# Machine Learning for Predicting Supplemental Aflibercept

Mihir Raja

August 28th, 2025

ADVERUM

# Agenda

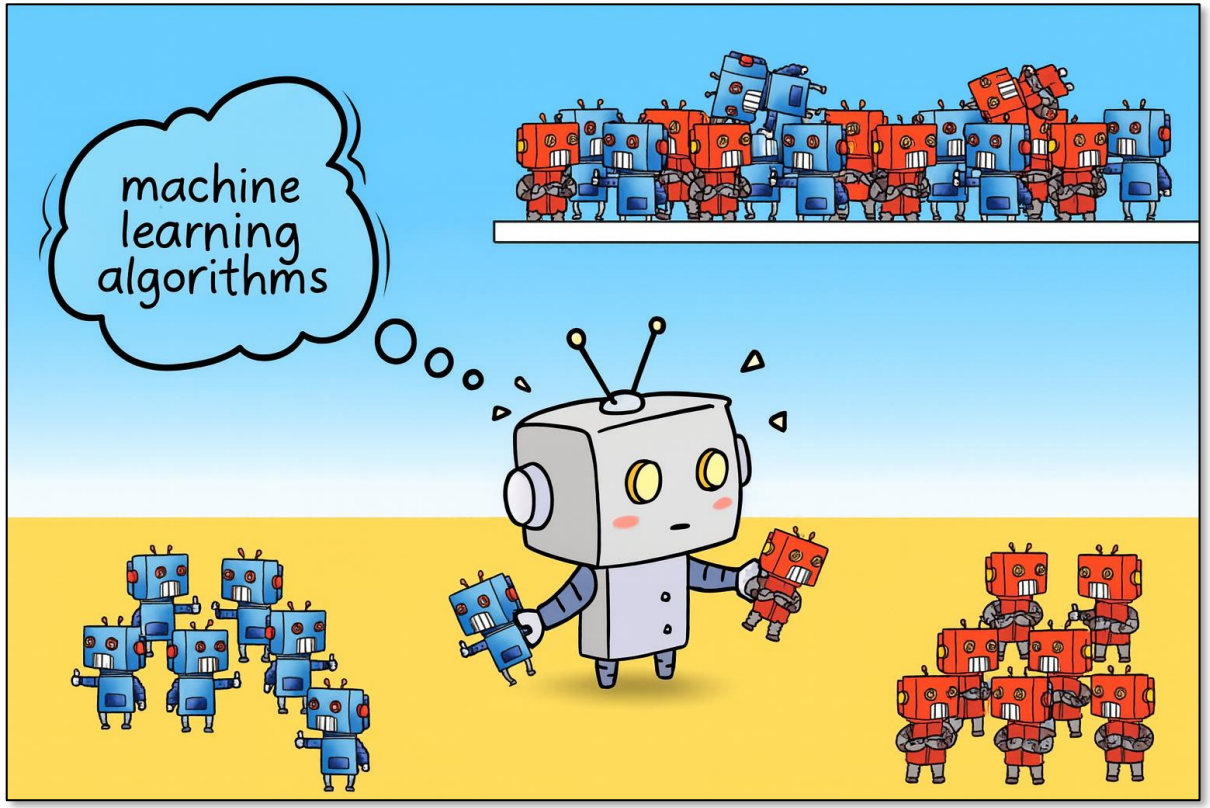
1. What is Machine Learning?
2. Purpose
3. Project Steps
4. Results
5. Limitations/Future Steps
6. Thank You



Confidential

## Basic Definition

- Programming computers to learn and improve from experience
- Typically involves finding patterns in data to make predictions or decisions
- Statistics is about inference and understanding relationships, while Machine Learning is about prediction + classification and optimizing performance





https://mljar.com/glossary/binary-classification/
Confidential

Anteater

Duck

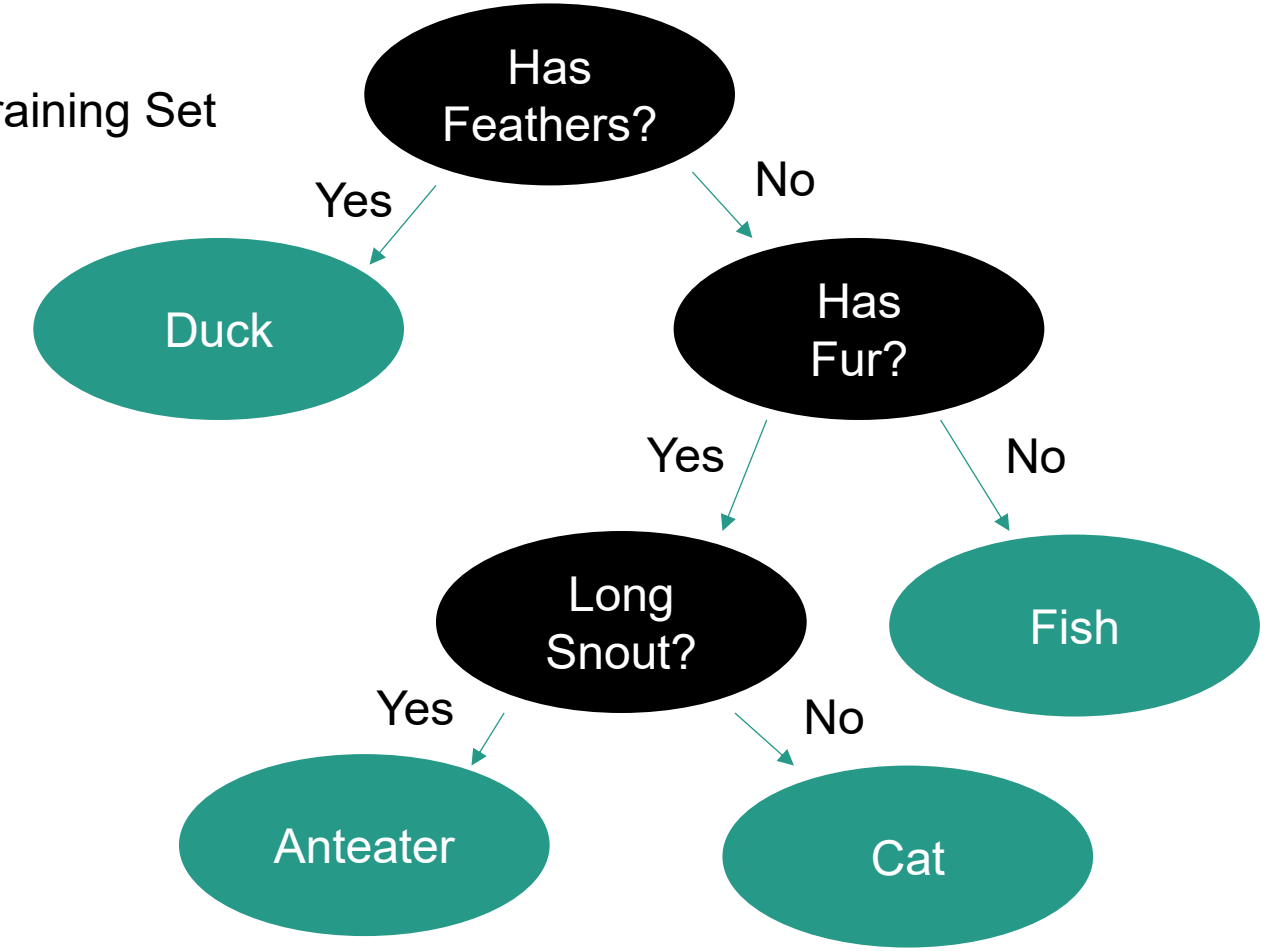
Fish

Cat

Training Set


?

Testing Set






Confidential

ADVERUM

Anteater

Duck

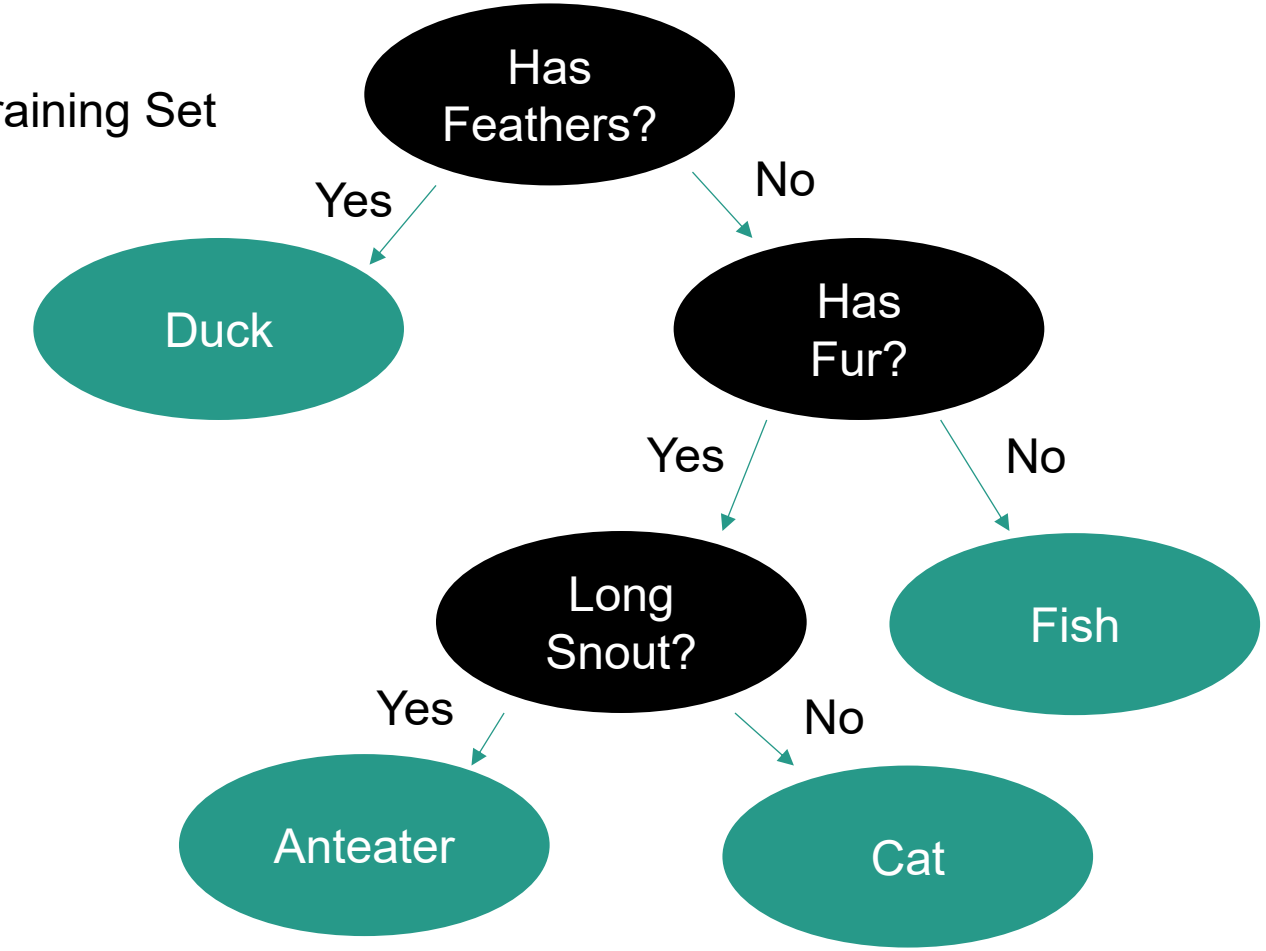
Fish

Cat

Training Set

?

Testing Set

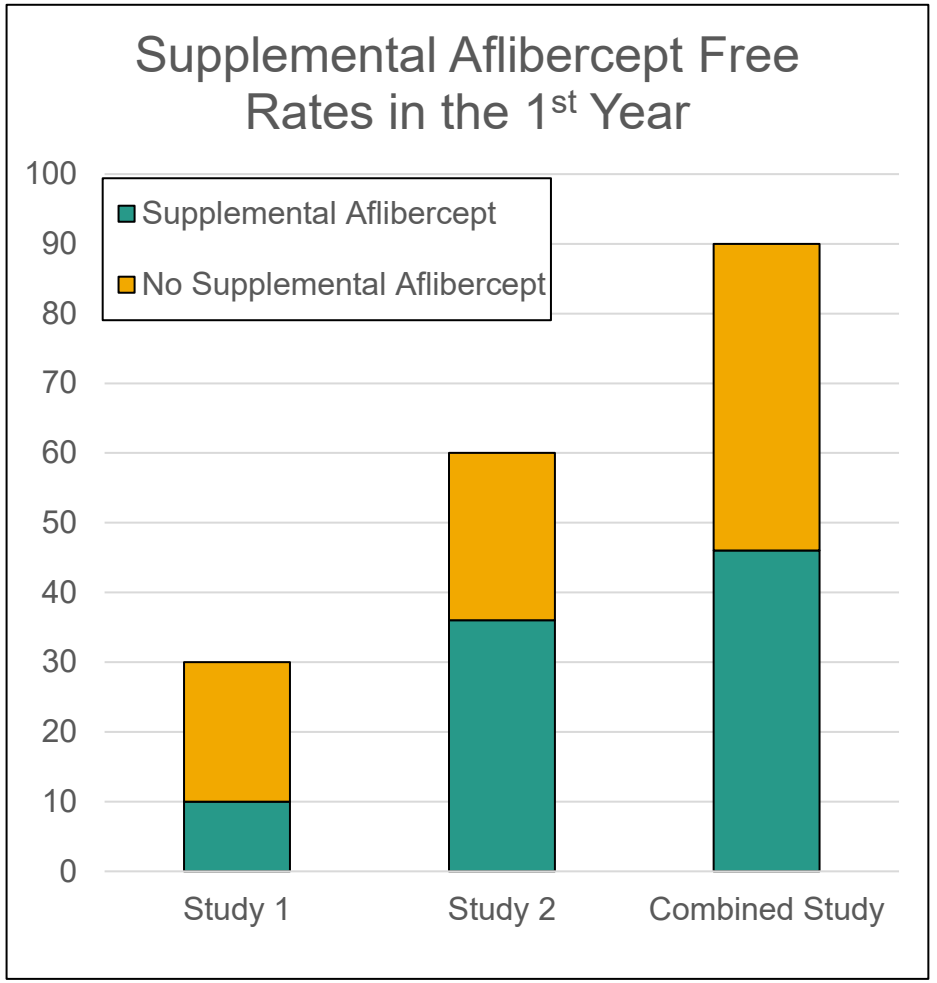





Confidential

ADVERUM

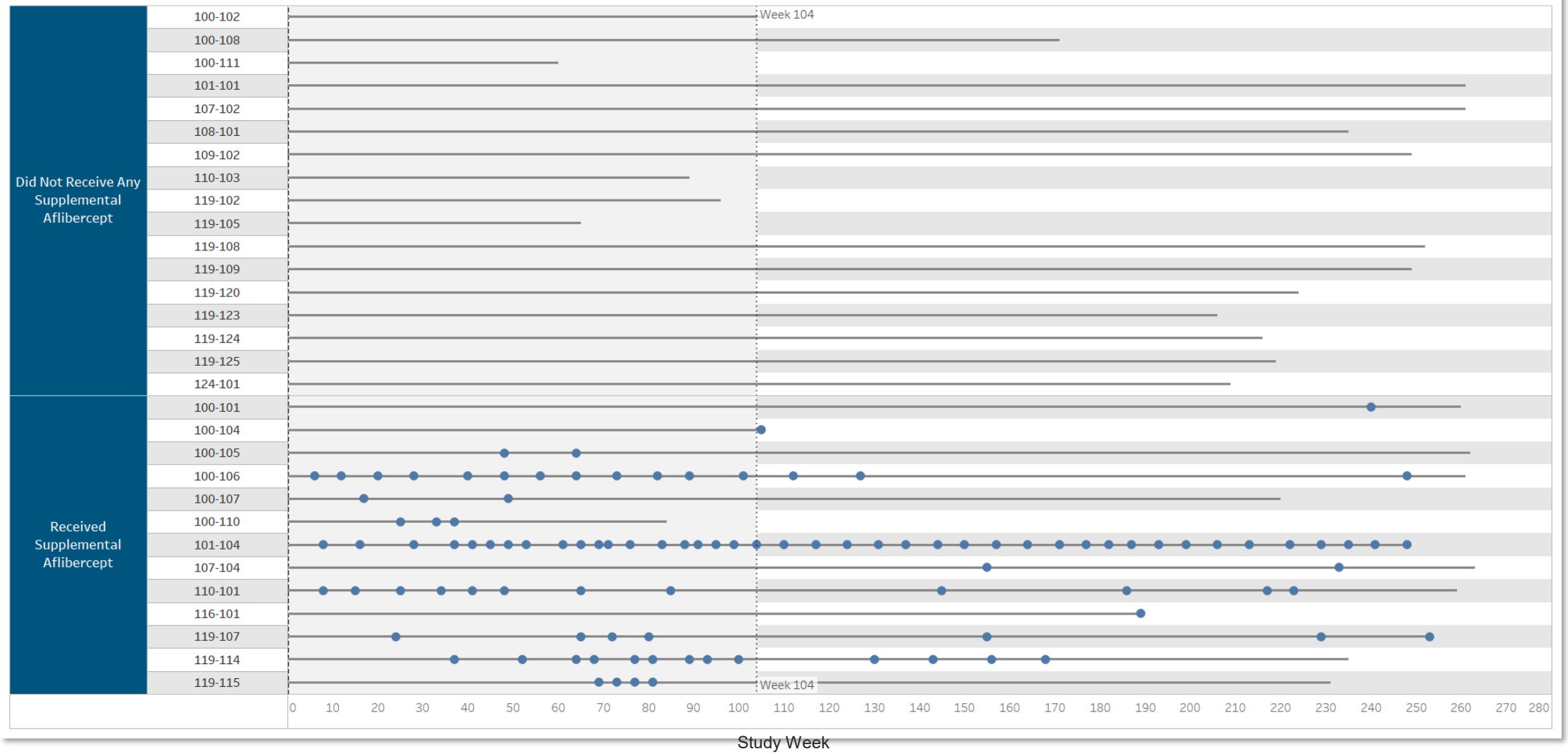
- Build a proof-of-concept ML pipeline that could predict supplemental aflibercept use based on Study 1 and Study 2 synthetic clinical data that mimics OPTIC and LUNA
  - Supplemental Aflibercept → addition Aflibercept treatments given to participants due to vision loss, increase in Central Subfield Thickness (CST), or new/worsening hemorrhage
- **End Goal:** Identify participants to monitor based on patterns from previous studies in Study 3 (which mimics ARTEMIS)
- Working with a 50/50 split with the data makes this end target more promising for ML applications





Confidential

## Adverum OPTIC and OPTIC EXT Supplemental Injection Swim Lane Plots






Confidential

Identify similar datapoints present in all three studies

Converting into usable dataset by joining and cleaning data sources

| Create usable tables with consistent data | |
|---|---|
| Study 1 + Study 2 | Study 3 w/ matching columns |



Confidential

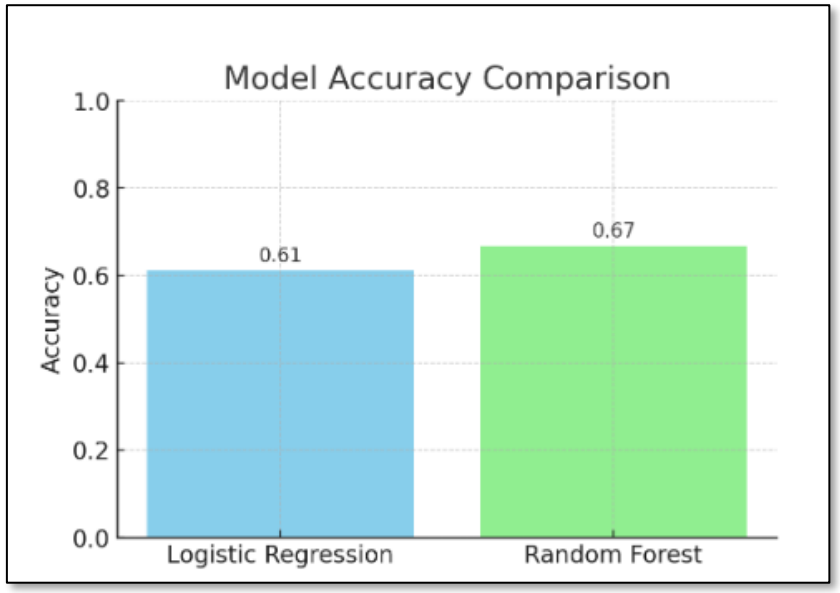
- The synthetic consolidated data contained a total of 90 subjects with both demographic and lab data
- Split the Study 1 + Study 2 combined data into separate training and testing sets
  - Training set (80% of the data)
  - Testing set (20% of the data)
- Run various types of algorithms using the training set to evaluate against the testing set
- **Reason**: Get an informed idea on which model would be best at predicting Supplemental Aflibercept use on Study 3 participants by 52 weeks

USER$MRAJA.PUBLIC  Settings

```
1   -- Rescue Aflibercept Retreatment
2   LEFT JOIN (
3     SELECT
4       PATIENT_ID,
5       MAX(EXYN) AS EXYN
6     FROM STUDY1.SUPPLEMENTAL_AFLIB
7     GROUP BY PATIENT_ID
8   ) RETRT
9     ON RETRT.PATIENT_ID = ADM.PATIENT_ID
10
11  -- Lab Results (screening only)
12  LEFT JOIN LAB_RESULTS LAB
13    ON LAB.SUBJECT_NUMBER = ID.SUBJECT_ID
14    AND LAB.VISIT_NAME = 'SCREEN'
15
16
17  GROUP BY
18    -- Patient Identifiers
19    PATIENT_DISPLAY_ID_FULL,
20    PATIENT_ID,
21    ID.SUBJECT_ID,
```


Confidential

- Model Performance from Simulated Tests:
  - Logistic Regression → 61% Accuracy
  - Random Forest Classifier → 67% Accuracy
- Developed a model that predicts if participants will need Supplemental Aflibercept treatments




jupyter ML_RA_PREDICTION Last Checkpoint: 6 seconds ago

File Edit View Run Kernel Settings Help

Code

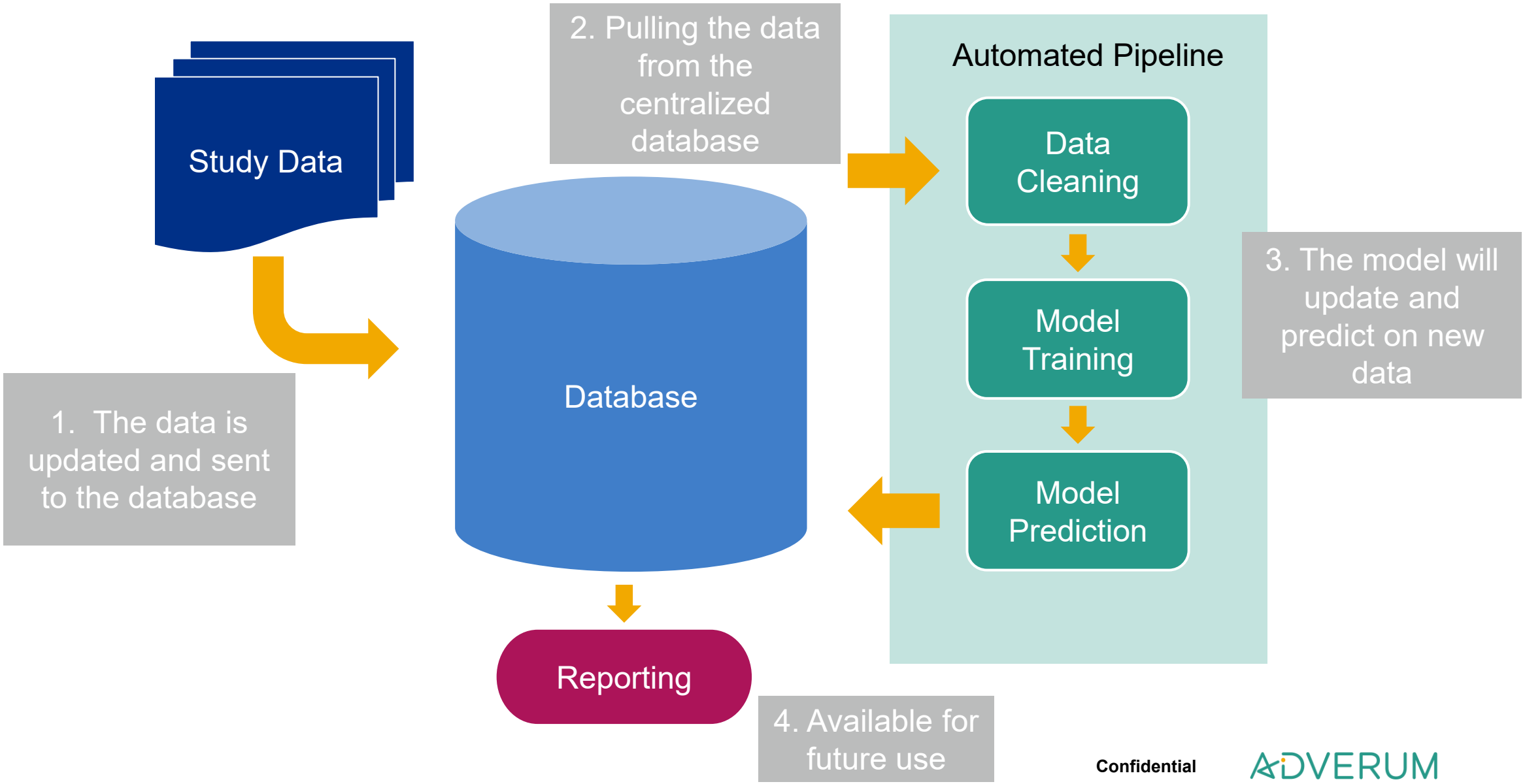
•[2]:

```
# Import libraries
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report, con
import joblib
# Load dataset
df = pd.read_csv(r".\Study1+Study2_combined_dataset.csv")

# Encode target variable
df['RESCUE_AFLIBERCEPT'] = df['RESCUE_AFLIBERCEPT'].astype(str).str.up
df = df[df['RESCUE_AFLIBERCEPT'].notna()]

# Store SUBJECT info
subject_ids = df['SUBJECT']

# Separate features and target
target = 'RESCUE_AFLIBERCEPT'
features = df.columns.drop([target])
```



Confidential

ADVERUM



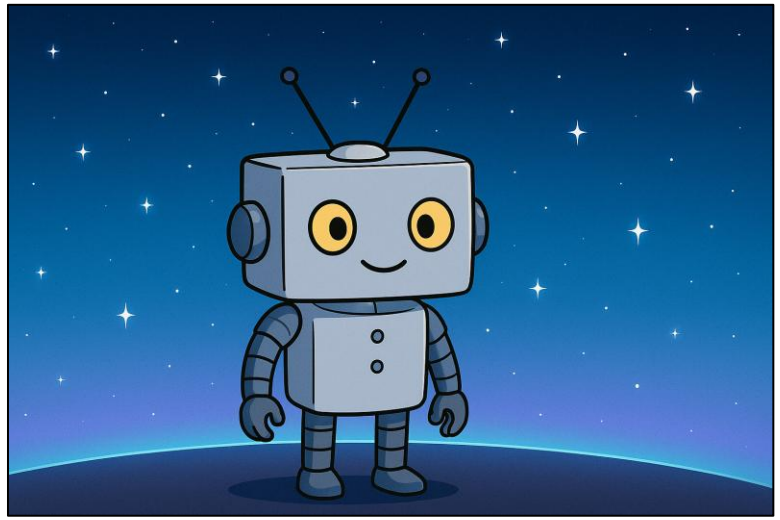



Confidential

ADVERUM

**Limitations:**

- ML Models are limited by the volume of data in the training set
  - Our synthetic data mimics the volume of available datasets

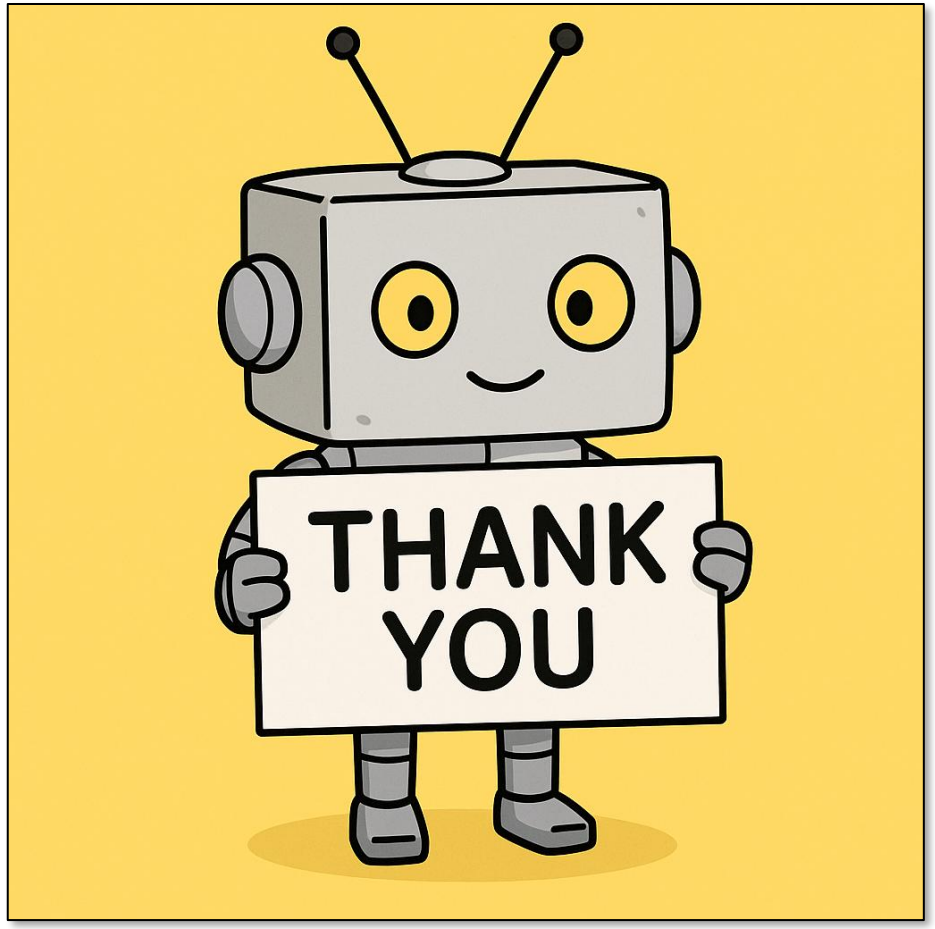


**Future Steps:**

- Deploy this model using real-world data
  - Repeat process with more training sets to increase accuracy
- Monitor participants with a high predictability of needing supplemental aflibercept
- Use ML Models to predict on different endpoints
- Analyze correlations which could inform screening/prescreening efforts in future studies


Confidential

- Special thanks to the Medical Affairs team (especially Phoebe and Jeff) for guiding me through the process as well as the other interns + the People team
- If you have any questions, or want to connect + reach out here is my contact:
  - mihir.raja@icloud.com
  - https://www.linkedin.com/in/mihiraja






Confidential