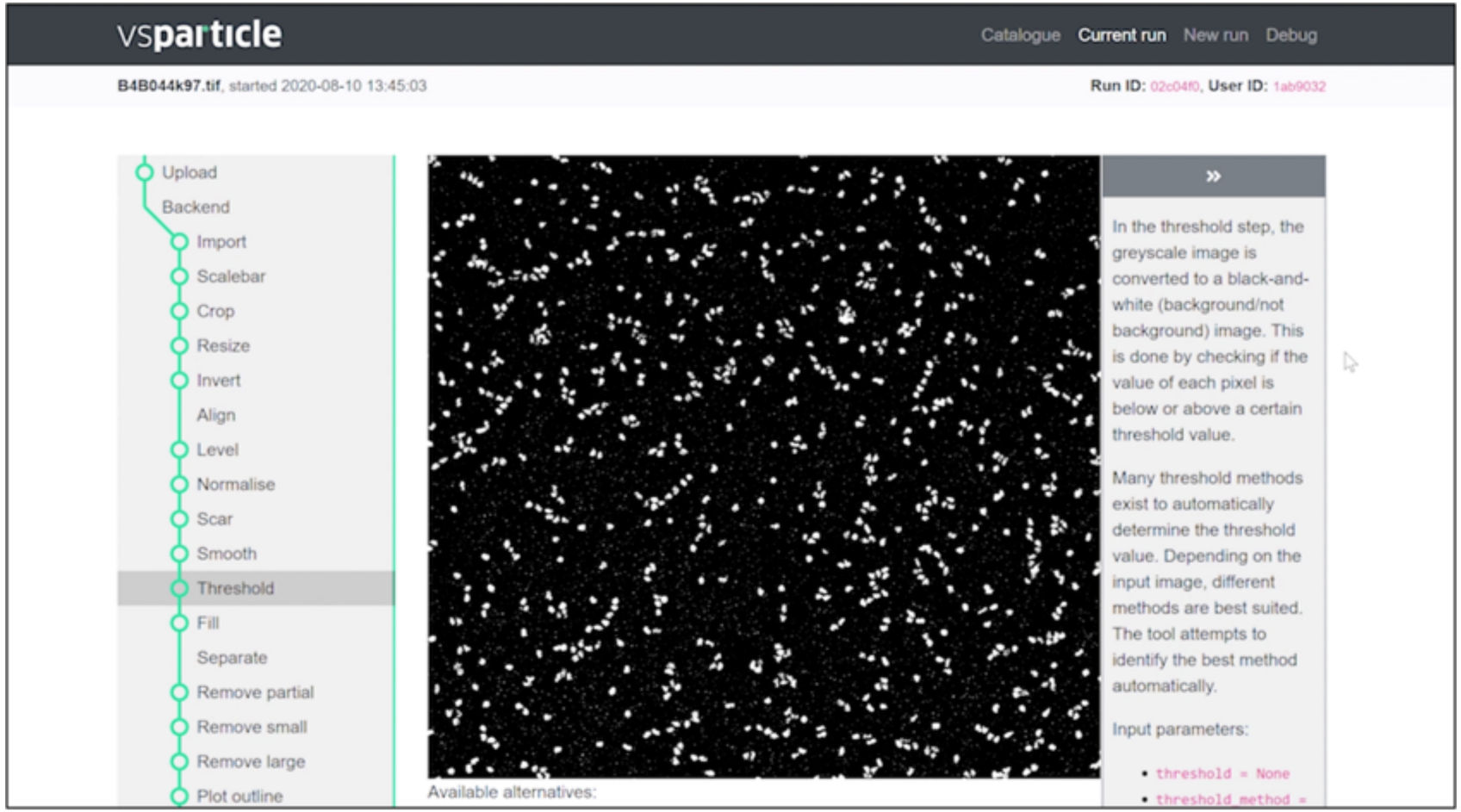
09/11 – 13/11

KLARA, YORAN, OSCAR

NANO — WEEK 10

OBJECTIVE

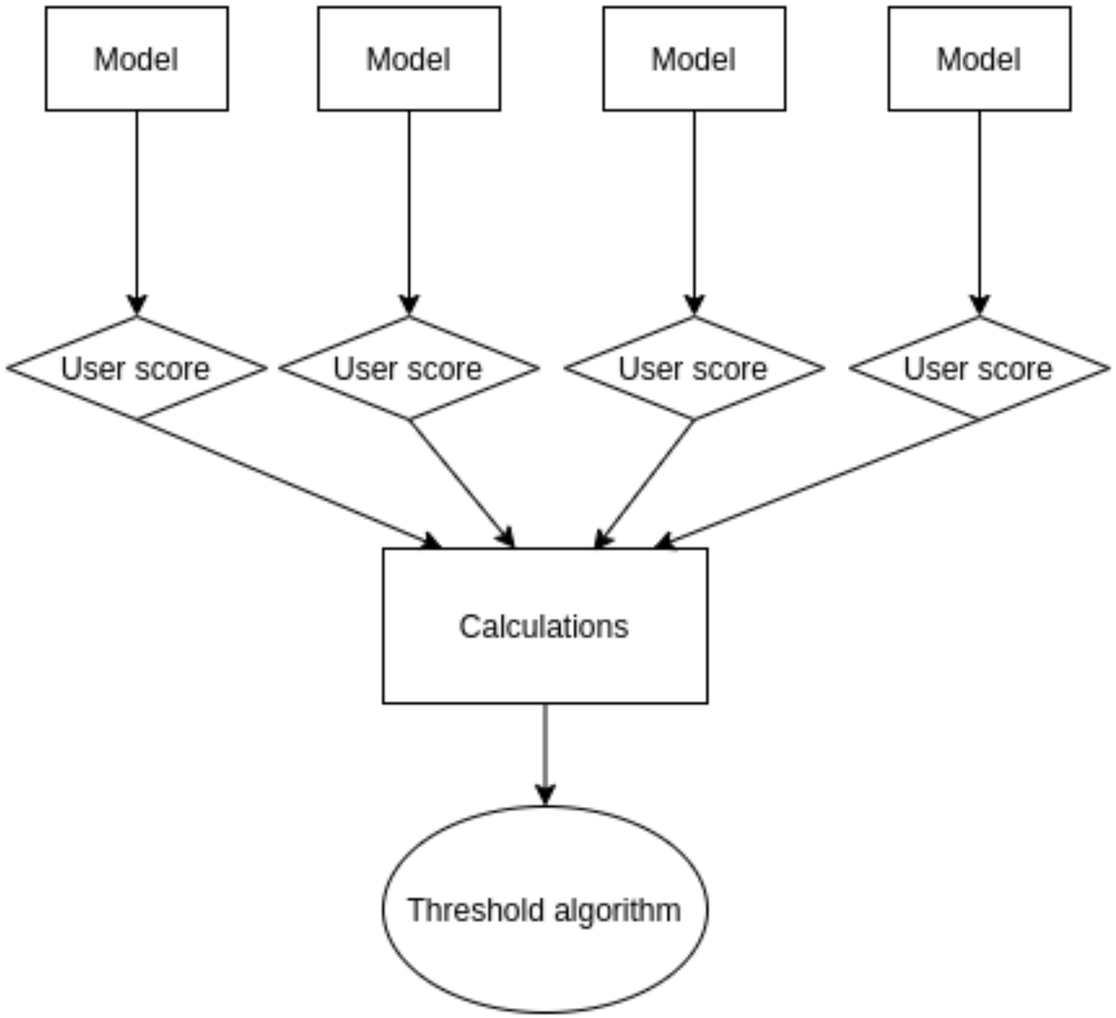
- VSParticle's image analysis software
- Create a model that predicts the thresholding method



DATASET

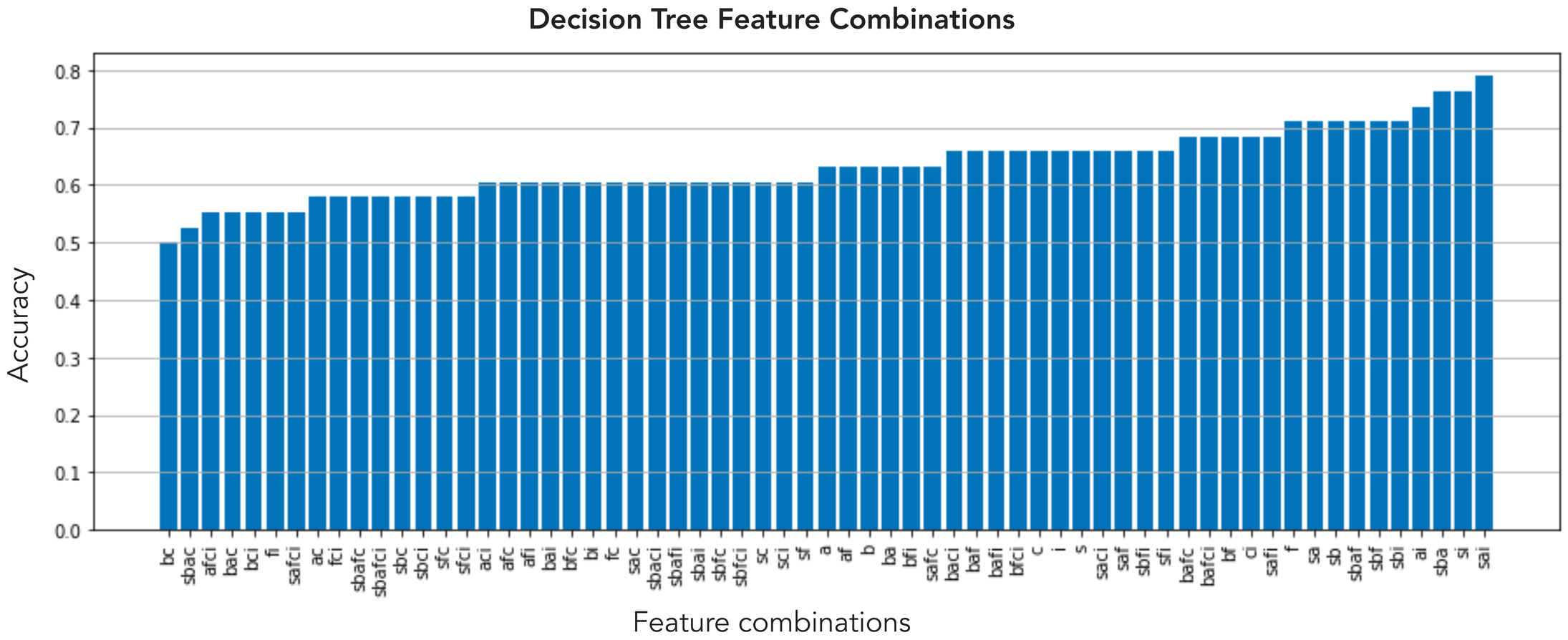
Contains:

- Used threshold algorithm
- Used algorithm feature scores
- User score



MODEL: BINARY DECISION TREE

- Separation, Area spread, Intensity

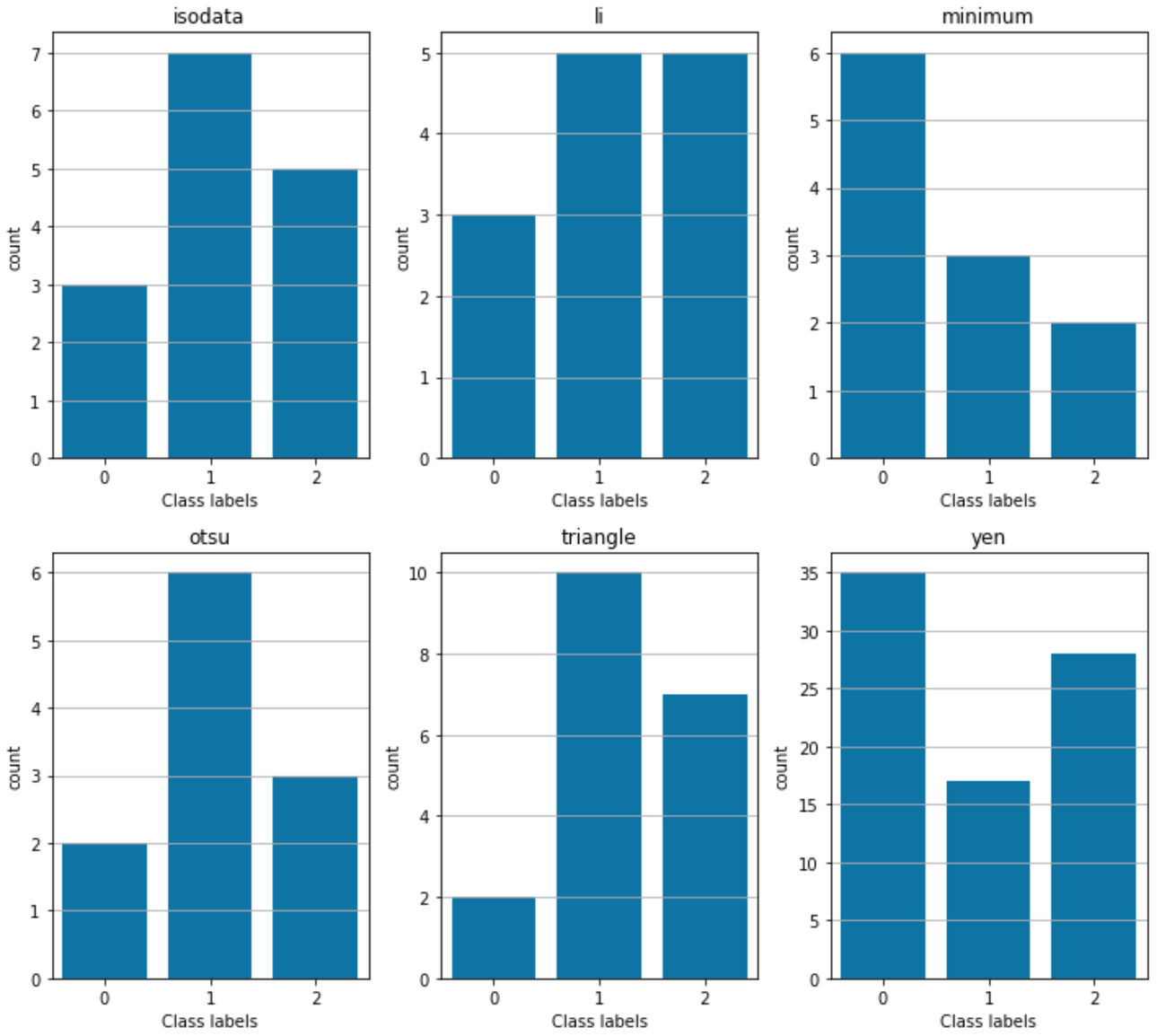


IMBALANCED DATA SOLUTIONS

- Random undersampling

- Random oversampling
- Cluster-based oversampling
- (M)SMOTE

- Non-linear classes



IMBALANCED DATA SOLUTIONS

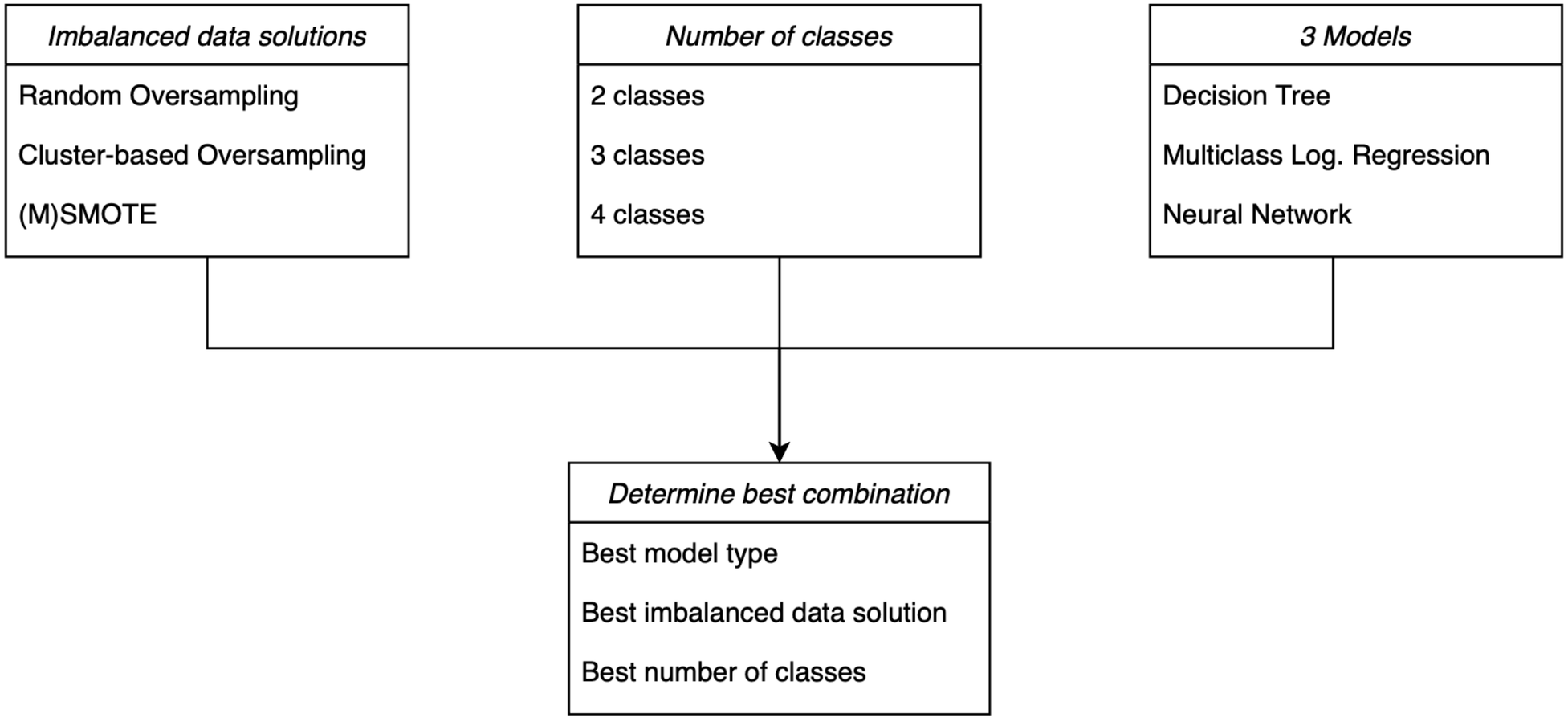
- Random undersampling → Not applicable

- Random oversampling
- Cluster-based oversampling
- (M)SMOTE

→ Determine best one

- Non-linear classes

APPROACH


Imbalanced data solutions
Random Oversampling
Cluster-based Oversampling
(M)SMOTE
Number of classes
2 classes
3 classes
4 classes
3 Models
Decision Tree
Multiclass Log. Regression
Neural Network
Determine best combination
Best model type
Best imbalanced data solution
Best number of classes

APPROACH

Possible outcomes:

Random Oversampling 2 classes Decision Tree	Random Oversampling 2 classes Neural Network	Random Oversampling 2 classes Multiclass Log. Regression
Cluster-based oversampling 2 classes Decision Tree	Cluster-based oversampling 2 classes Neural Network	Cluster-based oversampling 2 classes Multiclass Log. Regression
(M)SMOTE 2 classes Decision Tree	(M)SMOTE 2 classes Neural Network	(M)SMOTE 2 classes Multiclass Log. Regression
Random Oversampling 3 classes Decision Tree	Random Oversampling 3 classes Neural Network	Random Oversampling 3 classes Multiclass Log. Regression
Cluster-based oversampling 3 classes Decision Tree	Cluster-based oversampling 3 classes Neural Network	Cluster-based oversampling 3 classes Multiclass Log. Regression
(M)SMOTE 3 classes Decision Tree	(M)SMOTE 3 classes Neural Network	(M)SMOTE 3 classes Multiclass Log. Regression

RANKING PROBLEM INSTEAD OF PREDICTION PROBLEM

- Sounds good initially
- Not enough comparisons in the dataset
  - Mostly only between yen and one other algorithm for the same image
  - Example: yen + otsu, yen + isodata, yen + li are there,
    but not li + isodata

PROBLEMATIC DATASET

- Dataset has runs with same values but different user score
- Before we can do anything else, we need to fix it!

COMING WEEK

- Create new dataframe from json file
- Multiclass Logistic Regression model
- Combine models with number of classes and imbalanced data solutions

THANK YOU!

ANY QUESTIONS OR FEEDBACK?