Doing Something Useless
Slightly Faster: The State of the Art in Time Series Data Mining?

Eamonn Keogh

Computer Science & Engineering Department
University of California - Riverside
Riverside, CA 92521
eamonn@cs.ucr.edu
Outline of Talk

- Our Claim
- Results of our Survey
  - Size of test datasets
  - Number of rival methods considered
  - Diversity of test datasets
- Implementation Bias
  - What it is, why it matters
- Data Bias
  - What it is, why it matters
- Case Study: Similarity Measures
  - Subjective testing
  - Objective testing
- Concrete Suggestions
- Conclusions/Questions
Important Note

• We are anxious that this work should not be taken as being critical of the data mining community.

• **We note that several papers by the current first author are among the worst offenders in terms of weak experimental evaluation!!!**

• Our goal is simply to demonstrate that empirical evaluations in the past have often been inadequate, and we hope this work will encourage more extensive experimental evaluations in the future.
Our Claim

Much of the work in the time series data mining literature suffers from two types of experimental flaws, *implementation bias* and *data bias* (defined later).

Because of these flaws, much of the work has very little generalizability to real world problems.
In More Detail...

Many of the contributions offer an amount of “improvement” that would have been completely dwarfed by the variance that would have been observed by testing on many real world datasets, or the variance that would have been observed by changing minor (unstated) implementation details.
Time Series Data Mining Tasks

Indexing (Query by Content): Given a query time series $Q$, and some similarity/dissimilarity measure $D(Q,C)$, find the nearest matching time series in database DB.

Clustering: Find natural groupings of the time series in database DB under some similarity/dissimilarity measure $D(Q,C)$.

Classification: Given an unlabeled time series $Q$, assign it to one of two or more predefined classes.

Segmentation: Given a time series $Q$ containing $n$ datapoints, construct a model $\overline{O}$, from $K$ piecewise segments ($K \ll n$) such that $\overline{O}$ closely approximates $Q$. 
What do we want to do with the time series data?

Clustering
Classification

Motif Discovery
Rule Discovery

Query by Content

Visualization
Novelty Detection

\[ s = 0.5 \]
\[ c = 0.3 \]
Literature Survey

We read more than 340 papers, but we only included the subset of 56 papers actually cited in our paper when assessing statistics.

The subset was chosen based on the following criteria.

- Was the paper ever referenced?
- Was the paper published in a conference or journal likely to be read by a data miner?

In general the papers come from high quality conferences (SIG)KDD (11), ICDE (10), VLDB (5), SIGMOD/PODS (5), and CIKM (6).
A Cautionary Note

In presenting the results of the survey, we echo the caution of Prechelt, that “while high numbers resulting from such counting cannot prove that the evaluation has high quality, low numbers (suggest) that the quality is low”.

Finding 1: Size of Test Datasets

We recorded the size the test dataset for each paper. Where two or more datasets are used, we considered only the size of the largest.

The median size of the test databases was only 10,000 objects. Approximately 84% of the test databases are less than one megabyte in size.

This number only reflects the indexing papers, the other papers have even smaller sizes.
Finding 2: Number of Rival Methods

We recorded the number of rival methods to which the contribution of each paper is compared.

The median number is 1 (average is 0.91)

This number is even worse than it seems, because many of the strawman are very unrealistic. More about this later…

This number reflects all papers included in the survey.
Finding 3: Number of Test Datasets

We recorded number of different datasets used in the experimental evaluation.

The average is **1.85 datasets** (1.26 real and 0.59 synthetic)

In fact, this number may be optimistic, if you count stock market data as being the same as random walk data, then…

The average is **1.28 datasets**
Having seen the statistics, let us see why these low numbers are a real problem...
**Data Bias**

**Definition:** *Data bias* is the conscious or unconscious use of a particular set of testing data to confirm a desired finding.

**Example:** Suppose you are comparing Wavelets to Fourier methods, the following datasets will produce drastically different results…

![Graphs showing data bias](image)
Example of Data Bias: Who to Believe?

For the task of indexing time series for similarity search, which representation is best, the Discrete Fourier Transform (DFT), or the Discrete Wavelet Transform (Haar)?

• “Several wavelets outperform the DFT”.
• “DFT-based and DWT-based techniques yield comparable results”.
• “Haar wavelets perform slightly better than DFT”
• “DFT filtering performance is superior to DWT*”
Discrete Wavelet Transform

\[ [9 \ 7 \ 3 \ 5] \]

\[ I(x) = 9 \times \]
\[ + 7 \times \]
\[ + 3 \times \]
\[ + 5 \times \]

\[ I(x) = c_0^0 \phi_0^0(x) + d_0^0 \psi_0^0(x) + d_1^1 \psi_1^1(x) \]

\[ = 6 \times \]
\[ + 2 \times \]
\[ + 1 \times \]
\[ + -1 \times \]

Example of Data Bias: Who to Believe II?

To find out who to believe (if anyone) we performed an extraordinarily careful and comprehensive set of experiments. For example…

• We used a quantum mechanical device generate random numbers.
• We averaged results over 100,000 experiments!
• For fairness, we use the same (randomly chosen) subsequences for both approaches.

• More details in the paper…
Take another quick look at the conflicting claims, the next slide will tell us who was correct…

• “Several wavelets outperform the DFT”.

• “DFT-based and DWT-based techniques yield comparable results”.

• “Haar wavelets perform slightly better than DFT”

• “DFT filtering performance is superior to DWT*”
I tested on the Powerplant, Infrasound and Attas datasets, and I know DFT outperforms the Haar wavelet.
Stupid Flanders! I tested on the Network, ERPdata and Fetal EEG datasets and I know that there is no real difference between DFT and Haar.
Those two clowns are both wrong! I tested on the Chaotic, Earthquake and Wind datasets, and I am sure that the Haar wavelet outperforms the DFT.
Any claims about the relative performance of a time series indexing scheme that is empirically demonstrated on only 2 or 3 datasets should be viewed with suspicion.
**Implementation Bias**

**Definition:** *Implementation bias* is the conscious or unconscious disparity in the quality of implementation of a proposed approach, vs. the quality of implementation of the completing approaches.

**Example:** Suppose you want to compare your new representation to DFT. You might use the simple $O(n^2)$ DFT algorithm rather than spend the time to code the more complex $O(n \log n)$ radix 2 algorithm. This would make your algorithm run relatively faster.
Example of Implementation Bias: Similarity Searching

**Algorithm** sequential_scan(data, query)

*best_so_far*= \( \text{inf} \);
for every item in the database
    if euclidean_dist(data\(_i\), query) < best_so_far
        pointer_to_best_match = \(i\);
        best_so_far = euclidean_dist(data\(_i\), query);
    end;
end;

**Algorithm** accum = euclidean_dist(d, q);
accum = 0;
for \(i = 1\) to length(d)
    accum = accum + (d\(_i\) - q\(_i\))^2
end;
accum = sqrt(accum);

**Optimization 1:**
Neglect to take to square root

**Optimization 2:**
Pass the *best_so_far* into the euclidean_dist function, and abandon the calculation if *accum* ever gets larger than *best_so_far*
Results of a Similarity Searching Experiment on Increasingly Large Datasets

This trivial optimizations can have differences which are larger than many of the speedups claimed for new techniques.

This experiment only considers the main memory case, disk based techniques offer many other possibilities for implementation bias.
Another Example of Implementation Bias I

Researchers have found a difference between the indexing ability of Haar and PAA…

We believe that the reported result might be the result of an (unstated) implementation detail.

For normalized data, the first Haar coefficient is always zero.

What kind of difference would it make if we forgot that fact?

Let's find out!
Another Example of Implementation Bias II

We re-implemented the experiments, averaging the results over 100 experiments as the original authors did. We compared the results of two implementations, one that takes advantage of the “1st coefficient is zero”, and one that does not.

We tested on 50 datasets, and plotted the ratio of the results of both implementations as a histogram.

This minor (unstated) implementation detail can have effects larger than the claimed improvement!
The Bottom Line

• There are many many possibilities for implementation bias when implementing a strawman.

• Minor changes in these minor details can have effects as large as the claimed improvement.

• Experiments should be done in a manner which is completely free of implementation bias.

• We review such techniques in the paper.
Similarity Measures

Many papers in our survey have introduced a new similarity measure for time series, do they make a contribution?

- **61.9%** of the papers fail to show a single example of a matching time series under the measure.
- **52.3%** of the papers fail to compare their new measure to a single strawman.
- **85.7%** of the papers fail to demonstrate an objective test of their new measure.
38.1% of the papers do show an example of a query and a match, but what does that mean if we don’t see all the non-matches?
38.1% of the papers do show an example of a query and a match, but what does that mean if we don’t see all the non-matches? ... it means little
We believe that one of the best (subjective) ways to evaluate a proposed similarity measure is to use it to create a dendrogram of several time series from the domain of interest.
Objective Evaluation of Similarity Measures

We can use nearest neighbor classification to evaluate the usefulness of propose distance measures.

We compared 11 different measures to Euclidean distance, using 1-NN.

Some of the measures require the user to set some parameters. In these cases we wrapped the classification algorithm in a loop for each parameter, searched over all possible parameters and reported only the best result.

Cylinder-Bell-Funnel: This synthetic dataset has been in the literature for 8 years, and has been cited at least a dozen times. It is a 3-class problem; we create 128 examples of each class for these experiments.

Control-Chart: This synthetic dataset has been freely available for the UCI Data Archive since June 1998. It is a 6-class problem, with 100 examples of each class.
# Results: Classification Error Rates

<table>
<thead>
<tr>
<th>Approach</th>
<th>Cylinder-Bell-F'</th>
<th>Control-Chart</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Euclidean Distance</em></td>
<td>0.003</td>
<td>0.013</td>
</tr>
<tr>
<td>Aligned Subsequence</td>
<td>0.451</td>
<td>0.623</td>
</tr>
<tr>
<td>Piecewise Normalization</td>
<td>0.130</td>
<td>0.321</td>
</tr>
<tr>
<td>Autocorrelation Functions</td>
<td>0.380</td>
<td>0.116</td>
</tr>
<tr>
<td>Cepstrum</td>
<td>0.570</td>
<td>0.458</td>
</tr>
<tr>
<td>String (Suffix Tree)</td>
<td>0.206</td>
<td>0.578</td>
</tr>
<tr>
<td>Important Points</td>
<td>0.387</td>
<td>0.478</td>
</tr>
<tr>
<td>Edit Distance</td>
<td>0.603</td>
<td>0.622</td>
</tr>
<tr>
<td>String Signature</td>
<td>0.444</td>
<td>0.695</td>
</tr>
<tr>
<td>Cosine Wavelets</td>
<td>0.130</td>
<td>0.371</td>
</tr>
<tr>
<td>Hölder</td>
<td>0.331</td>
<td>0.593</td>
</tr>
<tr>
<td>Piecewise Probabilistic</td>
<td>0.202</td>
<td>0.321</td>
</tr>
</tbody>
</table>
The Bottom Line

Almost all papers that introduced a new distance measure as their primary contribution, failed to demonstrate the utility of their measure on a single objective or subjective test.

To the best of our knowledge there are no distance measures in the literature that are better than the decades old Euclidean Distance and Dynamic Time Warping.
The **Overall** Bottom Line

The time series data mining community is generally doing very poor quality work.
Concrete Suggestion I

Algorithms should be tested on a wide range of datasets, unless the utility of the approach is only been claimed for a particular type of data.

If possible, one subset of the datasets should be used to fine tune the approach, then a different subset of the datasets should be used to do that the actual testing. This methodology is widely used in the machine learning community to help prevent implementation and data bias.
Concrete Suggestion II

Where possible, experiments should be designed to be free of the possibility of implementation bias.

Note that this does not preclude the addition of extensive implementation testing.
Concrete Suggestion III

Novel similarity measures should be compared to simple strawman, such as Euclidean distance or Dynamic Time Warping. Some subjective visualization, or objective experiments should justify their introduction.
Concrete Suggestion III

Where possible, all data and code
used in experiments should be made
freely available to allow independent
duplication of findings

(as is the case for this paper)
Recent Advances in Mining Time Series Data...

Why most of them are not really “advances”...
Overview of Talk

• Introduction and background

• Why 90% of data mining papers make no contribution
  – Lack of reproducibility
  – Solving problems that don’t exist
  – Crippling the strawman
  – Overfitting (Arbitrarily Stupid Algorithms)

• What can be done?

• Conclusions and Questions
Claim 1: The vast majority of papers in time series data mining actually make no contribution

Claim 2: It does not have to be this way
Reproducibility

At the time of submission, my papers have a website associated with them. And the paper itself points to the website.

The website has *every* dataset available for download.

Anyone in the world that reviews or reads my papers can have the data, without having to rely on me reading their email, and sending them the data.
After asking 9 times over 18 months, I finally got the data. At the same time I got an email from the second author saying…

“The bad news for us is that the software clearly contains some bugs.. however, the bugs are less obvious that we would hope. Thus, we still haven't identified the bugs.” and “I suggested that we withdraw the TODS (journal extension) paper.”

My testing on the original data confirms that Chebyshev is no better than PAA, APCA, DFT etc (actually it is consistently slightly worse). I appreacte the authors (eventually) sharing the data.
Reproducibility Summary

Reproducibility is what differentiates *science* from *pseudoscience*.

If we are not willing to do reproducible work, we might as well be doing astrology.

Reproducibility is not a *bonus*, or *desirable* property. It should be treated as a *necessary* condition for publishing work.
Solving problems that don’t exist

• We have seen how to compare two time series of the same length, the Euclidean distance.

• Suppose the time series are of different lengths?

• We can just make one shorter or the other one longer..

\[ D(Q,C) \]

\[ C_{\text{new}} = \text{resample}(C, \text{length}(Q), \text{length}(C)) \]

It takes one line of matlab code.
Arbitrarily Stupid Algorithm for Prediction

A recent paper in SIGMOD claims to be able to do prediction in stock market data at about 68% accuracy.

The first author acknowledged to me that they were training and testing on the same data (but did not feel that this was a problem).

I gave them “stock data” to test on, they got 62% accuracy. But the data I gave them was really random walk!

My arbitrarily stupid algorithm for prediction has about the same performance as their approach…
Does all this really matter?

• Data mining is in danger of losing credibility if we allow such sloppy work.

• Most of us do work which is ultimately funded by the taxpayers in our home country, the taxpayers deserve better.

• Many of us waste time or opportunities because we cannot be sure of others results.

• Finally, we are supposed to be doing science, one of the noblest of all professions.
What Can We Do About This? I

• Make our own work 100% reproducible for reviewer and readers. (Must be *easily* reproducible)
• Expect (*insist*) on reproducibly as reviewers, editors, advisors and program chairs.
• Demonstrate that the problem at hand is a real problem, not just that we can solve it fast.
• Test in ways that make it hard to cripple the strawman.
• Stop overfitting! Get real data, do the kind of experiments that would satisfy a MLJ reviewer.
What Can We Do About This? II

• We should allow and encourage the publication of papers that do nothing but confirm others results.
• We should allow and encourage the publication of papers that try a good idea, but did not succeed.
• “Finney, Wakker, Kaelbling and Oates: The Thing that we Tried Didn't Work very Well: Deictic Representation in Reinforcement Learning. UAI 2002: 154-16”
An Important Note

The problem is “us”, not “you” or “them”.

• I cannot reproduce most of my experiments before 2000.
• My PAKDD00 paper accidentally cripples the main rival.
• My UAI99 paper does overfitting in some of the experiments.

• I am not claiming that my work is immune to the problems discussed in this talk.
• I am claiming that I want to do something about it.
• I am claiming that I welcome fair criticism of my work, and being kept to a high standard.
References


How to compare two classifiers: Steven L. Salzberg, On Comparing Classifiers: Pitfalls to Avoid and a Recommended Approach, Data Mining and Knowledge Discovery, v.1 n.3, p.317-328, 1997

The problem of Overfitting: C. Elkan. Magical Thinking in Data Mining: Lessons From CoIL Challenge 2000 .(KDD'01), pp. 426-43


Nice examples of people committed to reproducible research:
Roger Peng http://sandybox.typepad.com/reproducible/
Robert Gentleman “Reproducible Research: A Bioinformatics Case Study Statistical Applications in Genetics and Molecular Biology” ISSN: 1544-6115
Acknowledgements

• Many thanks to Luis Torgo, Pavel Brazdil, Rui Camacho, João Gama, Alípio Jorge and all the organizers of this conference.
• Note that I was somewhat deceptive in summarizing the contents of my talk to them, they are not responsible for any controversial comments made, and they do not necessarily endorse them.

Many people have offered suggestions and advice, I will refrain from mentioning them, but many thanks!
Questions?

Thanks to Michael Pazzani, Pedro Domingos, Dimitrios Gunopulos and the SIGKDD reviewers for their useful comments.

All datasets and code used paper can be found at..

www.cs.ucr.edu/~eamonn/TSDMA/index.html
Eamonn Keogh
Assistant Professor
Computer Science & Engineering Department
University of California - Riverside
Riverside, CA 92521

Phone: (951) 827-2032
Fax: (951) 827-3188
eamonn(AT)cs.ucr.edu

Welcome to my home page. You may be interested in:

- My publications
- My tutorials
- The UCR Time Series Data Mining Archive
- New! The UCR Time Series Classification/Clustering page has the largest collection of test datasets in the world.
- SAX is the best representation for most time series problems
- LB_Keogh is the best way to index time series, monitor streams and index shapes
- What does DNA look like?
<table>
<thead>
<tr>
<th>Name</th>
<th>First paper</th>
<th>Number of classes</th>
<th>Size of training set</th>
<th>Size of testing set</th>
<th>Time series Length</th>
<th>1-NN Euclidean Distance</th>
<th>1-NN Best Warping Window DTW (r)</th>
<th>1-NN DTW, no Warping Window DTW (r)</th>
<th>Your Method!</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synthetic Control</td>
<td>Pham</td>
<td>6</td>
<td>300 train</td>
<td>300 test</td>
<td>60</td>
<td>0.12</td>
<td>0.017 (6)</td>
<td>0.007</td>
<td></td>
</tr>
<tr>
<td>Gun-Point</td>
<td>Ratanamahatana</td>
<td>2</td>
<td>50</td>
<td>150</td>
<td>150</td>
<td>0.087</td>
<td>0.087 (0)</td>
<td>0.093</td>
<td></td>
</tr>
<tr>
<td>CBF</td>
<td></td>
<td>3</td>
<td>30</td>
<td>900</td>
<td>128</td>
<td>0.148</td>
<td>0.004 (11)</td>
<td>0.003</td>
<td></td>
</tr>
<tr>
<td>Face (all)</td>
<td>Xi</td>
<td>14</td>
<td>560</td>
<td>1,690</td>
<td>131</td>
<td>0.286</td>
<td>0.192 (3)</td>
<td>0.192</td>
<td></td>
</tr>
<tr>
<td>OSU Leaf</td>
<td>Gandhi</td>
<td>6</td>
<td>200</td>
<td>242</td>
<td>427</td>
<td>0.483</td>
<td>0.384 (7)</td>
<td>0.409</td>
<td></td>
</tr>
<tr>
<td>Swedish Leaf</td>
<td>Söderkvist</td>
<td>15</td>
<td>500</td>
<td>625</td>
<td>128</td>
<td>0.213</td>
<td>0.157 (2)</td>
<td>0.210</td>
<td></td>
</tr>
<tr>
<td>50Words</td>
<td>Rath</td>
<td>50</td>
<td>450</td>
<td>455</td>
<td>270</td>
<td>0.369</td>
<td>0.242 (6)</td>
<td>0.310</td>
<td></td>
</tr>
<tr>
<td>Trace</td>
<td>Riess</td>
<td>4</td>
<td>100</td>
<td>100</td>
<td>275</td>
<td>0.24</td>
<td>0.013 (3)</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>Two Patterns</td>
<td>Geurts</td>
<td>4</td>
<td>1,000</td>
<td>4,000</td>
<td>128</td>
<td>0.09</td>
<td>0.0015 (4)</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>Wafer</td>
<td>Olszewski</td>
<td>2</td>
<td>1,000</td>
<td>6,174</td>
<td>152</td>
<td>0.005</td>
<td>0.005 (1)</td>
<td>0.020</td>
<td></td>
</tr>
<tr>
<td>Face (four)</td>
<td>Ratanamahatana</td>
<td>4</td>
<td>24</td>
<td>88</td>
<td>350</td>
<td>0.216</td>
<td>0.114 (2)</td>
<td>0.170</td>
<td></td>
</tr>
<tr>
<td>Lightning-2</td>
<td>Eads</td>
<td>2</td>
<td>60</td>
<td>61</td>
<td>637</td>
<td>0.246</td>
<td>0.131 (6)</td>
<td>0.131</td>
<td></td>
</tr>
<tr>
<td>Lightning-7</td>
<td>Eads</td>
<td>7</td>
<td>70</td>
<td>73</td>
<td>319</td>
<td>0.425</td>
<td>0.288 (5)</td>
<td>0.274</td>
<td></td>
</tr>
<tr>
<td>ECG</td>
<td>Olszewski</td>
<td>2</td>
<td>100</td>
<td>100</td>
<td>96</td>
<td>0.12</td>
<td>0.12 (0)</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>Adiac</td>
<td>Jalba</td>
<td>37</td>
<td>390</td>
<td>391</td>
<td>176</td>
<td>0.389</td>
<td>0.391 (3)</td>
<td>0.396</td>
<td></td>
</tr>
<tr>
<td>Yoga</td>
<td>Xi</td>
<td>2</td>
<td>300</td>
<td>3,000</td>
<td>426</td>
<td>0.170</td>
<td>0.155 (2)</td>
<td>0.164</td>
<td></td>
</tr>
<tr>
<td>Fish (readme)</td>
<td>Lee</td>
<td>7</td>
<td>175</td>
<td>175</td>
<td>463</td>
<td>0.217</td>
<td>0.160 (4)</td>
<td>0.167</td>
<td></td>
</tr>
<tr>
<td>Plane</td>
<td>readme</td>
<td>7</td>
<td>105</td>
<td>175</td>
<td>144</td>
<td>0.038</td>
<td>0.0 (5)</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>Car</td>
<td>readme</td>
<td>4</td>
<td>60</td>
<td>60</td>
<td>577</td>
<td>0.267</td>
<td>0.23 (1)</td>
<td>0.267</td>
<td></td>
</tr>
<tr>
<td>Beef</td>
<td>Tony Bagnall</td>
<td>5</td>
<td>30</td>
<td>30</td>
<td>470</td>
<td>0.467</td>
<td>0.467</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>Coffee</td>
<td>Tony Bagnall</td>
<td>2</td>
<td>28</td>
<td>28</td>
<td>286</td>
<td>0.25</td>
<td>0.179</td>
<td>0.179</td>
<td></td>
</tr>
<tr>
<td>OliveOil</td>
<td>Tony Bagnall</td>
<td>4</td>
<td>30</td>
<td>30</td>
<td>570</td>
<td>0.133</td>
<td>0.167</td>
<td>0.133</td>
<td></td>
</tr>
</tbody>
</table>

*Please donate data!*
## Clustering:

<table>
<thead>
<tr>
<th>Name</th>
<th>First paper</th>
<th>Number of classes</th>
<th>Size of dataset</th>
<th>Time series Length</th>
<th>Euclidean Distance K-Means, Best of 10 Runs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synthetic Control</td>
<td>Pham</td>
<td>6</td>
<td>300+300</td>
<td>60</td>
<td>0.679</td>
</tr>
<tr>
<td>Gun-Point</td>
<td>Ratanamahatana</td>
<td>2</td>
<td>50+150</td>
<td>150</td>
<td>0.500</td>
</tr>
<tr>
<td>CBF</td>
<td></td>
<td>3</td>
<td>30+900</td>
<td>128</td>
<td>0.626</td>
</tr>
<tr>
<td>Face (all)</td>
<td>Xi</td>
<td>14</td>
<td>560+1,690</td>
<td>131</td>
<td>0.36</td>
</tr>
<tr>
<td>OSU Leaf</td>
<td>Gandhi</td>
<td>6</td>
<td>200+242</td>
<td>427</td>
<td>0.378</td>
</tr>
<tr>
<td>Swedish Leaf</td>
<td>Söderkvist</td>
<td>15</td>
<td>500+625</td>
<td>128</td>
<td>0.406</td>
</tr>
<tr>
<td>50Words</td>
<td>Rath</td>
<td>50</td>
<td>450+455</td>
<td>270</td>
<td>0.420</td>
</tr>
<tr>
<td>Trace</td>
<td>Rovero</td>
<td>4</td>
<td>100+100</td>
<td>275</td>
<td>0.485</td>
</tr>
<tr>
<td>Two Patterns</td>
<td>Geurts</td>
<td>4</td>
<td>1,000+4000</td>
<td>128</td>
<td>0.322</td>
</tr>
<tr>
<td>Wafer</td>
<td>Olszewski</td>
<td>2</td>
<td>1,000+6,174</td>
<td>152</td>
<td>0.625</td>
</tr>
<tr>
<td>Face (four)</td>
<td>Ratanamahatana</td>
<td>4</td>
<td>24+88</td>
<td>350</td>
<td>0.669</td>
</tr>
<tr>
<td>Lightning-2</td>
<td>Eads</td>
<td>2</td>
<td>60+61</td>
<td>637</td>
<td>0.611</td>
</tr>
<tr>
<td>Lightning-7</td>
<td>Eads</td>
<td>7</td>
<td>70+73</td>
<td>319</td>
<td>0.484</td>
</tr>
<tr>
<td>ECG</td>
<td>Olszewski</td>
<td>2</td>
<td>100+100</td>
<td>96</td>
<td>0.698</td>
</tr>
<tr>
<td>Adiacc</td>
<td>Jalba</td>
<td>37</td>
<td>390+391</td>
<td>176</td>
<td>0.384</td>
</tr>
<tr>
<td>Yoga</td>
<td>Xi</td>
<td>2</td>
<td>300+3,000</td>
<td>426</td>
<td>0.517</td>
</tr>
</tbody>
</table>
Review Questions

1. Name one way to help avoid the problem of implementation bias

2. Name one way to help avoid the problem of data bias