Quest Diagnostics: Predicting Disease From Longitudinal Laboratory Data

Antonin Dauvin and Michael Stern

Quest/ExamOne Advisors: Brian Lanzrath and Yan Yan
Faculty Advisor: Dimitris Bertsimas | PhD Advisor: Michael Li

Project Overview

ExamOne is Quest subsidiary that provides underwriting (risk assessment) solutions to insurers

Quest Diagnostics is the world's largest provider of clinical testing services

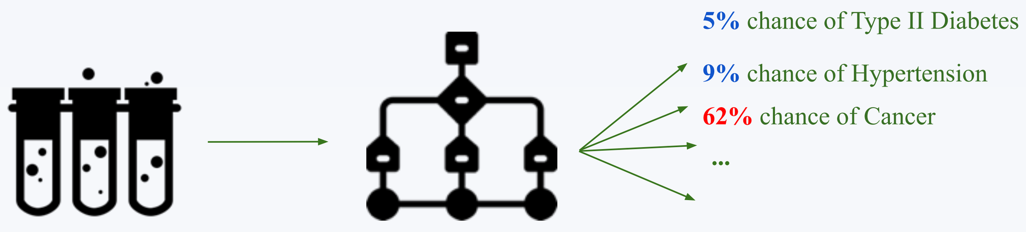
Pulse Check! Quest performs laboratory testing on 1 in 3 Americans

Core Data Science Objective

Input: Patient Laboratory History

Suite of Diagnostic Algorithms

Output: Likelihood of Disease



Business Need

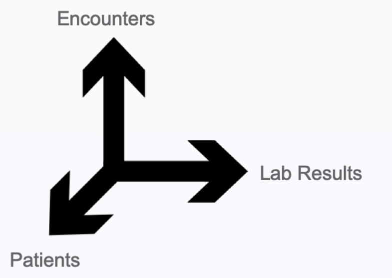
Both Quest and ExamOne have incomplete pictures of patients/applicants due to less-than-reliable diagnosis codes and the fact that people do not get lab tests every day

ExamOne needs analytics to:
- Discover conditions they didn't know were present
- Validate and give a confidence level on existing conditions

Quest needs analytics to:
- Obtain a more complete patient history
- Explore the use of forecasting future disease

Data

Lab Data's 3 Dimensions

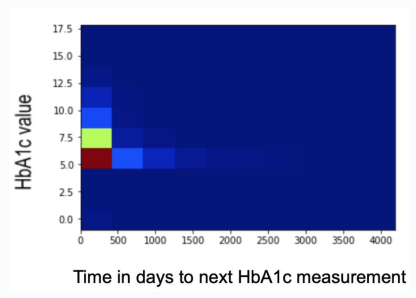


Key Facts
- **120 million** patients
- **2 billion** encounters
- **40 billion** individual lab results
- **3 types of data**: Demographics, Lab Results, Diagnoses

Takeaways From Data Exploration

- Bias exists in the dataset
- There is significant heterogeneity between patients
- Missingness is widespread
- Non-stationarity issues persist

Sickness Correlates with Measurement



Pulse Check! Quest has the largest known laboratory database in the world

Timelines

Spring
Exploratory data analysis, in-depth look at data bias, building of naive baseline models

Early June
Pre-processing; building of multiple data pipelines (e.g. time series)

Late June
Initial model development of a few conditions on 50K patients

Early July
Generalized models to work for all conditions; site-visit to Lenexa, Kansas

Late July
Parallelized code and transferred pipeline to the cloud; built interpretability mechanisms

Early August
Scaled data to 1M patients; improved models; developed of methodology for production

Late August
Final presentation, reporting, and production implementation for Quest/ExamOne

In need of a full check-up?

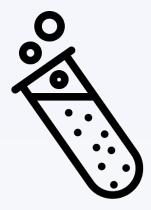
Ask us about the following topics that didn't make the poster
- Precision-Recall curves and methodology
- Dealing with non-stationarity
- Bias in the data
- Data Augmentation
- Bucketing system for ExamOne production
- Time series analysis

Approach

Accelerated Medical School: Translation of Relevant Jargon

LOINC code = Unique identifier for a lab test
ICD code = Unique identifier for a given condition
ICD Header = 3 first letters an ICD code that define broader classes of disease
Encounter = One "visit" to a laboratory (multiple lab tests per encounter)

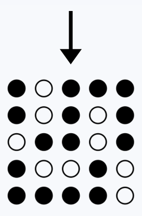
Preprocessing



Lab Results
- Selected top 200 LOINC codes
- Classified LOINC codes into categories (e.g. qualitative, numeric, semi-numeric) and processed accordingly

Diagnoses
- Customized aggregation of ICD codes
- Selected the 100 most common ICD codes and counted occurrences

Aggregation
- Transformed data by patient using aggregated stats (mean, median, max, min, standard deviation) for each lab result
- Created tensor for time series analysis (Patient x Encounter x Lab Test) and engineered time series features (e.g. days since last encounter)



Imputation
- More complex methods: Matrix Completion; specialized imputation based on clustering analysis of missing patterns
- Simpler methods: Mean Imputation; Filling Missing Values with '-1'

Defining The Targets

Target #1: Predicting Medical History

Predict if a patient *has ever been* diagnosed with any given disease

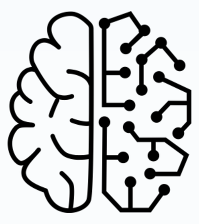
Target #2: Forecasting Future Disease

Predict if a patient *will be diagnosed* with any given disease within two years

Under the Microscope: We modeled 390 different diseases "aggregated" manually or at the header level

Methods

Classic Models
Multivariate Logistic Regression, Random Forest, Gradient Boosted Trees, Decision Trees.



More Complex Models
Optimal Classification Trees, Deep Neural Networks

For Time Series
Long-Short Term Memory, Recurrent Neural Networks

Attention Mechanism

It was also very important to Quest/ExamOne to have a sense of why any given patient received the prediction they did. To address this need, we developed something we called an ***attention mechanism*** and provide an example below

```
Patient ID: 194092265 | Target: E11 | Model:  RFC | Seed:  1
-----------------------------------------------------------------
Probability:  0.835 | Actual Value:  0
-----------------------------------------------------------------
Gender: Female | Age: 60
-----------------------------------------------------------------
Feature # 1 : Hgb A1c MFr Bld_max  |  Number of measurements made:  : 16.0
Value:  7.9

Feature # 2 : Hgb A1c MFr Bld_mean  |  Number of measurements made:  : 16.0
Value:  6.775

Feature # 3 : Hgb A1c MFr Bld_median  |  Number of measurements made:  : 16.0
Value:  6.65

Feature # 4 : Glucose SerPl-mCnc_mean  |  Number of measurements made:  : 15.0
Value:  111.2

Feature # 5 : Microalbumin Ur-mCnc_max  |  Number of measurements made:  : 3.0
Value:  3.9

Feature # 6 : Glucose SerPl-mCnc_max  |  Number of measurements made:  : 15.0
Value:  153.0
```

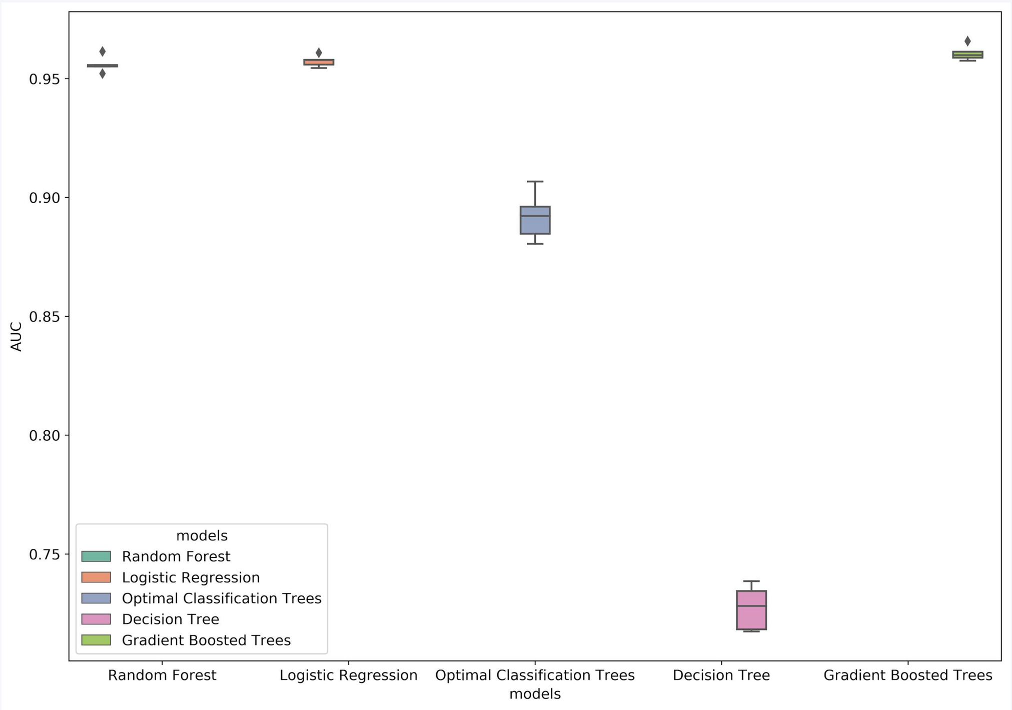
We wanted to understand better why our model was sometimes wrong but we also knew that because of Quest's labels, patients may have been wrongly diagnosed. Further investigation revealed most cases were like the one above:

Among cases where our model disagreed with the label, 90.1% were indeed diabetic according to the clinical definition

Results and Conclusions

Comparison of Model Performance

Sampling of Model Performance for Chronic Kidney Disease

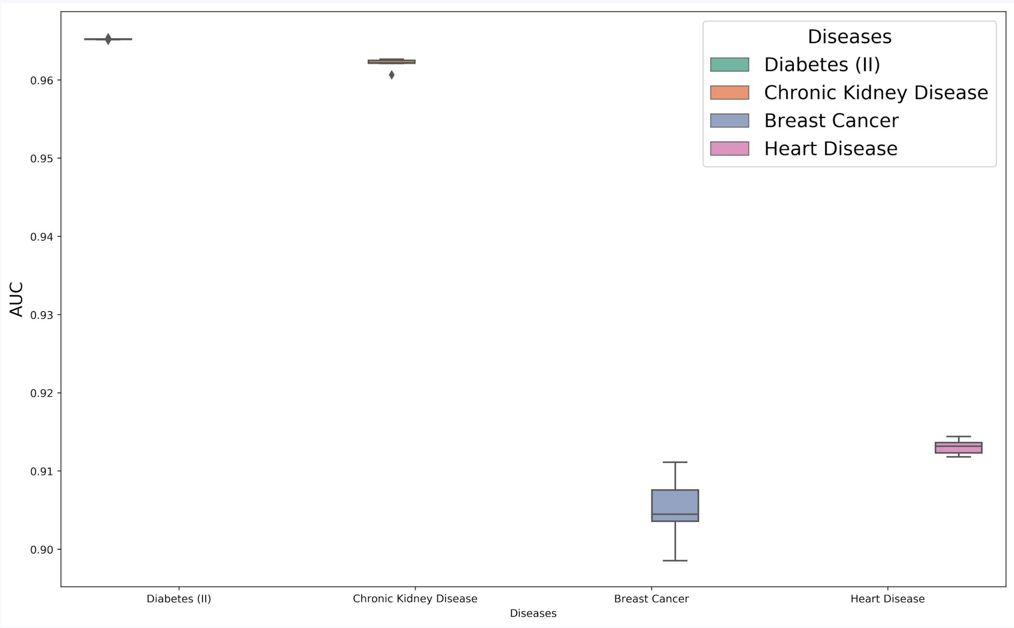


Chosen Treatment: ExamOne preferred Logistic Regression for interpretability and ease of implementation

#1 Predicting Medical History

230 of diseases have an AUC > 0.80
The median AUC for these diseases is **0.87**

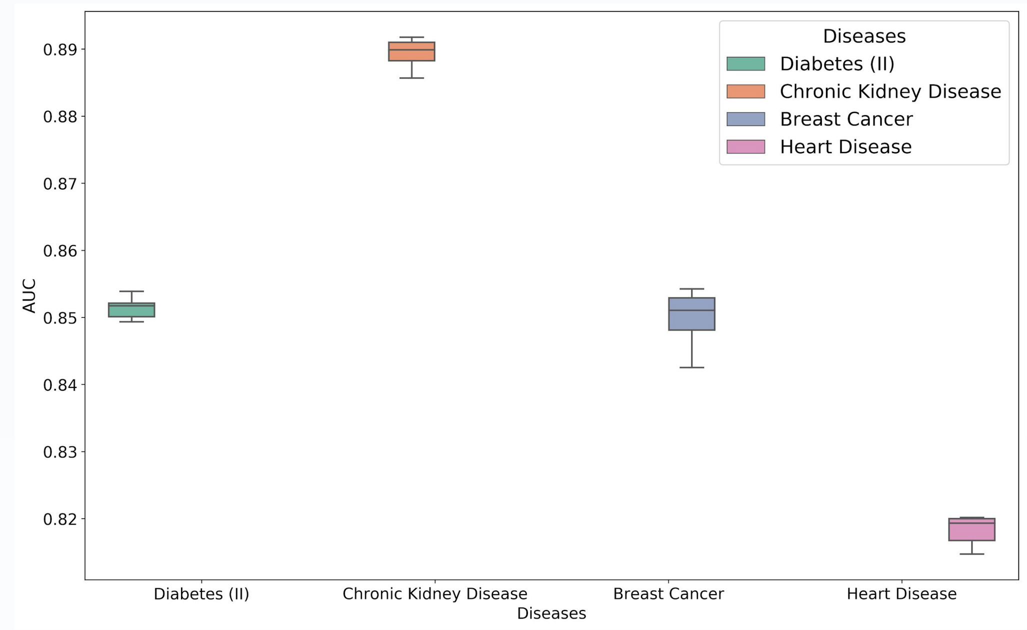
Sampling of Model Performance



#2 Forecasting Future Disease

103 diseases have an AUC > 0.80
The median AUC for these diseases is **0.86**

Sampling of Model Performance



Pulse Check! We expect the forecast of future disease to have lower performance compared to predicting existing medical history because there is inherently more uncertainty in how diseases will progress

Conclusion and Business Impact

- Target #1 significantly exceeded expectations and will be put into production by ExamOne before the end of 2019
- Target #2 provides an evidence-based proof of concept for the use of data science in Quest's core clinical business
- Interpretability mechanisms provide understanding and validation for all stakeholders

$12.6M
Estimated savings for end users

Recommended Next Steps

1. Explore the use of even **more complex models** for increased performance
2. Continue to **scale the data** sample and **customize models** for less common diseases and patient profiles
3. Invest in further exploring forecasting **future health** for **clinical applications** (likely the highest potential return for this work)