INTUITIVE®

Using Big Data Analytics Approaches to Analyze Aviation and Missile RAM Data

11/1/2023

Presenters

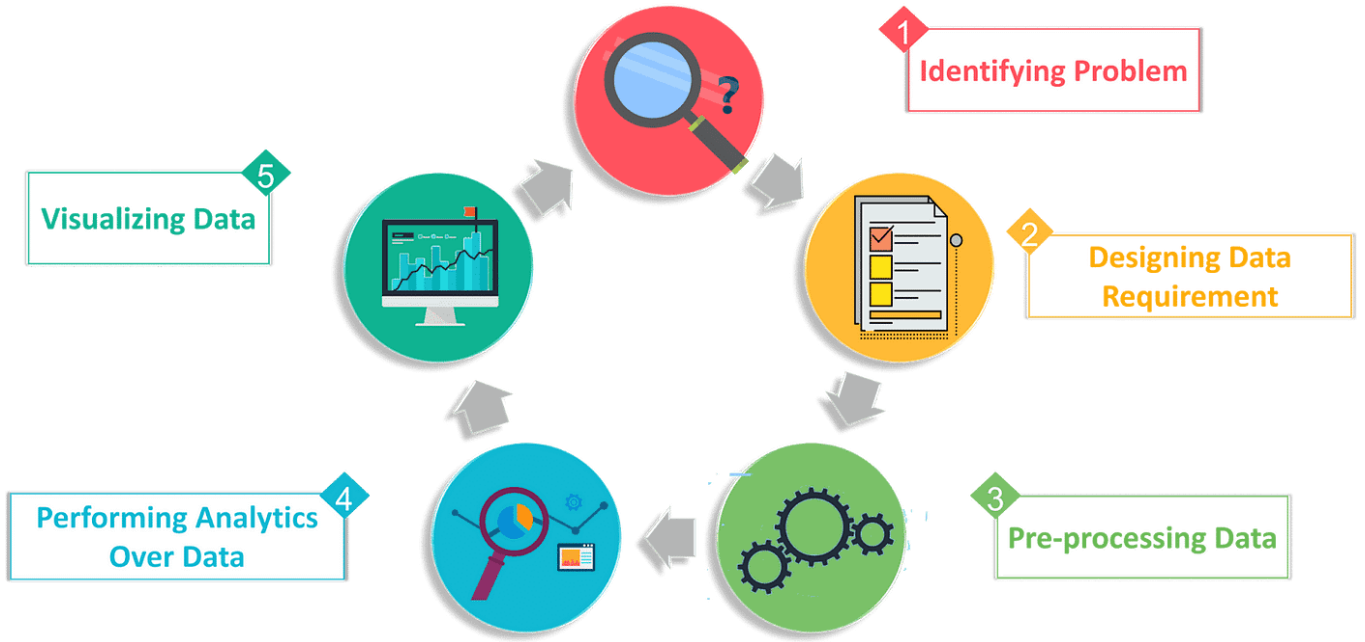
- Jordan Mills
- Kyle Russell
- Connor Green

Outline

- Introduction to Big Data Analytics
- Overview of Project
- Overview of Technical Methodologies/Processes/Tools
- Overview of Aviation Data
- Overview of Missile Data
- Overview of Missile Analysis
- Detailed Missile Analysis Discussion
- Missile Analysis Next Steps
- Conclusion

What is Big Data Analytics?

"Big data analytics describes the process of uncovering trends, patterns, and correlations in large amounts of raw data to help make data-informed decisions. These processes use familiar statistical analysis techniques – like clustering and regression – and apply them to more extensive datasets with the help of newer tools …. big data analytics methods are being used with emerging technologies, like machine learning, to discover and scale more complex insights." [1]



Courtesy of MultiTech [2]

1. https://www.tableau.com/learn/articles/big-data-analytics#:~:text=Big%20data%20analytics%20describes%20the,the%20help%20of%20newer%20tools
2. https://informationit27.medium.com/explain-big-data-analytics-5b814ce4390

Overview of Project

INTUITIVE®

Project Description

- Big Data Analytics Project supporting **Reliability, Availability, Maintainability (RAM) Division** within System Readiness Direction (SRD) under DEVCOM AvMC
- Involves using and maturing data analytics **methods** (dimensionality reduction, cluster analysis, natural language processing, neural nets) and **tools** to **find patterns/relationships** in the data and build models that **predict future events**
- Involves analyzing both **aviation** and **missile** data
  - **Validation** of aviation findings from 2022 and continued exploration of maintenance data
    - Data of interest includes maintenance events and on-board flight sensor data for the **UH-60 M aircraft**
  - **Exploration** of missile data to find patterns/relationships in the data
    - Data of interest includes component testing data, flight testing data, surveillance inspections and testing data, and field data to support the **Stockpile Reliability Program (SRP)**

Overview of Project

INTUITIVE®

Technical Objectives

- Mature and demonstrate analytics tools (powered by Artificial Intelligence/Machine Learning) that evaluate large data sets and find relationships within and between Army datasets
- Train supervised Machine Learning applications to proactively drive maintenance to increase platform availability
- Use derived relationships to develop and test models to improve operational readiness, aircraft certifications, condition-based maintenance, and lethality; therefore, increasing Army weapon system effectiveness

Overview of Project

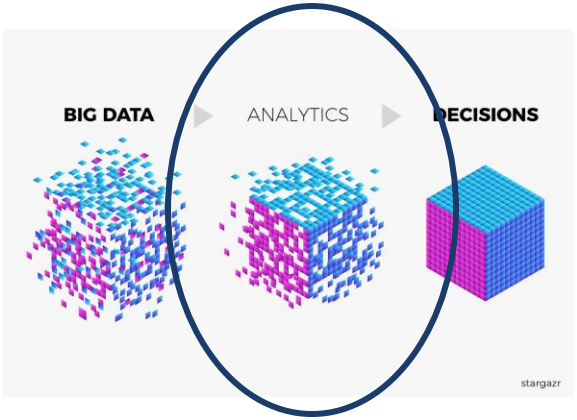
INTUITIVE®

Overarching Goals

- 1) **Develop methodologies** for combining multiple data sources to provide quantitative insight on negative impactors to platform/system availability/performance.
- 2) **Provide actionable data** and **predictive models** that can be used for the development of targeted Reliability Improvement Projects to improve readiness.
- 3) **Validate currently used testing regimes**, and/or **identify additional regimes and data** that can be used to improve operational lifetimes.
- 4) **Provide awareness** at all levels (enterprise to user) of the **environmental (parametric and regime) impacts** to the sustainment of new and legacy systems.
- 5) **Develop and document algorithms, workflow, and platform architecture**.

Some Technical Approaches/Concepts

- **Dimensionality reduction**
- **Cluster analysis**
- **Natural Language Processing**
- **Neural Nets, Transformers**



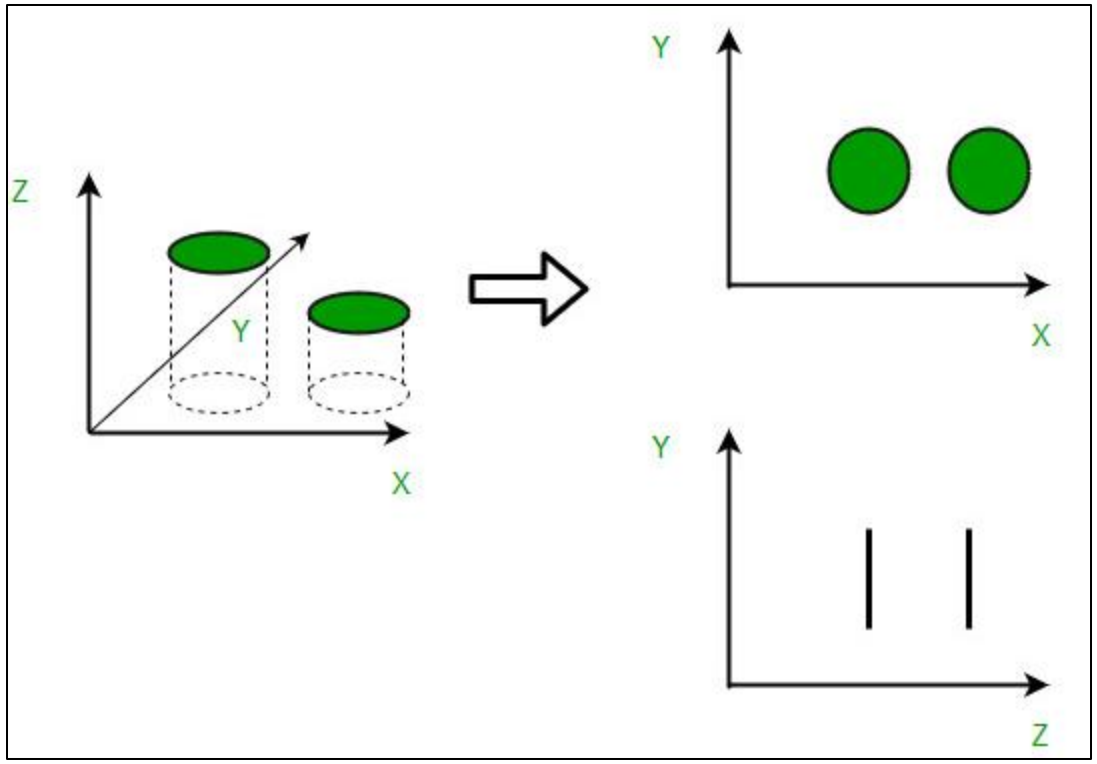
Courtesy of TowardsDataScience [1]

1. https://towardsdatascience.com/how-is-the-current-state-of-big-data-analytics-in-controlling-1273c725ac6a

Dimensionality reduction techniques, cluster analysis, application of neural nets and transformers, and Natural Language Processing (NLP) can all be used to find patterns and relationships within large data sets.

Dimensionality Reduction

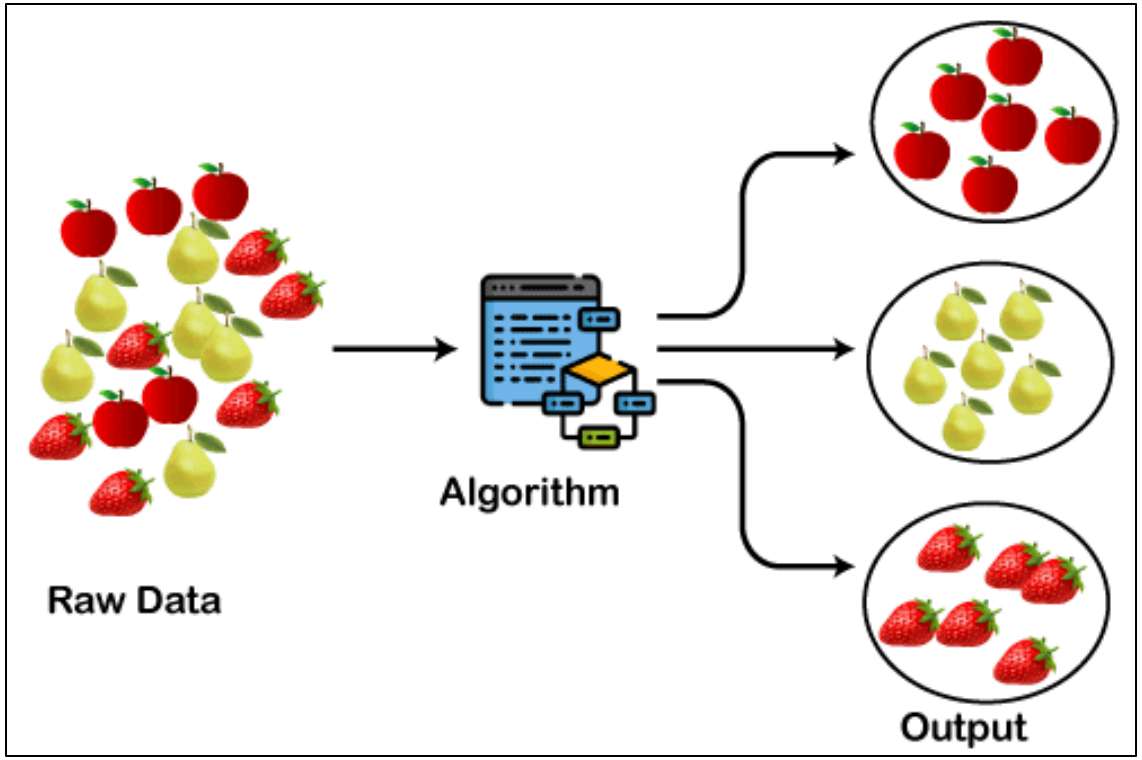
- This process calculates a dimensionally reduced representation of a dataset based on the many attributes in the original data.
- Imagine organizing your home
  - Place an item into a pile
  - Item goes with other similar items:
    - Function
    - Size
    - Color
    - Etc.
- Can ***FORM*** **CLUSTERS** in lower dimensional space.
- Useful for analyzing data with many features!



We use Dimensionality Reduction methods to reduce the Health Usage Monitoring System (HUMS) datasets prior to searching for correlations with the ACN (Maintenance) data

Cluster Analysis

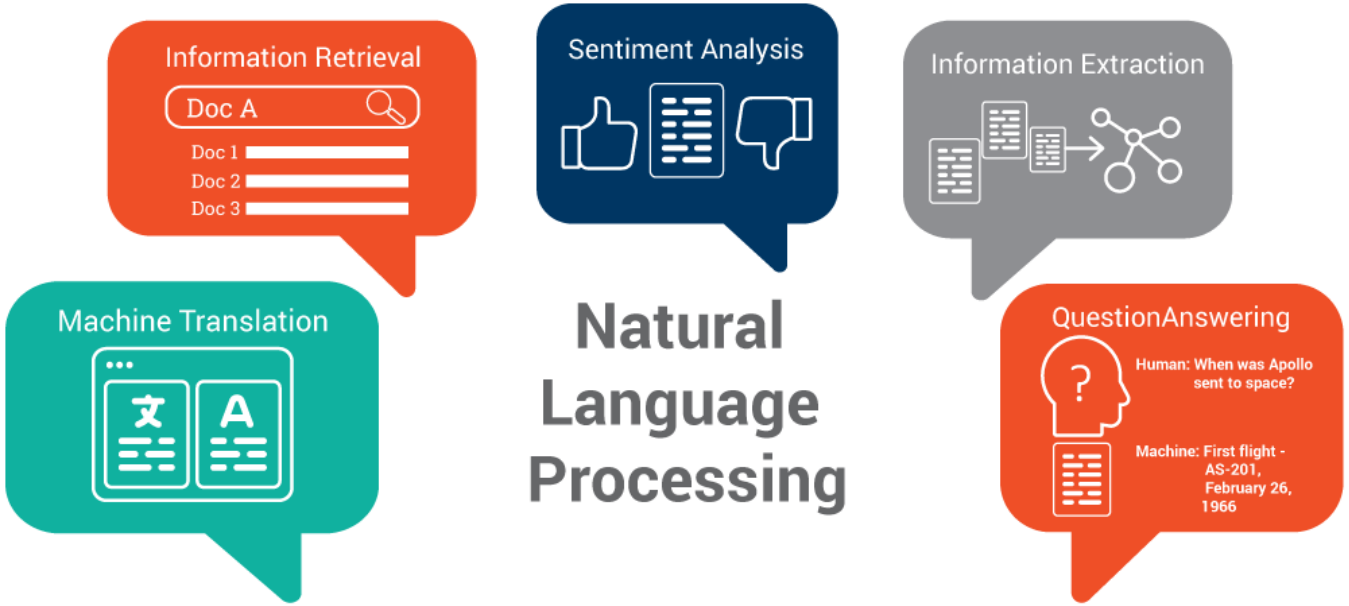
- Clustering **labels** data that is similar with the same label
- Many techniques
  - More supervised (k-means)
  - Less supervised (hdb-scan)



We use Cluster Analysis methods to allow the artificial intelligence/machine learning methods to find potential groupings of HUMS data that may be precursors of the maintenance events

Natural Language Processing

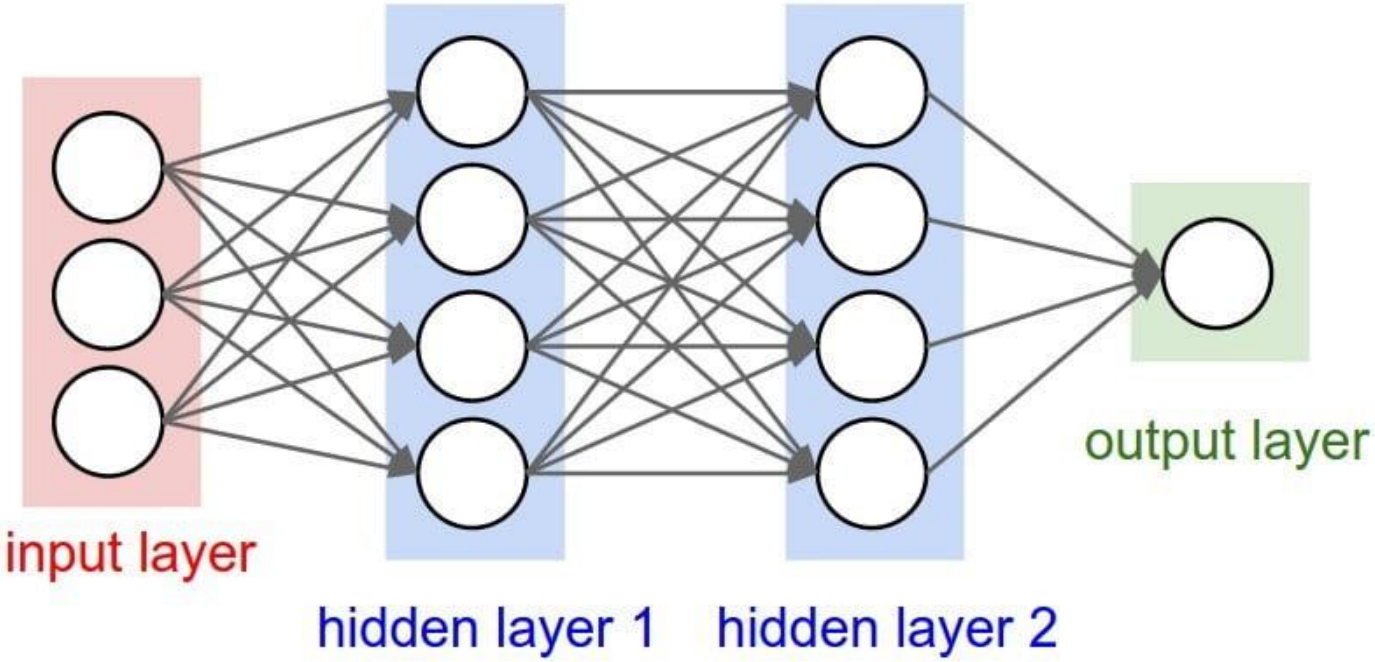
- Goal of Natural Language Processing is to allow computers to be able to understand language in a very similar way to humans



We use Natural Language Processing methods and algorithms to analyze and prepare the missile data sets prior to further mathematical examination

Neural Nets/Transformers

- Subset of machine learning concerned with using deep learning algorithms to train a model to predict behavior



We use Neural Nets and Transformers to develop predictive models that complement or improve pure statistics-based predictive models

More on Transformers

INTUITIVE®

- A transformer…
  - Deep Learning architecture, similar to recurrent networks, different mechanism
  - This architecture is used in NLP, computer vision, and audio processing
  - It has led to the development of pre-trained systems, such as Generative Pre-trained Transformers (GPTs)
- Applications
  - NLP, machine translation, document summarization and generation
  - Time series prediction (finance, weather, healthcare)
    - For healthcare – transformers can recognize the patterns of a patient developing a certain condition or disease
- What can we get from the transformer model?
    - The ability to recognize the patterns of an aircraft developing a certain condition that leads to a non-scheduled maintenance event

We are exploring the use of Transformers for predictive analytics…
using the HUMS data [channels] as predictors of ACN events.

Big Data is a Spectrum of Content

INTUITIVE®

…that requires specialized methods and tools to analyze and evaluate. We select and tailor the data science tools based on the data and outcome.

data type	modern data science methods
numbers	***computational algorithms, statistics, dimensionality reduction, multi-dimensional visualization, AI/ML***
models	***natural language processing, similarity analysis, statistics***
logs	***AI/ML behavior identification and characterization, auto-evaluation***
images	***computer vision, convolutional neural nets, AI/ML, virtual fiducials, image alignment, data and image fusion***
written text	***natural language processing, sentiment analysis, trend analysis, similarity analysis***

Our portfolio of tools

COMPREHEND™

NLP prototypes

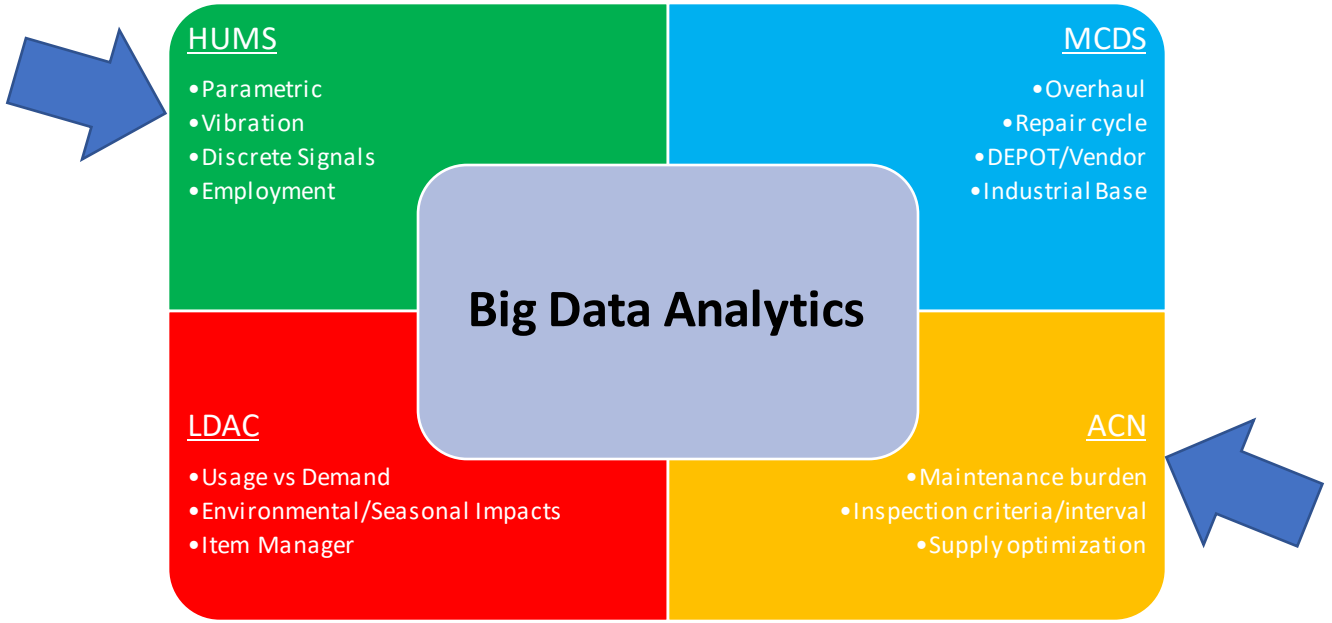
DECIDE™
Managed Cyber

CLARUS®
VIEWER

sentelligence™

Overview of Aviation Data

- Intent - combine/correlate data sources into a holistic picture to develop a prognostic toolset that services the enterprise at every level



- Expectation – quantify the relationship between how a platform is utilized to the maintenance/logistic burden that is generated
- Build a targeted supply/maintenance program for component level rebuild/overall and platform deep cycle maintenance programs such as RESET, RECAP & OCM
- Overall objective to reduce TOC and increase platform availability

We are analyzing the relationship between the Health Usage and Monitoring System (HUMS) and ACN datasets for this analysis effort.

Overview of Missile Data

COMPONENT TESTING

Sample of components tested in lab to ascertain essential performance parameters not available at all-up-round level.

- Almost all components are laboratory tested: Batteries, Motors, Warheads, Gyros/IMUs, Guidance Sections
- Critical data for trending/predicting performance PRIOR to the Soldier experiencing failures in the field

FLIGHT TESTING

A statistically representative sample of the tactical inventory is flown with flight test range controls and data collection.

- Flight testing is essential to measure system level reliability and identify failure modes that cannot be identified in the laboratory
- Determines if the system meets the user's in-flight reliability requirement

FIELD DATA

Firing and Malfunction Reports are required to be submitted by the user.

- Online entry system developed by DEVCOM AvMC has significantly improved reporting rates
- Analyzed for trends associated to manufacturing strata, age, firing scenario, etc.

SURVEILLANCE INSPECTIONS AND TESTING

Mobile capabilities allow us to non-destructively test missiles worldwide.

- Provides valuable parametric data for trend analysis. Not able to test "one-shot" components (thermal batteries, motors, warheads, etc)
- Directly improves readiness: immediately segregates failing missiles from the go-to-war inventory.

We are unifying the data into a single data set and analyzing the data holistically.

Overview of Missile Data

INTUITIVE®

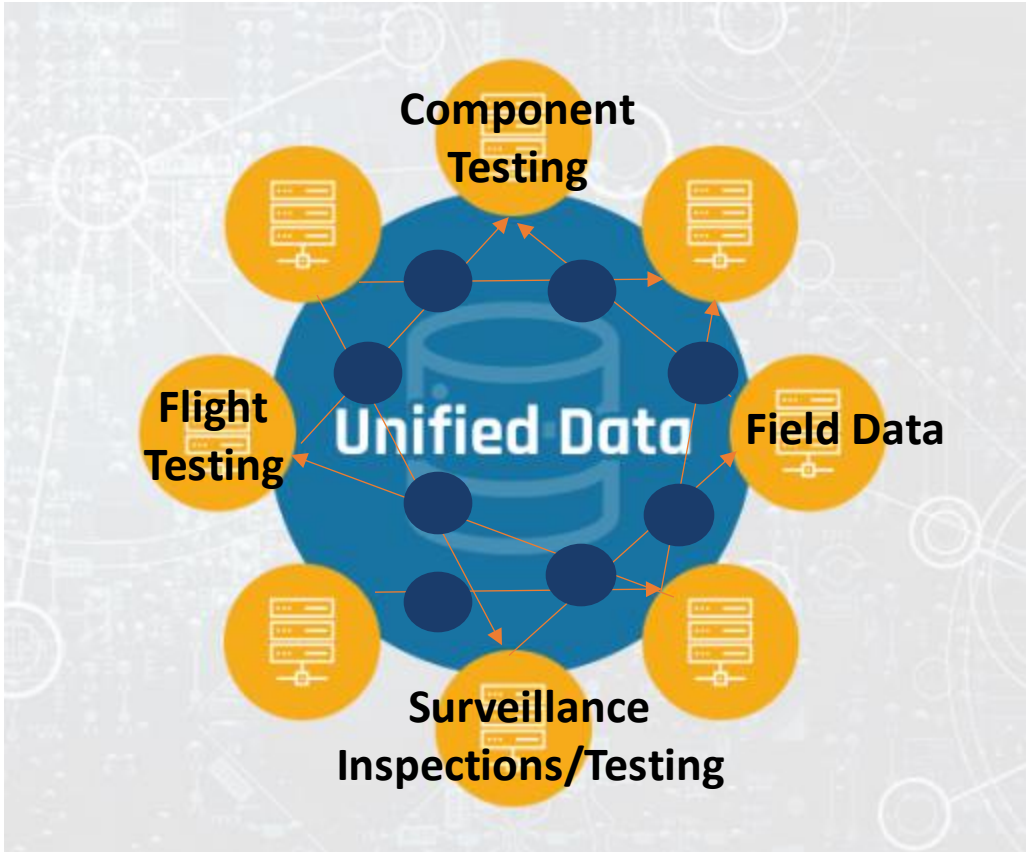
- We've received data from various data sources that falls into 4 main categories
  1. **Component Testing**
  2. **Flight Testing**
  3. **Field Data**
  4. **Surveillance Inspections/Testing**
- For **component testing data**, we have received data associated with the major components of the weapon systems including the Battery, Electronic Safe, Arm, Fire (ESAF) device, Accumulator, Control Section, Gyroscope, Warhead, Rocket Motor, and Guidance Section. We have also received the storage histories associated with these missiles prior to selection for test and test data recorded prior to disassembly.
- For **flight testing data**, we have received the pre-flight test data, in-flight test data, and storage histories associated with the missiles.
- For **field data**, we have received malfunction reports and firing reports.
- For **surveillance test data**, we have received laser log data.
- We have also received **support data** in the form of a lot identifier explanation, acronyms list, Weapon System Evolution briefing, and an SRP overview briefing.

We are analyzing a large amount of missile test data of varying types and sizes.

Overview of MSL Analysis – Unification Concept

INTUITIVE®

The various data sources have historically been analyzed separately. An initial effort of the team is to combine all available data into a comprehensive source.



Note: This image is not meant to depict actual relationships between missile datasets

Overview of Missile Analysis

INTUITIVE®

- Analysis began with accumulation of the various datasets into a **single unified** dataset.
- We used unsupervised artificial intelligence (**AI)** to reduce the unified data to **three dimensions** to allow for **visual inspection and forensics**.
- Performing visual forensics on the data allowed us to focus on specific aspects of the dataset.
- Visual forensics then led us to statistical analysis on subgroup representations for defective test results.

Our initial inspections led us to take two different approaches for unification of the missile data. We have ideas for more unique ways to evaluate the data.

Unification Plan for Missile Data

For successful analysis, data needs to be condensed into a single file with a unified structure for ease of use. In addition, including new files needs to be simple and automatic.

Received Data

Unification Script:

- ✓ Will automatically create a single master data file from everything provided
- ✓ Similar data fields will be unified in a standard format
- Some missing data will be added from context and engineering judgement
- Overly specific data will be simplified until it can be unified
- ✓ Newly provided data will only require this script to be run again to add it
- ✓ Fix some data entry errors

Search and Filter Tool:

- Builds family tree of similar missile results
- Builds timeline of missile results

Optional

Developed Analysis Tools and AI

The use of this schema will provide a unique way to analyze missile data from the various data sources.

Unification Current Status

From the original excel data, the 100s of unique columns are automatically unified into the following:

Category Columns:

- Missile Serial Number
- Missile Lot Number
- Date of Test
- Test Filename
- Test Name

Metric Columns:

- Test Verdict
- Age at Test

Including additional metrics is in progress

Using the Missile Lot Number, these columns:

Example of Lot Number: **MGP17A001-001R**

- Production Contractor
- Year of Production
- Month of Production
- Missile System (Longbow, HFII, etc.)
- Configuration (Basic, True ILV, Blast Frag, etc.)
- Buy Number (which round of purchases it belongs to)
- Lot Identifier (theatre deployment, or modifications)
- Lot Sequence Number
- Lot Suffix (blank unless there's a repair contract code or retrofit)

Missile Lot Numbers are being leveraged to provide additional context to existing data.

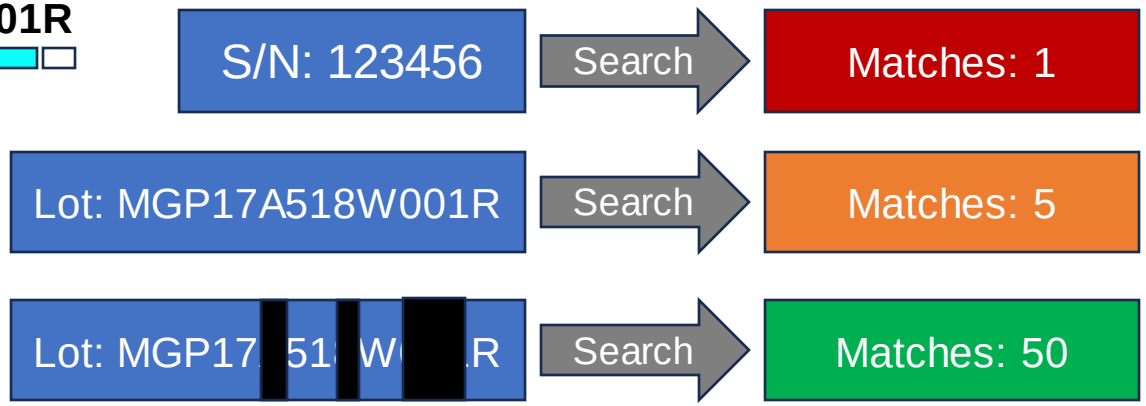
Unification Search Filter Plan

INTUITIVE®

By ignoring parts of a missile lot number, families of test results can be grouped in a variety of ways:

Example of Lot Number: **MGP17A518W001R**

- Contractor
- Year of Production
- Month of Production
- Missile System
- Configuration
- Buy Number
- Lot Identifier
- Lot Sequence Number
- Lot Suffix



Ignoring month, buy number, and sequence number creates a **missile test family tree** of similar age and make that has been exposed to the same harsh environment and has undergone a rocket motor retrofit.

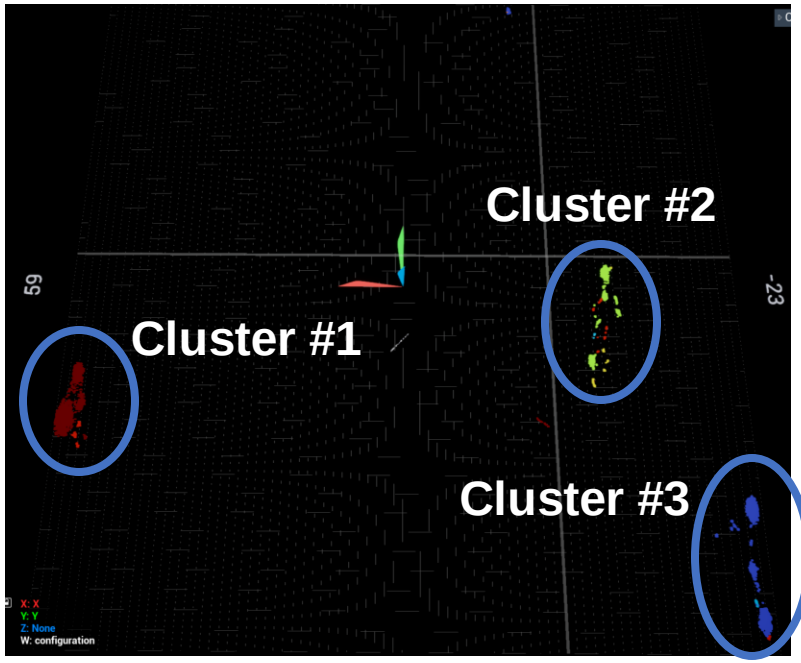
Ignoring lot identifier and suffix allows search results to form a **missile test timeline** where a group's deployment, modifications, and repairs/retrofits can be examined before and after.

Parts of the missile lot numbers can be used to filter and group data into family trees and timelines to improve data analysis efforts.

Preliminary Findings

Loading the unified file into CompreheND enables us to identify similarities and trends in missile data before investigating specific subsets

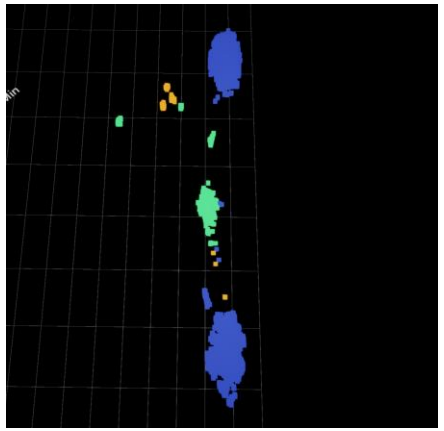
Configuration



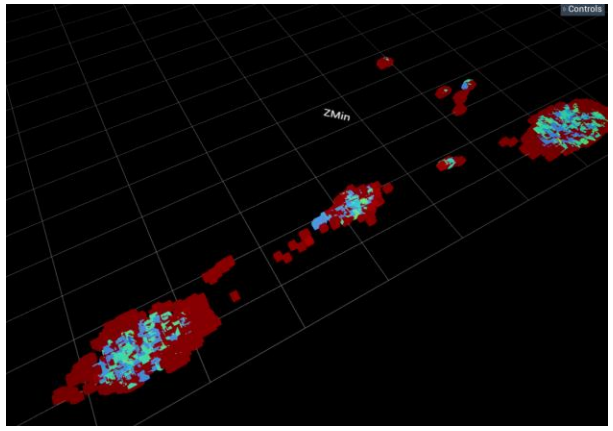
All images taken using CompreheND with the unification file of RAM MSL data loaded.

- **When colored for lot suffix categories**, interesting divisions were noticed within the clusters of similar missiles
- **When colored for filenames (50 total files)**, only **4 colors (4 files)** are heavily represented
- The next step is to **select subsets and compare them**

Lot Suffix



Filename

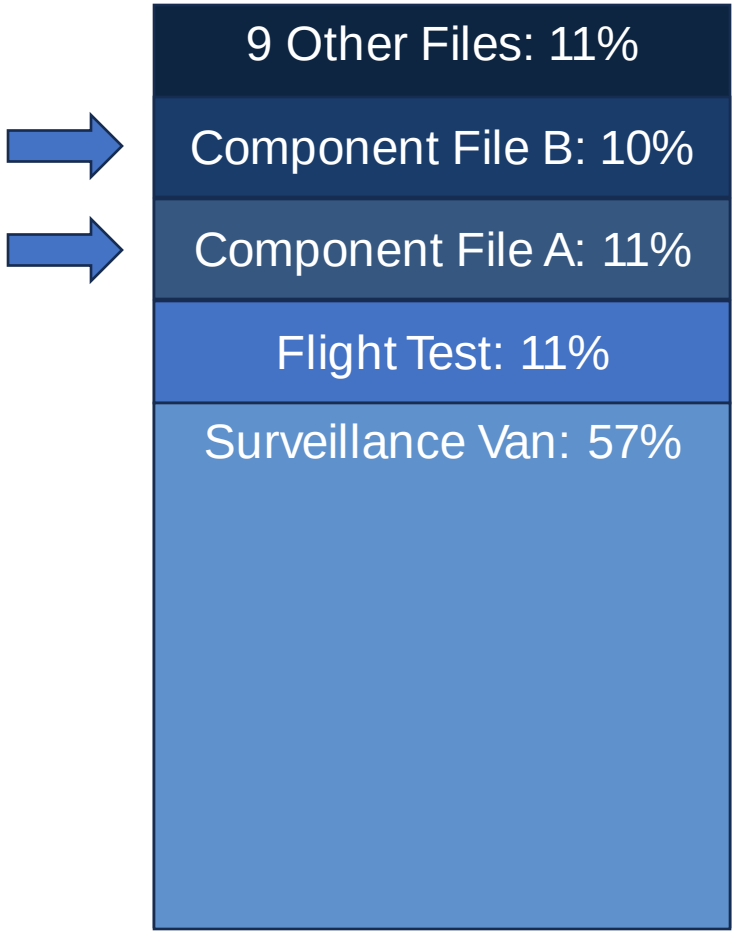


We were able to narrow down the amount of data that we were evaluating at one time – easier to understand, enables quicker turn-around time

Interesting Observation 1

Of a previous subset of missiles, a high portion of defective tests were flagged:

- 57% come from the surveillance van data
  - Primary form of non-destructive testing
- 11% come from the flight data
  - A full system test in tough conditions
- 11% come from a single component file (out of 44 files)
  - Testing on single part or function
- 10% come from a single component file (out of 44 files)
  - Testing on a single part or function

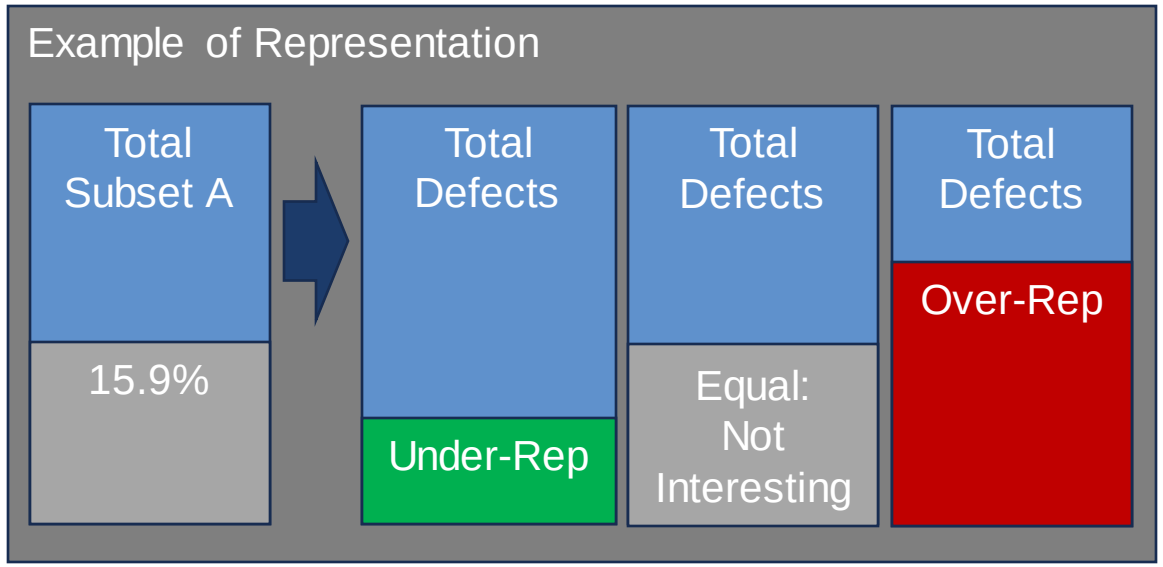


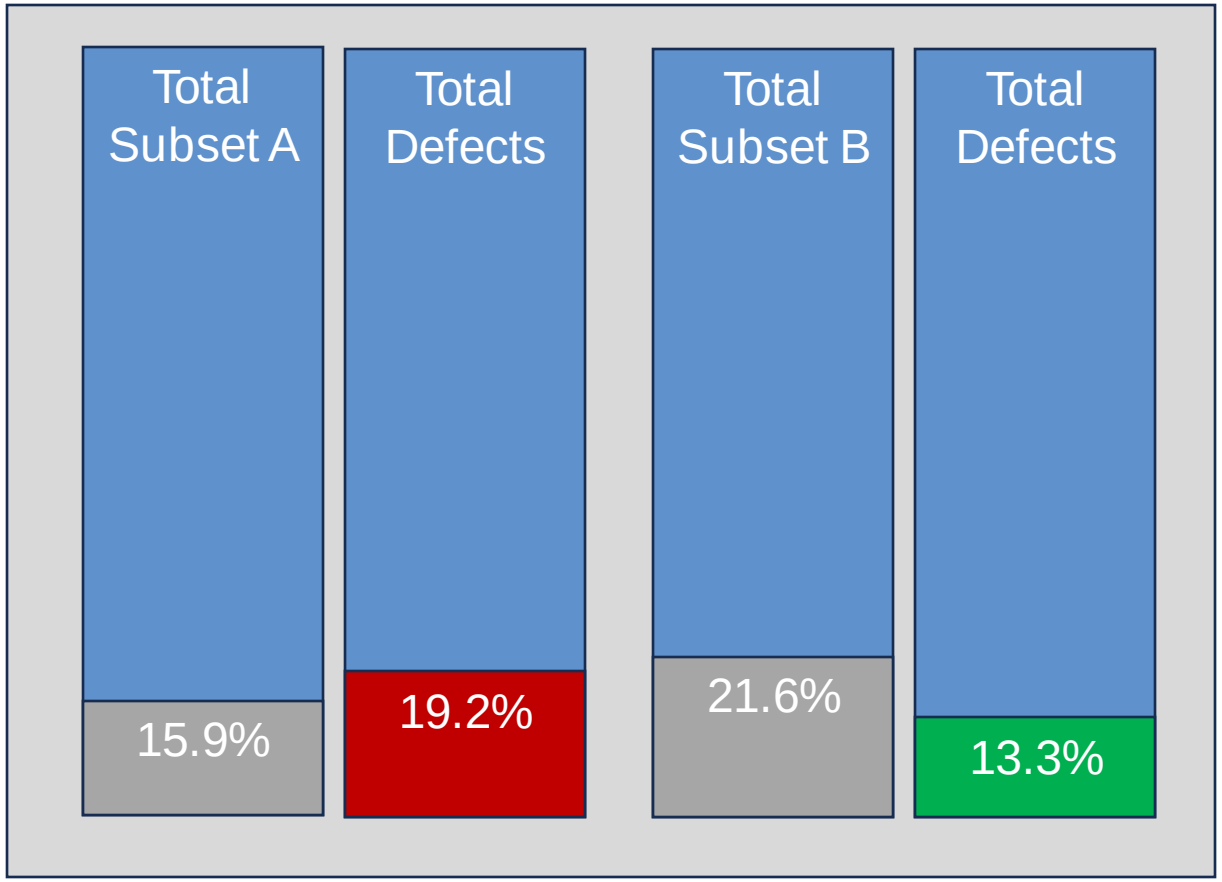
21% of defective conditions are a result of just two of the files that were merged into the unified data set. This indicates that these two documents should be further investigated because they are disproportionally affecting the data set

Interesting Observation 2

Representation in defects of subset A and subset B missiles:

- 15.9% of missiles are in subset A
- 19.2% of defects are in subset A
- 21.6% of missiles are in subset B
- 13.3% of defects are in subset B





Subset A had more defects than expected. Subset B had less defects than expected.

Missile Analysis Next Steps

INTUITIVE®

- **Validate** initial observations
- **Continue to unify data** across component testing, flight testing, field data, and surveillance van testing into a single unified data set
- **Identify additional observations** through exploration of the data

Conclusion

- As the amount of data that engineers and analysts need to evaluate increases, the use of **AI and ML** techniques become more critical.
- **Dimensionality reduction techniques, cluster analysis, application of neural nets and transformers, and NLP** can all be used to **find patterns and relationships within large data sets.**
- The domain of **RAM for aircraft and missiles** provides an ideal application for these big data methods.
- **For aviation data**, looking at relationships and patterns within and between maintenance events and on-board flight sensor data for the UH-60 M aircraft can provide added value for maintenance and logistics managers.
  - Using the ML-derived relationships between sensor data and subsequent maintenance events can reveal potential channels that function as indicators of upcoming non-scheduled maintenance events.
- **For missile system data**, looking at relationships and patterns within and between component testing data, flight testing data, field data, and surveillance inspections and testing data to support the Stockpile Reliability Program (SRP).
- We are working to **mature data science tools and techniques** to provide value to RAM stakeholders through the innovative analysis of aviation and missile big data sets.