Microsoft

Cortana Analytics Workshop

Sept 10 – 11, 2015 • MSCC

Predictive Maintenance in the IoT Era

Danielle Dean, PhD
Senior Data Scientist Lead

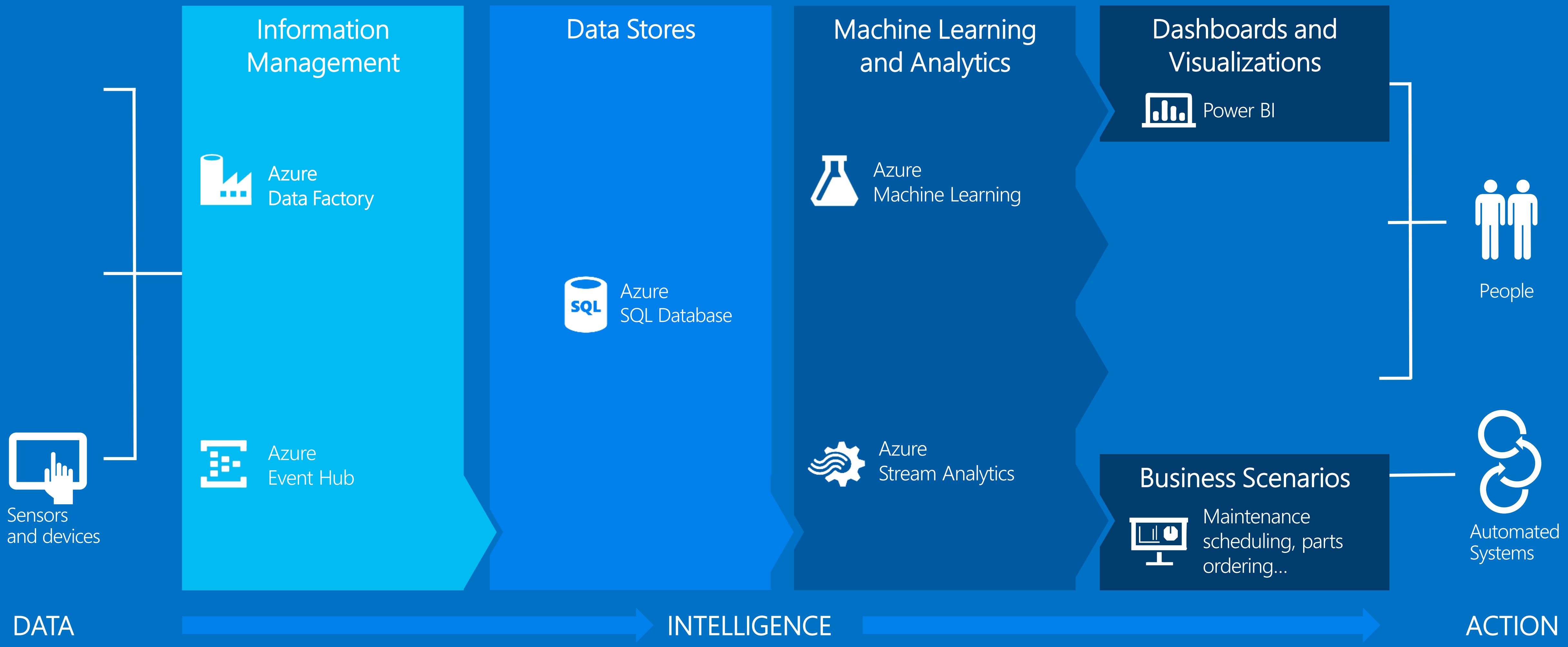
Twitter @danielleodean
aniedea@Microsoft.com

Microsoft

Cortana Analytics Suite

Transform data into intelligent action

Today's Talk


Information Management
Azure Data Factory
Azure Event Hub
Data Stores
Azure SQL Database
Machine Learning and Analytics
Azure Machine Learning
Azure Stream Analytics
Dashboards and Visualizations
Power BI
Business Scenarios
Maintenance scheduling, parts ordering...
Sensors and devices
People
Automated Systems
DATA
INTELLIGENCE
ACTION

Outline

- Predictive Maintenance Use Cases
- Building E2E solution with Cortana Analytics Suite
- Data
- Modeling, Evaluation

Predictive Maintenance Concepts

Important task in Internet of Things applications

Goal: improve production/maintenance efficiency

	Predictive Maintenance in IoT	Traditional Predicative Maintenance
Goal	Improve production and/or maintenance efficiency	Ensure the reliability of machine operation
Data	Data stream (time varying features), Multiple data sources	Very limited time varying features
Scope	Component level, System level	Parts level
Approach	Data driven	Model driven
Tasks	Failure prediction, fault/failure detection & diagnosis, maintenance actions recommendation, etc. Essentially any task that improves production/maintenance efficiency	Failure prediction (prognosis), fault/failure detection & diagnosis (diagnosis)

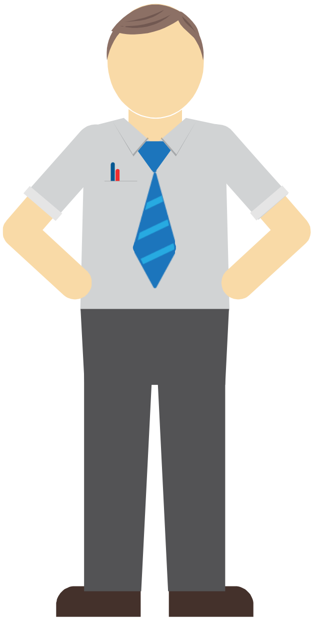
Predictive Maintenance Use Cases

Aerospace	Utilities	Manufacturing	Transportation & Logistics
What is the likelihood of delay due to mechanical issues?	When is my solar panel or wind turbine going to fail next?	Will the component pass the next stage of testing on factory floor or do I need to rework?	Should I replace the break disks in my car or can I wait for another month?
When is this aircraft component likely to fail next?	Which circuit breakers in my system are likely to fail in the next month?	What is the root cause of the test failure?	What maintenance task should I perform on my elevator?
	Is the ATM going to dispense the next 5 notes without failing?		

Aircraft Engine Demo

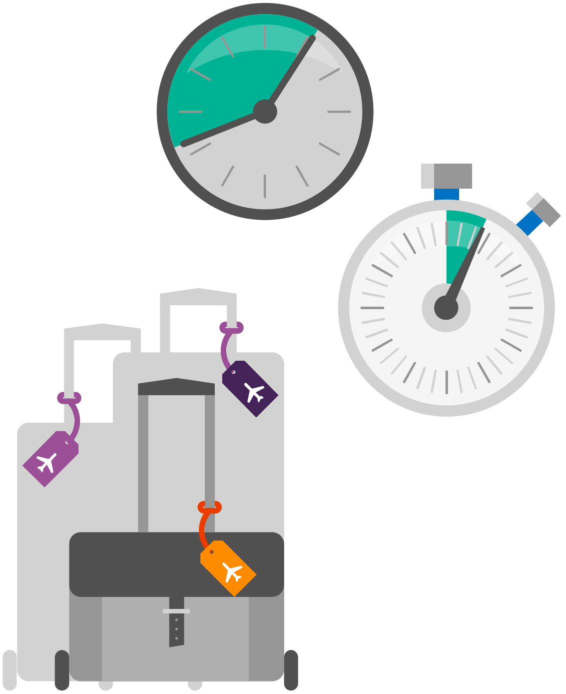
Scenario

This is Kyle.
Kyle manages a team that maintains aircrafts.



His job is to make sure that his 100 aircrafts are running properly & especially that the aircraft engines don’t need service.

Kyle wants to prevent delays due to mechanical issues so his customers will be happy.



Scenario

Sadly, engines occasionally show signs of problems & must be taken out of service for maintenance or replacement.

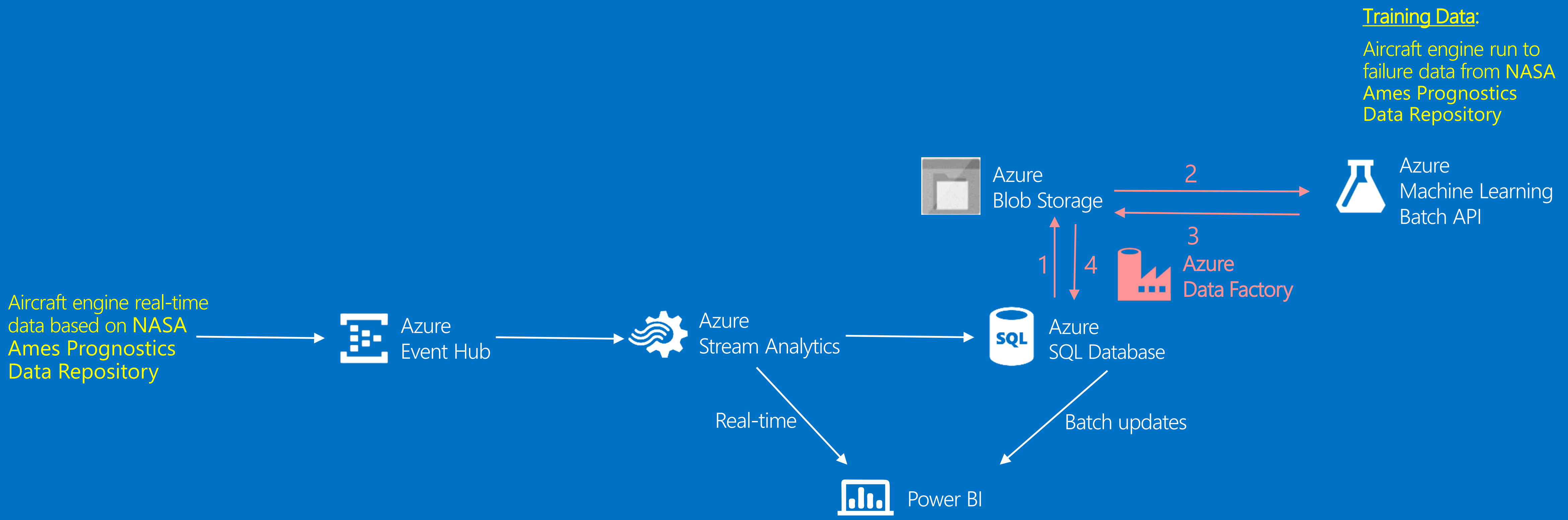
To eliminate this occurrence, Kyle must maintain operations & figure out the best way to utilize resources in order to minimize delays due to mechanical issues.

Questions & Solutions

Cortana Analytics to the Rescue!

1. How long did engines run in the past?
2. Which are showing signs of failure?
3. Which engines are likely to fail in the future?

Components


Training Data:
Aircraft engine run to failure data from NASA Ames Prognostics Data Repository
Azure Blob Storage
2
3
Azure Machine Learning Batch API
1
4
Azure Data Factory
Aircraft engine real-time data based on NASA Ames Prognostics Data Repository
Azure Event Hub
Azure Stream Analytics
SQL
Azure SQL Database
Real-time
Batch updates
Power BI

Sample training data

~20k rows,
100 unique engine id

id	cycle	setting1	setting2	setting3	s1	s2	s3	...	s19	s20	s21
1	1	-0.0007	-0.0004	100	518.67	641.82	1589.7		100	39.06	23.419
1	2	0.0019	-0.0003	100	518.67	642.15	1591.82		100	39	23.4236
1	3	-0.0043	0.0003	100	518.67	642.35	1587.99		100	38.95	23.3442
...	...										
1	191	0	-0.0004	100	518.67	643.34	1602.36		100	38.45	23.1295
1	192	0.0009	0	100	518.67	643.54	1601.41		100	38.48	22.9649
2	1	-0.0018	0.0006	100	518.67	641.89	1583.84		100	38.94	23.4585
2	2	0.0043	-0.0003	100	518.67	641.82	1587.05		100	39.06	23.4085
2	3	0.0018	0.0003	100	518.67	641.55	1588.32		100	39.11	23.425
...	...										
2	286	-0.001	-0.0003	100	518.67	643.44	1603.63		100	38.33	23.0169
2	287	-0.0005	0.0006	100	518.67	643.85	1608.5		100	38.43	23.0848

Sample testing data

~13k rows,
100 unique engine id

id	cycle	setting1	setting2	setting3	s1	s2	s3	...	s19	s20	s21
1	1	0.0023	0.0003	100	518.67	643.02	1585.29		100	38.86	23.3735
1	2	-0.0027	-0.0003	100	518.67	641.71	1588.45		100	39.02	23.3916
1	3	0.0003	0.0001	100	518.67	642.46	1586.94		100	39.08	23.4166
...	...										
1	30	-0.0025	0.0004	100	518.67	642.79	1585.72		100	39.09	23.4069
1	31	-0.0006	0.0004	100	518.67	642.58	1581.22		100	38.81	23.3552
2	1	-0.0009	0.0004	100	518.67	642.66	1589.3		100	39	23.3923
2	2	-0.0011	0.0002	100	518.67	642.51	1588.43		100	38.84	23.2902
2	3	0.0002	0.0003	100	518.67	642.58	1595.6		100	39.02	23.4064
...	...										
2	48	0.0011	-0.0001	100	518.67	642.64	1587.71		100	38.99	23.2918
2	49	0.0018	-0.0001	100	518.67	642.55	1586.59		100	38.81	23.2618
3	1	-0.0001	0.0001	100	518.67	642.03	1589.92		100	38.99	23.296
3	2	0.0039	-0.0003	100	518.67	642.23	1597.31		100	38.84	23.3191
3	3	0.0006	0.0003	100	518.67	642.98	1586.77		100	38.69	23.3774
...	...										
3	125	0.0014	0.0002	100	518.67	643.24	1588.64		100	38.56	23.227
3	126	-0.0016	0.0004	100	518.67	642.88	1589.75		100	38.93	23.274

Sample ground truth data

100 rows

RUL
112
98
69
82
91

Please refer to following link of doc for [Data description]() section

Conclusion

With the visualization prowess of Power BI, business owners can easily examine the performance of their entire company.

The Internet of Things and Stream Analytics connect data directly from the source to a dashboard to constantly track anomalies and asset performance in real-time.

Azure Machine Learning catches the problem before it becomes a problem. It streamlines operations without wasting resources.

Is the customer ready for ML?

The better the raw materials, the better the product.

Question is sharp.	Data measures what they care about.	Data is accurate.	Data is connected.	A lot of data.
E.g. Predict whether component X will fail in the next Y days	E.g. Identifiers at the level they are predicting	E.g. Failures are really failures, human labels on root causes	E.g. Machine information linkable to usage information	E.g. Will be difficult to predict failure accurately with few examples

Qualification Criteria

For ML-based solution:

1. Problem is predictive in nature
2. Clear path of action if potential failures detected
3. Data with sufficient quality
   - For predicting time left to failure, do you have failures or some proxy recorded?
   - Do you have enough failures to be able to model?
   - Is the “non-IoT” data in usable format?
   - Can the domain knowledge, such as timing of maintenance recordings, be translated into usable data for modeling?

Data Sources

FAILURE HISTORY

The failure history of a machine or component within the machine.

REPAIR HISTORY

The repair history of a machine, e.g. previous maintenance records, components replaced, maintenance activities performed. Maintenance types.

MACHINE CONDITIONS

The operation conditions of a machine, e.g. data collected from sensors.

MACHINE FEATURES

The features of machine or components, e.g. production date, technical specifications.

OPERATING CONDITIONS

Environmental features that may influence a machine's performance, e.g. location, temperature, other interactions.

OPERATOR ATTRIBUTES

The attributes of the operator who uses the machine, e.g. driver.

Feature Engineering

The process of creating features that provide better or additional predictive power to the learning algorithm.

id	cycle	setting1	setting2	setting3	s1	s2	s3	...	s19	s20	s21
1	1	-0.0007	-0.0004	100	518.67	641.82	1589.7		100	39.06	23.419
1	2	0.0019	-0.0003	100	518.67	642.15	1591.82		100	39	23.4236
1	3	-0.0043	0.0003	100	518.67	642.35	1587.99		100	38.95	23.3442
...	...										
1	191	0	-0.0004	100	518.67	643.34	1602.36		100	38.45	23.1295
1	192	0.0009	0	100	518.67	643.54	1601.41		100	38.48	22.9649
2	1	-0.0018	0.0006	100	518.67	641.89	1583.84		100	38.94	23.4585
2	2	0.0043	-0.0003	100	518.67	641.82	1587.05		100	39.06	23.4085
2	3	0.0018	0.0003	100	518.67	641.55	1588.32		100	39.11	23.425
...	...										
2	286	-0.001	-0.0003	100	518.67	643.44	1603.63		100	38.33	23.0169
2	287	-0.0005	0.0006	100	518.67	643.85	1608.5		100	38.43	23.0848

a1	a2	...	a21	sd1	sd2	...	sd21	RUL	label1	label2

60+ engineered features

Other potential features: change from initial value, velocity of change, frequency count over a predefined threshold

Example Feature Engineering Methods

1- Rolling aggregates:

For each labelled record of an asset, pick a rolling window of size w, compute rolling aggregate features for the periods before the labelling date and time of that record.

2- Lag features for short term:

For each labelled record of an asset, pick a window of size w and use tumbling windows to create aggregate features for the periods before the labelling date and time.

3- Lag features for long term:

For each labelled record, find aggregated features for a larger window than w reflecting the long term effects.

Create features that capture degradation over time.

Modelling Techniques

BINARY CLASSIFICATION

Predict failures within a future period of time

REGRESSION

Predict remaining useful life, the amount of time before the next failure

MULTICLASS CLASSIFICATION

Predict failures with their causes within a future time period.

Predict remaining useful life within ranges of future periods

ANOMALY DETECTION

Identify change in normal trends to find anomalies

Data Labeling

Regression

Binary classification

Multi-class classification

id	cycle	...	RUL	label1	label2
1	1		191	0	0
1	2		190	0	0
1	3		189	0	0
1	4		188	0	0
...			...	...	...
1	160		32	0	0
1	161		31	0	0
1	162		30	1	1
1	163		29	1	1
1	164		28	1	1
1	165		27	1	1
1	166		26	1	1
1	167		25	1	1
1	168		24	1	1
1	169		23	1	1
1	170		22	1	1
1	171		21	1	1
1	172		20	1	1
1	173		19	1	1
1	174		18	1	1
1	175		17	1	1
1	176		16	1	1
1	177		15	1	2
1	178		14	1	2
1	179		13	1	2
1	180		12	1	2
1	181		11	1	2
1	182		10	1	2
1	183		9	1	2
1	184		8	1	2
1	185		7	1	2
1	186		6	1	2
1	187		5	1	2
1	188		4	1	2
1	189		3	1	2
1	190		2	1	2
1	191		1	1	2
1	192		0	1	2

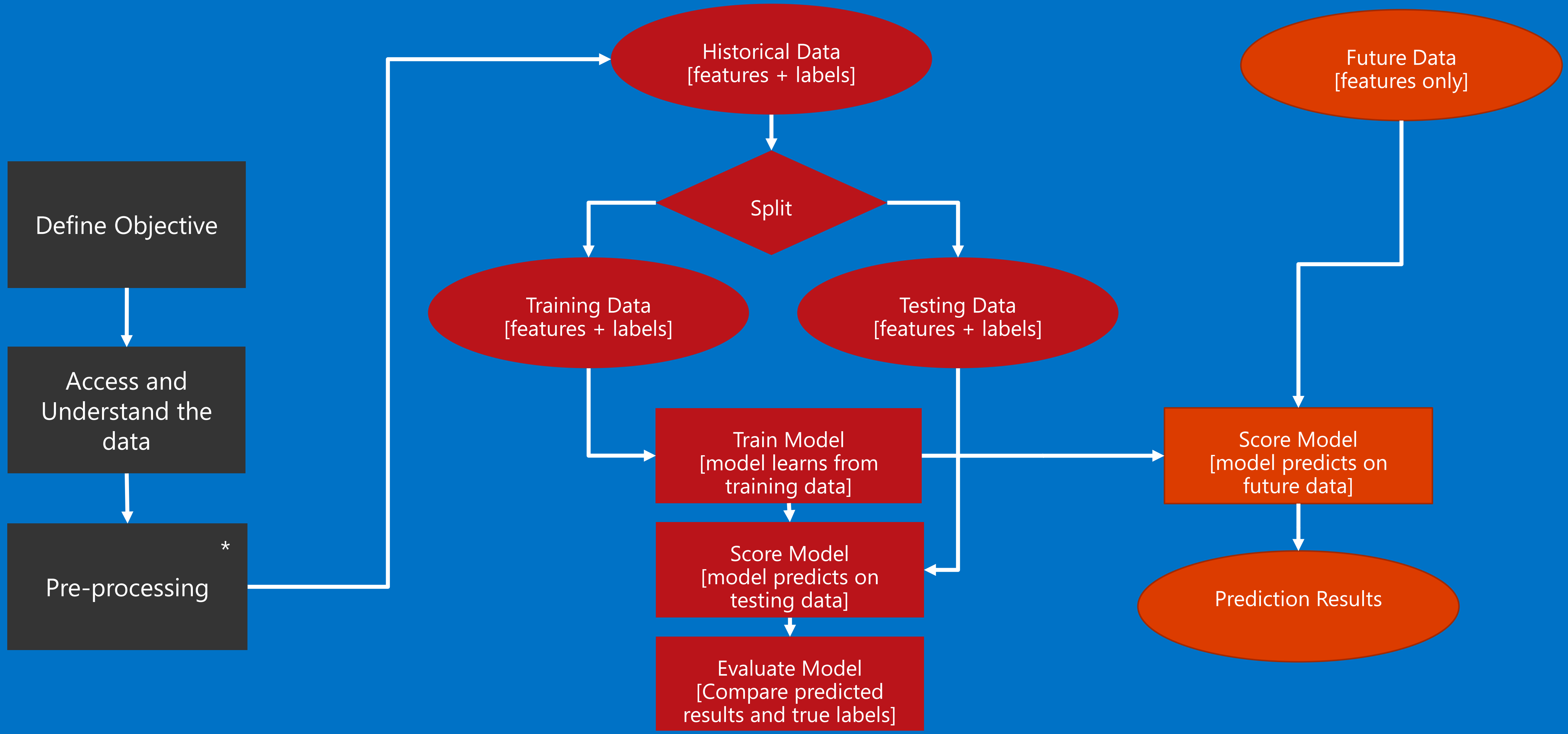
w1

w0

Predefined window size for classification models

w1 = 30
w0 = 15

ML Process



*Depending largely on size/complexity, may want to do pre-processing and/or feature/target construction before ingesting into AML Studio / AML API.

"Most IoT data are not used currently…

*the data that are used today are mostly for anomaly detection and control, not **optimization and prediction**, which provide the greatest value."*[1]

Go Dos

- Learn from Azure Machine Learning Gallery
  http://gallery.azureml.net (search "predictive maintenance")

Acknowledgements

We utilized the following publically available data to help us generate realistic data for this pre-configured solution. We received assistance in creating this solution as a result of this repository and the donators of the data.

“A. Saxena and K. Goebel (2008). "PHM08 Challenge Data Set", NASA Ames Prognostics Data Repository (http://ti.arc.nasa.gov/project/prognostic-data-repository), NASA Ames Research Center, Moffett Field, CA.”

Quote on Slide 23: McKinskey Global Institute, The Internet of Things: Mapping the Value beyond the hype

Microsoft