Thesis proposals

Ideas for research from social bots to beer

Florian Daniel
florian.daniel@polimi.it

February 25, 2019

Detection of harmful social bots

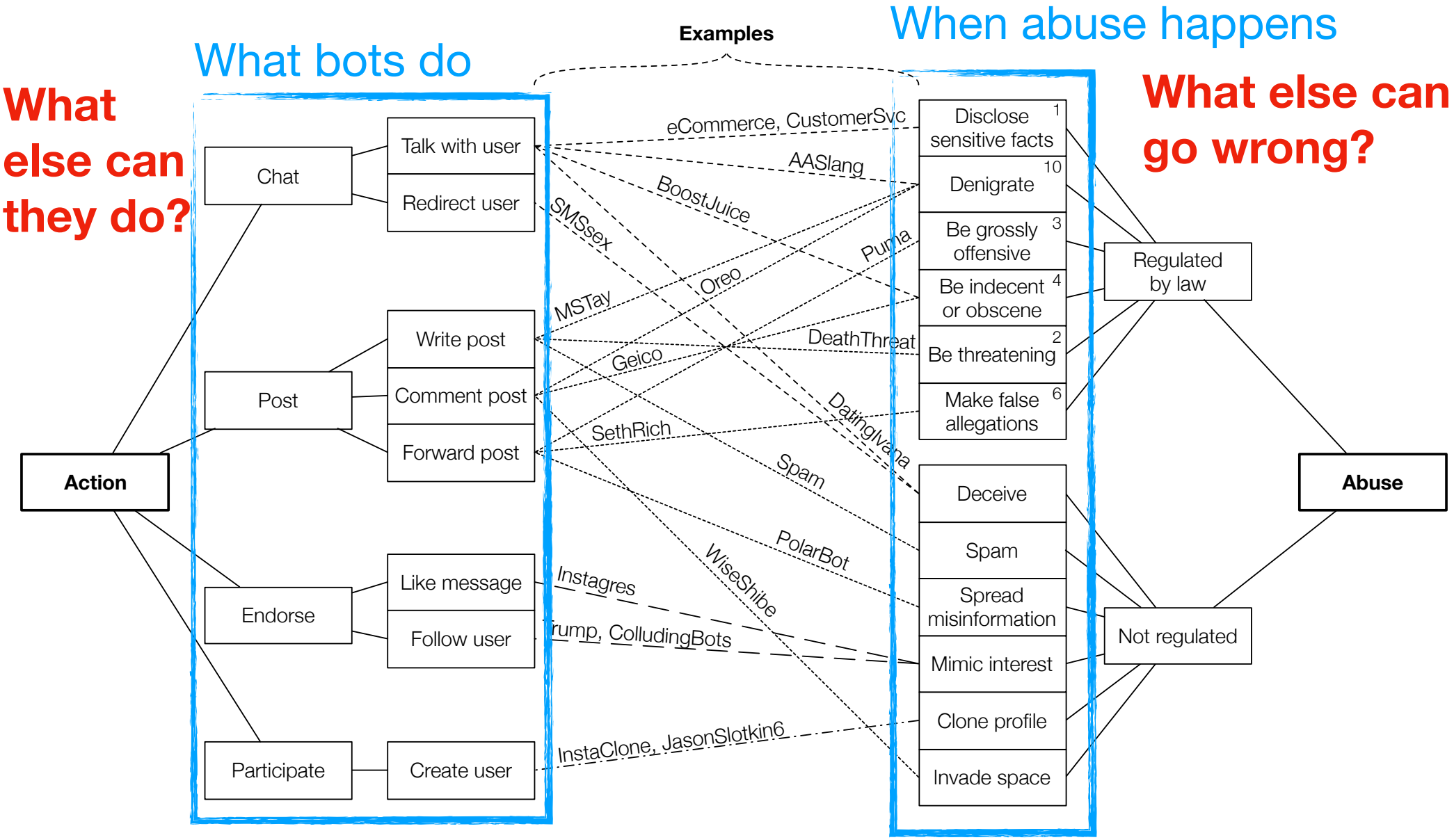
Identification of harmful communication patterns

Conversational screen readers

Domain-specific content extraction

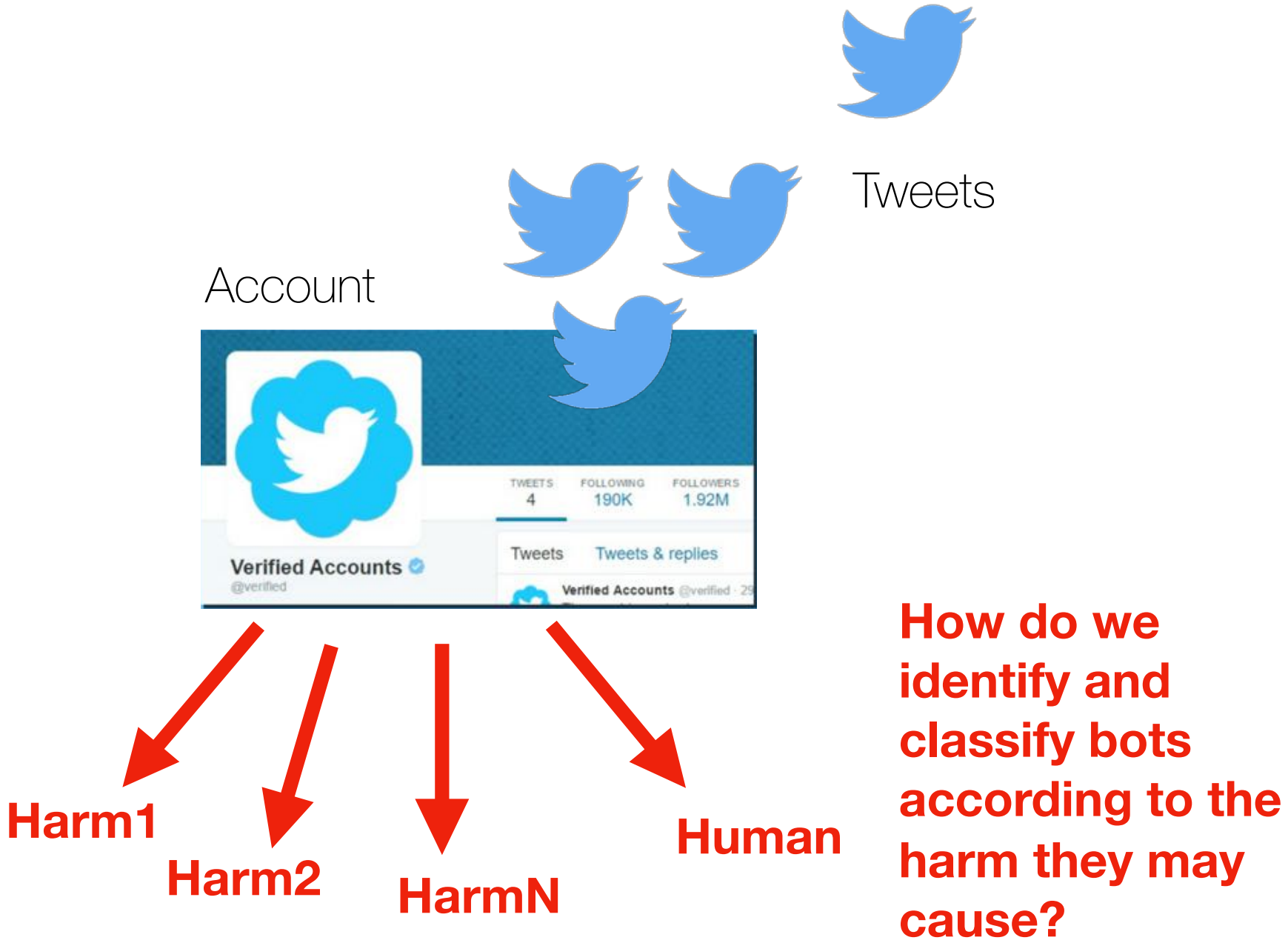
(Social) Bot = algorithmically driven entity that behaves like a human in online communications

Empirical study shows: bots may cause **harm** to humans



F. Daniel, C. Cappiello, B. Benatallah. Bots Acting Like Humans: Understanding and Preventing Harm. *IEEE Internet Computing*, 2019, accepted for publication. https://ieeexplore.ieee.org/document/8611348

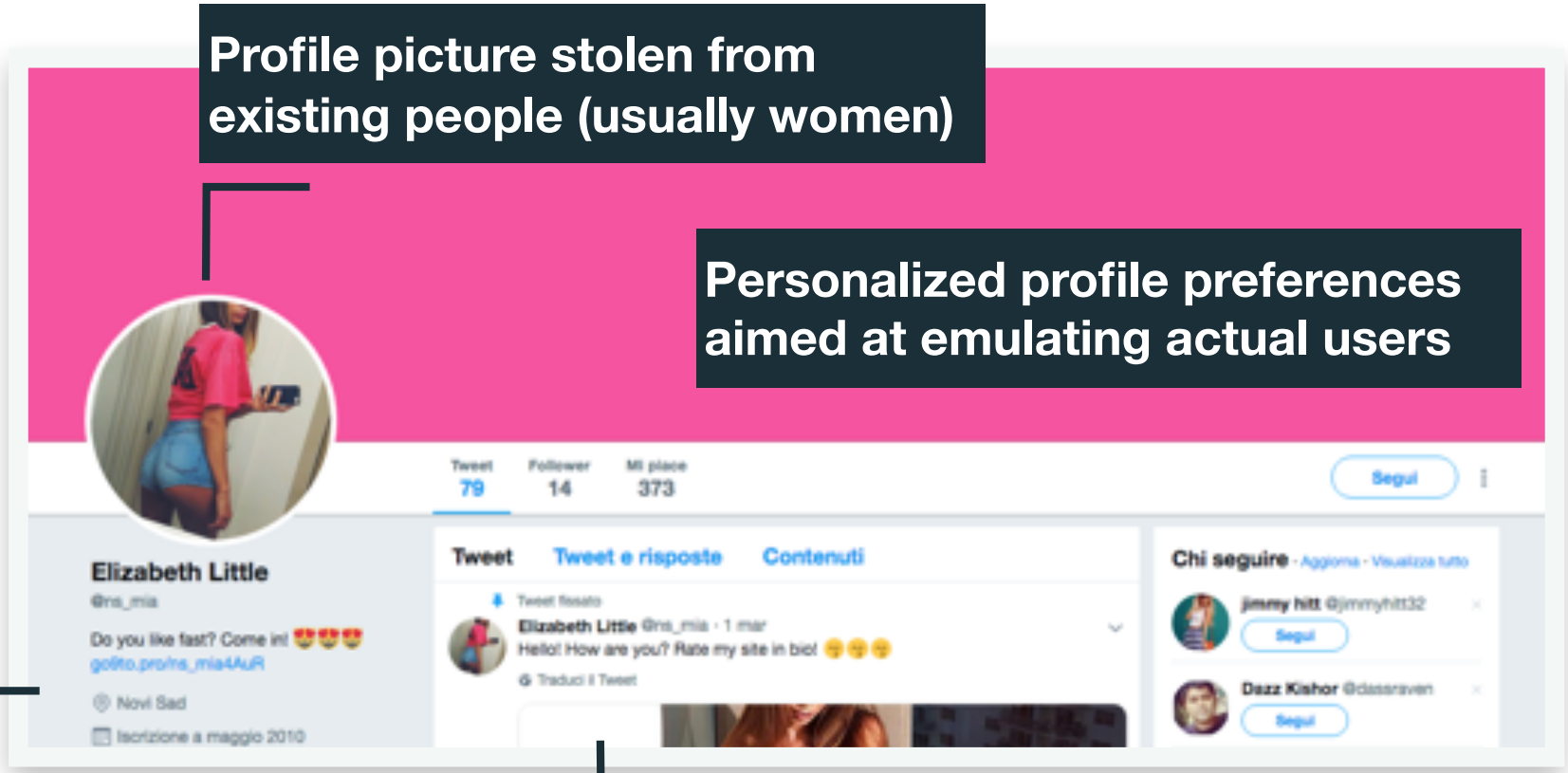
Detection of harmful social bots



How do we identify and classify bots according to the harm they may cause?

What has been done so far?

Not Safe For Work



Lorenzo Cannone, Matteo Di Pierro. *Detection and Classification of Harmful Bots in Human-Bot Interactions on Twitter*. MSc Thesis, Politecnico di Milano, December 2018.

News-Spreader

Propagandistic profile picture and tile

Propaganda hashtags and messages in description

Lots of (political) news retweeting

Lorenzo Cannone, Matteo Di Pierro. *Detection and Classification of Harmful Bots in Human-Bot Interactions on Twitter*. MSc Thesis, Politecnico di Milano, December 2018.

Spam-Bot

Profile tile with catchy messages

Profile picture with corporate logo

Corporate link in description paired with messages of job offers (product sales)

High number of tweets with high frequency

Repetitive tweets aimed at spamming URLs

Lorenzo Cannone, Matteo Di Pierro. *Detection and Classification of Harmful Bots in Human-Bot Interactions on Twitter*. MSc Thesis, Politecnico di Milano, December 2018.

Fake-Follower

Optional profile picture (usually missing)

Poor profile customization

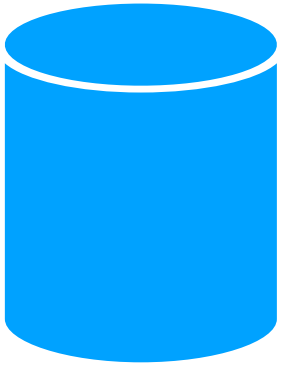
Poor description (usually empty)

Optional (re)tweeting activity

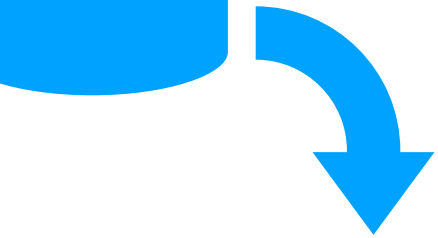
Lorenzo Cannone, Matteo Di Pierro. *Detection and Classification of Harmful Bots in Human-Bot Interactions on Twitter*. MSc Thesis, Politecnico di Milano, December 2018.

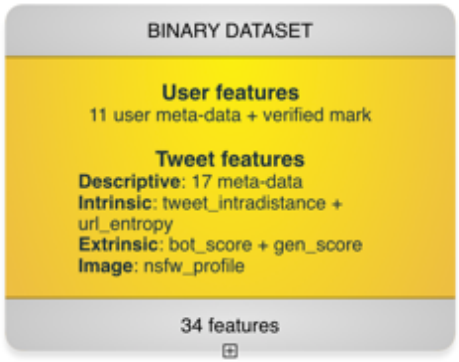
Chrome Archivio Modifica Visualizza Cronologia Preferiti Persone Finestra Guida
Mer 10 ott 11:28
BotBuster
localhost:4200/home
App Facebook Twitter YouTube G-Mail Welcome - BeeP Posta - lorenzo.ca... LinkedIn YouWeb full Amazon.it Tesi
BOTBUSTER
BUST IT

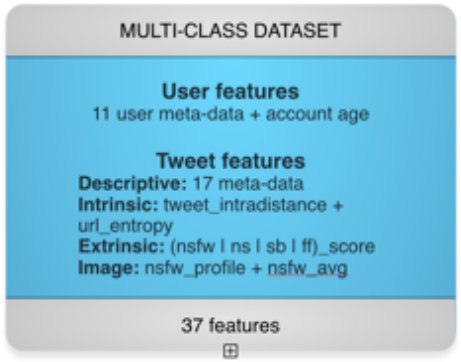
Thesis ingredients



Creation of datasets

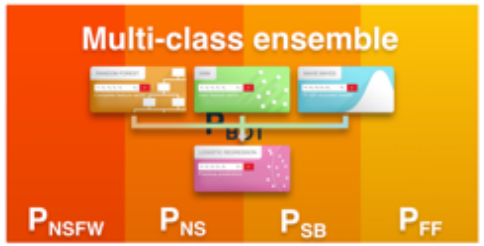




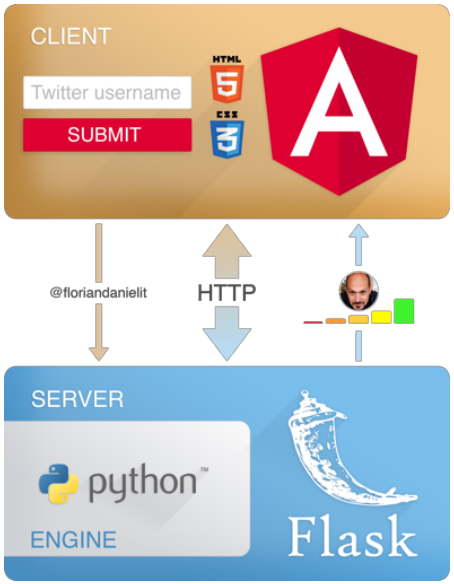


Feature selection and engineering

Bot-or-not classifier



Multi-class ensemble classifier



Web app

Can we add more types of harm?

How to train classes as users use BotBuster?

Identification of harmful communication **patterns**

Bot code repositories

GitHub

```
<div id="search" class="container-inline">
    <div class="form-item">

        <input type="text" maxlength="128"
name="search_theme_form_keys" id="edit-search-theme-form-keys"
size="15" value="" title="Enter the terms you wish to search
for." class="form-text" />
    </div><!-- /form-item -->

    <input type="image" src="/goband/files/icons/ico_search.jpg"
name="op" value="Search" id="search_icon"  />
    <input type="hidden" name="form_token" id="edit-search-theme-
form-form-token" value="26c6ef5d24a9b8af03b19f05131a19bc"  />
    <input type="hidden" name="form_id" id="edit-search-theme-
form" value="search_theme_form"  />
</div><!-- /container-inline -->
```

Which potential harms can we identify inside the code of the bots?

Harmful code **patterns**

Twitter

Python

Next: patterns search engine + web site for users

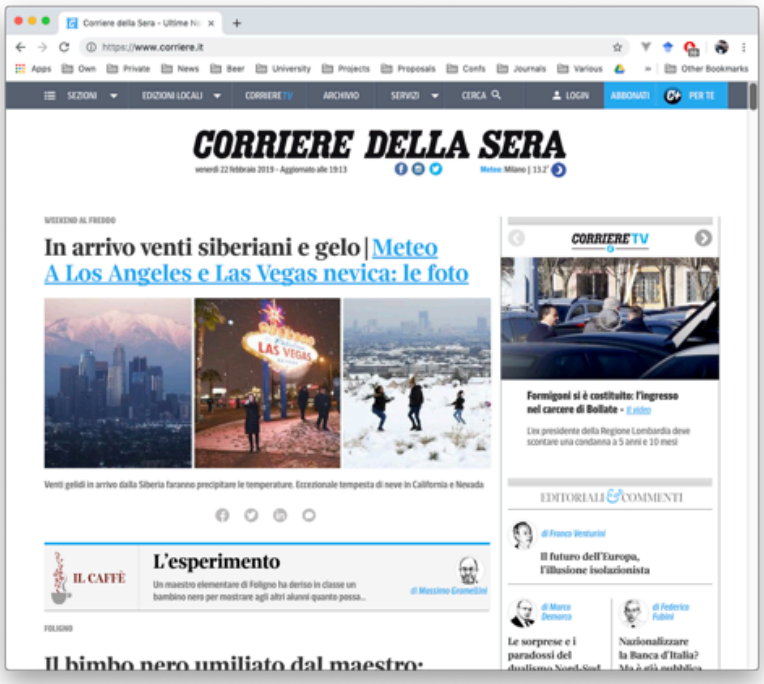
What else can we learn from the code of bots? Is it possible to trace back from messages to code?

Andrea Millimaggi and Florian Daniel. On Social Bots Behaving Badly: Empirical Study of Code Patterns on GitHub. Submitted for publication to ICWE 2019, under review.

Action	Pattern	Disclose sensitive facts	Denigrate	Be grossly offensive	Be indecent or obscene	Be threatening	Make false allegations	Deceive	Spam	Spread misinformation	Mimic interest	Clone profile	Invade space
Follow	Indiscriminate follow	—	C	C	C	—	—	—	—	—	GO	—	—
	Whitelist-based follow	—	⊘	⊘	⊘	—	—	—	—	—	GO	—	—
	Blacklist-based follow	—	⊘	⊘	⊘	—	—	—	—	—	GO	—	—
	Phantom follow	—	C	C	C	—	—	—	—	—	GO	—	—
Like	Indiscriminate like	—	C	C	C	—	—	—	—	—	GO	—	—
	Whitelist-based like	—	⊘	⊘	⊘	—	—	—	—	—	GO	—	—
	Blacklist-based like	—	⊘	⊘	⊘	—	—	—	—	—	GO	—	—
	Mass like	—	C	C	C	—	—	—	GO	—	GO	—	—
Tweet	Fixed-content tweet	⊘	⊘	⊘	⊘	⊘	⊘	—	GO	—	—	—	—
	AI-generated tweet	C	C	C	C	C	C	—	GO	C	—	GO	—
	Trusted source tweet	⊘	⊘	⊘	⊘	⊘	⊘	—	GO	—	—	—	—
Mention	Indiscriminate mention	—	GO	GO	GO	—	—	—	GO	—	GO	—	—
	Opt-in mention	—	GO	GO	GO	—	—	—	⊘	—	—	—	—
	Targeted mention	—	C	—	—	—	C	—	—	—	—	—	—
	Whitelist-based mention	—	⊘	⊘	⊘	—	—	—	—	—	GO	—	—
	Blacklist-based mention	—	⊘	⊘	⊘	—	—	—	—	—	GO	—	—
Retweet	Indiscriminate retweet	—	C	C	C	C	C	C	—	C	GO	—	—
	Whitelist-based retweet	—	⊘	⊘	⊘	⊘	⊘	⊘	—	⊘	GO	—	—
	Blacklist-based retweet	—	⊘	⊘	⊘	⊘	⊘	⊘	—	⊘	GO	—	—
	Mass retweet	—	C	C	C	C	C	C	GO	C	GO	—	—
Talk to	Indiscriminate talk	—	—	—	—	—	—	—	GO	—	GO	—	—
	Fixed-content talk	⊘	⊘	⊘	⊘	⊘	⊘	—	GO	—	GO	—	—
	AI-generated talk	C	C	C	C	C	C	—	GO	C	GO	GO	—
	Talk with opt-in	—	—	—	—	—	—	—	⊘	—	⊘	—	—
	Targeted talk	—	—	—	—	—	—	—	GO	—	GO	—	—
Pause	Mimic human	—	—	—	—	—	—	GO	—	—	—	—	—
	Satisfy API contraints	—	—	—	—	—	—	—	GO	—	—	—	—
Store	Store persistently	T	—	—	—	—	—	—	—	—	—	T	—

GO *Enables* ⊘ *Prevents* C *Vulnerable to content abuse* T *Vulnerable to trust abuse*

Conversational screen readers



Can we extract conversational knowledge from webpages?

What about “talking” to websites?

Domain-specific **content extraction**

ALL-GRAIN

Batch size: 5 gallons (19 liters)
Brewhouse efficiency: 72%
OG: 1.054
FG: 0
IBUs: 25
ABV: 7% ABV

MALT/GRAIN BILL

8 lb Pilsner malt
1 lb flaked corn
1 lb flaked rice

HOPS AND ADDITIONS SCHEDULE

.17 oz Mosaic [12.25% AA] at 15 minutes
2.5 oz Mosaic [12.25% AA] at Whirlpool
7.5 oz Mosaic [12.25% AA] at Dry Hop

YEAST

Neutral Ale yeast
.65 ml Amyloglucosidase

DIRECTIONS AND BREWER'S NOTES

Social Kitchen and Brewery's brewmaster Kim Sturdavant says: "I would encourage folks to add t
shoot for a mash temp of 143 - 146 to not denature the enzyme. The equivalent of 20mls / 100#
the enzyme 1/3 of the way into mashing to ensure the mash isn't too hot to denature the enzyme
mash have an hour-long rest to let the enzyme do its work."

"The other option would be to add the enzyme to the kettle while lautering, then hold off on heating the kettle until all wort is in and has 30 minutes of contact with the enzyme at around 145 degrees (lower would be fine, too)."

"I know a lot of homebrewers just throw grain on top of all their mash water, so this option makes more sense to them. I preferred not to mash this way, when I was a homebrewer... I think the conversion temp is more consistent adding grain and water at the same time, I also like being able to react to the consistency so I can end up the thickness I want."

"The enzyme will denature in the boil, but it's already done its work."

"That would all mean that the amylo in the fermentation is no longer needed. I'm finding better results with fermentation character and hop aromatics by having no enzyme present in the fermentation."

Lastly, it is very important to add nutrient to the boil as well as 1/2 way through fermentation (with something like BSG's Startup) since the wort will be almost entirely glucose, there is not a lot of nutrition for the yeast.

The Grist

The grist described by California brewers is either entirely pale malt or pale malt with only the slightest amount of specialty malts. Do not include crystal malts (or dextrine malts) in your formulation as these add body and sweetness to the beer, the opposite of what you are trying to achieve. Either US 2-row pale malt, English 2-row pale malt, or a European Pilsner malt could be used as the base malt. Small amounts of either Munich or Vienna malt could enhance the malt character without adding body or sweetness. The complex carbohydrates ("dextrins") in all the malts will be greatly, or entirely, converted to simple sugars, but the malt flavors from the husks of the kilned base malts will persist.

Higher gravity beers are going to be harder to dry out. And, even if you accomplish that, they are more likely to seem like "rocket fuel" than beer. I would suggest shooting for an OG in the high pale ale range (SG 1.045–1.060) to low IPA range (SG 1.056–1.070). However, there are no rules here, so feel free to experiment.

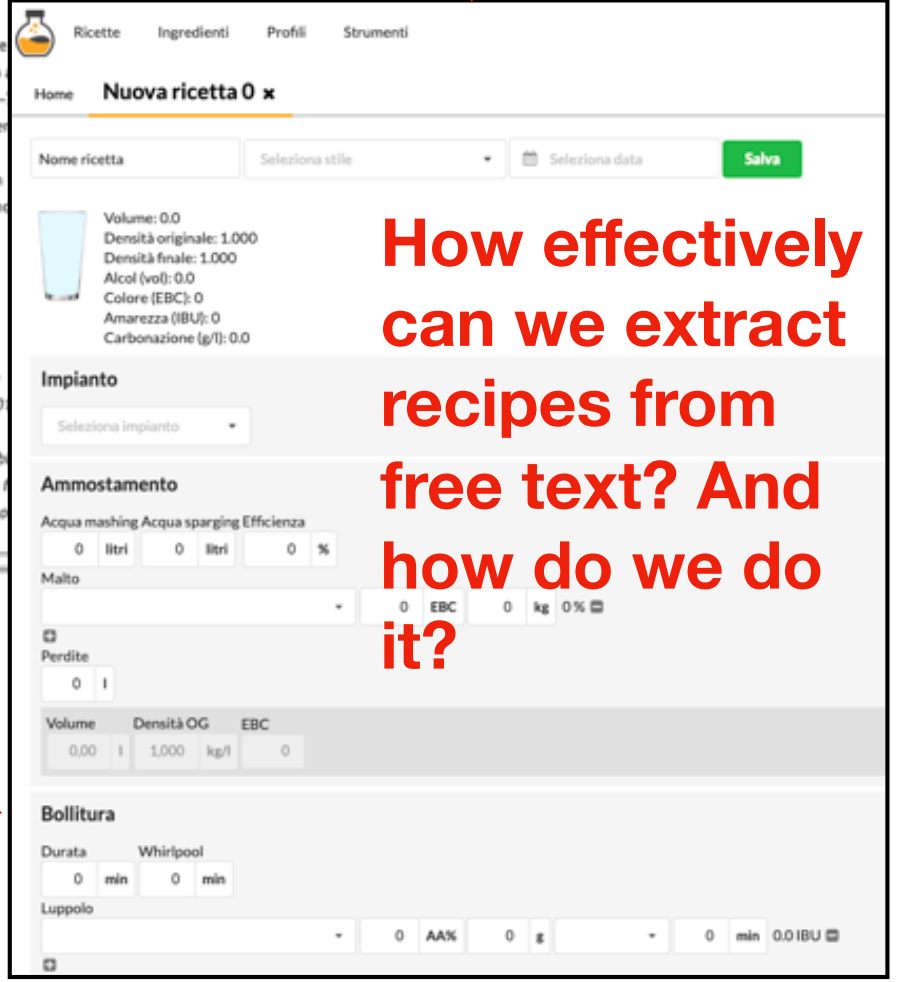
The Mash

There are several ways to produce highly fermentable
wort that will yield a dry beer. A single infusion mash
the low end of saccharification range (148–162 °F/64–
°C), say 148–150 °F (64–66 °C), will produce a dry beer
Stirring the mash well a few times and extending the
mash time to 90 minutes may help a bit. A step mash
with a lengthy rest — 15 minutes to 2 hours — around
140 °F (60 °C) will produce an even drier beer.

Tomorrow, I'll describe how to use the enzyme amyloglucosidase to make a bone dry beer.

If you enjoy Beer & Wine Journal, please consider supporting us by purchasing one of my books, which include "Home Brew Recipe Bible," by Chris Colby (20
Page Street Publishing) and "Methods of Modern Homebrewing," by Chris Colby (2017, Page Street Pub
Amazon (linked) and Barnes and Noble. You can also
carries them through Indiebound. You can also suppo
Thank you.

HomebrewLab
DESIGN YOUR BREWING EXPERIMENTS



How effectively can we extract recipes from free text? And how do we do it?

Typical data science steps

Online tool

Domain understanding

Data collection

Manual inspection of data

Hypotheses formulation

Feature engineering, data labeling

Algorithm engineering (from AI/machine learning to statistics)

Validation and hypotheses verification

Soft skills

Proposals

Template

Home
Research
Projects
Publications
Books
For students
Teaching
Thesis proposals
- Harmful social bots
- Conversational screen readers
- Conversational data exploration
- Service-oriented computing on blockchain
- Blockchain-based service level agreements
- Domain-specific content extraction

Thesis template and formalities
Curriculum vitae
Contact

Follow me on ResearchGate

© Copyright 2012-today by Florian Daniel

Thesis/project proposals

Are you looking for a project for your Master thesis or for some other project you would like to carry out during your studies, for example for the BSc course Progetto di Ingegneria Informatica? Below you can find some **research topics** I am or would like to work on. All these can be turned into projects and theses and are meant to allow you to get a glimps behind the scenes of the work of a professor and to show you that there is more than teaching in a university. Methodologically, typically these topics are to be approached using a **data science** and/or **web/software engineering** methodology.

If you spot something you like, drop me an email or get in touch with me! I'll be happy to discuss possible options to work together. Also, yes, you can work in groups of two and write and present your thesis together.

What you need: programming skills, knowlege of web technologies, data management skills, a brain and passion.

What you get: new competences, a thesis (or tesina), a project, personal satisfaction and recognition of your work.

Writing your thesis: yes, a thesis is also about writing and writing skills. Here some thoughts on writing and a possible template for your thesis.

Harmful social bots

Topics: social bots, social networks, data science, ethics

Bots are algorithmically driven entities that behave like humans in online communications. They are increasingly infiltrating social conversations on the Web or in chat apps. If not properly prevented, this presence of bots may cause harm to the humans they interact with, e.g., they may offend or discriminate people. The goal of this research is to understand which types of harm and which abuses may happen, whether abuses can be considered intentional or not, whether it is possible to prevent them and, if yes, how.

Research questions and possible thesis/project topics

- **Bot detection:** Given a specific type of abuse, e.g., discrimination, spamming or mimicking interest, how can we identify bots that may be vulnerable to this type of abuse? Answering this question requires following a data science methodology and may require the use of machine learning and AI algorithms, online/reinforcement learning techniques, network analysis techniques, and similar. Different social networks may be studied. The expected output are algorithms or methods able to classify social network accounts based on the harm they may cause.
- **Code analysis:** Here the idea is to start from the code of bots shared online, e.g., in GitHub, and to study which abusive code patterns can be identified, which effects they may have, and how to spot bots that implement them. Again, answering these questions requires a data science approach and may imply the

Florian Daniel

http://www.floriandaniel.it