Data Labeling Instructions
(Panichella Dataset)

Introduction:

This document is meant to help you understand the dataset and what is expected from you in the data labelling step. Please read the instructions carefully.

Dataset:

You'll find a 'dataset.sql' file that was created for you. This SQL file contains 3439 user reviews collected from 17 different applications with the following metadata:

Column	Explanation
id	The review id (3439 reviews in total)
app_name	The application name (17 different applications in total)
version	The application version. An application may have multiple versions
userid	The id of the user who wrote the review
date	The review date
rating	The review rating (1-5)
title	The review title
text	The review body
Has_user_experience	Binary (0/1) label that you are expected to complete in the dataset labelling step. Please read the instructions in the data labelling section carefully.
Has_feature_request	
Has_bug_report	

Please note that this a real-world dataset so you may have empty values for any of the review columns. Please make sure you handle this properly.

Labels:

As shown above, we have four label columns. In this section, a brief description of each is given:

Has_User_Experience

To set this binary column's value to 1, a review must contain at least a single **sentence** that discusses or expresses emotions towards a specific aspect of the application. They key idea here is that it describes a specific aspect/feature of the app that the user likes/hates, i.e., to capture reviews that tell the developers what features/aspects of the app are liked/hated by their users.

Review	Explanation	Has User Experience?
The **UI** is amazing	The sentence clearly describes a specific aspect of the app, i.e., the UI. In other words, the user is giving us information that he likes the UI or User Interface of the application	Yes
I like the **predictive text**	The expresses that he likes the aspect of the app where 'predictive text' is used	Yes
I love this app	The sentence does NOT describe a specific aspect/feature of the app	No

Has_feature_request column:

To set this binary column's value to 1, a review must contain at least a single sentence expressing ideas, suggestions or needs for improving or enhancing the app or its functionalities.

Review	Explanation
It's a pity it doesn't support Chinese.	The user is asking to add Chinese support
Messing around, wish there was a paint bucket, couldn't fill in face...	The user wants a 'paint bucket' feature added to the app
Facebook Login? Nope. App immediately deleted.	The user wants a 'facebook login' feature added to the app

<u>Has_bug_report column</u>

To set this binary column's value to 1, a review must contain at least a single sentence describing technical issues with the app or unexpected behaviors.

Review	Explanation
The pop-up ads block the actual game area.	The user is describing a technical issue. The ads are blocking the game view
When I put in a photo to edit. I want it to be full screen. After the edit the image just looks weird and small. That's a problem I've had with the app.	The user is describing a technical issue. The edit feature seem to have a bug
Really enjoyed it till IOS8, **can't get it to load properly now**.	The user seem to have problems opening the application after iOS version 8

Data Labelling Instructions:

First, you're expected to read the description of each application, go over their screen images, and make sure you **understand** the application **functionality** and **purpose**. To help you do so, here is a list of all the applications and a link to help better understand what they are all about:

Application	Link (read description and go over screen images)	#
Blinq	https://play.google.com/store/apps/details?id=com.sharemore.cust.blinq	299
CSTP	https://play.google.com/store/apps/details?id=it.cstp.app	35
Doodle Pairs	https://itunes.apple.com/gb/app/doodle-pairs-free/id431135894?mt=8	31
Karaoke – Sing Me	https://play.google.com/store/apps/details?id=com.antosdr.karaoke_free	23
Karaoke – Sing Me (Free)		94
Lifelog	https://play.google.com/store/apps/details?id=com.sonymobile.lifelog	129
Minesweeper Reloaded	https://itunes.apple.com/us/app/minesweeper-reloaded/id477031499?mt=8	69
Movie Creator	https://play.google.com/store/apps/details?id=com.sonymobile.moviecreator.rmm	84
Picturex	https://play.google.com/store/apps/details?id=picturex.Client.Android	120
powernap	https://www.microsoft.com/en-us/p/powernapp/9wzdncrdmxlv?activetab=pivot%3aoverviewtab	1050
Sheep-O-block	https://itunes.apple.com/gb/app/sheep-o-block/id415394498?mt=8	35
sketch_beta	https://play.google.com/store/apps/details?id=com.sonymobile.sketch	130
Stone Flood	https://itunes.apple.com/us/app/stone-flood/id475159342?mt=12	804
trackID_beta	App shut down. Here's a community page to help understand what it was: https://plus.google.com/communities/115935263324989133575	131
video_beta	App shut down. The is simply a video editor.	343
Weight Track	https://itunes.apple.com/us/app/monitor-your-weight/id413313086?mt=8	20
wififiletransfer_free	https://play.google.com/store/apps/details?id=com.smarterdroid.wififiletransfer	42

Second, you're expected to go over each review and provide a 0/1 value for each of the three label columns. For example, given the following review:

id	30
app_name	Doodle Pairs
version	1.50
userid	Teacher K-12 Gifted
date	2012-01-17
rating	1
title	Distracting ads
text	Used to be an excellent game, but ads ruin the free version.

You're expected to provide the following labels:

Has_information_giving	1
Has_feature_request	0
Has_bug_report	0

Why were those labels selected?

Well, the user is providing information on a specific aspect of the application, i.e., not liking the ads (user experience). The review also does not have any feature requests and no application technical problems (no feature request and no bug report).

Important note:

Please be aware that those label columns are inclusive. This means that a review can be labelled as 'has_information_giving" and as "has_feature_request" at the same time as a review may consist of several sentences each may fall under a different label. You're not supposed to pick one label per review, but rather analyze each sentence and find all the labels that apply to the given review.