phuse·share
connect·share·advance

Single Day Event

NLP in Pharma: Applicability & Interpretability

Sherrine Eid & Sundaresh Sankaran,
SAS Institute

Accelerate ANALYTIC INNOVATIONS throughout the PHARMACEUTICAL LIFE CYCLE

Clinical Research

Real-World Evidence

Manufacturing and Supply Chain

Commercial

Natural Language Processing (NLP) plays a role.

Manual Review is Time-consuming & Resource-intensive.

- Clinical Trial Summaries
- Adverse Event Narratives
- Other Healthcare Text Data

Improve **Patient-centricity & effectiveness of clinical trials**.

Challenges

Clinical Reports consist of structured and unstructured data

Critical information mixed with supporting details

- Need to identify essential information of interest

Text not directly extractable

- OCR and information extraction techniques (a.k.a. Concepts and Facts)

Data not in chronological order

- Date, time and other positional markers to be leveraged

References to dates and times not explicit

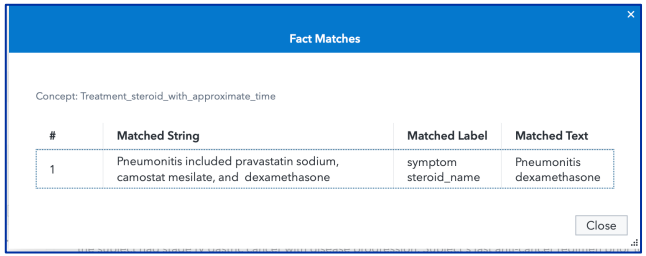
- Derivations based on business rules and logic

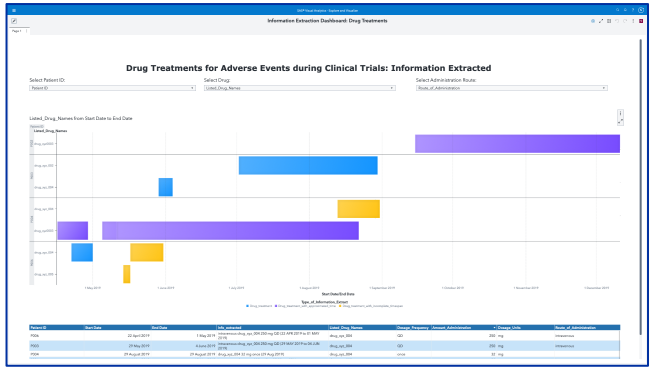
Link narrative data to other structured information

- Identify possible anomalies / errors in recording

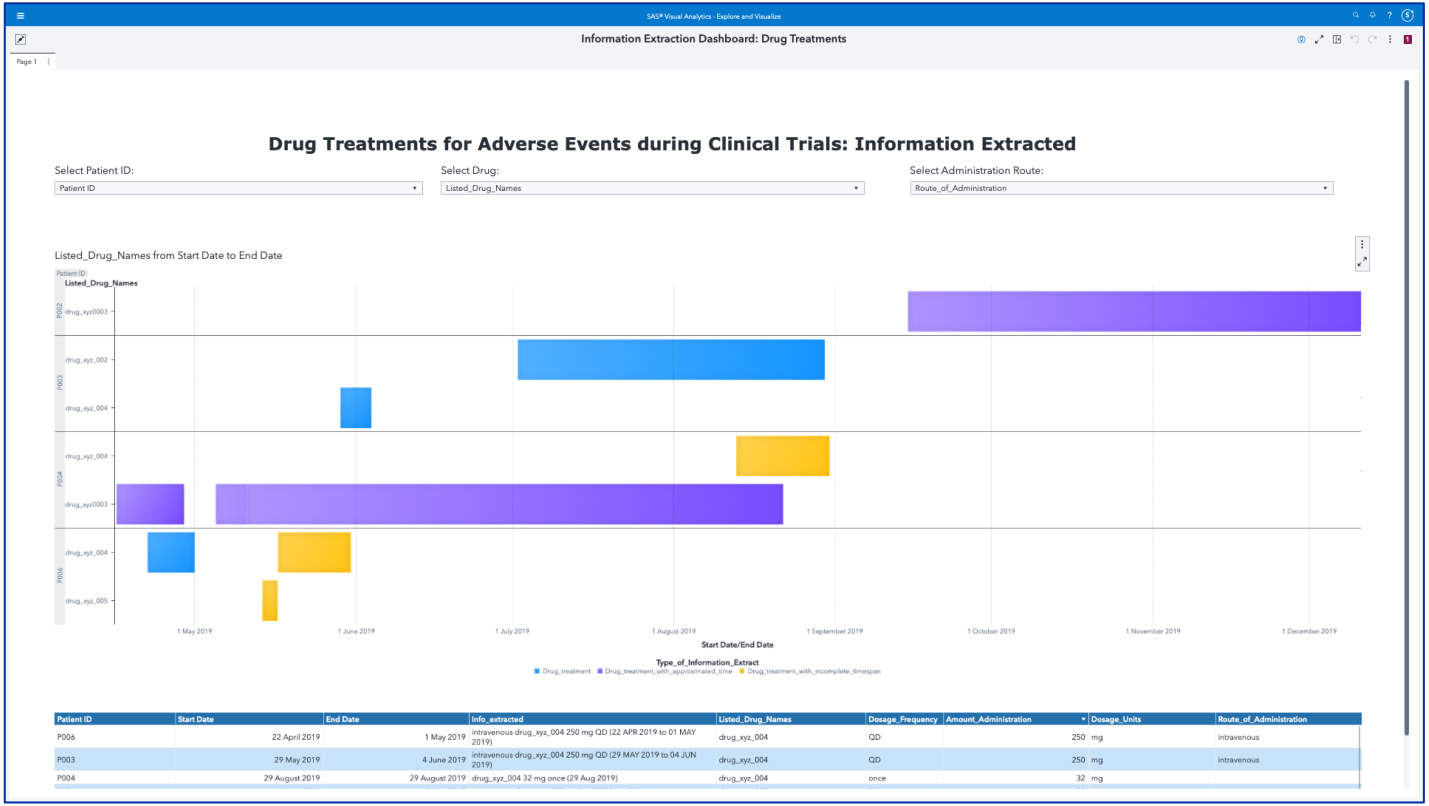
Solution Overview

Goal : Timely, accurate, and actionable insights from unstructured data





Temporal View for Patients & Procedures



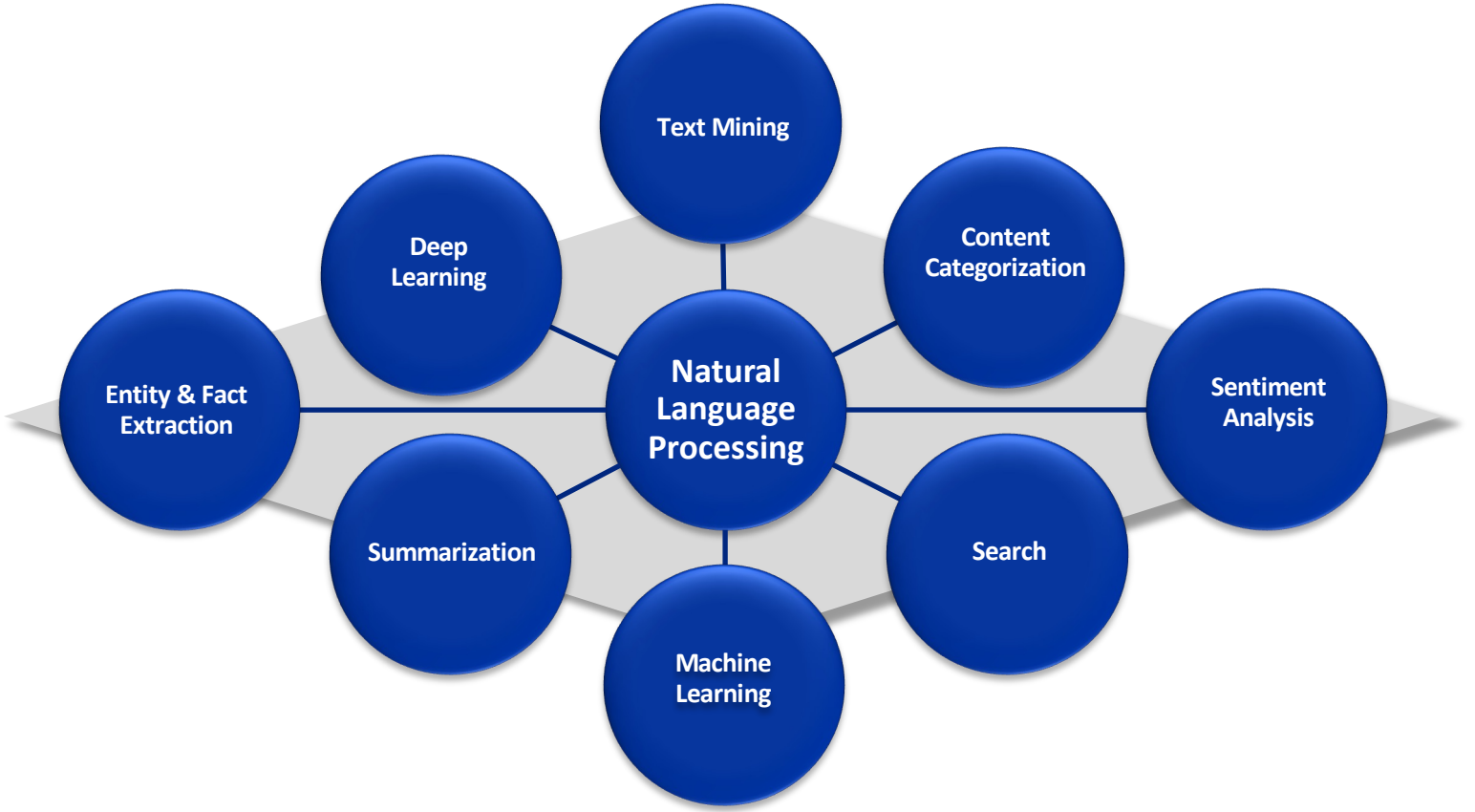
Patient ID	Start Date	End Date	Info_extracted	Listed_Drug_Names	Dosage_Frequency	Amount_Administration	Dosage_Units	Route_of_Administration
P006	22 April 2019	1 May 2019	intravenous drug_xyz_004 250 mg QD (22 APR 2019 to 01 MAY 2019)	drug_xyz_004	QD	250	mg	intravenous
P003	29 May 2019	4 June 2019	intravenous drug_xyz_004 250 mg QD (29 MAY 2019 to 04 JUN 2019)	drug_xyz_004	QD	250	mg	intravenous
P004	29 August 2019	29 August 2019	drug_xyz_004 32 mg once (29 Aug 2019)	drug_xyz_004	once	32	mg	

What is Natural Language Processing and Text Analytics?

What is Text Analytics?

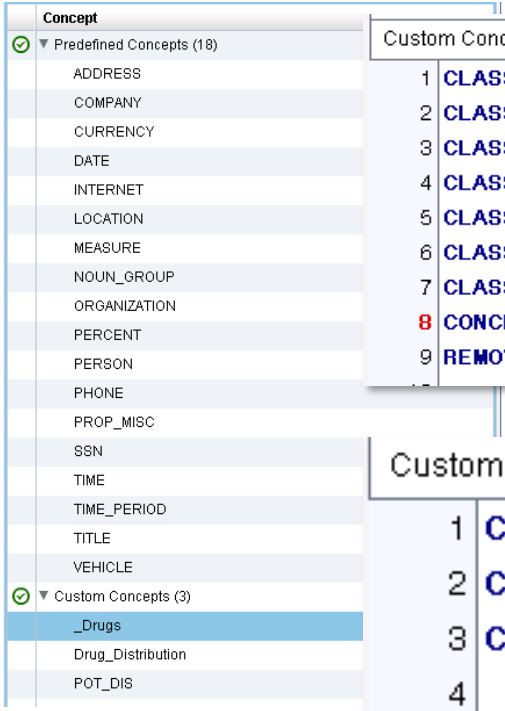
*"Using technology to **scale** the human acts of **reading, organizing, and quantifying** freeform text in meaningful ways."*

SAS Text Analytics Capabilities



What is Entity and Fact Extraction?

Capabilities to extract relevant information from text including people, places, and things.
Provides disambiguation. Goes beyond ctrl-F.



Custom Concept > _Drugs

```
1 CLASSIFIER: Marijuana
2 CLASSIFIER: Opioid
3 CLASSIFIER: pain pills
4 CLASSIFIER: OxyContin
5 CLASSIFIER: Norco
6 CLASSIFIER: Percocet
7 CLASSIFIER: Adderall
8 CONCEPT: _w _w pot
9 REMOVE_ITEM:(ALIGNED, "_c{_Drugs}", "POT_DIS")
```

Custom Concept > _Drugs

1 matched string

The suspect was caught with a quantity of pot that he was intending to sell.

The suspect was caught with a stolen pot that he was intending to sell.

Custom Concept > POT_DIS

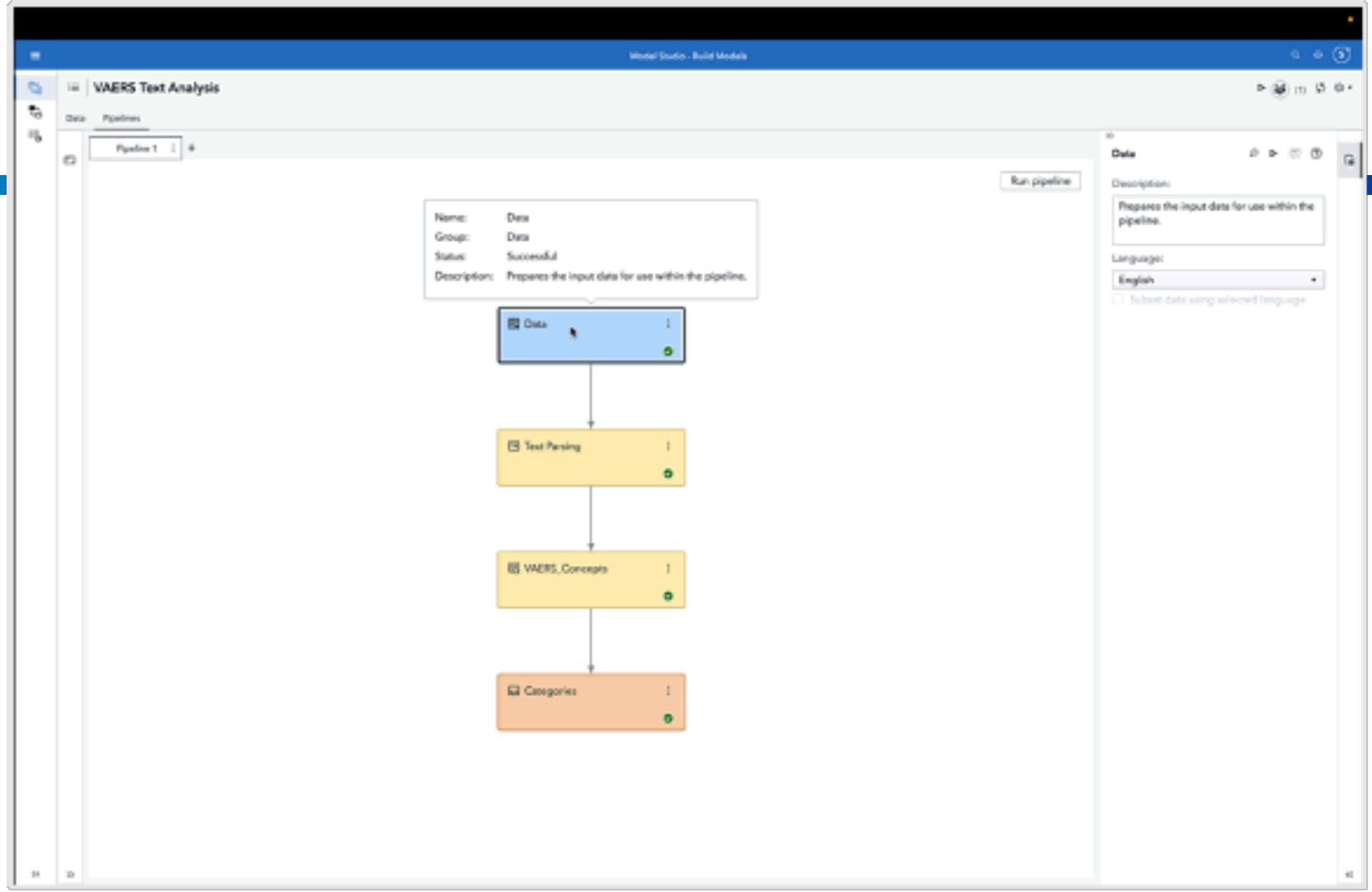
```
1 CLASSIFIER: a pot
2 CLASSIFIER: a flower pot
3 CONCEPT: a _w pot
4
```

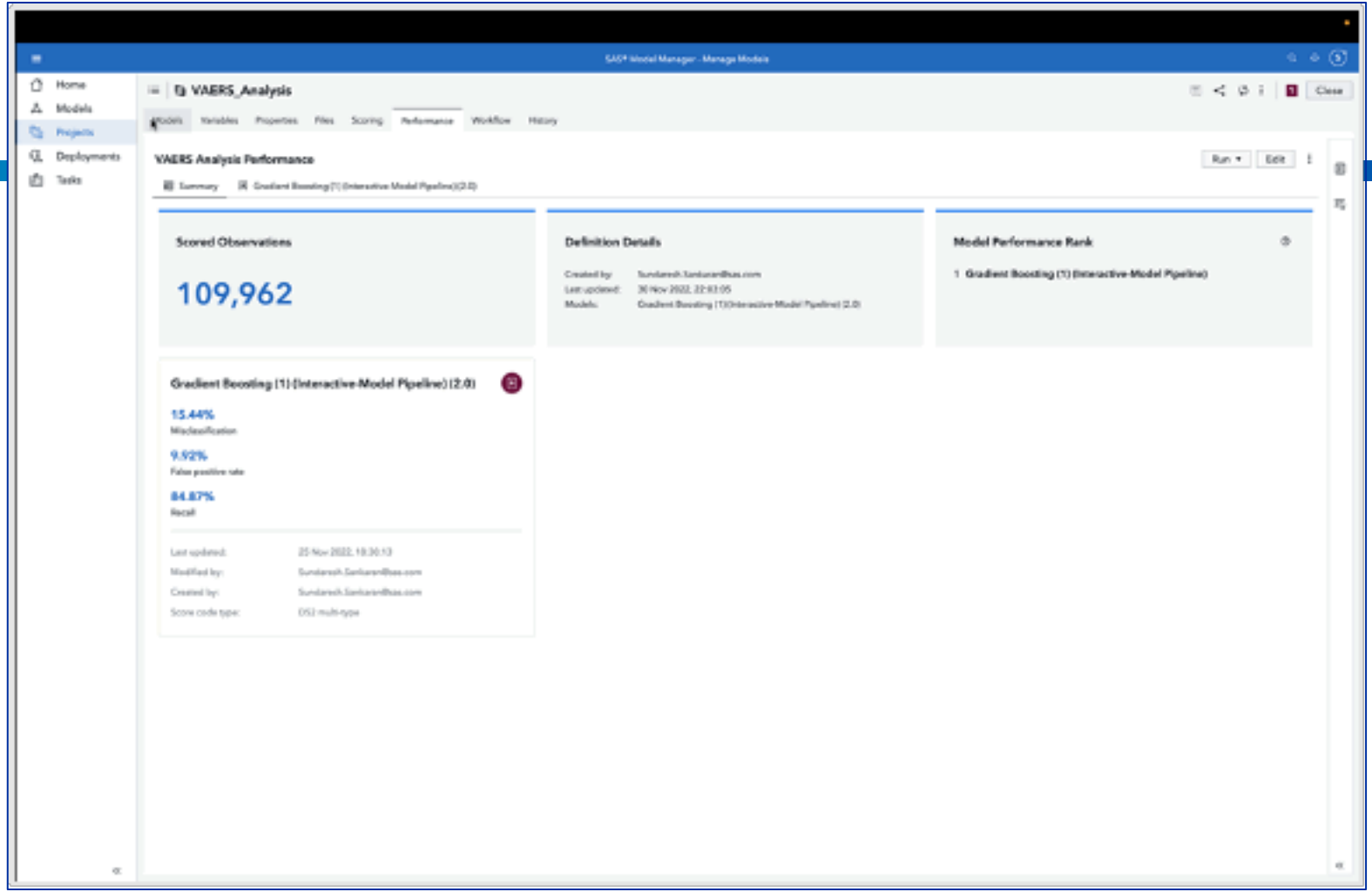
Custom Concept > POT_DIS

1 matched string

The suspect was caught with a quantity of pot that he was intending to sell.

The suspect was caught with a stolen pot that he was intending to sell.


Model Studio - Build Models
VAERS Text Analysis
Data
Pipelines
Pipeline 1
Run pipeline
Name: Data
Group: Data
Status: Successful
Description: Prepares the input data for use within the pipeline.
Data
Text Parsing
VAERS_Concepts
Categories
Data
Description:
Prepares the input data for use within the pipeline.
Language:
English
Subset data using selected language


SAS® Model Manager - Manage Models
Home
Models
Projects
Deployments
Tasks
VAERS_Analysis
Close
Models
Variables
Properties
Files
Scoring
Performance
Workflow
History
VAERS Analysis Performance
Run
Edit
Summary
Scored Observations
109,962
Definition Details
Model Performance Rank
1 Gradient Boosting (1) (Interactive Model Pipeline)
Gradient Boosting (1) (Interactive Model Pipeline) (2.0)
15.44%
Misclassification
9.92%
False positive rate
84.87%
Recall

The Global Healthcare Data Science Community

Contact Channels

UK +44 1843 609600
office@phuse.global
phuse.global

Social Media

@phusetwitta
/phusebook
/phusetube
/company/phuse