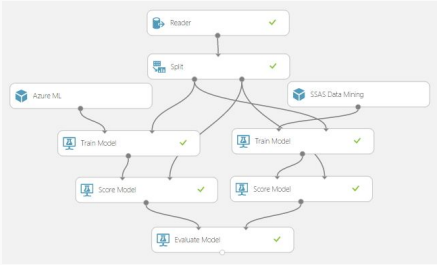
IoTFuse 2019 Data Workshop

Creating Predictive Web Services with Azure ML

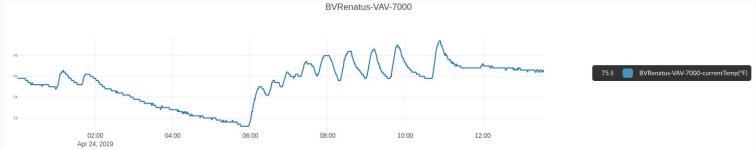
Presented by Madhu Tennakoon

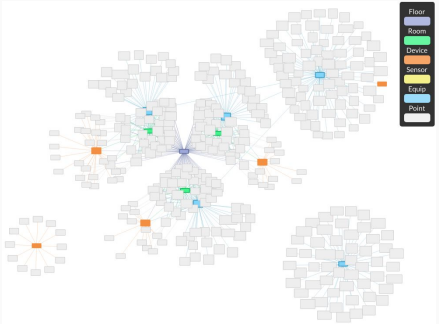
Goals

- Designing and deploying a predictive maintenance web service (1hour 30mins)
  - Experimentation on Jupyter Notebook and Azure ML Studio
  - Deploy a predictive web service
  - Design an interactive dashboard



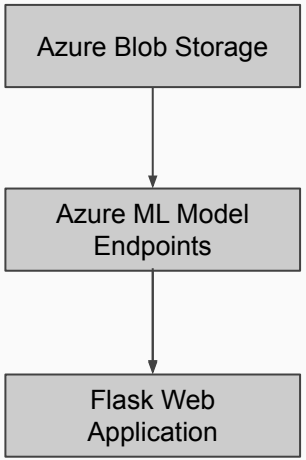
- Demo: Better IoT data analytics with Project Haystack (30mins)
  - Standardizing building IoT data: Project Haystack
  - Haystack API on Azure
  - Visualizing building data





What we will need

1. NASA C-MAPSS Dataset - sample dataset (simulates our IoT data)
2. Azure ML Studio Workspace
3. Azure Blob Storage - store data and metrics
4. If you're working locally:
    a. Anaconda environment
    b. Jupyter Notebook
5. Flask - Python web application development



- Data storage
- Demo purposes only!
- Consider InfluxDB, CosmosDB

- Predictive models
- Deployed as web services
- Accessed via http endpoint with access-key

- Custom visualizations
- Model evaluation

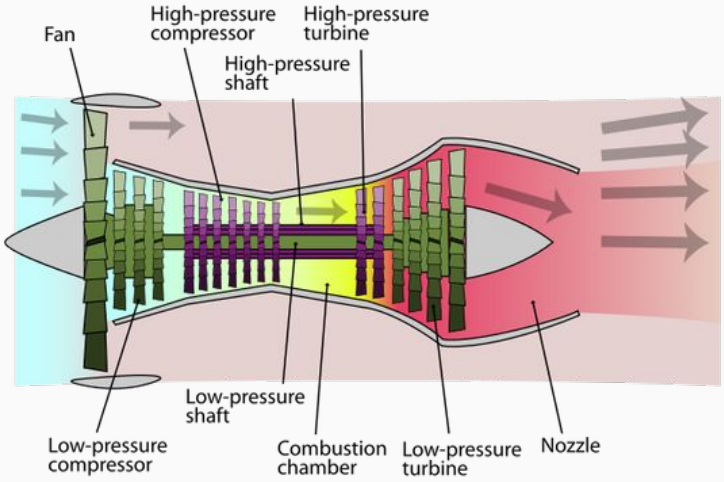
Part I: Machine Learning Experimentation

ML experimentation workflow

1. Understanding your data
2. Formulating the problem - is it a classification problem or a regression problem?
3. Data preparation - Which sensors are useful? What features do we extract? How do we remove noise? How do we create training and testing data?
4. Learning and predicting - neural network, decision tree, ensemble methods, e.g. random forests, etc.
5. Performance evaluation - choosing the right metrics; accuracy, precision, recall

Understanding your data

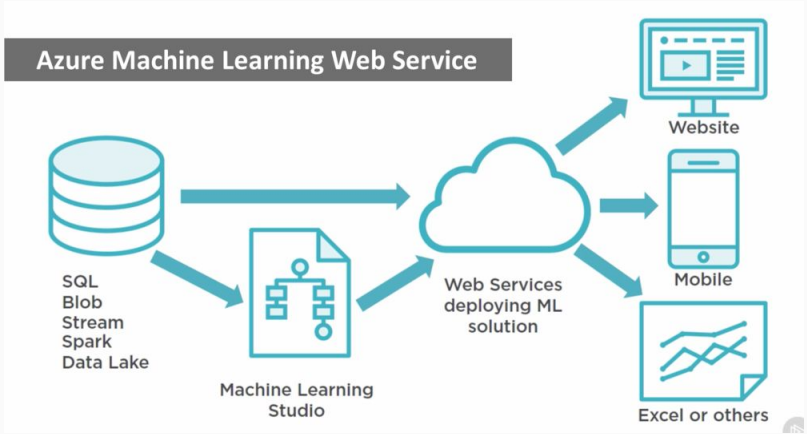
- What are you trying to predict?
- Which datasets do you use?
- How many fault modes exist?
- Managing categorical and numeric data types
- What assumptions are being made of the data?



Part 2: Model Deployment

Model deployment

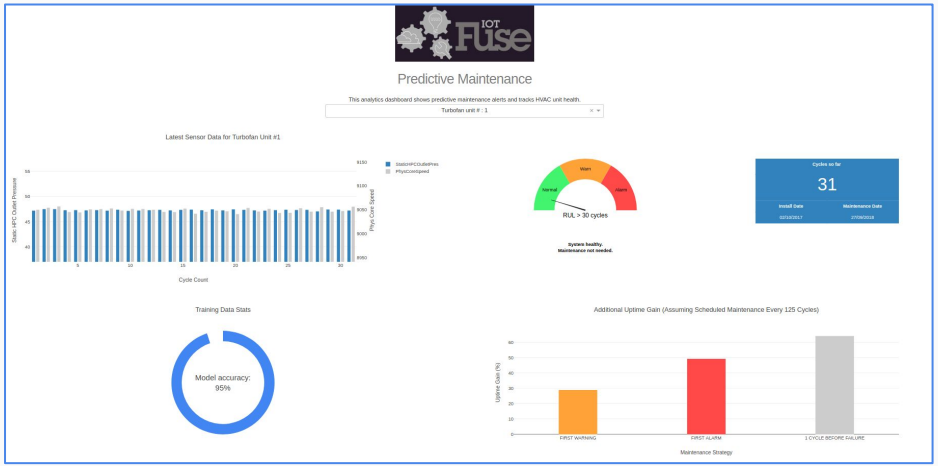
1. Creating a low-latency HTTP endpoint that can be accessed using an authorization token
2. Web service input: **features**
3. Web service output: **prediction**



Part 3: Creating visualizations

Visualizations

1. Designing a dashboard with Plotly Dash
2. Displaying sensor data
3. Creating widgets to show metrics



Show & Tell:

Using Project Haystack to streamline IoT analytics

What is Project Haystack?

- We have hundreds of thousands of endpoints streaming real-time sensor data. How do we make sense of it?
- Most operational data has poor semantic modeling and must be mapped before value creation
- Haystack is an open-source initiative to streamline data from the Internet of Things
- Standardized semantic data models
- Translation: all your data speaks the same language, and is self-describing
- Enables more cost-effective and powerful data analytics
- https://project-haystack.org/

What is Project Haystack?

- We have hundreds of thousands of endpoints streaming real-time sensor data. How do we make sense of it?
- Most operational data has poor semantic modeling and must be mapped before value creation
- Haystack is an open-source initiative to streamline data from the Internet of Things
- Standardized semantic data models
- Translation: all your data speaks the same language, and is self-describing
- Enables more cost-effective and powerful data analytics
- https://project-haystack.org/