

TEST AND TRAIN ECLIPSE APPLICATION

PREDICTION MODEL

ECLIPSE VERSION 3.2

STATISTIC

- ▶ More than 2.7 million line of code
- ▶ And on version 3.2, more than 1.2 line of code had add
- ▶ Average file age is around 1000 weeks, or about 20 years
- ▶ The oldest file was created about 35 years ago

ECLIPSE VERSION 3.2

CHARACTERISTICS

Application predict's characteristics

- ▶ Should have about 1000 bug exist on this version (3.2)
- ▶ Calculate as 13.652% through all POST session
- ▶ Need to write more than 1660 lines, to occurred any bug
- ▶ Efficiency of the prediction is 0.77



CARET PACKAGE

METRIC

RANDOM FOREST (RF)

R MODEL CLASSIFICATION

```
fitControl ← trainControl(  
  method = "repeatedcv",  
  number = 9,  
  repeats = 4  
)
```

Definition

repeatedcv = repeated training/test splits.

number = controls with the number of folds in K -fold cross-validation.

repeats = applied only to repeated K -fold cross-validation.


```
model.lda ← train(POST ~ .,  
  data = shrunked_data,  
  method = "rf",  
  tuneGrid = data.frame(mtry = 2),  
  trControl = fitControl  
)
```

Definition

shrunked_data = filter data from training file.

mtry = Number of variables randomly sampled at each split.

trControl = A list of values that define how this function acts.