Regression or Classification?

Ankur Garg, Chinmoy Baruah, Sanket Shahane
Group - i

Interesting Question

- Two typical types of problems: Regression and Classification

- Classification:
  - Target Variable is categorical. Example: Is it going to rain tomorrow ?
- Regression:
  - Target Variable is numeric. Example: Cost of a house, Temperature prediction

- But there can be one more case

Interesting Question

- What if your target variable doesn’t belong to just one type.
- Should we handle it as a regression problem or a classification problem ?
- Let’s find out.

Bug Priority Prediction

- Bug Priority is an important factor in software maintenance life cycle.
- In Bugzilla, Bug Priority can range from P1 to P5 {P1, P2, P3, P4, P5}.
- These are not simply classes because there is an ordinal relationship.
  - P1 > P2 > P3 > P4 > P5
- Can be represented as 1, 2, 3, 4, 5. But values are discrete. Not a straightforward regression problem.

Problem Definition

- How to predict Bug Priority ?
- Using Regression ? as suggested by our base paper. Or as a standard classification problem ? Does one method outperform the other ?
- What effect does Parameter Tuning have on such methods ?

What do we care about

- Conclusions supported by statistical evidence.
- Effect of Hyperparameter tuning.
- Model Complexity.

Literature Review

- Paper - “DRONE: Predicting Priority of Reported Bugs by Multi-Factor Analysis”
  - By Yuan Tian, David Lo
  - IEEE SM in 2015

- Treat the bug prediction as a regression problem.

Literature Review - Data Description

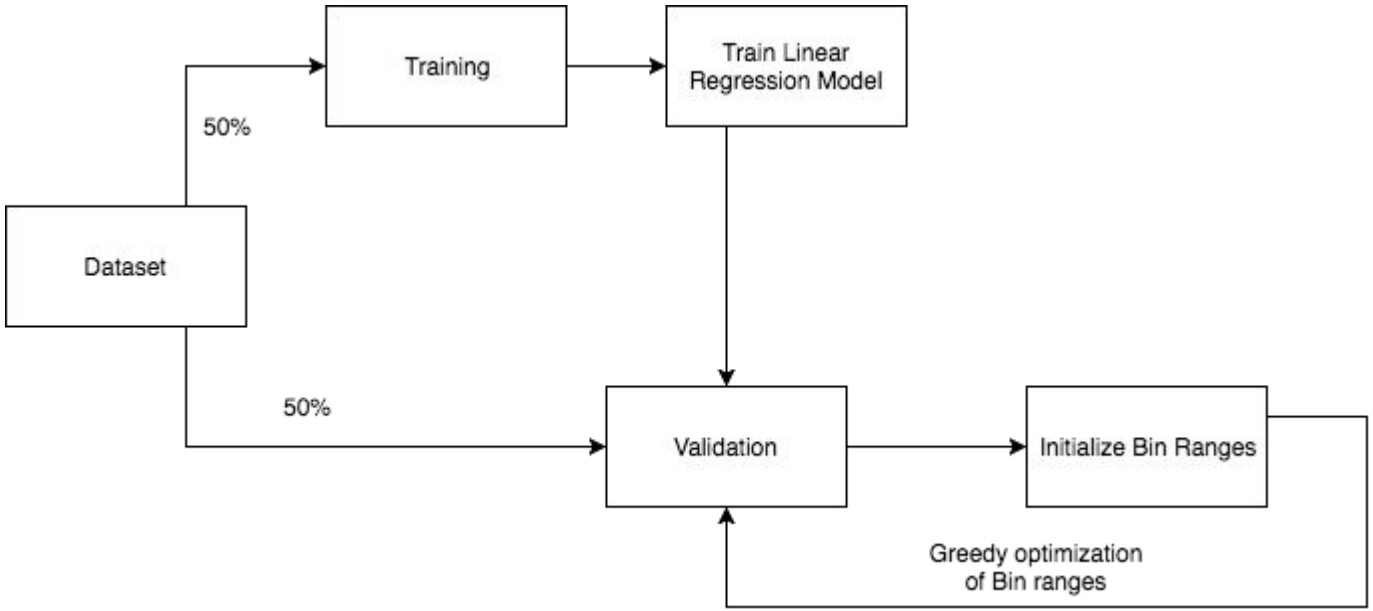
- Bug Reports from Eclipse bugzilla.
  - From 2001-10-10 to 2007-12-14
- ~103k samples. 11 Features (Severity, Date, Description etc.)
- Derived Features.

Literature Review - Data Description

- Derived Features:
  - Temporal Features: How many bug reports before this bug ?
  - Author Features: How many bugs before this by this author ?
  - Product Features: How many bugs before this for this component ?
  - Summary: Count Vectors

- After feature generation:
  - 39 Features (Temporal + Author + Product)
  - ~18k sized count vector (Summary)

Literature Review - Methodology



Literature Review - Algorithms & Claims

- DRONE, Naive Bayes, SVM, SeverisPrio+
- Claims:
  - Regression based DRONE better than SVM, SeverisPrio+
  - Naive Bayes couldn’t run to completion for comparison

Literature Review - Critique

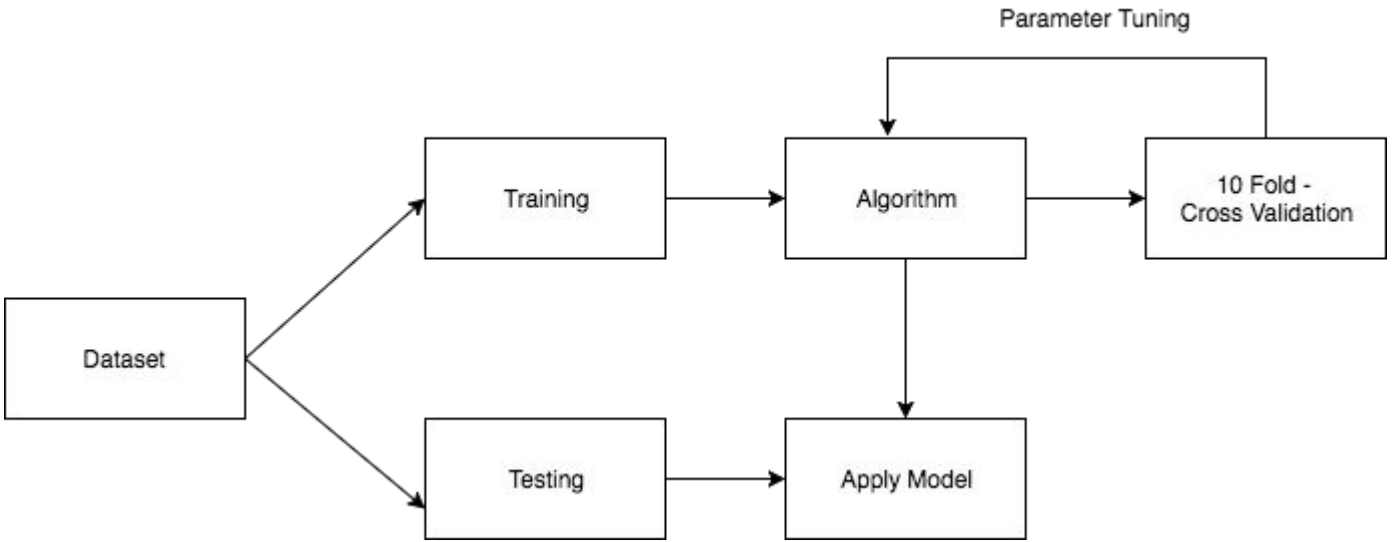
- No statistical evidence/evaluation provided for learner comparison.
- No Parameter Tuning. Off the shelf learners used with default parameters.
- Naive Bayes is fast and has low memory footprint. It should run to completion.

Research Questions

- Does hyper parameter tuning improve the results of classification algorithms ?
- Is there statistical evidence to support that DRONE is better than classification algorithms ?
- Can Naive Bayes run to completion on the given dataset ?

Our Methodology

- Comparison of DRONE with - Naive Bayes, Random Forest, SVM



Parameter Tuning

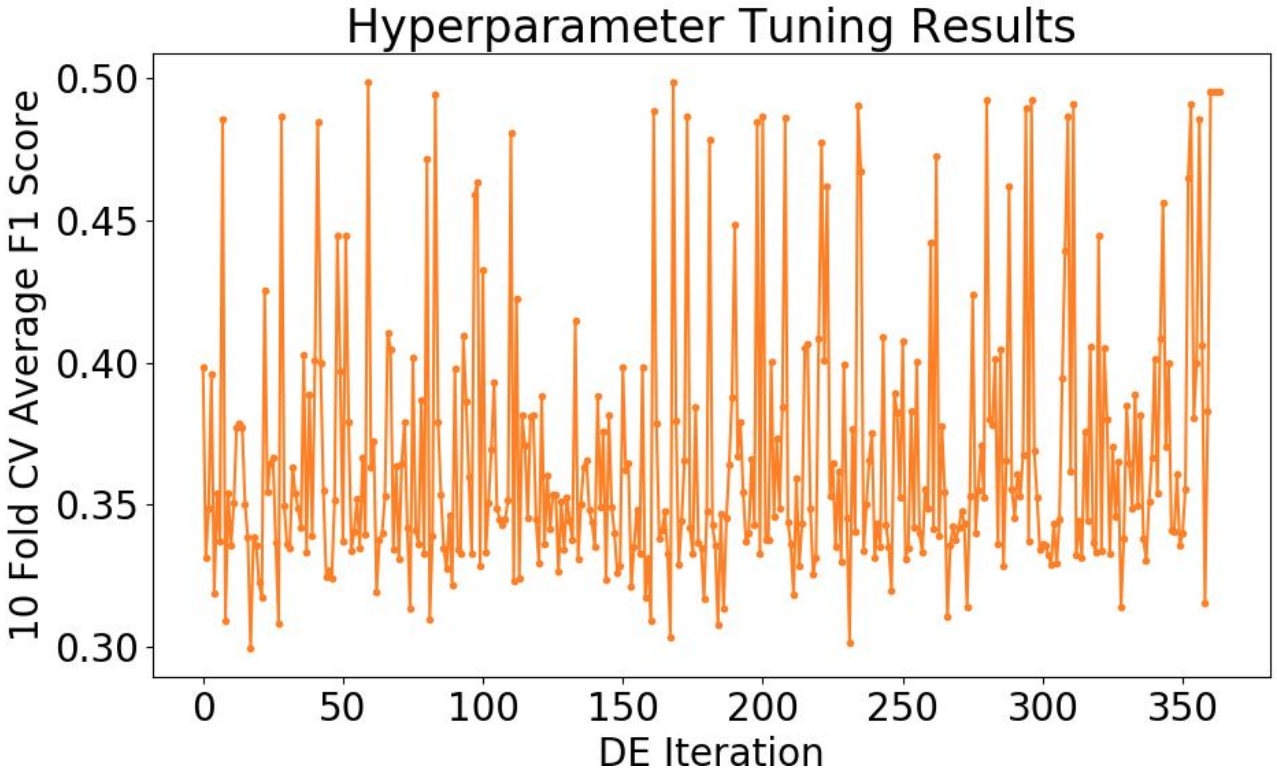
- Differential Evolution for tuning hyperparameters of Random Forest Classifier.
- Parameters tuned:
  - Number of trees, Minimum samples for splitting, Minimum samples in a leaf
- DE parameters:
  - Strategy: ‘rand2bin’
  - Frontier Size: 30
  - Mutation (f): (0.5, 1.9)
  - Recombination (crossover): 0.7
  - Max Iterations: 3

Results

Time Taken:
45 mins

Intel i7
8 cores
16 GB RAM

RF1: 46, 13, 1
RF2: 29, 10, 1
RF3: 19, 13, 1

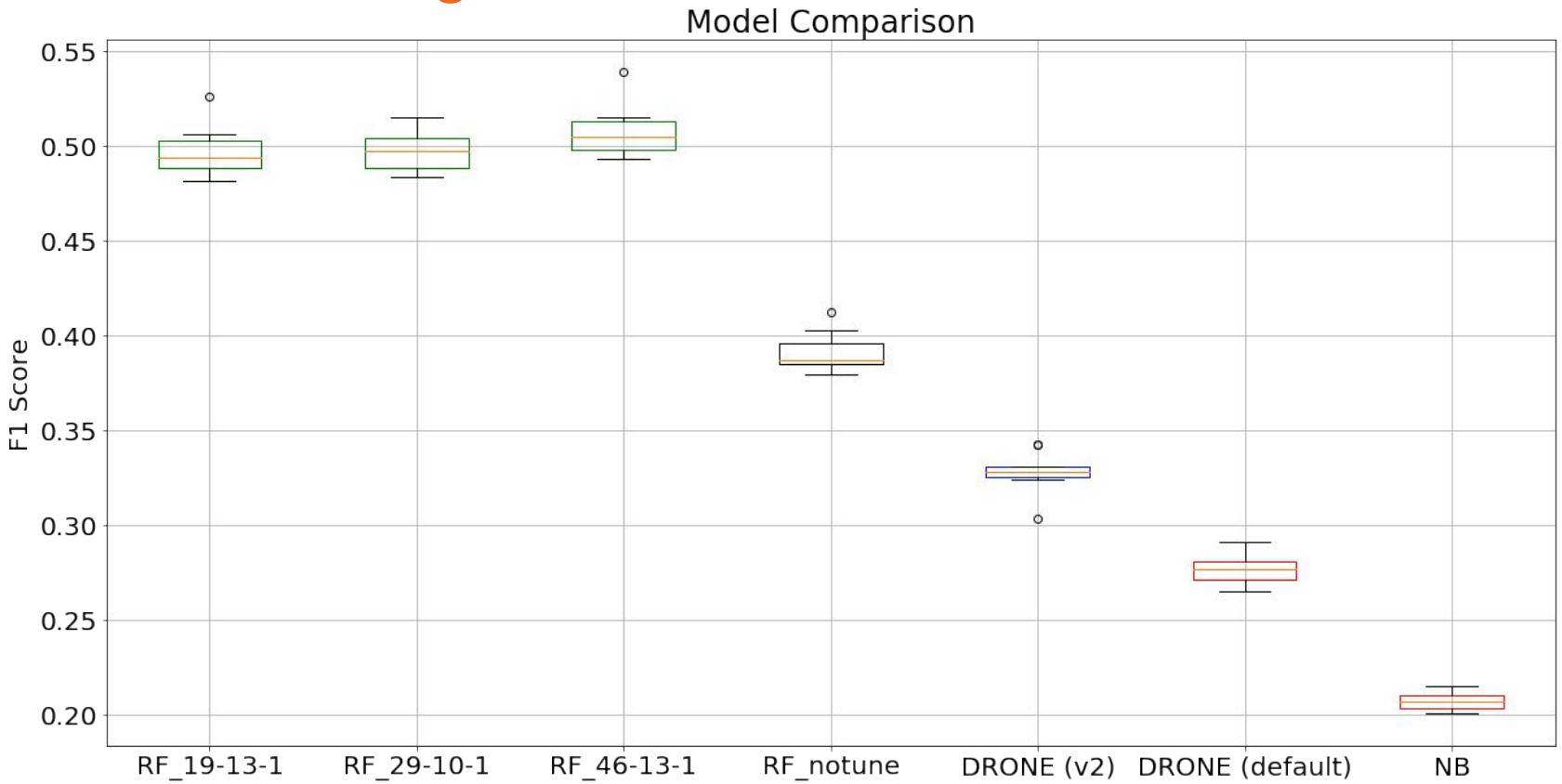


Results

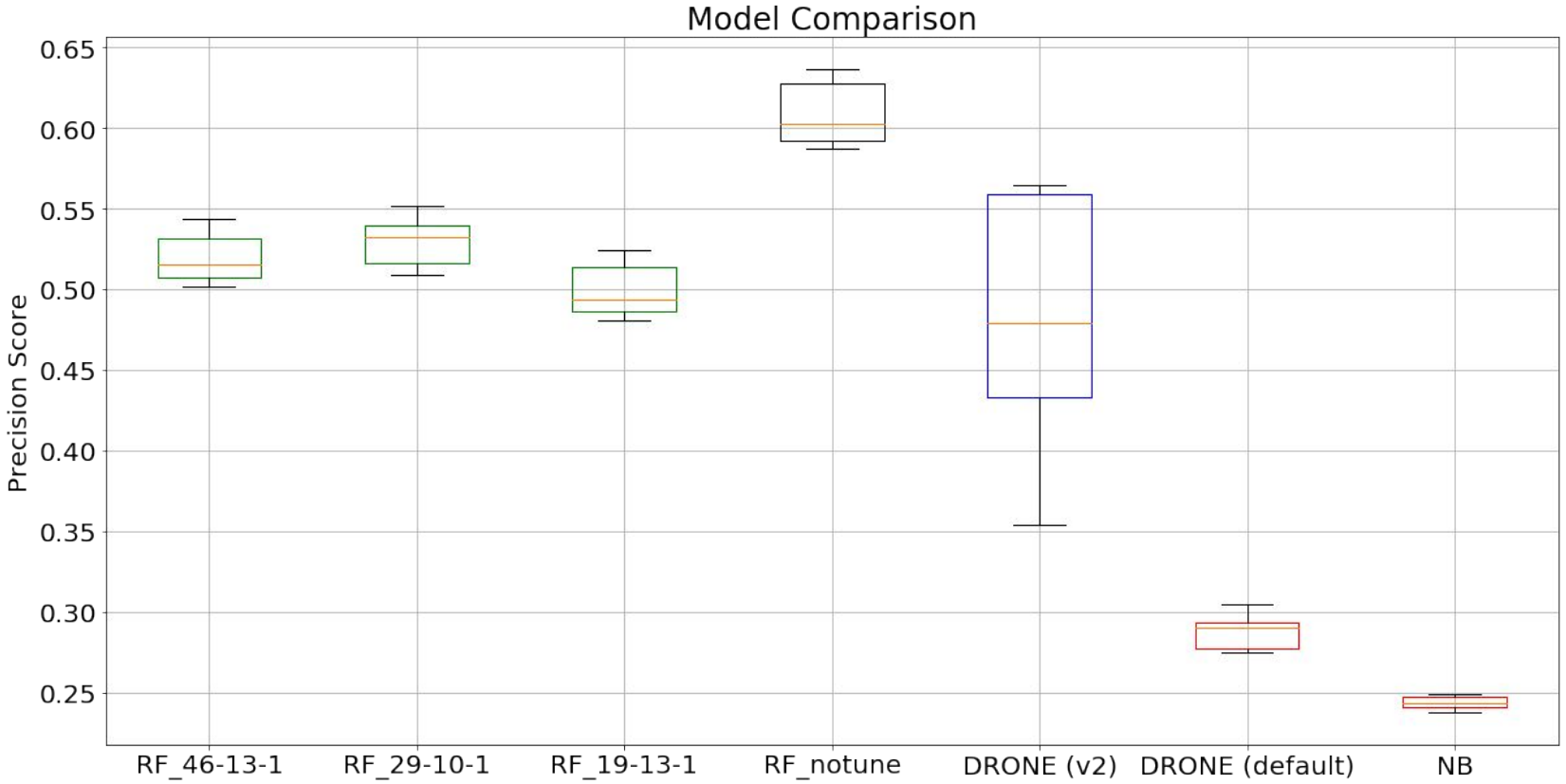
```
Ankurs-MacBook-Pro:correct_results iankurgarg$ cat results_for_stats.txt | python stats.py

rank ,         name ,    med   ,  iqr
----------------------------------------------------
   1 ,           nb ,    0   ,    0 (*              |             ), 0.20,  0.20,  0.21,  0.21,  0.21
   2 , drone_default ,    0   ,    0 (     -*-       |             ), 0.27,  0.27,  0.28,  0.28,  0.29
   3 ,     drone_v2 ,    0   ,    0 (          -*   |             ), 0.32,  0.33,  0.33,  0.33,  0.34
   4 ,    rf_notune ,    0   ,    0 (               |*-           ), 0.38,  0.39,  0.39,  0.40,  0.41
   5 ,     rf_19131 ,    0   ,    0 (               |         *-  ), 0.48,  0.49,  0.50,  0.50,  0.53
   5 ,     rf_29101 ,    0   ,    0 (               |         *   ), 0.48,  0.49,  0.50,  0.50,  0.51
   5 ,     rf_46131 ,    0   ,    0 (               |         *-- ), 0.49,  0.50,  0.50,  0.51,  0.54
Ankurs-MacBook-Pro:correct_results iankurgarg$
```

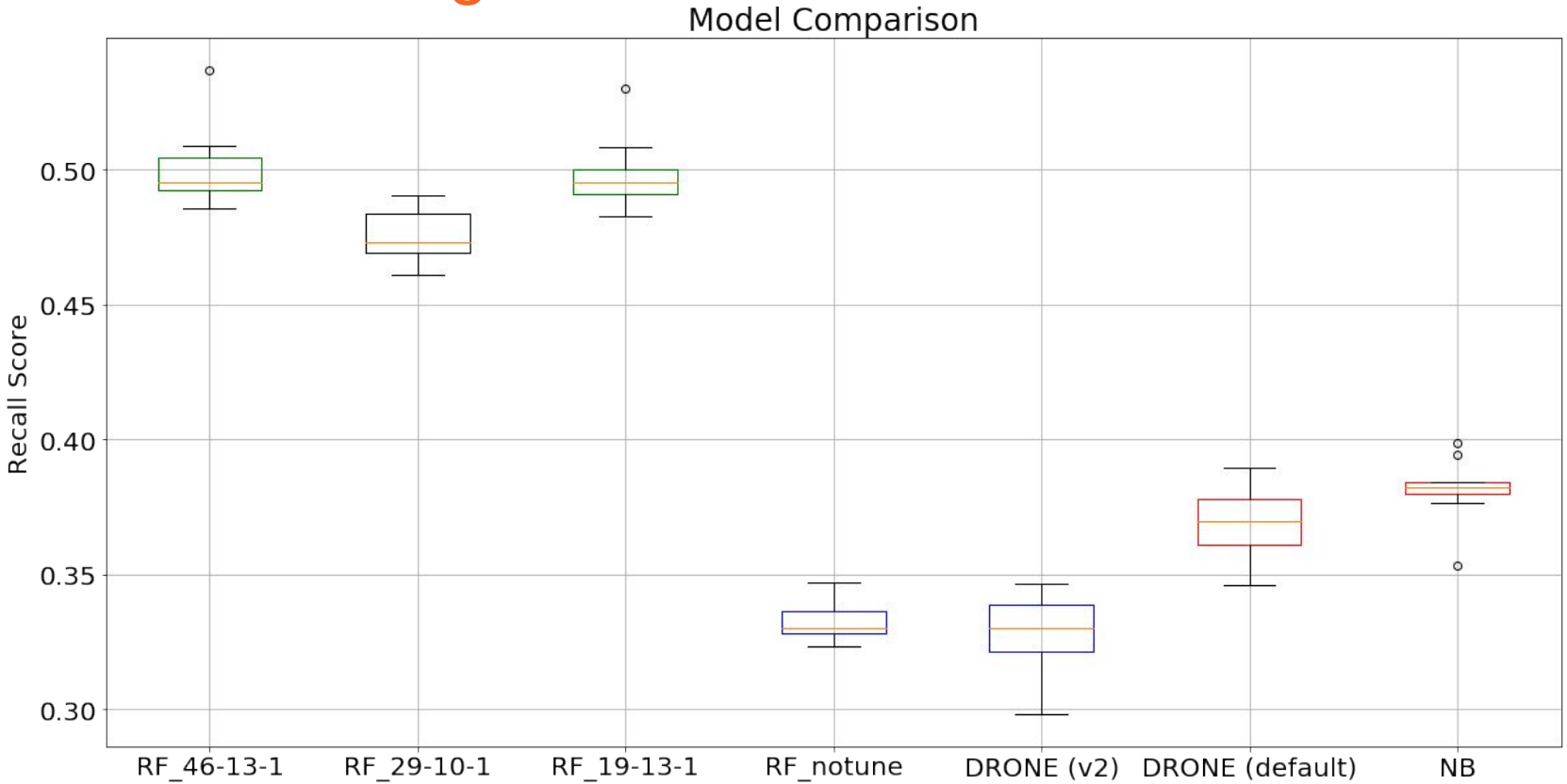
Results - Average F1 Score



Results - Average Precision

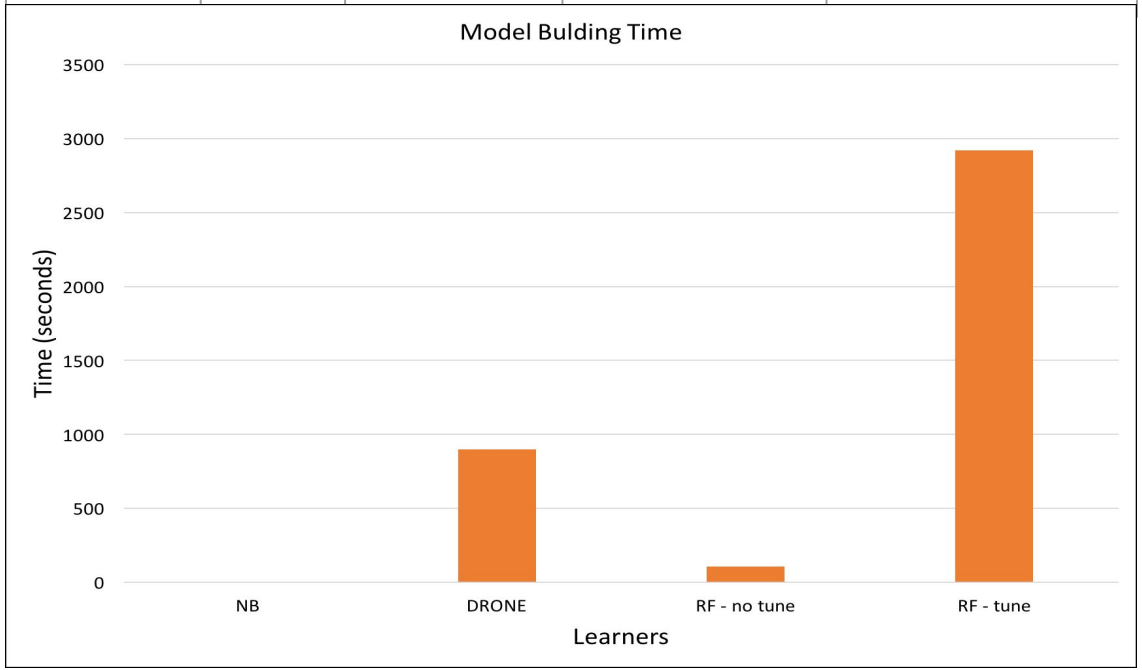


Results - Average Recall



Results - Time to train learners

Naive Bayes	DRONE	RF: no tuning	RF: tuning	SVM (1 Iter)
3.8 sec	15 mins	1 min 47 sec	3 mins 41 secs + 45 mins tuning	5 hrs 52 mins



Observations

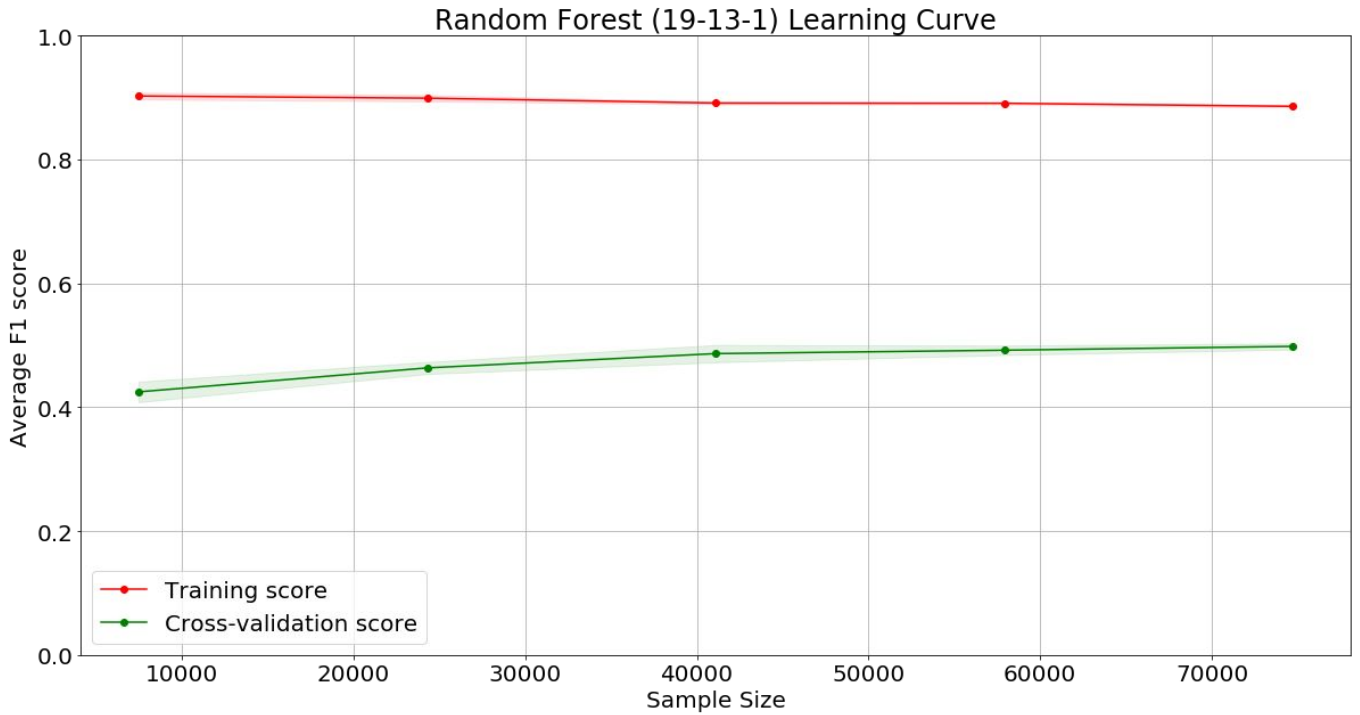
- Statistically significant change in F1 Score for Random Forest after parameter tuning.
- Random Forest with and without tuning performs better than DRONE
- Simplification of DRONE binning method yields significantly better results than DRONE.

Model Complexity

- From Differential Evolution on RF, selected three parameter settings with similar results (F1 Score = 0.49).
  - Num trees: 46, Min samples Split: 13, Min leaf samples: 1
  - Num trees: 29, Min samples Split: 10, Min leaf samples: 1
  - Num trees: 19, Min samples Split: 13, Min leaf samples: 1

- No Statistical Difference based on T-test on cross validation results.

Model Stability



The chosen one !

- Random Forest after parameter tuning:
  - Num trees: 19, Min samples Split: 13, Min leaf samples: 1

Results - Comparison with DRONE

DRONE			
Class	**Precision**	**Recall**	**F1 Score**
P1	29.4%	22.4%	25.4%
P2	20.0%	14.8%	16.99%
P3	88.7%	79.2%	83.7%
P4	0.0%	0.0%	0.0%
P5	6.5%	6.55%	11.8%
Average	**28.9%**	**36.4%**	**27.6%**

Random Forest (19, 13, 1)			
Class	**Precision**	**Recall**	**F1 Score**
P1	34.7%	42.0%	38.1%
P2	34.8%	32.5%	33.7%
P3	91.4%	91.6%	91.5%
P4	37.7%	30.6%	33.8%
P5	47.2%	48.1%	47.7%
Average	**49.2%**	**49%**	**49%**

Answers to Research Questions

1. Does hyper parameter tuning improve the results of classification algorithms ?

- Parameter Tuning can have significant impact on results.
- Parameter Tuning may yield unnecessarily complex models.
    - Use statistical tests to choose similar model of simpler complexity.

Answers to Research Questions

2. Is there statistical evidence to support that DRONE is better than classification algorithms ?

- Statistical tests important for comparing learners.
- Random Forest with Parameter tuning performs better than DRONE.
- No statistical evidence that DRONE is better than standard classification.

Answers to Research Questions

3. Can Naive Bayes run to completion on the given dataset ?

- Yes. Infact, Naive Bayes fastest among all algorithms compared.

Thank You