

AWS
re:Invent

LFS401

Cluster adverse events with Amazon Comprehend Medical & Amazon SageMaker

Ujjwal Ratan

Principal Solutions Architect
Amazon Web Services

Aaron Friedman

Principal Partner Solutions Architect
Amazon Web Services

Agenda

Business problem definition

Solution overview

Introduction to key services

Hands-on workshop

Business problem

Adverse events analysis and reporting

- An adverse drug event (ADE) involves harms to patients caused by medication use*
- ADEs are the leading type of nonsurgical adverse event occurring in hospitals in the United States, with an estimated 1.6 million events in 2010*
- According to a report published by The Agency for Healthcare Research and Quality (AHRQ):*

“Overall, patients hospitalized with an ADE have an increased length of stay, higher costs, and increased risk of in-hospital death compared with those not experiencing an ADE.”

* <https://www.hcup-us.ahrq.gov/reports/statbriefs/sb234-Adverse-Drug-Events.pdf>

An example: Oxycodone

Treatment Options > Pain > Oxycodone > Side Effects

 Print  Share

DRUG STATUS

Oxycodone Side Effects

Medically reviewed by Drugs.com. Last updated on Dec 27, 2018.

[Overview](#) **Side Effects** [Dosage](#) [Professional](#) [Tips](#) [Interactions](#) [More](#) 

[Consumer](#) | [Professional](#) | [Managing Side Effects](#)

-  **Availability**
Prescription only
-  **Pregnancy & Lactation**
Risk data available 
-  **CSA Schedule***
High potential for abuse 
-  **Approval History**
Drug history at FDA 
-  **WADA Class**
Anti-Doping Classification 

In Summary

Commonly reported side effects of oxycodone include: constipation, drowsiness, nausea, pruritus, and vomiting. Other side effects include: asthenia, diaphoresis, and xerostomia. See below for a comprehensive list of adverse effects.

How can we analyze adverse events

- Most adverse events are reported as free text/notes without any structure, so analyzing it using conventional techniques is difficult



About Patient

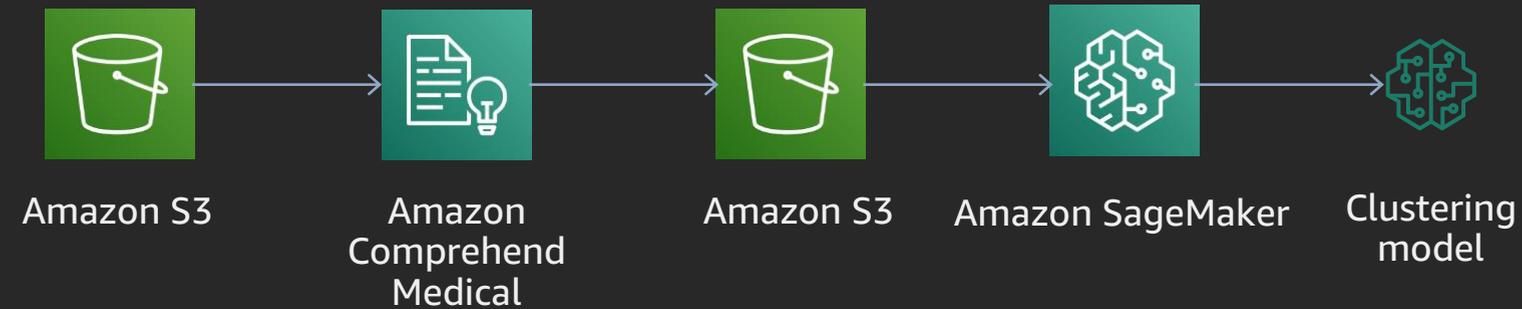
- This complicates aggregate analysis to identify patterns in adverse events data
- Machine learning can help

What types of techniques can help

- Natural language processing can extract key entities and relationships from unstructured text
- Unsupervised-learning approaches like dimensionality reduction and clustering allow unstructured data to be represented in ways that are easier to analyze
- These techniques allow users to get important insights about how drugs react, and also paves the way for supervised learning techniques by providing a labelled dataset

Solution overview

Workflow



Raw data

The raw data files are uploaded into an S3 bucket for the workshop. The dataset we will use is from the University of California, Irvine, Machine Learning repository available [here](#).

Raw data processing

The raw data is processed to extract adverse event terms from the raw clinical notes using Amazon Comprehend Medical. The processed data is stored back into Amazon Simple Storage Service (Amazon S3).

Model training and deployment

A clustering model is trained and deployed using Amazon SageMaker. The model allows grouping of drugs that have similar reported adverse events.

Our data source



UCI  [About](#) [Citation Policy](#) [Donate a Data Set](#) [Contact](#)

Machine Learning Repository
Center for Machine Learning and Intelligent Systems

Repository Web 

[View ALL Data Sets](#)

Drug Review Dataset (Drugs.com) Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: The dataset provides patient reviews on specific drugs along with related conditions and a 10 star patient rating reflecting overall patient satisfaction.

Data Set Characteristics:	Multivariate, Text	Number of Instances:	215063	Area:	Life
Attribute Characteristics:	Integer	Number of Attributes:	6	Date Donated	2018-10-04
Associated Tasks:	Classification, Regression, Clustering	Missing Values?	N/A	Number of Web Hits:	54025

Introduction to key services: Amazon Comprehend Medical

Customer problems

1.2 billion unstructured clinical documents **created per year**

Critical information trapped in these documents

Difficult to extract insights

Extended Explanation:

One of the important ways to improve patient care and accelerate clinical research is by understanding and analyzing the insights and relationships that are trapped in free-form medical text, including hospital admission notes and a patient's medical history.

Today this is achieved by writing and maintaining a set of customized rules for natural language processing software, which are complicated to build, time-consuming to maintain, and fragile. A change to a single classification code name, for example, can impact dozens of hard-coded rules and failing to update a single one of them can result in missed or incorrect data. Machine learning can change all that with models that can reliably understand the medical information in unstructured text, identify meaningful relationships, and improve over-time.

Amazon Comprehend Medical

Medical Named Entity and
Relationship Extraction (NERe API)

Protected Health Information Identification
(PHId API)*

Entities

- Medication
- Medical condition
- Test, Treatments, and Procedures
- Anatomy
- Protected Health Information (PHI)

Relationship extraction

- Medication and dosage
- Test and result
- And more

Entity traits

- Negation
- Diagnosis, Sign, or Symptom

Service is HIPAA Eligible and "Stateless" (no customer data stored)

*This API extracts Protected Health Information only at lower cost

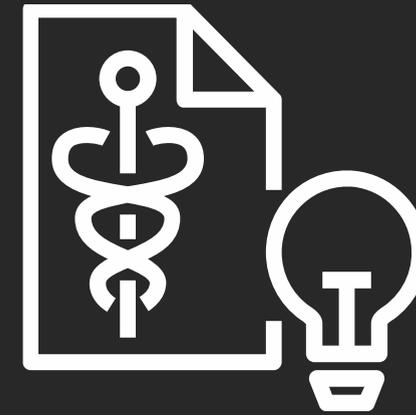
Distill a complex process into a simple API call



Amazon Comprehend

A natural language processing (NLP) service that uses machine learning to find insights and relationships in text

The service identifies the language of the text; extracts key phrases, places, people, brands, or events; understands how positive or negative the text is; analyzes text using tokenization and parts of speech; and automatically organizes a collection of text files by topic. You can also use AutoML capabilities in Amazon Comprehend to build a custom set of entities or text classification models that are tailored uniquely to your organization's needs.



Amazon Comprehend Medical

For extracting complex medical information from unstructured text, you can use Amazon Comprehend Medical

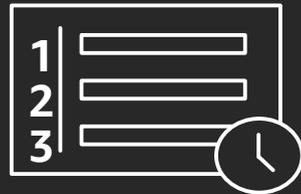
The service can identify medical information, such as medical conditions, medications, dosages, strengths, and frequencies from a variety of sources like doctor's notes, clinical trial reports, and patient health records. Amazon Comprehend Medical also identifies the relationship among the extracted medication and test, treatment, and procedure information for easier analysis. For example, the service identifies a particular dosage, strength, and frequency related to a specific medication from unstructured clinical notes.

Introduction to key services: Amazon SageMaker

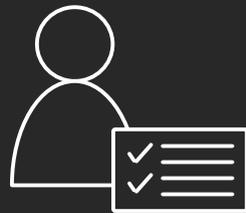
Amazon SageMaker: Build, train, and deploy machine learning (ML) models at scale



Collect and
prepare training
data



Choose and
optimize your
ML algorithm



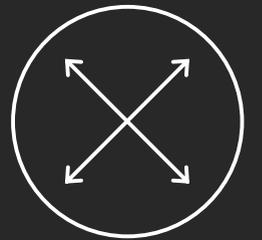
Set up and
manage
environments
for training



Train and
tune ML models



Deploy models
in production

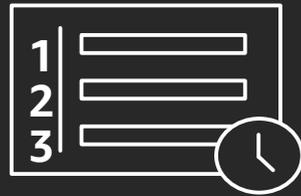


Scale and manage
the production
environment

Amazon SageMaker: Build, train, and deploy ML models at scale

Pre-built
notebooks
for common
problems

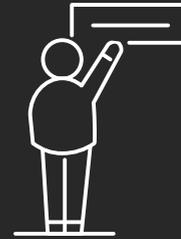
Collect and
prepare training
data



Choose and
optimize your
ML algorithm



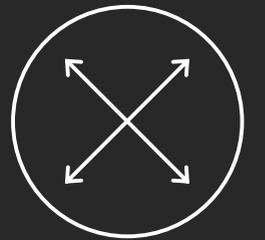
Set up and
manage
environments
for training



Train and
tune ML models



Deploy models
in production



Scale and manage
the production
environment

Amazon SageMaker: Build, train, and deploy ML models at scale

Pre-built notebooks for common problems

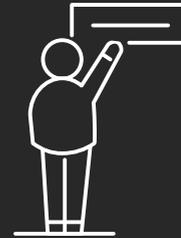
Collect and prepare training data

Built-in, high performance algorithms

Choose and optimize your ML algorithm



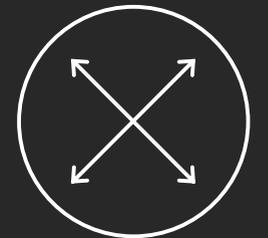
Set up and manage environments for training



Train and tune ML models



Deploy models in production



Scale and manage the production environment

Amazon SageMaker: Build, train, and deploy ML models at scale

Pre-built notebooks for common problems

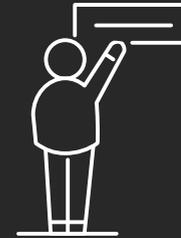
Collect and prepare training data

Built-in, high performance algorithms

Choose and optimize your ML algorithm

One-click training on the highest performing infrastructure

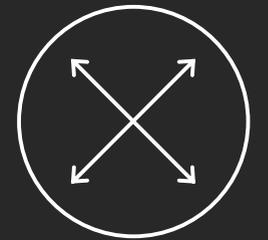
Set up and manage environments for training



Train and tune ML models



Deploy models in production



Scale and manage the production environment

Amazon SageMaker: Build, train, and deploy ML models at scale

Pre-built notebooks for common problems

Collect and prepare training data

Built-in, high performance algorithms

Choose and optimize your ML algorithm

One-click training on the highest performing infrastructure

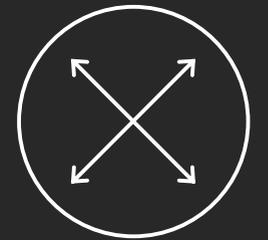
Set up and manage environments for training

Model optimization

Train and tune ML models



Deploy models in production



Scale and manage the production environment

Amazon SageMaker: Build, train, and deploy ML models at scale

Pre-built
notebooks
for common
problems

Collect and
prepare training
data

Built-in, high
performance
algorithms

Choose and
optimize your
ML algorithm

One-click
training on the
highest
performing
infrastructure

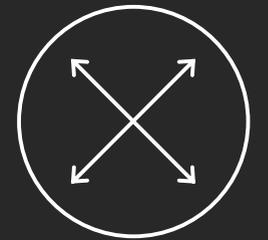
Set up and
manage
environments
for training

Model
optimization

Train and
tune ML models

One-click
deployment

Deploy models
in production



Scale and manage
the production
environment

Amazon SageMaker: Build, train, and deploy ML models at scale

Pre-built notebooks for common problems

Collect and prepare training data

Intuit Inc.

SIEMENS

Built-in, high performance algorithms

Choose and optimize your ML algorithm



Dow Jones & Company, Inc.

One-click training on the highest performing infrastructure

Set up and manage environments for training

National Football League



Model optimization

Train and tune ML models



One-click deployment

Deploy models in production



SONY

Fully managed with auto-scaling for 75% less

Scale and manage the production environment



CONVOY



Hands-on workshop

Prerequisites

- A laptop with internet access and a browser
- Basic understanding of Python
- Basic understanding of AWS services like AWS Identity and Access management (IAM), Amazon Simple Storage Service (Amazon S3), and AWS CloudFormation
- Understanding of machine learning

Setup guide

1. Be sure to log out of your other AWS account(s)
2. Navigate to <https://dashboard.eventengine.run/login>
3. Enter the 12-digit hash in front of you
4. Click on "AWS Console." Then click "Open AWS Console."
5. Confirm your region is Ohio (us-east-2). Can be found in top bar on right.
6. Navigate to Amazon SageMaker notebook instances. Open Jupyter Notebook
7. Start with "0.prep-lfs401.ipynb"

Thank you!

Ujjwal Ratan

ujjwalr@amazon.com

Aaron Friedman

ajfriedm@amazon.com



Please complete the session survey in the mobile app.