FIU Applied Research Center

Advanced Automated Machine Learning (AAML) System

Presented by Dr. Himanshu Upadhyay

Florida International University

FIU
Applied Research Center

Test and Evaluation / Science and Technology Program
Cyberspace Test Technology

Advanced Automated Machine Learning (AAML) System

Developed by: Florida International University

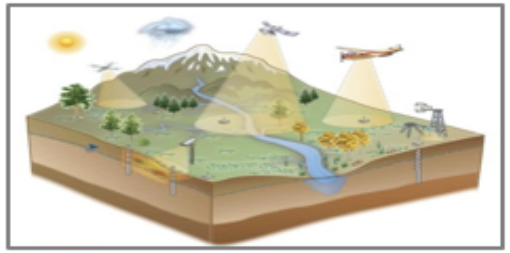
Research Sponsored by: Department of Defense (DOD) - Test Resource Management Center (TRMC)

FIU Applied Research Center

Artificial Intelligence / Machine Learning

ARTIFICIAL INTELLIGENCE / MACHINE LEARNING

APPLICATION DOMAIN



SENSOR / IMAGERY DATA FROM DOE-EM SITES

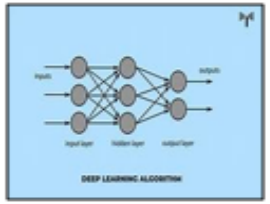
AI / ML TECHNOLOGY

TensorFlow

Keras

python

APACHE Spark



AI / ML FRAMEWORKS AND ALGORITHMS

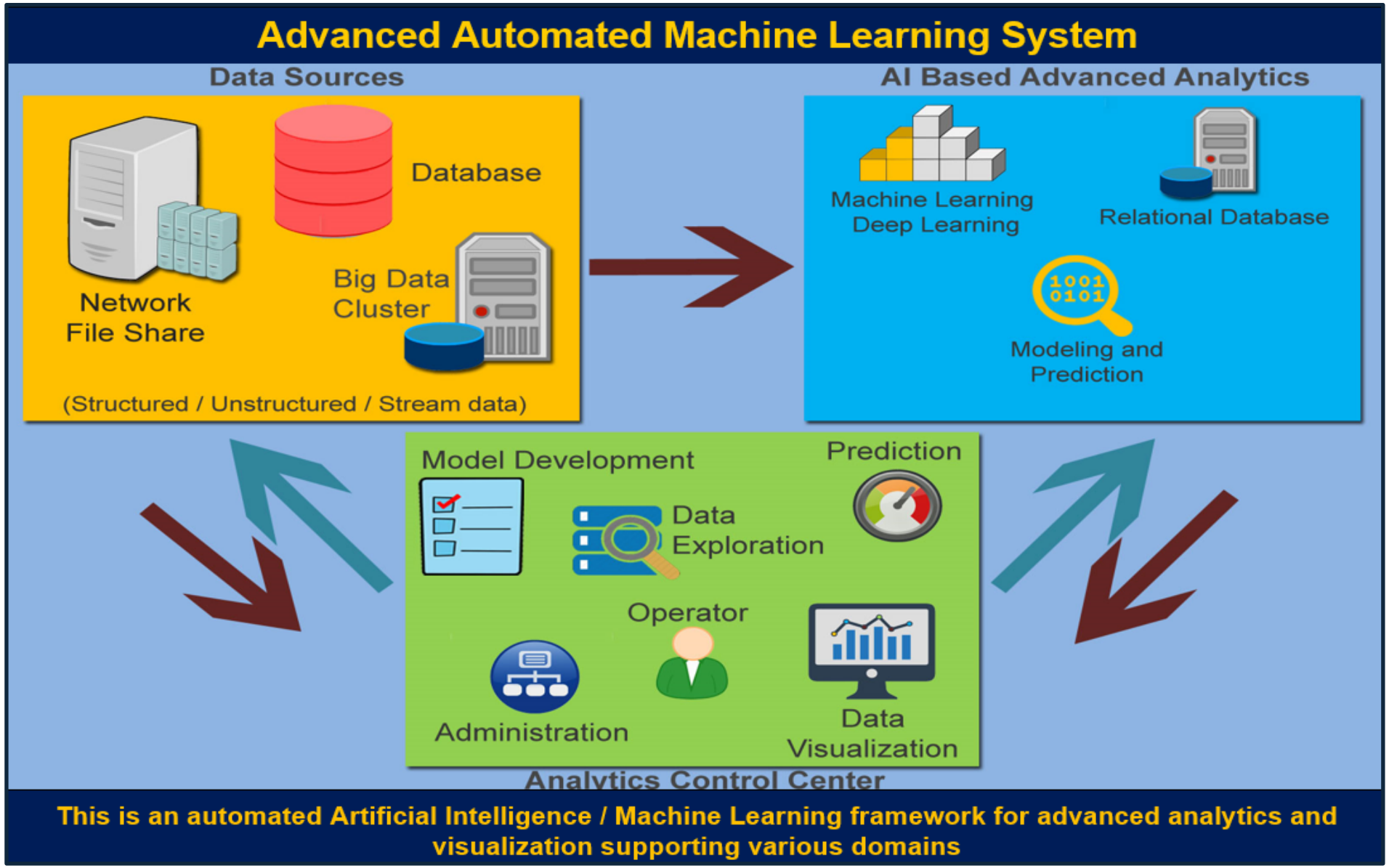
OPERATIONALISE AI / ML MODEL

DATABASE / MACHINE LEARNING SERVERS / BIG DATA CLUSTER

IT INFRASTRUCTURE AND APPLICATIONS

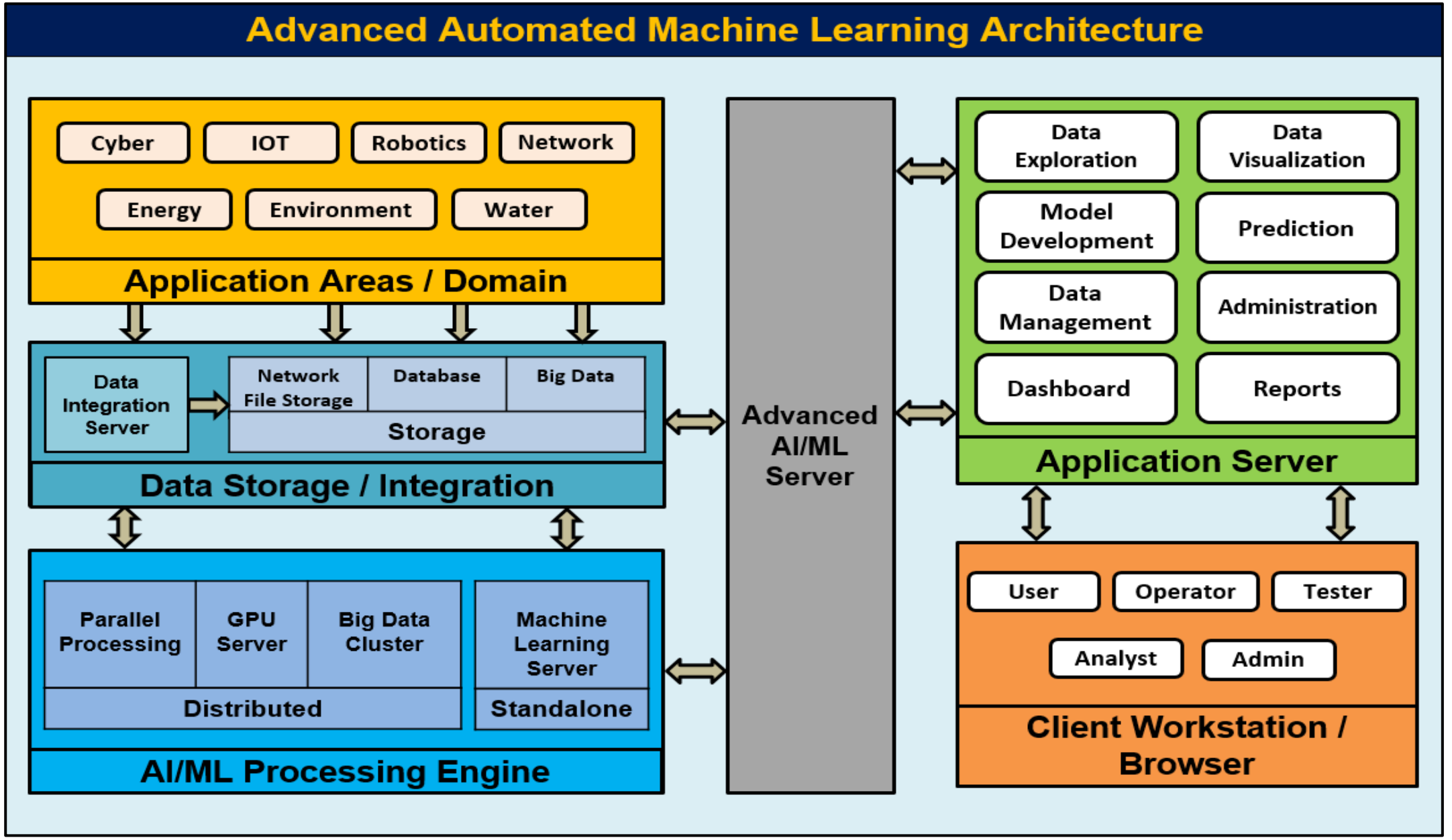
FIU
Applied Research Center

AAML Overview

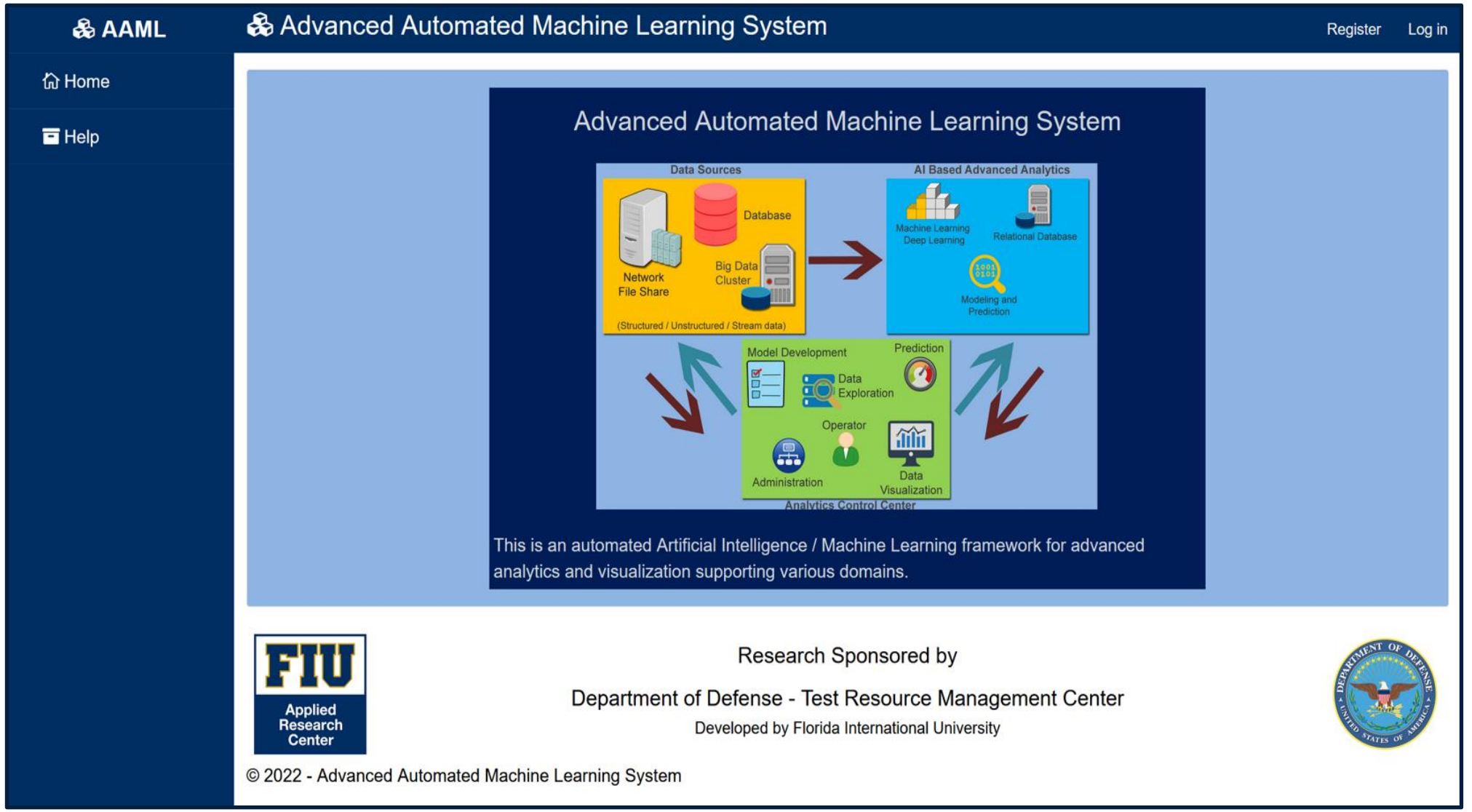


FIU
Applied Research Center

AAML Architecture

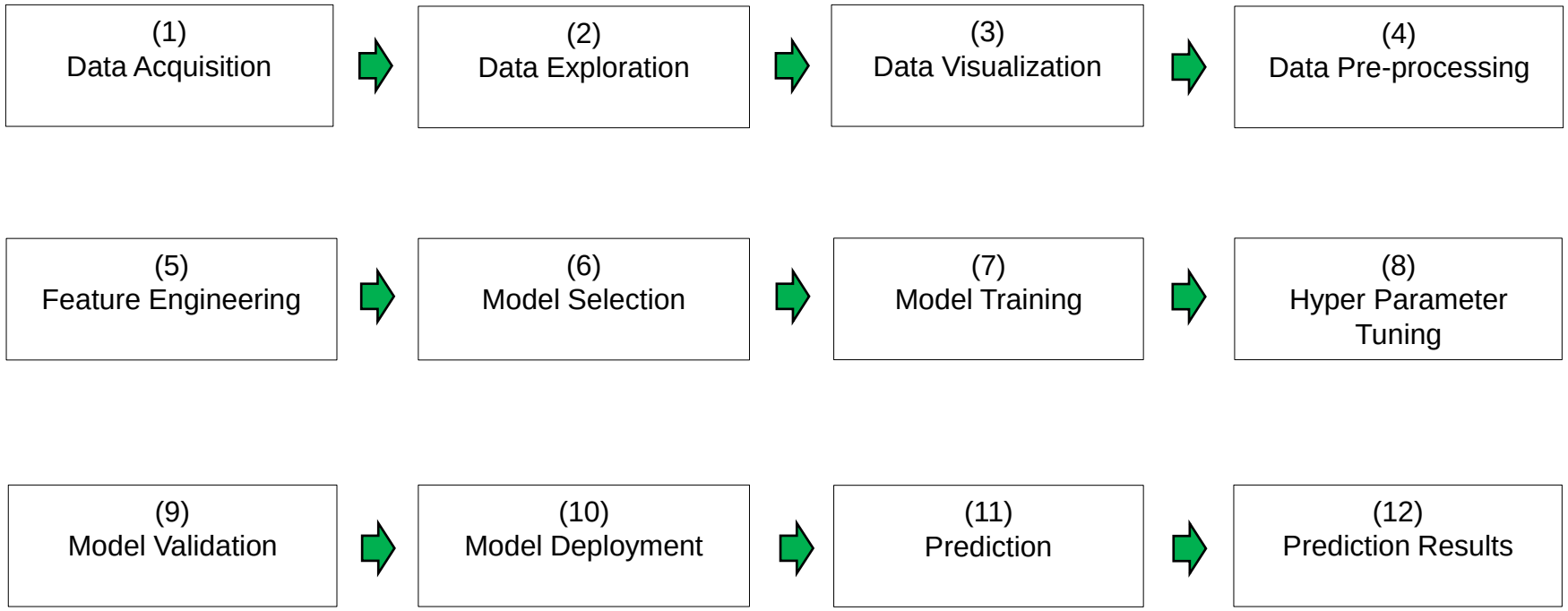


AAML Home Page



FIU
Applied Research Center

Data Science Workflow

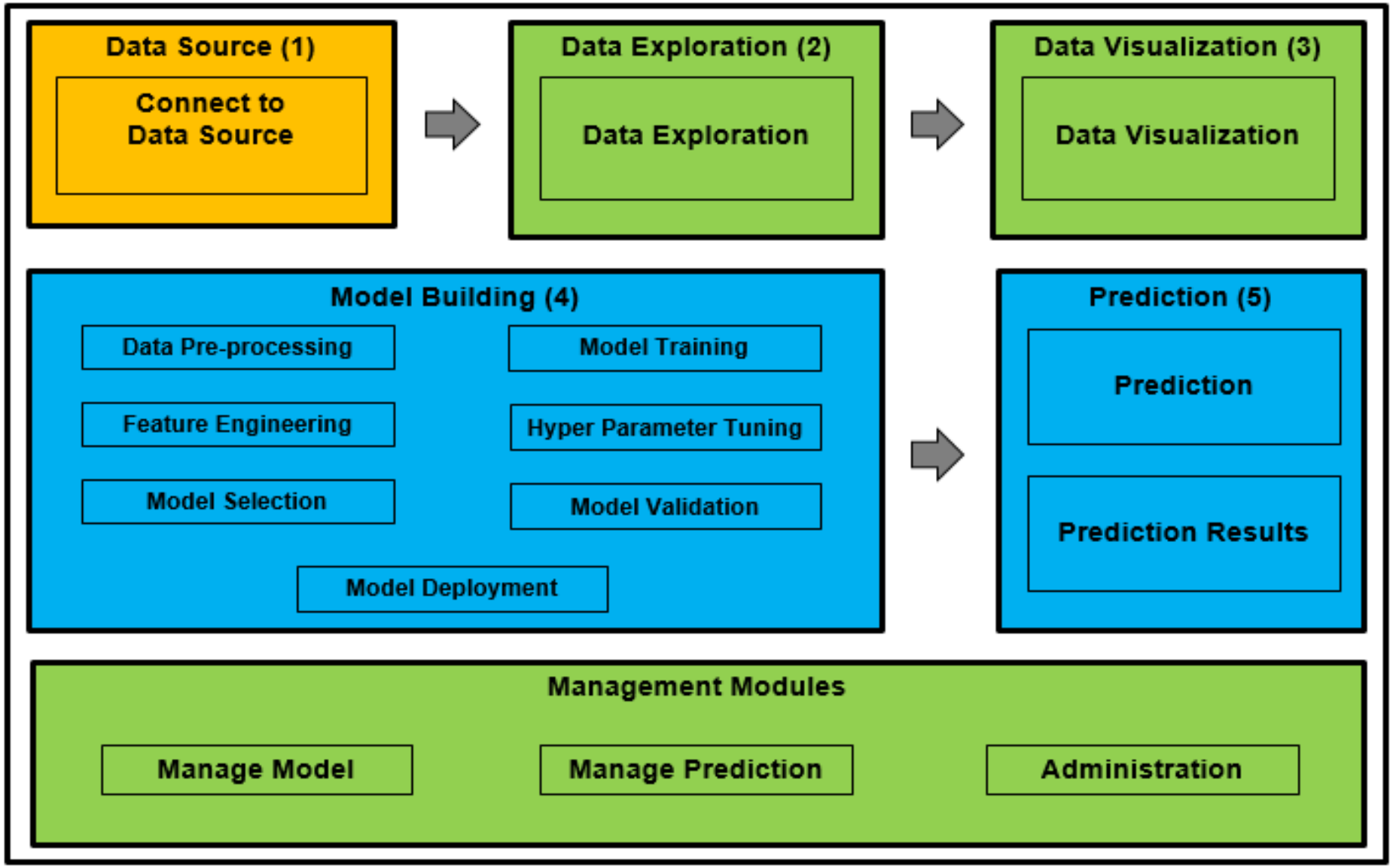


Machine Learning Problems

Classification Regression Time-Series Anomaly Detection

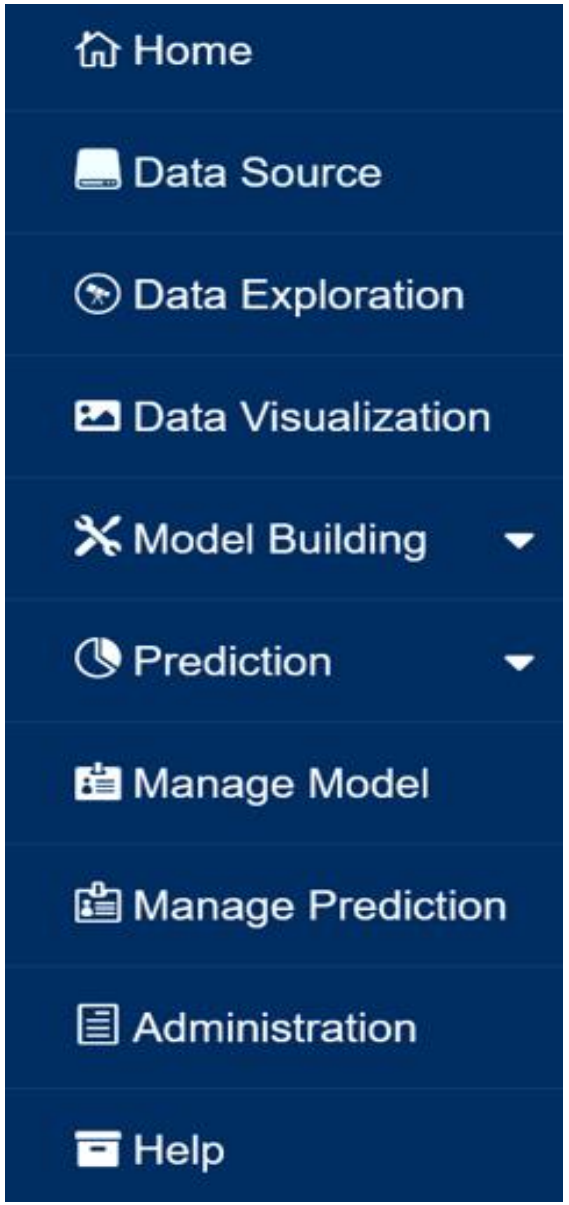
FIU
Applied Research Center

AAML Workflow



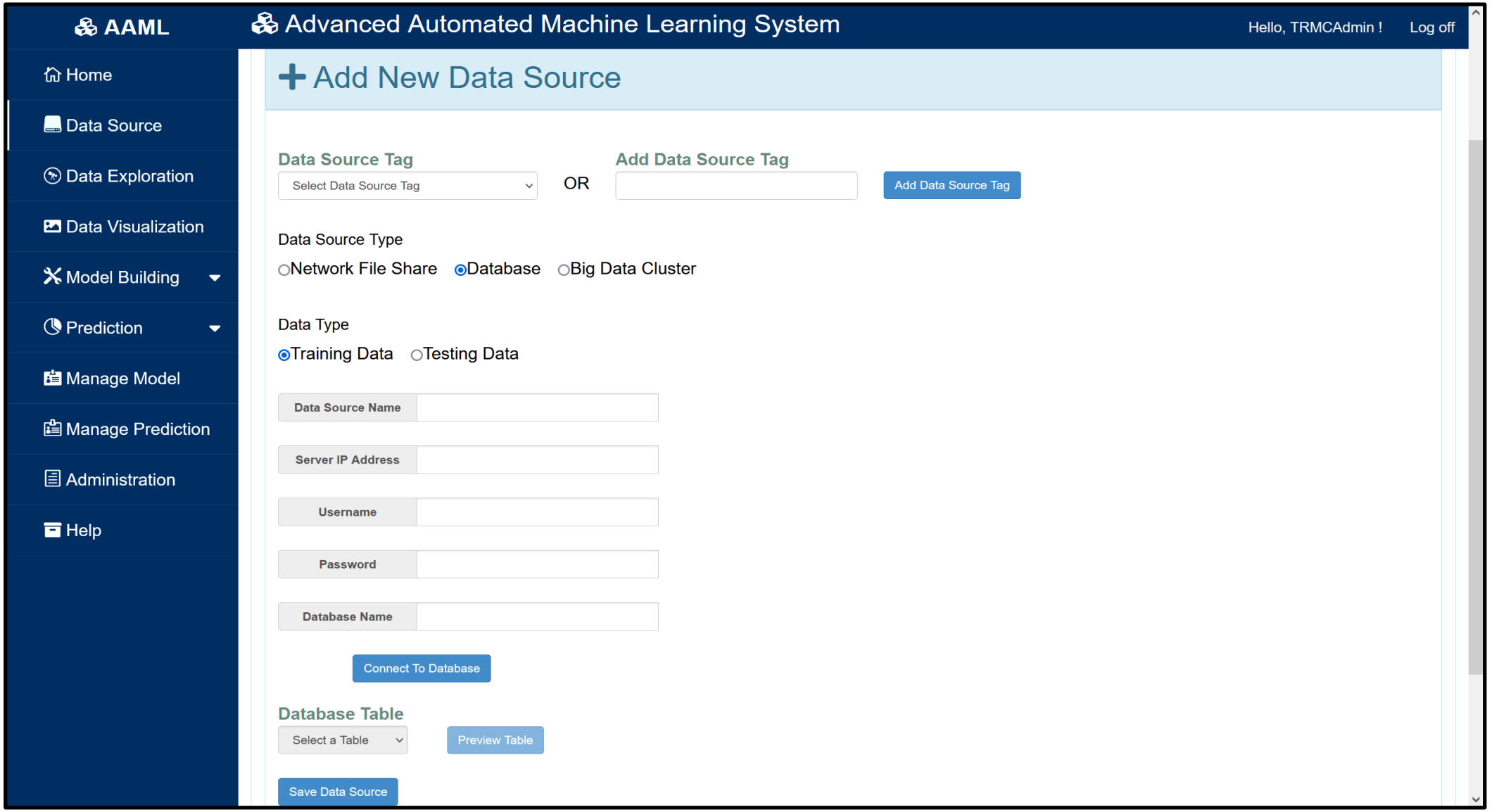
FIU
Applied Research Center

AAML Modules

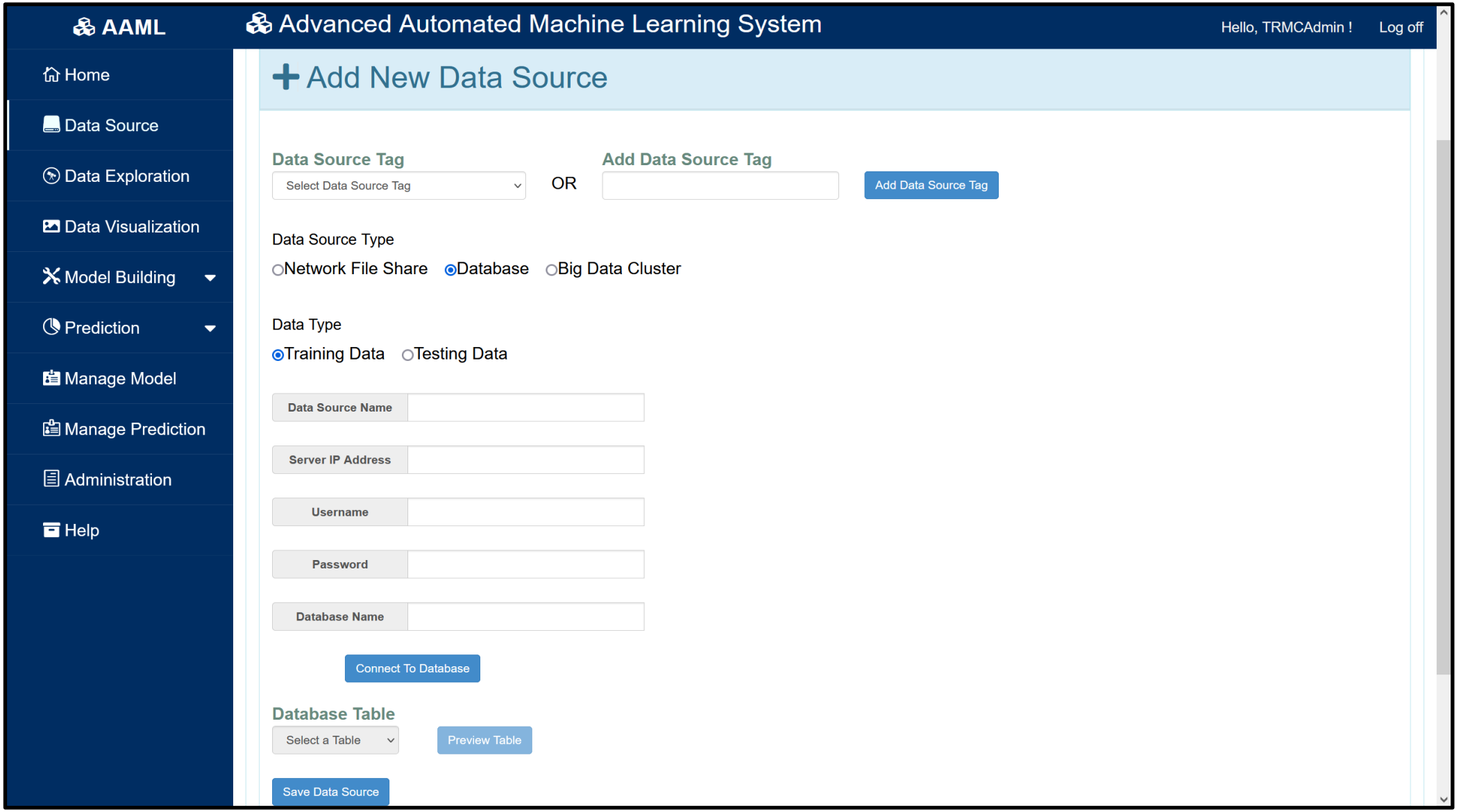


FIU Applied Research Center

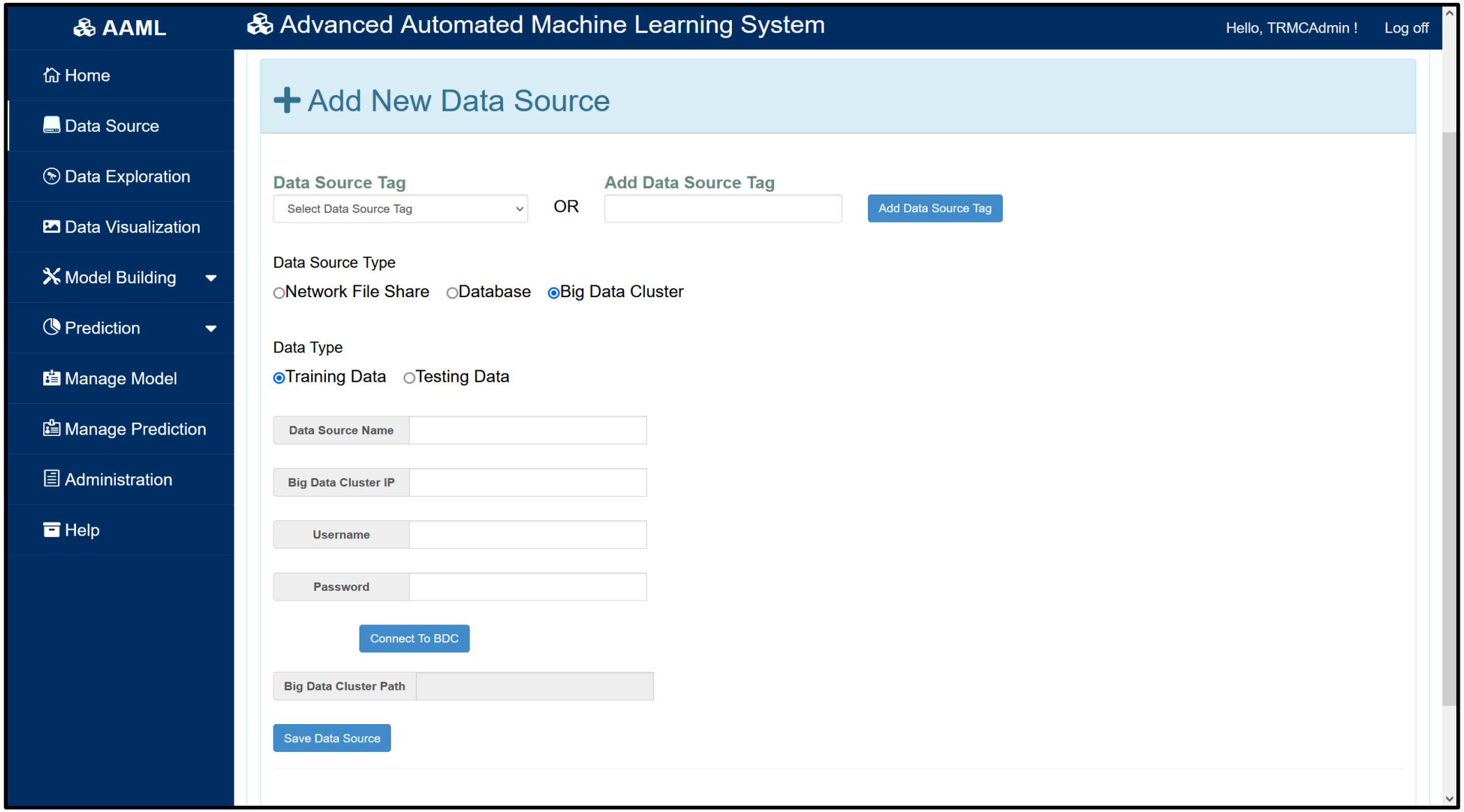
Add Data Source – Network File Share



Add Data Source – Database

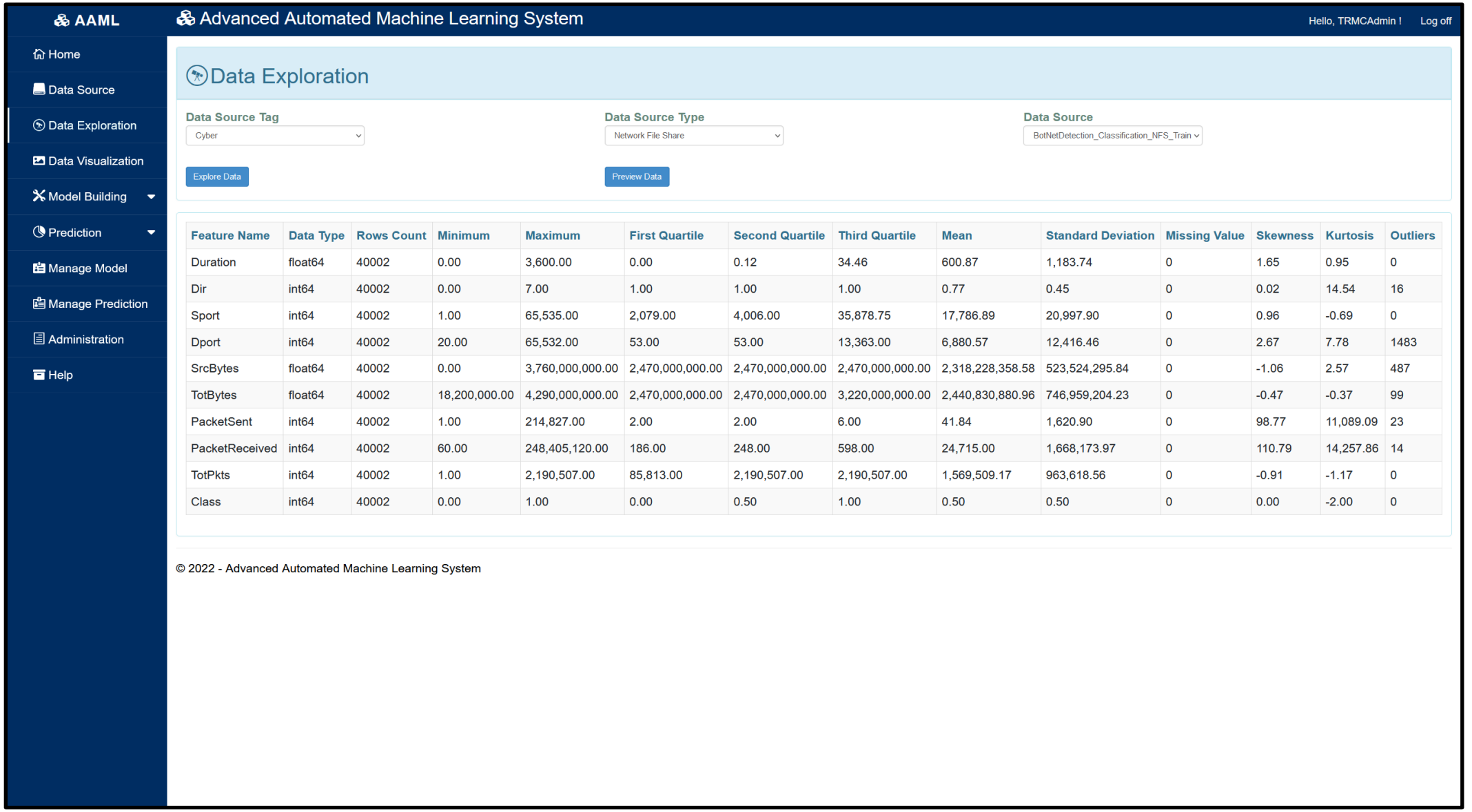


Add Data Source – Big Data Cluster



FIU Applied Research Center

Data Exploration



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

FIU Applied Research Center

Data Visualization - Features

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

Data Visualization

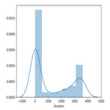
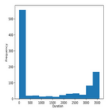
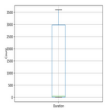
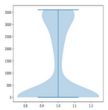
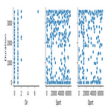
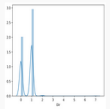
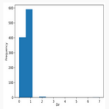
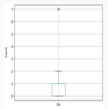
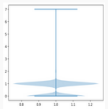
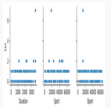
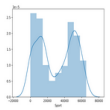
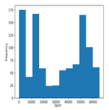
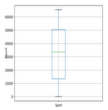
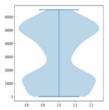
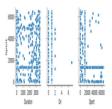
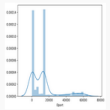
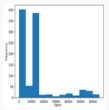
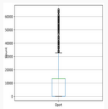
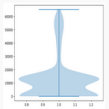
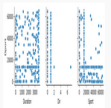
Data Source Tag: Cyber

Data Source Type: Network File Share

Data Source: BotNetDetection_Classification_NFS_Trair

Visualization Type: (•) Features ( ) Dataset

Visualize Data | Preview Data

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

Data Visualization - Dataset

FIU Applied Research Center

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

Data Visualization

Data Source Tag: Cyber

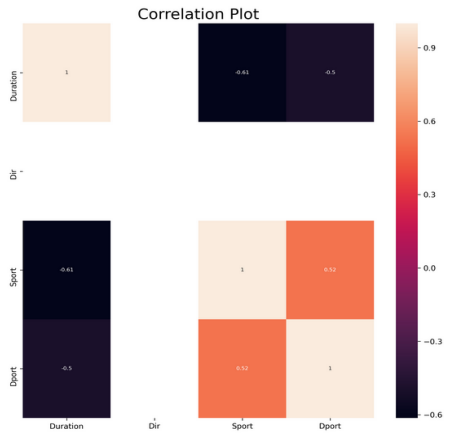
Data Source Type: Network File Share

Data Source: BotNetDetection_Classification_NFS_Trair

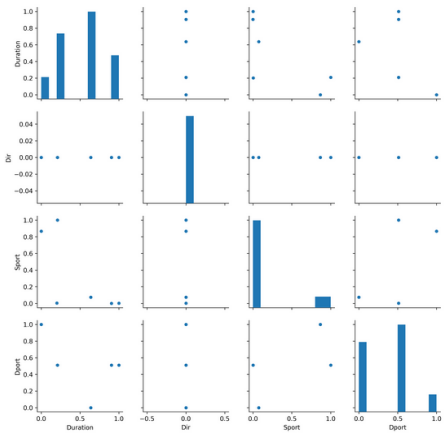
Visualization Type: ○ Features ◉ Dataset

Visualize Data | Preview Data

Correlation Plot

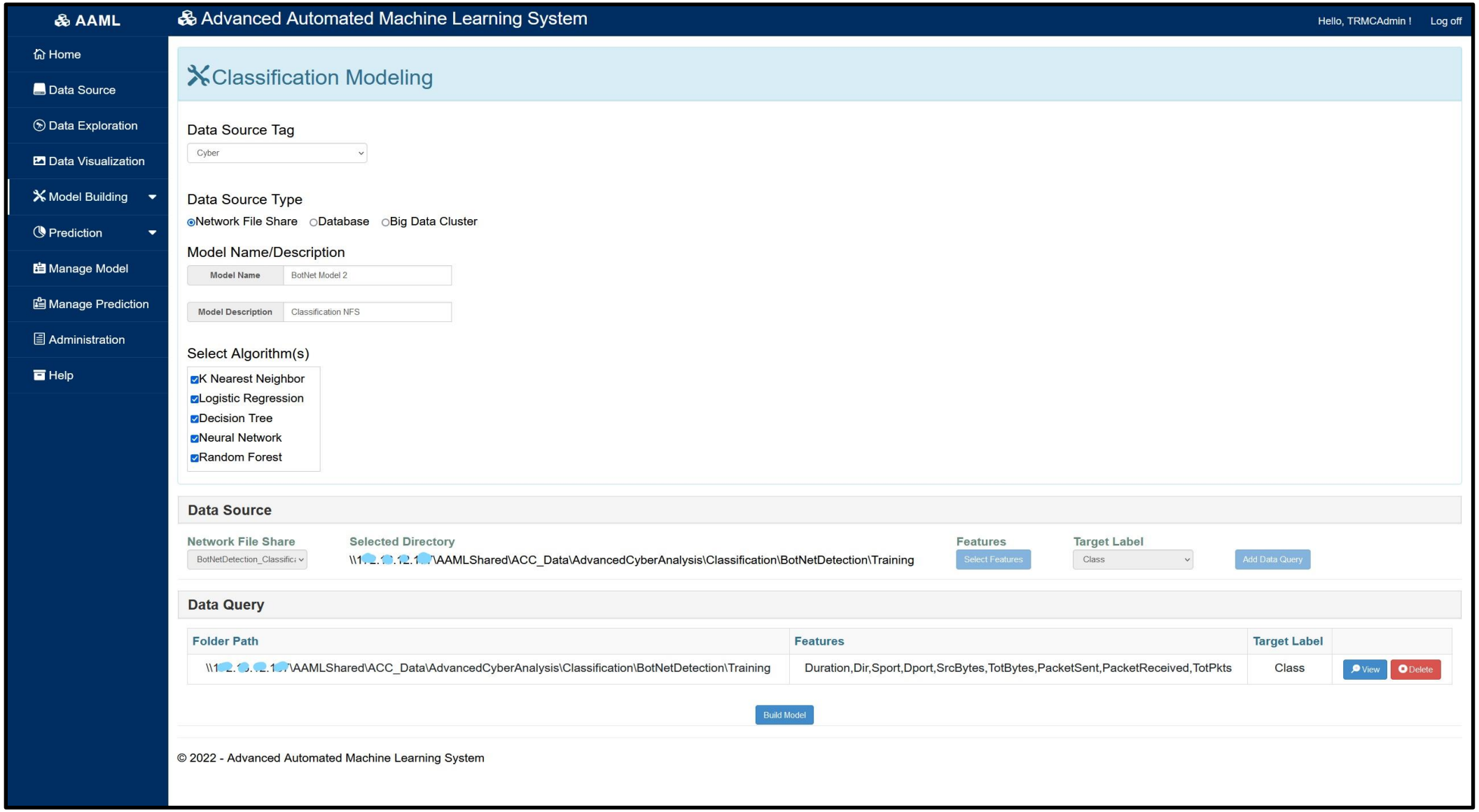


Pair Plot



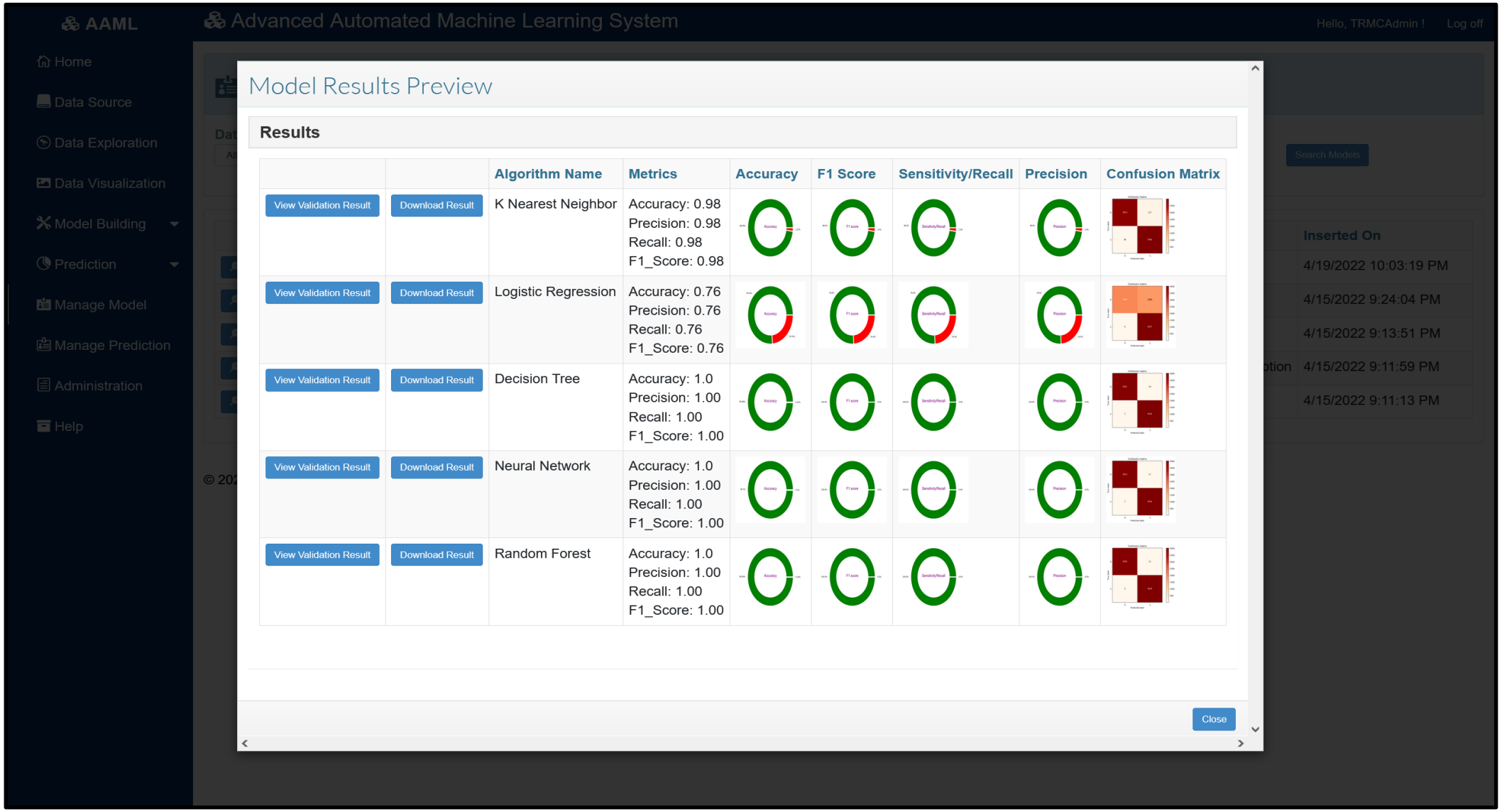
FIU Applied Research Center

Model Building



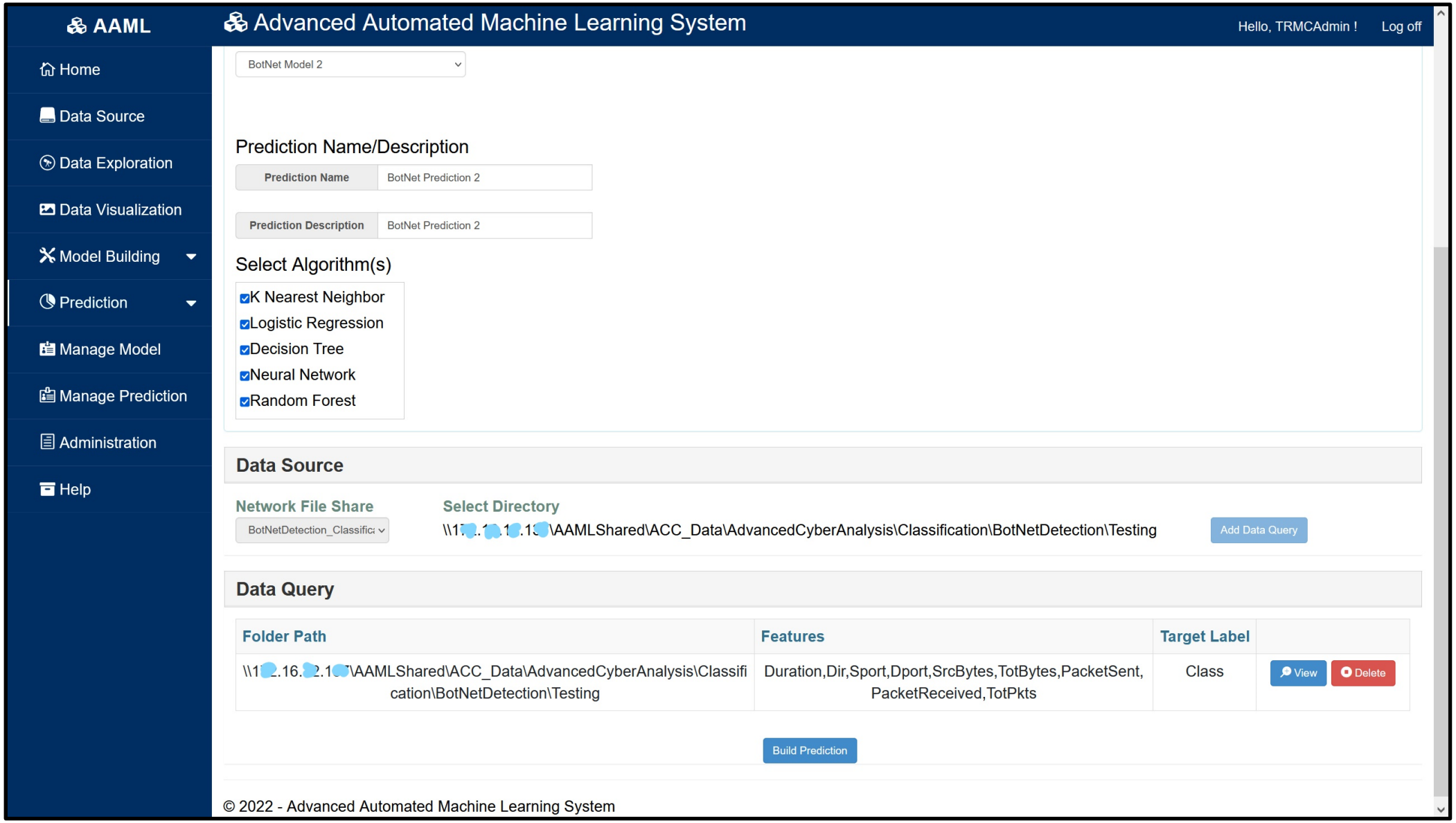
FIU Applied Research Center

Model Results



FIU Applied Research Center

Prediction



FIU Applied Research Center

Prediction Results

AAML

Advanced Automated Machine Learning System

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

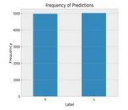
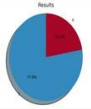
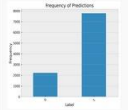
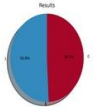
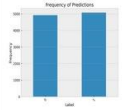
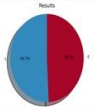
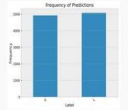
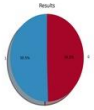
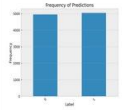
\\1[redacted]\AAMLShared\ACC_Data\AdvancedCyberAnalysis\Classification\BotNetDetection\Testing

Duration,Dir,Sport,Dport,SrcBytes,TotBy[redacted]PacketReceived,TotPkts

BotNet Prediction 2_5_Random Forest.csv
Completed — 770 KB
Show all downloads

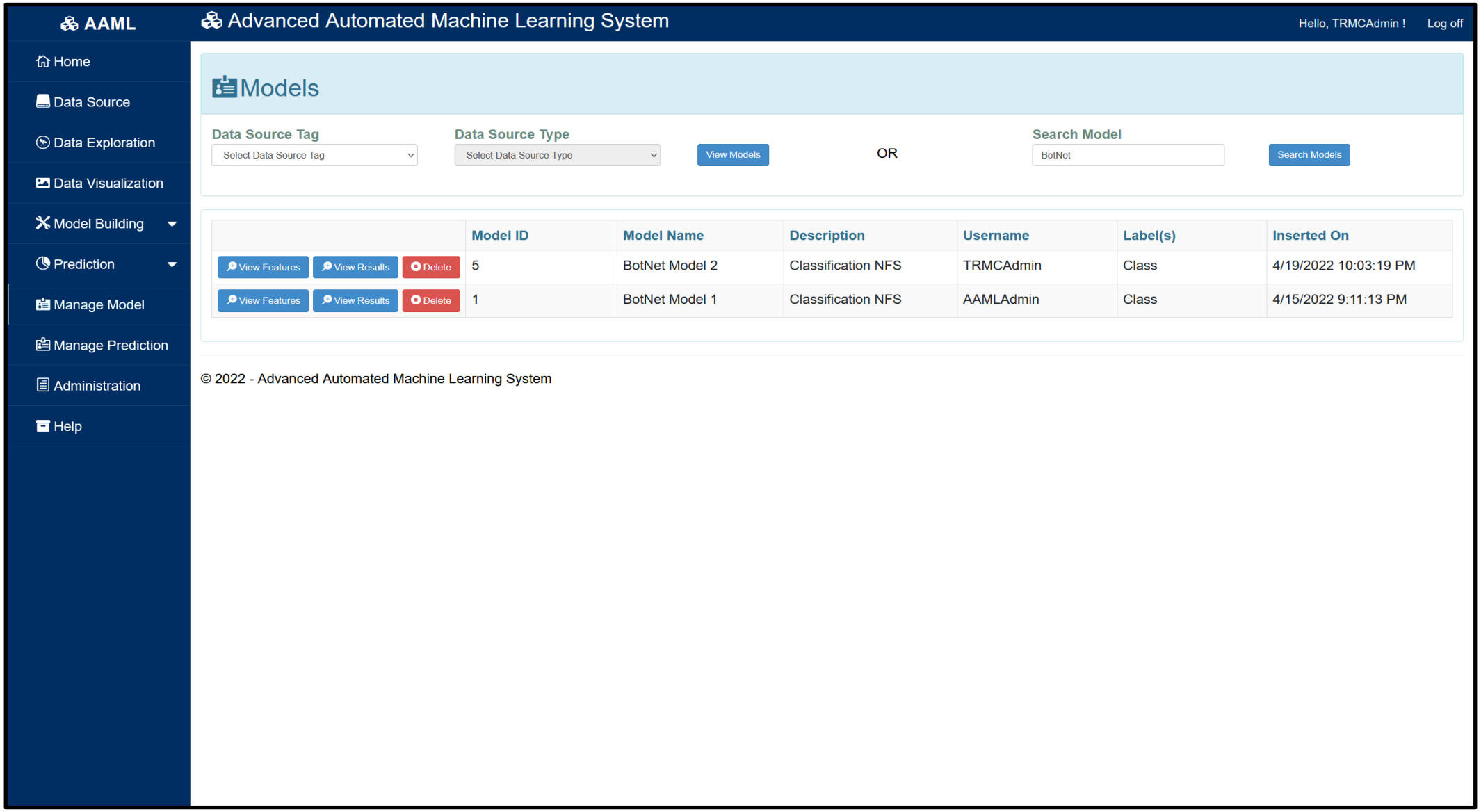
Build Prediction

Results

		Algorithm Name	Prediction Distribution	Frequency
View Prediction Result	Download Result	K Nearest Neighbor		
View Prediction Result	Download Result	Logistic Regression		
View Prediction Result	Download Result	Decision Tree		
View Prediction Result	Download Result	Neural Network		
View Prediction Result	Download Result	Random Forest		

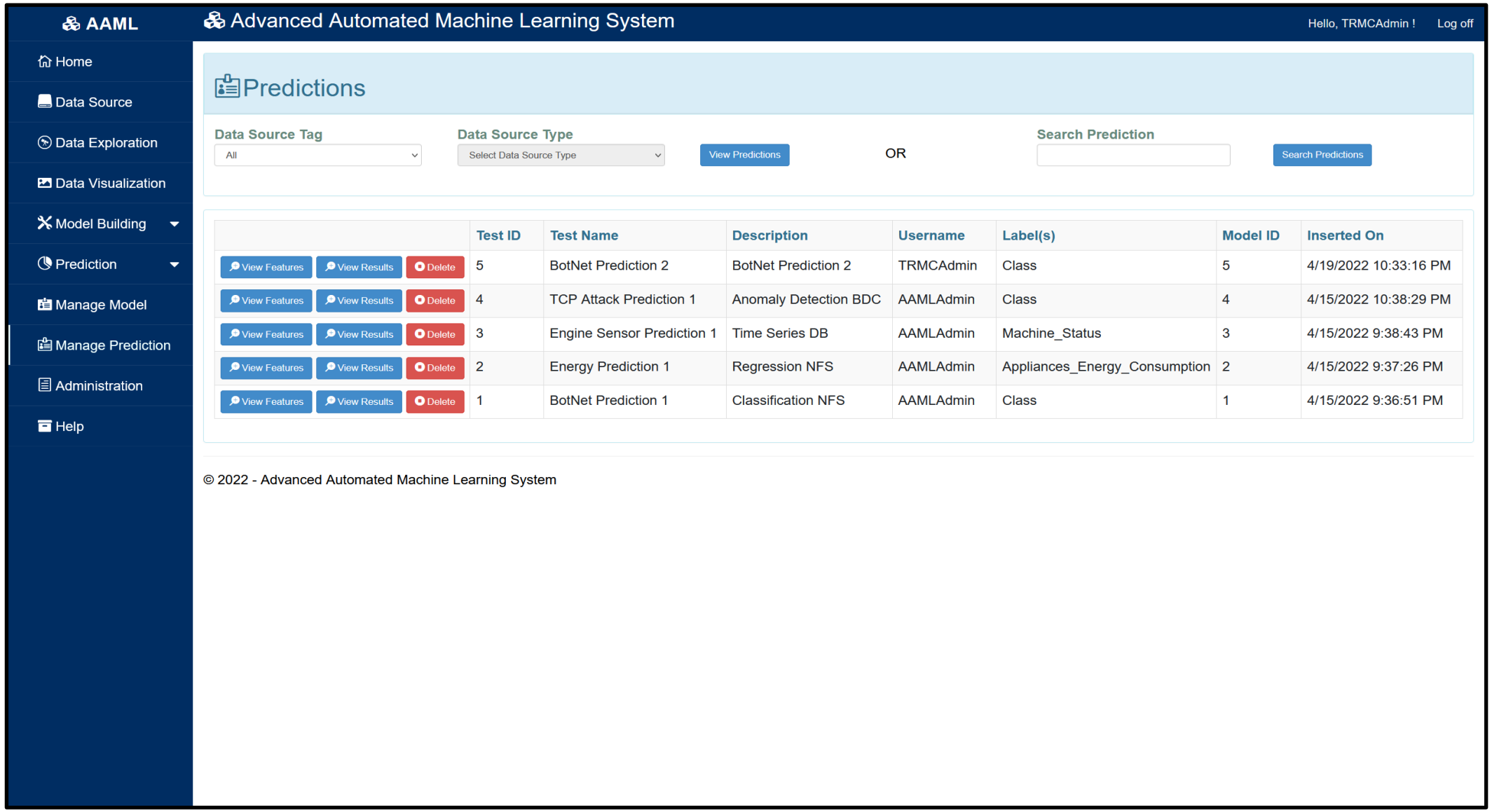
FIU Applied Research Center

Manage Model



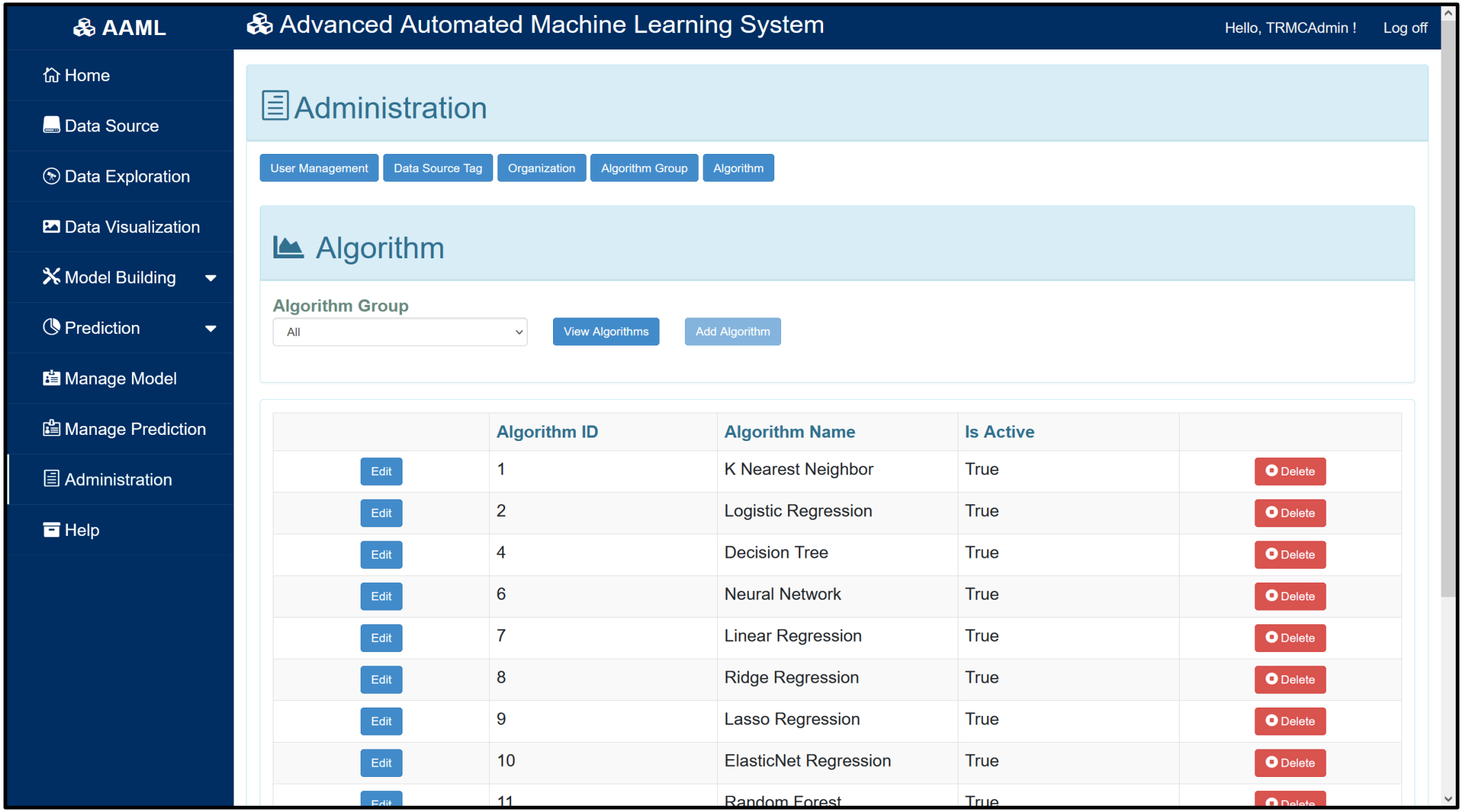
FIU Applied Research Center

Manage Prediction



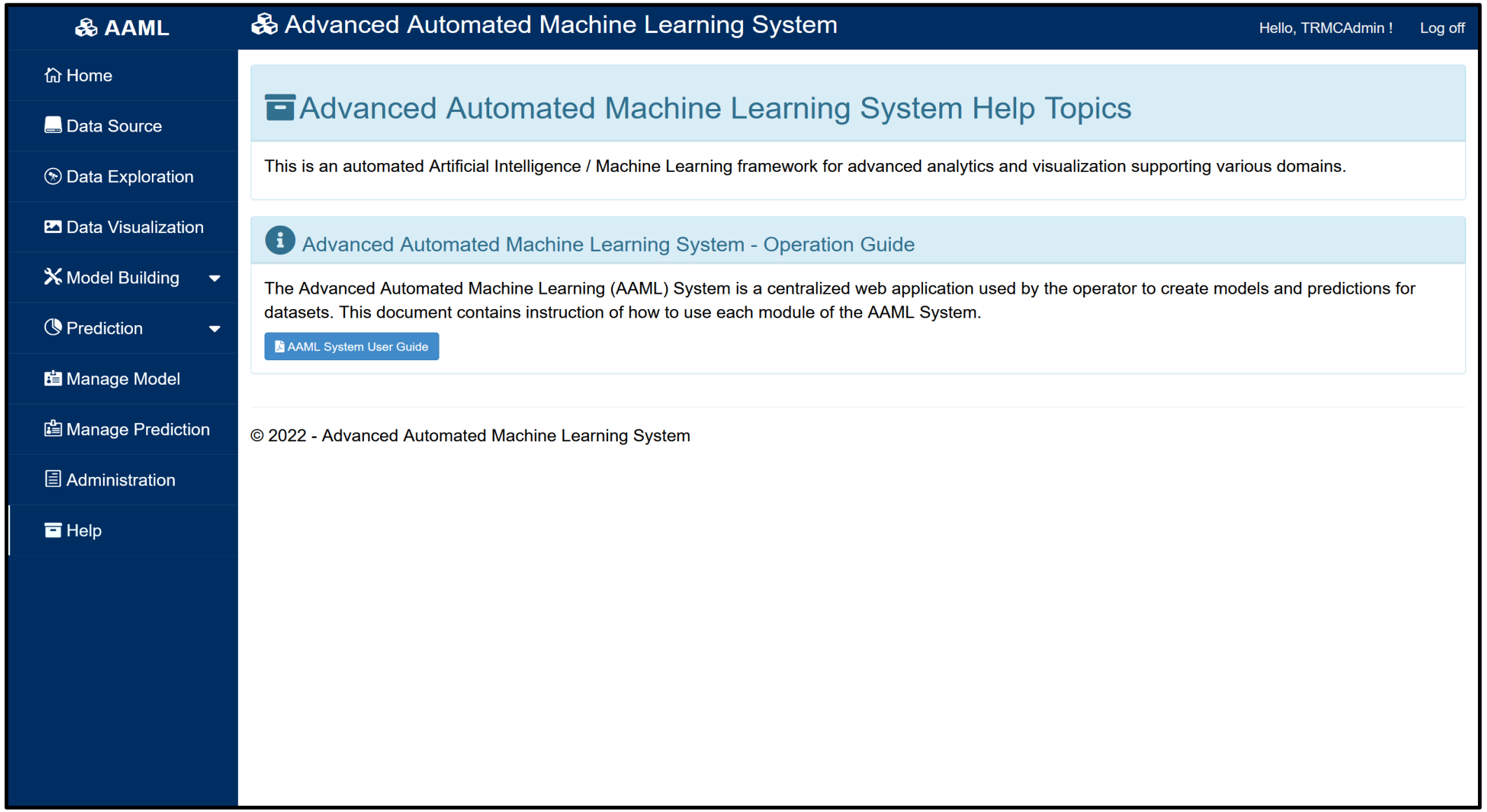
FIU Applied Research Center

Administration



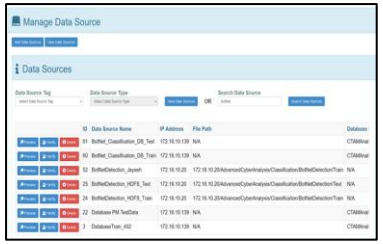
FIU
Applied Research Center

AAML Help

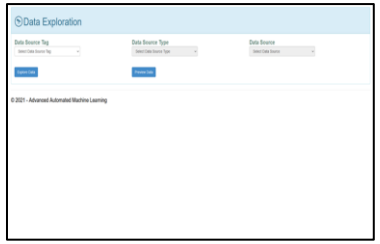


AAML System Recap

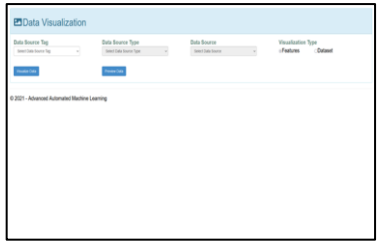
Manage Data Source (1)



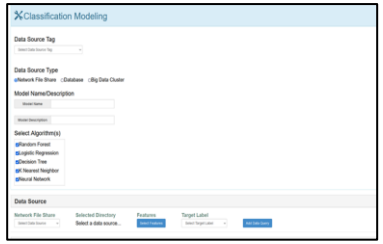
Data Exploration (2)



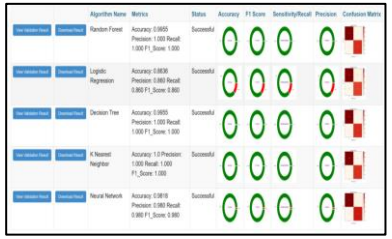
Data Visualization (3)



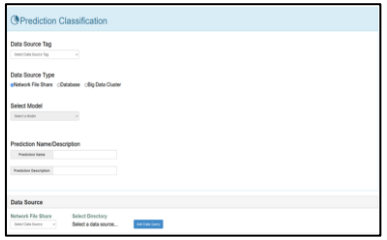
Model Building (4)



Model Validation (5)



Prediction (6)



Prediction Results (7)

FIU Applied Research Center

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

FIU Applied Research Center

Artificial Intelligence Support to DOE EM

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

FIU Applied Research Center

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

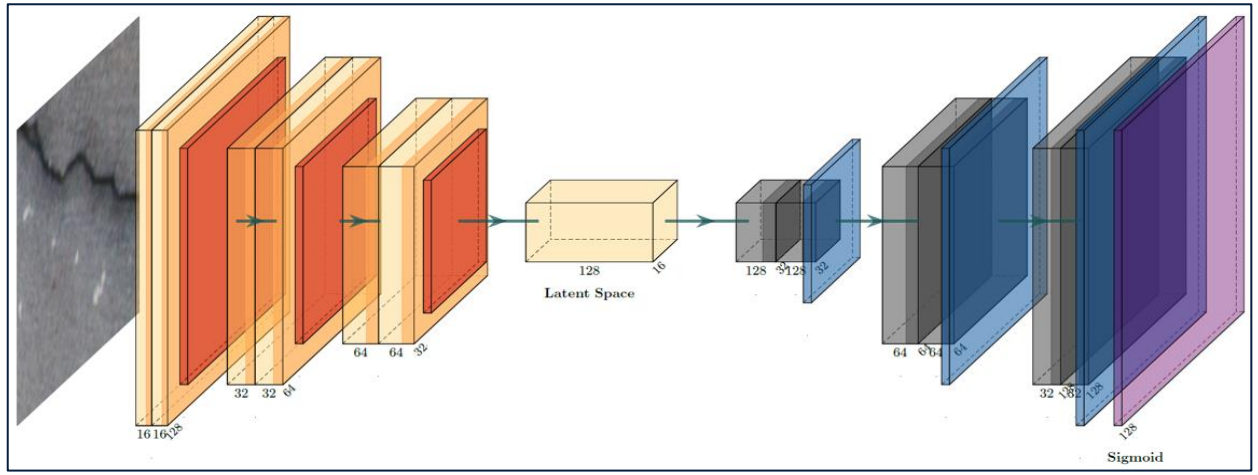
Site Needs:

- 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.

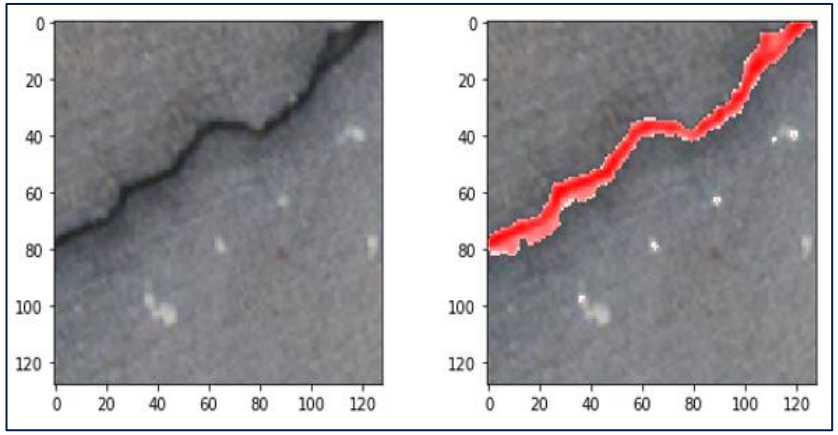
Objectives:

- 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



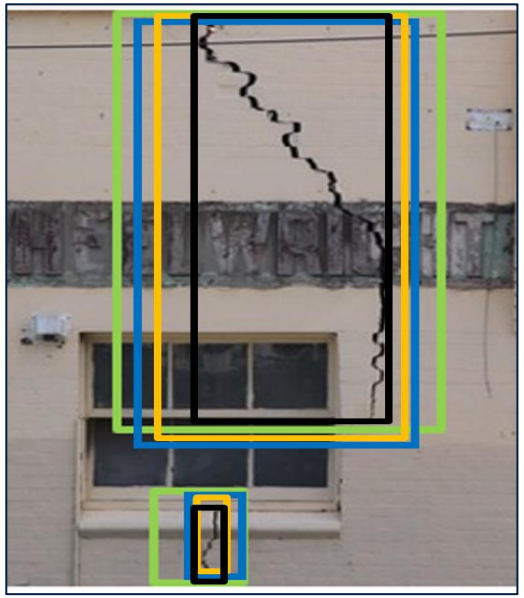
Crack Detection and Heat map

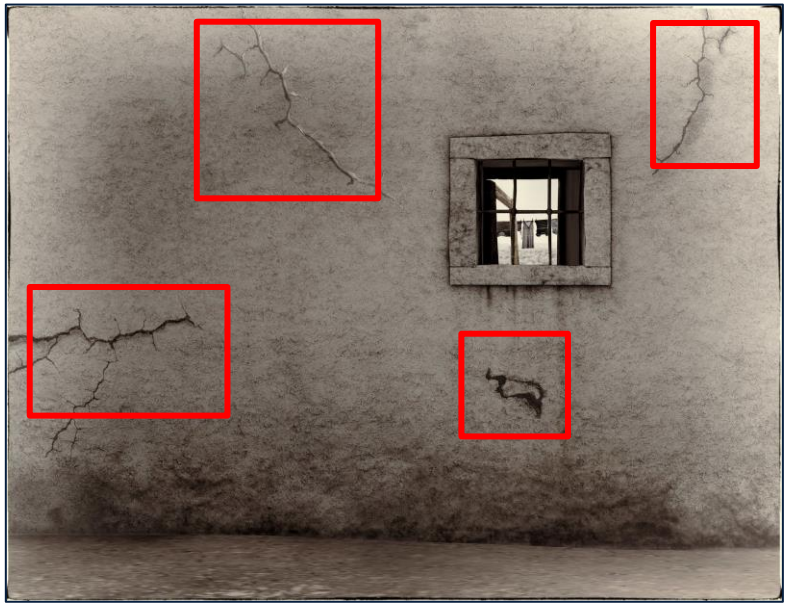
Black
Ground truth
- Manually done.
- Used for accuracy.

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

FIU Applied Research Center

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

Site Needs:

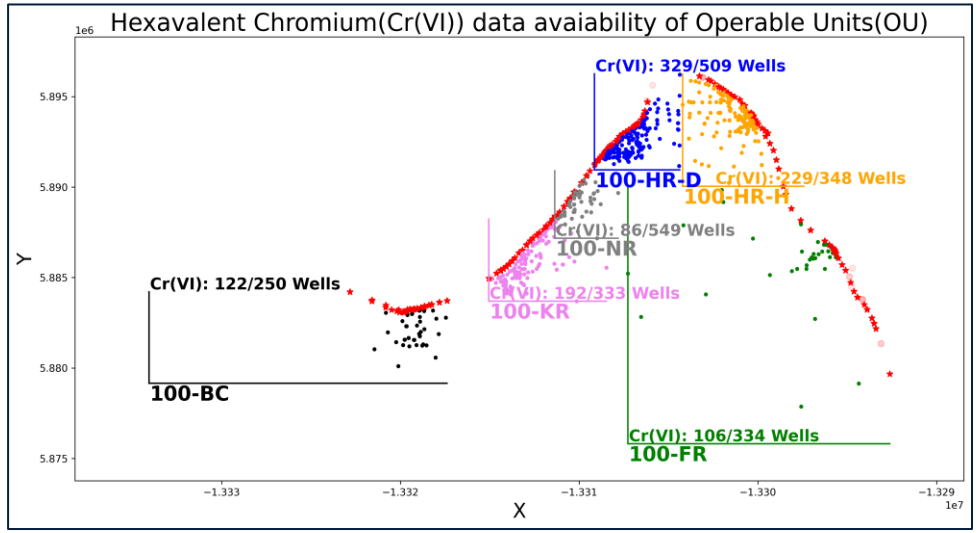
- To identify temporal and spatial relationships of subsurface chromium transport that reduces uncertainties in the conceptual site model (CSM).
- Data analysis & prediction using Artificial Intelligence and Machine Learning algorithms.

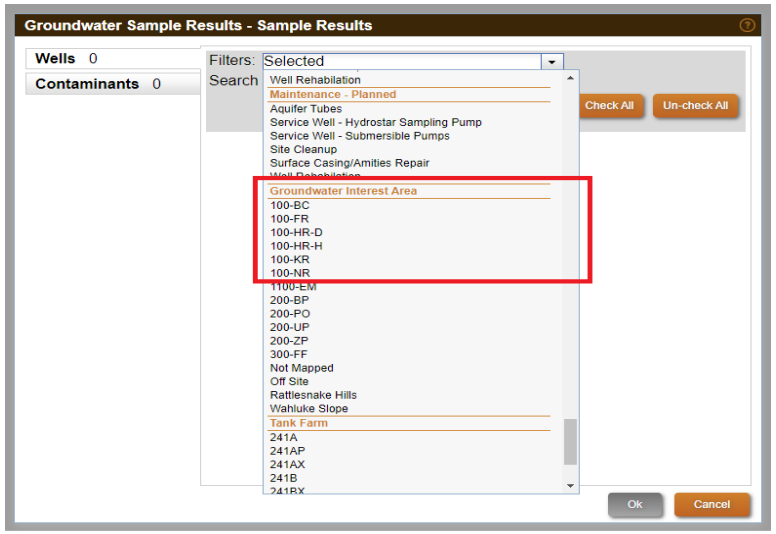
Objectives:

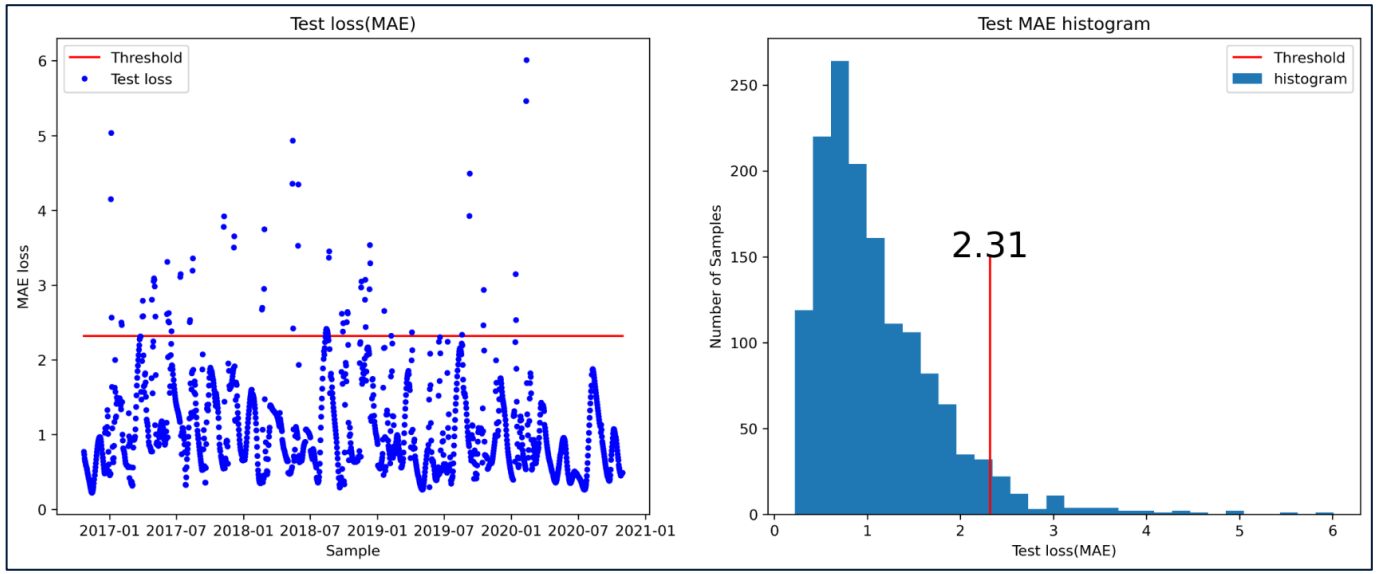
- Data source identification and pre-processing for data cleansing, discretization, and transformation.
- Perform exploratory data analysis using state-of-art statistical methods and various machine learning algorithms.
- Develop AI/ML models to explore spatiotemporal relationships of subsurface hexavalent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

FIU Applied Research Center

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

FIU Applied Research Center

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

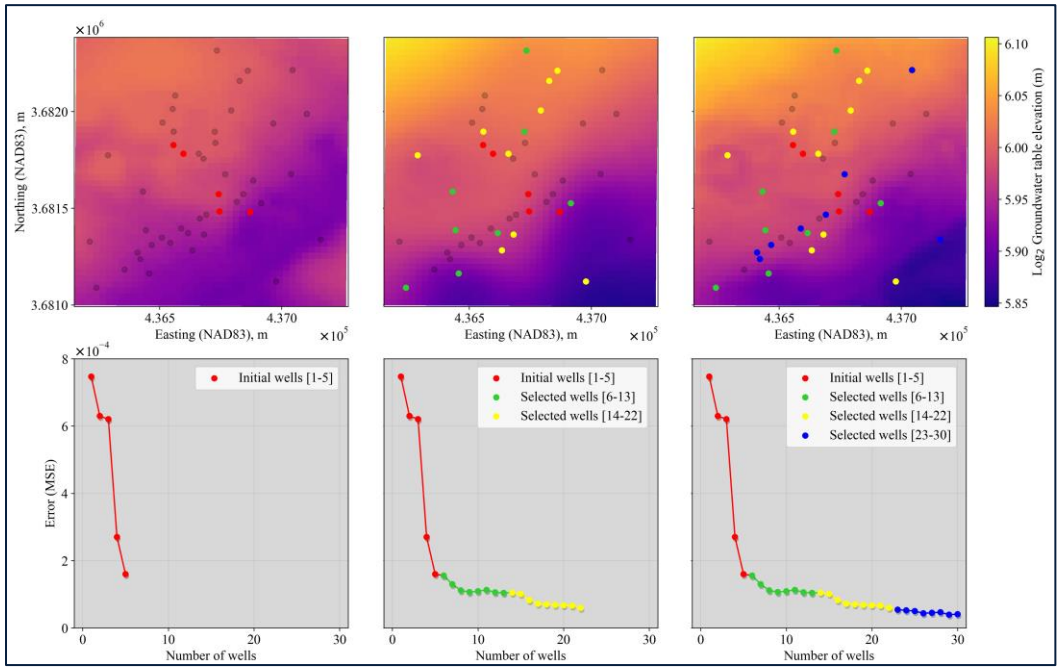
Site Needs:

- 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.

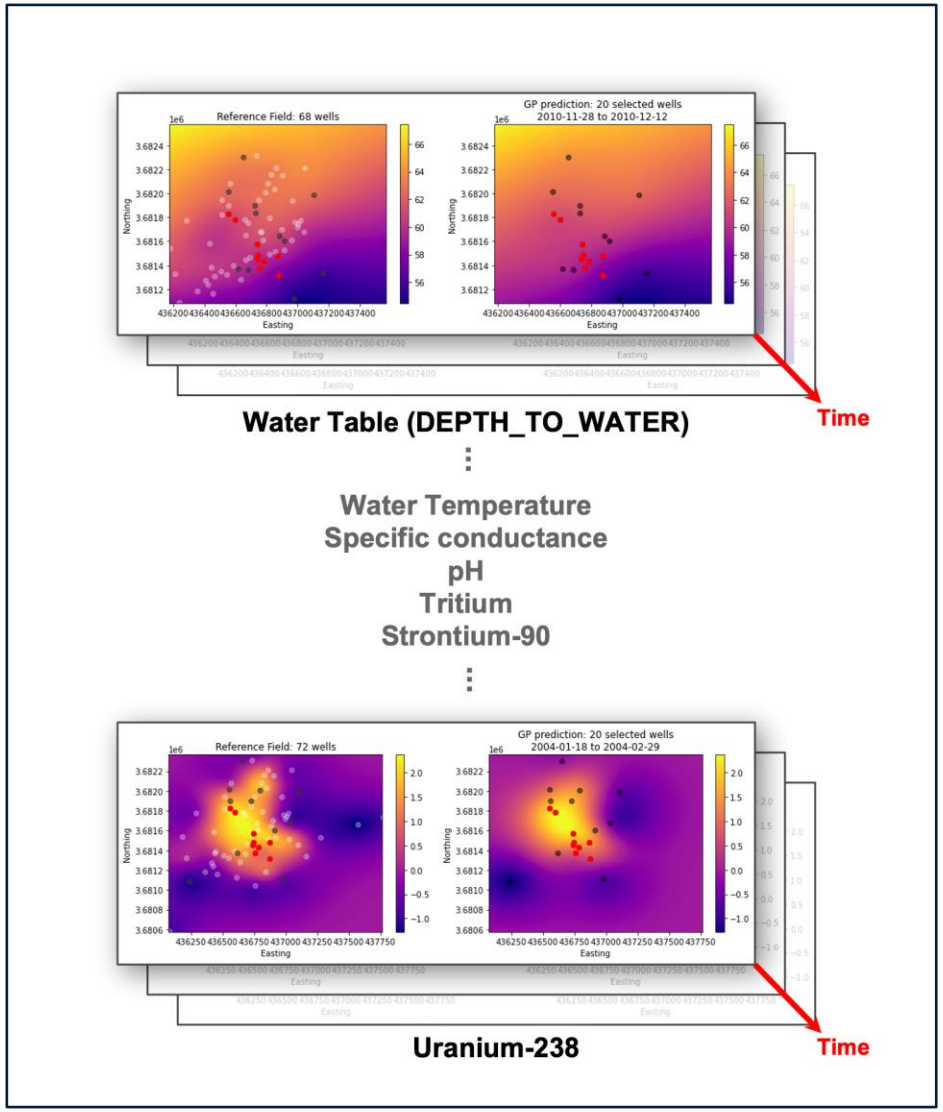
Objectives:

- 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)

FIU Applied Research Center

Advanced Automated Machine Learning System Demonstration

Thank You. Questions?