Presented to:
RADS

Computer Learning Algorithm for Records Evaluation Overview

TECHNOLOGY DRIVEN. WARFIGHTER FOCUSED.

Presented by:

RAM Engineering & System Assessment Division

9/6/2017

Who is AMRDEC?

US Army Aviation and Missile Research, Development, and Engineering Center provides increased responsiveness to the nation's Warfighters through aviation and missile capabilities and life cycle engineering solutions.

- **Headquartered at Redstone Arsenal, AL**
- **5 Directorates**
- **9,000 scientist and engineers**
- **$2.45 billion in reimbursable funding, FY16**
- **$339 million in Science and Technology funding, FY16**

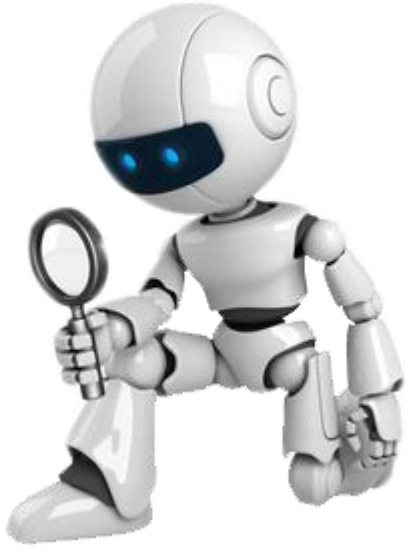
Strategic Readiness – provide aviation and weapons technology and systems solutions to ensure victory on the battlefield

Future Force – develop and mature Science and Technology to provide technical capability to our Army's (and nation's) aviation and weapons systems

Soldiers & People – develop the engineering talent to support both Science and Technology and materiel enterprise

Briefing Outline

- What is CLARE?
- Logic
- Road Map

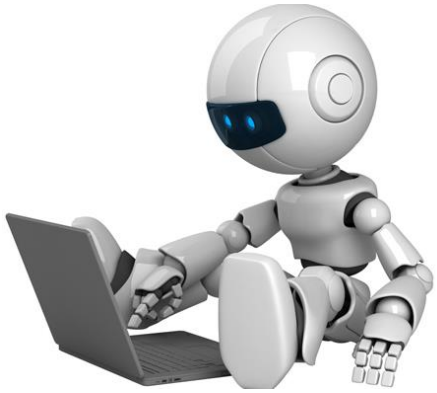


What is CLARE?

What is CLARE?

Computer
Learning
Algorithm for
Records
Evaluation

- **CLARE is an Supervised Interactive Machine Learning Algorithm (SIMLA) that learns records association by modeling past scored records. This is used to propagate scored information to blank records.**
- **This scoring propogation enables…**
  - **B grade work at high speed**
  - **Increased speed of A level work**
  - **Characterization of data**

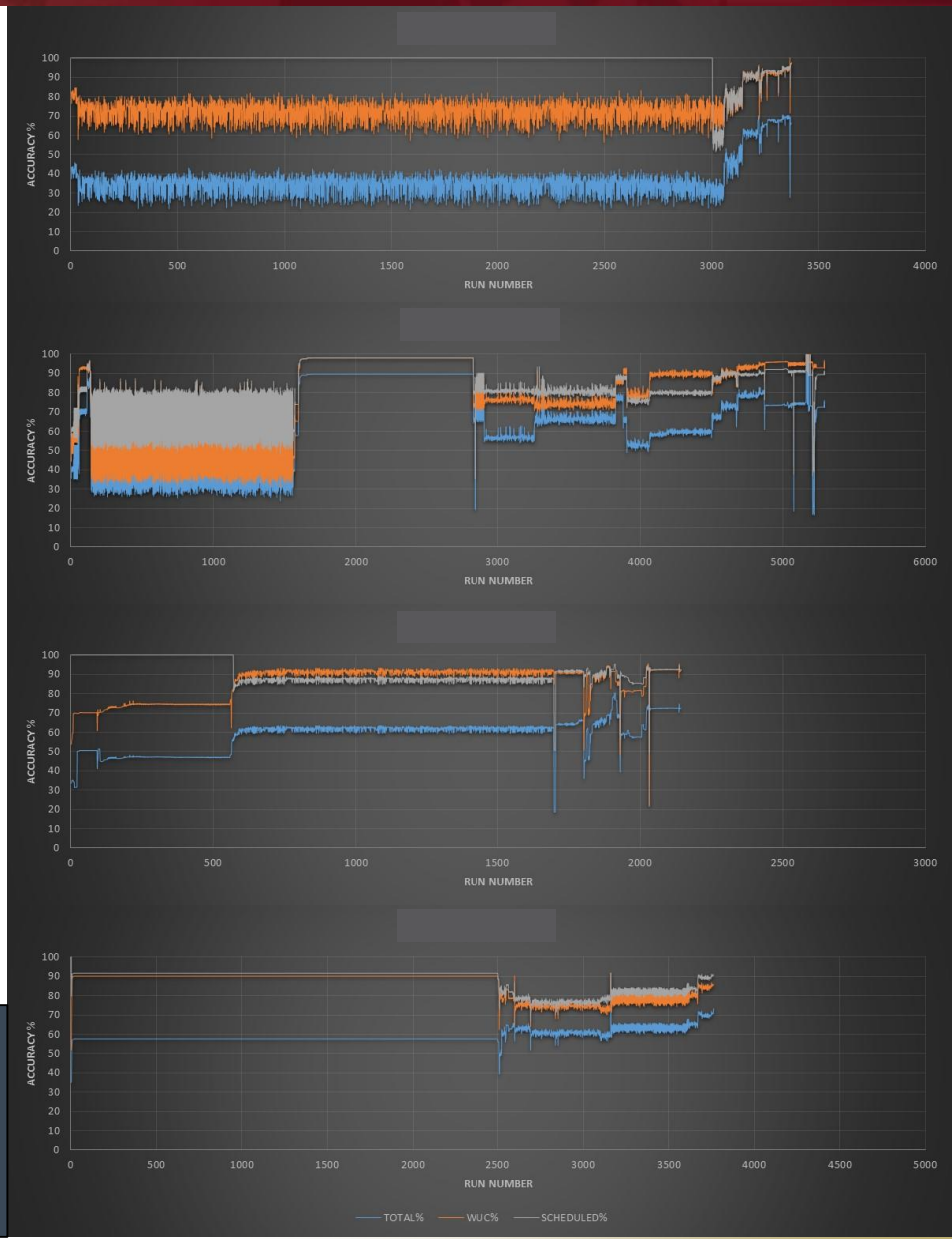


Because CLARE is continuously learning and changing her model to be more accurate, she requires little maintenance and allows for global changes can be made very quickly to the data.

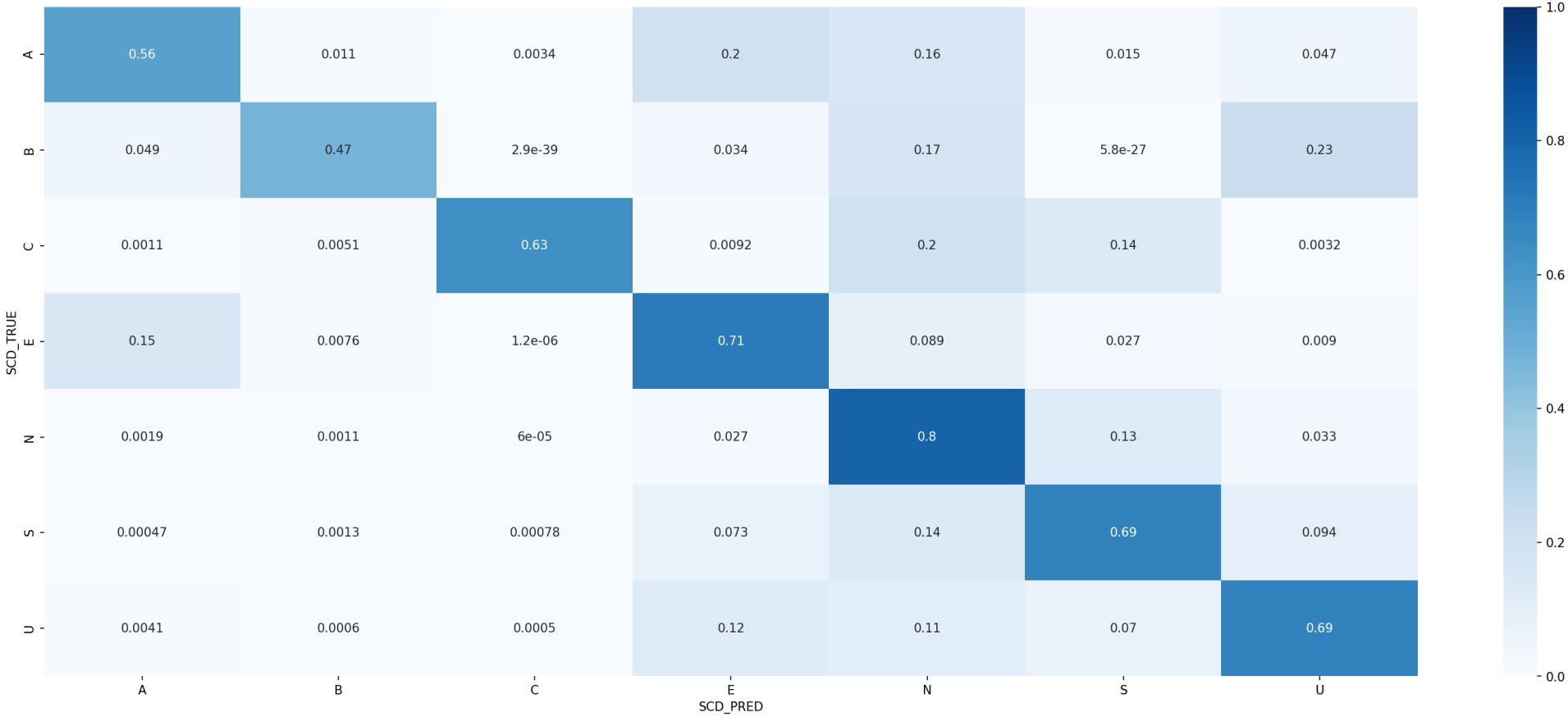
CLARE Development

- **CLAREs performance is measured in three fields:**
  - Weighted accuracy: sum(m=1 to N) (# of correct predictions of class m / # of class m)
  - Accuracy: # correct / # in sample
  - Per class metrics as appropriate
    - True positive rate: correct positives / real positives
    - True positive accuracy: correct positives / predicted positives

Accuracy:
Total %
WUC %
Scheduled %



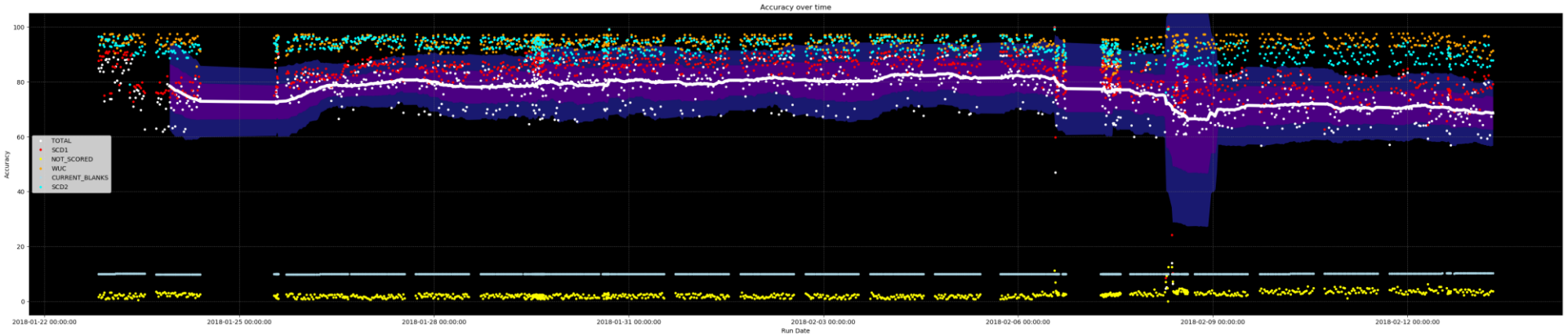
Weighted Accuracy: Confusion Matrix

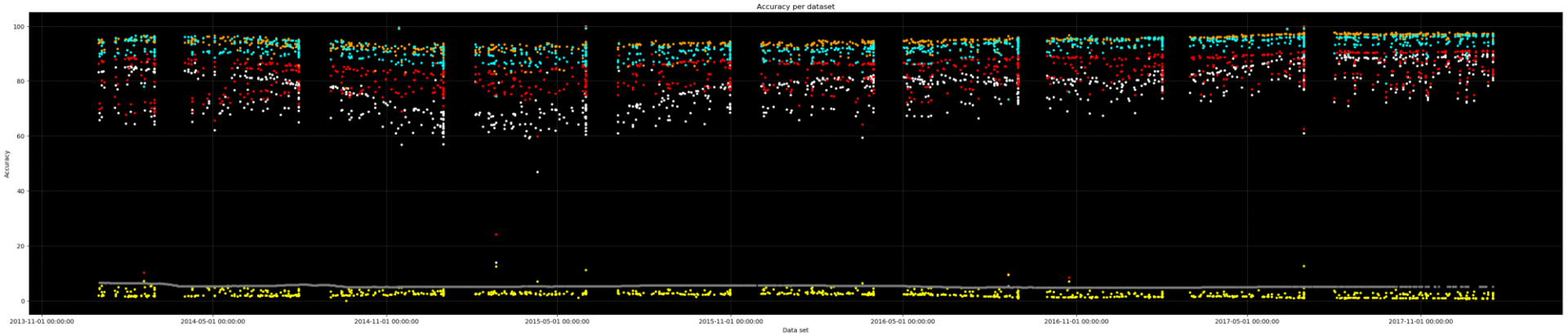


A confusion matrix is used for weighted accuracy and helps to determine the classes that often share traits.

A strong diagonal is an indication that your algorithm is on the right track and is doing well.

Accuracy

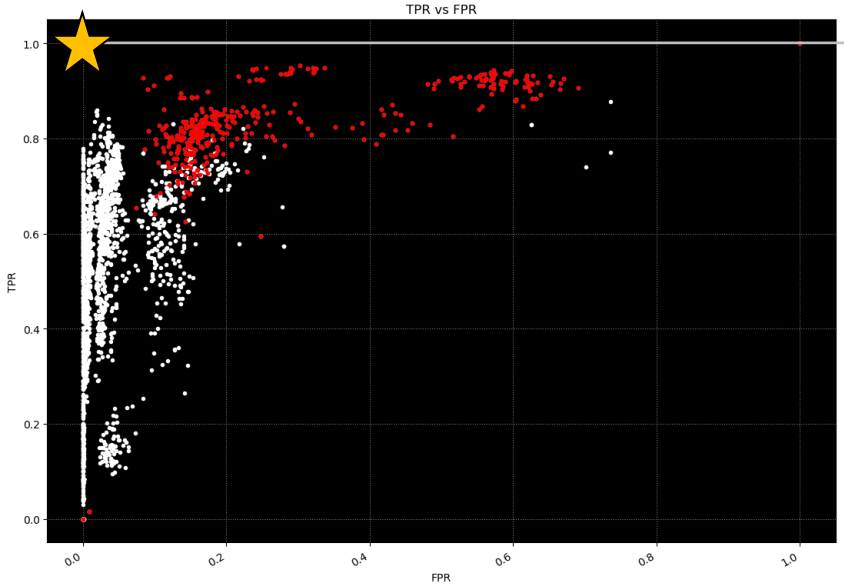


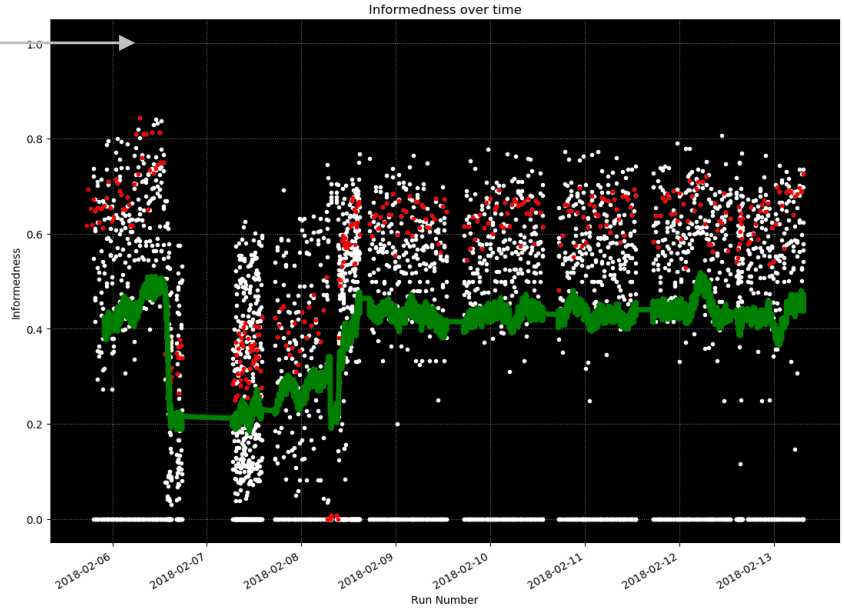


The white line/dots in the above charts shows the line by line accuracy of the model developed by CLARE.

The top is an accuracy measured by run date and the bottom is an accuracy measured by record date in the data sample.

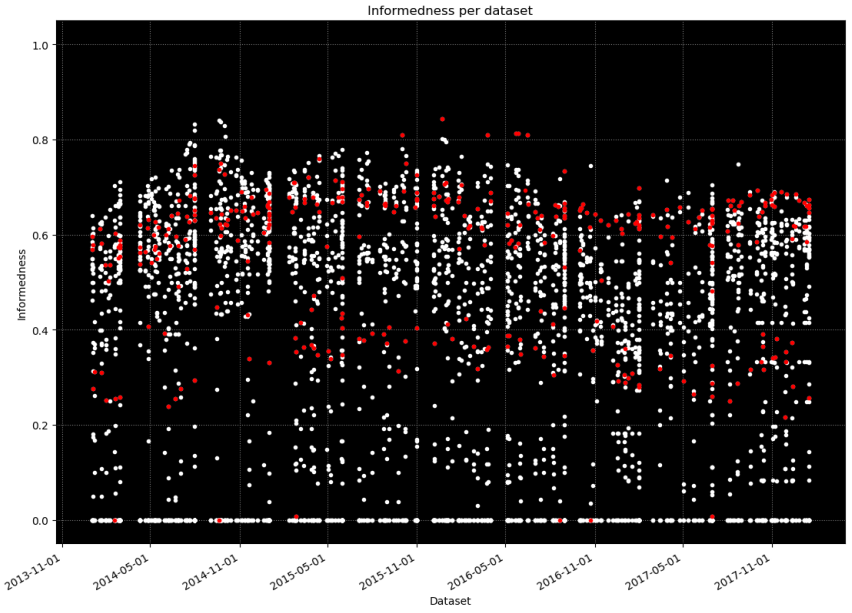
TPR, FPR, and Informedness





The True Positive Rate (FPR) and False Positive Rate (FPR) for each class is plotted in three distinct fashions.

These allow for interpretation of how the algorithm is learning and how much it can be determined to know about a particular class measured by informedness.



Logic and Progression

Logic

Label	Narrative
Apple	The apple fell from the tree.

Removing Non-Descriptive words

Apple	apple fell tree

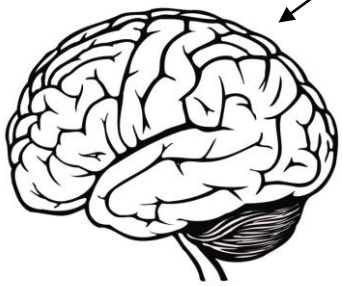
- System
- Next Higher Assembly
- Failure Mode

Logic

Label	Narrative
Apple	The apple fell from the tree.

Removing Non-Descriptive words

Apple	apple fell tree



Database

Logic

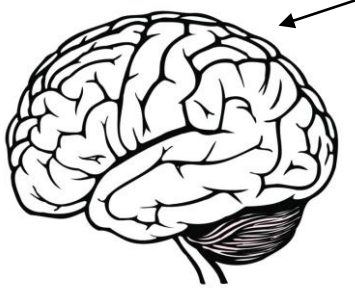
Label	Narrative
Apple	The apple fell from the tree.

Removing Non-Descriptive words

Apple	apple fell tree

New unlabeled data

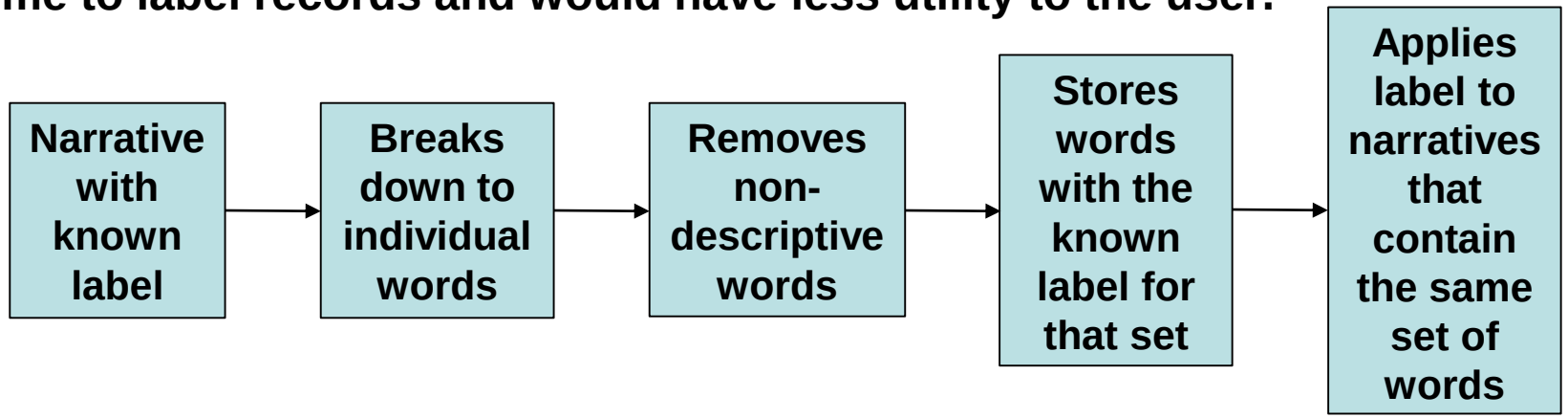
Apple	The red apple fell from tree number 7.



Database

Current Learning Process (Simplified)

The current method of learning is limited due to the application and limitations of hardware. A more complex algorithm would consume more time to label records and would have less utility to the user.



The above process is iterated and a log of the True Positives and False positives is kept for each "description" of a label. For each iteration, only the label with the highest word count, highest True Positive Accuracy, and whose description fits in the narrative, is used to assign a label.

Limitations:

- Does not consider word order
- Some descriptions can be too specific

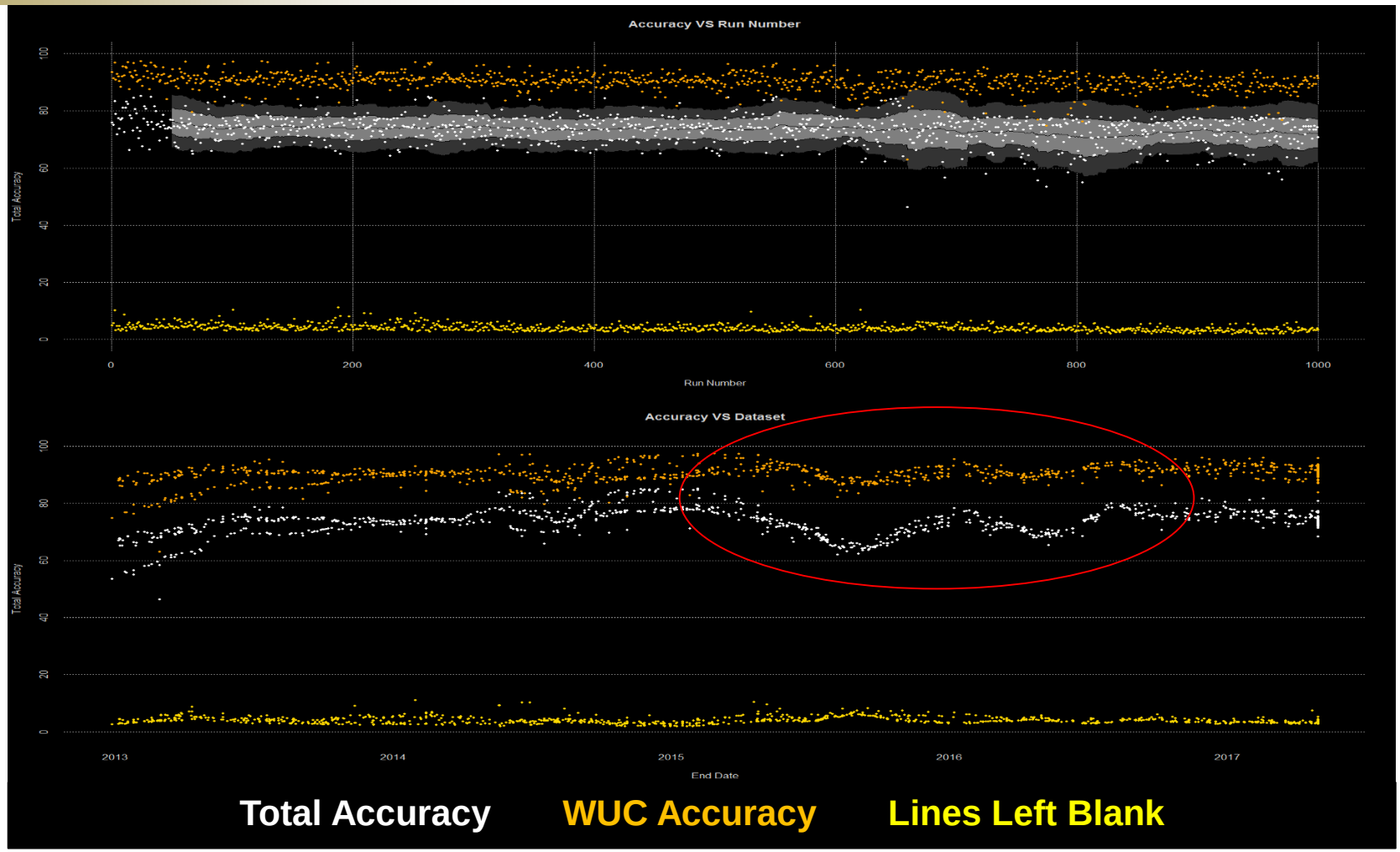
Sampling Theory Discussion Points

Current methods of Machine Learning recommend putting aside a set of labeled data to use as a test set.

- Does this solve the problem of "How do we know"?
- What is the best model that can be built from a subset of all data?

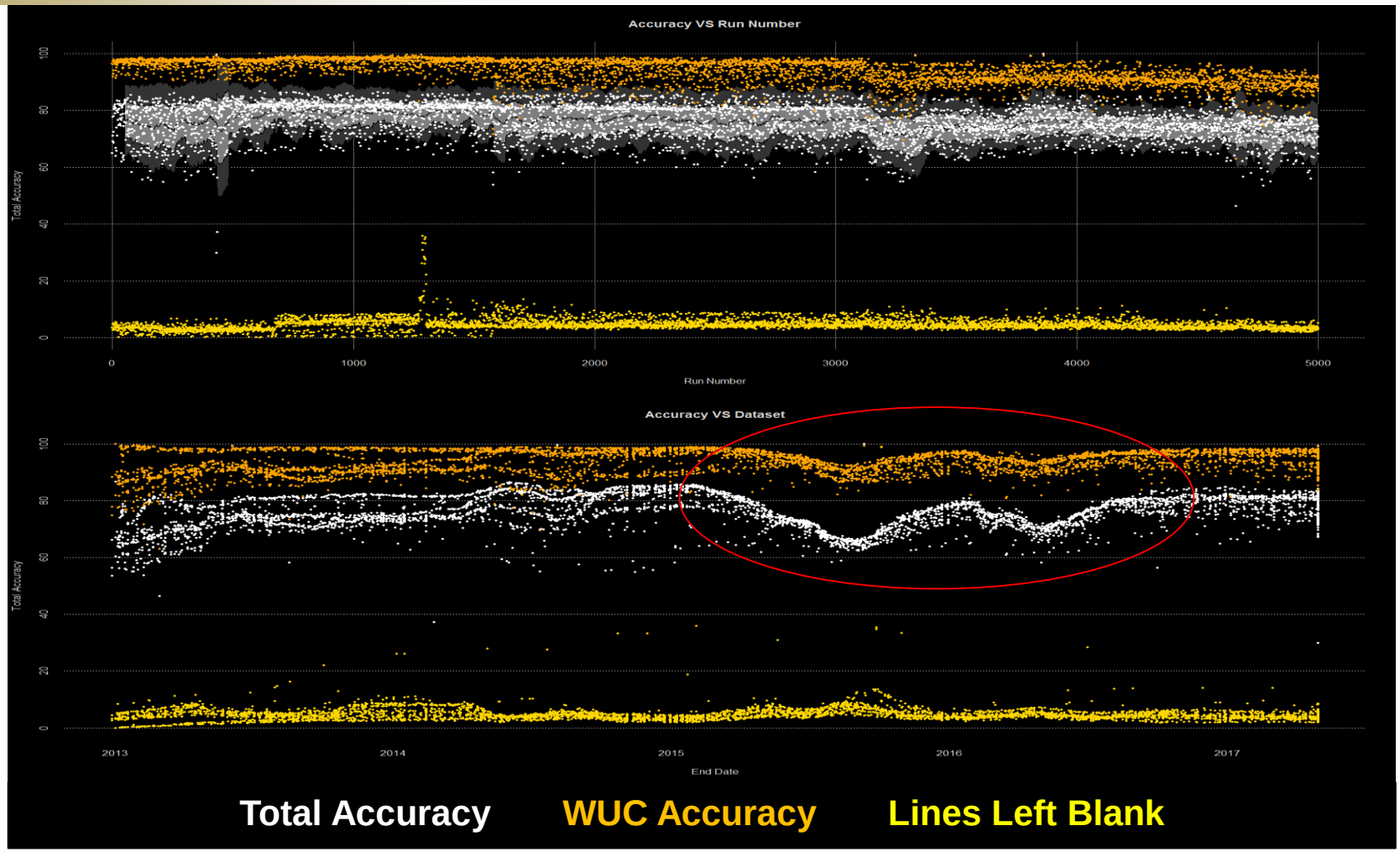
Current approach is to iterate learning over the entire set of labeled data by randomly sampling subsets of the data. This will produce a model that is a mean of all possible models. This model will not be fit to any particular data points in the data set, but will instead be a fit for things that are mostly true about the data (true more than 50% of the time). The idea is to provide the best and most generalized fit for all currently classified data.

Data Characterization last 1000 runs



Data Characterization and Modeling

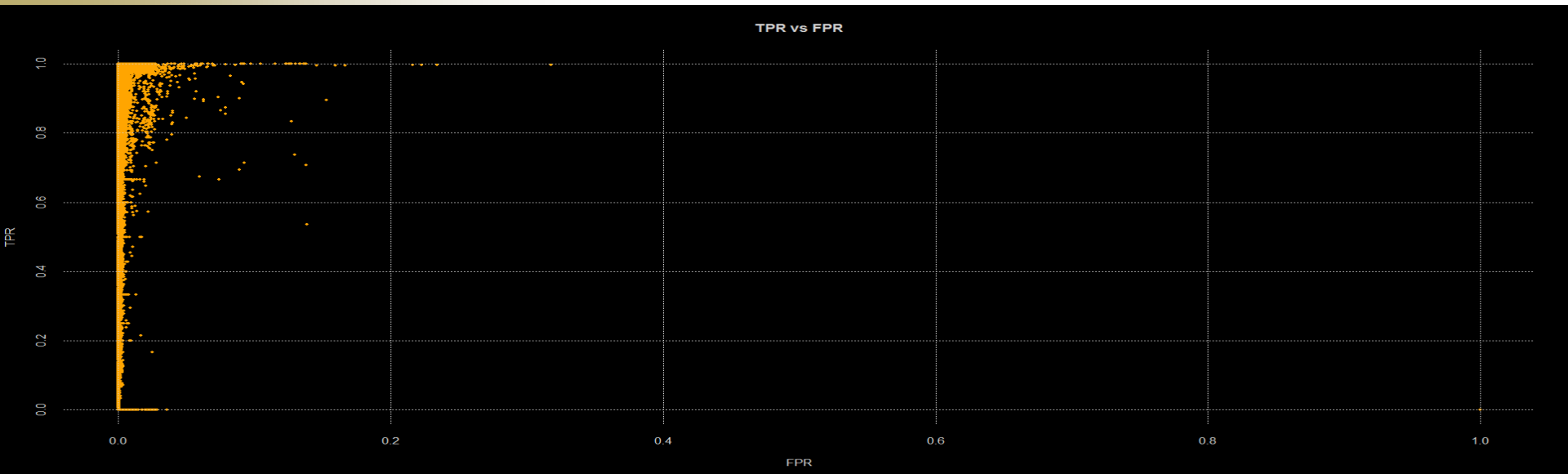
Data Characterization last 5000 runs

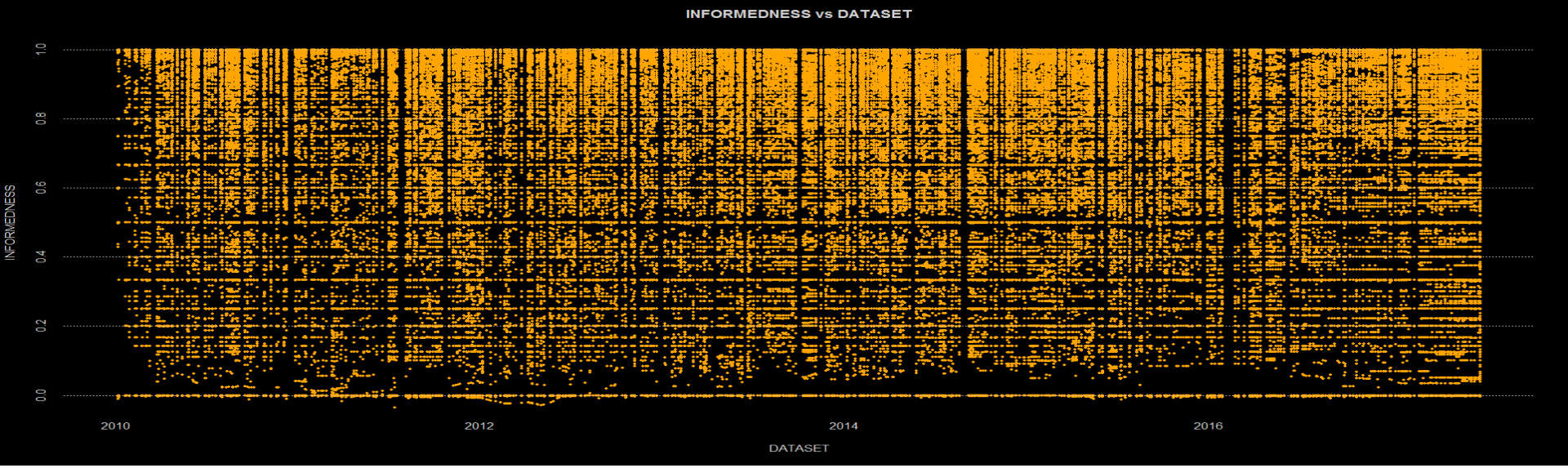


Total Accuracy WUC Accuracy Lines Left Blank

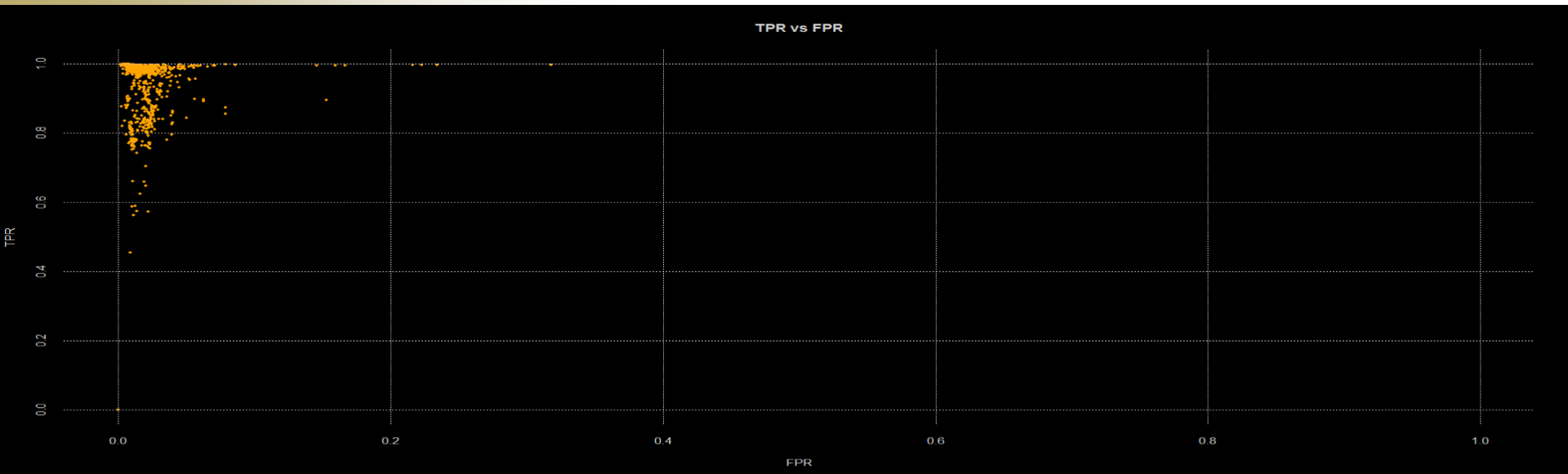
Data Characterization and Modeling

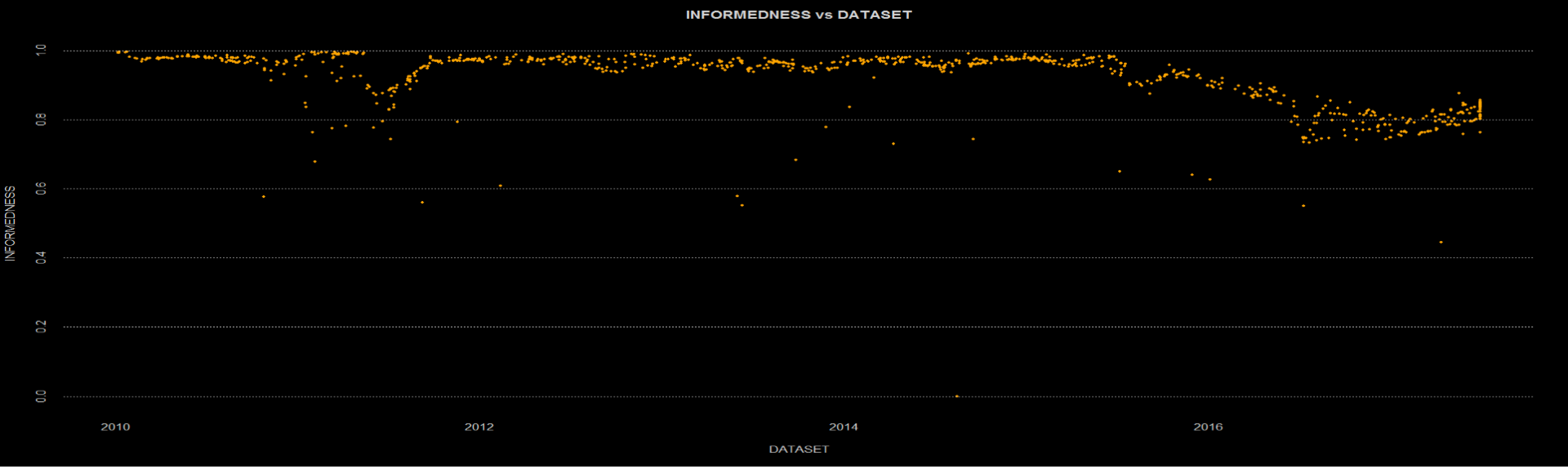
Informedness of all labels (previous 2 weeks)



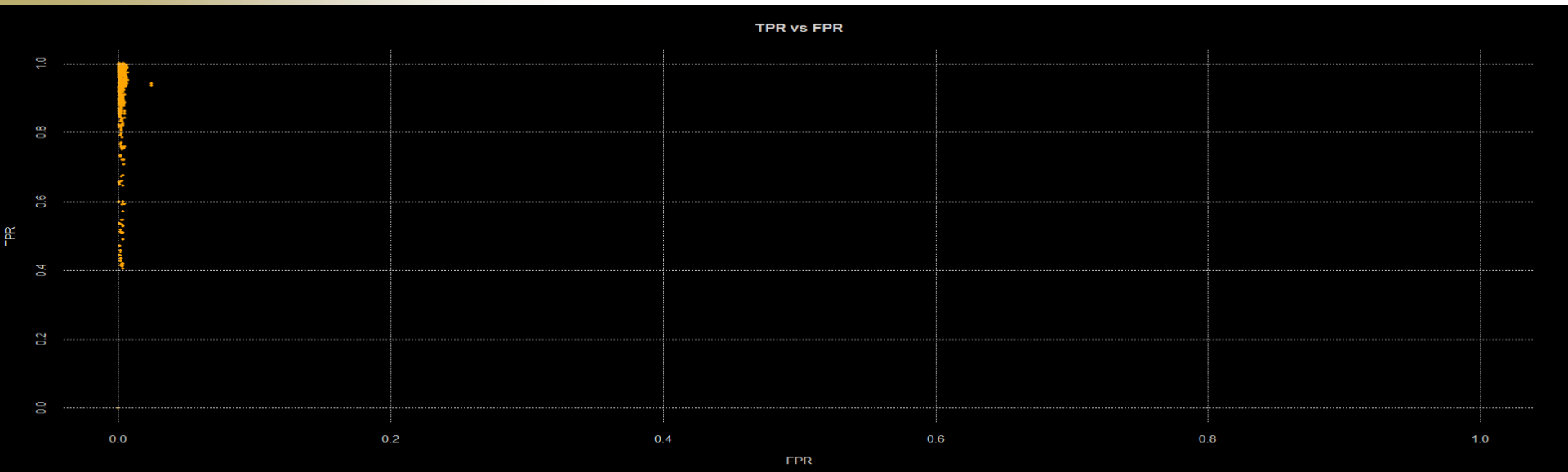


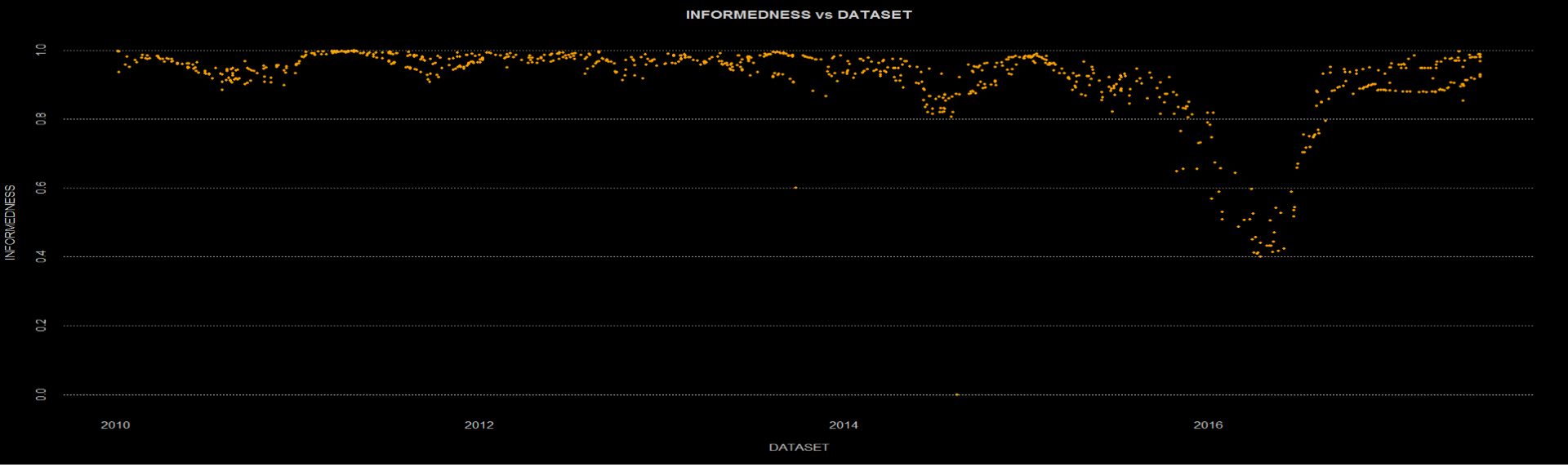
Informedness of “00” (previous 2 weeks)





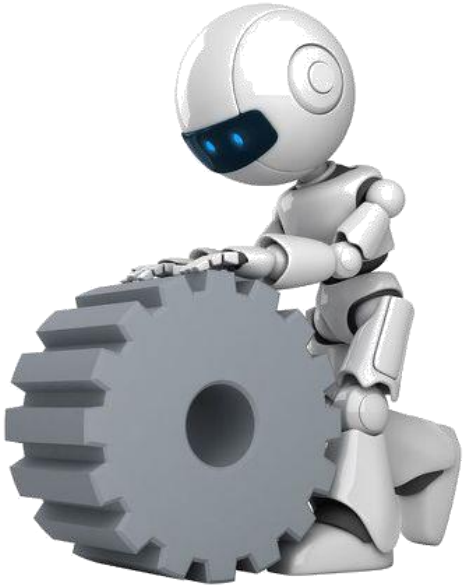
Informedness of “05A02” (previous 2 weeks)





CLARE Road Map

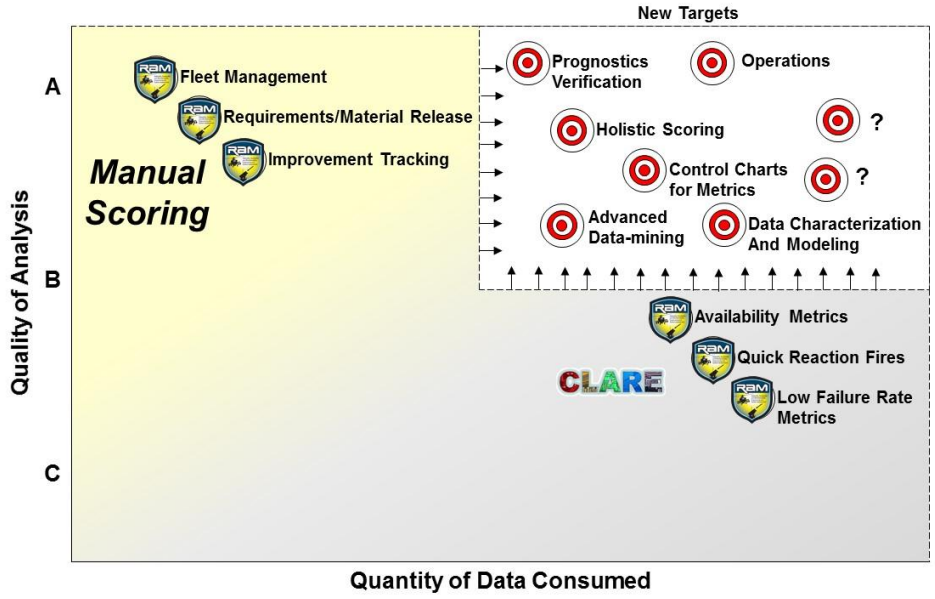
- ✓ Create Machine Learning Scoring System:
  - CLARE's base program is up and running for multiple platforms.
- ✓ Beta test CLARE and train up to production level accuracy:
  - >75% accurate on all platforms
- ❑ Implement CLARE data in: (Current State)
  - ✓ Availability Metrics and Associated Tools
  - ✓ Multi-Source Meshing
  - ✓ Phase Marking
  - ✓ QC integration
  - ❑ Scoring integration (FY 2017-2018)
- ❑ Receive Model Accreditation (AED & ERDC FY 2018)
- ❑ Scoring (FY 2018-2019)
- ❑ Anomaly Detection (FY 2018-2019)

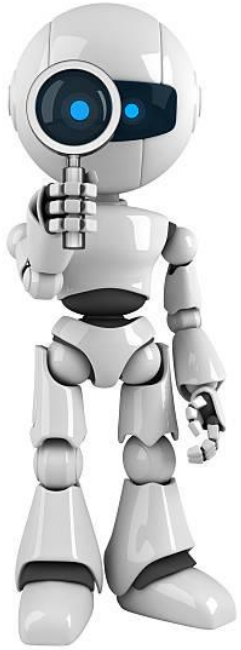


Summary

CLARE greatly expands and enhances our system assessment capabilities and has the potential for even greater utility.

- **Increase the amount of information gathered from the data**
- **Expand our customer base**
- **Lower cost of analysis**





Questions?

AMRDEC Web Site
www.amrdec.army.mil

Facebook
www.facebook.com/rdecom.amrdec

YouTube
www.youtube.com/user/AMRDEC

Twitter
@usarmyamrdec

Public Affairs
AMRDEC-PAO@amrdec.army.mil