

Label File

Code File

Text-Based Analysis

Image-Based Analysis

1) Read

2) Shuffle

3) Interpret

4) Split into training  
and testing stratified  
to class frequency

5) Train

6) Test

code file path

*dimension scores*

readability

understand-  
ability

complexity

multiclass (arg-  
max on Likert)

multiclass (arg-  
max on Likert)

multiclass (arg-  
max on Likert)

binary classes

binary classes

binary classes

*architectures*

raw text

*tokeni-  
zation*

BERT  
tokenizer

RoBERTa  
tokenizer

CodeBERT  
tokenizer

TF-IDF

512-char  
chunks

512-char  
chunks

512-char  
chunks

syntax-  
highlighted PDF

quadratic  
images

flat  
array

BERT

RoBERTa

CodeBERT

Naïve Bayes

SVM

SVM

AlexNet

*Grid Search  
Cross  
Validation*

Stem-  
ming

camelCase  
Splitting

N-grams range:  
(1, 1), (2, 2), (3, 3),  
(1, 2), (2, 3), (1, 3)

Kernels:  
Polynomial,  
RBF, Sigmoid

*evaluation criteria*

Accuracy

Precision

Recall

F1-Score

Matthews Corre-  
lation Coefficient

Mean Squa-  
red Error