

# Information Modelling & Analysis – Project 2

Claudio Maggioni

## Code Repository

The code and result files, part of this submission, can be found at

- Repository: <https://github.com/infoMA2023/project-02-bug-prediction-maggion>
- Commit ID: **TBD**

## Data Pre-Processing

I use the sources of the CLOSURE repository that were already downloaded using the command:

```
defects4j checkout -p Closure -v 1f -w ./resources
```

and used the code in the following subfolder for the project:

```
./resources/defects4j-checkout-closure-1f/src/com/google/javascript/jscomp/
```

relative to the root folder of the repository. The resulting csv of extracted, labelled feature vectors can be found in the repository at the following path:

```
./metrics/feature_vectors_labeled.csv
```

relative to the root folder of the repository.

## Feature Vector Extraction

I extracted **291** feature vectors in total. Aggregate metrics about the extracted feature vectors, i.e. the distribution of the values of each code metric, can be found in Table 1.

Table 1: Distribution of values for each extracted code metric.

<b>Metric</b>	<b>Min</b>	<b>Average</b>	<b>Max</b>
BCM	0	13.4124	221
CPX	0	5.8247	96
DCM	0	4.8652	176.2
EX	0	0.1134	2
FLD	0	6.5773	167
INT	0	0.6667	3
MTH	0	11.6529	209
NML	0	13.5622	28
RET	0	3.6735	86
RFC	0	107.2710	882
SZ	0	18.9966	347
WRD	0	314.4740	3133

## Feature Vector Labelling

After feature vectors are labeled, I determine that the dataset contains **75** buggy classes and **216** non-buggy classes.

## Classifiers

In every subsection below, describe in a concise way which different hyperparameters you tried for the corresponding classifier, and report the corresponding precision, recall and F1 values (for example in a table or an ITEMIZE-environment). Furthermore, for every type of classifiers, explicitly mention which hyperparameter configuration you chose (based on above reported results) to be used in further steps, and (in one or two sentences), explain why these hyperparameters may outperform the other ones you tested..

### Decision Tree (DT)

### Naive Bayes (NB)

### Support Vector Machine (SVP)

### Multi-Layer Perceptron (MLP)

### Random Forest (RF)

## Evaluation

### Output Distributions

Add a boxplot showing mean and standard deviation for **Precision** values on all 6 classifiers (5 trained + 1 biased)

Add a boxplot showing mean and standard deviation for **Recall** values on all 6 classifiers (5 trained + 1 biased)

Add a boxplot showing mean and standard deviation for **F1** values on all 6 classifiers (5 trained + 1 biased)

### Comparison and Significance

For every combination of two classifiers and every performance metric (precision, recall, f1) compare which algorithm performs better, by how much, and report the corresponding p-value in the following subsubsections:

Table 2: Pairwise Wilcoxon test for precision for each combination of classifiers.

	DecisionTreeClassifier	GaussianNB	MLPClassifier	RandomForestClassifier	SVC
BiasedClassifier	0	0	0	0	0
DecisionTreeClassifier	–	0.0893	0.4012	0	0
GaussianNB	–	–	0.0348	0	0
MLPClassifier	–	–	–	0	0
RandomForestClassifier	–	–	–	–	0

Table 3: Pairwise Wilcoxon test for recall for each combination of classifiers.

	DecisionTreeClassifier	GaussianNB	MLPClassifier	RandomForestClassifier	SVC
BiasedClassifier	0	0	0	0	0
DecisionTreeClassifier	–	0	0.0118	0.3276	0
GaussianNB	–	–	0	0	0
MLPClassifier	–	–	–	0.0001	0

	DecisionTreeClassifier	GaussianNB	MLPClassifier	RandomForestClassifier	SVC
RandomForestClassifier	–	–	–	–	0

Table 4: Pairwise Wilcoxon test for the F1 score metric for each combination of classifiers.

	DecisionTreeClassifier	GaussianNB	MLPClassifier	RandomForestClassifier	SVC
BiasedClassifier	0	0	0	0	0
DecisionTreeClassifier	–	0	0.4711	0	0
GaussianNB	–	–	0	0	0
MLPClassifier	–	–	–	0	0
RandomForestClassifier	–	–	–	–	0

### F1 Values

- 
- ...

### Precision

(same as for F1 above)

### Recall

(same as for F1 above)

### Practical Usefulness

Discuss the practical usefulness of the obtained classifiers in a realistic bug prediction scenario (1 paragraph).

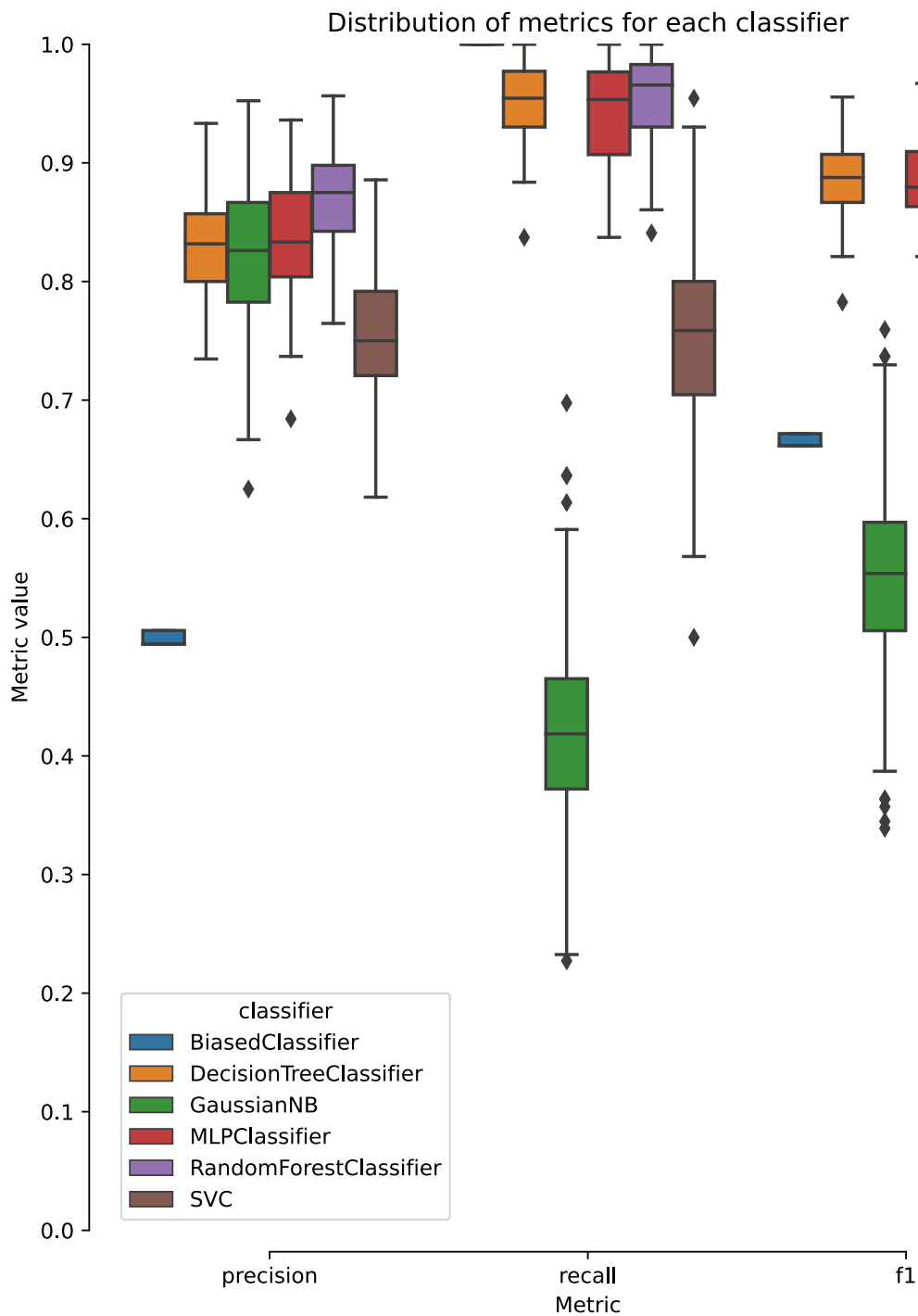


Figure 1: Precision, Recall and F1 score distribution for each classifier for 20-times cross validation.