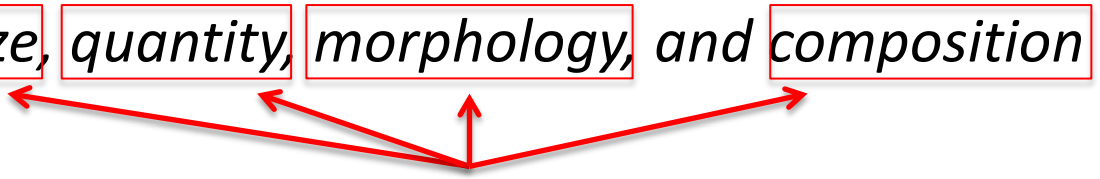
Australian Government
Department of Defence
Science and Technology

Using K-Nearest Neighbours (KNN) Machine Learning Technique to Classify Archived Helicopter Wear Debris Data

Eric Lee (DST Group)

ADF Wear Debris Analysis (WDA) Laboratory

- **Joint Laboratory** between Defence Aviation Safety Authority (DASA) Engine Structural Integrity (ESI) and Defence Science and Technology Group (DST Group)
  - *Free service to ADF Air assets (Gas Turbine Engine, Gearboxes, and other mechanical systems)*
  - *Not "Oil" analysis, but metallic particles*
  - *Determine the size, quantity, morphology, and composition of particles*



Features for Machine Learning

- **Why?**
  - ***To improve the availability and safety of the Air asset without incurring the cost of removing the whole part***

ADF Wear Debris Analysis (WDA) Laboratory

- Since the creation of the WDA Lab:
  - *Filter Debris Analysis (FDA) for Fixed Wing and Rotary Wing ADF air assets*
  - *A large number of Wear Debris Analysis Reports (PDF)*
    - *Size and Quantity (According to ASTM 7898-14)*
    - *Shape and Features (According to ASTM 7898-14)*
    - *Material*

Attention		Sample Date	
Aircraft		Report Date	
Component		Serial No	
TSN		Comments	
ADF-WDA Lab Ref		Objective Ref	

References:
A. ASTM 7898-14 Standard Practice for Lubrication & Hydraulic Filter Debris Analysis for Machinery Condition Monitoring
B. Metal Map

Location	Quantity [1] (Extreme 100+ Particles; Large 26 to100 Particles; Few 6 to 25 Particles; Small 1 to 5 Particles)		Shape and Features [2]	Material [3,4]	Typical Source (Ref. B)
	Size Range	100-250 µm			
		250-500 µm			
		500-1000 µm			
		1000+ µm			
		Total			

What Next?

- Health and Usage Monitoring Systems (HUMS) Community:
  - *A lot of data been recorded and archived, but no one is looking into the data*
  - ***Why?***
    - ***"Because there is too much data for any person to go through manually"***
- Buzzwords:
  - *Big Data*
  - *Data Fusion*
  - *Data Analytic/Mining (e.g. Machine Learning)*

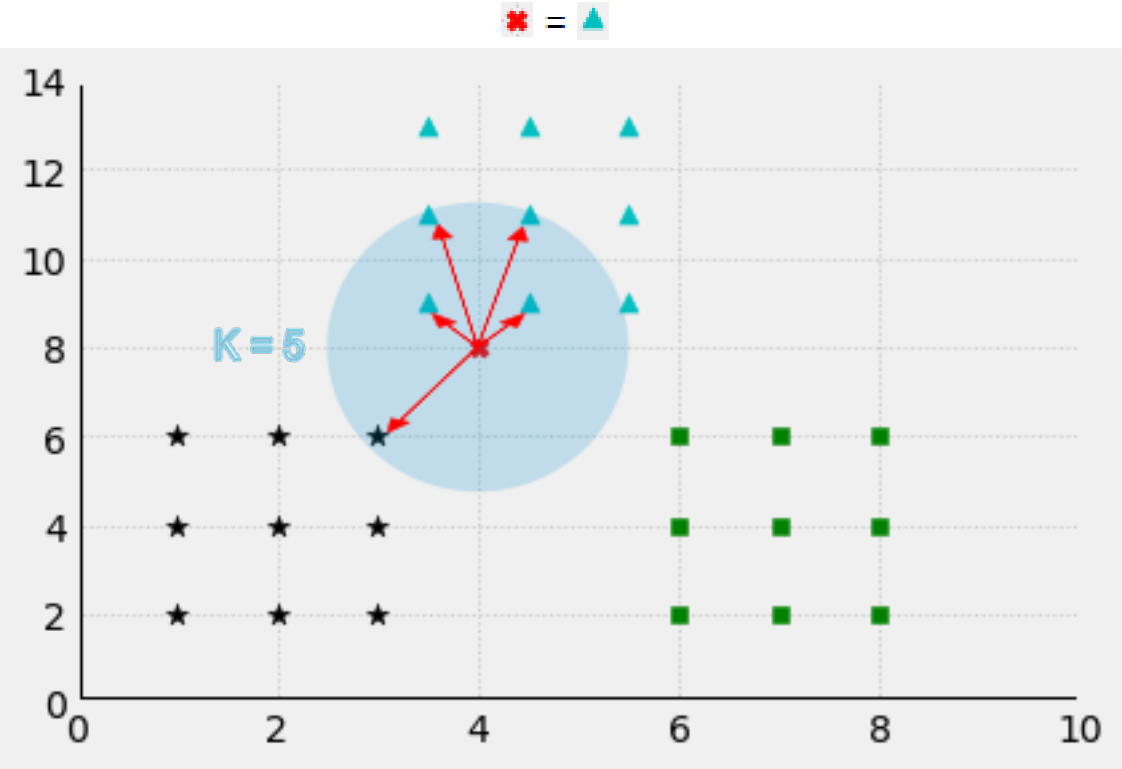
Machine Learning (ML)

- Machine Learning:
  - *One of the many methods that can perform Data Analytics*
- What is Machine Learning?
  - *In a simple description:* ***ML is based on the idea that systems can learn from data, identify patterns, and make decisions with minimal human intervention***
- Aerospace Community:
  - *ML is the popular buzzword in recent years*
  - *Hardly any tangible working examples have been shown*
  - *Can it practically assist with large size data issues or a buzzword that eventually fades away?*

Machine Learning (ML)

- A ML technique: K - Nearest - Neighbours (KNN)
  - *Considered by many as an entry level ML technique*
  - *A supervised learning algorithm (Features and Labels)*
  - *Classifies the given dataset and learns patterns discovered*
  - *With learned knowledge perform classification on a given "Unseen" observation*
- KNN, Pros and Cons:
  - ***Pros:*** *No assumptions about dataset, Simple algorithm, Relatively high accuracy, and Versatile*
  - ***Cons:*** *Computationally intensive, High memory requirement, Stores all learned data, Slow when dataset is huge, and sensitive to irrelevant features*

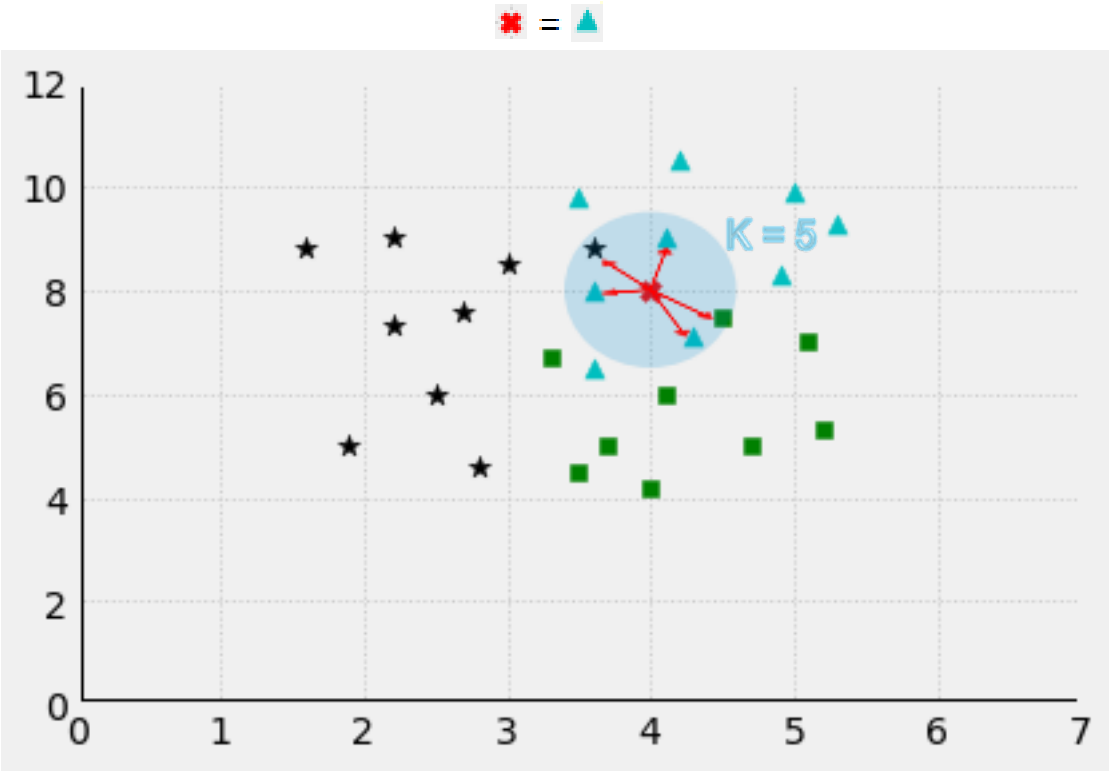
How Does KNN Work? (Linear Dataset)



Distances to the nearest five points can be calculated using:

- *Manhattan distance*
- *Minkowski distance*
- *Euclidean distance (Most Widely Used for KNN)*

How Does KNN Work? (Non-linear Dataset)



WDA Lab Main Gearbox (MGB) Records for KNN

- Database creation:
  - *Converted 307 ADF Helicopter MGB WDA pdf reports*
  - *Each database entry contains: Size, Quantity, Shape, Debris Compositions, and Typical Source*

Size: Very Large (1000+ µm) = 4, Large (500-1000 µm) = 3, Medium (250-500 µm) = 2, small (100-250 µm) = 1	**Quantity:** Extreme (100+ Particles) = 4, Large (26 to 100 Particles) = 3, Few (6 to 25 Particles) = 2, Small (1 to 5 Particles) = 1	Curlicue = 1, Flake (silver) = 2, Chunk = 3, Strand (wire) = 4	**% Ferrous debris (SEM, EDS, and XRF):** Titanium (Ti), Vanadium (V), Chromium (Cr), Manganese (Mn), Iron (Fe), Nickel (Ni), Molybdenum (Mo), Silicon (Si), Cadmium (Cd)									**% Non-Ferrous debris:** Silver (Ag), Chlorine (Ci), Tantalum (Ta), Calcium (Ca) Carbon (C), Sulfur (S), Phosphorus (P), Copper (Cu), Lead (Pb), Tin (Sn), Magnesium (Mg), Antimony (Sb), Zinc (Zn), Aluminium (Al)													**From System Metal Map:** Bearing cage metal = 1, Gear steels = 2, Silver plating metal = 3, Not part of Metal Map = 4
Size	**Quantity**	**Shape (Morphology)**	**Ti**	**V**	**Cr**	**Mn**	**Fe**	**Ni**	**Mo**	**Si**	**Cd**	**Ag**	**Ci**	**Ta**	**Ca**	**C**	**S**	**P**	**Cu**	**Pb**	**Sn**	**Mg**	**Sb**	**Zn**	**Al**	**Typical Source**
4	2	4	0.00	0.30	2.80	0.60	95.40	0.00	0.90	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	4
2	1	2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3
3	3	2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3
4	2	2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3
4	1	4			19.20	0.00	60		0.70	0.70			0.00	0.00	0.00		0.00	0.00	0.00			0.00	0.00		0	4
3	1				00	0.80			0.00	0.00			00	0.00	0.0		0	0.00	0.00			0.00	0.0			4

Features

Labels or Classes

Setup KNN Analysis

- Quarantine three data entries for trial purpose:

Size	Quantity	Shape (Morphology)	Ti	V	Cr	Mn	Fe	Ni	Mo	Si	Cd	Ag	Ci	Ta	Ca	C	S	P	Cu	Pb	Sn	Mg	Sb	Zn	Al	Typical Source
Silver is commonly used as a plating on bearing cages																										
2	1	2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3
EX53 is used for pinions of MGB																										
4	1	2	0.00	0.00	1.10	0.50	87.70	2.10	5.10	1.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.30	0.00	0.00	0.00	0.00	0.00	0.00	2
AISI 4130 steel is found in nuts, liners, miscellaneous components, cages of most critical bearings																										
4	1	4	0.00	0.00	1.10	0.80	98.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1

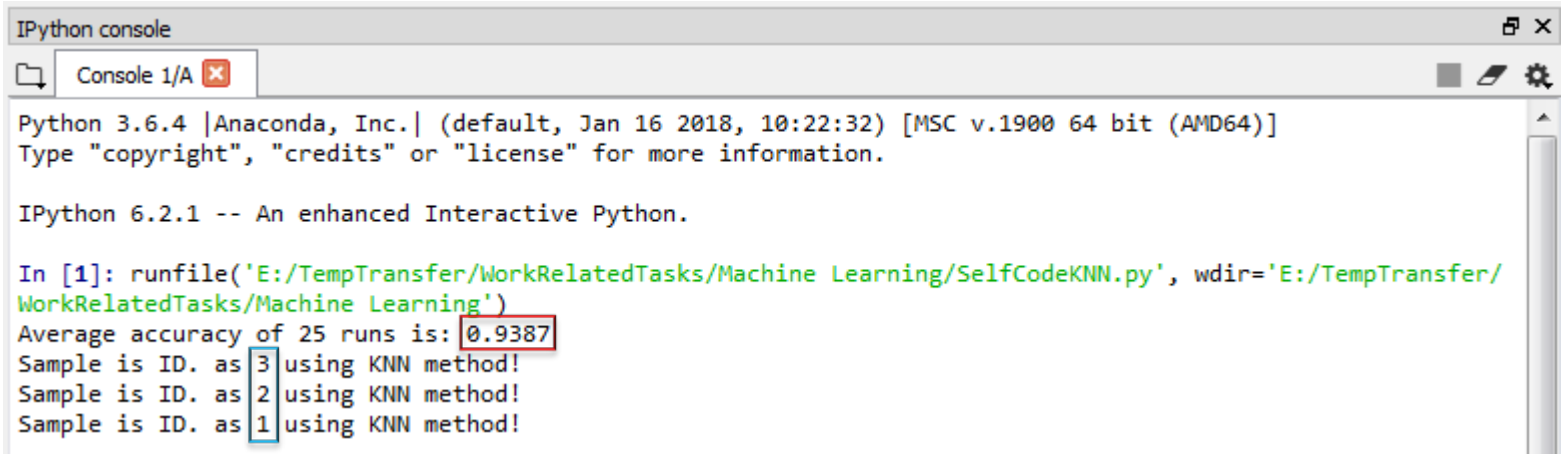
- Database (304 entries) split for training and testing:
  - Randomly select 80% of the database data for training
    - Learning the patterns and classifying the database
  - Rest of the 20% data for testing
    - From knowledge learned in the 80% of data, test the knowledge with rest of the 20% data for accuracy
- Test the learned knowledge on unseen data entries
  - The three quarantined data entries

Wear Debris KNN Analysis Results

- Three Unseen entries:

Size	Quantity	Shape (Morphology)	Ti	V	Cr	Mn	Fe	Ni	Mo	Si	Cd	Ag	Ci	Ta	Ca	C	S	P	Cu	Pb	Sn	Mg	Sb	Zn	Al
Silver is commonly used as a plating on bearing cages																									
2	1	2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EX53 is used for pinions of MGB																									
4	1	2	0.00	0.00	1.10	0.50	87.70	2.10	5.10	1.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.30	0.00	0.00	0.00	0.00	0.00	0.00
AISI 4130 steel is found in nuts, liners, miscellaneous components, cages of most critical bearings																									
4	1	4	0.00	0.00	1.10	0.80	98.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

- KNN Determined:



Summary

- Real world case has been tested
- ML (KNN) appears to have great potential
- Other types of real world example should be tested (those that have potential to benefit ADF)

Next Step

- Investigate ML techniques (such as Support Vector Machine, SVM) that can handle database in Terabytes.

Questions?