

**FIU**

Applied Research  
Center



# Advanced Automated Machine Learning (AAML) System

Presented by Dr. Himanshu Upadhyay  
Florida International University

Worlds  
**Ahead**

*Advancing the research and academic mission of Florida International University*



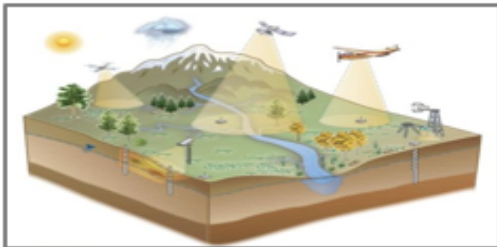
**Test and Evaluation / Science and Technology  
Program  
Cyberspace Test Technology**

***Advanced Automated Machine Learning (AAML)  
System***

***Developed by: Florida International University  
Research Sponsored by: Department of Defense (DOD) -  
Test Resource Management Center (TRMC)***

## ARTIFICIAL INTELLIGENCE / MACHINE LEARNING

### APPLICATION DOMAIN



**SENSOR / IMAGERY DATA FROM DOE-EM SITES**

### AI / ML TECHNOLOGY



**AI / ML FRAMEWORKS AND ALGORITHMS**

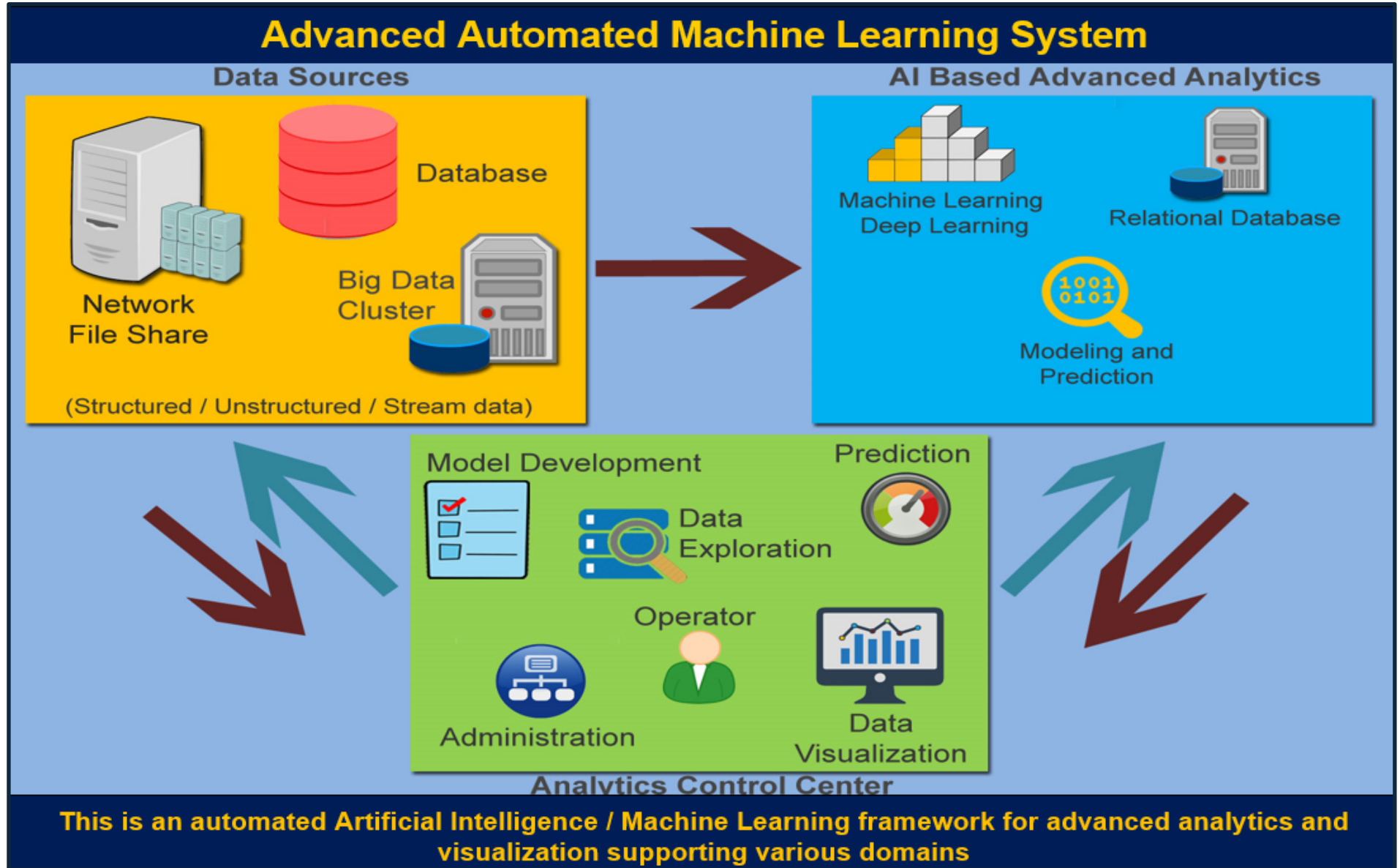
### OPERATIONALISE AI / ML MODEL

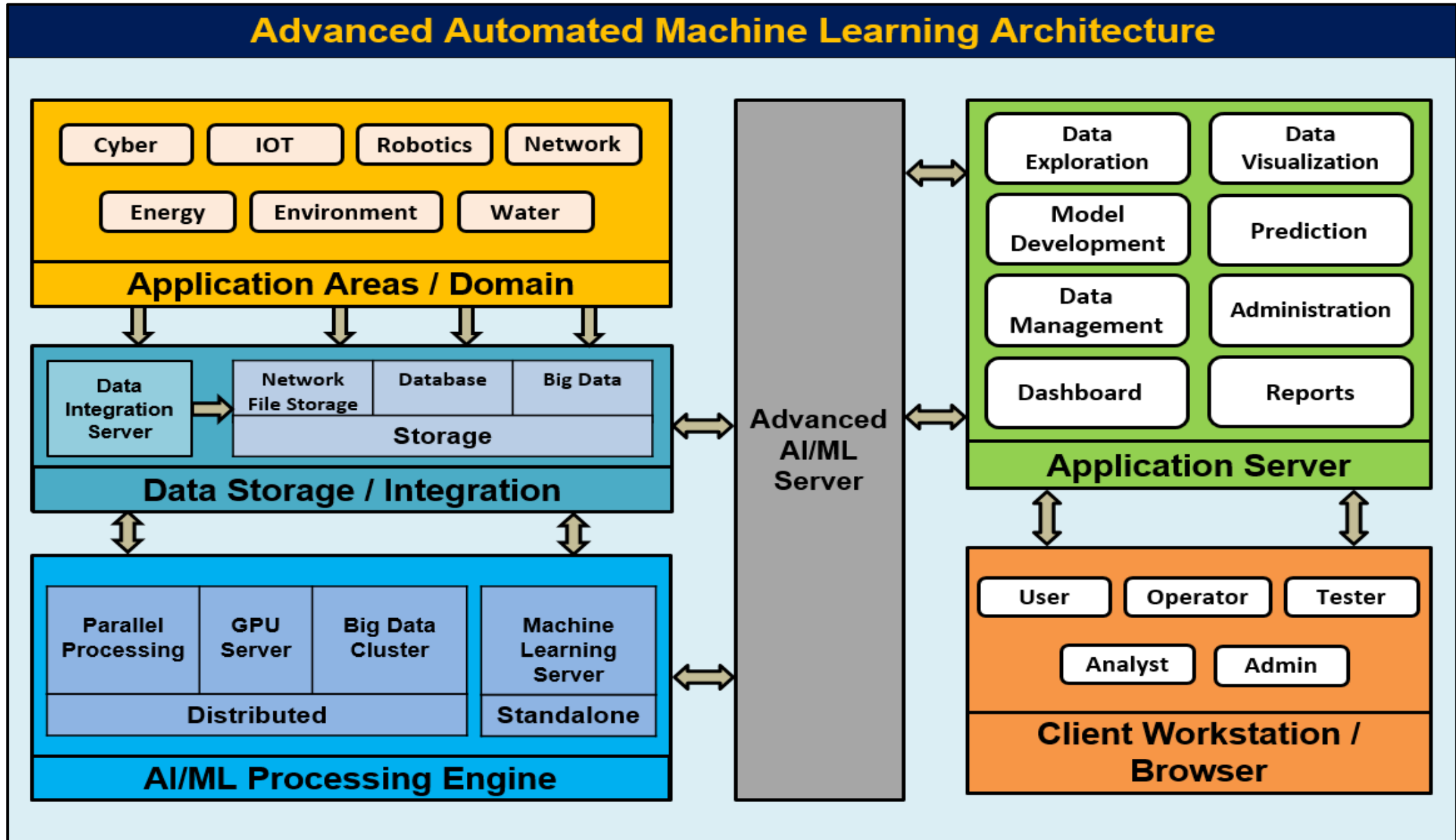


**DATABASE / MACHINE LEARNING SERVERS / BIG DATA CLUSTER**

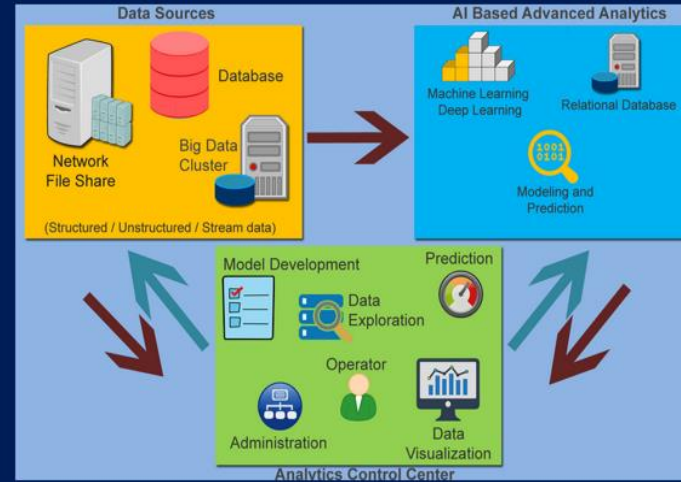
**IT INFRASTRUCTURE AND APPLICATIONS**

# AAML Overview





## Advanced Automated Machine Learning System



This is an automated Artificial Intelligence / Machine Learning framework for advanced analytics and visualization supporting various domains.



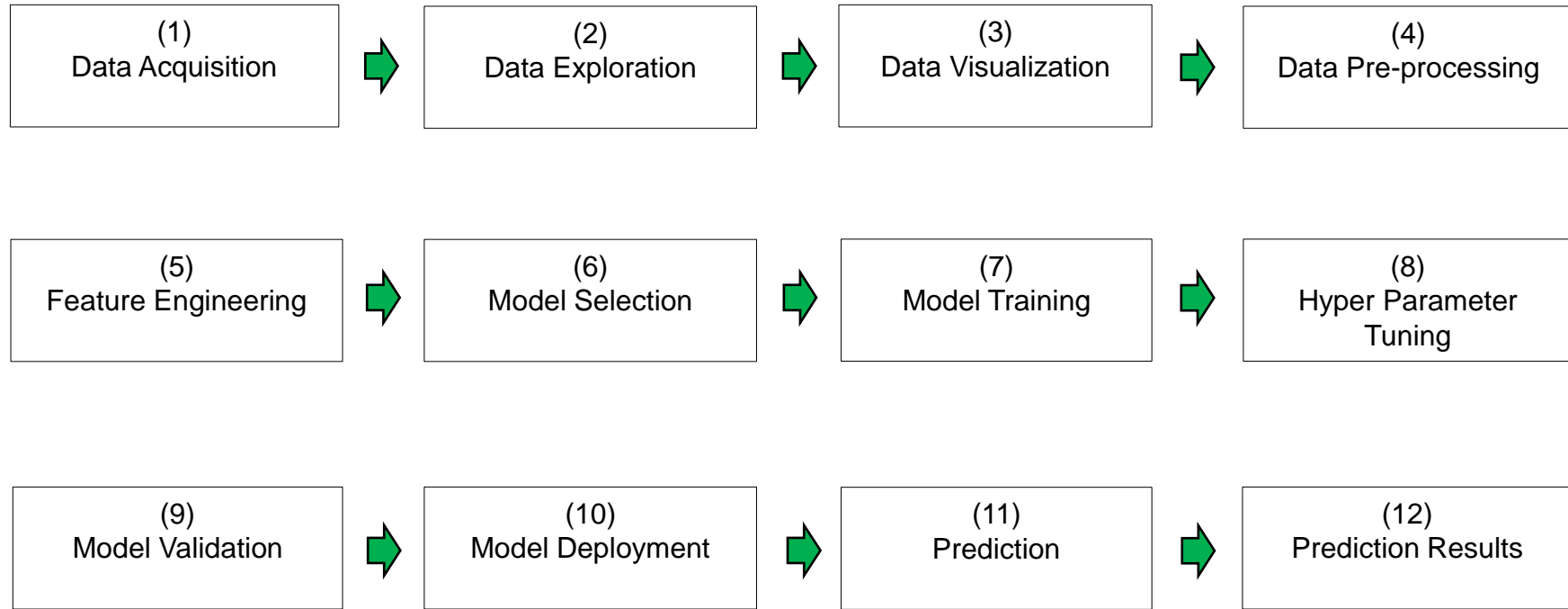
Research Sponsored by

Department of Defense - Test Resource Management Center

Developed by Florida International University



# Data Science Workflow



## Machine Learning Problems

Classification

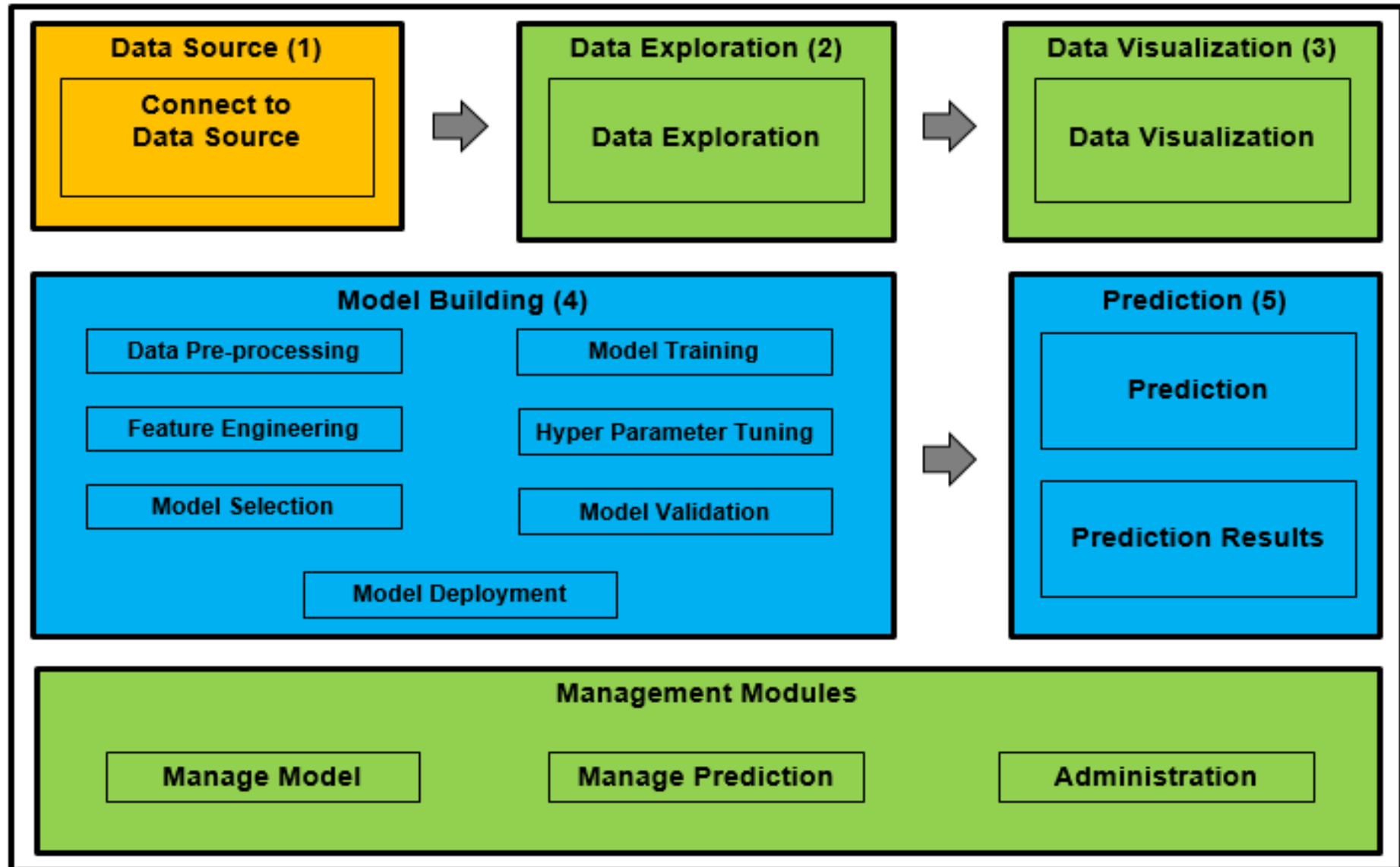
Regression

Time-Series

Anomaly Detection



# AAML Workflow





# AAML Modules

 Home

 Data Source

 Data Exploration

 Data Visualization

 Model Building ▼

 Prediction ▼

 Manage Model

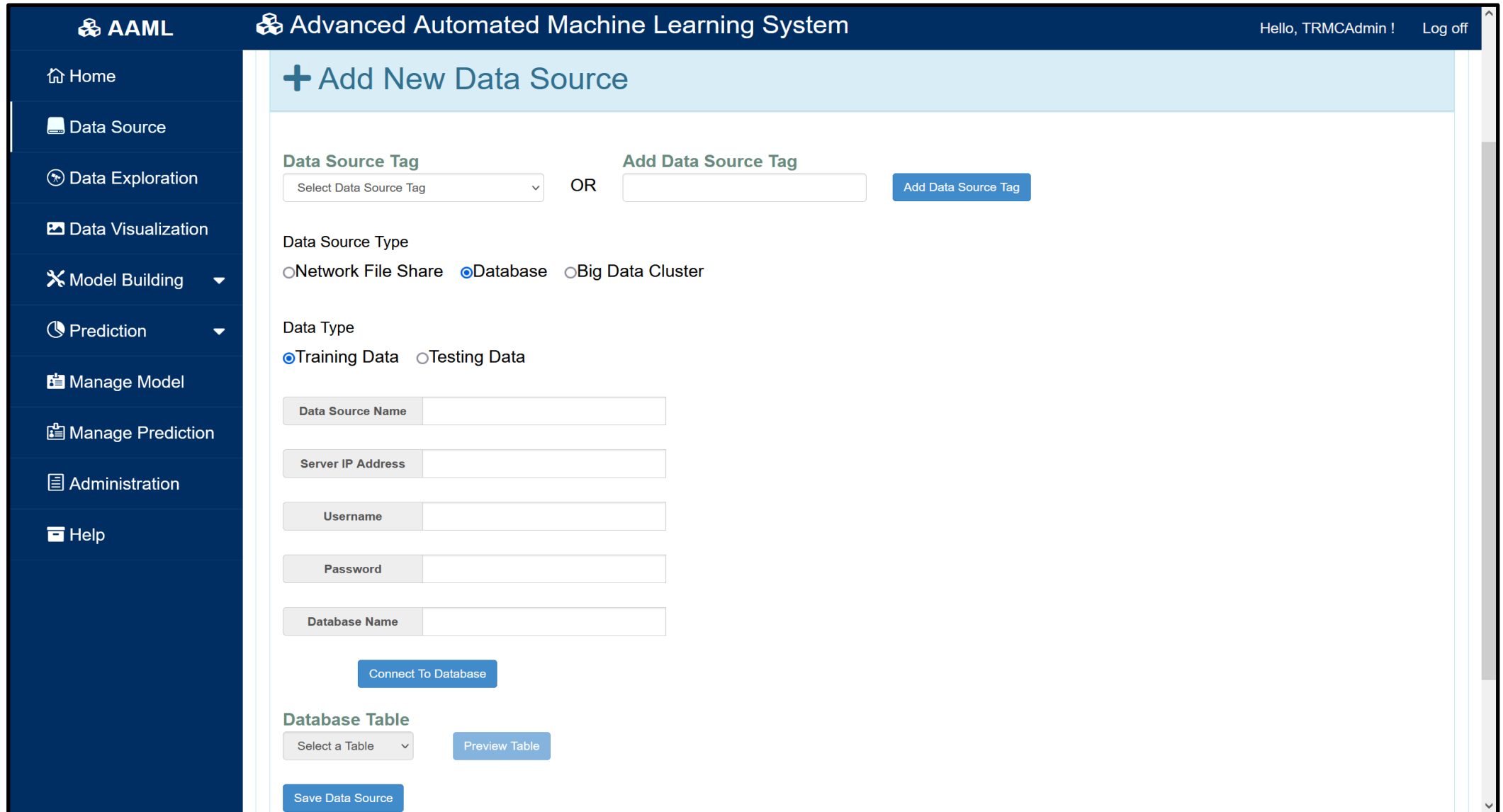
 Manage Prediction

 Administration

 Help



# Add Data Source – Network File Share



The screenshot displays the 'Add New Data Source' page in the AAML system. The interface includes a dark blue sidebar with navigation options: Home, Data Source, Data Exploration, Data Visualization, Model Building, Prediction, Manage Model, Manage Prediction, Administration, and Help. The main content area is titled 'Add New Data Source' and contains the following form elements:

- Data Source Tag:** A dropdown menu labeled 'Select Data Source Tag' and an 'Add Data Source Tag' button.
- Data Source Type:** Radio buttons for 'Network File Share', 'Database' (selected), and 'Big Data Cluster'.
- Data Type:** Radio buttons for 'Training Data' (selected) and 'Testing Data'.
- Input Fields:** Text boxes for 'Data Source Name', 'Server IP Address', 'Username', 'Password', and 'Database Name'.
- Buttons:** 'Connect To Database' and 'Preview Table'.
- Database Table:** A dropdown menu labeled 'Select a Table' and a 'Preview Table' button.
- Save Button:** 'Save Data Source'.

The top navigation bar shows 'AAML Advanced Automated Machine Learning System' and the user 'Hello, TRMCAAdmin!' with a 'Log off' link.



# Add Data Source – Database

**AAML** **Advanced Automated Machine Learning System** Hello, TRMCAdmin! [Log off](#)

## + Add New Data Source

**Data Source Tag**  OR **Add Data Source Tag**  [Add Data Source Tag](#)

**Data Source Type**  
 Network File Share  Database  Big Data Cluster

**Data Type**  
 Training Data  Testing Data

**Data Source Name**

**Server IP Address**

**Username**

**Password**

**Database Name**

[Connect To Database](#)

**Database Table**  
 [Preview Table](#)

[Save Data Source](#)



# Add Data Source – Big Data Cluster

**AAML**    **Advanced Automated Machine Learning System**    Hello, TRMCAAdmin !    Log off

- Home
- Data Source
- Data Exploration
- Data Visualization
- Model Building
- Prediction
- Manage Model
- Manage Prediction
- Administration
- Help

## + Add New Data Source

**Data Source Tag**  OR **Add Data Source Tag**

**Data Source Type**  
 Network File Share    Database    Big Data Cluster

**Data Type**  
 Training Data    Testing Data

**Data Source Name**

**Big Data Cluster IP**

**Username**

**Password**

**Big Data Cluster Path**



AAML
Advanced Automated Machine Learning System
Hello, TRMCAAdmin ! [Log off](#)

- [Home](#)
- [Data Source](#)
- [Data Exploration](#)
- [Data Visualization](#)
- [Model Building](#)
- [Prediction](#)
- [Manage Model](#)
- [Manage Prediction](#)
- [Administration](#)
- [Help](#)

## Data Exploration

**Data Source Tag**

**Data Source Type**

**Data Source**

Feature Name	Data Type	Rows Count	Minimum	Maximum	First Quartile	Second Quartile	Third Quartile	Mean	Standard Deviation	Missing Value	Skewness	Kurtosis	Outliers
Duration	float64	40002	0.00	3,600.00	0.00	0.12	34.46	600.87	1,183.74	0	1.65	0.95	0
Dir	int64	40002	0.00	7.00	1.00	1.00	1.00	0.77	0.45	0	0.02	14.54	16
Sport	int64	40002	1.00	65,535.00	2,079.00	4,006.00	35,878.75	17,786.89	20,997.90	0	0.96	-0.69	0
Dport	int64	40002	20.00	65,532.00	53.00	53.00	13,363.00	6,880.57	12,416.46	0	2.67	7.78	1483
SrcBytes	float64	40002	0.00	3,760,000,000.00	2,470,000,000.00	2,470,000,000.00	2,470,000,000.00	2,318,228,358.58	523,524,295.84	0	-1.06	2.57	487
TotBytes	float64	40002	18,200,000.00	4,290,000,000.00	2,470,000,000.00	2,470,000,000.00	3,220,000,000.00	2,440,830,880.96	746,959,204.23	0	-0.47	-0.37	99
PacketSent	int64	40002	1.00	214,827.00	2.00	2.00	6.00	41.84	1,620.90	0	98.77	11,089.09	23
PacketReceived	int64	40002	60.00	248,405,120.00	186.00	248.00	598.00	24,715.00	1,668,173.97	0	110.79	14,257.86	14
TotPkts	int64	40002	1.00	2,190,507.00	85,813.00	2,190,507.00	2,190,507.00	1,569,509.17	963,618.56	0	-0.91	-1.17	0
Class	int64	40002	0.00	1.00	0.00	0.50	1.00	0.50	0.50	0	0.00	-2.00	0

© 2022 - Advanced Automated Machine Learning System



AAML
Advanced Automated Machine Learning System
Hello, TRMCAAdmin! [Log off](#)

- [Home](#)
- [Data Source](#)
- [Data Exploration](#)
- [Data Visualization](#)
- [Model Building](#)
- [Prediction](#)
- [Manage Model](#)
- [Manage Prediction](#)
- [Administration](#)
- [Help](#)

## Data Visualization

**Data Source Tag**

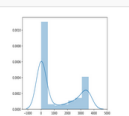
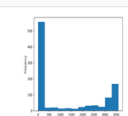
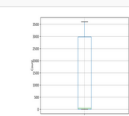
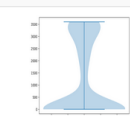
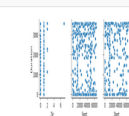
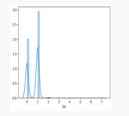
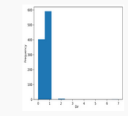
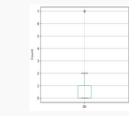
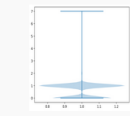
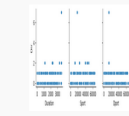
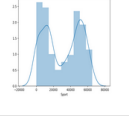
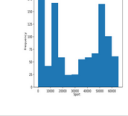
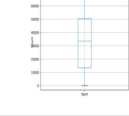
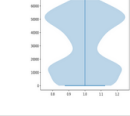
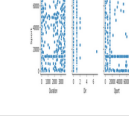
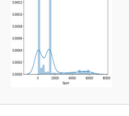

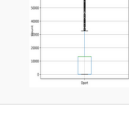
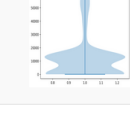
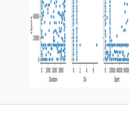
**Data Source Type**

**Data Source**

**Visualization Type**

 Features  Dataset

Visualize Data
Preview Data

Feature Name	Distribution Plot	Histogram Plot	Box Plot	Violin Plot	Pair Plot
Duration					
Dir					
Sport					
Dport					

© 2022 - Advanced Automated Machine Learning System



Home

Data Source

Data Exploration

Data Visualization

Model Building

Prediction

Manage Model

Manage Prediction

Administration

Help

## Data Visualization

Data Source Tag

Cyber

Data Source Type

Network File Share

Data Source

BotNetDetection\_Classification\_NFS\_Train

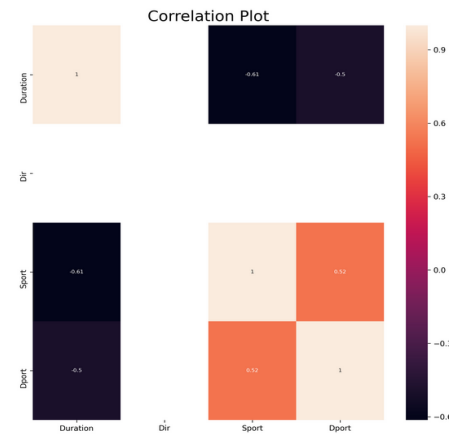
Visualization Type

Features Dataset

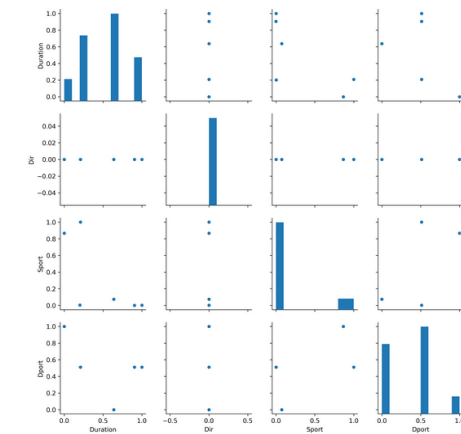
Visualize Data

Preview Data

### Correlation Plot



### Pair Plot



AAML
Advanced Automated Machine Learning System
Hello, TRMCAAdmin! [Log off](#)

- [Home](#)
- [Data Source](#)
- [Data Exploration](#)
- [Data Visualization](#)
- [Model Building](#)
- [Prediction](#)
- [Manage Model](#)
- [Manage Prediction](#)
- [Administration](#)
- [Help](#)

## Classification Modeling

**Data Source Tag**

**Data Source Type**  
 Network File Share    Database    Big Data Cluster

**Model Name/Description**

<b>Model Name</b>	BotNet Model 2
<b>Model Description</b>	Classification NFS

**Select Algorithm(s)**

- K Nearest Neighbor
- Logistic Regression
- Decision Tree
- Neural Network
- Random Forest

**Data Source**

<b>Network File Share</b>	<b>Selected Directory</b>	<b>Features</b>	<b>Target Label</b>
<input type="text" value="BotNetDetection_Classific"/>	\\10.2.0.1\AAMLShared\ACC_Data\AdvancedCyberAnalysis\Classification\BotNetDetection\Training	<input type="button" value="Select Features"/>	<input type="text" value="Class"/>

**Data Query**

Folder Path	Features	Target Label	
\\10.2.0.1\AAMLShared\ACC_Data\AdvancedCyberAnalysis\Classification\BotNetDetection\Training	Duration,Dir,Sport,Dport,SrcBytes,TotBytes,PacketSent,PacketReceived,TotPkts	Class	<input type="button" value="View"/> <input type="button" value="Delete"/>

© 2022 - Advanced Automated Machine Learning System





AAML
Advanced Automated Machine Learning System
Hello, TRMCAAdmin! [Log off](#)

- [Home](#)
- [Data Source](#)
- [Data Exploration](#)
- [Data Visualization](#)
- [Model Building](#)
- [Prediction](#)
- [Manage Model](#)
- [Manage Prediction](#)
- [Administration](#)
- [Help](#)

## Model Results Preview

**Results**

		Algorithm Name	Metrics	Accuracy	F1 Score	Sensitivity/Recall	Precision	Confusion Matrix
<a href="#">View Validation Result</a>	<a href="#">Download Result</a>	K Nearest Neighbor	Accuracy: 0.98 Precision: 0.98 Recall: 0.98 F1_Score: 0.98					
<a href="#">View Validation Result</a>	<a href="#">Download Result</a>	Logistic Regression	Accuracy: 0.76 Precision: 0.76 Recall: 0.76 F1_Score: 0.76					
<a href="#">View Validation Result</a>	<a href="#">Download Result</a>	Decision Tree	Accuracy: 1.0 Precision: 1.00 Recall: 1.00 F1_Score: 1.00					
<a href="#">View Validation Result</a>	<a href="#">Download Result</a>	Neural Network	Accuracy: 1.0 Precision: 1.00 Recall: 1.00 F1_Score: 1.00					
<a href="#">View Validation Result</a>	<a href="#">Download Result</a>	Random Forest	Accuracy: 1.0 Precision: 1.00 Recall: 1.00 F1_Score: 1.00					

[Close](#)

[Search Models](#)

**Inserted On**

- 4/19/2022 10:03:19 PM
- 4/15/2022 9:24:04 PM
- 4/15/2022 9:13:51 PM
- 4/15/2022 9:11:59 PM
- 4/15/2022 9:11:13 PM



AAML
Advanced Automated Machine Learning System
Hello, TRMCAdmin! [Log off](#)

- [Home](#)
- [Data Source](#)
- [Data Exploration](#)
- [Data Visualization](#)
- [Model Building](#)
- [Prediction](#)
- [Manage Model](#)
- [Manage Prediction](#)
- [Administration](#)
- [Help](#)

BotNet Model 2

**Prediction Name/Description**

Prediction Name	BotNet Prediction 2
Prediction Description	BotNet Prediction 2

**Select Algorithm(s)**

- K Nearest Neighbor
- Logistic Regression
- Decision Tree
- Neural Network
- Random Forest

**Data Source**

<p><b>Network File Share</b></p> <span>BotNetDetection_Classific</span>	<p><b>Select Directory</b></p> <span>\\192.168.2.1\AAMLShared\ACC_Data\AdvancedCyberAnalysis\Classification\BotNetDetection\Testing</span>	<a href="#">Add Data Query</a>
---	--	--------------------------------

**Data Query**



Folder Path	Features	Target Label	
\\192.168.2.1\AAMLShared\ACC_Data\AdvancedCyberAnalysis\Classification\BotNetDetection\Testing	Duration,Dir,Sport,Dport,SrcBytes,TotBytes,PacketSent,PacketReceived,TotPkts	Class	<a href="#">View</a> <a href="#">Delete</a>

[Build Prediction](#)

© 2022 - Advanced Automated Machine Learning System




# Prediction Results

 **AAML**
 **Advanced Automated Machine Learning System**

\\10.10.10.137\AAMLShared\ACC\_Data\AdvancedCyberAnalysis\Classification\BotNetDetection\Testing

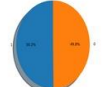
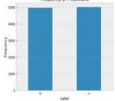
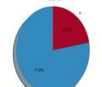
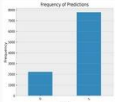
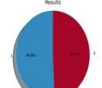
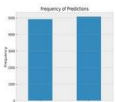
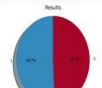

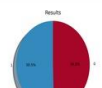

Duration,Dir,Sport,Dport,SrcBytes,TotBy  
PacketReceived,TotPkts

 BotNet Prediction 2\_5\_Random Forest.csv  
Completed — 770 KB

[Show all downloads](#)

Build Prediction

### Results

		Algorithm Name	Prediction Distribution	Frequency
<span style="background-color: #007bff; color: white; padding: 2px 5px; border-radius: 3px;">View Prediction Result</span>	<span style="background-color: #007bff; color: white; padding: 2px 5px; border-radius: 3px;">Download Result</span>	K Nearest Neighbor		
<span style="background-color: #007bff; color: white; padding: 2px 5px; border-radius: 3px;">View Prediction Result</span>	<span style="background-color: #007bff; color: white; padding: 2px 5px; border-radius: 3px;">Download Result</span>	Logistic Regression		
<span style="background-color: #007bff; color: white; padding: 2px 5px; border-radius: 3px;">View Prediction Result</span>	<span style="background-color: #007bff; color: white; padding: 2px 5px; border-radius: 3px;">Download Result</span>	Decision Tree		
<span style="background-color: #007bff; color: white; padding: 2px 5px; border-radius: 3px;">View Prediction Result</span>	<span style="background-color: #007bff; color: white; padding: 2px 5px; border-radius: 3px;">Download Result</span>	Neural Network		
<span style="background-color: #007bff; color: white; padding: 2px 5px; border-radius: 3px;">View Prediction Result</span>	<span style="background-color: #007bff; color: white; padding: 2px 5px; border-radius: 3px;">Download Result</span>	Random Forest		

© 2022 - Advanced Automated Machine Learning System



AAML
Advanced Automated Machine Learning System
Hello, TRMCAdmin! [Log off](#)

- [Home](#)
- [Data Source](#)
- [Data Exploration](#)
- [Data Visualization](#)
- [Model Building](#)
- [Prediction](#)
- [Manage Model](#)
- [Manage Prediction](#)
- [Administration](#)
- [Help](#)

## Models

**Data Source Tag**

**Data Source Type**

View Models

OR

**Search Model**

Search Models

	Model ID	Model Name	Description	Username	Label(s)	Inserted On
<a href="#">View Features</a> <a href="#">View Results</a> <a href="#">Delete</a>	5	BotNet Model 2	Classification NFS	TRMCAdmin	Class	4/19/2022 10:03:19 PM
<a href="#">View Features</a> <a href="#">View Results</a> <a href="#">Delete</a>	1	BotNet Model 1	Classification NFS	AAMLAdmin	Class	4/15/2022 9:11:13 PM

© 2022 - Advanced Automated Machine Learning System



AAML
Advanced Automated Machine Learning System
Hello, TRMCAdmin! [Log off](#)

- [Home](#)
- [Data Source](#)
- [Data Exploration](#)
- [Data Visualization](#)
- [Model Building](#)
- [Prediction](#)
- [Manage Model](#)
- [Manage Prediction](#)
- [Administration](#)
- [Help](#)

## Predictions

**Data Source Tag**

**Data Source Type**

View Predictions

OR

**Search Prediction**

Search Predictions

	Test ID	Test Name	Description	Username	Label(s)	Model ID	Inserted On
<a href="#">View Features</a> <a href="#">View Results</a> <a href="#">Delete</a>	5	BotNet Prediction 2	BotNet Prediction 2	TRMCAdmin	Class	5	4/19/2022 10:33:16 PM
<a href="#">View Features</a> <a href="#">View Results</a> <a href="#">Delete</a>	4	TCP Attack Prediction 1	Anomaly Detection BDC	AAMLAdmin	Class	4	4/15/2022 10:38:29 PM
<a href="#">View Features</a> <a href="#">View Results</a> <a href="#">Delete</a>	3	Engine Sensor Prediction 1	Time Series DB	AAMLAdmin	Machine_Status	3	4/15/2022 9:38:43 PM
<a href="#">View Features</a> <a href="#">View Results</a> <a href="#">Delete</a>	2	Energy Prediction 1	Regression NFS	AAMLAdmin	Appliances_Energy_Consumption	2	4/15/2022 9:37:26 PM
<a href="#">View Features</a> <a href="#">View Results</a> <a href="#">Delete</a>	1	BotNet Prediction 1	Classification NFS	AAMLAdmin	Class	1	4/15/2022 9:36:51 PM

© 2022 - Advanced Automated Machine Learning System



AAML
Advanced Automated Machine Learning System
Hello, TRMCAAdmin! [Log off](#)

- [Home](#)
- [Data Source](#)
- [Data Exploration](#)
- [Data Visualization](#)
- [Model Building](#)
- [Prediction](#)
- [Manage Model](#)
- [Manage Prediction](#)
- [Administration](#)
- [Help](#)

## Administration

User Management
Data Source Tag
Organization
Algorithm Group
Algorithm

### Algorithm

Algorithm Group

All

View Algorithms
Add Algorithm

	Algorithm ID	Algorithm Name	Is Active	
<a href="#">Edit</a>	1	K Nearest Neighbor	True	<a href="#">Delete</a>
<a href="#">Edit</a>	2	Logistic Regression	True	<a href="#">Delete</a>
<a href="#">Edit</a>	4	Decision Tree	True	<a href="#">Delete</a>
<a href="#">Edit</a>	6	Neural Network	True	<a href="#">Delete</a>
<a href="#">Edit</a>	7	Linear Regression	True	<a href="#">Delete</a>
<a href="#">Edit</a>	8	Ridge Regression	True	<a href="#">Delete</a>
<a href="#">Edit</a>	9	Lasso Regression	True	<a href="#">Delete</a>
<a href="#">Edit</a>	10	ElasticNet Regression	True	<a href="#">Delete</a>
<a href="#">Edit</a>	11	Random Forest	True	<a href="#">Delete</a>



 Home

 Data Source

 Data Exploration

 Data Visualization

 Model Building 

 Prediction 

 Manage Model

 Manage Prediction

 Administration

 Help

## Advanced Automated Machine Learning System Help Topics

This is an automated Artificial Intelligence / Machine Learning framework for advanced analytics and visualization supporting various domains.

### Advanced Automated Machine Learning System - Operation Guide

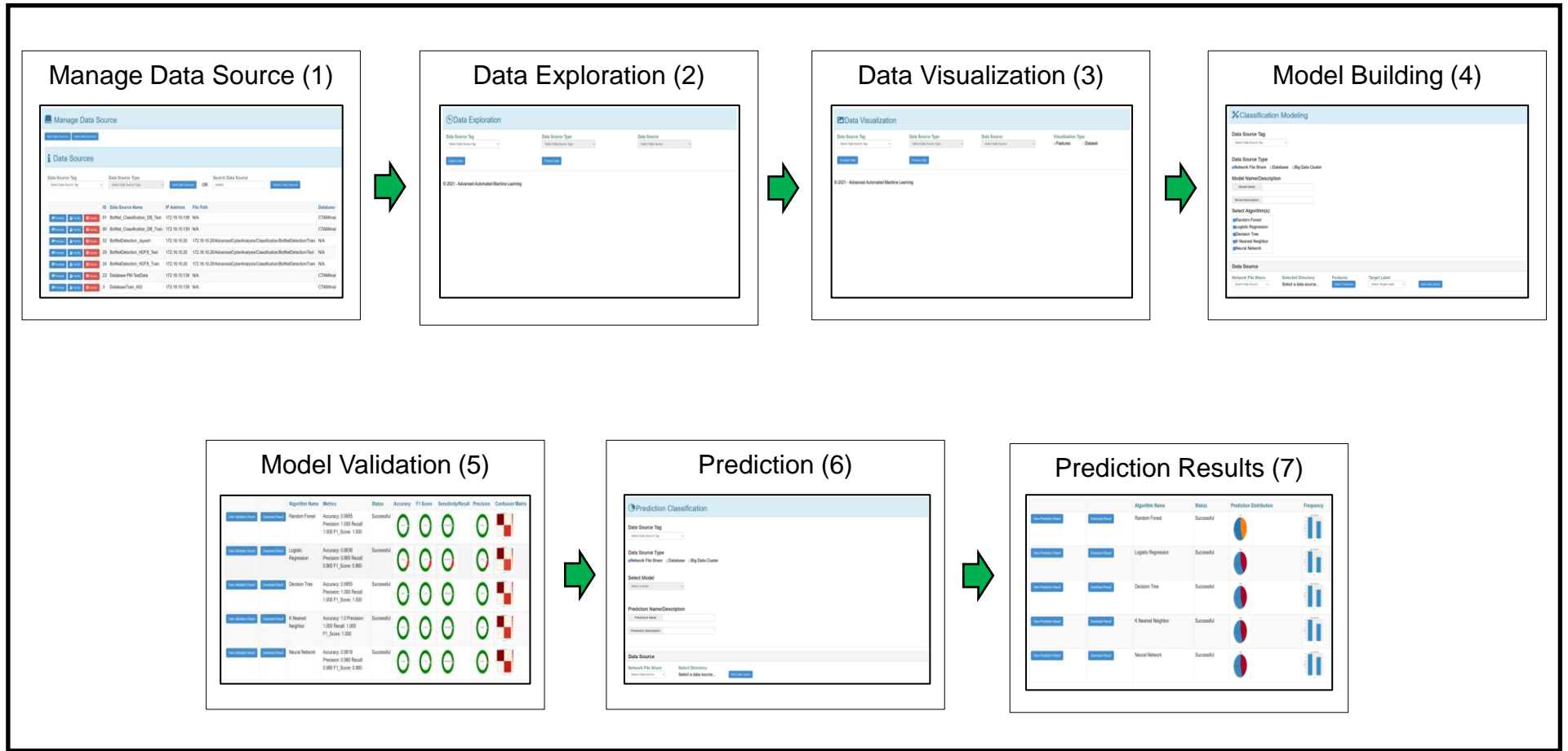
The Advanced Automated Machine Learning (AAML) System is a centralized web application used by the operator to create models and predictions for datasets. This document contains instruction of how to use each module of the AAML System.

 [AAML System User Guide](#)

© 2022 - Advanced Automated Machine Learning System



# AAML System Recap





# AAML System Summary

- Automation of Machine Learning model development and prediction in few steps
- Application can be used with minimal machine learning knowledge
- Dynamic connectivity to existing data sources in network file share, database and big data cluster
- Explore and visualize datasets prior to building model and prediction
- Access to the historical model and prediction results



**Task 6: AI for EM Problem Set (D&D):** Structural Health Monitoring of D&D Facility to Identify Cracks and Structural Defects for Surveillance and Maintenance

**Task 7: AI for EM Problem Set (Soil & GW):** Exploratory Data Analysis and Machine Learning Model for Hexavalent Chromium [Cr (VI)] Concentration in 100-H Area

**Task 8: AI for EM Problem Set (Soil & GW):** Data Analysis and Visualization of Sensor Data from the Wells at the SRS F-Area using Machine Learning



## Task 6: AI for EM Problem Set (D&D): Structural Health Monitoring of D&D Facility to Identify Cracks and Structural Defects for Surveillance and Maintenance

### Site Needs:

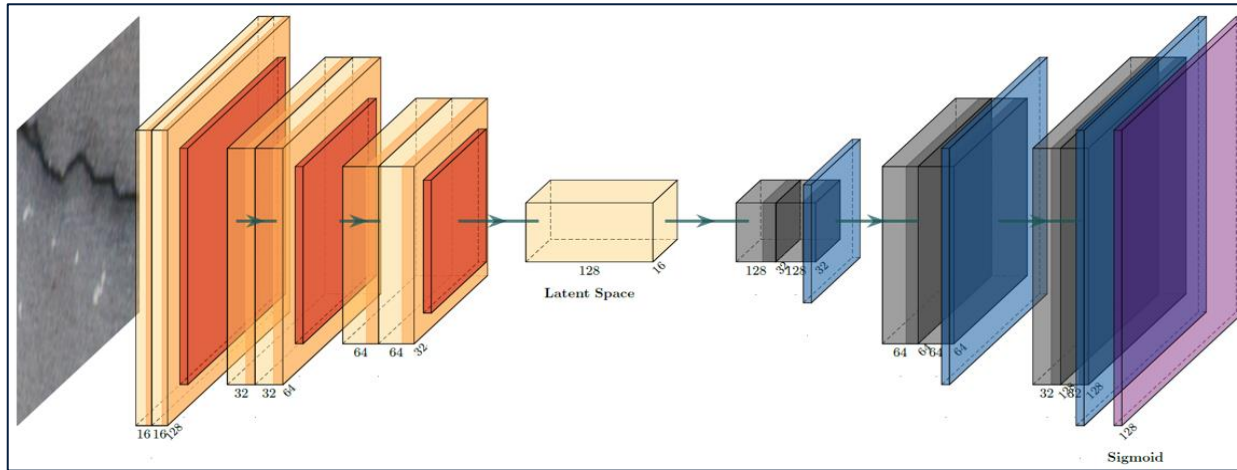
- Assess the structural integrity of aging facilities in support of ongoing surveillance and maintenance (S&M) across the DOE complex.
- Adequate inspections and data collection / analysis to be performed on a ongoing basis.

### Objectives:

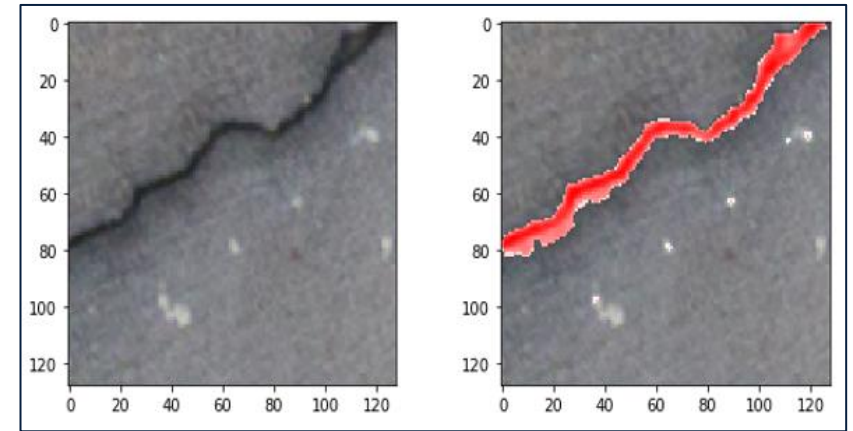
- Investigating specific applications of artificial intelligence and big data technologies to solve DOE-EM problem sets.
- Data Collection from the simulated testbed at FIU facility using imagery devices like Camera and Lidar.
- Implement Structural Health Monitoring (SHM) using Deep Learning (DL) algorithms to identify cracks on surfaces.
- Implementation of Auto Encoders (AE) with Convolutional Neural Network (CNN) layers and post image processing.
- Implementation of YOLO algorithm for the Object Detection.
- Deploy trained Machine Learning and Deep learning models in iOS devices on site.



## Task 6: AI for EM Problem Set (D&D): Structural Health Monitoring of D&D Facility to Identify Cracks and Structural Defects for Surveillance and Maintenance

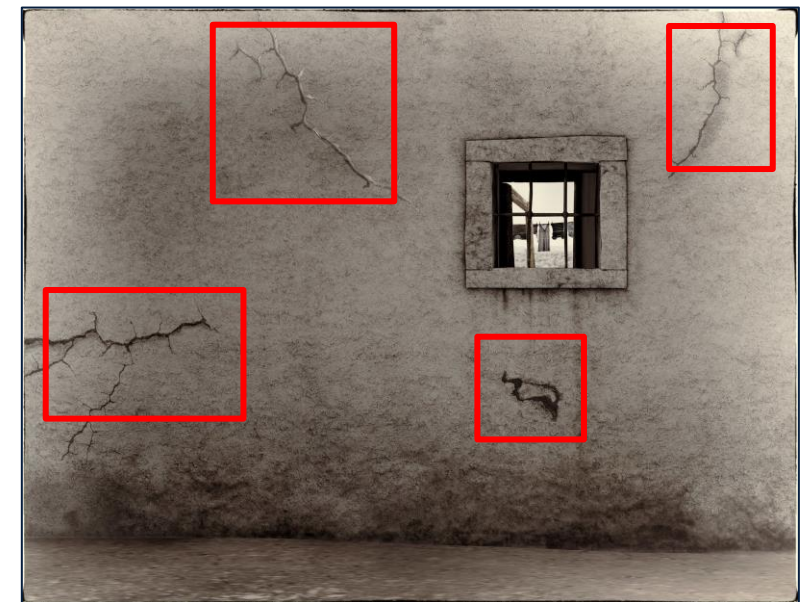
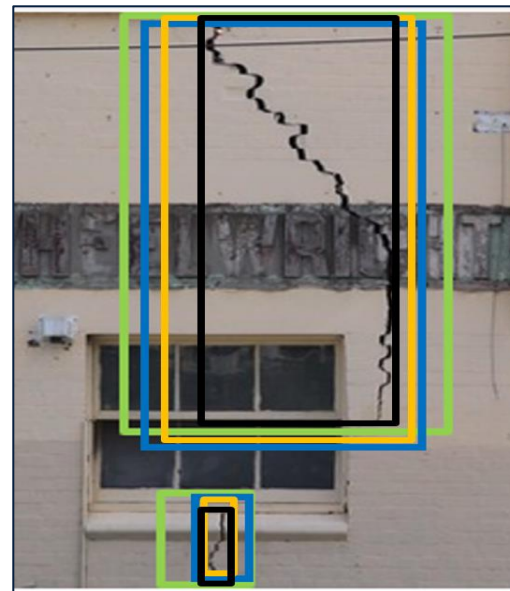


**Convolutional AutoEncoder (CAE) Deep Learning Architecture**



**Crack Detection and Heat map**

- Black**
  - Ground truth
  - Manually done.
  - Used for accuracy.
- Green**
  - Early model
  - No optimization.
  - No data augmentation.
- Blue**
  - Intermediate model
  - Initial optimizations.
  - Some data augmentation.
- Yellow**
  - Final model
  - Complete data augmentation.
  - Best one!



**Crack Detection Results**

**Site Needs:**

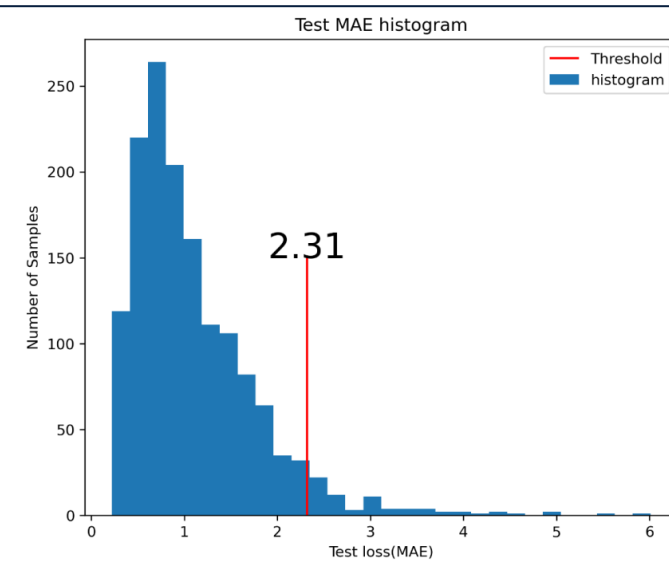
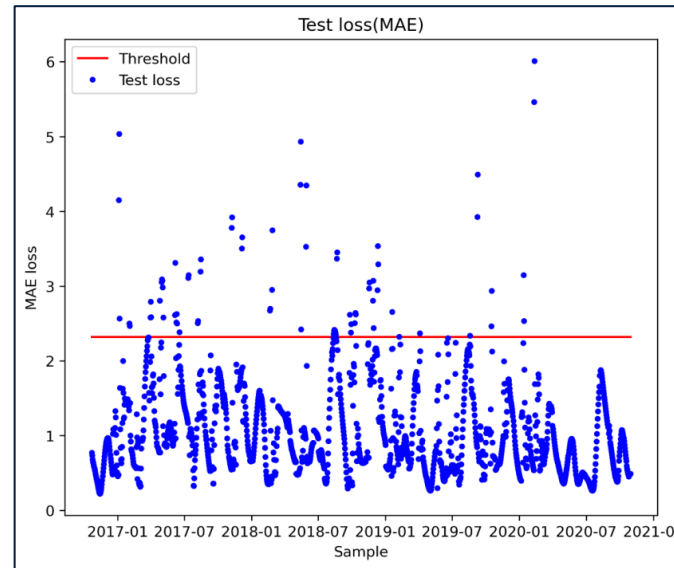
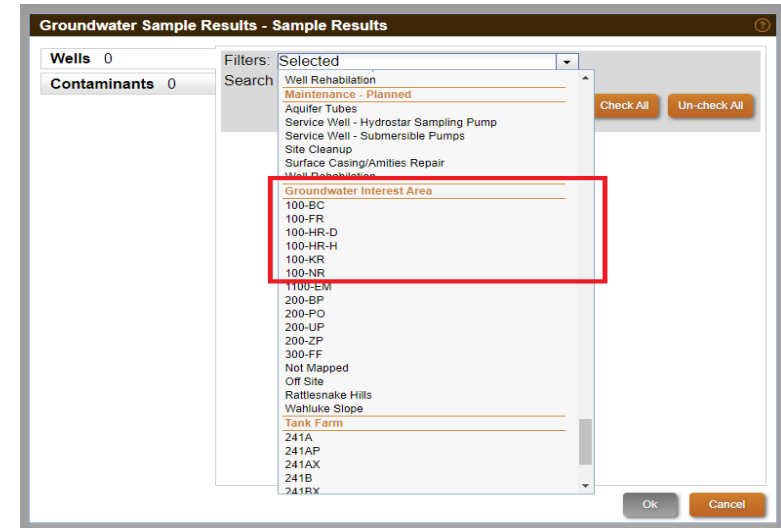
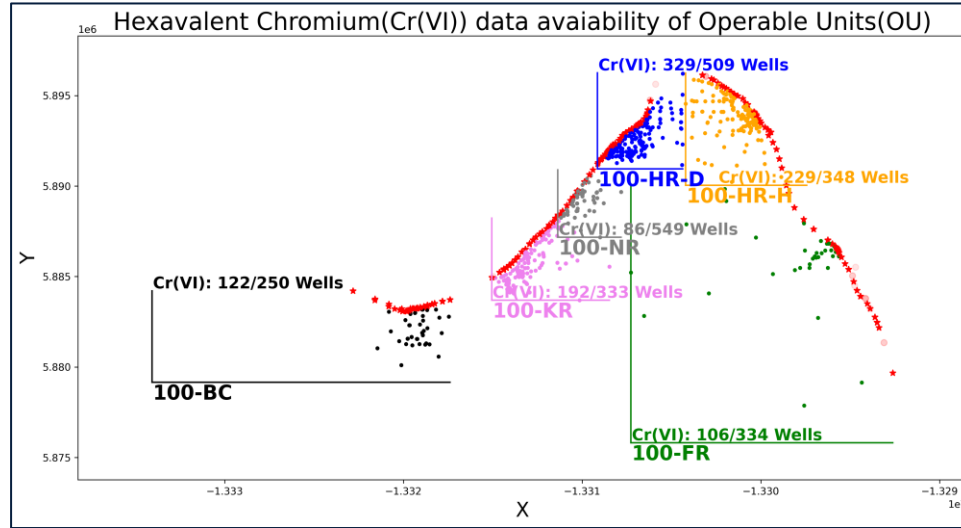
- To identify temporal and spatial relationships of subsurface chromium transport that reduces uncertainties in the conceptual site model (CSM).
- Data analysis & prediction using Artificial Intelligence and Machine Learning algorithms.

**Objectives:**

- Data source identification and pre-processing for data cleansing, discretization, and transformation.
- Perform exploratory data analysis using state-of-art statistical methods and various machine learning algorithms.
- Develop AI/ML models to explore spatiotemporal relationships of subsurface hexavalent chromium transport.



## Task 7: AI for EM Problem Set (Soil & GW): Exploratory Data Analysis and Machine Learning Model for Hexavalent Chromium [Cr (VI)] Concentration in 100-H Area



LSTM AutoEncoder / Decoder Based Anomaly Detection Results



**Site Needs:**

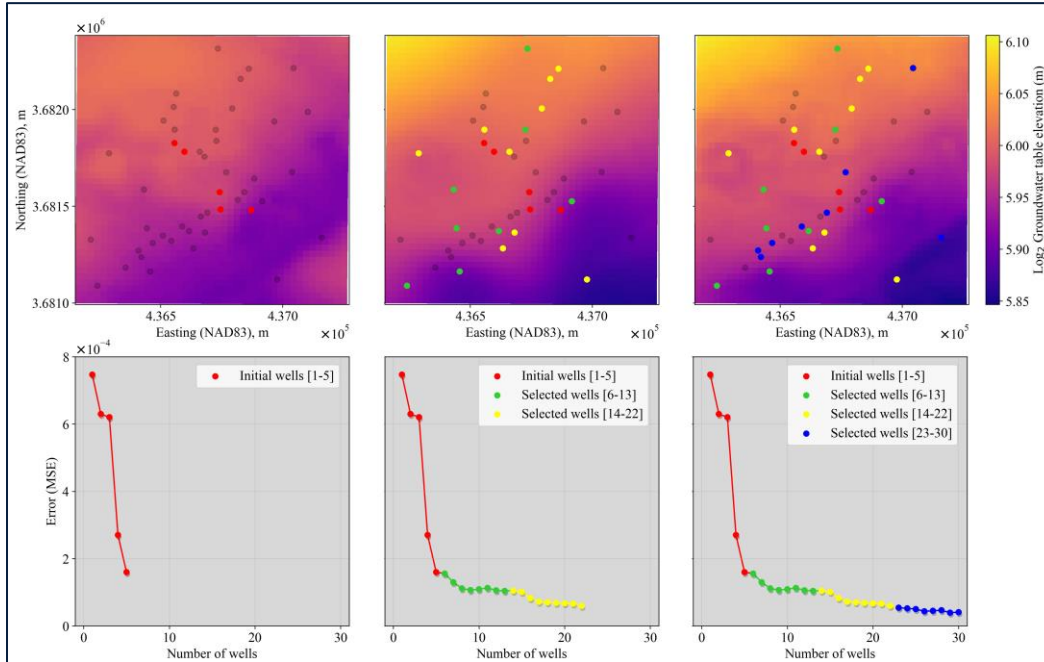
- Develop machine learning tools to automate the monitoring and forecasting of contaminant transport dynamics at the Savannah River Site (SRS) F-Area to support DOE-EM's goal for long time monitoring of contaminated groundwater sites.

**Objectives:**

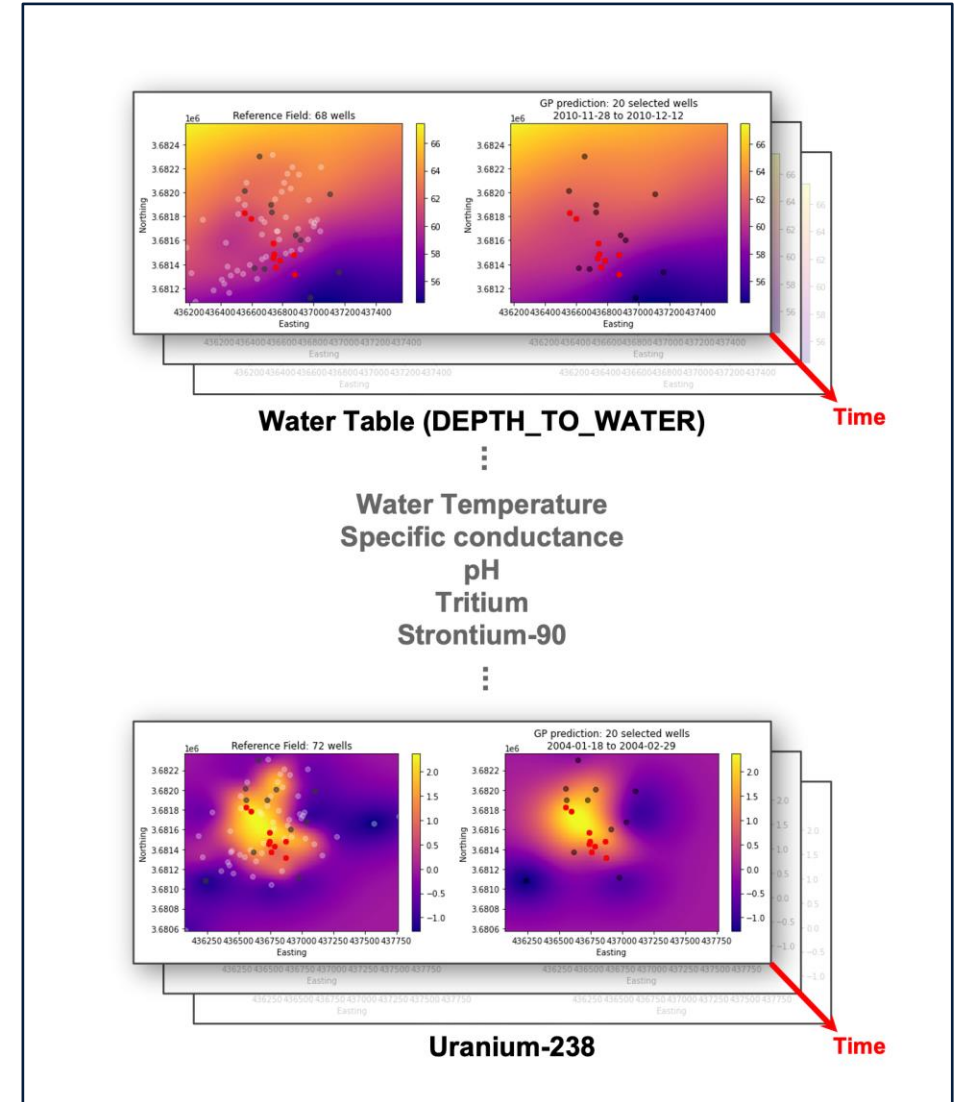
- Develop data exploration tools for understanding the spatial and temporal distribution of the F-Area dataset.
- Develop a spatial interpolation approach for estimating a plume.
- Examine proxy variables at the site.
- Develop a sensor placement optimization approach for identify subset of wells that captures the overall plume dynamic.



## Task 8: AI for EM Problem Set (Soil & GW): Data Analysis and Visualization of Sensor Data from the Wells at the SRS F-Area using Machine Learning



**Water Table optimization (single timestep)**



**Multi-analyte optimization (Multiple timesteps)**





# **Advanced Automated Machine Learning System Demonstration**





Thank You. Questions?