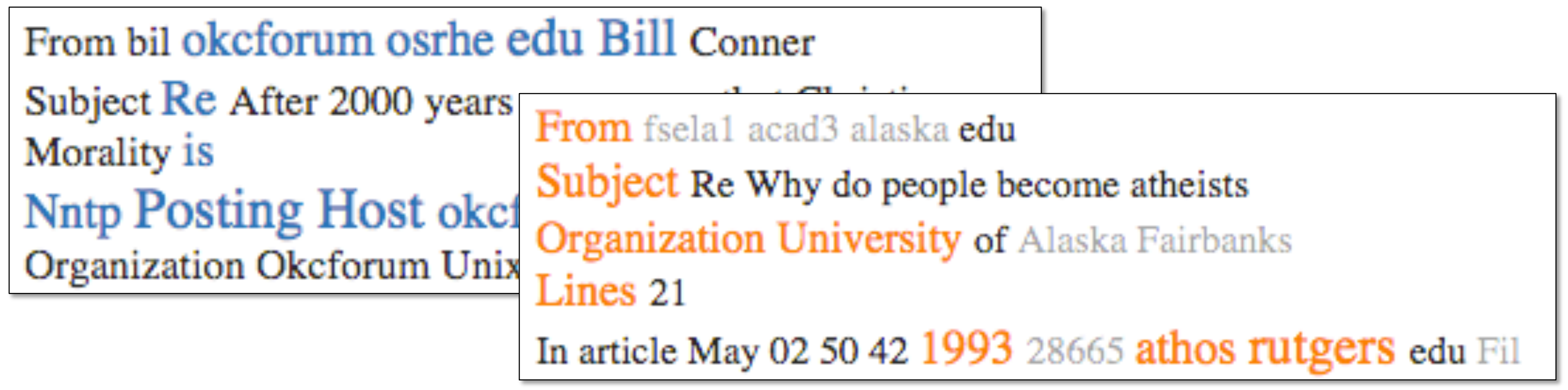
A Visualization Tool for Human-in-the-loop Machine Learning

CSE 512 Final Project, Marco Túlio Ribeiro and Brian Dolhansky

Problem

Practitioners don’t understand what their systems are learning. Could they be learning **irrelevant information**?



From bil okcforum osrhe edu Bill Conner
Subject Re After 2000 years
Morality is
Nntp Posting Host okc
Organization Okcforum Unix

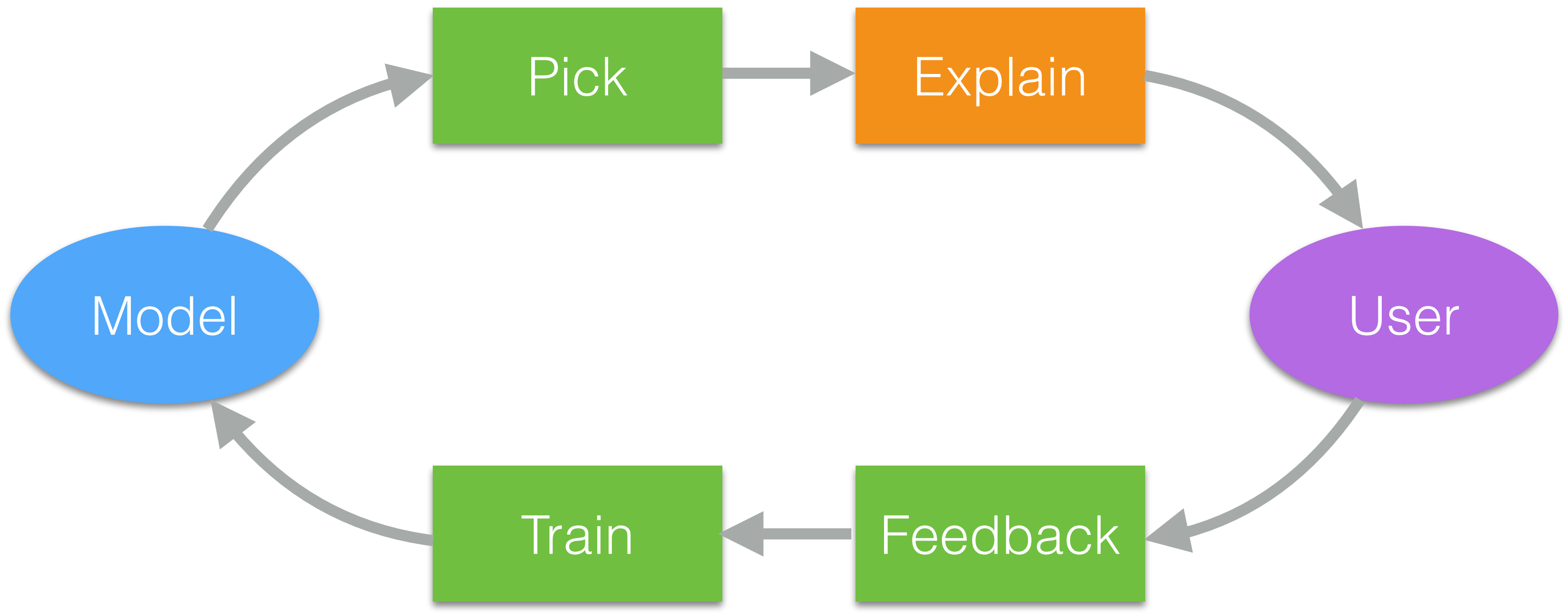
From fselal acad3 alaska edu
Subject Re Why do people become atheists
Organization University of Alaska Fairbanks
Lines 21
In article May 02 50 42 1993 28665 athos rutgers edu Fil

Accuracy with headers, quotes: ~**93%**, without: ~**84%**.
Can we help users spot and fix this?

Motivation and approach

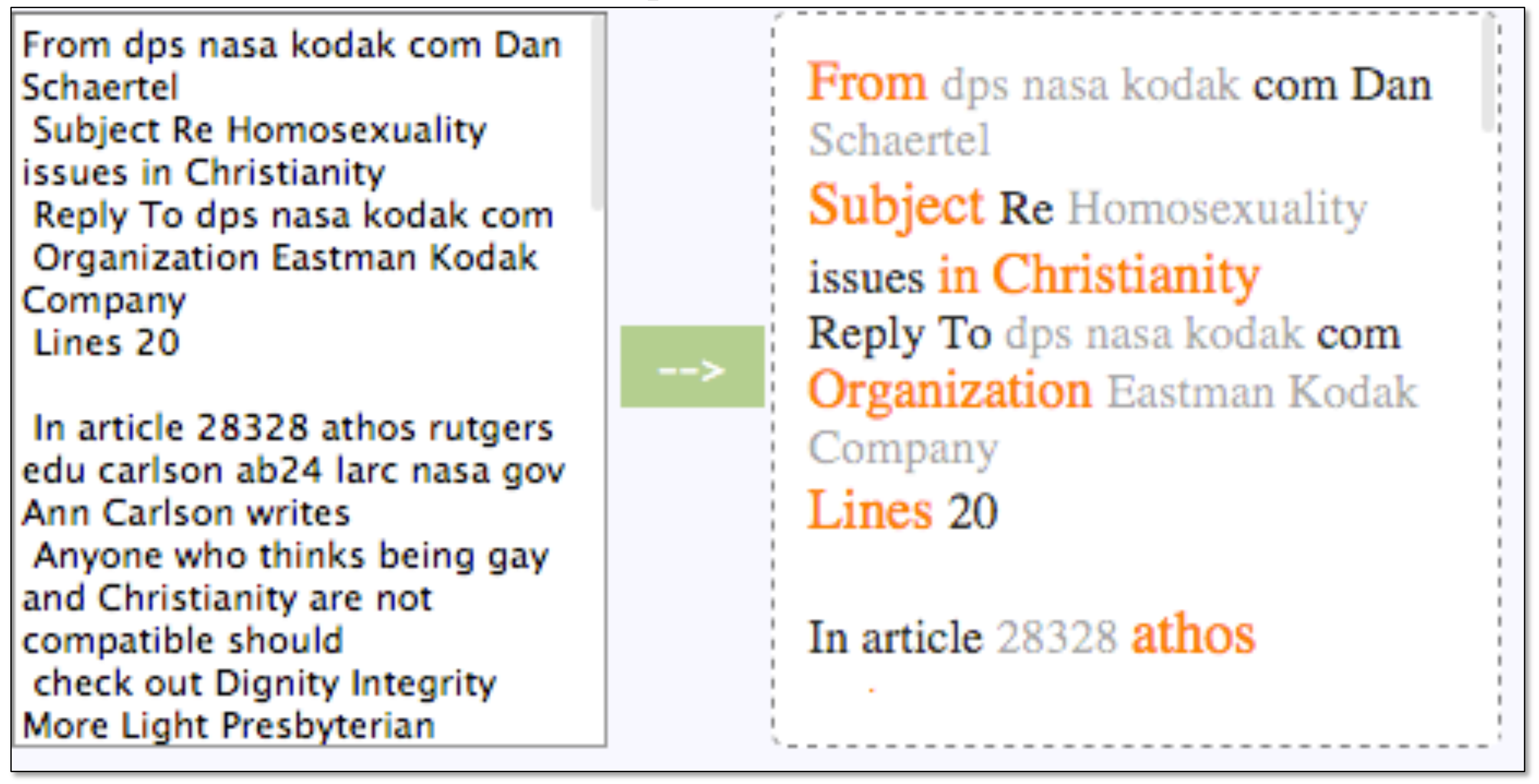
Our tool attempts to:

1. Visualize aspects of a dataset and model to enhance understanding of what a model is learning
2. Make it easy for human intervention that forces a model to learn the “right” information, leading to better generalization



Our system

Exploration

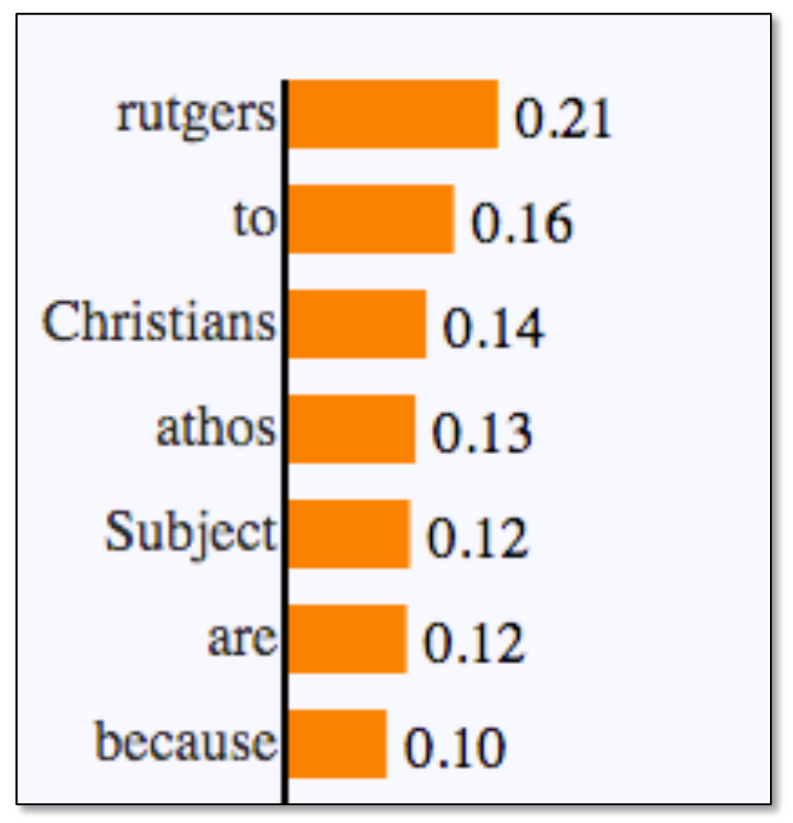


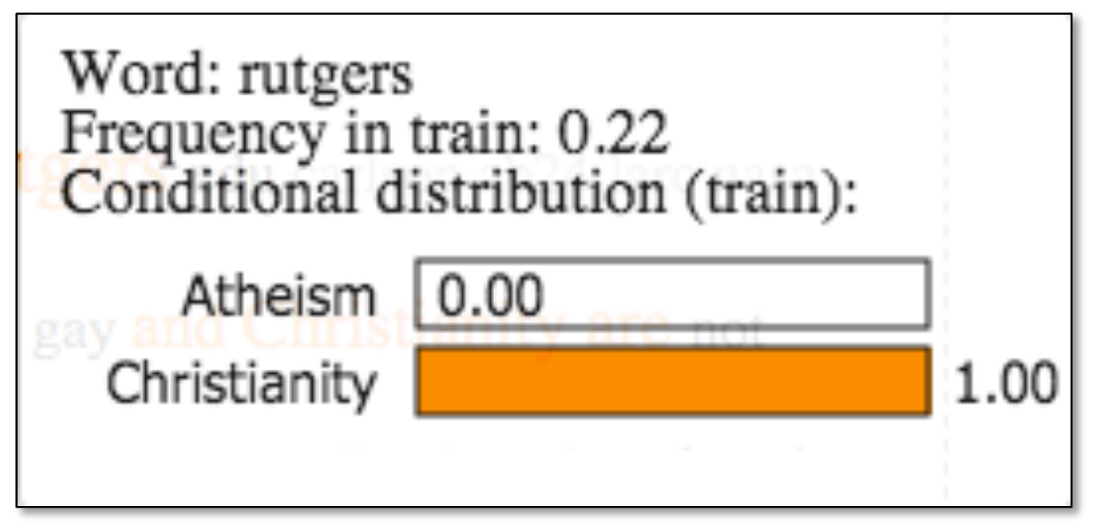
From dps nasa kodak com Dan Schaertel
Subject Re Homosexuality issues in Christianity
Reply To dps nasa kodak com
Organization Eastman Kodak Company
Lines 20

In article 28328 athos rutgers edu carlson ab24 larc nasa gov Ann Carlson writes
Anyone who thinks being gay and Christianity are not compatible should
check out Dignity Integrity More Light Presbyterian

From dps nasa kodak com Dan Schaertel
Subject Re Homosexuality issues in Christianity
Reply To dps nasa kodak com
Organization Eastman Kodak Company
Lines 20

In article 28328 athos

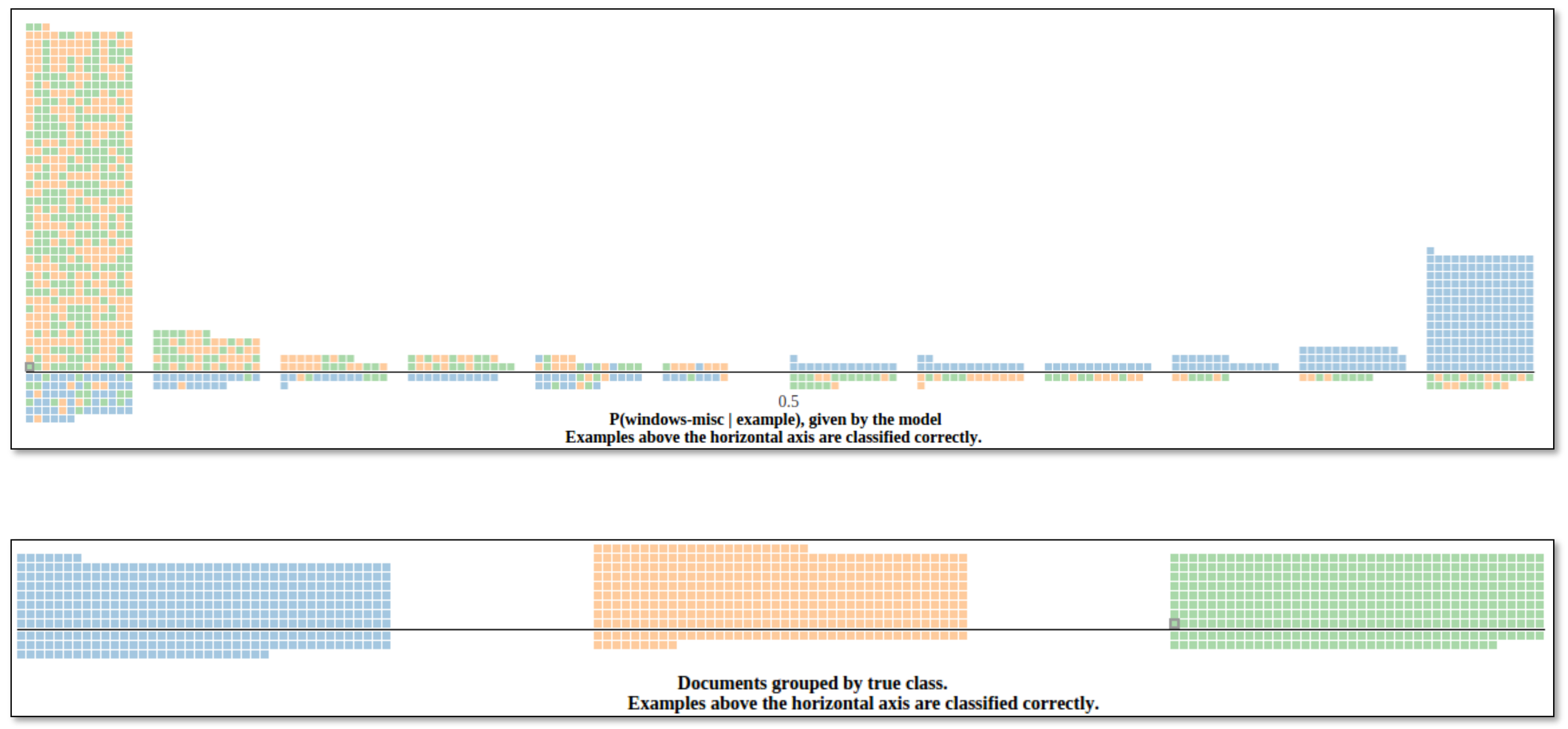




Word: rutgers
Frequency in train: 0.22
Conditional distribution (train):

Atheism	0.00
Christianity	1.00

Feature-level importance visualizations

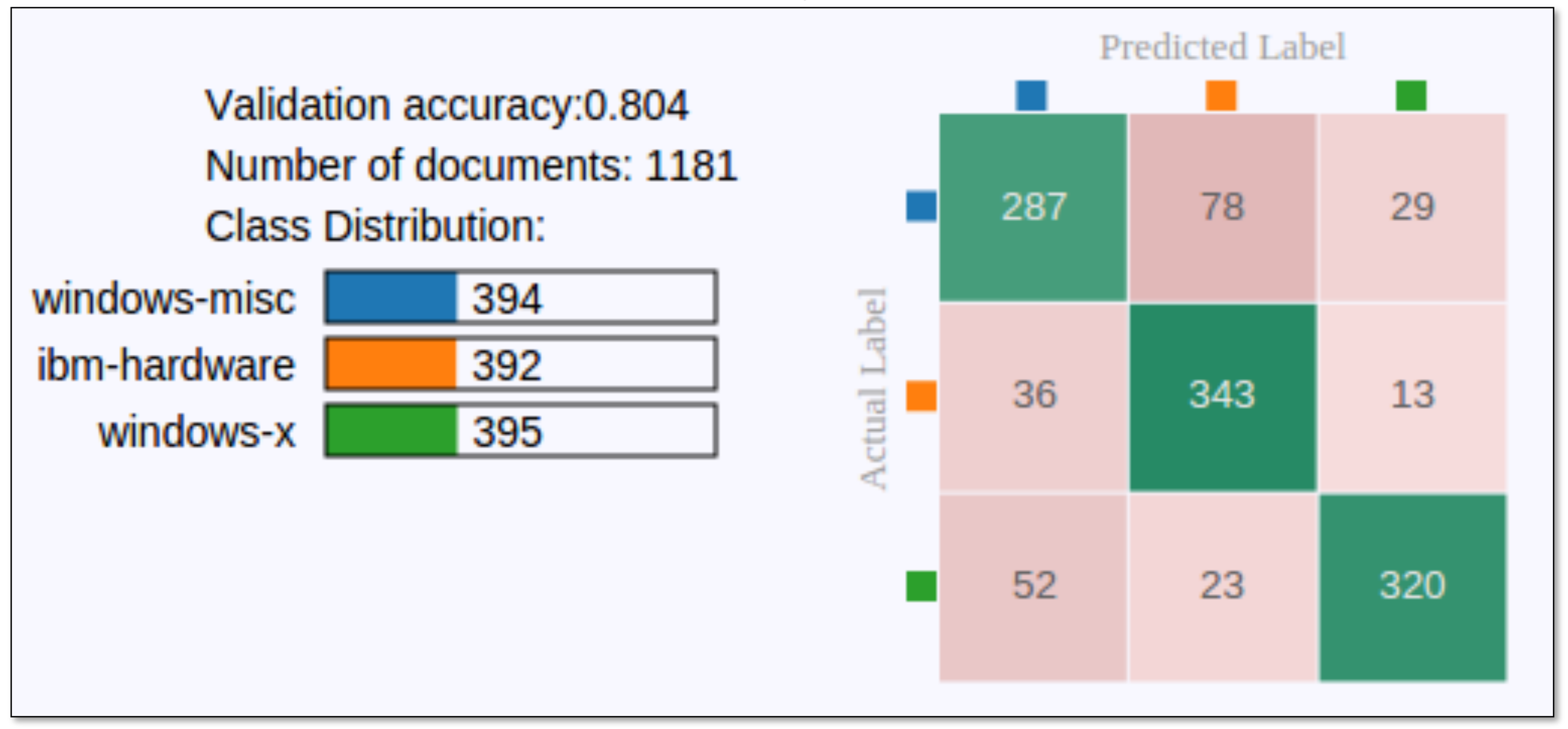


P(windows-misc | example), given by the model
Examples above the horizontal axis are classified correctly.

Documents grouped by true class.
Examples above the horizontal axis are classified correctly.

The databin (interactive)

Analysis



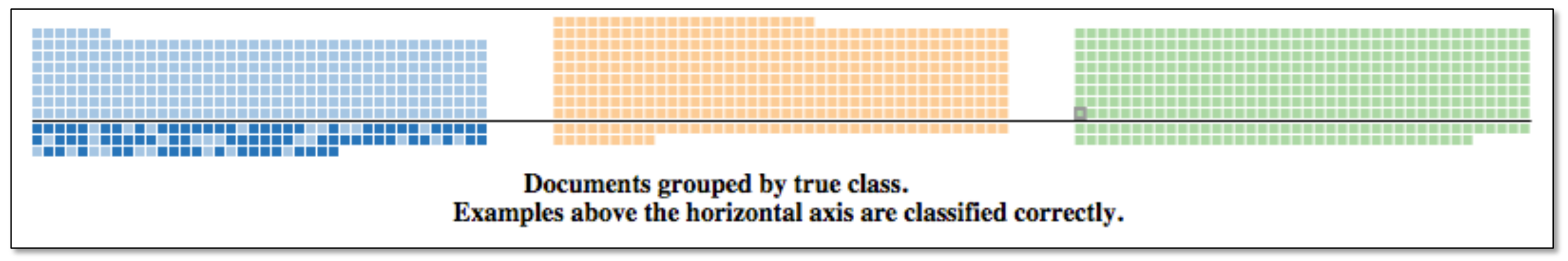
Validation accuracy:0.804
Number of documents: 1181
Class Distribution:

windows-misc	394
ibm-hardware	392
windows-x	395

Predicted Label

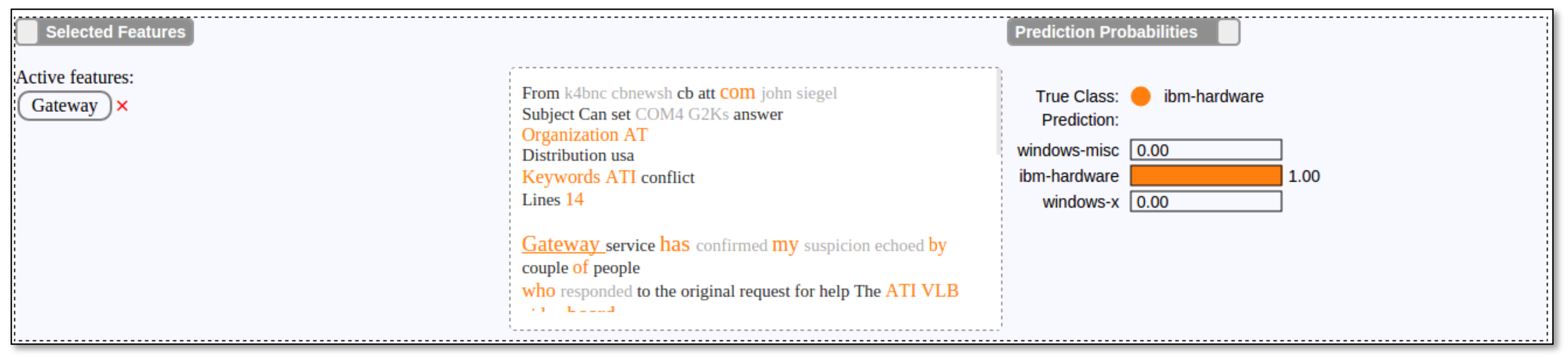
Actual Label	🟦	🟧	🟩
🟦	287	78	29
🟧	36	343	13
🟩	52	23	320

Overall statistics (interactive)



Documents grouped by true class.
Examples above the horizontal axis are classified correctly.

“Brushed” mistakes shown in databin



Active features:
Gateway ×

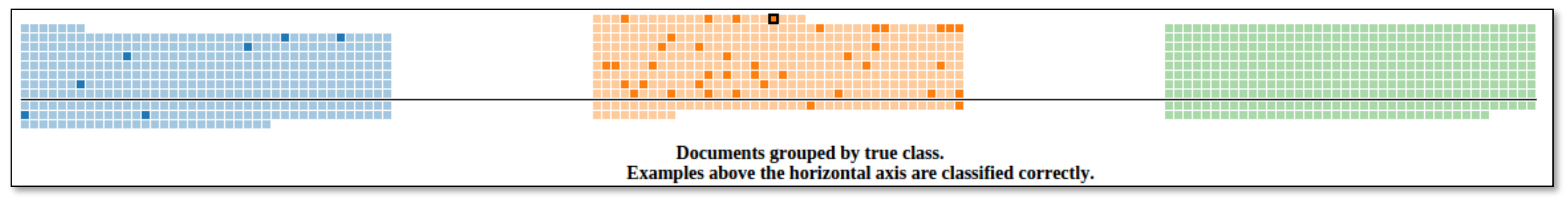
From k4bnc cbnewsh cb att com john siegel
Subject Can set COM4 G2Ks answer
Organization AT
Distribution usa
Keywords ATI conflict
Lines 14

Gateway service has confirmed my suspicion echoed by couple of people
who responded to the original request for help The ATI VLB

True Class: ● ibm-hardware
Prediction:

windows-misc	0.00
ibm-hardware	1.00
windows-x	0.00

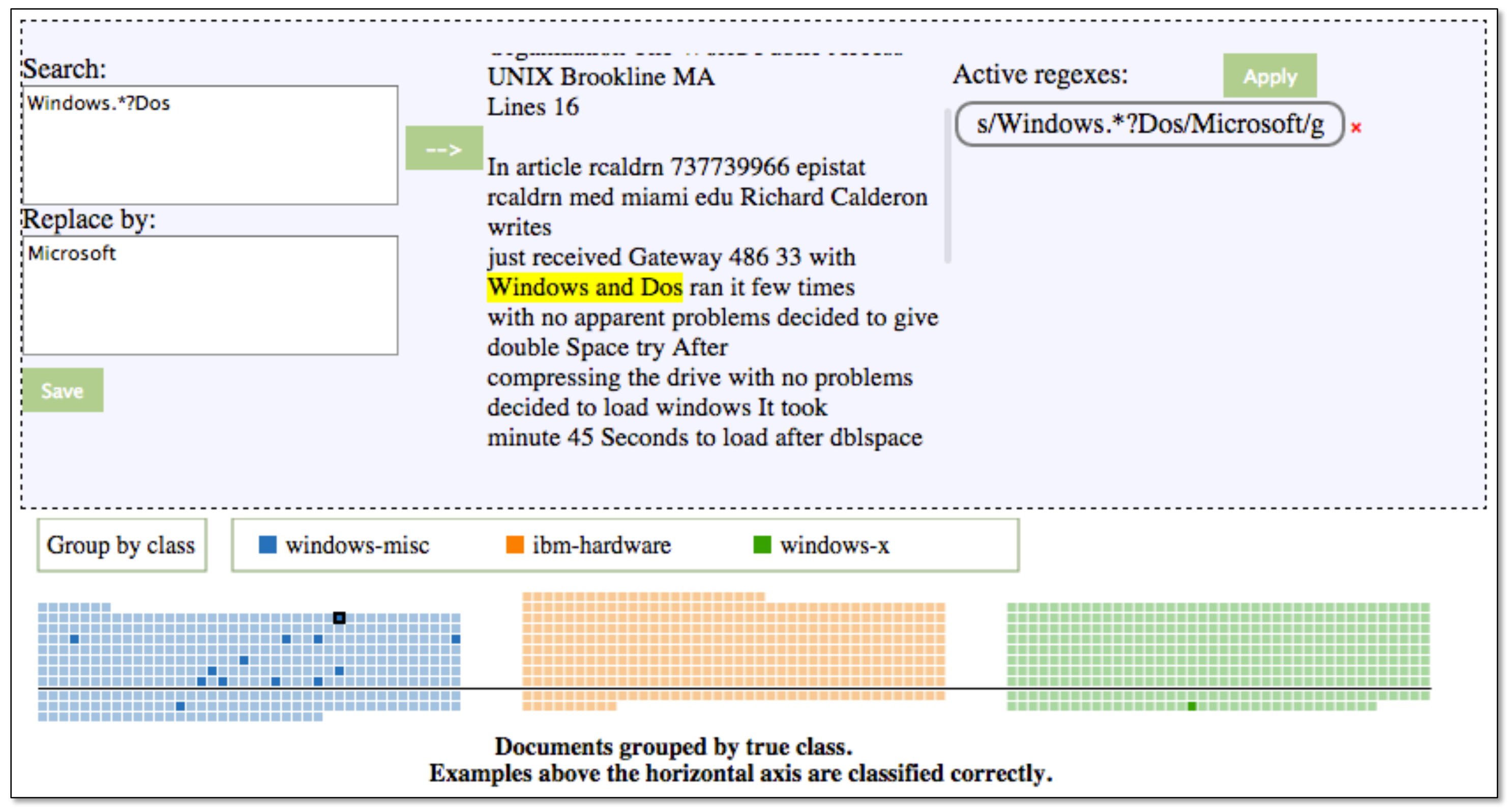
Examining specific features



Documents grouped by true class.
Examples above the horizontal axis are classified correctly.

“Brushed” features shown in databin

Feedback



Search:
Windows.*?Dos

Replace by:
Microsoft

UNIX Brookline MA
Lines 16

In article rcaldrn 737739966 epistat
rcaldrn med miami edu Richard Calderon
writes
just received Gateway 486 33 with
Windows and Dos ran it few times
with no apparent problems decided to give
double Space try After
compressing the drive with no problems
decided to load windows It took
minute 45 Seconds to load after dblspace

Active regexes:
s/Windows.*?Dos/Microsoft/g ×

Documents grouped by true class.
Examples above the horizontal axis are classified correctly.

Modifying a dataset and model (retraining) with search and replace

Future Work

1. Handle other types of data: tabular, images, custom datasets
2. Clustering databin points for bigger datasets
3. Novice and domain expert user studies
4. Allow for other kinds of feedback:
    - Feature engineering
    - Tuning parameters
    - Choosing models