National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

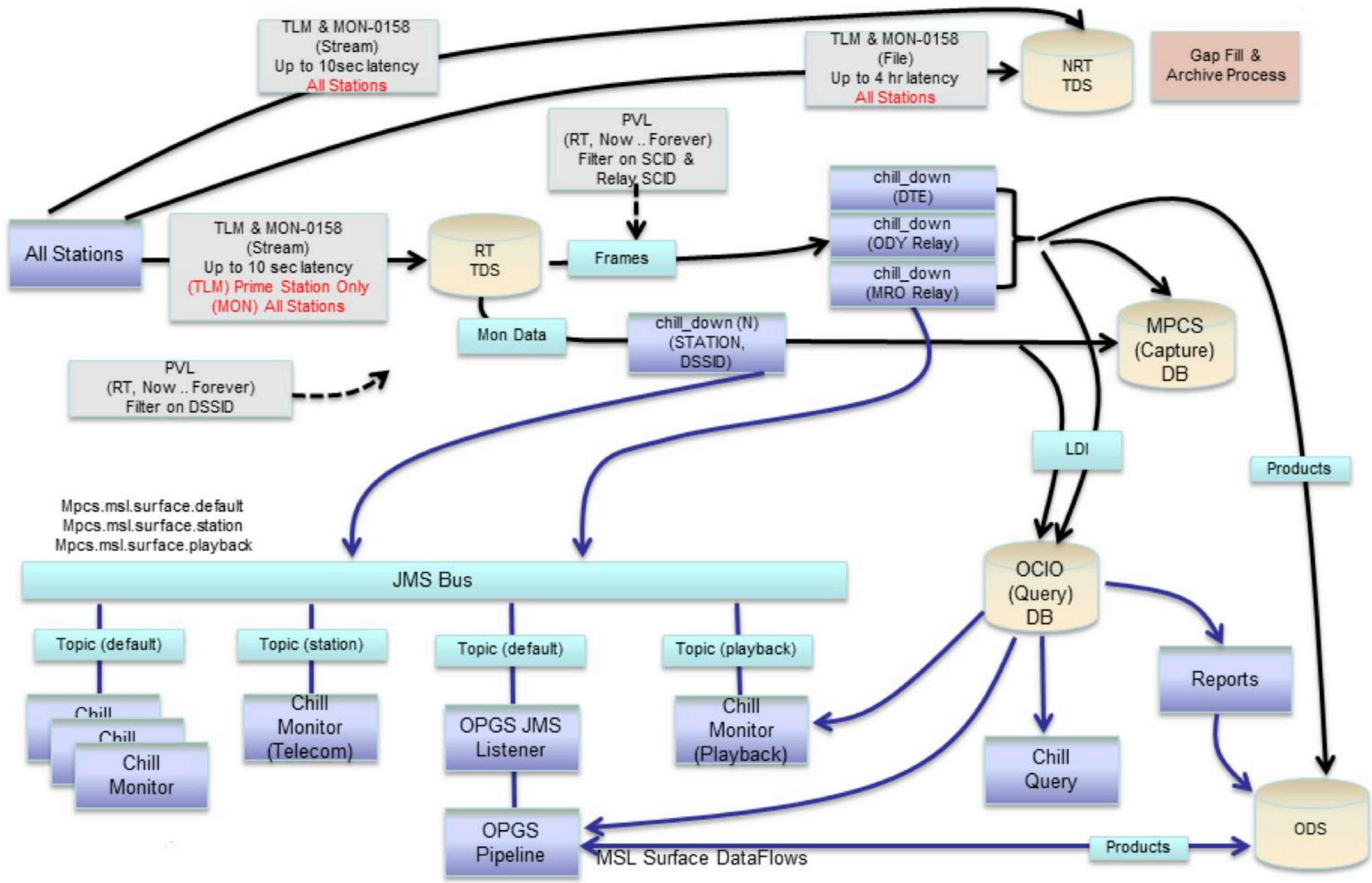
Automated Data Accountability for Mars Science Lab

March 4, 2020

Brian Kahovec
Dr. S. Ryan Alimo
Dr. Dariush Divsalar
Jet Propulsion Laboratory, California Institute of Technology

Overview of the MSL Ground Data System

- The MSL Ground Data System is complex

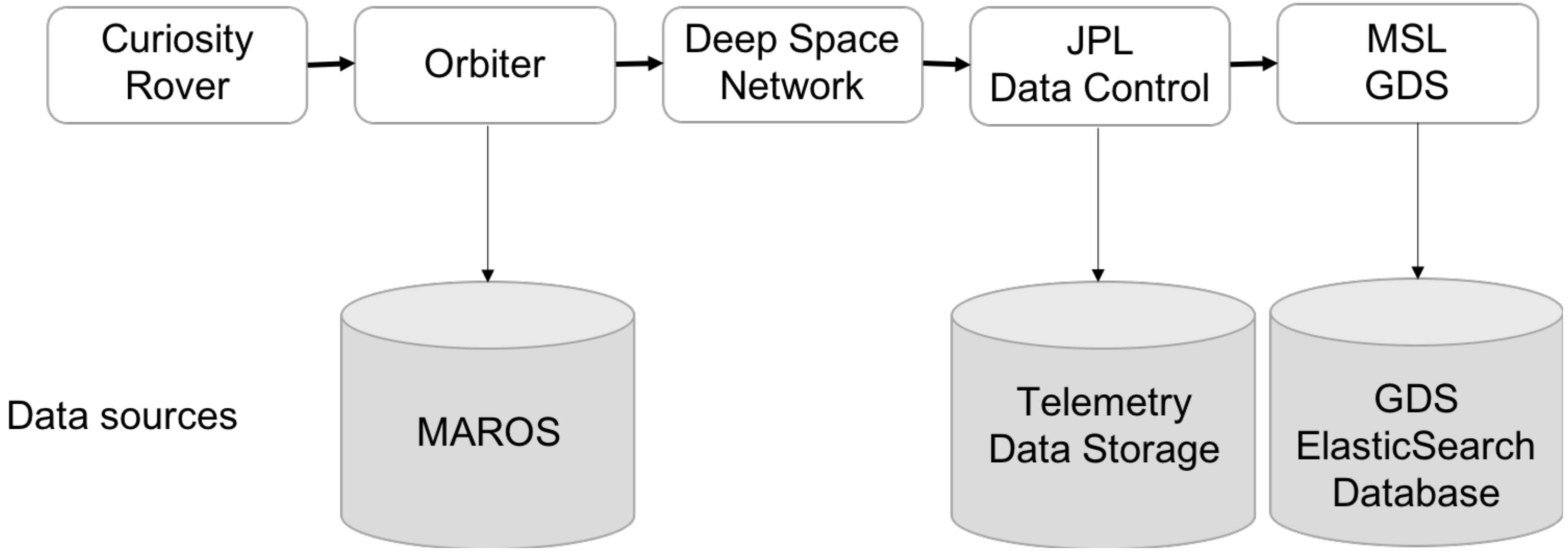


National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Overview of the MSL Downlink Process

- We simplified the problem and identified where data is available



Approach

- Data Collector gathers data from various APIs
- Signal Processor formats the data and computes features
- Machine Learning Algorithms train on historical data
  - Classify each downlink as complete or incomplete
  - Detect anomalies in real-time data

Dataset Description

- Examples of types of data available
  - Predicted Data Volume of the Downlink
  - Actual Data Volume at each step in Downlink Process
  - Predicted start and end time of the Downlink
  - Timestamp received at each step in Downlink Process
  - The orbiter used to transmit the data
  - Elevation of the orbiter
  - The DSN station that received the data
  - Number of in-sync frames
  - Number of out-of-sync frames

Automated Feature Analysis

- Three Different Methods
  - P-score
  - Variance
  - Random Forest
- Important Features
  - Differences in Data Volume
  - Difference between Actual and Predicted Start Time
  - Out-of-sync frames
- Non-important Features
  - Orbiter ID
  - Deep Space Station ID

Current Techniques

- GDSA Dashboard labels passes as complete or incomplete
  - No longer reliable in operations

Sol 2433					
34330	MRO_MSL_2019_161_04	Complete		401.112	344.419
34331	MRO_MSL_2019_162_01	Complete		357.719	403.643
34331	MRO_MSL_2019_162_01	Incomplete	Wrong ERT Times	357.809	403.643
44330	TGO_MSL_2019_162_01	Complete		236.015	58.931

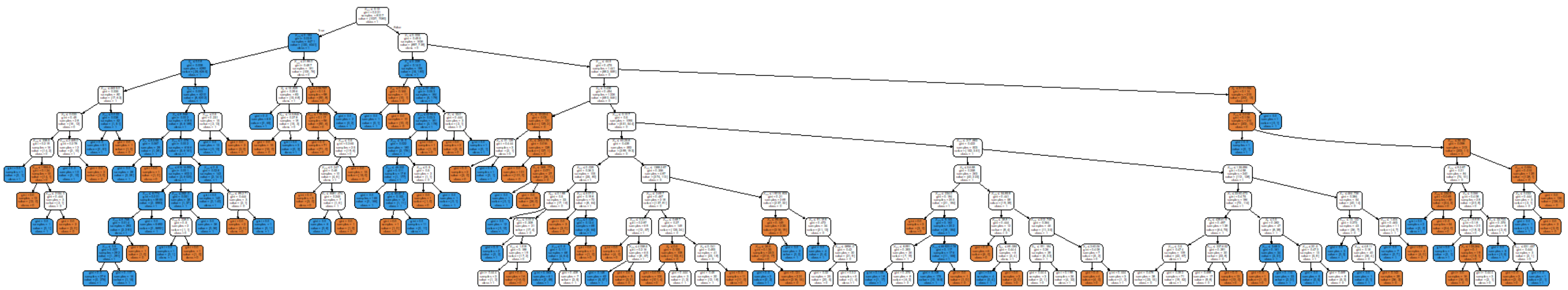
  - Overall Accuracy: 91.9%

	Precision	Recall	f1-score	Support
0	0.74	0.55	0.63	1141
1	0.94	0.97	0.95	7867
Avg / Total	0.91	0.92	0.91	9008

		Dashboard	
		0	1
Actual	0	625	516
Actual	1	218	7649

Results

- Random Forest Classifier
  - Image of one Decision Tree



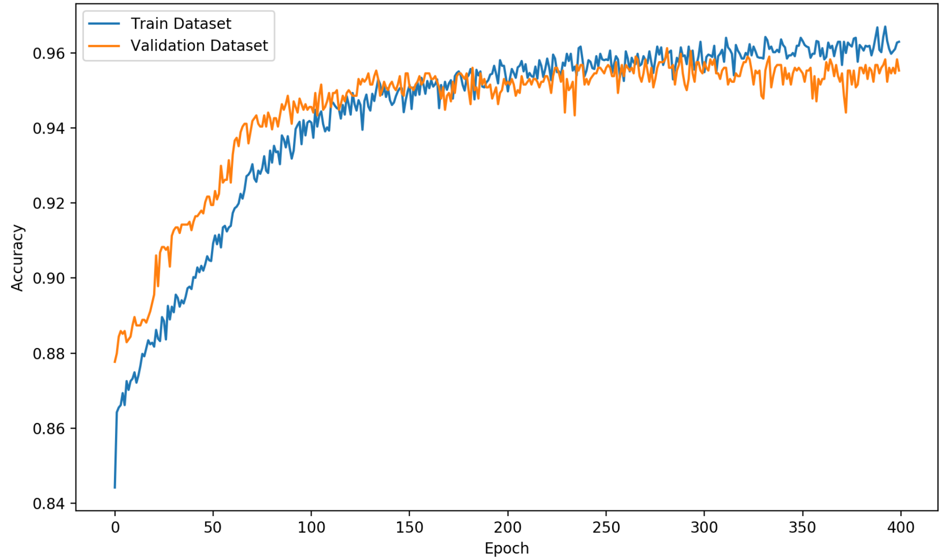
  - Overall Accuracy: 98.3%

	Precision	Recall	f1-score	Support
0	0.94	0.93	0.93	114
1	0.99	0.99	0.99	787
Avg / Total	0.98	0.98	0.98	901

		Decision Tree	
		0	1
Actual	0	106	8
	1	7	780

Results

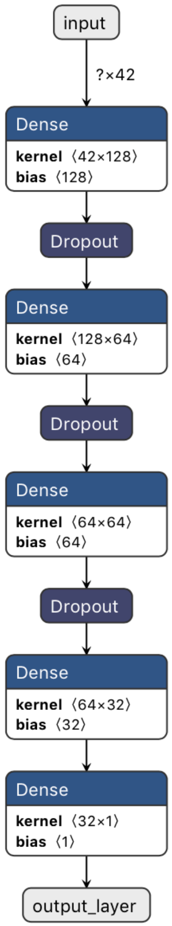
- Deep Neural Network
  - 97% training accuracy, 95.5% validation accuracy



  - Overall Accuracy: 95.1%

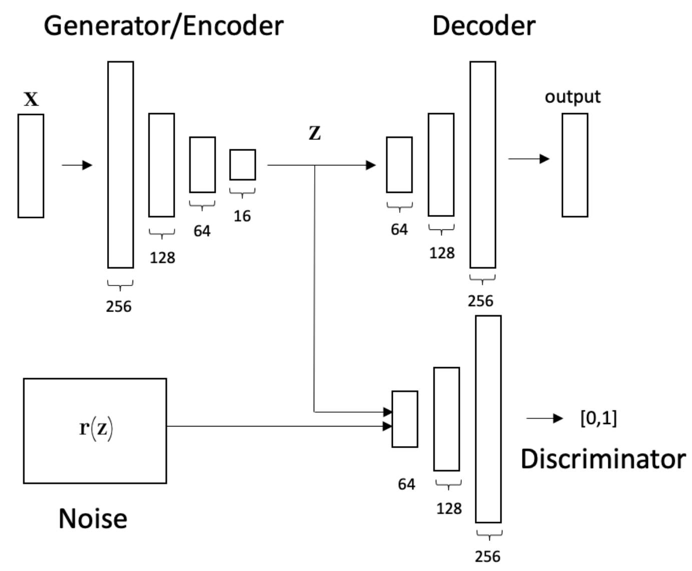
	Precision	Recall	f1-score	Support
0	0.84	0.77	0.80	176
1	0.97	0.98	0.97	1175
Avg / Total	0.98	0.98	0.98	1351

		Decision Tree	
		0	1
Actual	0	135	41
	1	25	1150



Results

- Anomaly Detection
  - Adversarial Autoencoder
    - Imposed Gaussian Distribution on the latent space



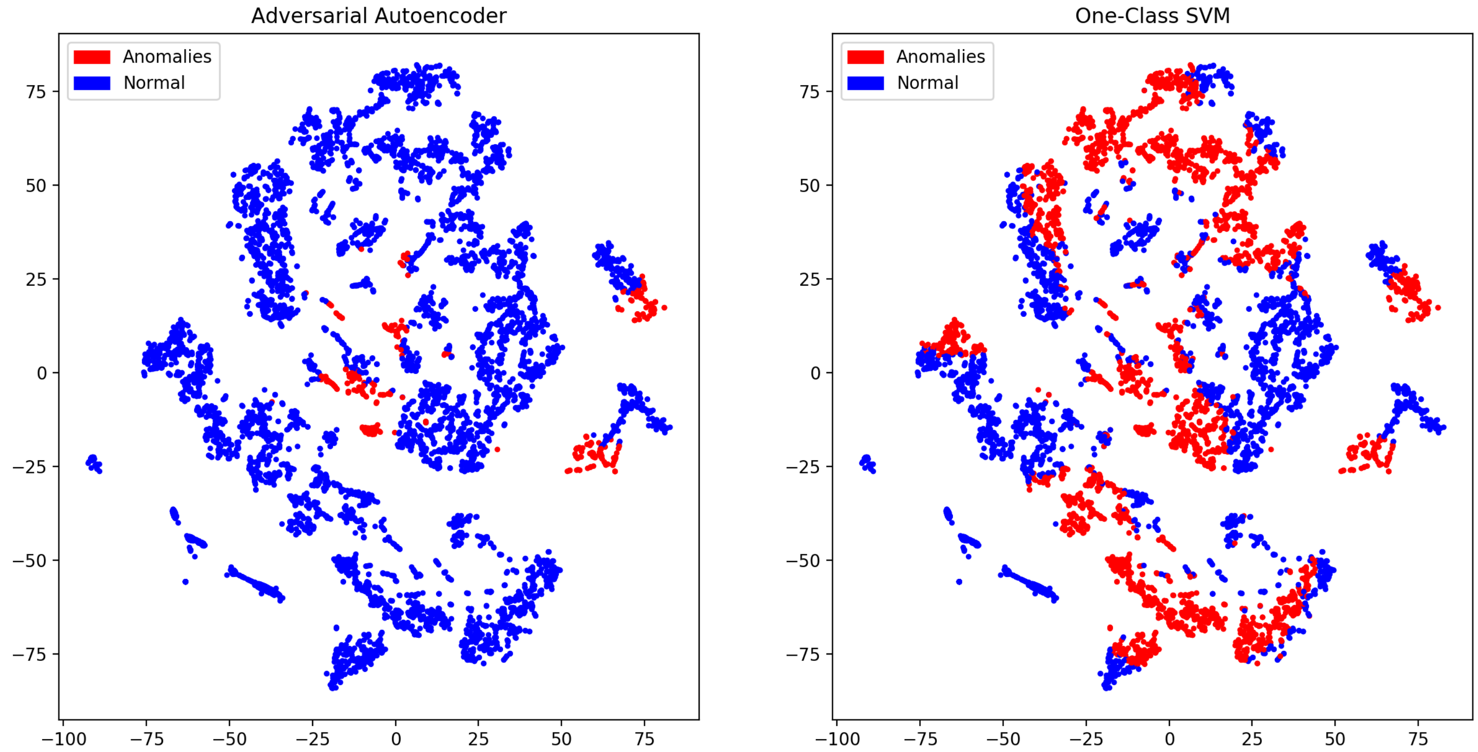
  - One-Class Support Vector Machine (SVM)
    - Sigmoid Kernel yielded the best results

$$k(x, y) = \tanh(\alpha x^T y + c)$$

  - Autoencoders and reconstruction thresholds are not well-suited for our classification problem. Our other methods (NN, Decision Tree) produced more accurate results.

Results

- Anomalies found in the Training Dataset



Results

- Anomalies found in the Testing Dataset