# AN OVERVIEW OF USER FEEDBACK CLASSIFICATION APPROACHES

**Rubens Santos, Eduard C. Groen, Karina Villela**

<table>
<thead>
<tr>
<th>ML Algorithm</th>
<th>ML Feature</th>
<th>BOW</th>
<th>BOF</th>
<th>TF-IDF</th>
<th>$\chi^2$</th>
<th>n-Gram</th>
<th>NLP-Heur.</th>
<th>AUR-BOW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve Bayes</td>
<td></td>
<td>18</td>
<td>8</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>37</td>
</tr>
<tr>
<td>Bayesian Network</td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logistic Regression</td>
<td></td>
<td>9</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td></td>
<td></td>
<td>19</td>
</tr>
<tr>
<td>k-Nearest Neighbors</td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Support Vector Machines</td>
<td></td>
<td>16</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td></td>
<td></td>
<td>30</td>
</tr>
<tr>
<td>DT – Single Tree / C4.5</td>
<td></td>
<td>10</td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>26</td>
</tr>
<tr>
<td>DT – Boosted</td>
<td></td>
<td>4</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>DT – Random Forest</td>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>DT – Bagging</td>
<td></td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Neural Networks</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td><strong>No. of pairs</strong></td>
<td></td>
<td>63</td>
<td>2</td>
<td>29</td>
<td>3</td>
<td>9</td>
<td>30</td>
<td>3</td>
</tr>
</tbody>
</table>
What do we want?

A BENCHMARKING
of user feedback classification approaches for RE (CrowdRE)

When do we have it?

Now, see...

the differences between the approaches we found actually make it kind of difficult to make a proper comparison that tells us reliably which approaches may be better suited for RE so that we are still several steps away from performing a benchmarking which may require researchers to re-do analyses or to provide us with their data in order for us to perform those analyses ourselves for their results to be comparable on the various levels that these analyses currently differ to such great extents...not now...
The Idea of Our Benchmarking is Simple…

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>✔️</td>
</tr>
<tr>
<td>2</td>
<td>✔️</td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>✔️</td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

Wrench image source: MidnightLightning / WikiMedia Commons, public domain
...The Reality of this Benchmarking is Difficult...

** Algorithms are used in combination with different combinations of other NLP techniques, including primary and secondary machine learning features, semi-supervised classification algorithms, and pre-processing techniques.

** Datasets** differ, among other things, in size (number of entries), object granularity (sentence vs. review), sources covered (e.g., app stores, social media), and mean text object size.

** Analyses** use different classification categories according to different definitions and gold standards.
Datasets differ, among other things, in size (number of entries), object granularity (sentence vs. review), sources covered (e.g., app stores, social media), and mean text object size.

Hurdle 3 and further:
- Comparing datasets
- Assessing the influence of NLP techniques
- Aligning analyses
- Etc.

Hurdle 1: A taxonomy for user feedback classifications

Analyses use different classification categories according to different definitions and gold standards

Hurdle 2: An overview of user feedback classification approaches

Systematic Literature Review

Benchmarking

Algorithms are used in combination with different combinations of other NLP techniques, including primary and secondary machine learning features, semi-supervised classification algorithms, and pre-processing techniques

...But We Are Doing This Benchmarking

Focus of this presentation
Systematic Literature Review

- Conducted according to Kitchenham, with an SLR protocol specifying:
  - objectives / research questions,
  - a search strategy with inclusion/exclusion criteria & a search string,
  - a data extraction strategy.

➤ Note: The SLR is not the main focus of this presentation!
- We’re showing a “byproduct” in a preliminary form
  - Focusing only on the second hurdle that we had to overcome
- We wanted to get this material out there, so you can work with it!
SLR: Objectives

**Overall Objective:** What are the state-of-the-art automated approaches for assisting the task of requirements extraction from user feedback acquired from the crowd, and which NLP techniques and features do they use?

- **Objective 1:** Regarding requirements elicitation from user feedback acquired from the crowd, what are the state-of-art automated approaches for classifying user feedback?

- **Objective 2:** How do such approaches classify user feedback?
  - **Objective 2.1:** What are the different sets of categories in which user feedback is classified?
  - **Objective 2.2:** Which automated techniques are used?
  - **Objective 2.3:** What are the characteristics of the user feedback these approaches aim to classify?
SLR: Paper Search

Performed March 2018 (+ December 2018)

1,219 papers
  + 14 papers
  • Found by search string

146 papers
  • After filtering by title & abstract

40 papers
  + 3 papers
  • After filtering by intro & conclusion

**EC1:** not English
**EC2:** before 2013
**EC3:** not peer-reviewed

**IC1:** filters out irrelevant user feedback
**IC2:** classifies into predetermined categories

**EC4:** not RE / unrelated title
**EC5:** not on req. extraction from user feedback
**EC6:** tool not (usable) for requirement extraction
**EC7:** tool does not process textual user feedback
**EC8:** manual processing without automation
SLR: Data Extraction from 43 Papers

1. **Dataset-related information**
   - E.g., dataset size in number of entries, object granularity, sources, mean text object size

2. **NLP techniques applied** ➔ Classification approach comparison
   - E.g., algorithms, parsers, ML features, text pre-processing techniques

3. **User feedback classification categories**
   - E.g., name, definition, rationale/goal
Research Focuses on Machine Learning Algorithms

- The SLR found **43 papers** on user feedback classification in RE (CrowdRE)

- Analysis of NLP techniques:
  - **86% used ML algorithms**
  - Mostly several (1 to 14; 3.8 average) → comparative experiments

![Type of Classification](image-url)
Systematic Mapping of Machine Learning Techniques 1/2

- **Primary ML features** that represent the text according to word count or related methods
  - E.g., Bag of Words, Term Frequency – Inverse Document Frequency, Bag of Frames

- **Secondary ML features** that represent specific aspects of the text or metadata. They yield low scores when used on their own, but can help achieve greater efficiency & quality in combination with primary features.
  - E.g., length, sentiment score, star rating

- **Semi-supervised classification algorithms**
  - E.g., Expectation-Maximization, Self-Training, Rasco

- **Pre-processing techniques**
  - E.g., stop words removal, synonym unification, stemming, lemmatization, special characters removal, abbreviation transformation, negation handling
## Frequency of ML Algorithm + ML Technique Pair

<table>
<thead>
<tr>
<th>ML Algorithm</th>
<th>ML Feature</th>
<th>BOW</th>
<th>BOF</th>
<th>TF-IDF</th>
<th>$\chi^2$</th>
<th>n-Gram</th>
<th>NLP-Heur.</th>
<th>AUR-BOW</th>
<th>No. of pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve Bayes</td>
<td>BOW</td>
<td>18</td>
<td>1</td>
<td>8</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>37</td>
</tr>
<tr>
<td>Bayesian Network</td>
<td>BOF</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>TF-IDF</td>
<td>9</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td></td>
<td>19</td>
</tr>
<tr>
<td>k-Nearest Neighbors</td>
<td>$\chi^2$</td>
<td>3</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Support Vector Machines</td>
<td>n-Gram</td>
<td>16</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td></td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>DT – Single Tree / C4.5</td>
<td>NLP-Heur.</td>
<td>10</td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>DT – Boosted</td>
<td>AUR-BOW</td>
<td>4</td>
<td>1</td>
<td>6</td>
<td>4</td>
<td></td>
<td></td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>DT – Random Forest</td>
<td></td>
<td>4</td>
<td>2</td>
<td></td>
<td>2</td>
<td>2</td>
<td></td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>DT – Bagging</td>
<td></td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Neural Networks</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>No. of pairs</strong></td>
<td></td>
<td>63</td>
<td>2</td>
<td>29</td>
<td>3</td>
<td>9</td>
<td>30</td>
<td>3</td>
<td>139</td>
</tr>
</tbody>
</table>

* One measurement uses BOF – Boolean; all others use BOF-TF.
## Frequency of ML Algorithm + ML Technique Pair

<table>
<thead>
<tr>
<th>ML Algorithm</th>
<th>ML Feature</th>
<th>BOW</th>
<th>BOF</th>
<th>TF-IDF</th>
<th>$\chi^2$</th>
<th>n-Gram</th>
<th>NLP-Heur.</th>
<th>AUR-BOW</th>
<th>No. of papers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve Bayes</td>
<td></td>
<td>18</td>
<td></td>
<td>8</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>27</td>
</tr>
<tr>
<td>Bayesian Network</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td></td>
<td>9</td>
<td></td>
<td>2</td>
<td>2</td>
<td>6</td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>k-Nearest Neighbors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Support Vector Machines</td>
<td></td>
<td>16</td>
<td></td>
<td>6</td>
<td>1</td>
<td>6</td>
<td></td>
<td></td>
<td>22</td>
</tr>
<tr>
<td>DT – Single Tree / C4.5</td>
<td></td>
<td>10</td>
<td></td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td>DT – Boosted</td>
<td></td>
<td>4</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>DT – Random Forest</td>
<td></td>
<td>4</td>
<td></td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>DT – Bagging</td>
<td></td>
<td>1</td>
<td></td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Neural Networks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

* One measurement uses BOF – Boolean; all others use BOF-TF.
## Number of "Feature Request" Measurements

<table>
<thead>
<tr>
<th>ML Algorithm</th>
<th>BOW</th>
<th>BOF</th>
<th>TF-IDF</th>
<th>$\chi^2$</th>
<th>n-Gram</th>
<th>NLP-Heur.</th>
<th>AUR-BOW</th>
<th>Measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve Bayes</td>
<td>9 *</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>1 *</td>
<td>2</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>Bayesian Network</td>
<td>2 *</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2 **</td>
<td>1</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>7 *</td>
<td>1</td>
<td>3</td>
<td>1 *</td>
<td>2</td>
<td>1</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>k-Nearest Neighbors</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Support Vector Machines</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>DT – Single Tree / C4.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>DT – Boosted</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>DT – Random Forest</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Neural Networks</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

* Includes one $F_1$ measure for which no precision/recall values could be obtained
** Includes two $F_1$ measures for which no precision/recall values could be obtained
### $F_\beta$ Measures for “Feature Request”

<table>
<thead>
<tr>
<th>ML Algorithm</th>
<th>BOW</th>
<th>BOF</th>
<th>TF-IDF</th>
<th>$\chi^2$</th>
<th>n-Gram</th>
<th>NLP-Heur.</th>
<th>AUR-BOW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve Bayes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.65</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.40</td>
</tr>
<tr>
<td>k-Nearest Neighbors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.53</td>
</tr>
<tr>
<td>Support Vector Machines</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.68</td>
</tr>
<tr>
<td>DT – Single Tree / C4.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.64</td>
</tr>
<tr>
<td>DT – Boosted</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.66</td>
</tr>
<tr>
<td>DT – Random Forest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.84</td>
</tr>
<tr>
<td>DT – Bagging</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.66</td>
</tr>
<tr>
<td>Neural Networks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.39</td>
</tr>
<tr>
<td><strong>Averages</strong></td>
<td>0.68</td>
<td>0.81</td>
<td>0.55</td>
<td>0.68</td>
<td>0.68</td>
<td>0.70</td>
<td>0.64</td>
</tr>
</tbody>
</table>

**Task-based $\beta_T$, calculated as $\frac{1}{\lambda}$**

$$F_\beta = (1 + \beta^2) \times \frac{P \times R}{(\beta^2 \times P) + R}$$
The Most Popular…

- **User feedback analysis approach:** Machine Learning
  - Only few use dictionaries, regular expressions or parsing

- **ML algorithms:** NB, SVM, LR, and DT (esp. Single Tree)
  - Probably because they provide a relatively large degree of control over the supervised ML
  - 12 clusters found in total

- **Primary ML features:** BOW and TF-IDF
  - Probably because of their versatile nature
  - 7 clusters found in total

- **ML models:** NB + BOW, SVM + BOW, NB + TF-IDF
  - Probably because of their tool support and familiarity
User Feedback Analysis Research for RE: Still Got a Long Long Way to Go

- $F_\beta$ Measures for “Feature Requests” were surprisingly **moderate**
  - Especially assuming publication bias (best possible outcomes)
  - Four ML models had $F_\beta > 0.85$, but for just one measurement

- All popular ML algorithms can potentially result in **good-quality results**
  - Study characteristics we did not investigate seem to have a strong impact on classification efficiency

- **Strong variance** in ML models used & study set-ups
  - Research is still exploring appropriate ML models
By Not Taking Inspiration from Other Works, CrowdRE Research is Missing Out on Opportunities!

- **NLP Heuristics and $n$-Grams** have been shown to contribute to better results by introducing context information into the classification task.

- No research has picked up on adaptations of BOW in CrowdRE research that yielded good results: **Bag of Frames** (P20) and **Augmented User Reviews – BOW** (P24).

- Other works may obtain better results for **Bayesian Network** as in P43, or **Neural Networks** (or another Deep Learning approach) than in P14.

- Works investigating **non-ML approaches** for RE suggest that carefully designed heuristics may in some cases also provide accurate results (i.e., high precision), but not necessarily contribute to higher recall.

- Researchers could help others if they provide a **rationale** for their choice of techniques, which we hardly saw in research.
Implications and Outlook

- Our findings can help you (and us) make a more informed choice of appropriate ML algorithms and ML features to achieve better user feedback classification for RE.

- **No decisive conclusions** about the most suitable ML models.
  - Factors other than the ones we considered in this work appear to have had a strong influence on the performance of the ML models.
  - We did find that good results have been attained with the most often used ML algorithms, especially when used in combination with appropriate primary and secondary ML features.

- The current landscape is still one of exploration into the most suitable techniques, but progress is hindered by a lack of cross-fertilization.
  - Research does not pick up on promising findings in other works to investigate whether these approaches work well in their context.
Future Work

For our benchmarking study, these findings further fuel the need for an evaluation of user feedback analysis techniques for different purposes.

On the other hand, the potential of non-ML approaches reported in some works should not be ignored either.

This work was descriptive in nature and was limited to a comparison of only the ML algorithms and primary ML features.

More prescriptive results could be obtained through an assessment of which study-specific aspects impact performance most strongly.

E.g., addressed goals & problems; dataset type/quality/size; classification categories chosen; gold standard composed; additional ML techniques used (semi-supervised classification algorithms / pre-processing techniques / secondary ML features).

Due to the study’s set-up, we did not investigate the performance of ML models in other contexts within and outside of RE.
## Thank you!

<table>
<thead>
<tr>
<th>ML Algorithm</th>
<th>ML Feature</th>
<th>BOW</th>
<th>BOF</th>
<th>TF-IDF</th>
<th>$\chi^2$</th>
<th>n-Gram</th>
<th>NLP-Heur.</th>
<th>AUR-BOW</th>
<th>No. of papers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve Bayes</td>
<td></td>
<td>18</td>
<td>1</td>
<td>8</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>27</td>
</tr>
<tr>
<td>Bayesian Network</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td></td>
<td>9</td>
<td></td>
<td>2</td>
<td></td>
<td>2</td>
<td>6</td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>k-Nearest Neighbors</td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Support Vector Machines</td>
<td></td>
<td>16*</td>
<td>1</td>
<td>6</td>
<td></td>
<td>1</td>
<td>6</td>
<td></td>
<td>22</td>
</tr>
<tr>
<td>DT – Single Tree / C4.5</td>
<td></td>
<td>10</td>
<td></td>
<td>6</td>
<td></td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>17</td>
</tr>
<tr>
<td>DT – Boosted</td>
<td></td>
<td>4</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>DT – Random Forest</td>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>2</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>DT – Bagging</td>
<td></td>
<td>1</td>
<td></td>
<td>2</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Neural Networks</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

**No. of papers:** 23, 1, 12, 1, 1, 5, 8, 1
## Thank you!

<table>
<thead>
<tr>
<th>ML Algorithm</th>
<th>BOW</th>
<th>BOF</th>
<th>TF-IDF</th>
<th>$\chi^2$</th>
<th>n-Gram</th>
<th>NLP-Heur.</th>
<th>AUR-BOW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve Bayes</td>
<td><img src="image1" alt="Heatmap" /></td>
<td><img src="image2" alt="Heatmap" /></td>
<td><img src="image3" alt="Heatmap" /></td>
<td><img src="image4" alt="Heatmap" /></td>
<td><img src="image5" alt="Heatmap" /></td>
<td><img src="image6" alt="Heatmap" /></td>
<td>0.65</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td><img src="image7" alt="Heatmap" /></td>
<td><img src="image8" alt="Heatmap" /></td>
<td><img src="image9" alt="Heatmap" /></td>
<td><img src="image10" alt="Heatmap" /></td>
<td><img src="image11" alt="Heatmap" /></td>
<td><img src="image12" alt="Heatmap" /></td>
<td>0.40</td>
</tr>
<tr>
<td>k-Nearest Neighbors</td>
<td><img src="image13" alt="Heatmap" /></td>
<td><img src="image14" alt="Heatmap" /></td>
<td><img src="image15" alt="Heatmap" /></td>
<td><img src="image16" alt="Heatmap" /></td>
<td><img src="image17" alt="Heatmap" /></td>
<td><img src="image18" alt="Heatmap" /></td>
<td>0.53</td>
</tr>
<tr>
<td>Support Vector Machines</td>
<td><img src="image19" alt="Heatmap" /></td>
<td><img src="image20" alt="Heatmap" /></td>
<td><img src="image21" alt="Heatmap" /></td>
<td><img src="image22" alt="Heatmap" /></td>
<td><img src="image23" alt="Heatmap" /></td>
<td><img src="image24" alt="Heatmap" /></td>
<td>0.68</td>
</tr>
<tr>
<td>DT – Single Tree / C4.5</td>
<td><img src="image25" alt="Heatmap" /></td>
<td><img src="image26" alt="Heatmap" /></td>
<td><img src="image27" alt="Heatmap" /></td>
<td><img src="image28" alt="Heatmap" /></td>
<td><img src="image29" alt="Heatmap" /></td>
<td><img src="image30" alt="Heatmap" /></td>
<td>0.64</td>
</tr>
<tr>
<td>DT – Boosted</td>
<td><img src="image31" alt="Heatmap" /></td>
<td><img src="image32" alt="Heatmap" /></td>
<td><img src="image33" alt="Heatmap" /></td>
<td><img src="image34" alt="Heatmap" /></td>
<td><img src="image35" alt="Heatmap" /></td>
<td><img src="image36" alt="Heatmap" /></td>
<td>0.66</td>
</tr>
<tr>
<td>DT – Random Forest</td>
<td><img src="image37" alt="Heatmap" /></td>
<td><img src="image38" alt="Heatmap" /></td>
<td><img src="image39" alt="Heatmap" /></td>
<td><img src="image40" alt="Heatmap" /></td>
<td><img src="image41" alt="Heatmap" /></td>
<td><img src="image42" alt="Heatmap" /></td>
<td>0.84</td>
</tr>
<tr>
<td>DT – Bagging</td>
<td><img src="image43" alt="Heatmap" /></td>
<td><img src="image44" alt="Heatmap" /></td>
<td><img src="image45" alt="Heatmap" /></td>
<td><img src="image46" alt="Heatmap" /></td>
<td><img src="image47" alt="Heatmap" /></td>
<td><img src="image48" alt="Heatmap" /></td>
<td>0.66</td>
</tr>
<tr>
<td>Neural Networks</td>
<td><img src="image49" alt="Heatmap" /></td>
<td><img src="image50" alt="Heatmap" /></td>
<td><img src="image51" alt="Heatmap" /></td>
<td><img src="image52" alt="Heatmap" /></td>
<td><img src="image53" alt="Heatmap" /></td>
<td><img src="image54" alt="Heatmap" /></td>
<td>0.39</td>
</tr>
<tr>
<td><strong>Averages</strong></td>
<td>0.68</td>
<td>0.81</td>
<td>0.55</td>
<td>0.68</td>
<td>0.68</td>
<td>0.70</td>
<td>0.63</td>
</tr>
</tbody>
</table>