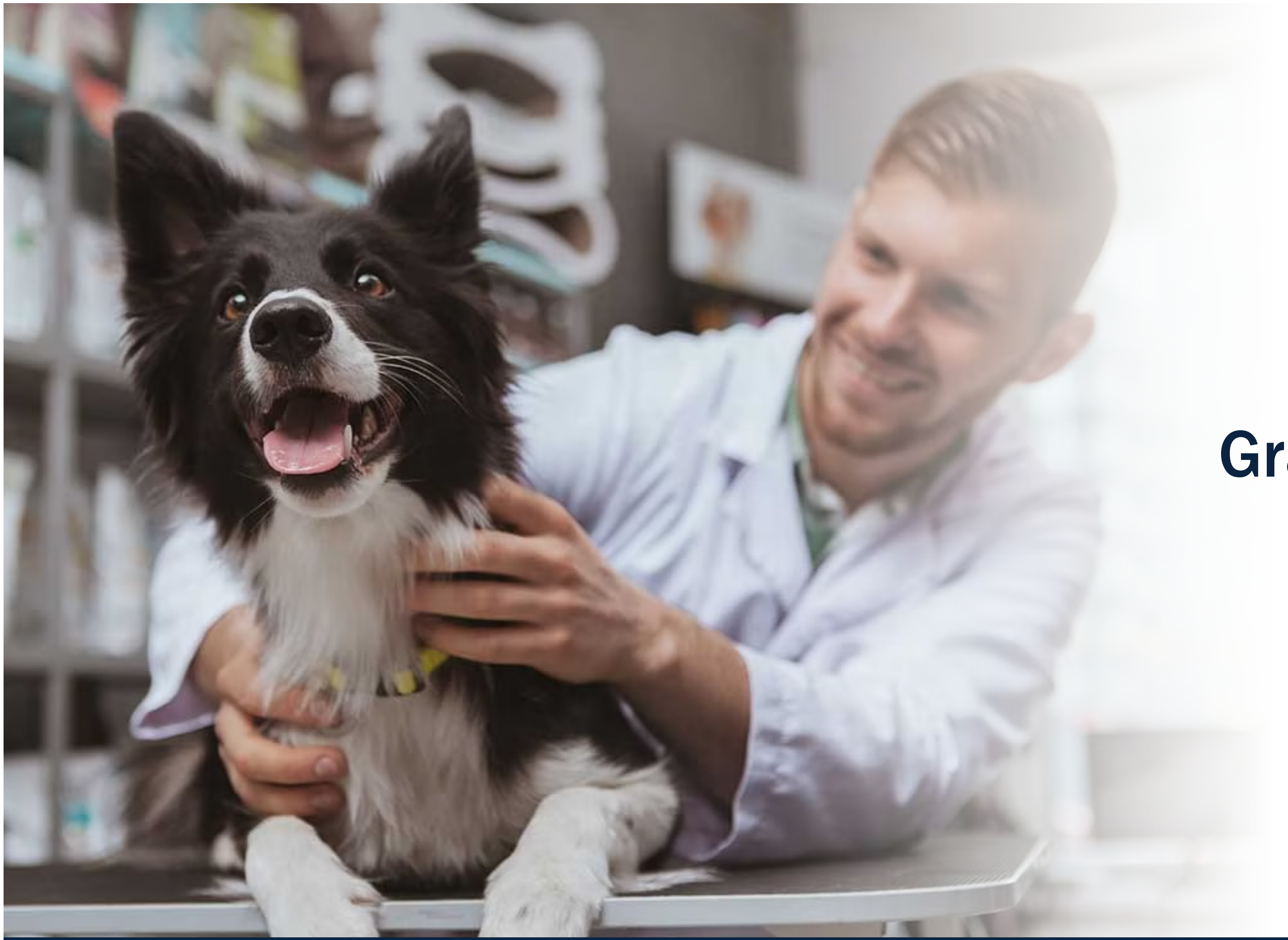


Data Science Graduate Capstone

University of Notre Dame
April 2023

IN PARTNERSHIP WITH:

Banfield Pet Hospital

Banfield PET HOSPITAL

UNIVERSITY OF NOTRE DAME

DATA SCIENCE

Meet the Team

DAN CHERNOFF

RICK MILLIKEN

SAMUEL PEPPER

NORA FLOTT

DISHA MAC

RYAN MACIEJ

Support Team

Banfield PET HOSPITAL	Christina Malone
Banfield PET HOSPITAL	Jai Thomas
UNIVERSITY OF NOTRE DAME	Chris Frederick

Objectives

Confirm Patient's Heartworm Preventative Treatment Status

- Determine if Banfield is accurately recording Heartworm Preventative in designated (structured) fields.
  - Provided by Banfield (is_provided_flg = Y)
  - Provided by Client (client_provided_flg = Y)
  - HWP not being Provided (Both Flags = N)
- Develop Natural Language Processing (NLP) Model to evaluate medical notes and determine if HWP is being administered.

Purpose ML Ops Framework

- Suggest sustainable Machine Learning Operations (ML Ops) framework for Banfield to utilize for further data science projects.

Agenda

- Data Structure & EDA
- Modeling
- ML Ops Infrastructure
- Ethical Considerations

Data Structure & EDA

Clashmore Mike

- Notre Dame Legend
- Prime Heartworm Preventative Candidate

Vet Visits for Clashmore Mike

What the Data Looks Like

Visit Date	Visit Notes (Predictor)	Visit ID	HWP Provided (Response)
Mar 2, 2018	Foxtail Removal	VST00001	None
Sep 30, 2018	Ate Nat'l Championship Pennant, induced vomiting, Pennant was recovered	VST00002	None
Nov 12, 2018	Annual Exam – Client is providing HWP from previous vet.	VST00003	Client
Mar 1, 2019	Provided heartworm test, client still providing HWP from previous vet.	VST00004	Client
May 29, 2019	Bordetella Vaccine Administered - Nasal	VST00005	None
Nov 18, 2019	Annual Exam – Client is administering Heartgard 51-100# from previous vet.	VST00006	None *Should be Client*
Nov 19, 2021	Annual Exam – Refilled 6 mo Heartgard Plus	VST00007	Banfield
May 19, 2022	Bordetella Vaccine Administered – Nasal, Client using other pets HWP, no refill needed	VST00008	Client

Pet Visit Table

Preventative Care Table

Data Obstacles

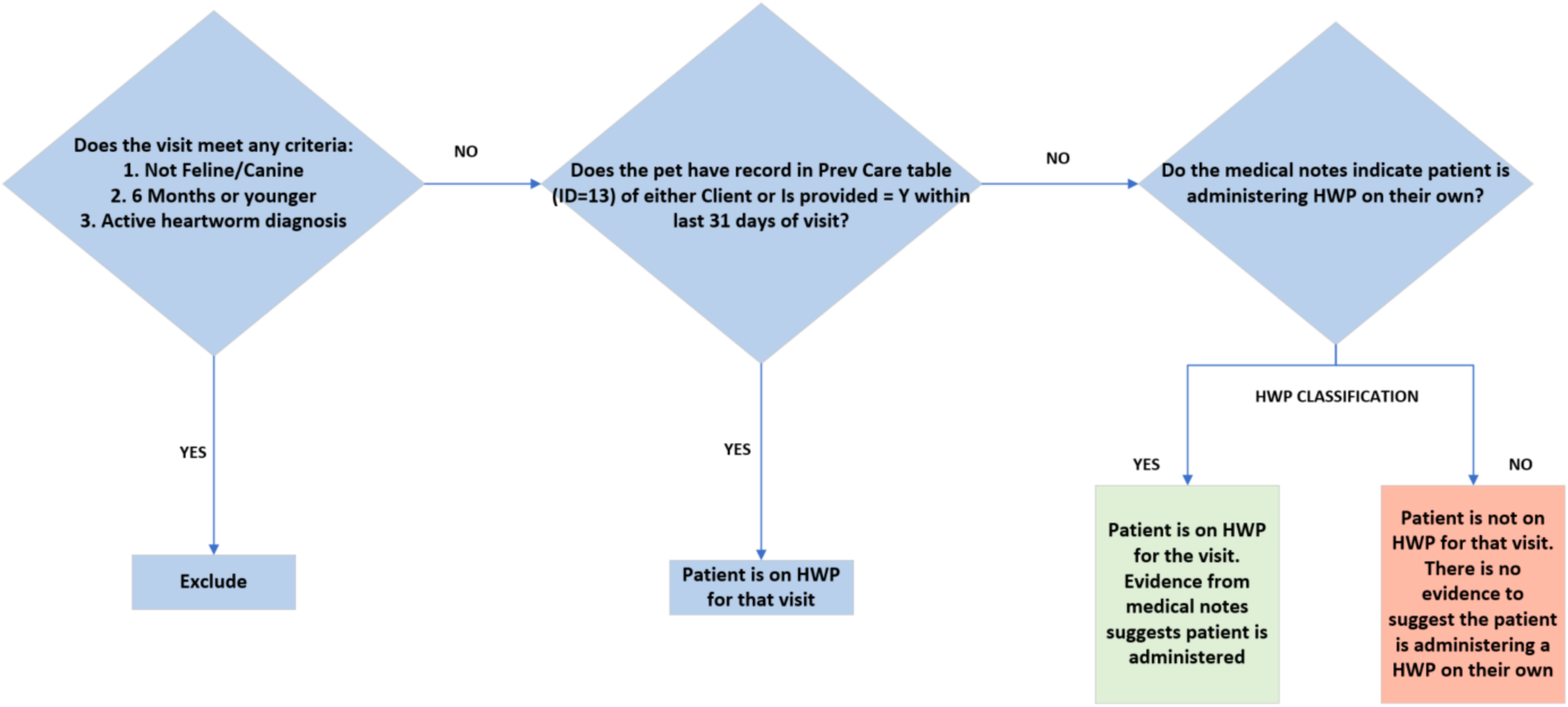
Preventative Care Table

Pet ID	Create Date	Visit ID	Banfield Provided	Client Provided
PT1842	Nov 13, 2018	-999999	N	Y
PT1842	Nov 19, 2021	VST00007	N	N
PT1842	Nov 19, 2021	VST00007	N	N
PT1842	Nov 19, 2021	VST00007	N	Y

Challenge: For all HWP Preventative Care records flagged as Client Provided, no Visit ID to join on
Solution: “Fuzzy Joined” on Pet ID & 1 day lag between visit date in visit table and create date in preventative care table.

Challenge: For some visits in preventative care table, multiple conflicting records exist for HWP.
Solution: If any contain yes, prioritize that record.

Process Diagram for Classification



Does the visit meet any criteria:
1. Not Feline/Canine
2. 6 Months or younger
3. Active heartworm diagnosis

YES

Exclude

NO

Does the pet have record in Prev Care table (ID=13) of either Client or Is provided = Y within last 31 days of visit?

YES

Patient is on HWP for that visit

NO

Do the medical notes indicate patient is administering HWP on their own?

HWP CLASSIFICATION

YES

Patient is on HWP for the visit. Evidence from medical notes suggests patient is administered

NO

Patient is not on HWP for that visit. There is no evidence to suggest the patient is administering a HWP on their own

Examples of Actual Medical Notes

(PDF of Full Medical Note)

Modeling

Note on Data Used

Modeling Approaches

Keyword Filtering

- Simple to Build
- Searches for HWP Keywords
- Does not take context into consideration
- Keyword List may be incomplete

NLP Model

- Complex Build
- Built on Pretrained Models specific to Medical field
- Hard to interpret what classification criteria is
- Can Typically Produce Much Better Results

Measures Used in Model Selection & Improvement

Performance Metric

Accuracy	How often a model makes correct predictions.	$\frac{TP + TN}{TP + FP + TN + FN}$
Precision Positive Predictive Value (PPV)	How often a model correctly predicts positive outcomes. High Precision will decrease False Positives	$\frac{TP}{TP + FP}$
Recall Sensitivity True Positive Rate (TPR)	How often a model correctly identifies positive outcomes.	$\frac{TP}{TP + FN}$

First Keyword Matching Model
(Baseline Model)

Model Criteria:

If a note contains a HWP Keyword
Classify as Positive
Else Negative

Confusion Matrix:

ACTUAL	PREDICTED: PROVIDED	PREDICTED: NOT PROVIDED
PROVIDED	283 True Positive	560 False Negative
NOT PROVIDED	7 False Positive	804 True Negative

Model Metrics:

Accuracy	65.7%
Precision	**97.6%**
Recall	33.4%

What is BERT NLP?

***B**idirectional **E**ncoder **R**epresentations from **T**ransformers*

"The chicken is ready to eat."

"I marinated the chicken overnight and cooked it for several hours. Now the chicken is ready to eat."

"The chicken is ready to eat after marinating overnight and cooking it for several hours."

Unidirectional NLP Models may struggle with one of the statements above. BERT, and its bidirectional superpowers would be able to better understand the context either way it was written.

BERT NLP Model

Model Criteria:

BERT sentence encodings with transfer learning from BioBERT fed to binary classification neural network. Trained on client medical notes where HWP provided by client and notes where client not providing any HWP.

Confusion Matrix:

		PROVIDED	NOT PROVIDED
ACTUAL	PROVIDED	264 True Positive	3 False Negative
	NOT PROVIDED	0 False Positive	237 True Negative

PREDICTED

Note: *Smaller volume of cases in confusion matrix due to Train/ Test splitting not required in Keyword Models*

Model Metrics:

Accuracy	99.4%
Precision	**100%**
Recall	98.9%

Why is BERT NLP Performing So Well?

System Generated Notes Highly Correlated with Negative Cases

(Back to the Full Medical Note)

Second Keyword Matching Model
(Prioritizing Negative Cases)

Model Criteria:

If a note contains “not given”
Classify as Negative
Else Positive

Confusion Matrix:

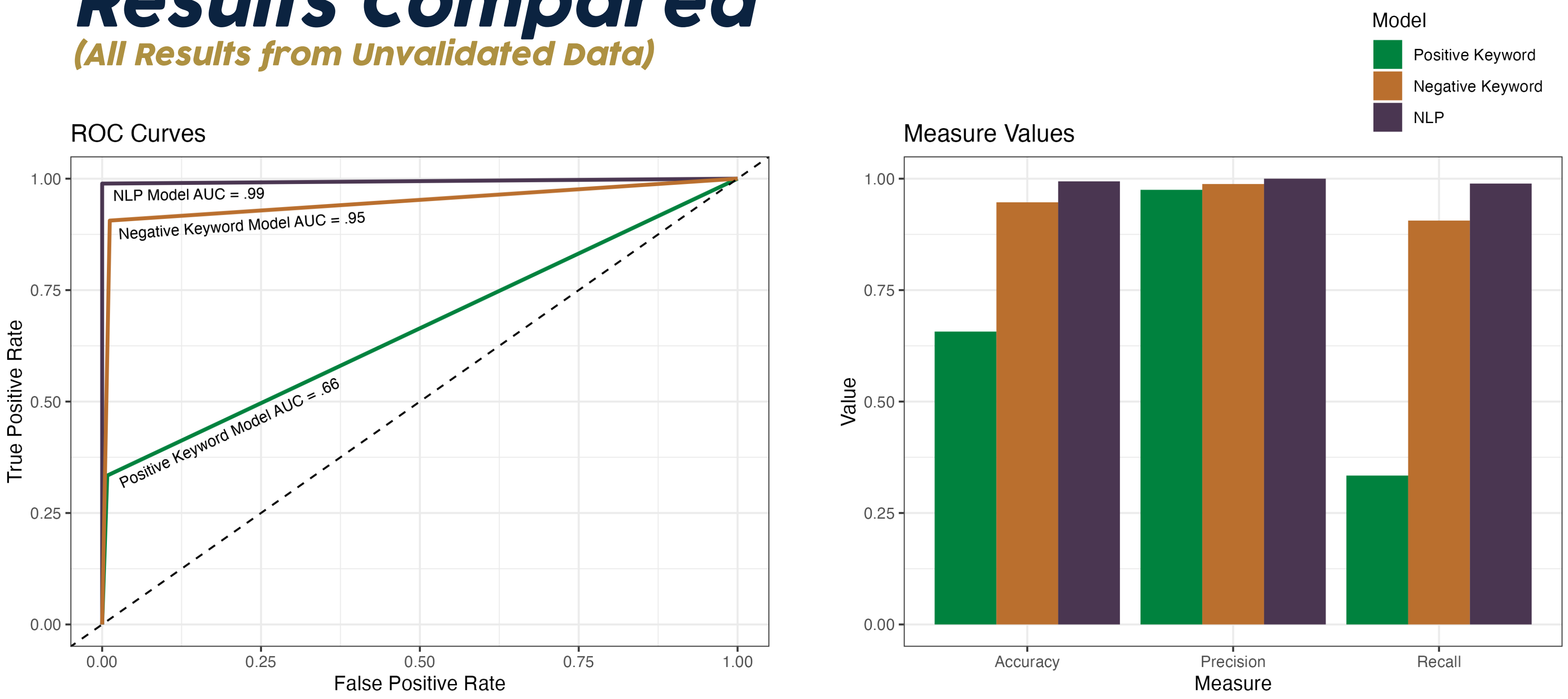
		PREDICTED	
		PROVIDED	NOT PROVIDED
ACTUAL	PROVIDED	764 True Positive	79 False Negative
	NOT PROVIDED	9 False Positive	802 True Negative

Model Metrics:

Accuracy	94.7%
Precision	**98.8%**
Recall	90.6%

Results Compared

(All Results from Unvalidated Data)



ROC Curves

Measure Values

Next Steps

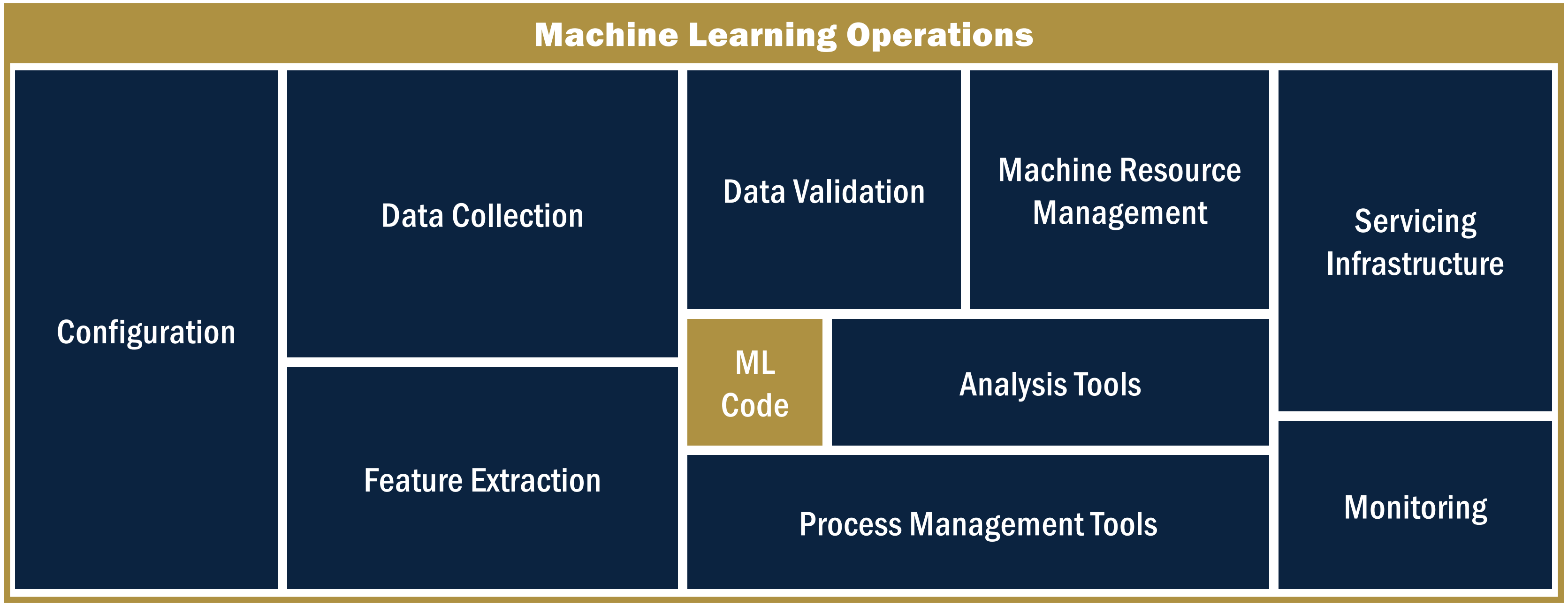
1. Focus model text on either:
   - Model Focused on specific template if it can be used consistently throughout organization.
   - RegEx Model extracted specific part of notes (i.e. “Subjective” part of S.O.A.P Notes) to reduce noise and key in on signal in the data.

2. Larger Validated Dataset for Training
   - Including Both Positive and Negative Cases

3. Better understanding of relationship between Preventative Care table and Pet Visit Table for records where labeled Client Provided.

ML Ops

Machine Learning Code is a Small Component of MLOps



Machine Learning Operations

Configuration

Data Collection

Feature Extraction

Data Validation

Machine Resource Management

ML Code

Analysis Tools

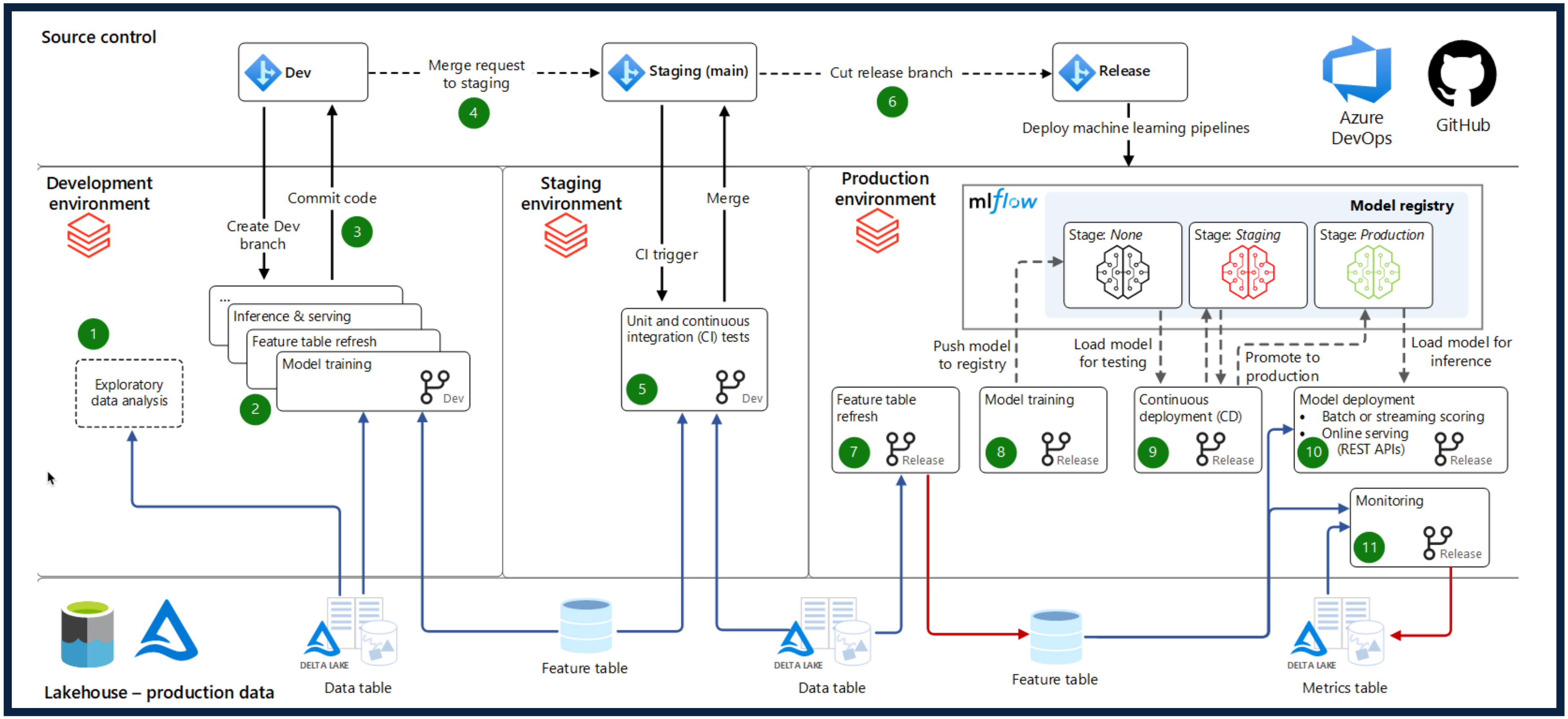
Process Management Tools

Servicing Infrastructure

Monitoring

Azure Based Technologies Map

Functionality	Technology
Execution Environment	Azure Databricks
Feature Store	Databricks Feature Store
Model Registry	ML Flow (Databricks)
Source Control	Azure DevOps •Azure Repos •Azure Pipelines
Storage	Azure Data Lake Storage Gen 2 •Delta Lake format to support Time Travel
Secrets Management	Azure Key Vault



Source control

Dev

Merge request to staging

4

Staging (main)

Cut release branch

6

Release

Deploy machine learning pipelines

Azure DevOps

GitHub

Development environment

Create Dev branch

Commit code

3

...

Inference & serving

Feature table refresh

Model training

Dev

2

1

Exploratory data analysis

Staging environment

CI trigger

Merge

Unit and continuous integration (CI) tests

5

Dev

Production environment

mlflow

Model registry

Stage: *None*

Stage: *Staging*

Stage: *Production*

Push model to registry

Load model for testing

Promote to production

Load model for inference

Feature table refresh

7

Release

Model training

8

Release

Continuous deployment (CD)

9

Release

Model deployment

- Batch or streaming scoring
- Online serving (REST APIs)

10

Release

Monitoring

11

Release

Lakehouse – production data

DELTA LAKE

Data table

Feature table

DELTA LAKE

Data table

Feature table

DELTA LAKE

Metrics table

Model

Model transition

Reads

Writes

Microsoft Azure

Code is Converted to Models and Promoted Through Environments

Ethical Considerations

Ethical Considerations

User Rights & Privacy

Data retention plan (closed hospitals, deceased pets, client departures)

Rights to be forgotten (client removal request)

Model does not reveal clients

Results do not impact privacy

Security

Data (Partitions)

Modeling

Infrastructure

Transparency

Risks assessed and communicated

Clear language utilized

Help guides provided

Quality

Unvalidated data

Model assumptions evaluated

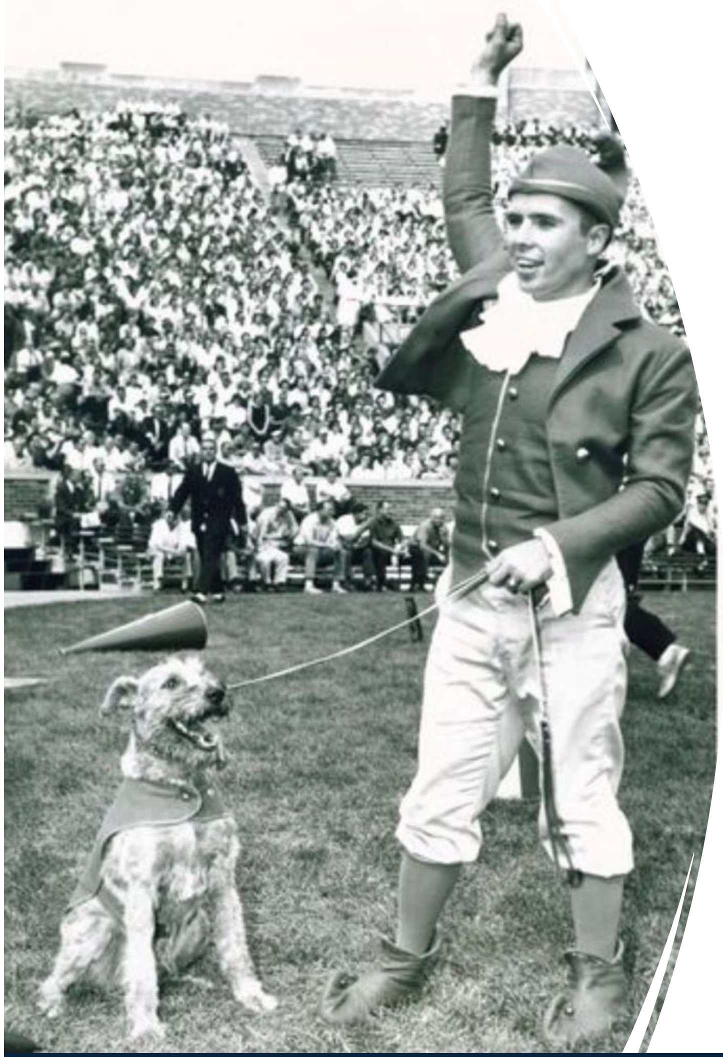
Subject Matter Experts' interpretation lacking

Bias Mitigation & Fairness

More dogs than cats, No Bias detected

Modeling downstream impacts have not been evaluated

Suggested Resource: *https://www.aiethicist.org/frameworks-guidelines-toolkits*



Thank You