AI and Machine Learning: Turn Your Data into Actionable Insights

February 20, 2020
Agenda

Welcome and Introductions
  • Claudia Ellison, Director of Programs, eHealth Initiative

Presenters
  • Cole Erdmann, Director, Clinical Intelligence, Cerner Corporation
  • Navneet Srivastava, Senior Solutions Architect, Amazon Web Services
  • William Feaster, MD, MBA, Vice President, Chief Health Information Officer at CHOC Children’s Hospital
  • Janos Hajagos, Ph.D., Chief of Data Analytics, Stony Brook Medicine
Housekeeping

- All participants are muted

- Use the **Q&A** box to ask a question related to the presentation

- Use the chat box is for *technical difficulties* and other questions / comments

Presentation slides are in the eHI resource Center

[https://www.ehidc.org/resources](https://www.ehidc.org/resources)
eHI’s Mission

To serve as the industry leader in convening executives and multi-stakeholder groups to identify best practices that transform healthcare through the use of technology and innovation
Current Areas of Focus

<table>
<thead>
<tr>
<th>Cost Transparency</th>
<th>Understanding FHIR / APIs / Da Vinci</th>
<th>Protection vs. Access</th>
<th>Non-traditional data sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior Authorization</td>
<td>Policy</td>
<td>HIPAA Part 2</td>
<td>SDOH &amp; PGHD</td>
</tr>
<tr>
<td>Workflow</td>
<td>Cybersecurity</td>
<td>EHR Data for Clinical trials</td>
<td></td>
</tr>
<tr>
<td>Info Blocking</td>
<td>Medical Device Security</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Value Based Care
Interoperability
Privacy & Security
Analytics & SDOH
eHealth Resource Center

Thousands of Resources

• Best Practices
• Reports
• Surveys
• Policy Briefings
• Comment Letters

Analytics
Examine how healthcare data can provide insight across claims, care, clinical, and more.

Medication Adherence
Understand barriers and solutions for medication use.

Privacy & Cybersecurity
Explore the ways in which patient and health system data is being protected.

Consumers
Examine how individuals and families experience healthcare.

Policy
Stay up-to-date with what’s happening with healthcare policy and how it affects stakeholders.

Interoperability
Discover how healthcare technology works together.

Precision Medicine
Examine how customized medical care is evolving.

Telemedicine

Value-Based Care
Discover how patient-centered care is changing healthcare.
Who does eHI work with?

In 2019 eHI worked with
- 1,400+ Payers & Providers
- 3,000+ Stakeholders

How do we reach stakeholders?
- Webinars
- Roundtables
- Task Forces / Work Groups
- Conferences
- Meetings
- HIMSS
eHI Events and Meetings (https://www.ehiddc.org/events)

• eHI VIP Networking Reception at HIMSS
  Tuesday, March 10, 2020
  5:30 – 7:00 pm
  OCCC, EF Overlook
This webinar was made possible through the generosity and support of

Cerner
Today’s Speakers

Cole Erdmann
Director, Clinical Intelligence, Cerner Corporation

Navneet Srivastava
Senior Solutions Architect, Amazon Web Services

William Feaster, MD, MBA
Vice President, Chief Health Information Officer at CHOC Children’s Hospital

Janos Hajagos, Ph.D.
Chief of Data Analytics, Stony Brook Medicine
## Solving data science challenges

### Complex data
- Manage multiple data sources
- Aggregate and normalize disparate data models

### Disparate infrastructure
- Meet demand by automatically provisioning and de-provisioning workloads in an elastic, adaptable environment
- Scale from large, complex to smaller, more targeted analysis

### Limited adoption
- Utilize industry-leading, sophisticated tools
- Access data and models via an easy-to-use, secure environment
Data science ecosystem

HealtheDataLab

Securely access all HealtheIntent data

Build predictive models and algorithms

Implement in workflow

Validate models and algorithms

Single environment
Data lakes and analytics from AWS

- Open and comprehensive
- Secure
- Scalable and durable
- Cost-effective

Machine learning → Analytics → Data lake on AWS

On-premises data movement → Real-time data movement
Data lake components

Metadata

User access

Data movement

Analytics and machine learning

Security/governance

Catalog & Search
Access and search metadata

Access & User Interface
Give your users easy and secure access

Central Storage
Secure, cost-effective Storage in Amazon S3

Data Ingestion
Get your data into S3
Quickly and securely

Protect and Secure
Use entitlements to ensure data is secure and users' identities are verified

Processing & Analytics
Use of predictive and prescriptive analytics to gain better understanding
Apache Hadoop on the data lake

- Distributed processing
- Diverse analytics
  - Batch/script (Hive/Pig)
  - Interactive (Spark, Presto)
  - Real-time (Spark)
  - Machine learning (Spark)
  - NoSQL (HBase)
- For many use-cases
  - Log and clickstream analysis
  - Machine learning
  - Real-time analytics
  - Large-scale analytics
  - Genomics
  - ETL
Enterprise-grade Hadoop and Spark

Scale to any size

- Scale compute (EMR) & storage (S3) independently
- Store and process any amount of data—PB to EBs
- Provision one, hundreds or thousands of nodes
- Auto-scaling
HealtheDataLab architecture
Security is **job zero**

- **People & process**
- **System**
- **Network**
- **Physical**

**Familiar security model**

**Validated and driven by customers’ security experts**

**Benefits all customers**
Shared responsibility model

Customers have their choice of security configuration in the cloud.

AWS is responsible for the security of the cloud.

Customer data

Platform, applications, identity & access management

Operating system, network & firewall configuration

Client-side data
- Encryption & data
- Integrity authentication

Server-side encryption
- (file system and/or data)

Networking traffic
- Protection (encryption integrity, identity)

Software

<table>
<thead>
<tr>
<th>Compute</th>
<th>Storage</th>
<th>Database</th>
<th>Networking</th>
</tr>
</thead>
</table>

Hardware/AWS global infrastructure

<table>
<thead>
<tr>
<th>Regions</th>
<th>Edge locations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Availability zones</td>
<td></td>
</tr>
</tbody>
</table>
Implementing a cloud computing solution for advanced analytics and data science in health care

William W. Feaster, MD, MBA
V.P., Chief Health Information Officer
Children’s Hospital Orange County
Objectives

• Explore the scope and challenges of utilizing health care data
• Explain how we have applied advanced analytics and data science to predict readmissions
• Describe our prior on-premises and current cloud-based computational environments that aids the work of the data scientist
• Illustrate how CHOC Children’s has applied advanced analytics and data science back into the EHR and process of care
Disclaimer
Health care is awash in a sea of data
Health care data

• From the EHR
  • The good news is all patient data in the EHR are readily available at the point of care
  • The bad news is all patient data in the EHR are readily available at the point of care
  • 40% in notes

• Images – radiology, CT, MRI, ultrasound

• Lab data
  • “Omics” of various types – genomic markers for disease risk, pharmacogenomics to guide prescribing, tumor genomics to guide therapies

• Streaming data from acute care monitors, wearables, continuous glucose monitors, etc.

• Finance data, billing data, claims data, satisfaction data
Predictive analytics aim to take advantage of EHR data through computational data science

- Takes into account all available data on all qualifying patients
- Explores associations and statistical interactions between all variables
- Uses a data-driven and statistically sound approach to determine the most important set of variables that impact care for the event of interest
- Predicts the future occurrence of an event or classify patients at high risk of poor outcomes
- Guides caregivers to intervene on patients identified to be at high risk and improves quality of care outcomes
Data science
Applying predictive analytics to an identified clinical question (pre HealtheDataLab)

1. What data is needed?
2. Identify data sources
3. Extract data with CCL/SQL
4. Load data into secondary DB
5. Run analysis with R, SAS, etc
6. Assess AUC
7. Implement algorithm in EHR
8. Need better performance
9. Good performance
Applying data science to predict readmissions
Reducing readmissions

- While CHOC has had great success in applying guidelines to care, and have reduced length of stay and readmissions for certain diseases, our overall 7-day readmissions (solutions for Patient Safety metric) hasn’t decreased for the overall organization, hovering at around 4% or near the national average for children’s hospitals.

- We’ve tried many ways to improve this number without success.

- We’re now turning to data science...
The goal of our original model

Build the most predictive model for 30-day readmissions under the most stringent set of inclusion/exclusion criteria
30-day readmissions

- Study takes into account variables novel to those studies, including:
  - Rothman Index
  - Prior readmissions and whether current admission is a readmission
  - Medication classes
- 30 day predictive algorithm actually predicts 7 day readmissions better than a model specific to 7 days
Performance of the 30-day readmission model

- AUC: 0.79 [0.77, 0.80]
  - Highest for pediatric readmission models with encounters for chemotherapy excluded

- Operating threshold:
  - We chose the following thresholds on analyzing the performance of the model and predicted probabilities:
    - **High risk** – Predicted probability greater than or equal to 0.22
    - **Medium risk** – Predicted probability between 0.11 and 0.22
    - **Low risk** – Predicted probability less than 0.11

- Among “high risk” patients, predicted 3 will be flagged daily of which 1 is guaranteed to be readmitted in the absence of an intervention
Initial clinical application of readmission model

• We initiated the use of our published model via spreadsheets utilized by our inpatient case management staff
  • Those at high readmission risk worked daily from admission to discharge
    • They determine interventions required
• Performance of the model is actually better than we predicted
• For patients identified as:
  • High risk, 52% were readmitted in 30 days
  • Medium risk, 26% were readmitted in 30 days
  • Low risk, 5% were readmitted in 30 days
• Follow up measurements of 7-day readmissions show the rate trending down (statistically significant drop)
Exploring data with HealtheDataLab
In our ideal data science environment

• We would not be limited by access to data
• Data elements needed for analysis would be easy to identify
• We would not have to spend an extended time preprocessing this data
• We would not be limited by computing resources
• We would have access to all available big data analysis tools
Enter HealtheDataLab

- Cerner approached us to be its development partner for a new, client-facing cloud computing and big data analytics and insights platform, powered by AWS

- Single data science environment
  - Complex and disparate data are easily stored and accessible
  - Common big data analytic tool set is readily available
  - Elastic MapReduce for high performance distributed computing
HealtheDataLab data

• Our data from our EHR and other data sources in HealtheIntent® is in HealtheDataLab™ in both identified and deidentified form
  • Data mapped to S3 bucket (encrypted data lake) utilizing FHIR data definitions where available
• Cerner Real-World Data® database – deidentified database of 68.7 million patients, 503.8 million encounters across 600 health care facilities
• We’ve loaded other data into the S3 bucket
  • Rothman Index data to validate prior readmission model
  • Mimic III ICU database – deidentified data from MIT, 46,520 ICU patients and 61,532 ICU stays
Predictive Analytics in HealtheDataLab

- Access control
- Implement Algorithm in EHR
- Assess AUC
- Needs better performance
- Build models using Spark MLlib
- Retrieve data via Spark/Hive as gold standard for big data processing
- Pre-process and transform data using the Spark SQL and DataFrame tools
- More security configurations
- S3 bucket data lake

Access control
Our readmission work in HealtheDataLab

• Validate the environment by replicating prior published study (M0)
• Provide methodological improvement over our previous study (M1)
  • Exposure to more patients and more data elements
• Re-run the analysis on all readmissions, not just unplanned to prepare for work on the HF database (M2)
• Run a multi-center model for predicting general all-cause 30-day readmissions among pediatric-age patients (patients less than 18 years) using the HF database (M3)
  • 1.4 million children hospitalized at 48 hospitals over a 10 year period with higher pediatric volumes in the HF database
The ROC curve illustrates the performance of the models. Model M1 (AUC = 0.8226), Model M2 (AUC = 0.8756), and Model M3: RandomForest (AUC = 0.8446) are compared against the diagonal line representing a random guess.
Current implementation of our readmission predictive model

- We’ve refined our readmission model as described above
- The real challenge is bringing the results of these models back to caretakers in the EHR
  - Patient lists can replace the spreadsheets for care managers
  - Patient specific results can be used in patient care
  - Alerts related to those results can fire as appropriate
Integration into the EHR

Cerner Millennium® database

On premise EDW

On premise server

Cerner EHR application
Patient example
RESEARCH ARTICLE

A Statistical-Learning Model for Unplanned 7-Day Readmission in Pediatrics

Louis Ehwerhemuepha, PhD, Karen Pugh, MSN, RNC, Alex Grant, BS, Sharief Taraman, MD, Anthony Chang, MD, MBA, MPH, MS, Cyril Rakovski, PhD, William Feaster, MD, MBA.

HOSPITAL PEDIATRICS Volume 10, Issue 1, January 2020
Since onboarding HealtheDataLab

- We’re currently developing numerous predictive models for asthma outcomes, early diagnosis of sepsis after ED triage, ICU transfers, rising risk prediction, among other projects
- 6 more papers under current publisher review
- 8 more will be submitted over the next several months
- The real challenge is bringing the results of these models back to caretakers in the EHR
HealtheDataLab: a next-generation analytic environment

Janos G. Hajagos, Ph.D.
Chief of Data Analytics
Stony Brook Medicine
Research Assistant Professor
Department of Biomedical Informatics
Stony Brook University
Stony Brook, New York
Stony Brook Medicine

- Academic medical center (600 beds)
  - Two consecutive years “America’s 100 Best Hospitals” by Health Grades
- Two community hospitals on the eastern end of Long Island
- DSRIP award (2015) managing the Medicaid population of Suffolk County, New York
- Research campus of the State University of New York
- Leading department of Biomedical Informatics combining Engineering and Medicine
Population health: provider map and Medicaid diabetic patients
No single tool or environment

- Data for population health from multiple sources
  - EHRs
  - Administrative systems and claims
  - Public sources
    - ACS: American Community Survey (census data)
    - NPPES: National Plan & Provider Enumeration (NPI database)
- Need tools which can combine data from multiple sources
- Efficiently query large volumes of data
- Build machine learning models with cutting-edge algorithms
Jupyter notebooks are widely used.
Apache SPARK
Data flow

Registries

EHRs

Claims

HealtheIntent platform

Vertica

Tableau

HealtheDataLab

S3

Data syndication

HIDUU tool

AWS

Apache SPARK

Train models

Apply models

Scikit Python runtime
Hemoglobin A1C registry measure prediction

Generate features with Apache Spark

```python
noNull = pd.DataFrame([{'empl_id':empl_id, 'grp_yr':grp_yr},
                      {'service_date_first': first_service_date,
                       'service_date_last': last_service_date,
                       'service_date_first_days': (last_service_date - first_service_date).days,
                       'AIC_first': first_AIC,
                       'AIC_last': last_AIC,
                       'non_diabetic_first': 1 if first_AIC < 6.5 else 0,
                       'diabetic_controlled_first': 1 if first_AIC < 6.5 and first_AIC < 9.0 else 0,
                       'diabetic_uncontrolled_first': 1 if first_AIC > 6.0 else 0,
                       'non_diabetic_last': 1 if last_AIC < 6.5 else 0,
                       'diabetic_controlled_last': 1 if last_AIC < 6.5 and last_AIC < 9.0 else 0,
                       'diabetic_uncontrolled_last': 1 if last_AIC > 6.0 else 0,
                       'gender': gender,
                       'ethnicity': ethnicity,
                       'dbs': dbs,
                       'age': age,
                       'condition_id_disp': condition_id_disp}, index=[0])
```

Test model performance

<table>
<thead>
<tr>
<th>S.No</th>
<th>Model</th>
<th>Parameters</th>
<th>Training AUC-ROC</th>
<th>Testing AUC-ROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Decision Tree Classifier (standard(scale=True), PCA(components=4))</td>
<td>0.9065</td>
<td>0.934</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Random Forest Classifier (random_state=42, min_samples_split=10)</td>
<td>0.7294</td>
<td>0.7176</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>AdaBoost Classifier</td>
<td>0.7019</td>
<td>0.6587</td>
<td></td>
</tr>
</tbody>
</table>
Current projects in progress

• Development of inpatient diabetic scoring criteria for at-risk patients
  • Moved prototyped algorithm in Python to HealtheDataLab

• Aortic aneurysm screening
  • Determine at risk population and associated documents to extract
  • Working with Cerner engineering on making radiology notes accessible in HealtheDataLab

• Development of predictive models for coded diagnosis
  • Data normalization and processing medications, labs, and vitals
  • Future model development and testing with Cerner Real-World Data™
Q&A

Cole Erdmann
Director, Clinical Intelligence, Cerner Corporation

Navneet Srivastava
Senior Solutions Architect, Amazon Web Services

Dr. William Feaster, MD, MBA
Vice President, Chief Health Information Officer at CHOC Children’s Hospital

Janos Hajagos, Ph.D.
Chief of Data Analytics, Stony Brook Medicine
Thank you for participating in today’s webinar!

We would like to thank Cerner for supporting eH’s educational initiatives!