Clinician Layer**
  - Translate analysis into care decisions

Data Flow Diagram


Security Layer
Authenticate via VPN+MFA
RBAC Check
Log to SIEM
IDS/IPS Monitor
Data Scientist Layer
Connect to EHR Database
Extract raw Diabetes data
Send data to ETL Pipeline
Store raw Data in Lakehouse
Register Data in Azure ML
Expose csv data through Flask API
Load to Jupyter Notebook
Clean and feature engineer data
Create visualizations and insights from clean data
Split cleaned dataset to train and test sets
Train and evaluate models
Select model and generate predictions
Generate report
Clinician Layer
Access EHR Portal
View Clinician Dashboard
View Readmission Risk Report
Assess Patient Risk
Update Care Plan
Feedback to EHR Database

Data Retrieval

1. Downloaded the CSV file, “Diabetes 130-US Hospitals for Years 1999-2008”
2. Saved it to a folder in our local device
3. Copied the file’s path
4. Used pd.read_csv() to access the data using the path

Exploratory Data Analysis

1. Insights into Target Variable
2. Regarding Metrics
3. Checking for Missing Data
4. Feature Categorization
5. Bivariate Analysis
6. Correlation Between Numerical Variables

Data Cleaning

Cleaning

1. Drop unnecessary columns + create new columns
2. Get rid of duplicates
3. Check and address missing data
4. Combine NO and >30 together in the readmitted column

Regression

1. Convert categorical features to numeric
2. Used a 70-15-15 data split percentage for training, validation, and test
3. Over-sampled the data as that led to the better model

9 Models Tested

Logistic Regression	Decision Tree	Random Forest
KNN	Linear SVC	Gradient Boosting
Catboost	Stochastic Gradient Descent	XGB

Final Model – Random Forest

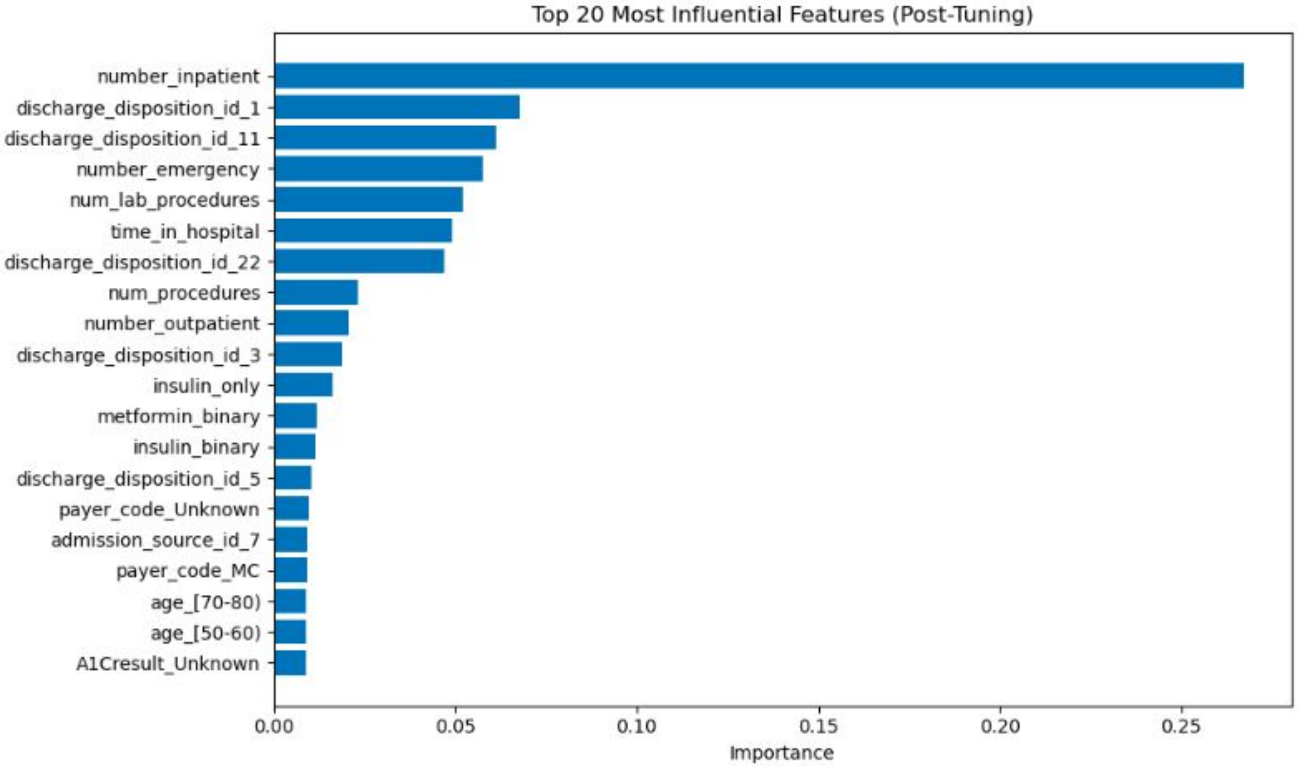
Top 3 Models

	1) Random Forest	2) Logistic Regression	3) Linear SVC
AUC	**0.687**	0.686	0.684
Accuracy	**0.634**	0.684	0.882
Recall	**0.641**	0.579	0.109
Precision	**0.177**	0.191	0.373
Specificity	**0.633**	0.697	0.977
Prevalence	**0.110**	0.110	0.110

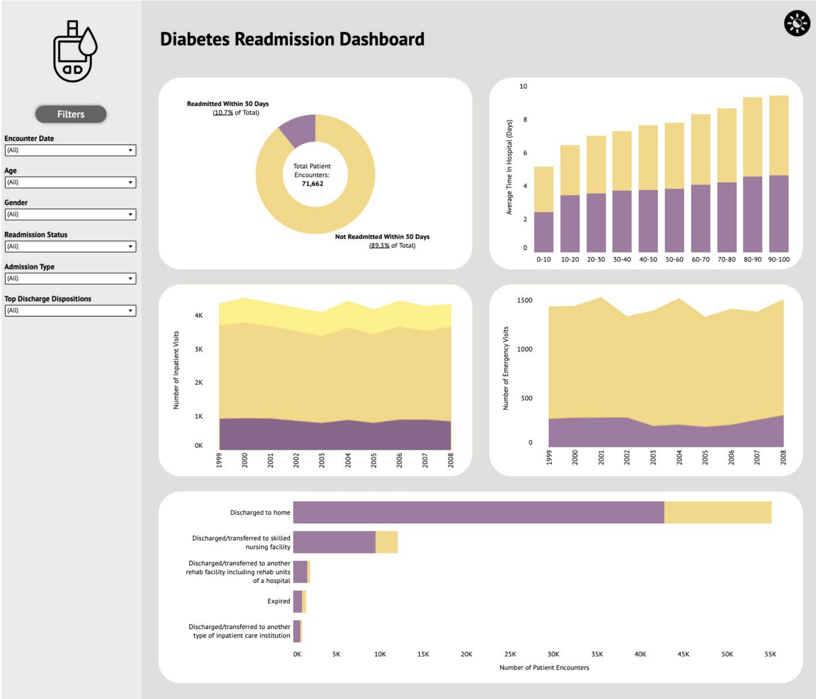
Random Forest – Hyperparameter Tuning

AUC	**0.692**
Accuracy	**0.667**
Recall	**0.610**
Precision	**0.188**
Specificity	**0.674**
Prevalence	**0.110**

Top 20 Most Influential Predictive Features



Visualizing Results



Diabetes Readmission Dashboard

Filters

Encounter Date: (All)

Age: (All)

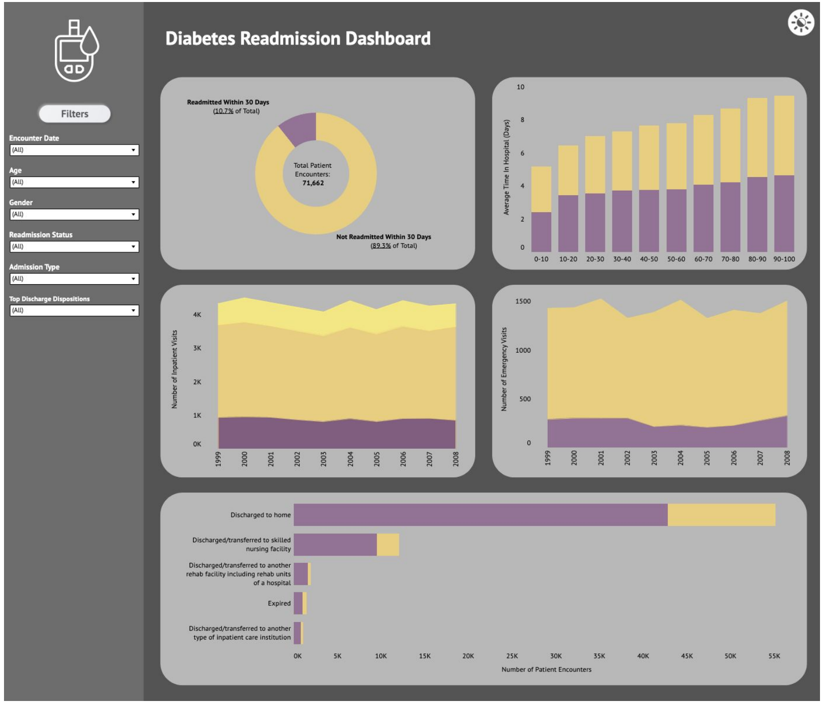
Gender: (All)

Readmission Status: (All)

Admission Type: (All)

Top Discharge Dispositions: (All)

Readmitted Within 30 Days
(10.7% of Total)
Total Patient Encounters: 71,662
Not Readmitted Within 30 Days
(89.3% of Total)
Average Time in Hospital (Days)
0
2
4
6
8
10
0-10
10-20
20-30
30-40
40-50
50-60
60-70
70-80
80-90
90-100
Number of Inpatient Visits
0K
1K
2K
3K
4K
Number of Emergency Visits
0
500
1000
1500
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
Discharged to home
Discharged/transferred to skilled nursing facility
Discharged/transferred to another rehab facility including rehab units of a hospital
Expired
Discharged/transferred to another type of inpatient care institution
0K
5K
10K
15K
20K
25K
30K
35K
40K
45K
50K
55K
Number of Patient Encounters



Diabetes Readmission Dashboard

Filters

Encounter Date: (All)

Age: (All)

Gender: (All)

Readmission Status: (All)

Admission Type: (All)

Top Discharge Dispositions: (All)

Readmitted Within 30 Days
(10.7% of Total)
Total Patient Encounters: 71,662
Not Readmitted Within 30 Days
(89.3% of Total)
Average Time in Hospital (Days)
0
2
4
6
8
10
0-10
10-20
20-30
30-40
40-50
50-60
60-70
70-80
80-90
90-100
Number of Inpatient Visits
0K
1K
2K
3K
4K
Number of Emergency Visits
0
500
1000
1500
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
Discharged to home
Discharged/transferred to skilled nursing facility
Discharged/transferred to another rehab facility including rehab units of a hospital
Expired
Discharged/transferred to another type of inpatient care institution
0K
5K
10K
15K
20K
25K
30K
35K
40K
45K
50K
55K
Number of Patient Encounters

Real World Implications

- Prioritize high risk patient populations
- Strengthen preventive care in outpatient settings
- Identify barriers to metformin treatment
- Promote diabetes medication optimization
- Improve comprehensive discharge planning

Thank You!

Any Questions?