Privacy Analytics
Toward Data-Driven Privacy

Pradeep K. Murukannaiah

Department of Software Engineering, GCCIS
Rochester Institute of Technology
Research Overview: Engineering Social Applications

- Software Engineering
  - Requirements engineering

- Artificial Intelligence
  - Agents Analytics (ML, NLP, SNA)

- Privacy
- Security
- Trust

- Social Computing
  - Human factors
  - Social norms
  - Arguments
  - Crowdsourcing

Murukannaiah (RIT)
This talk is NOT about
Privacy-preserving data analytics

This talk is about
Data analytics for preserving privacy

Objective/Societal  Privacy Expectations  Subjective/Individual
Collaborators

PrIncipedia

- Dr. Jessica Staddon, Google (formerly, NC State)
- Dr. Heather Lipford, UNC Charlotte
- Dr. Bart Knijnenburg, Clemson University
- Graduate Students: Chinmaya Dabral, Karthik Sheshadri, Esha Sharma

Privacy-Aware Agents

- Dr. Munindar P. Singh, NC State
- Dr. Jose Such, King’s College London
- Dr. Ricard Fogues, Universitat Politecnica de Valencia
- Graduate Students: Nirav Ajmeri, Hui Gao
Outline

PrIncipedia: Privacy Incidents Database
  - Definition and Scope
  - Automation

SoSharp: Privacy-Aware Personal Agents
  - Multiuser Sharing Scenarios
  - Data Collection
  - An Inference Model
Outline

PrIncipedia: Privacy Incidents Database
  Definition and Scope
  Automation

SoSharp: Privacy-Aware Personal Agents
  Multiuser Sharing Scenarios
  Data Collection
  An Inference Model
Motivation

Consider a few questions . . .

- When did awareness of cyberbullying increase in the US?
  - How many known cyberbullying incidents in 2016?
- How many public privacy incidents involved [insert large Internet company] last year?
- What types of privacy incidents are increasing in frequency?
- How many privacy incidents are caused by software bugs?
- What types of privacy incidents are successfully litigated?
Motivation

Consider a few questions...

- When did awareness of cyberbullying increase in the US?
  - How many known cyberbullying incidents in 2016?
- How many public privacy incidents involved [insert large Internet company] last year?
- What types of privacy incidents are increasing in frequency?
- How many privacy incidents are caused by software bugs?
- What types of privacy incidents are successfully litigated?

How would you go about answering these questions?
Vision: Privacy Incidents Database (PrIncipedia)

A collaboratively maintained repository of privacy incidents

- Crowdsourced, but moderated
- Includes who, what, when, where and why of incidents
- Support privacy analytics
- Visualize incident trends
Similar incidents

... despite *WhatsApp*’s lack of ads, its privacy policy allows it to periodically scan the mobile address book of its users and upload the numbers to its server, albeit without names attached to those numbers. [Forbes, 2/21/14]

... The new terms of service also explain that *Uber* could start accessing and storing your address book contacts to facilitate social interactions, such as sending special offers to your friends and family. [Refinery29, 6/5/15]
Different problem, same root cause

I don't want people seeing what I'm listening to
2013-08-07 08:22 PM
Hello,
I don't want Spotify posting on my Facebook what I'm listening to. The slider bar is telling me that sharing songs I listen to on Facebook is currently off. However I have had multiple friends who are not following me on Spotify tell me that they can see what I am listening to. How can I make Spotify stop posting to my Facebook?
PrIncipedia: Analytics

Same design flaw: Unexpected Defaults

By default, everyone can see who your friends are in the Friends section of your timeline. To change this setting:

1. Go to your timeline
2. Click on the Friends unit (under your cover photo)
3. Click the Edit button at the top of the page and use the audience selector to choose who can see your friend list on your timeline

Remember that mutual friends will still be visible to others.
PrIncipedia: Who should Care and Why?

Organizations

- Identify and respond to privacy incidents

Research community

- Analyze privacy incidents to inform technology improvements

Government

- Inform privacy law and policies
Privacy vs. Security Incidents Databases

Security Incidents

- Security alerts
- PII leaks
- Data breaches
- Malware
- Viruses

Privacy Analytics
Privacy vs. Security Incidents Databases

Security Incidents
- Security alerts
- Security updates
- Malware
- Viruses

Privacy Incidents
- PII leaks
- Data breaches
- Social media oversharing
- Reidentification
- Cyberbullying
- Surveillance
### Privacy Incidents Database

**Presidential candidate web sites found to often be poor on privacy (e.g. no privacy policy or inappropriate policies).**

- #2015 #World #PresidentialCandidateSites #PrivacyPolicy #Citizens

<table>
<thead>
<tr>
<th>Date</th>
<th>Keywords</th>
<th>Description</th>
<th>Source</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015-09</td>
<td>#whsmith</td>
<td>#citizen(s)</td>
<td><a href="http://www.businessinsider.com">www.businessinsider.com</a></td>
<td>09-2015- PersonalInfoLeak-WHSmith</td>
</tr>
<tr>
<td></td>
<td>#pi-leak</td>
<td>#uk</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015-09</td>
<td>#countrygovernment</td>
<td>#citizen(s)</td>
<td><a href="http://www.concordmonitor.com">www.concordmonitor.com</a></td>
<td></td>
</tr>
<tr>
<td></td>
<td>#RemovedPrivacyToolOrFeature</td>
<td>#usa</td>
<td></td>
<td>09-2015- PrivacyToolRemoved</td>
</tr>
<tr>
<td></td>
<td>#lawenforcement</td>
<td>#usa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015-09</td>
<td>#Google</td>
<td>#citizen(s)</td>
<td><a href="http://www.jdsupra.com">www.jdsupra.com</a></td>
<td></td>
</tr>
<tr>
<td></td>
<td>#countrygovernment</td>
<td>#RightToBeForgotten</td>
<td></td>
<td>09-2015- GoogleCriminalConviction</td>
</tr>
<tr>
<td></td>
<td>#citizen(s)</td>
<td>#uk</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>#press</td>
<td>#usa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015-09</td>
<td>#Facebook</td>
<td>#adolescent</td>
<td><a href="http://www.centralmaine.com">www.centralmaine.com</a></td>
<td></td>
</tr>
<tr>
<td></td>
<td>#citizen(s)</td>
<td># revengeporn</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>#photoaccess</td>
<td>#usa</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>#world</td>
<td>#attack</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015-09</td>
<td>#Google</td>
<td>#internetcompany</td>
<td><a href="http://www.theverge.com">www.theverge.com</a></td>
<td></td>
</tr>
<tr>
<td></td>
<td>#citizen(s)</td>
<td>#ads</td>
<td></td>
<td>09-2015- PrivacyPolicy-GoogleAds</td>
</tr>
<tr>
<td></td>
<td>#world</td>
<td>#BusinessDecision</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015-09</td>
<td>#PresidentialCandidates</td>
<td>#citizen(s)</td>
<td><a href="http://www.nbcnews.com">www.nbcnews.com</a></td>
<td></td>
</tr>
<tr>
<td></td>
<td>#privacypolicy</td>
<td>#usa</td>
<td></td>
<td>09-2015- PrivacyPolicy-PresCandidateSites</td>
</tr>
<tr>
<td></td>
<td>#accident</td>
<td>#usa</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
PrIncipedia: Challenges

Definition and Scope

- How to define a privacy incident so as to capture all incidents of interest?

Automation

- How can we automate the process of populating the incidents database?

Sustainability

- How can we engage researchers and practitioners to contribute to PrIncipedia, turning it into a self-sustaining endeavor?
Outline

PrIncipedia: Privacy Incidents Database
Definition and Scope
Automation

SoSharp: Privacy-Aware Personal Agents
Multiuser Sharing Scenarios
Data Collection
An Inference Model
Defining Privacy Incidents

Defining privacy incidents is nontrivial

- Attitudes vary by cultures and context
- Perceptions vary by stakeholder groups (e.g., end users, product developers, policy makers, and legal scholars)

We conduct a first study on end user understanding of privacy incidents involving 482 Amazon Mechanical Turk (AMT) users

Methodology

1. **Working definition**
Formulate a working definition of privacy incident

2. **Research questions (RQs)**
Formulate RQs corresponding to different aspects of the definition

3. **User study**
Provide examples and counterexamples to end users

4. **Definition validation**
Measure precision of the definition based on users’ perceptions
A privacy incident is

1. An instance of accidental or unauthorized collection, use or exposure of sensitive information, OR,

2. An event that creates the perception that unauthorized collection, use or exposure of sensitive information may happen, AND,

3. Involves information that is either being collected, used or shared in digital form.
Research Questions

RQ1  Do end users perceive the collection, usage and/or sharing of sensitive data to be closely associated with privacy incidents?

RQ2  Do end users recognize events in which it is perceived or anticipated that sensitive data is or will be collected, used or shared, as privacy incidents?

RQ3  Are there privacy events beyond the scope of the working definition (e.g., the release of privacy-enhancing products or privacy laws) that end users recognized as incidents?
User Study: Questionnaire (Partial)

Read the news article in the URL below and answer the following questions.
${articleUrl}$

1. Is this article primarily about a privacy incident?
   - Yes
   - No
   - Not sure

2. Is this article primarily concerned with the collection, use or exposure of sensitive information?
   - Yes
   - No
   - Not sure

3. Is this article primarily concerned with information that is either being collected, used or shared in digital form?
   - Yes
   - No
   - Not sure

4. Consider now that for an article to be primarily about privacy incident, the answers to the previous two questions must be yes. Given this, do you think the article is primarily about a privacy incident?
   - Yes
   - No
   - Not sure
User Study: Positive Examples (Set $P$)

**Data Breach**  Blippy allows some credit card numbers to be indexed by search engines

**Emerging Technology**  Police departments building DNA databases of potential suspects

**Surveillance**  PA SD remotely activates cameras to locate school laptops

**Targeting**  Facebook uses visits to sites with like button to target ads

**Revenge Porn**  Woman posts revenge porn pictures; convicted under UK law
User Study: Negative Examples (Set N)

Security Breach  Security of an electronic road sign is compromised
Physical Security Breach  Security breach at a Donald Trump rally
Celebrity Privacy Request  Professional athlete asks for privacy as he enters drug rehab
Wildlife Privacy  Park visitors disturb privacy of animals during mating season
Privacy Standards  EFF announces a stronger DNT standard
Security/Privacy Law  Companies warn about privacy implications of cybersecurity bill
Participants and Measurements

Study Units

- $|P| = 204$ and $|N| = 63$

Participants

- 482 unique Amazon Mechanical Turk users from the US
- At least three unique responses per HIT
- Maximum of three HITs per participant
- Payment: USD 0.7 per HIT

Measurements

- Measure Precision($P$) and Precision($N$)
- Measure initial and final Precision
Findings: Collection/Usage/Sharing (RQ1)

- Initial Precision($P$) = 0.799, and Final Precision($P$) = 0.794
- Initial Precision($N$) = 0.549, and Final Precision($N$) = 0.600

Participants view sensitive data collection, usage and/or sharing as core to their understanding of the term “privacy incident”

Participants more likely to see the collection, usage and sharing aspects as sufficient for a privacy incident, than as necessary
Findings: Anticipated/Perceived Privacy Harm (RQ2)

13 articles about emerging technologies (e.g., DNA testing) in $P$
  - Initial Precision($P$) = 0.769, and Final Precision($P$) = 0.769

5 articles about privacy policy changes in $P$
  - Initial Precision($P$) = 0.714, and Final Precision($P$) = 0.643

Participants are more inclined than privacy law to view perceived or anticipated privacy issues as grounds for an incident
Findings: Definition Scope (RQ3)

- Only 3 of the 13 articles about legal privacy events were viewed as non-incidents.
- 2 of the 3 articles about digital security incidents that do not involve a privacy breach were reported to be privacy incidents.

Participants do not consistently view privacy law events and security breaches that do not involve privacy breaches as non-incidents.

- None of the 4 such articles about cyberbullying, involving sexual assault, were identified as privacy incidents.

Cyberbullying events are not consistently viewed as privacy incidents.
14 articles from $P$ that contain the term “privacy” were reported to not be privacy incidents by a majority of participants.

4 of the 25 articles in $N$ that were reported to be privacy incidents, do not contain the term “privacy”.

Participants do not seem to consider the presence of the term “privacy” as a necessary or sufficient condition to treat an event as privacy incident.
Summary

- High and stable precision for $P$

Our working definition is compatible with end users’ perceptions about privacy incidents

- Increase in precision for $N$ after definition
- Substantial disagreement remains in $N$

Determine if disagreements suggest a way to improve the definition
Outline

PrIncipedia: Privacy Incidents Database
  Definition and Scope
  Automation

SoSharp: Privacy-Aware Personal Agents
  Multiuser Sharing Scenarios
  Data Collection
  An Inference Model
Bootstrapping the PrIncipedia

- Information about privacy incidents is available from a variety of sources
  - news articles, blog posts and social media
- Privacy incidents may have differentiating features
  - Sentiment, entities, keywords, authors
- Many document will mention privacy, but the term “privacy” itself is not indicative of an incident

Can we train an automated classifier for efficient, large-scale identification of privacy incidents?

A Semi-Automated Framework

The classifier tags information as potentially related to privacy incident

Training Data

- Source: News articles from New York Times and Guardian
  - Selected via crafted keywords as well as randomly
  - Manually labelled as privacy ($P$) or not privacy ($N$)
- $|P| = 543$ and $|N| = 781$
Natural Language Processing Pipeline

Text Preprocessing
- Sentence split
- PoS tagging
- Lemmatization
- Content word retention

Feature Engineering
- Tokenization
  - Unigrams
  - Uni & bigrams
- Scoring
  - TF
  - TF-IDF
- Selection
  - Mutual information
  - Top-K Features

Classification
- Naive Bayes
- SVM
- Random Forests

Unique words (before preprocessing) 59,864
Unigrams (after preprocessing) 34,537
Bigrams (after preprocessing) 449,988
Overfitting: Feature Selection based on Information Gain

F1-measure (Privacy class)

Number of features

- Unigrams (TF)
- Unigrams (TF-IDF)
- Uni and Bigrams (TF)
- Uni and Bigrams (TF-IDF)
## Accuracy Comparison

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Privacy Class</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>0.863</td>
<td>0.924</td>
</tr>
<tr>
<td>SVM</td>
<td>0.962</td>
<td>0.897</td>
</tr>
<tr>
<td>Random Forests</td>
<td>0.931</td>
<td>0.820</td>
</tr>
<tr>
<td>Privacy keyword</td>
<td>0.918</td>
<td>0.700</td>
</tr>
<tr>
<td>Privacy &amp; Solove keywords</td>
<td>0.664</td>
<td>0.930</td>
</tr>
<tr>
<td>Top-3 keywords</td>
<td>0.721</td>
<td>0.932</td>
</tr>
</tbody>
</table>
PrIncipedia: Summary

Definition and Scope

- How to define a privacy incident so as to capture all incidents of interest?

Automation

- How can we automate the process of populating the incidents database?

Sustainability

- How can we engage researchers and practitioners to contribute to PrIncipedia, turning it into a self-sustaining endeavor?
Outline

PrIncipedia: Privacy Incidents Database
  Definition and Scope
  Automation

SoSharp: Privacy-Aware Personal Agents
  Multiuser Sharing Scenarios
  Data Collection
  An Inference Model
Subjective Privacy Requirements

- Vary from one individual to another, one context to another

How can we assist an individual in making data-driven privacy decisions?
Privacy-Aware Personal Agents

A personal agent acts and interacts on behalf of its human user.

Outline

PrIncipedia: Privacy Incidents Database
Definition and Scope
Automation

SoSharp: Privacy-Aware Personal Agents
Multiuser Sharing Scenarios
Data Collection
An Inference Model
Multiuser Sharing Scenarios

19. If @Person tweets and uses @Other somewhere in the tweet ("I watched the football and saw @Other score!")}, then even if @Other doesn’t follow @Person, @Other will also receive a Notification of the tweet by @Person (unless they have first Blocked or Muted tweets from @Person). However, subject to the point below, at-mentions are sent to all of @Person’s followers.

20. If, however, @Person begins their tweet with a username (@Stranger), this particular type of at-mention is called an ‘at-reply’ (because it is in the same format as a Reply to a tweet, in that it starts with a username).

Picture Sharing Example

- A piece of information concerns Alice, Bob, and Charlie
  - Charlie finds the information as sensitive
  - Alice shares the information via her social media account

Alice’s action may violate Charlie’s privacy
Reasoning about Privacy in Multiuse Scenarios

Factors influencing sharing decisions.

Context

- Users share information in distinctive social contexts
- Relationships, sensitivity, and sentiment

Preferences

- Users may have preferences on how they want to share information
- Share with all, common friends, or among themselves

Arguments

- Users may have justifications for their preferences
Arguments

Argument help make sharing decisions in multiuser scenarios

Alice I want to share this picture with all because *we had a great time*

Charlie I want to keep this picture among ourselves because *I do not want my boss to think that I left work early*

How can we find a representative set of arguments to study their influence in multiuser scenarios?
Argumentation Schemes

Arguments used in everyday conversation fall into a small number of argumentation schemes [Walton et al.]

Argument from positive consequences

If $A$ is brought about, then good consequences will happen

$A$ should be brought about

Argument from exceptional case

Doing $a$ is the established rule for $A$

The case of $a$ is an exception

$a$ need not do $A$
Outline

PrIncipedia: Privacy Incidents Database
  Definition and Scope
  Automation

SoSharp: Privacy-Aware Personal Agents
  Multiuser Sharing Scenarios
  Data Collection
  An Inference Model
Crowdsourced Data Collection

Objective: Understanding how context, preferences, and arguments help sharing decisions in multiuser scenarios

- Directly ask users about the factor influences, e.g.,
  - To what extent does sensitivity influence a sharing decision?
  - If one user argues positive consequence and another argues negative consequence, whose preferred policy should be applied?

- Identify cliques of users and observe how they make decisions
  - Logistically challenging
  - May be privacy invasive
  - May yield observations biased toward nonsensitive items

Simulate scenarios: Depict specific context, preferences, and arguments
Scenarios: Parameter Space

Context
- Relationships: \{family, friends, colleagues\}
- Sensitivity: \{high, low\}
- Sentiment: \{positive, negative\}

Preferences
- \{Share with all, Share with common friends, Share among themselves\}

Arguments
- \{Positive consequence, Negative consequence, Exceptional case\}
Example Scenario: Context

Aiko (C) took the picture above with her colleagues Ichiro and Katsu and, a French volunteer at the tsunami relief center.

Rating

Identify the relationship between Aiko, Ichiro, and Katsu and rate the sensitivity and sentiment of the picture.

Case 1: Context

Consider that Aiko wants to upload this picture to her social media account. What sharing policy should she apply for the picture?
Example Scenario: Preferences

Aiko (C) took the picture above with her colleagues Ichiro and Katsu and, a French volunteer at the tsunami relief center.

Case 2: Preferences

- **Aiko.** Share among ourselves
- **Ichiro.** Share among ourselves
- **Katsu.** Share with all

Considering the context and users’ preferences, what sharing policy should Aiko apply for the picture?
Example Scenario: Arguments

Aiko (C) took the picture above with her colleagues Ichiro and Katsu and, a French volunteer at the tsunami relief center.

Case 3: Arguments

- **Aiko.** This was one of the worst disasters ever. Share among ourselves.
- **Ichiro.** Tsunami was a disaster and our hand gestures are not appropriate; people may get the wrong idea. Share among ourselves.
- **Katsu.** The picture shows the difficult situation of survivors; sharing this picture can encourage people to help. Share with all.

Considering the context, users’ preferences, and arguments, what sharing policy should Aiko apply for the picture?
Number of Scenarios

- Three users in a scenario
- Restriction: Not all three use the same policy-argument combination

12 pictures, 6 policy-argument combinations for each of first two users, and 5 policy-argument combinations for the third user = 2,160 scenarios

- Aim to get at least two responses per scenario

Conduct the study on Amazon Mechanical Turk (MTurk)

- 988 unique workers, each answered 5 scenarios
- Attention check: Count the number of faces in each picture
- Reward: USD 2 for each worker
Outline

PrIncipedia: Privacy Incidents Database
  Definition and Scope
  Automation

SoSharp: Privacy-Aware Personal Agents
  Multiuser Sharing Scenarios
  Data Collection
  An Inference Model
A Classifier for Predicting Appropriate Sharing Policy

- The crowdsourced data can be used to bootstrap the classifier to make predictions off-the-shelf
- The classifier can be retrained once user-specific data is available
### Results: Influence of Contextual Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficients</th>
<th>Context (Case 1)</th>
<th>Preferences (Case 2)</th>
<th>Arguments (Case 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>All</td>
<td>Common</td>
<td>All</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>−5.485** −4.174**</td>
<td>−2.045** −1.508**</td>
<td>−1.707** −1.38**</td>
<td></td>
</tr>
<tr>
<td>Sentiment</td>
<td>−0.415 −0.424</td>
<td>−0.352 0.057</td>
<td>−0.193 0.048</td>
<td></td>
</tr>
<tr>
<td>R = colleagues</td>
<td>−0.525** −1.345**</td>
<td>−0.355 −0.457**</td>
<td>−0.245 −0.543**</td>
<td></td>
</tr>
<tr>
<td>R = friends</td>
<td>1.162** −0.464**</td>
<td>0.507** −0.165</td>
<td>0.659** −0.154</td>
<td></td>
</tr>
</tbody>
</table>

- Among contextual variables, sensitivity has the highest influence on the selected sharing policy.
- The influences of contextual variables reduce when preferences and arguments are added.
Results: Users’ Perceptions Contextual Variables

- Sensitivity
- Sentiment
- Relationship
## Results: Influence of Preferences

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficients</th>
<th>Preferences</th>
<th>Arguments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>All</td>
<td>All</td>
</tr>
<tr>
<td>Least restrictive policy = All</td>
<td>1.021**</td>
<td>0.013</td>
<td>1.197**</td>
</tr>
<tr>
<td><strong>Most restrictive policy = Self</strong></td>
<td>−1.421**</td>
<td>−2.461**</td>
<td>−1.31**</td>
</tr>
<tr>
<td>Majority policy = All</td>
<td>0.674**</td>
<td>−0.452*</td>
<td>0.766**</td>
</tr>
<tr>
<td>Majority policy = Common</td>
<td>−0.174</td>
<td>−0.195</td>
<td>−0.054</td>
</tr>
<tr>
<td>Majority policy = Self</td>
<td>−0.584*</td>
<td>−1.07**</td>
<td>−0.536</td>
</tr>
</tbody>
</table>

The most restrictive policy, not the majority policy, has the highest influence on the selected sharing policy.
### Results: Influence of Arguments

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficients</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Common</td>
<td></td>
</tr>
<tr>
<td>Positive supporting least restrictive policy</td>
<td>-1.612**</td>
<td>-1.257**</td>
<td></td>
</tr>
<tr>
<td>Negative supporting least restrictive policy</td>
<td>-1.728**</td>
<td>-0.168**</td>
<td></td>
</tr>
<tr>
<td>Exceptional supporting least restrictive policy</td>
<td>-1.334**</td>
<td>-1.043**</td>
<td></td>
</tr>
<tr>
<td>Positive supporting most restrictive policy</td>
<td>1.806**</td>
<td>1.354**</td>
<td></td>
</tr>
<tr>
<td>Negative supporting most restrictive policy</td>
<td>1.082*</td>
<td>1.269**</td>
<td></td>
</tr>
<tr>
<td>Exceptional supporting most restrictive policy</td>
<td>1.532**</td>
<td>1.446**</td>
<td></td>
</tr>
<tr>
<td><strong>Positive supporting majority policy</strong></td>
<td>2.641**</td>
<td>1.728**</td>
<td></td>
</tr>
<tr>
<td>Negative supporting majority policy</td>
<td>-0.111</td>
<td>0.671**</td>
<td></td>
</tr>
<tr>
<td><strong>Exceptional supporting majority policy</strong></td>
<td>-2.385**</td>
<td>-2.53**</td>
<td></td>
</tr>
</tbody>
</table>

The majority policy, supported with positive consequence or exceptional case arguments, has the highest influence on the selected sharing policy.
## Results: Cross-Validating the Classifier

<table>
<thead>
<tr>
<th>Context</th>
<th>Precision</th>
<th>Recall</th>
<th>$F_1$</th>
<th>Class</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.539</td>
<td>0.379</td>
<td>0.445</td>
<td>all</td>
<td>26.5%</td>
</tr>
<tr>
<td></td>
<td>0.493</td>
<td>0.472</td>
<td>0.482</td>
<td>common</td>
<td>36.0%</td>
</tr>
<tr>
<td></td>
<td>0.588</td>
<td>0.736</td>
<td>0.654</td>
<td>self</td>
<td>37.5%</td>
</tr>
<tr>
<td></td>
<td><strong>0.541</strong></td>
<td><strong>0.546</strong></td>
<td><strong>0.537</strong></td>
<td>weighted mean</td>
<td>—</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Preferences</th>
<th>Precision</th>
<th>Recall</th>
<th>$F_1$</th>
<th>Class</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.329</td>
<td>0.047</td>
<td>0.082</td>
<td>all</td>
<td>15.4%</td>
</tr>
<tr>
<td></td>
<td>0.606</td>
<td>0.622</td>
<td>0.614</td>
<td>common</td>
<td>40.8%</td>
</tr>
<tr>
<td></td>
<td>0.610</td>
<td>0.779</td>
<td>0.684</td>
<td>self</td>
<td>43.8%</td>
</tr>
<tr>
<td></td>
<td><strong>0.565</strong></td>
<td><strong>0.602</strong></td>
<td><strong>0.563</strong></td>
<td>weighted mean</td>
<td>—</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Arguments</th>
<th>Precision</th>
<th>Recall</th>
<th>$F_1$</th>
<th>Class</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.347</td>
<td>0.122</td>
<td>0.180</td>
<td>all</td>
<td>16.1%</td>
</tr>
<tr>
<td></td>
<td>0.610</td>
<td>0.640</td>
<td>0.624</td>
<td>common</td>
<td>39.2%</td>
</tr>
<tr>
<td></td>
<td>0.646</td>
<td>0.769</td>
<td>0.702</td>
<td>self</td>
<td>44.7%</td>
</tr>
<tr>
<td></td>
<td><strong>0.584</strong></td>
<td><strong>0.614</strong></td>
<td><strong>0.588</strong></td>
<td>weighted mean</td>
<td>—</td>
</tr>
</tbody>
</table>
Results: Users’ Confidence

\[ \hat{x}(\text{Arguments}) > \hat{x}(\text{Preferences}) > \hat{x}(\text{Context}) \]

- Arguments
- Preferences
- Context
Privacy Analytics: Summary

PrIncipedia

SoSharp