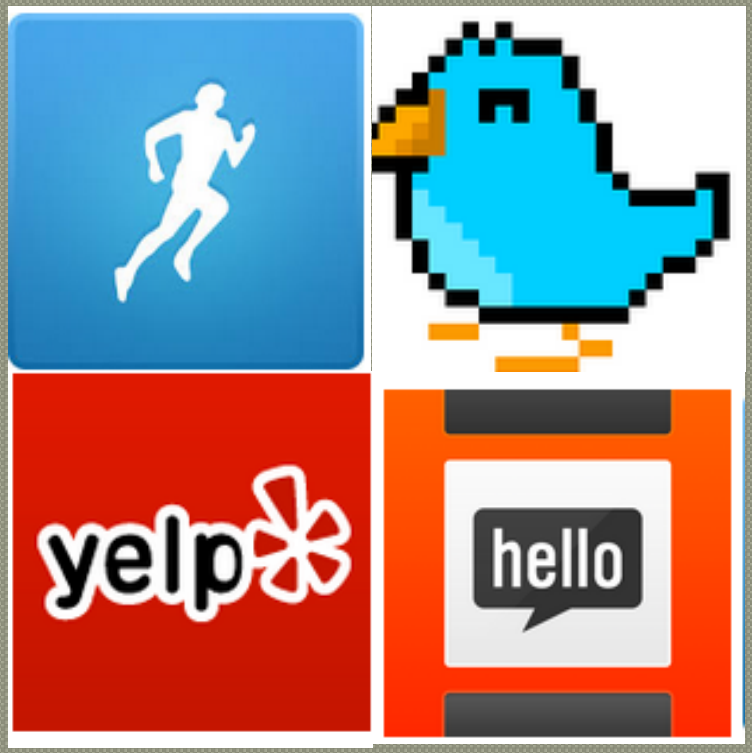
Pebble App Search for Pebble Technology

Swati Jain



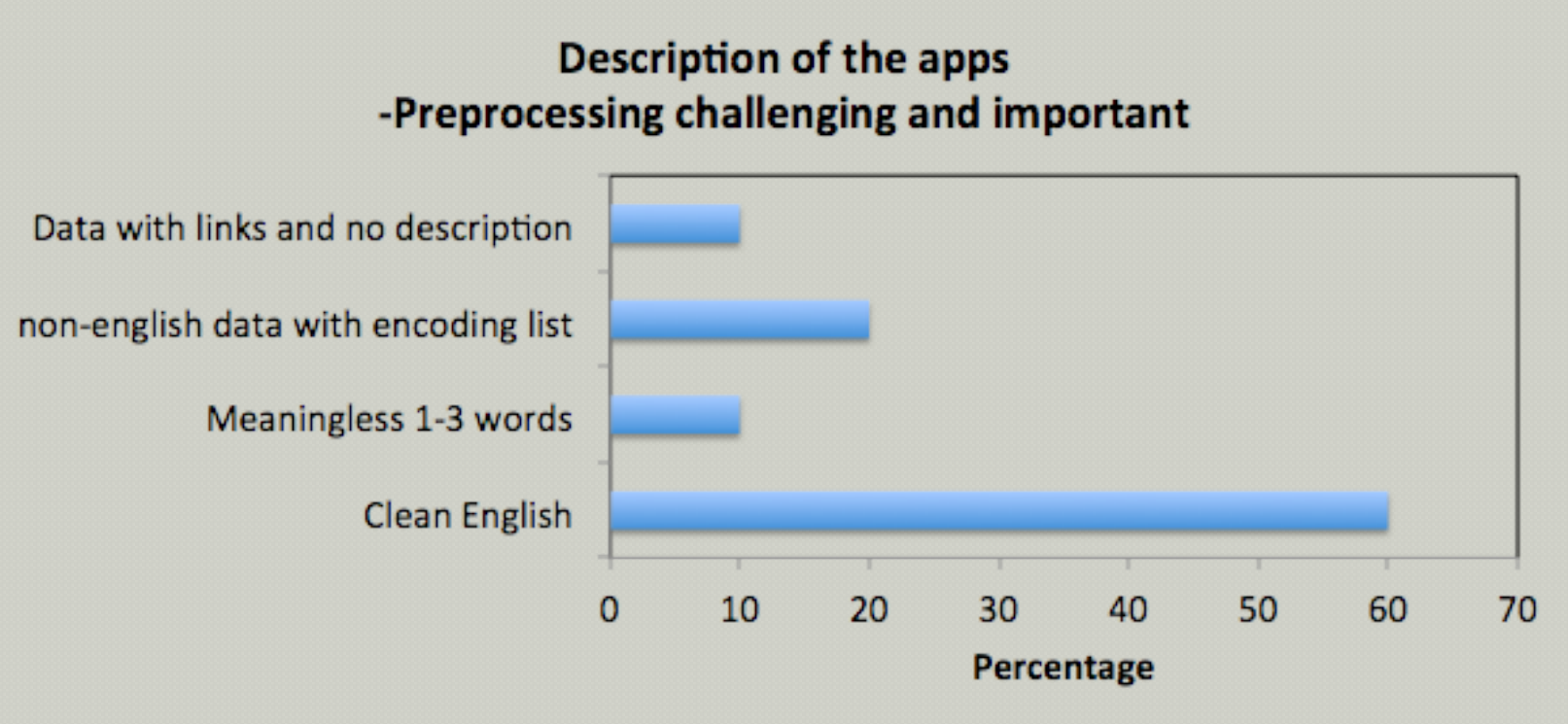
Problem to solve: Apps are not searchable

- 8000 apps in the Database
- No useful preset categories: *Tools & Utilities, Get Some Apps, Daily, Index, Games*

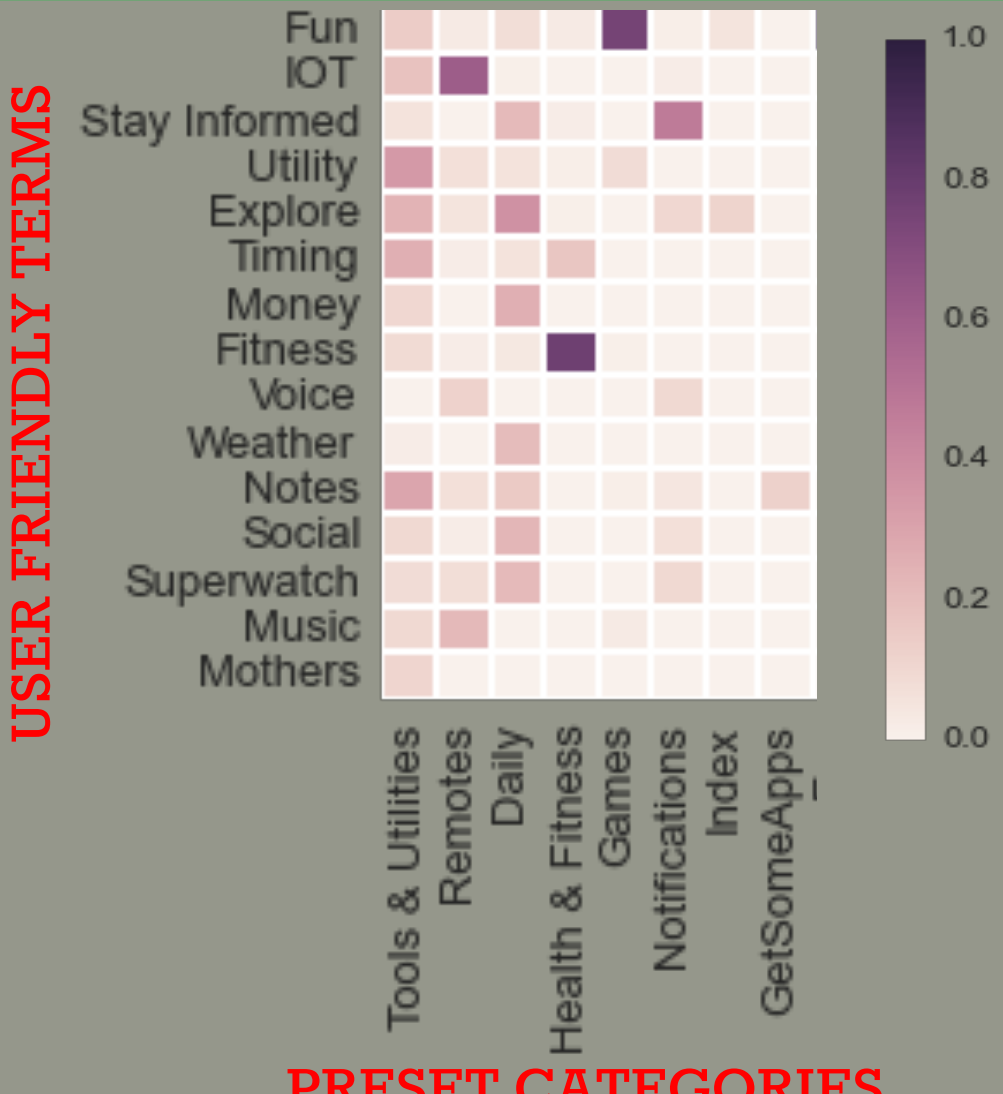
Data Story

8000 apps

- Preset Categories
- Description



Correlation between “user-friendly” and Preset categories



Functionality 1: App Browsing by categories and “Other similar apps”

8000 Apps

Pre-label 10 % apps: 15 categories and 72 subcategories

Divide into test and train

Bootstrap samples

Natural Language processing -- Translation, Tokenization, TFIDF

Ensemble of Machine Learning Models

Precision

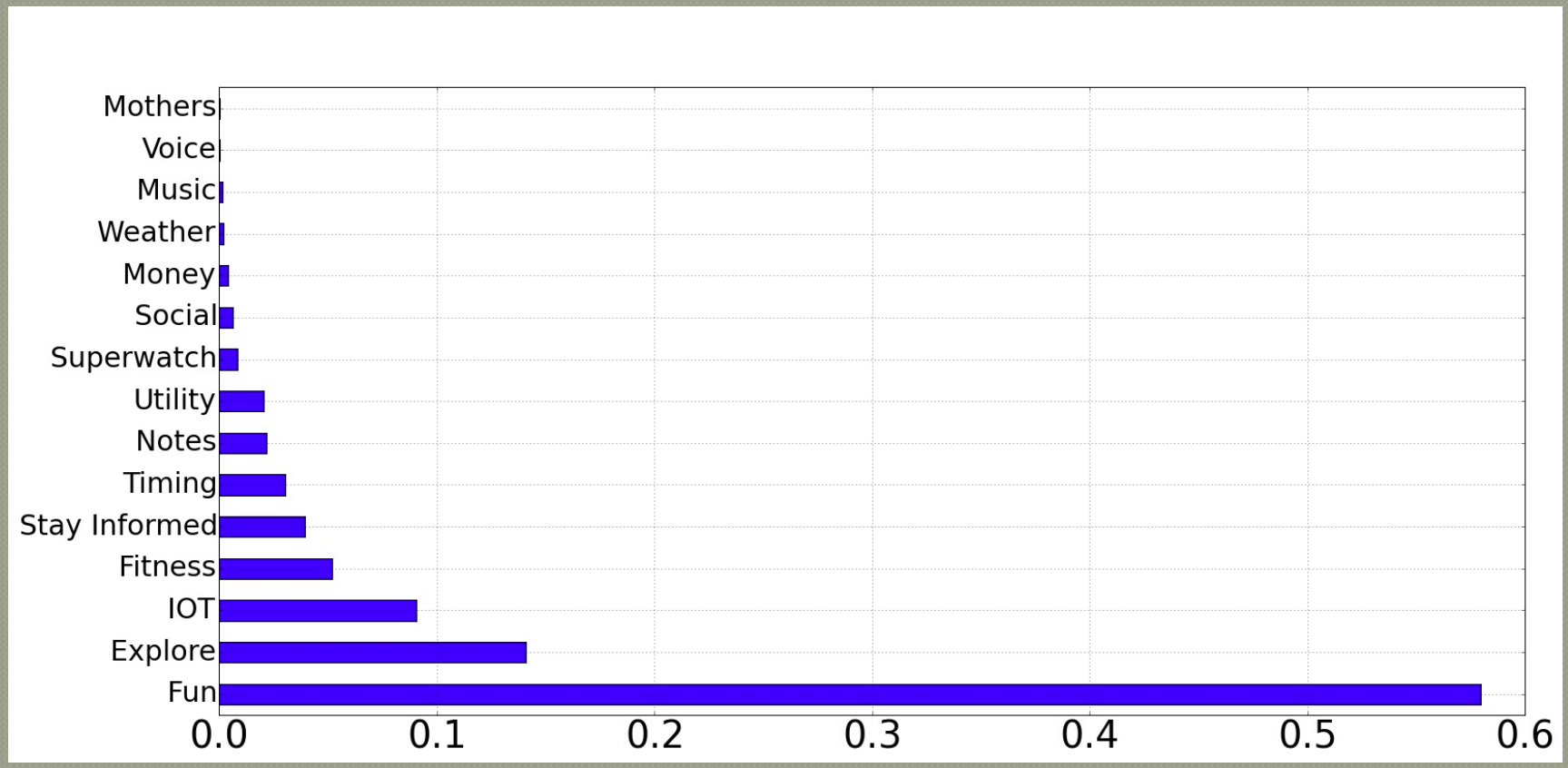
--- Categories: 62 % (random : 6 %)

--- Subcategories: 44 % (random: 1.3 %)

python™
Natural Language Analyses with NLTK

scikit learn

8000 apps now automatically categorized



Functionality 2: Text Search

- Unsupervised classification method used: Similarity between new text and corpus
  - Translation with google API
  - Tokenization with regular expressions
  - Stemming
  - TF-IDF
  - Cosine Similarity between sparse matrices generated from TF-IDF

SEARCH : SMS without taking out your phone
Pebble Communicator
Take Photo
Pebble Messenger
Pebble Tools (SMS)

Value Added for Business

- App browsing and search
  - Improves user experience
  - Increases user retention
- Allows for scalability for the growing business
- Categorization helps to develop business insights
- Brings it at the same level as the competitors

Swati Jain, PhD

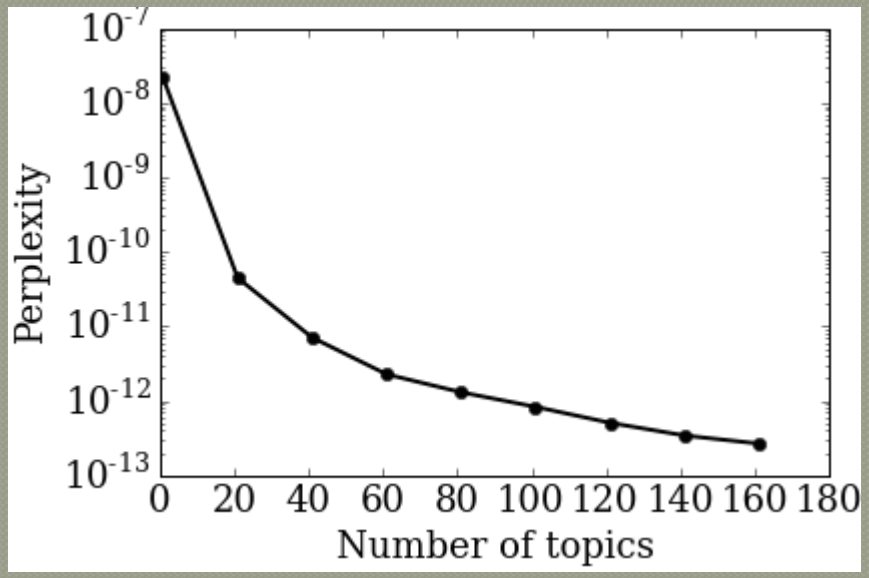
DNV·GL

-- Analytics and predictive
Data Science for Risk Assessment

UN Volunteers
inspiration in action

Functionality 2: App Search

- Method 2: LDA
  - Translation with google API
  - Tokenization with regular expressions
  - Stemming
  - Topic classification using LDA
  - Cosine Similarity between Text
- Not as good performance as TFIDF
- But combined with TFIDF to rank apps



Data Story

8000 apps

- Preset Categories
- Description

Data Type	Example of Description	% of data in this category
Clean English	*Pebble IR blaster*: Allows the Pebble to use your phone's IR blaster to turn on the TV, switch channels, or even change the colors of LED lights.	~ 60 %
Meaningless 1-3 words	*Nanga: Nanga*	~10 %
Data from other languages with encoding lost	*Twistoast: Twistoast est une application ri©servi©e aux utilisateur*	~20 %
Data requiring more processing	*Title:* Tile Thanks to http://www.watchface-generator.de/	~10%

Functionality 1: App Browsing by categories and “Other similar apps”

- 9 % of the data manually labeled into 15 categories and 72 subcategories
- Preprocessing:
  - Translation with google API
  - Tokenization with regular expressions
  - Stemming
  - TFIDF
- Machine Learning:
  - Bootstrapping
  - Several models tested
  - Ensemble of Logistic Regression models used

Predicted Feature	Algorithm Precision on Test Data (20 % of the prelabeled data)
Category (15 categories)	0.62 (Random Precision: 0.06)
Subcategory (72 subcategories)	0.44 (Random precision: 0.013)