# Finding EKRAN-2 debris using Random Forrest and K-Nearest Neighbors

VU Hoang Long, Hongru CHEN, and Toshiya HANADA
Kyushu University, Fukuoka, Japan

July 23, 2023

KYUSHU UNIVERSITY

# Overview

| Introduction | Method | Result | Summary |
| --- | --- | --- | --- |
| • Objective<br>• Reasons<br>• Background | • Principle of ML<br>• Data Generating<br>• Evaluating Indexes | • Scores<br>• Graphs<br>• Discussion<br>• Comparision | • Conclusion and Insight<br>• Future work<br>• Acknowledgement<br>• Reference |

# Introduction

# Objective

Design a Machine Learning Model to find EKRAN-2 debris fragments from observational data

- What is EKRAN-2? And their debris fragment?
- Why go to find them?
- Why machine learning?

# EKRAN-2 Satellites

- Soviet-Russian Geostationary satellited launched in September 1977.
- Used for the world first direct-to-home TV service.
- Lost contact in June 1978.
- Explosion revealed in 1992.

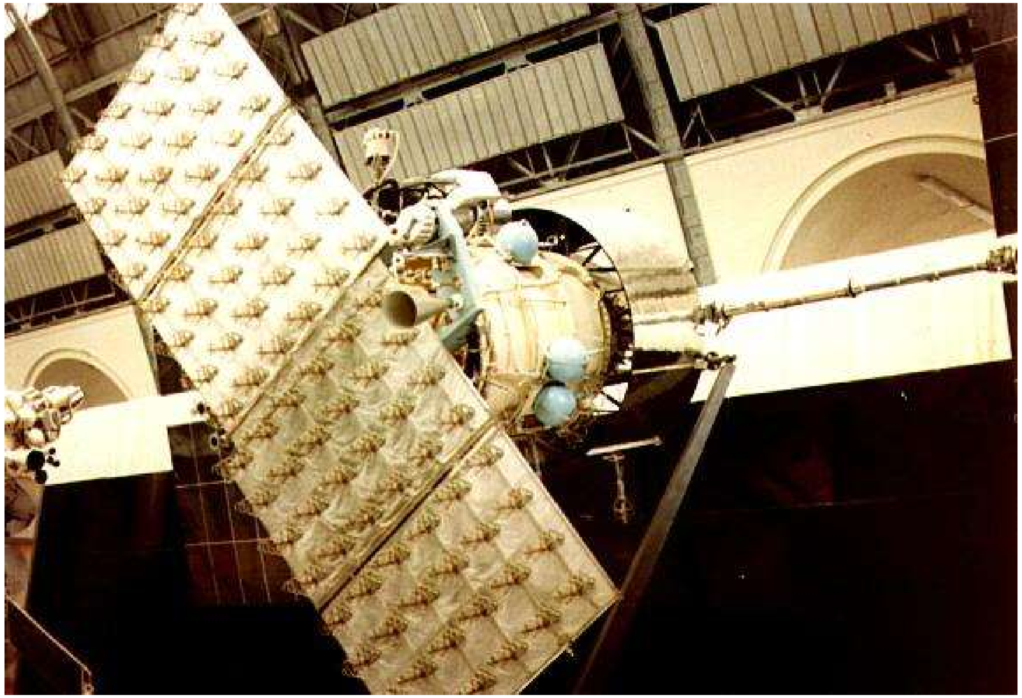


Figure 1. EKRAN satellite

http://www.astronautix.com/e/ekran.html

# EKRAN-2 debris

- A nickel-hydrogen battery explosion.
- Contribute more debris in GEO.
- Destroyed during longitude control.

  ⇒More difficult to analyze fragments.
- Four fragments are being tracked.
- Lose sight of fragment 1977-092L.
- There are still unknown fragments.

| INTLDES | OBJECT_NAME |
|---|---|
| 1977-092A | EKRAN-2 |
| 1977-092H | EKRAN-2 DEB |
| 1977-092J | EKRAN-2 DEB |
| 1977-092K | EKRAN-2 DEB |
| 1977-092L | EKRAN-2 DEB |

Figure 2. EKRAN-2 fragments International Designators

⇒ Need to find them.

# General debris identification method

1. Determine the position of an observed moving object.
2. Propagate orbits of various objects.
3. Compare the observed and predicted values.
4. Object is identified if the difference is less than 10 arc minute.

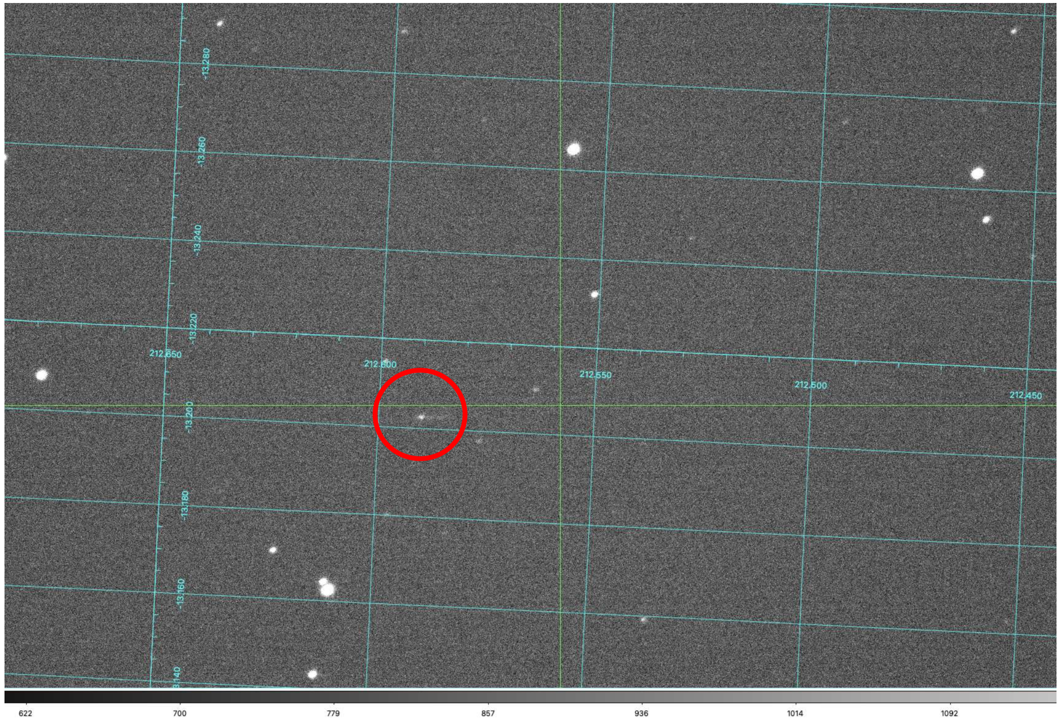


Figure 3. Fragment 1977-092J observed.
The crosshair shows propagated position.

# Problem with current method

- Take a lot of time and calculating power.
  - EACH observed object requires many orbits' propagation.
  - 30000 objects recorded every night.
- Can have multiple or none propagated orbit satisfy the condition.
- Not a good method if you are interested in only one specific object.
  - Need to go through all observational data with different TLE.
  - Will not help if we are looking for unknown object with limited TLE knowledge.

# A different method is necessary.

- New method’s goals:
  - Able to quickly filter through observational data.
  - Give highly accurate predictions.
  - Can self-improve.

An algorithm can learn from itself?

# Why machine learning?

- Machine Learning
  - A subset of Artificial Intelligence.
  - Machine learn from data and experience without explicit teaching.
  - Getting more attention recently.
  - Have application in prediction and classifications

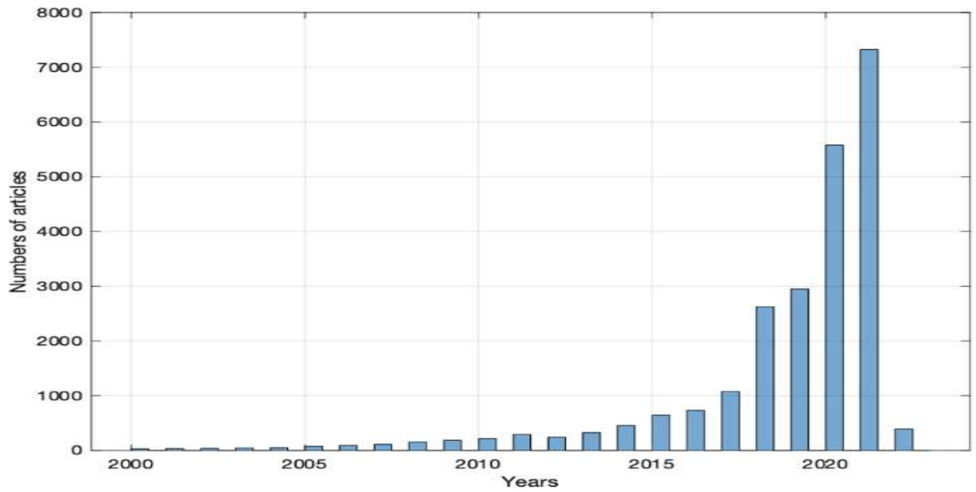
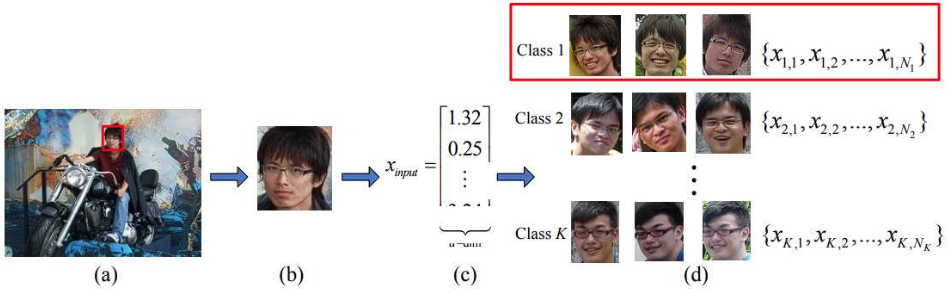
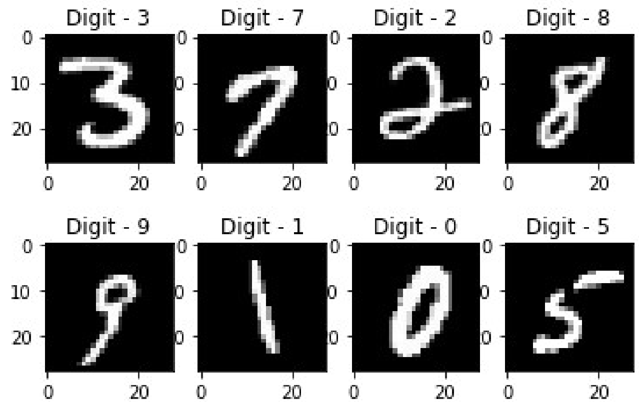



Figure 4. Numbers of ML research increases fast in recent year

- Practice example:




http://disp.ee.ntu.edu.tw/~pujols/Face%20Recognition-survey.pdf




https://medium.com/@himanshubeniwal/handwritten-digit-recognition-using-machine-learning-ad30562a9b64

# Method

# Section Overview

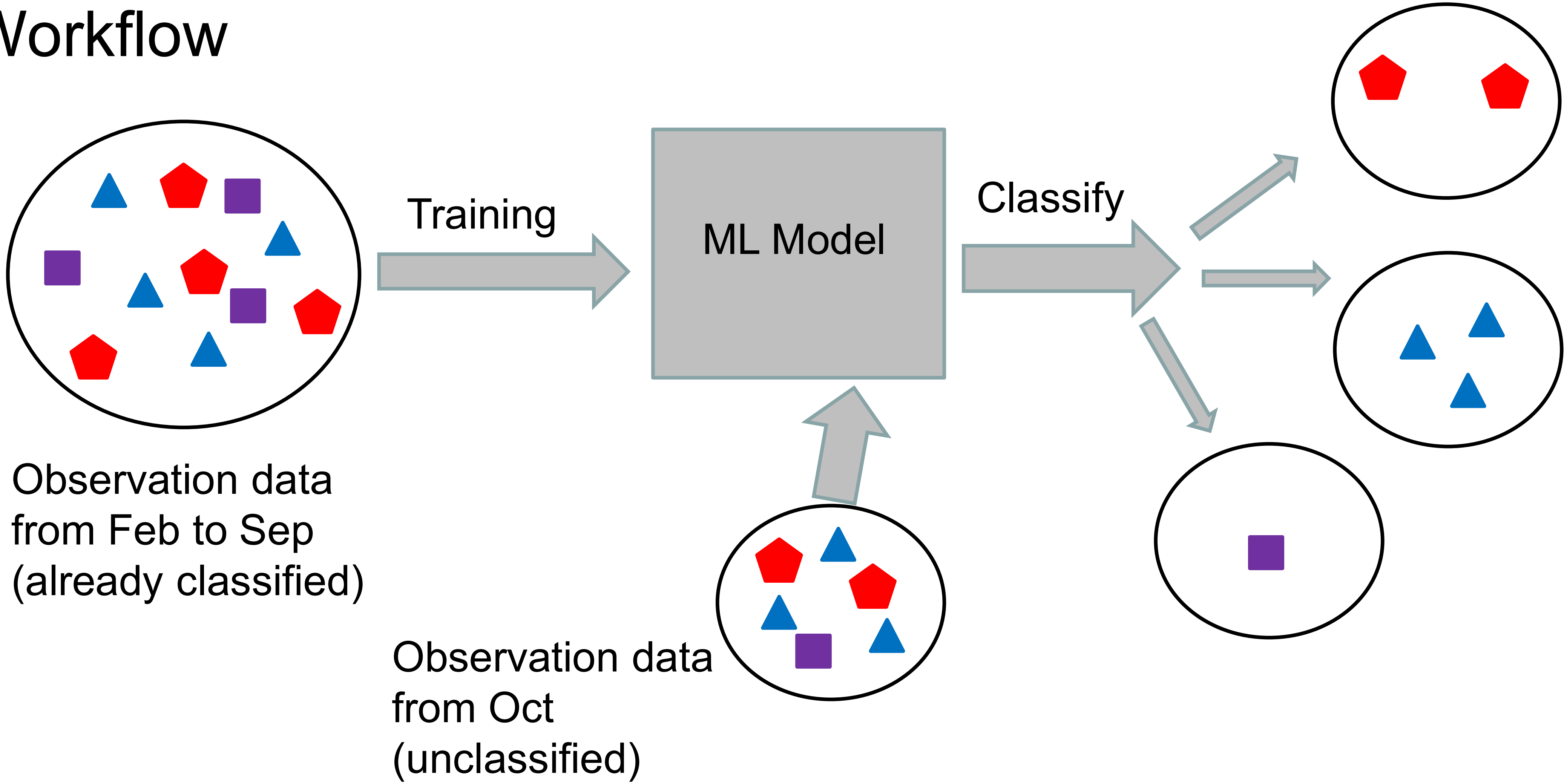
- Method's demonstration.

- Two ML models:
  - K-Nearest Neighbors.
    - Principle
    - Advantage
    - Miscellaneous
  - Random Forest.
    - Principle
    - Advantage
    - Miscellaneous

- Data Generating
  - Training
  - Evaluating

- Evaluating Indexes

# Workflow


Training
ML Model
Classify
Observation data
from Feb to Sep
(already classified)
Observation data
from Oct
(unclassified)

# First ML model: K-Nearest Neighbor

# K-Nearest Neighbor

- Use proximity known points to classify.
- Assume that similar points are near each other.
- Major vote dictates the label.

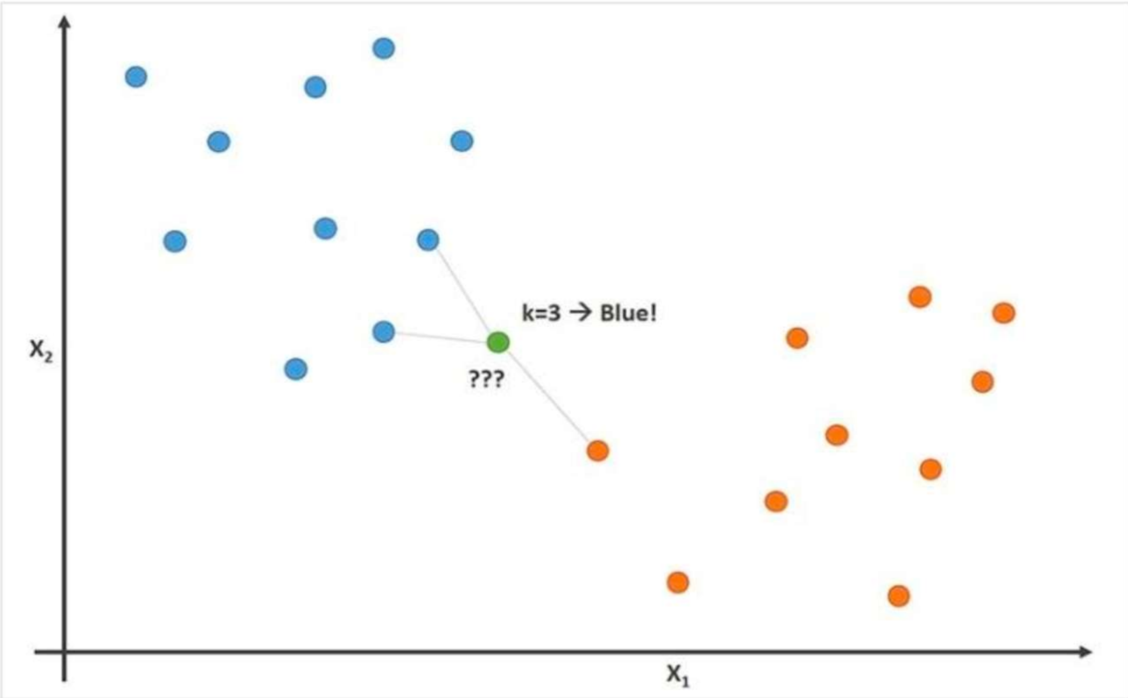



Figure 5. Green point is labelled as blue since 2 out of 3 nearest labelled points are blue.

# Distance Function

- To determine which points are closest to a given query point.
- Manhattan and Euclidean distance functions are studied.

Euclidean:

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (y_i - x_i)^2}$$

Manhattan:

$$d(x,y) = \sqrt{\sum_{i=1}^{n} |y_i - x_i|}$$

$y, x$: two different points
$i$: different feature

# Advantages of K-NN

- Simplicity
  - Easy to interpret algorithm
  - Two parameters to tune
- Fast run time
- Reasonable accuracy

# Miscellaneous

- Sensitive to the scale of variables

- Example

Euclidean distance function:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(y_i - x_i)^2}$$

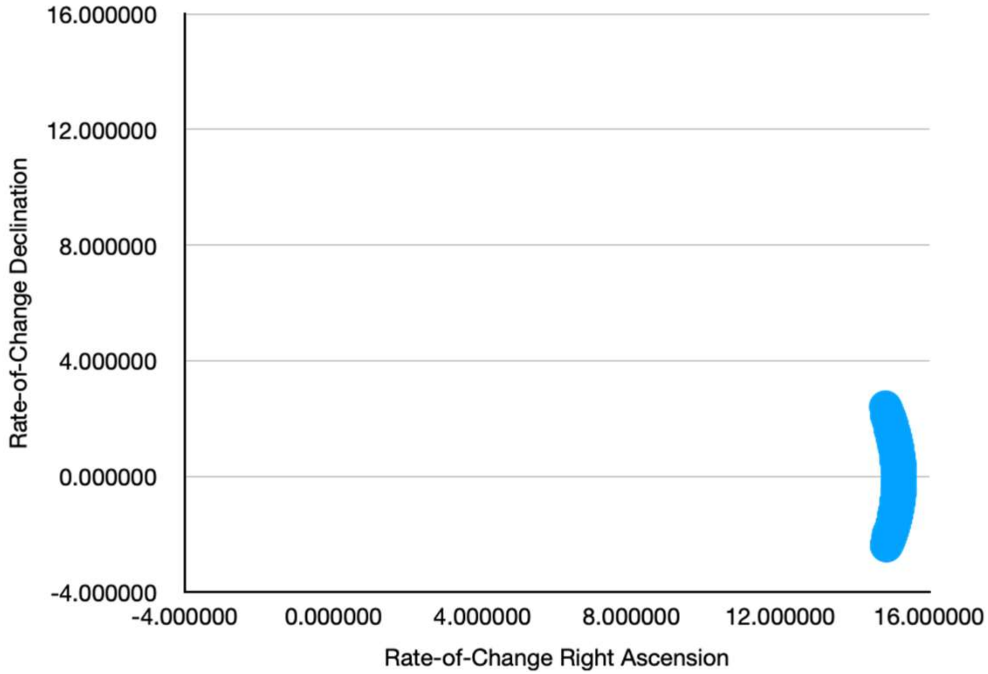



Figure 6a. Unscaled data from 1977-092J

# Data Scaling

- Make value range from 0 to 1.
- Make variables distribute around their mean value.

Min-Max Scaling

$$\bar{x} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

$x$: original value
$\bar{x}$: scaled value

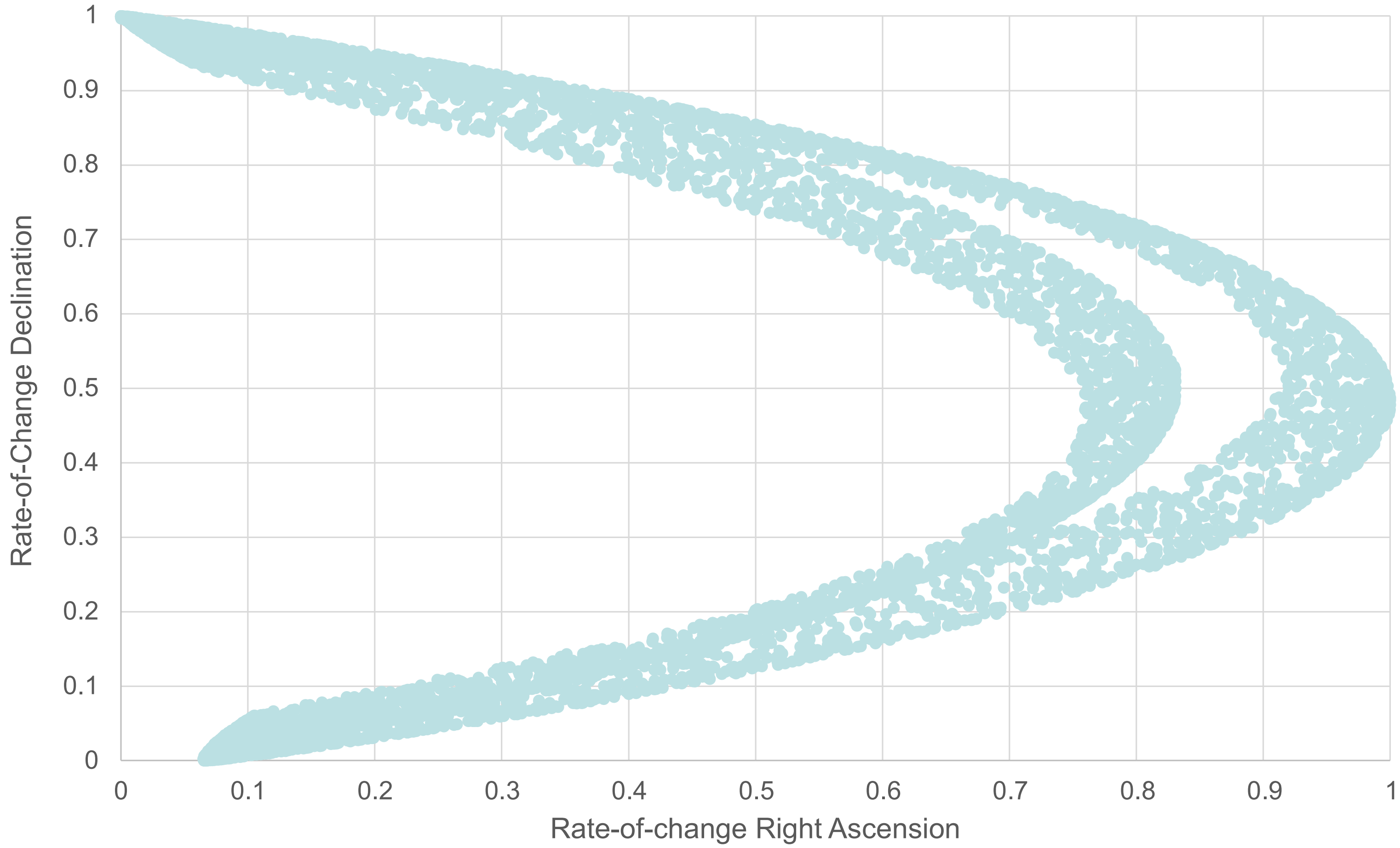



Figure 6b. Scaled data from 1977-092J

# Second ML model: Random Forest

# Random Forest

- A tree-family method.
- Consist of many decision trees.
- Make decisions based on the purity/chaotic of dataset.

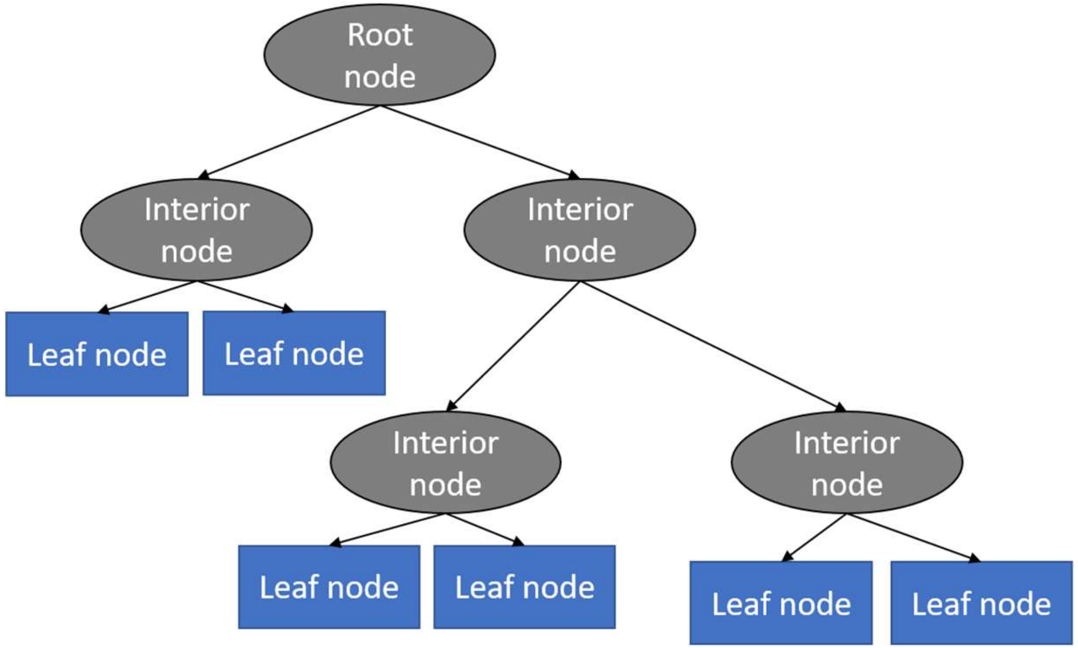



Figure 7. A simple Decision Tree

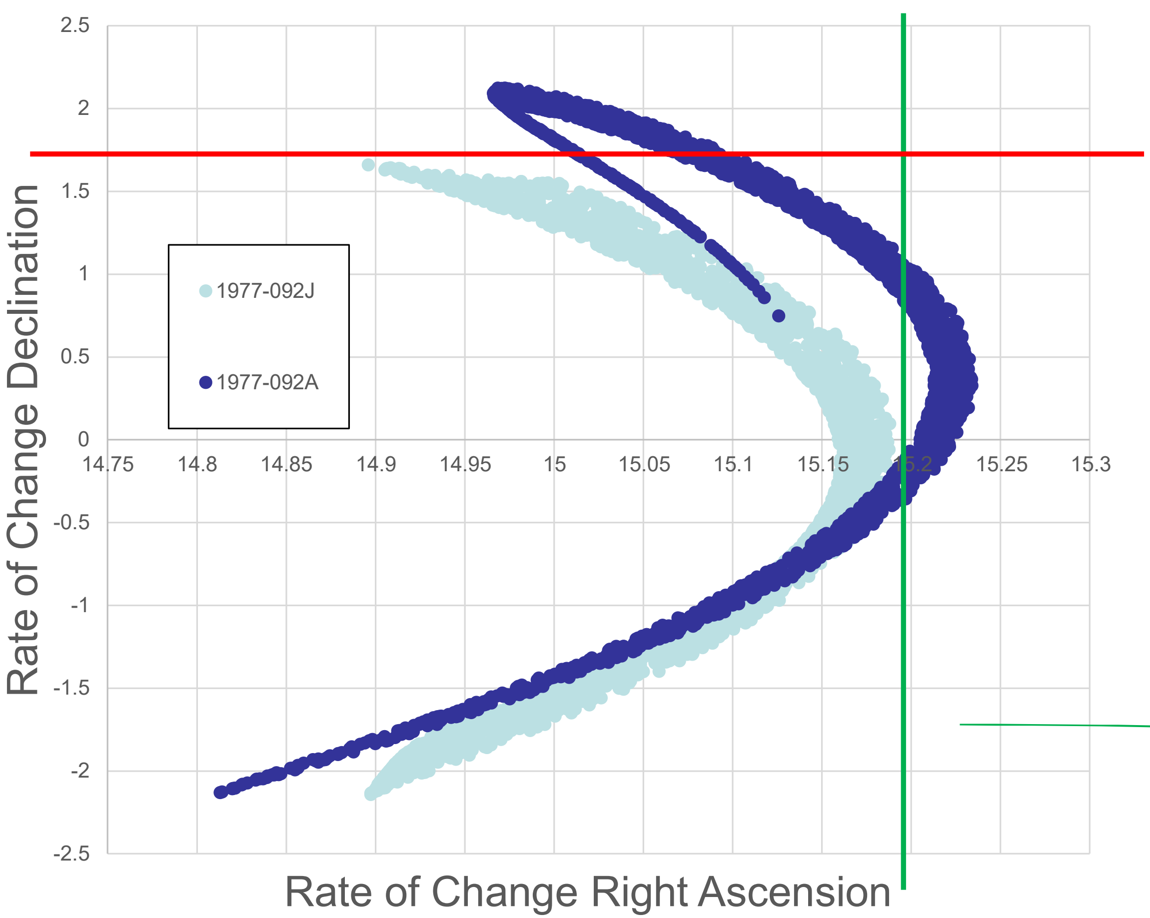
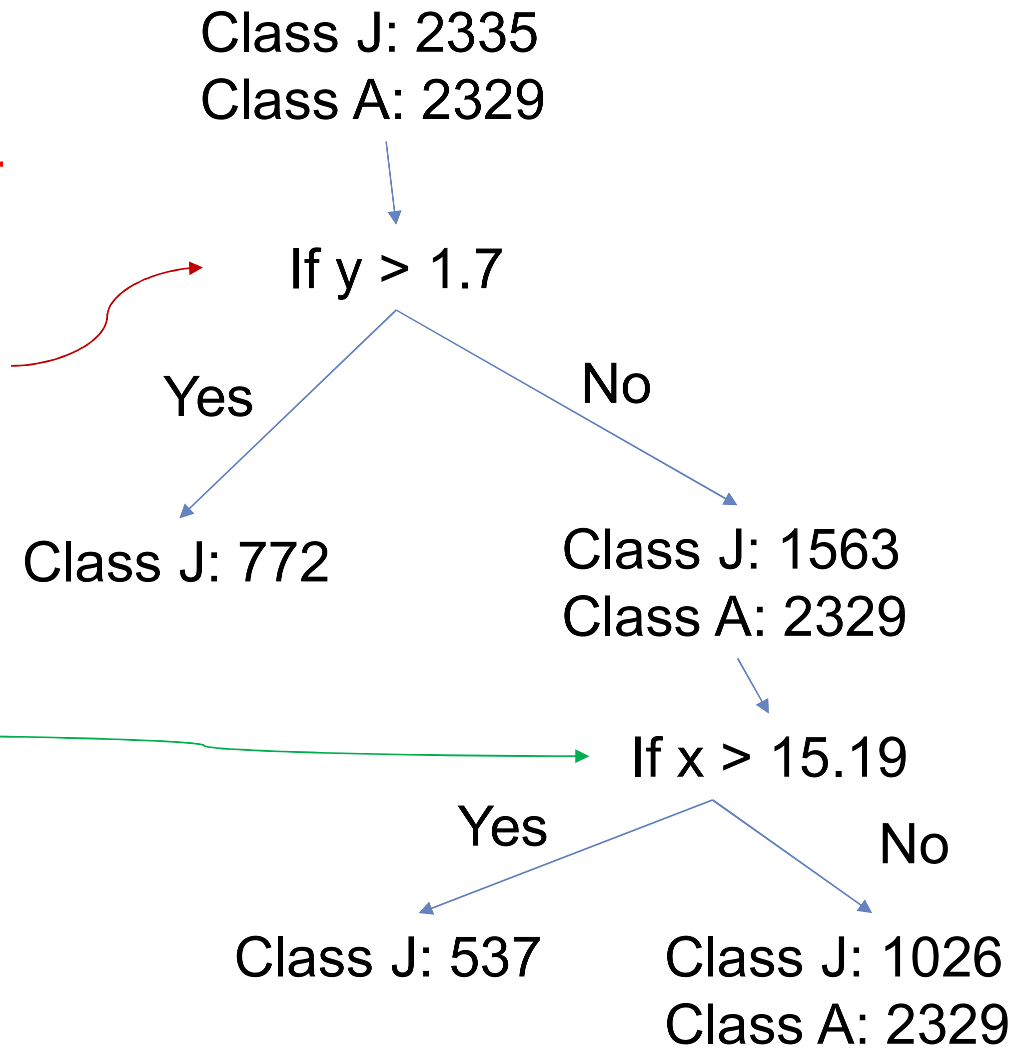



Figure 8. A simplified classification example between 1977-092J and 1977-092A



*So how can decision tree know how to pick condition?*

- Entropy:

$$E = \sum -p_i \log_2 p_i$$

$p_i$ is possibility of class $i$

- Information gain:

$$IG = E_{parent} - \sum \omega_i E_{children}$$

$\omega_i$ is number of points

→ Maximize IG

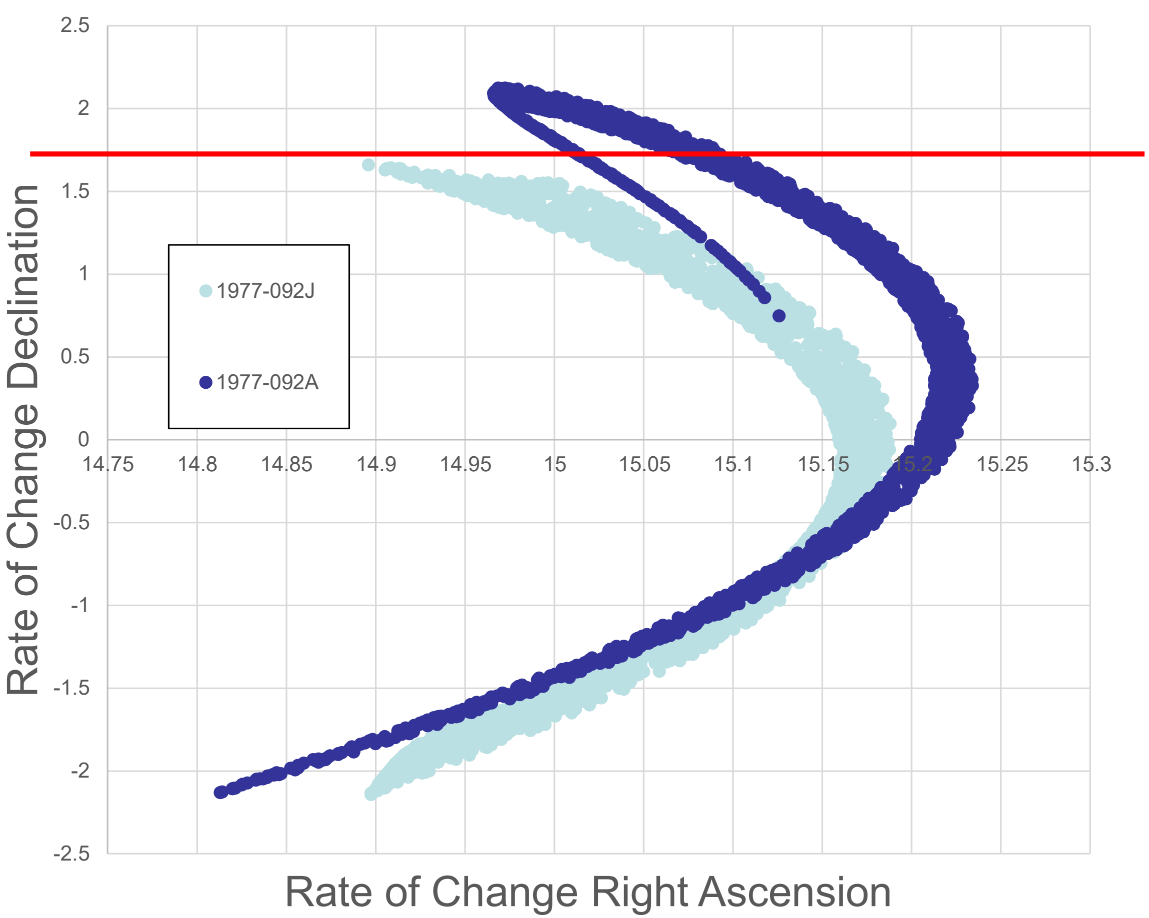



Figure 8. A simplified classification between 1977-092J and 1977-092A

Class J: 2335
Class A: 2329

$E = 0.9999$

If y > 1.7

Yes

Class J: 772

$E_1 = 0$

No

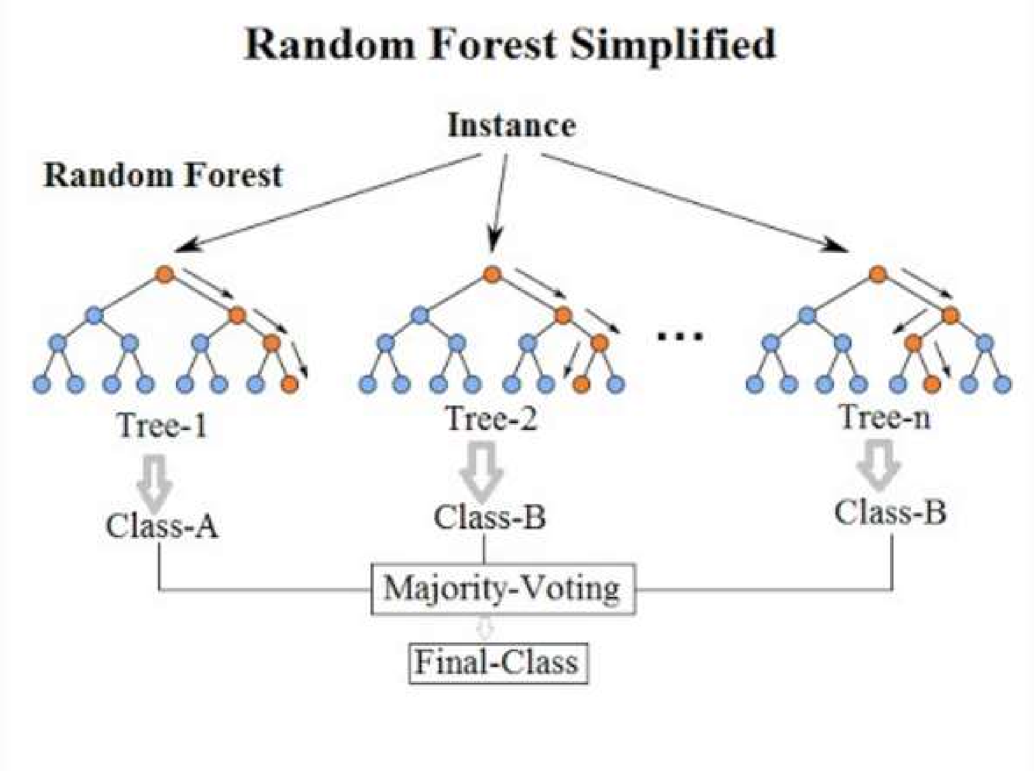
Class J: 1563
Class A: 2329

$E_2 = 0.972$

IG = 0.188

# Advantage of Random Forest

- High precision
- Robust to noise and errors
- Two parameters to tune
- No need to scale data




https://community.tibco.com/wiki/random-forest-template-tibco-spotfirer-wiki-page

# Evaluating Indexes

- Precision: The confidence of predictions

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

- Recall: The sensitivity of model

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

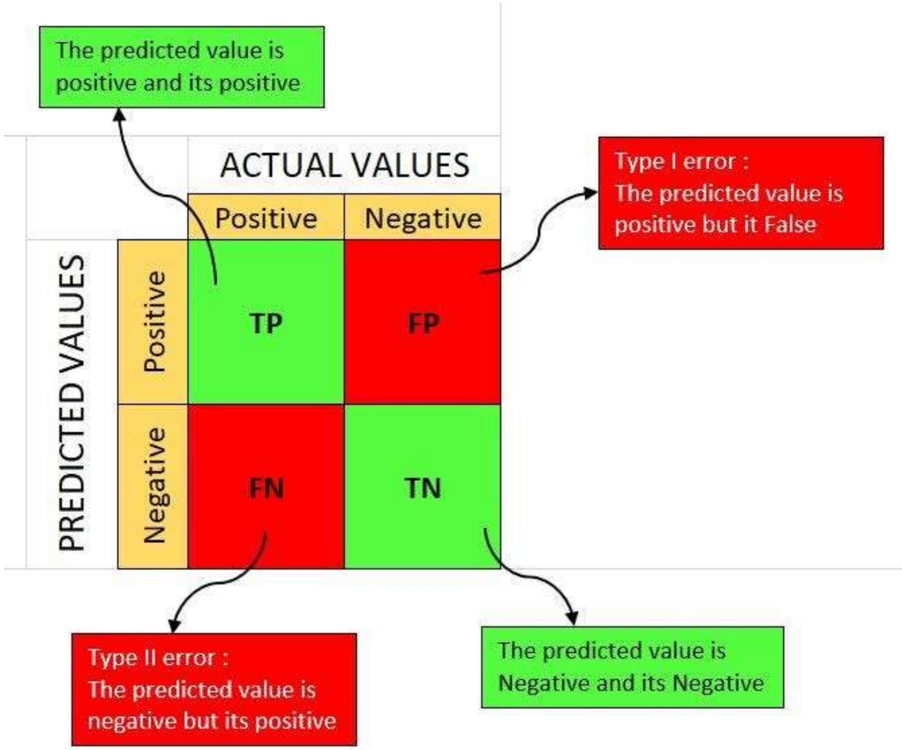



Figure 9. Confusion matrix

# Data generating

- Simulated from TLE files of Ekran-2 debris and 50 random satellites.

- Observatory is Kiso, Japan:
  35.7972° N, 137.6254° E

Figure 10. Kiso Observatory

TLE files → SGP4 → TEME coordinates → ECI to Topocentric →

TEME coordinates: $time, x, y, z, v_x, v_y, v_z$

$time$
$dec$
$RA$
$\dot{dec}$
$\dot{RA}$

*Like how observation data is recorded*

- Sample satellites have similar inclination and period

- 210000+ samples

- 8 input features

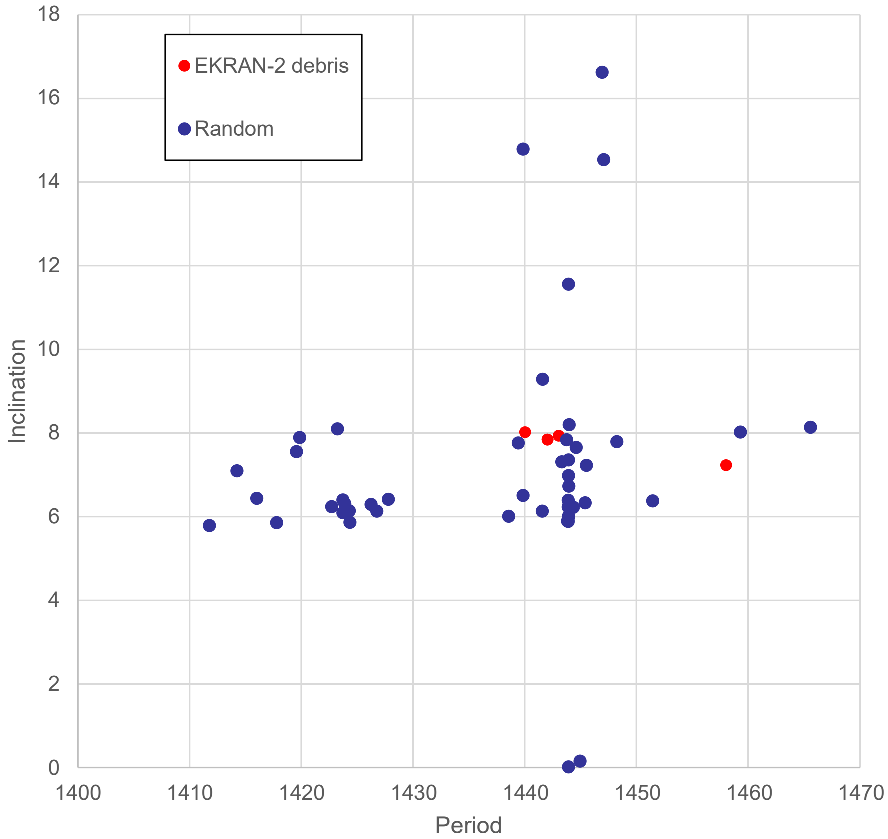



Figure 11. Orbital elements of sample satellites

# Training Set and Evaluating Set

| Set no. | Period of collected data | Number of collected data |
|---|---|---|
| T1 (Training set 1) | August to September | 31605 |
| T2 | July to September | 63152 |
| T3 | June to September | 92810 |
| T4 | May to September | 125691 |
| T5 | April to September | 156778 |
| T6 | March to September | 188472 |
| T7 | February to September | 217095 |
| E1 (Evaluation set) | September to October | 29326 |

# Result

# K-NN method:

- Best parameter search result:
  - N-number: varying between 5 to 9 (6)
  - Distance function: Manhattan

- Training run time:
  - Fixed parameter: < 30 sec
  - Random find best parameter: ~2 mins
    - n-number values range: 1 – 20

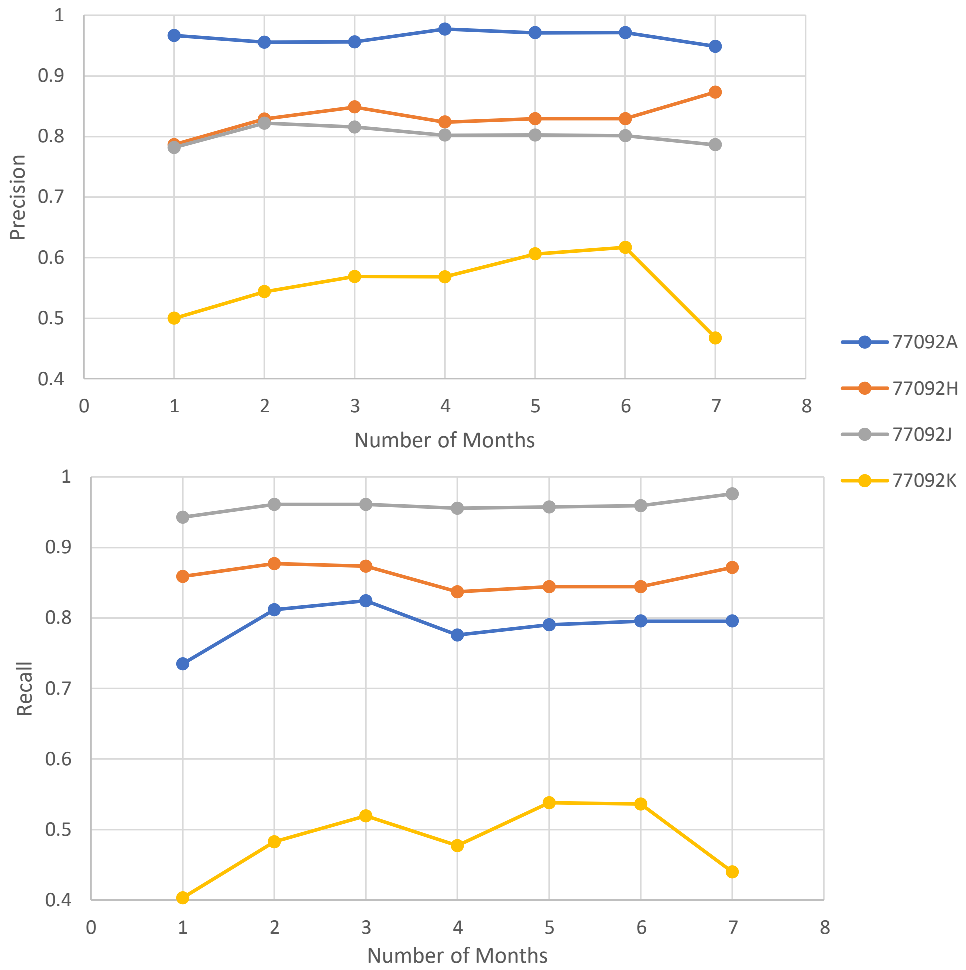



Figure 12. K-NN result

# Random Forest

- Best parameter search result:
  - Tree Depth: varying between 15 to 21
  - Number of Trees: ~210 trees

- Training run time:
  - Fixed parameter: 1 to 2 mins.
  - Random find best parameter: ~13 mins
    - Number of Trees range: 50 to 500
    - Depth values range: 1 – 25

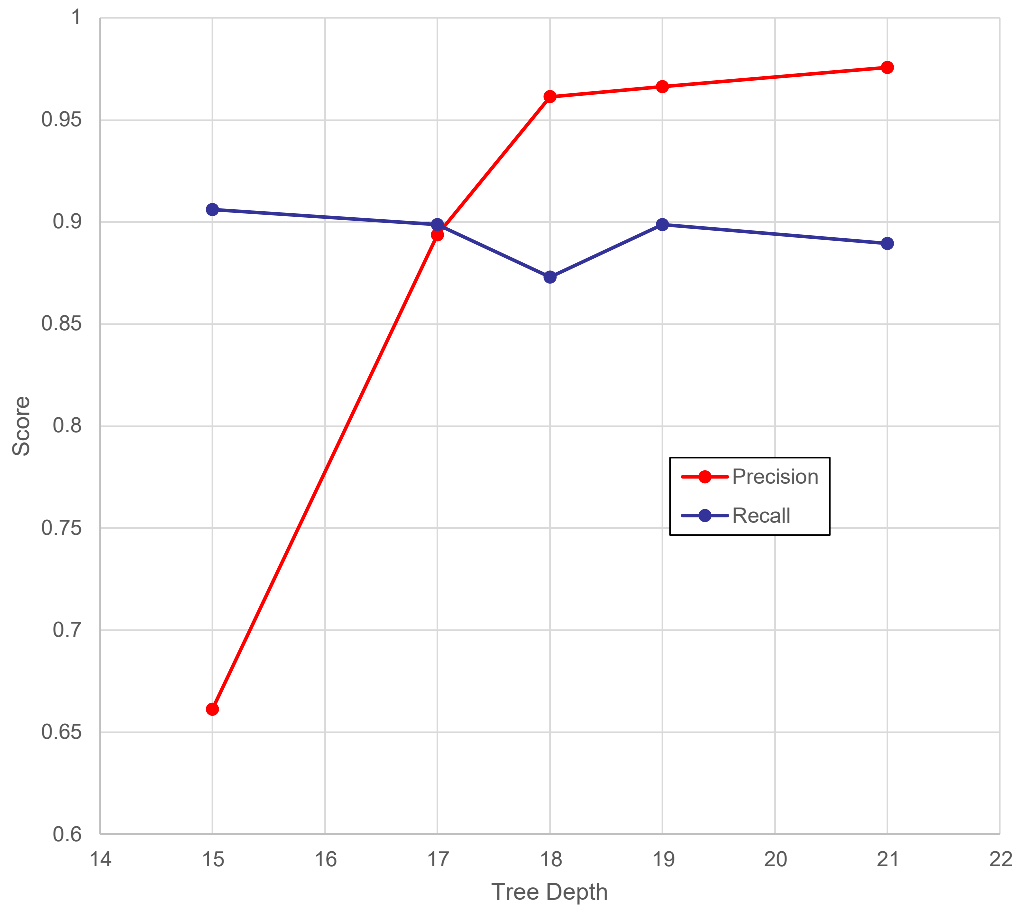
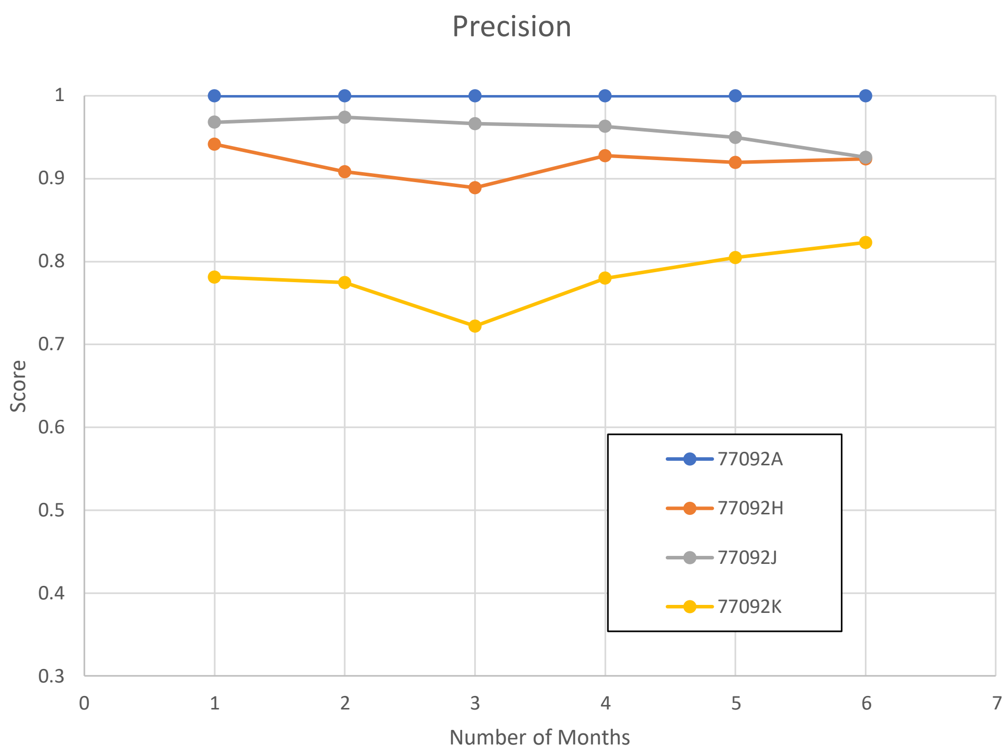



Figure 13. Effect of Tree Depth on result of 1977-092J debris fragment



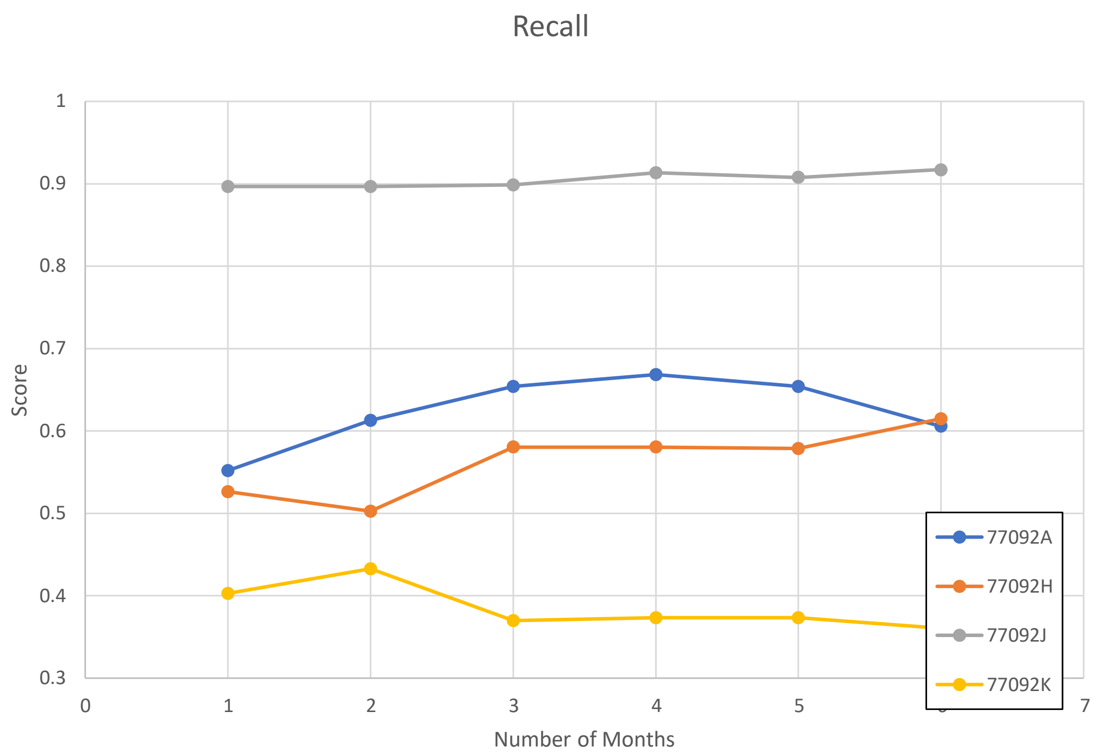




Figure 14. RF result
Tree Depth: 19 and 210 Trees

# Comparison

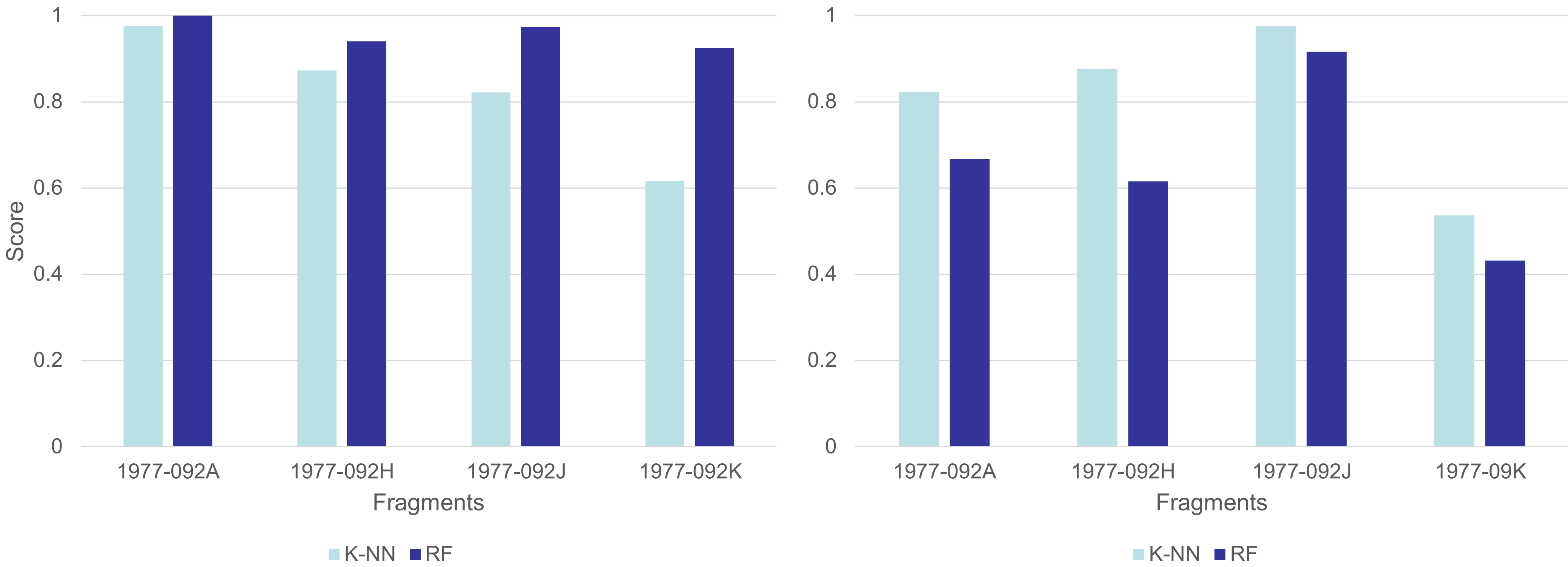



Figure 15. Comparison of Precision (left) and Recall (right) between two algorithms

# Summary

# Conclusion

- Both methods perform well and have their own strengths.
  - Better Precision: Random Forest.
  - Better Recall: K-NN.
  - Better Training time: K-NN.
- Precision score is often higher than Recall score.
- Increase number of months for training data is not always a net positive.

# Future work

## Objective

Using trained Machine Learning model to find unknown EKRAN-2 debris fragments from observational data

- Model is trained to identify fragments A, H, J, K.
- Delta-V of the explosion is small => Unknow fragments have similar orbital elements with know fragments.

# Acknowledgement

- Professor Sako and the team who working on Tomo-e Gozen System in Kiso Observatory.

# Reference


- *User guide: Contents*. scikit. (n.d.). https://scikit-learn.org/stable/user_guide.html

- Krebs, Gunter D. “Ekran (11F647)”. Gunter's Space Page. Retrieved July 24, 2023, from https://space.skyrocket.de/doc_sdat/ekran.htm

- Nagisa INOKUCHI, H. H. (2023). *Investigation into Fragmentation of the Russian EKRAN 2 Spacecraft.*

# Thank you for listening