Bug Prediction Made Easy

Ensar Guneşdoğdu

Supervisor: Haidar Osman

Software Composition Group Seminar
26 May 2015

Problem

Independent Variables

Source Code Metrics

History Version Metrics

Organizational Metrics

Input

Prediction Model

Regression

Probability

Binary

Cache

Output

Dependent Variable

Class

Bug Proneness

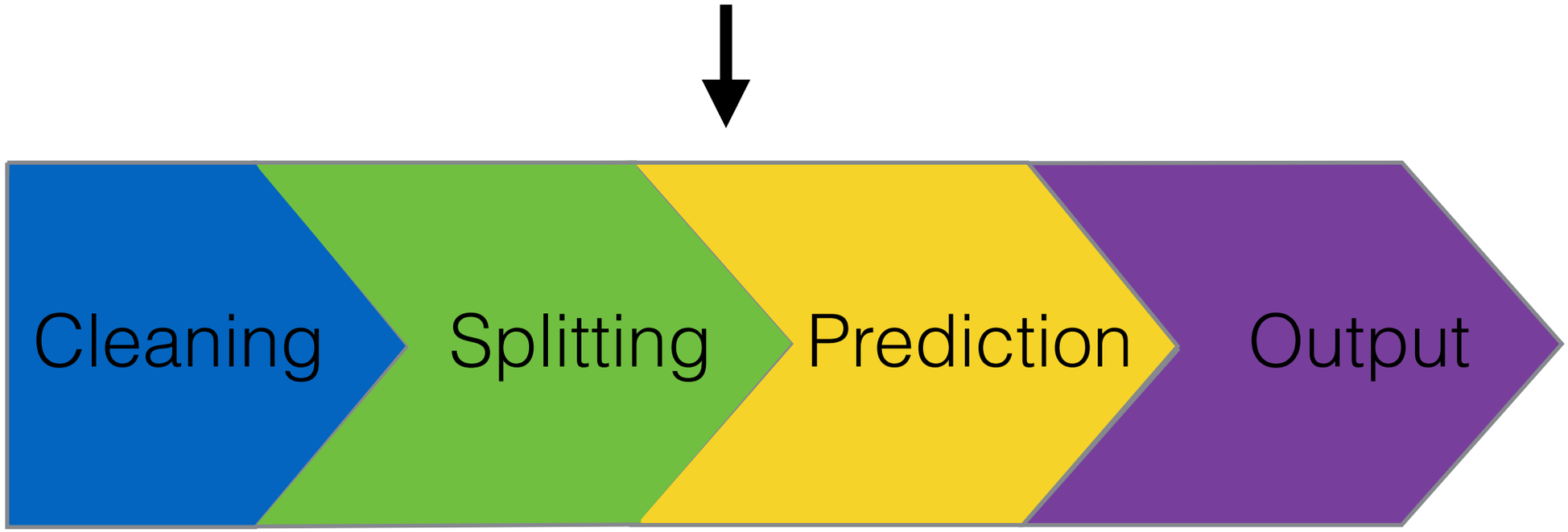
Number of Bugs

Bug Density

Anteater

How it works

Configuration XML & Datasets



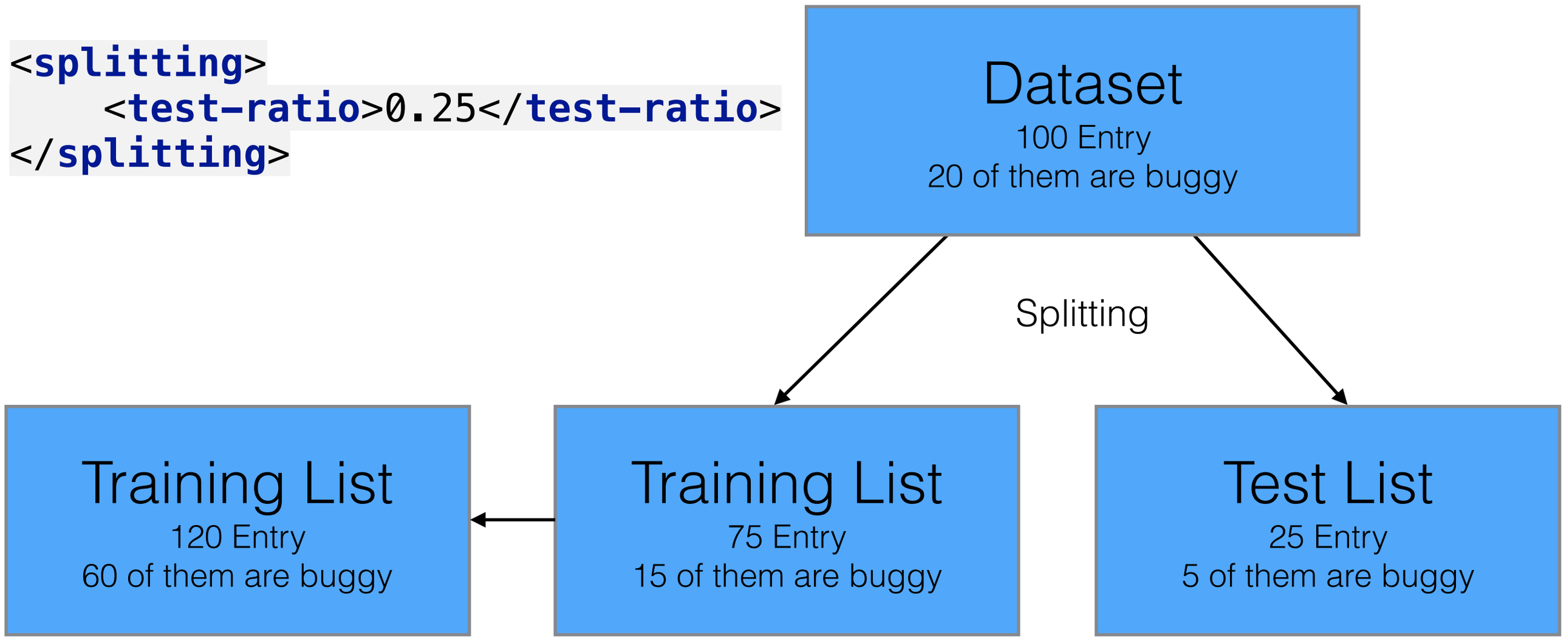
CSV and Excel results

```
<cleanup>
   <rows-to-remove>
      <item>
         <header>numberOfLinesOfCode</header>
         <value>0</value>
      </item>
   </rows-to-remove>

   <columns-to-remove>
      <item>
         <header>majorBugs</header>
      </item>
   </columns-to-remove>

   <class-name-column>
      <header>classname</header>
   </class-name-column>

   <loc-column>
      <header>numberOfLinesOfCode</header>
   </loc-column>

   <target-columns>
      <header>bugs</header>
   </target-columns>

</cleanup>
```

Cleaning Step

Splitting

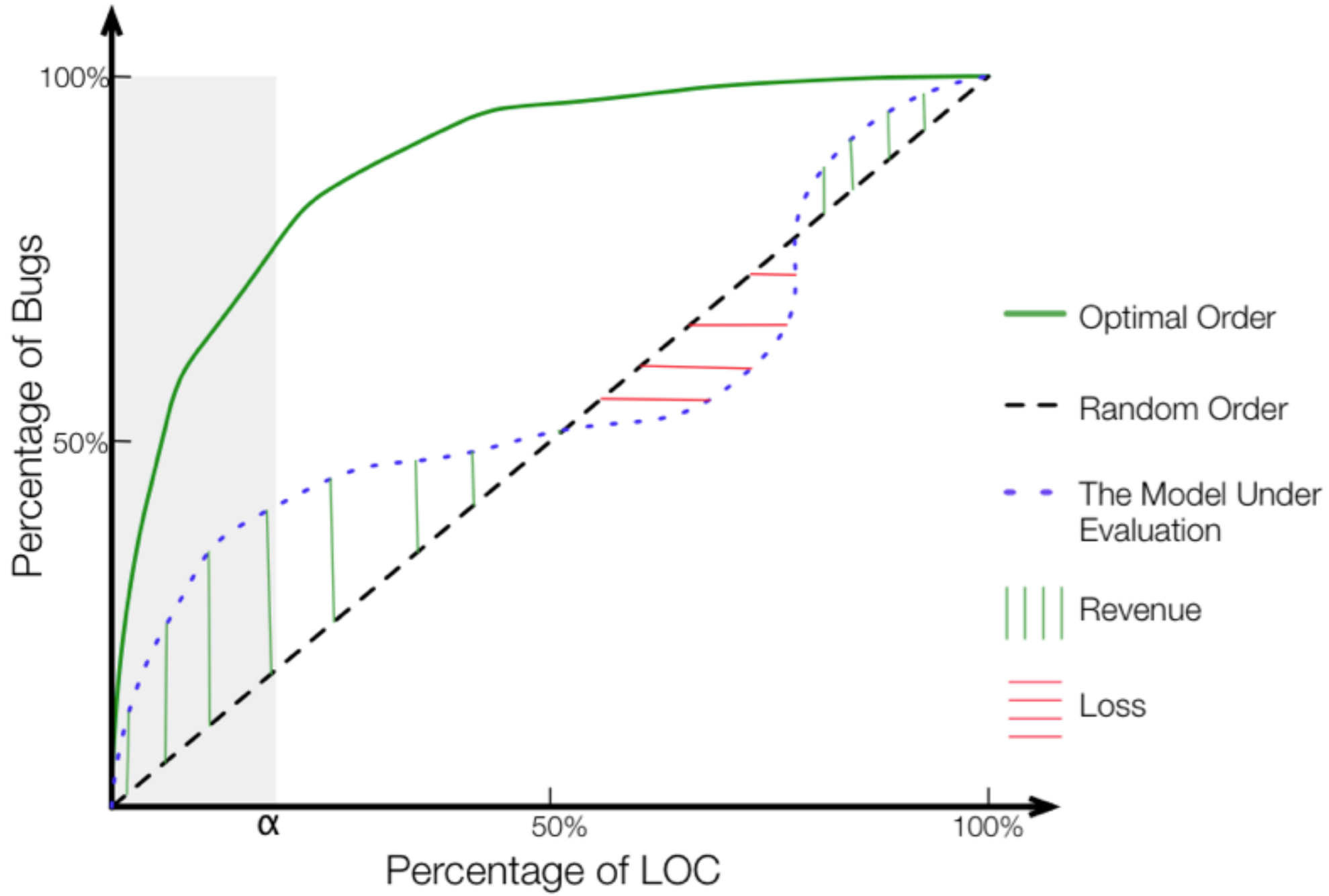
```
<splitting>
    <test-ratio>0.25</test-ratio>
</splitting>
```



```
<predictions>
   <mode>
      <name>buggy-class</name>
      <model>SMO</model>
      <model>MultilayerPerceptron</model>
   </mode>

   <mode>
      <name>bug-density</name>
      <model>MultilayerPerceptron</model>
   </mode>

   <mode>
      <name>bug-count</name>
      <model>LinearRegression</model>
      <model>MultilayerPerceptron</model>
   </mode>

   <mode>
      <name>bug-proneness</name>
      <model>RandomForest</model>
      <model>IBk</model>
      <model>J48</model>
   </mode>
</predictions>
```

Prediction Step

Cost Effectiveness



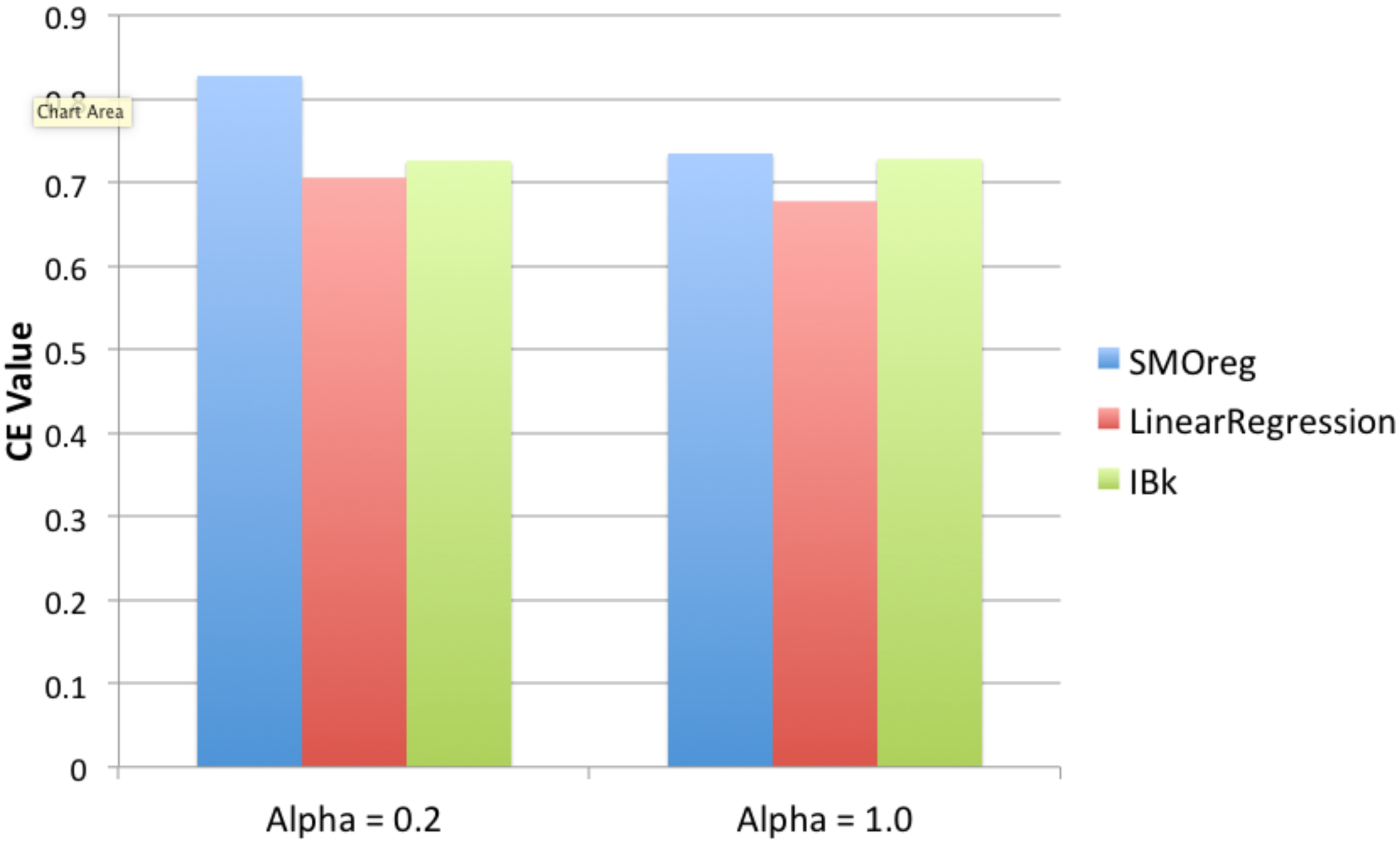
$$CE_\alpha(model) = \frac{AUC_\alpha(model) - AUC_\alpha(random)}{AUC_\alpha(optimal) - AUC_\alpha(random)}$$

Output Step

```
<outputs>
    <ceList>
        <value>0.2</value>
        <value>0.3</value>
        <value>0.4</value>
        <value>1.0</value>
    </ceList>
</outputs>
```

Predicting the number of bugs from change metrics in Equinox

Predicting the number of bugs from change metrics in Equinox



```
<rows-to-remove>
  <item>
    <header>numberOfLinesOfCode</header>
    <value>0</value>
  </item>
</rows-to-remove>
<columns-to-remove>
  <item>
    <header>majorBugs</header>
  </item>
</columns-to-remove>
<class-name-column>
  <header>classname</header>
</class-name-column>
<loc-column>
  <header>numberOfLinesOfCode</header>
</loc-column>
```

NCD-02-A	New Caledoniz	0.68388	12.97743	8.5228	0.169121	1.64142
NCD-02-B	New Caledoniz	0.55044	12.05182	9.8241	0.175576	1.3264
NCD-03-A	New Caledoniz	0.6672	12.75075	8.5371	0.166539	1.70774
NCD-03-B	New Caledoniz	0.20016	3.72133	2.3881	0.049058	0.46424
NCD-04-A	New Caledoniz	0.6672	13.24189	8.4799	0.16783	1.59168
NCD-04-B	New Caledoniz	0.60048	11.57957	7.5647	0.152338	1.47562
NCD-05-A	New Caledoniz	0.38364	7.44266	4.6189	0.096825	0.829
NCD-05-B	New Caledoniz	0.68388	13.43079	8.5514	0.169121	1.62484
NCD-06-A	New Caledoniz	0.65052	12.61852	8.3655	0.16783	1.62484
NCD-06-B	New Caledoniz	0.46704	11.74958	4.6475	0.122645	0.7461
NCD-07-A	New Caledoniz	0.28356	47.0361	3.432	0.069714	0.84558
DUP-NCD-07-A	New Caledoniz	0.28356	44.9582	3.3605	0.067132	0.829
NCD-07-B	New Caledoniz	0.68388	13.4119	8.3226	0.163957	1.77406
NCD-08-A	New Caledoniz	0.6672	13.16633	8.2225	0.162666	1.59168
NCD-08-B	New Caledoniz	0.6672	12.93965	8.5085	0.169121	1.87354
NCD-09-A	New Caledoniz	0.68388	12.95854	8.8803	0.170412	1.70774
NCD-09-B	New Caledoniz	0.6672	12.93965	9.1234	0.179449	1.7409
NC2010P01	New Caledoniz	0.53376	13.4119	3.1317	0.1291	0.72952
NC2010P03	New Caledoniz	0.80064	12.59963	7.722	0.163957	1.27666
NC2010P05-2	New Caledoniz	0.53376	11.50401	5.0479	0.131682	0.76268
NC2010P05-3	New Caledoniz	0.61716	14.96088	6.9784	0.151047	1.84038