

Software Requirements

L5. Mining User Feedback



Christoph Stanik



Prof. Walid Maalej



Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Overview

1

Motivation, definition, and conceptual model

2

Mining requirements-related feedback

3

Identifying and matching app features

4

Summary

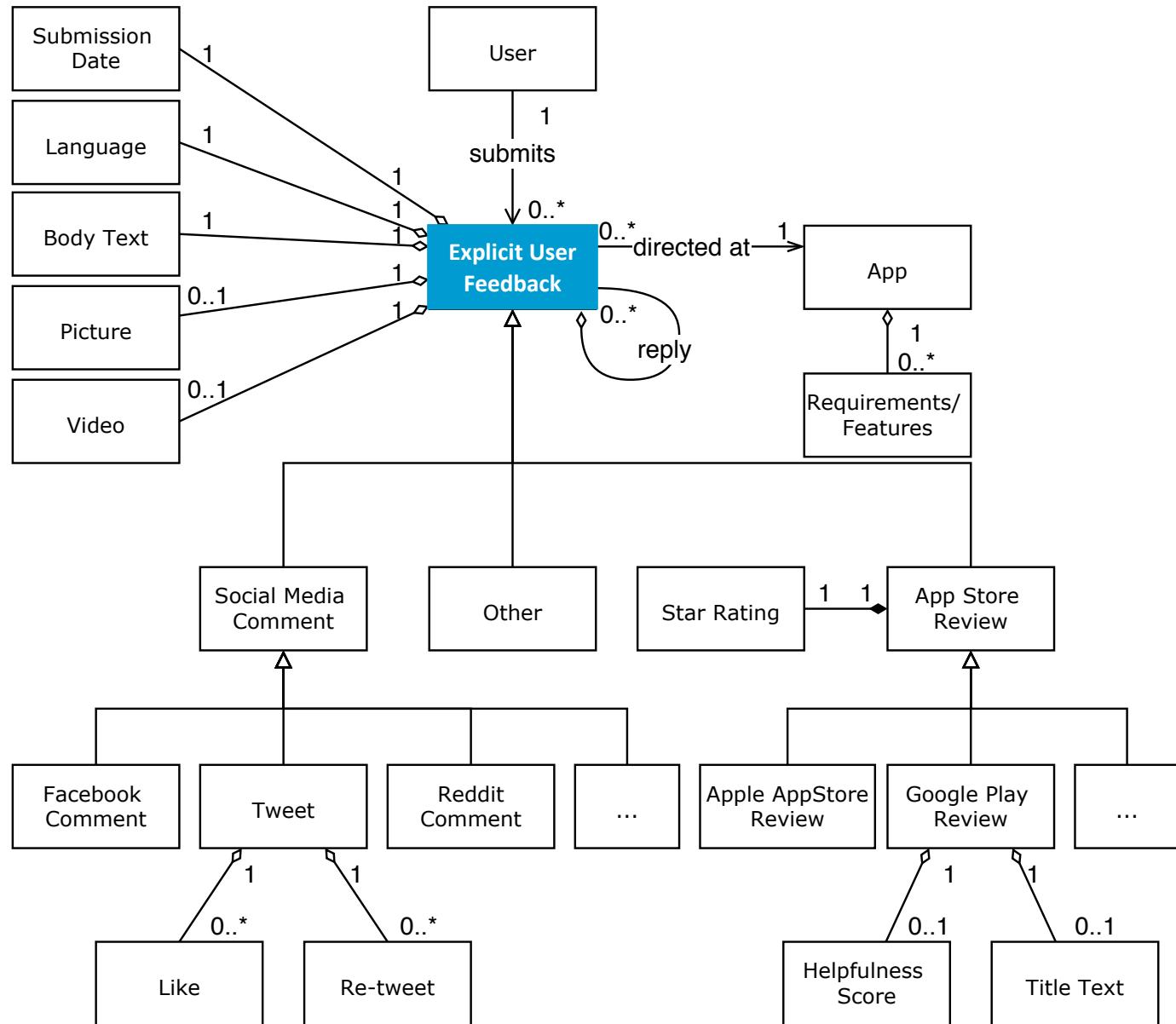
What is explicit user feedback?

“Explicit User Feedback is the [written opinion of a user](#). In its base-form, it is freely-written text addressing any aspect of a software or service provided by an entity (e.g., company or single developer). It may have a numerical rating attached that expresses the users’ overall satisfaction and experience.”

Explicit
User
Feedback

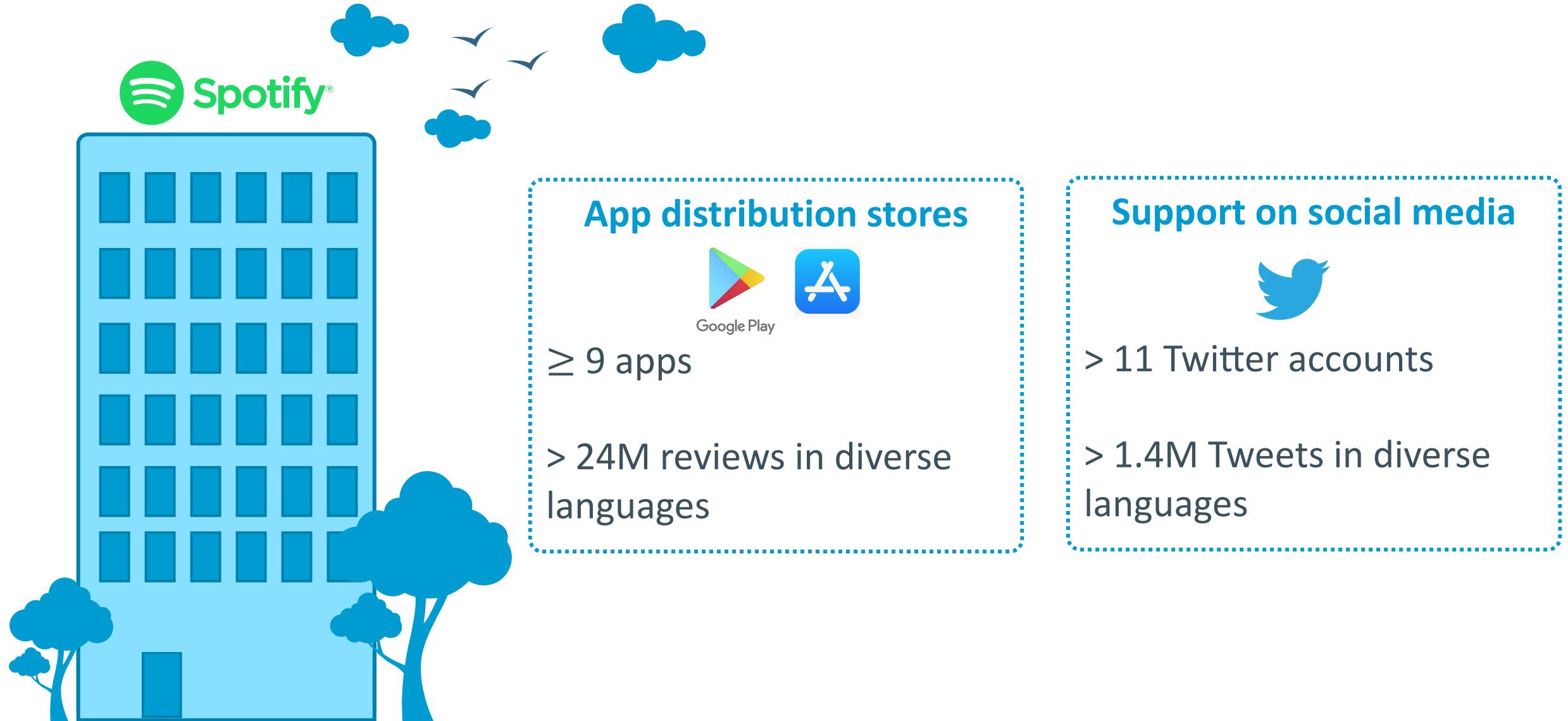
- is freely-written (unstructured) text
- contains the opinion of a user (subjective feedback)
- source for requirements-related information

Conceptual model

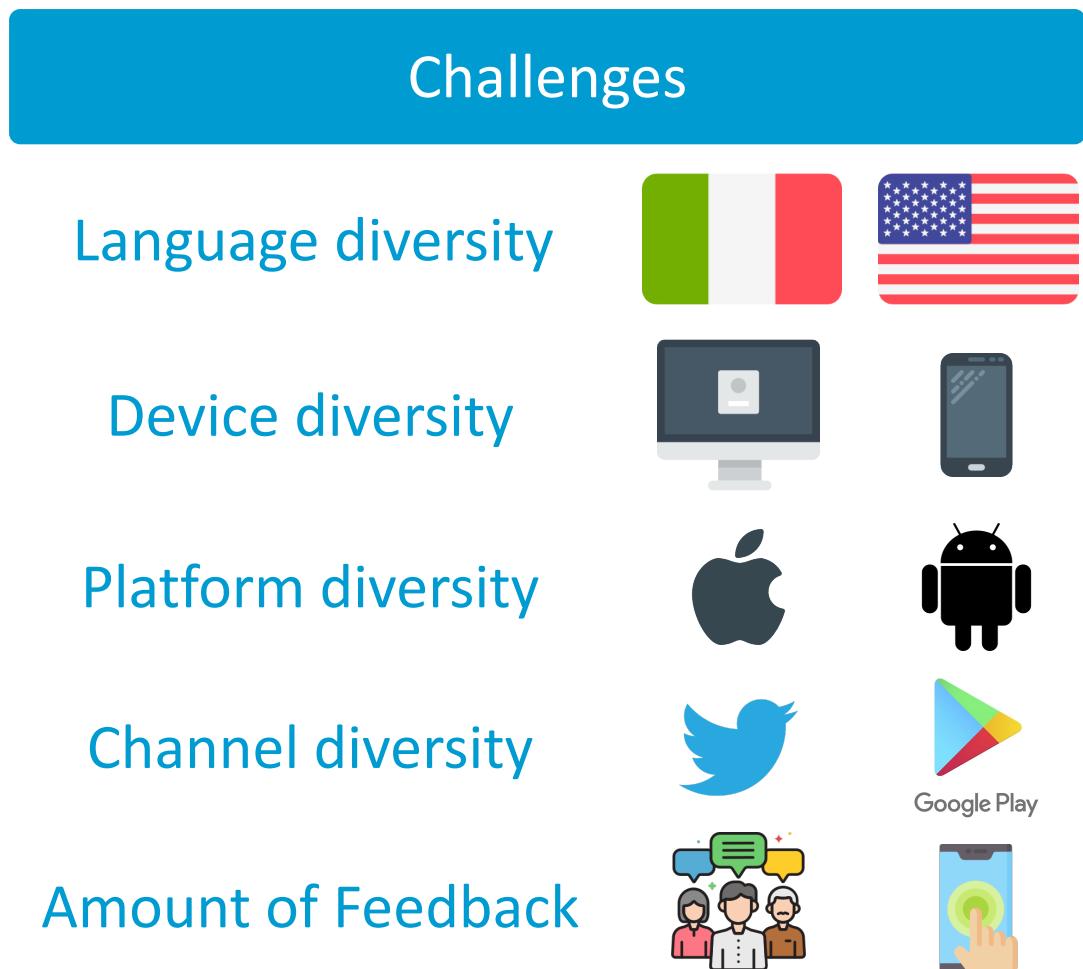


Why is
Explicit User Feedback
important?

Exemplary situation at Spotify



Understanding user needs is difficult!



Types of User Feedback

Explicit Feedback

(e.g., Reviews, Social Media)



Bojan Alempijevic ★★★★★



Doesnt work with sound
enhancements. With others,
such as MS Groove and even

William and Justine Robinsc ★★★★★



I am always listening to music,
constantly switching from my
headphones, to my car stereo

Sean Lyons ★★★★★



Found a bug. Having an issue
with the 'create playlist' section.
Everytime I tap to create a

Robert Stones ★★★★★



The app itself is good but you
cannot pay for the Family
Subscription by Direct Debit as

Implicit Feedback

(e.g., Interaction Data)

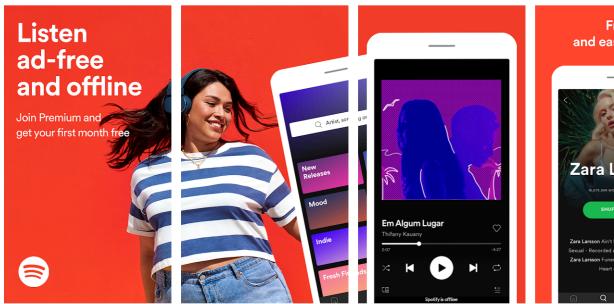


session_id	packagename	Type	Internet	event_type	longitude	latitude	Data sent
mhma124ua_1594	com.facebook.orca	MOBILE	TRUE	change content	-97.54606246	35.42715311	13401779
2tcbl7bqg_18027	com.lge.launcher3	WIFI	TRUE	change content	-94.20432142	45.56927969	0
r4opbgu5a_3434	com.facebook.katana	WIFI	TRUE	change content	-80.65248462	35.50420719	581203
phackhfw_1620	com.android.chrome	WIFI	TRUE	change content	-76.30209865	36.89856881	0
r4opbgu5a_4237	com.facebook.katana	WIFI	TRUE	scroll view	-80.65241874	35.50404454	581203
mhma124ua_1594	com.facebook.orca	MOBILE	TRUE	change content	-97.54606246	35.42715311	13401779
2tcbl7bqg_18027	com.lge.launcher3	WIFI	TRUE	change content	-94.20432142	45.56927969	0
r4opbgu5a_3434	com.facebook.katana	WIFI	TRUE	change content	-80.65248462	35.50420719	581203
phackhfw_1620	com.android.chrome	WIFI	TRUE	change content	-76.30209865	36.89856881	0
r4opbgu5a_4237	com.facebook.katana	WIFI	TRUE	scroll view	-80.65241874	35.50404454	581203

Where can we get explicit user feedback from?



Spotify: Listen to new music, podcasts, and songs
 Spotify Ltd. Music & Audio
 ★★★★ 18,454,289
 USK: Ages 12+
 Contains Ads
 This app is compatible with some of your devices.



With Spotify, you can play millions of songs and podcasts for free. Listen to the songs and podcasts you love and find music from all over the world.

- Discover new music, albums, and podcasts
- Search for your favorite song, artist, or podcast
- Enjoy playlists made just for you
- Make and share your own playlists
- Find music for any mood and activity
- Listen on your mobile, tablet, desktop, PlayStation, Chromecast, TV, and speakers

Play music for free on your mobile and tablet with Spotify. Listen to music, albums, playlists, and podcasts wherever you are.

With Spotify, you have access to a world of free music, curated playlists, artists, and podcasts you love. Discover new music, podcasts and listen to your favorite artists, albums, or create a playlist with the latest songs to suit your mood.

Spotify also offers free music, curated playlists and thousands of podcasts you can't find anywhere else. Find music from your favorite artists, and listen to new music for free.

- Free music and podcasts made easy – Listen to a playlist, album or play music by any artist on shuffle mode

Listen to music and podcasts on your tablet for free
 • Play any song, artist, podcast, album, or playlist and enjoy a personalised music experience

- Some users give feedback, including a freely-written text and a numerical rating

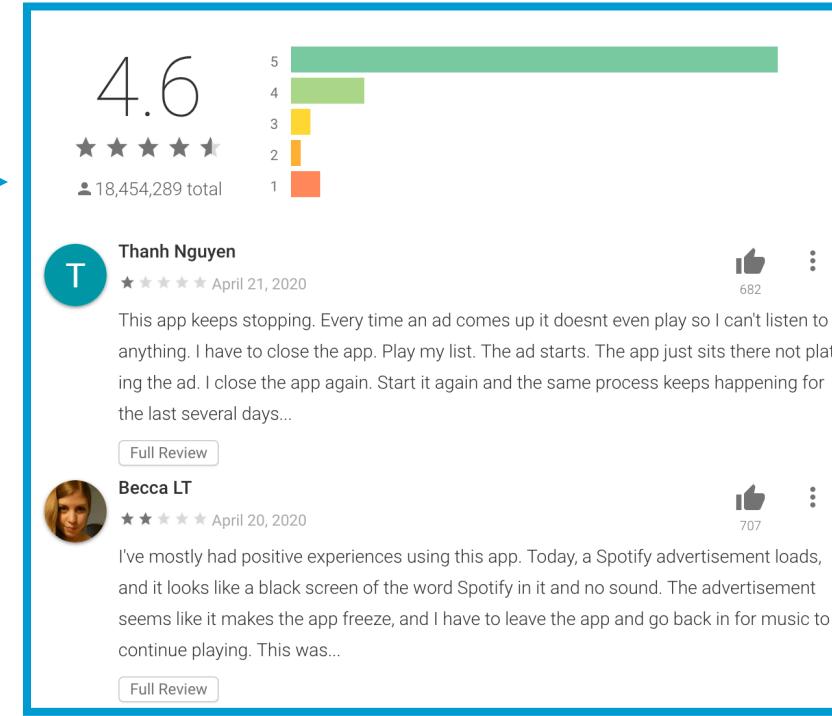


- App reviews contain valuable requirements-related information (e.g., bug reports)



- Stakeholders use app descriptions to describe functional requirements (features)

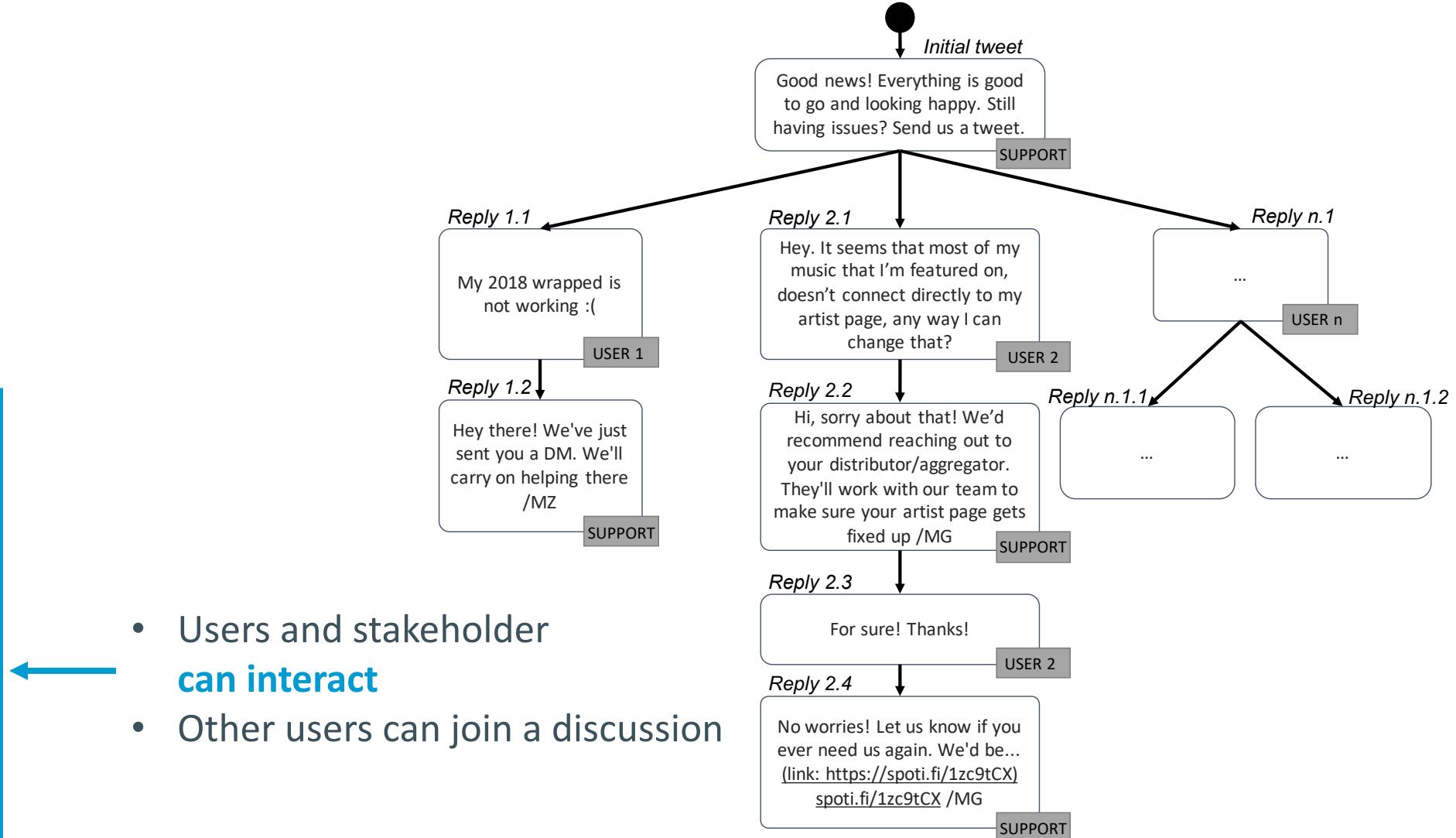
- However the descriptions are often incomplete



Where can we get explicit user feedback from? Twitter

The screenshot shows the official Snapchat Support account (@snapchatsupport) on Twitter. The profile picture is the yellow Snapchat logo. The bio reads: "Snapping, snacking, and supporting! Discover tips, find answers to common questions, and get even more help on our support site: support.snapchat.com". The account has 688.2K followers and joined March 2014. Recent tweets include:

- A reply to a user locked out of their account, directing them to a link for login issues.
- A reply to a user who can't log in, suggesting they use email to reset the password.
- A reply asking for a screenshot of a login prompt.



- Users and stakeholder **can interact**
- Other users can join a discussion

Topics in user feedback

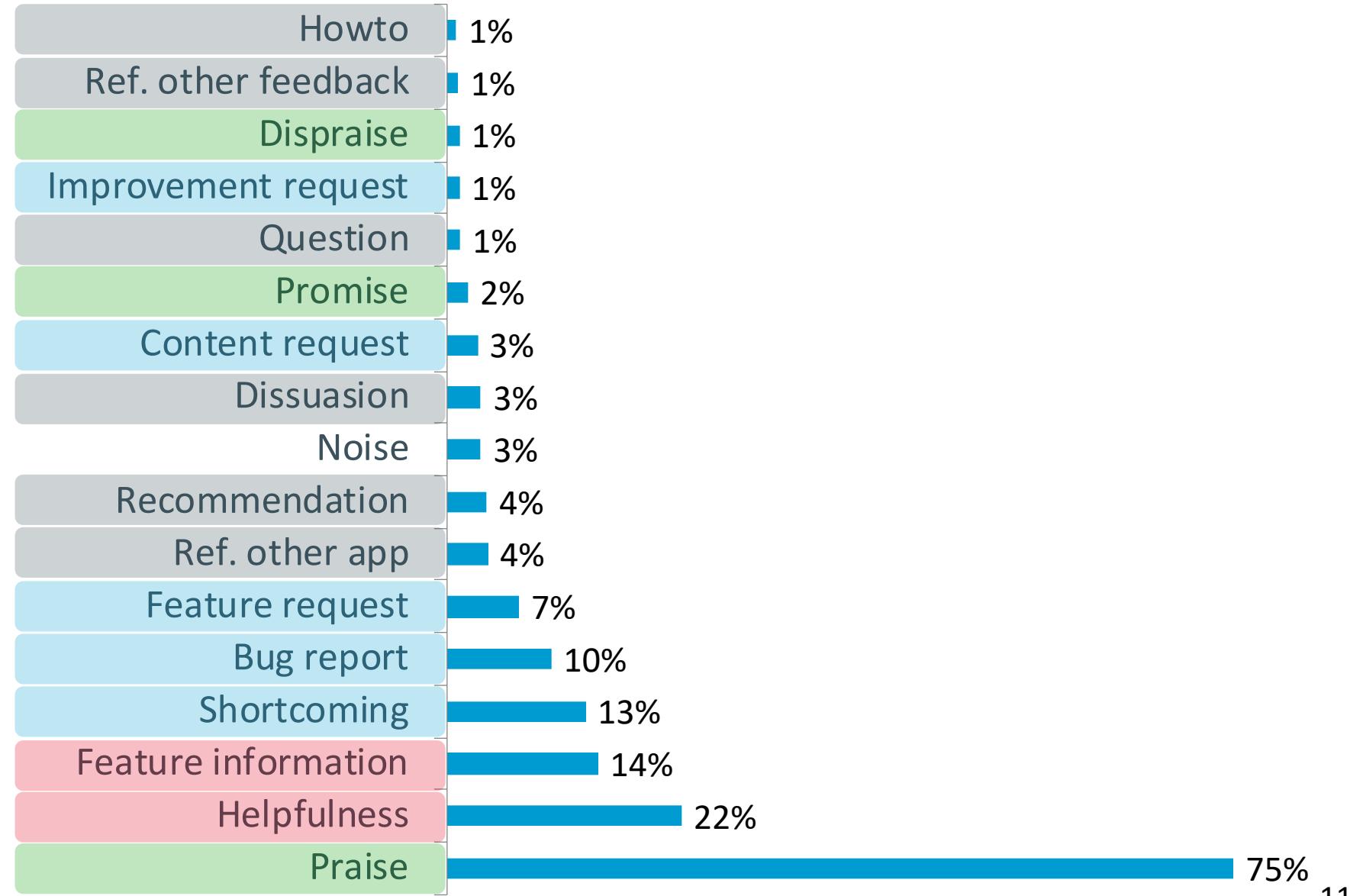
Requirements

Community

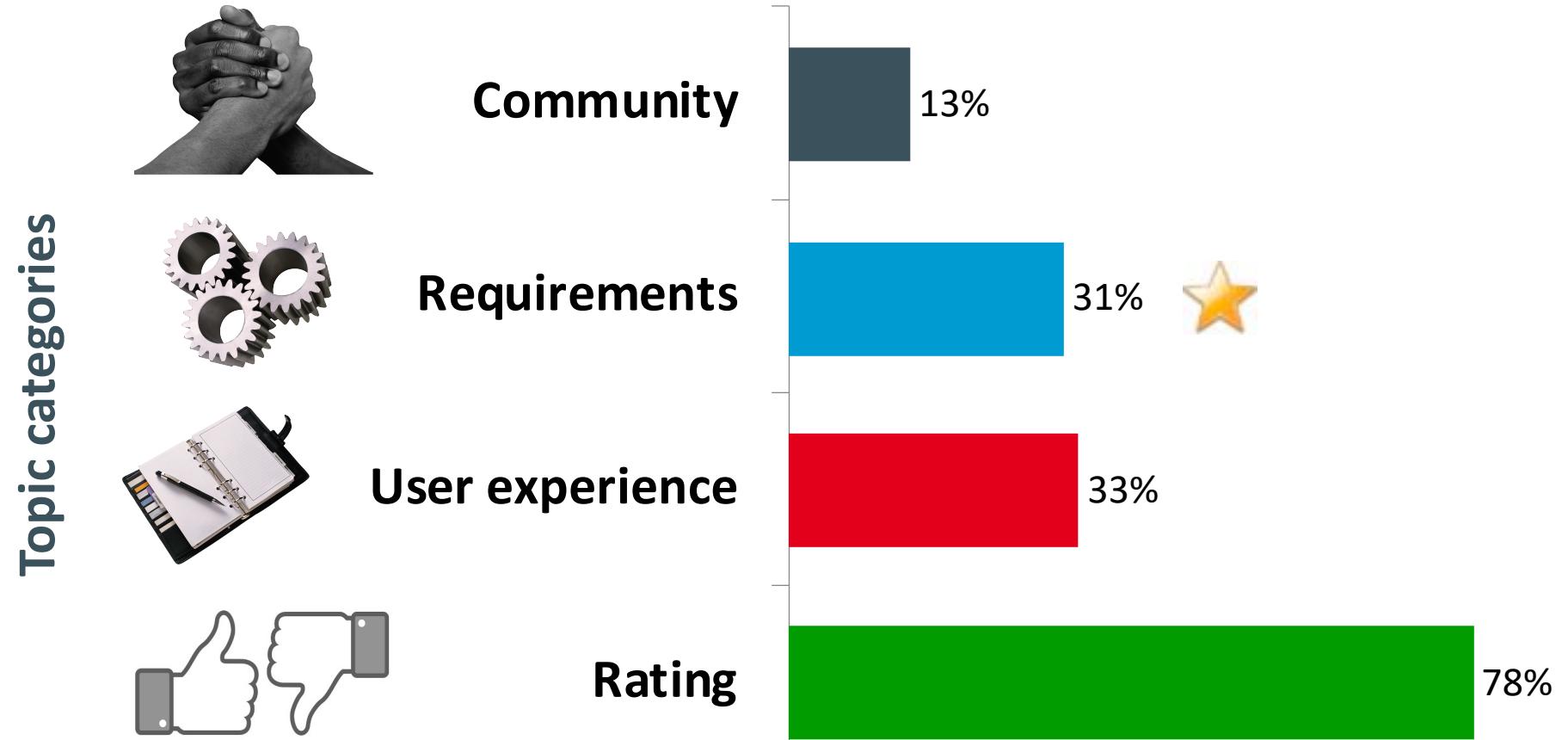
Rating

User Experience

[Pagano and Maalej 2013]



Frequency of user feedback by topic category



% of entries that includes at least one of the associated topics

Similar findings [Galvis and Winbladh 2013]

Overview

1

Motivation, definition, and conceptual model

2

Mining requirements-related feedback

3

Identifying and matching app features

4

Summary

Major challenges

Amount of feedback



Feedback channels



Google Play



Feedback in many languages



I am not going to read all of that!

I only want to read relevant feedback.



Should we care about user feedback?



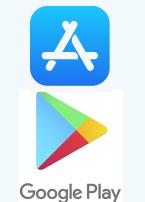
Users write about what they like and dislike



Popular apps such receive about **4,000 reviews and about 31,000 tweets on Twitter daily**
(Pagano and Maalej, 2013; Guzman et al., 2016)



51% tweets contain technical information
42% tweets are bugs, feature short comings or requests
(Williams and Mahmoud, 2017; Guzman et al., 2017)



10% app reviews report bugs/problems
7% requests more features
(Pagano and Maalej, 2013)

Feedback category examples

Irrelevant/Noise:

- Very bad.
- Cool. It is the best app ever.
- This game sucks.



This is not helpful!

Problem Report:

[...] I loved SoundHound before the last update. Now I can't view my history. I've tried several times [...] I just give up. Back to using Shazam. I hope this gets fixed.



Problem reports help us to fix issues early and to validate our existing requirements.

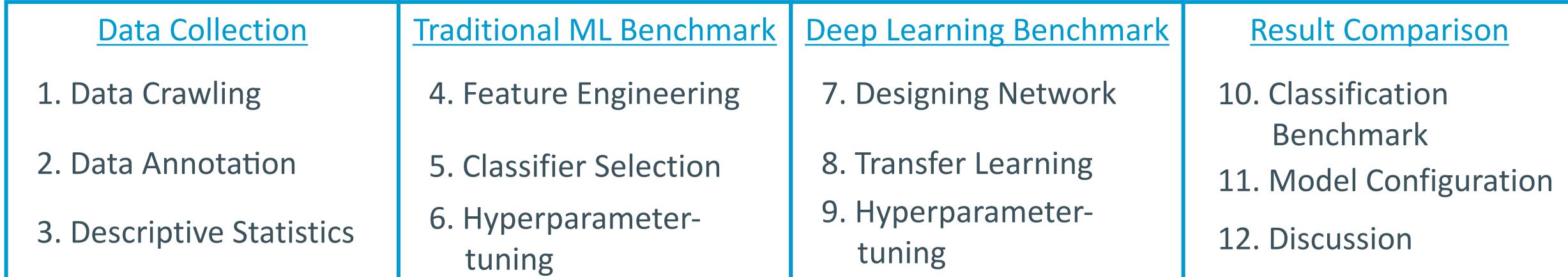
Inquiry/Feature Request:

You can't even view your sorted contact groups. That needs to change. What's the point?

Inquiries may help us in the elicitation to find new requirements.

Filter tweets and app reviews into
Problem Reports, Inquiries, and
Irrelevant

Study Design and Data



Study Data			
Class	App Reviews (EN)	Tweets (EN)	Tweets (IT)
# problem reports	1,437	2,933	4,414
# inquiries	1,100	1,405	2,594
# irrelevant	3,869	6,026	9,794
TOTAL	6,406	10,364	15,802

The importance of high-quality data



Needed for good classification results

Time vs. budget vs. availability of experts

Crowdsourcing vs. domain experts

How to reach high quality?

Coding guide

Goal: create a shared understanding of the task.

- Explains each class
- Gives positive and negative examples
- Language check by a native
- Created in cooperation with stakeholders

Pilot study

Goal: test the coding guide and coder.

- Curated data with easy and difficult cases
- Define metrics (e.g., inter-coder agreement)

Intercoder-agreement

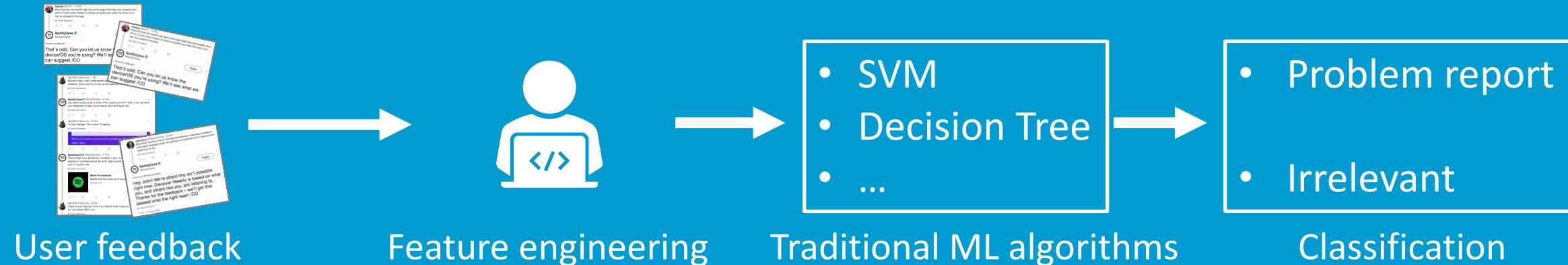
Goal: statistical check of the data quality

Metrics:

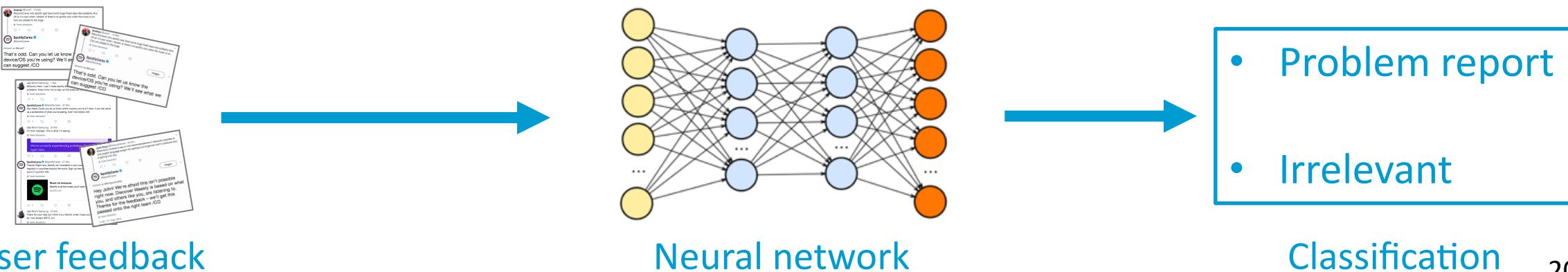
- Simple mean
- Majority vote
- Kappa statistics
- Correlation coefficient
 - Pearson's r
 - Kendall's τ
 - Spearman's ρ

Traditional Machine Learning vs Deep Learning

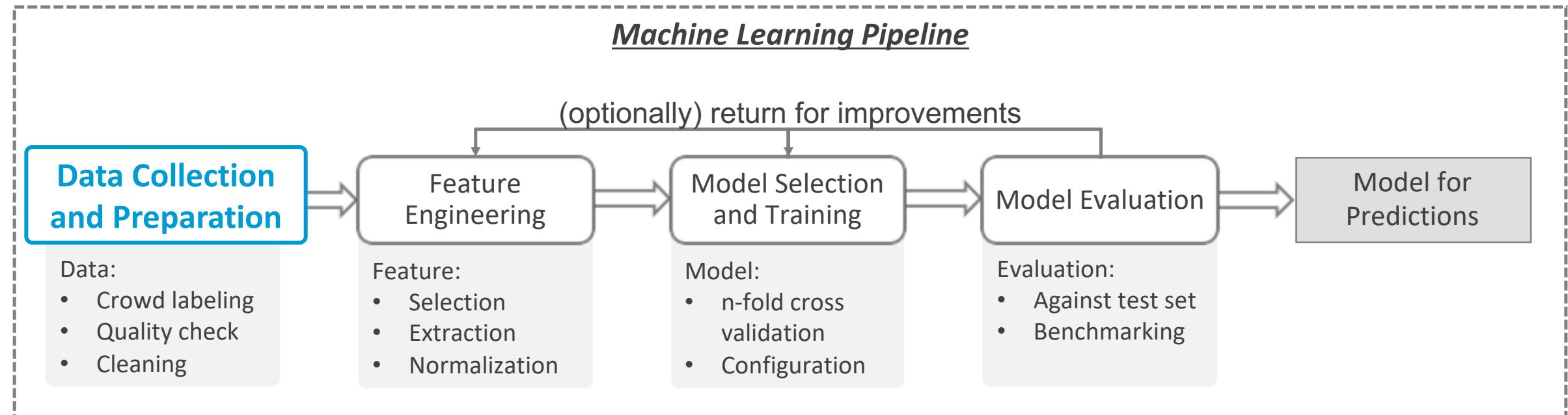
Traditional machine learning



End-to-end learning



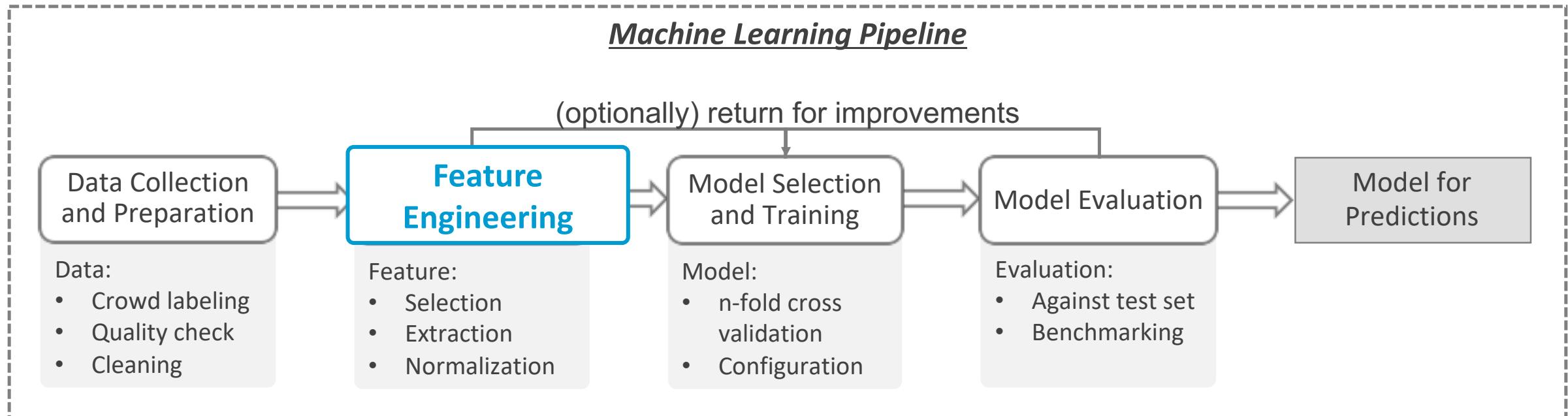
Our Traditional Machine Learning Approach



Cleaning:

- | | |
|---|---------------------------------------|
| 1. To lower case | ➤ Feature, feature, FEAture → feature |
| 2. Perform masking (e.g., for accounts) | ➤ @MyCompany → [account] |
| 3. Lemmatization | ➤ See, saw, seen, seeing → see |

Our Traditional Machine Learning Approach



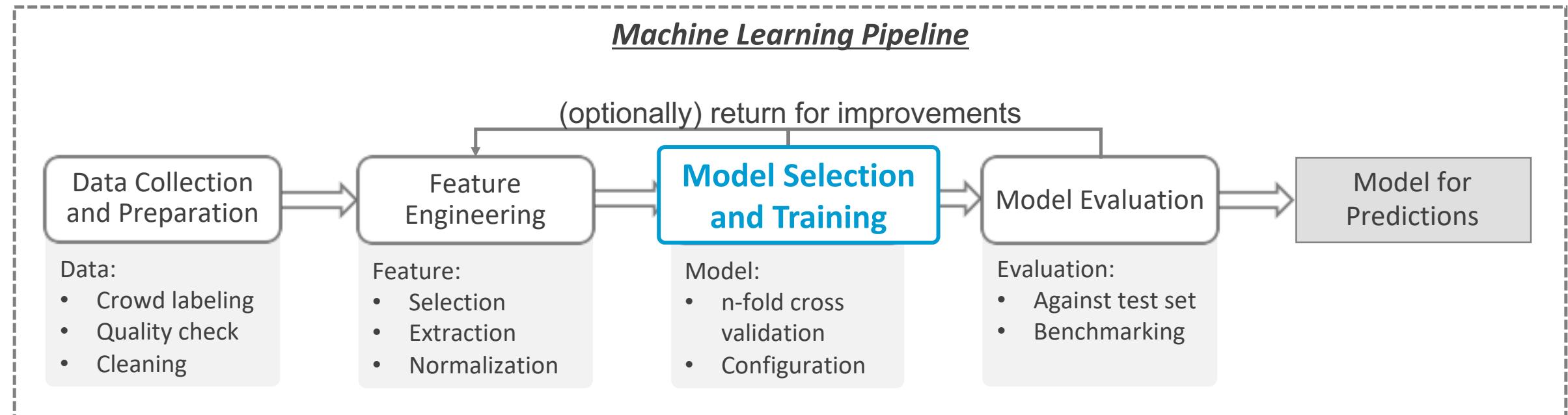
Feature Engineering:

- Only text-based features
- Total features:
 - 1.047 app reviews, EN
 - 1.281 tweets, EN
 - 1.301 tweets, IT

Features:

- | | |
|----------------------|------------|
| • # words | • Keywords |
| • # stopwords | • Tense |
| • Negative sentiment | • Tf-idf |
| • Positive sentiment | • fastText |

Our Traditional Machine Learning Approach

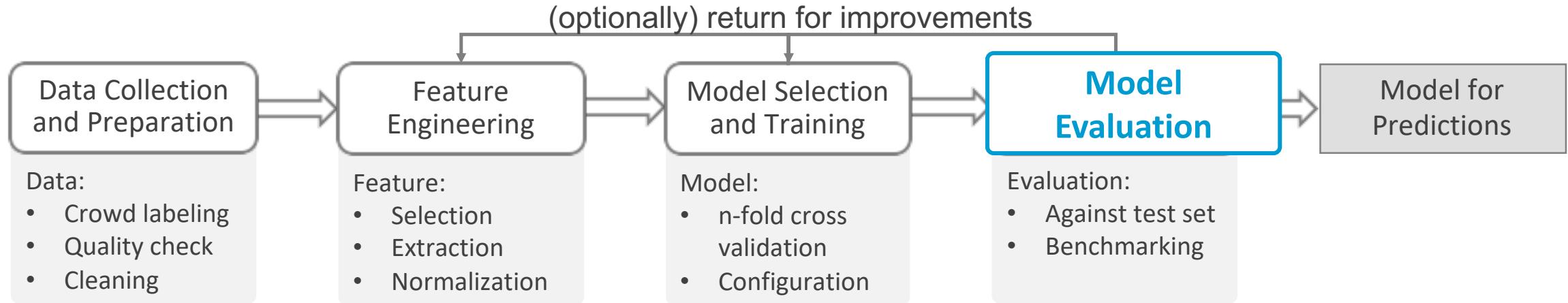


Experiment Configuration

- Dim. 1:** Best performing features
- Dim. 2:** Effect of scaling features
- Dim. 3:** Hyper-parameter tuning
- Dim. 4:** Effect of data balancing
- Dim. 5:** Machine learning algorithms

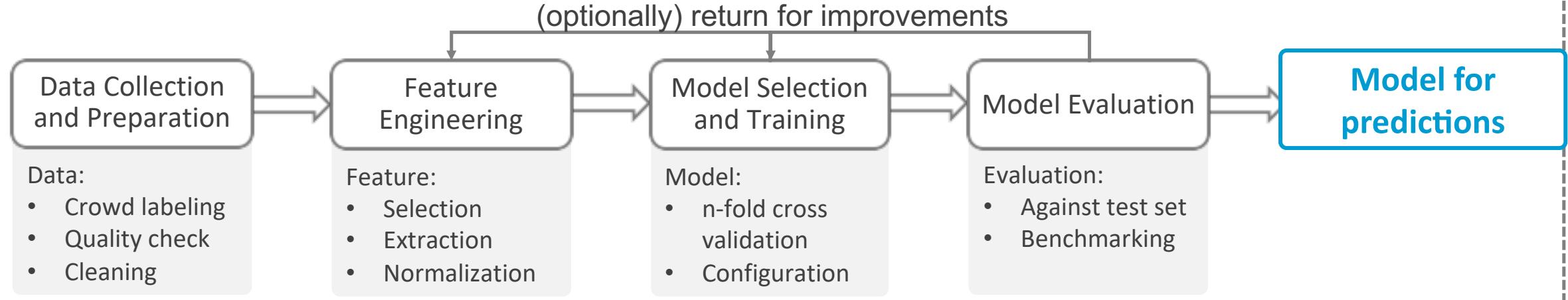
Machine learning experiments

Machine Learning Pipeline

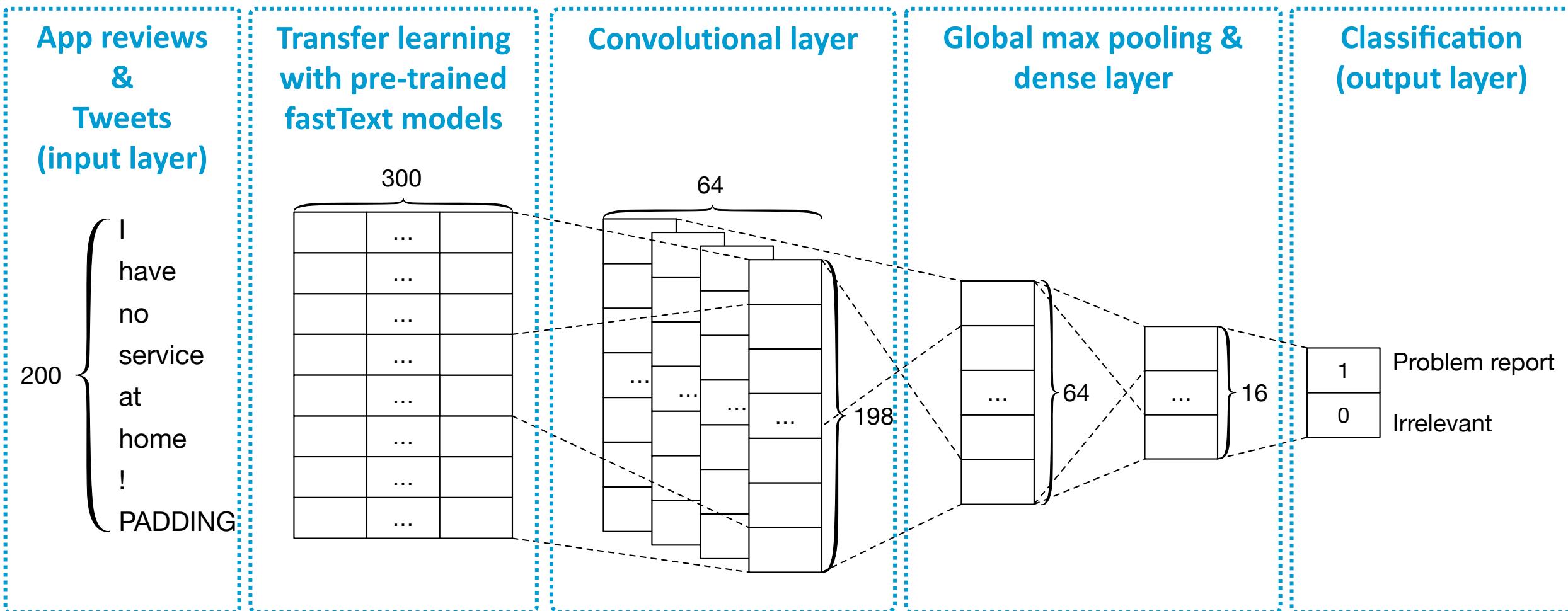


Machine learning experiments

Machine Learning Pipeline



Our Deep Learning Approach



Classification Results

Evaluation Metrics

- Precision
- Recall
- F1
- ROC AUC

Precision	Recall	F1	ROC AUC
0.55	1.00	0.71	0.47



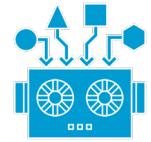
What can go wrong when only reporting F1?

Benchmark Results

		App Review (EN)		Tweet (EN)		Tweet (IT)	
	Class	f1	AUC	f1	AUC	f1	AUC
Traditional ML	problem	.79	.85	.59	.72	.65	.83
	inquiry	.72	.85	.43	.73	.60	.82
	irrelevant	.89	.86	.74	.69	.83	.73
Deep Learning	problem	.52	.82	.46	.74	.59	.84
	inquiry	.74	.94	.40	.75	.54	.83
	irrelevant	.85	.90	.72	.75	.81	.86

Summarized results

Irrelevant



F1 up to .89



.94

Problem Report

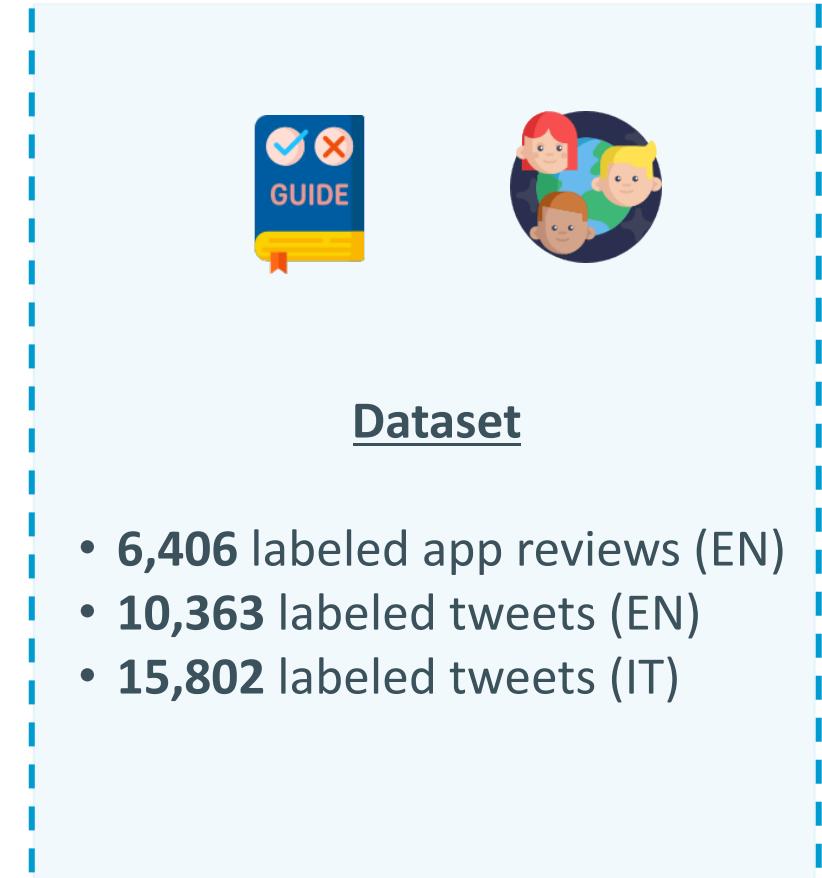
F1 up to .79

.89

Inquiry

F1 up to .74

.93



Some Findings

General

- High quality data is crucial. If the annotators don't understand their task, how shall a machine do so?
- Carefully select the machine learning performance metrics as their interpretation may be tricky

Traditional Machine Learning

- Works well with few data
- Benefits much from expert knowledge
- High effort in feature engineering
- Supporting more languages requires much effort

Deep Learning

- Excels with much data (but not necessarily)
- Does not require domain expert knowledge
- No feature engineering
- Relies on pre-trained models such as fastText

What is it good for? Field of applications

Filter



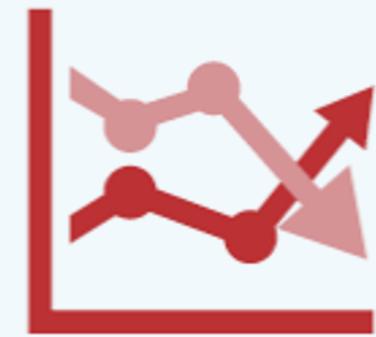
Allocate and assign



Compare



Track over time



Overview

1

Motivation, definition, and conceptual model

2

Mining requirements-related feedback

3

Identifying and matching app features

4

Summary

The data we work with

App descriptions

Facebook

By Facebook, Inc.

Open iTunes to buy and download apps.

Description

Keeping up with friends is faster than ever.

- See what friends are up to
- Share updates, photos and video
- Get notified when friends like and comment on your posts
- Watch and interact with live video
- Play games and use your favorite apps
- Buy and sell items and services

Read our Data Use Policy, Terms and other important info in

App reviews



Imogen Jackson June 20, 2017

★★★★★

It's good... Apart from the WiFi and Mobile Data issues: it doesn't send/receive messages via WiFi, only Mobile Data; which annoys me. I don't wallop in excess data so it is difficult for me to message my friends, considering none of them have any credit to text me. That's what I liked about this app when I had it before these updates, is that I could message them freely. Now I can't.

I advertise our apps by documenting its features.

But... which of our app feature are addressed?



**Extract and match app features
stakeholders document and users
address in their feedback**

Explicit Feedback to Requirements: Study Design

Approach:



Explicit Feedback to Requirements: Results

App Description

Facebook

By Facebook, Inc.

Open iTunes to buy and download apps.

Description

Keeping up with friends is faster than ever.

- See what friends are up to
- Share updates, photos and video
- Get notified when friends like and comment on your posts
- Watch and interact with live video
- Play games and use your favorite apps
- Buy and sell items and services

Read our Data Use Policy, Terms and other important info in:

App Reviews



Imogen Jackson June 20, 2017

★★★★★

It's good... Apart from the WiFi and Mobile Data issues: it doesn't send/receive messages via WiFi, only Mobile Data; which annoys me. I don't wallop in excess data so it is difficult for me to message my friends, considering none of them have any credit to text me. That's what I liked about this app when I had it before these updates, is that I could message them freely. Now I can't.

1. Extract App Features

- Share pictures
- Follow my friends
- Send messages
- See notifications

2. Match App Features

Approach	Eval. set	P.	R.	F1
Harman et al.	197	.28	.29	.28
Our approach		.56	.43	.49

F1 of .63

- Share photos
- Share videos
- Send messages
- Group chats

Approach	Eval. set	P.	R.	F1
Guzman and Maalej	244	.22	.28	.25
Our approach		.24	.71	.48

The Process: 1. Data Preparation

Tokenize sentences

Remove explanations

Filter noise (e.g., URLs, contact data)

Remove subordinates

Remove custom stopwords

Part-of-Speech tagging

The Process: 1. Data Preparation

Tokenize sentences

Remove explanations

Filter noise (e.g., URLs, contact data)

Remove subordinates

Remove custom stopwords

Part-of-Speech tagging

I love this app



"I", "love", "this", "app"

The Process: 1. Data Preparation

Tokenize sentences

Remove explanations

Filter noise (e.g., URLs, contact data)

Remove subordinates

Remove custom stopwords

Part-of-Speech tagging

The app allows you to upload PDFs
(you will love it!)



The app allows you to upload PDFs

The Process: 1. Data Preparation

Tokenize sentences

Remove explanations

Filter noise (e.g., URLs, contact data)

Remove subordinates

Remove custom stopwords

Part-of-Speech tagging

please email us at: android-support@xxxx.com
or follow us on twitter: http://XXXX



[completely removed]

The Process: 1. Data Preparation

Tokenize sentences

Remove explanations

Filter noise (e.g., URLs, contact data)

Remove subordinates

Remove custom stopwords

Part-of-Speech tagging

The app let you upload images, so that you don't have to worry about losing them



The app let you upload images

The Process: 1. Data Preparation

Tokenize sentences

Remove explanations

Filter noise (e.g., URLs, contact data)

Remove subordinates

Remove custom stopwords

Part-of-Speech tagging

Please fix this app, it always crashes when I upload images



Please fix this app, it always crashes when I upload images

The Process: 1. Data Preparation

Tokenize sentences

Remove explanations

Filter noise (e.g., URLs, contact data)

Remove subordinates

Remove custom stopwords

Part-of-Speech tagging

I love to send images



I_PRP love_VBP to_TO send_VB
images_NNS

The Process: 2. Feature Extraction

Analyze sentence

Extract raw
app features

Apply POS patterns

The Process: 2. Feature Extraction

Analyze sentence

Extract raw
app features

Apply POS patterns

verbs

nouns

Send and receive images, audio, and messages

punctuation and conjunctions

The Process: 2. Feature Extraction

Analyze sentence

Extract raw
app features

Apply POS patterns

- **Case 1:** “send and receive attachments”.
- **Case 2:** “View documents, PDFs, photos, and videos”.
- **Case 3:** “Discuss and annotate notes and drafts”.
- **Case 4:** “Use camera capture to easily scan and comment on pieces of paper, including printed documents, business cards, handwriting and sketches”.
- **Case 5:** “Write, collect and capture ideas as searchable notes, notebooks, checklists and to-do lists”.

App features:

Discuss notes, discuss drafts

Annotate notes, annotate drafts”.

The Process: 2. Feature Extraction

Analyze sentence

Extract raw
app features

Apply POS patterns

Examples:

POS Pattern	Example app feature
Noun Noun	Group conversations
Verb Noun	Send messages
Adjective Noun	Precise location

The Process: 3. Feature Matching

Single term
matching

Synonym matching

Semantic similarity
matching

The Process: 3. Feature Matching

Single term
matching

Synonym matching

Semantic similarity
matching



Take picture – take picture



capture picture– take picture

The Process: 3. Feature Matching

Single term
matching

Synonym matching

Semantic similarity
matching



Take photo – capture image



Send new email – receive emails

The Process: 3. Feature Matching

Single term
matching

Synonym matching

Semantic similarity
matching

Feature 1	Feature 2	Cosine Similarity
Compose new email	Create new email	0.85
Taking image	Capture image	0.73
Send attachments	Send attached files	0.63
Add image	Delete image	0.38

Threshold was set to 0.7

What is it good for? Field of applications

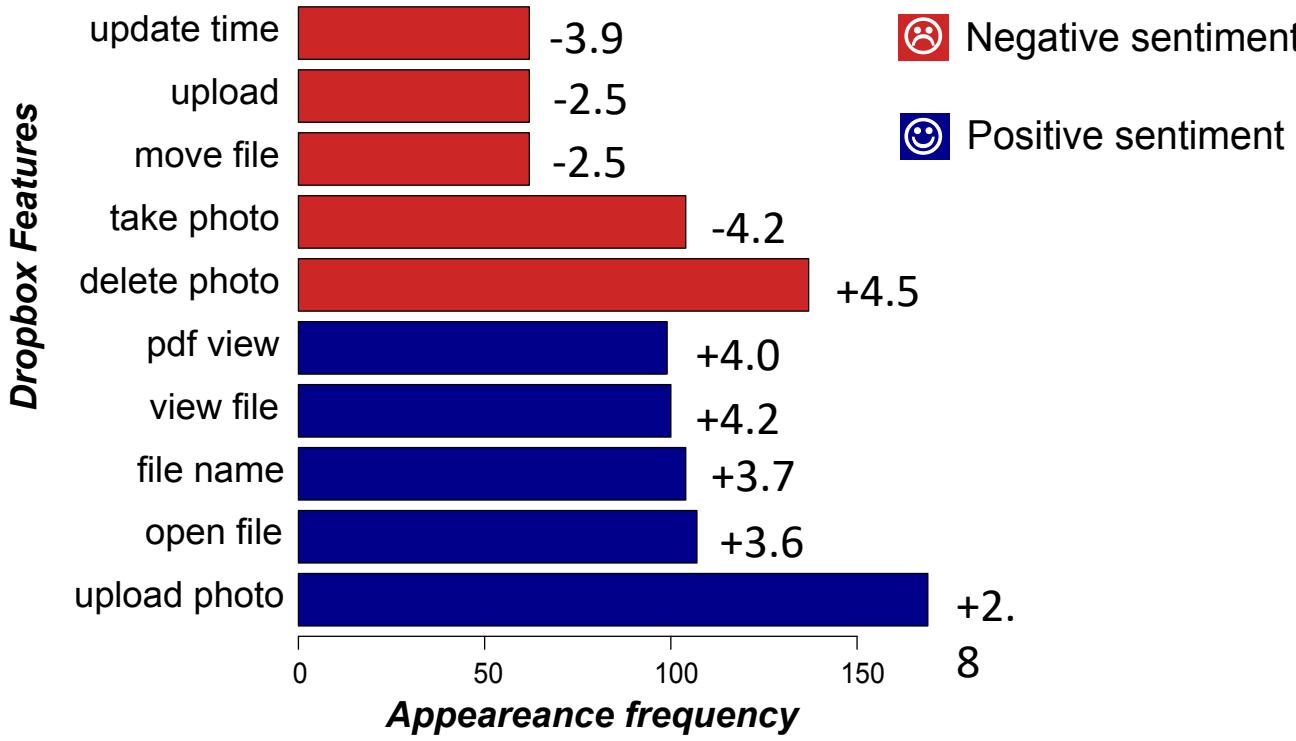
Filter



Combine with
previous study



Advanced analytics



[Guzman and Maalej 2014]

Overview

1

Motivation, definition, and conceptual model

2

Mining requirements-related feedback

3

Identifying and matching app features

4

Summary

Summary

- 01 Explicit user **feedback contains valuable** information for requirements engineering like problem reports and feature requests
- 02 We can identify feedback categories using machine learning. These **algorithms need high quality input**. Use coding guides, pilot studies, and check the intercoder-agreement.
- 03 A **content analysis** can help you to draw **important conclusions** about the data. Stakeholders currently do this manually, but by abstracting your insights you can create automated approaches.
- 04 By **continuously analyzing** user feedback to improve their app, stakeholders can **increase user satisfaction**.

Literature

- Christoph Stanik, Marlo Haering, Walid Maalej: Classifying Multilingual User Feedback using Traditional Machine Learning and Deep Learning. In Proceedings International Requirements Engineering Conference Workshops, 2019.
- Timo Johann, Christoph Stanik, Alireza Alizadeh B., Walid Maalej: SAFE: A Simple Approach for Feature Extraction from App Descriptions and App Reviews. In Proceedings International Requirements Engineering Conference, 2017.
- Walid Maalej, Maleknaz Nayebi, Timo Johann, Gunther Ruhe. Towards Data-Driven Requirements Engineering, in IEEE Software Special Issue on the Future of Software Engineering, 2016.
- Walid Maalej, Hadeer Nabil. Bug Report, Feature Request, or Simply Praise? On Automatically Classifying App Reviews. In Proceedings of the 23rd IEEE International Requirements Engineering Conference, 2015.
- Emitza Guzman, Walid Maalej. How Do Users Like this Feature? A Fine Grained Sentiment Analysis of App Reviews. In Proceedings of the 22nd IEEE International Requirements Engineering Conference, 2014.
- Dennis Pagano, Walid Maalej. User Feedback in the AppStore: An Empirical Study. In Proceedings of the 21st. IEEE International Requirements Engineering Conference, 2013.