

# MBTI-16 Personality Analysis using Emojis

By Wenshuo Qin, You Peng, and Rupei Huang



## Introduction

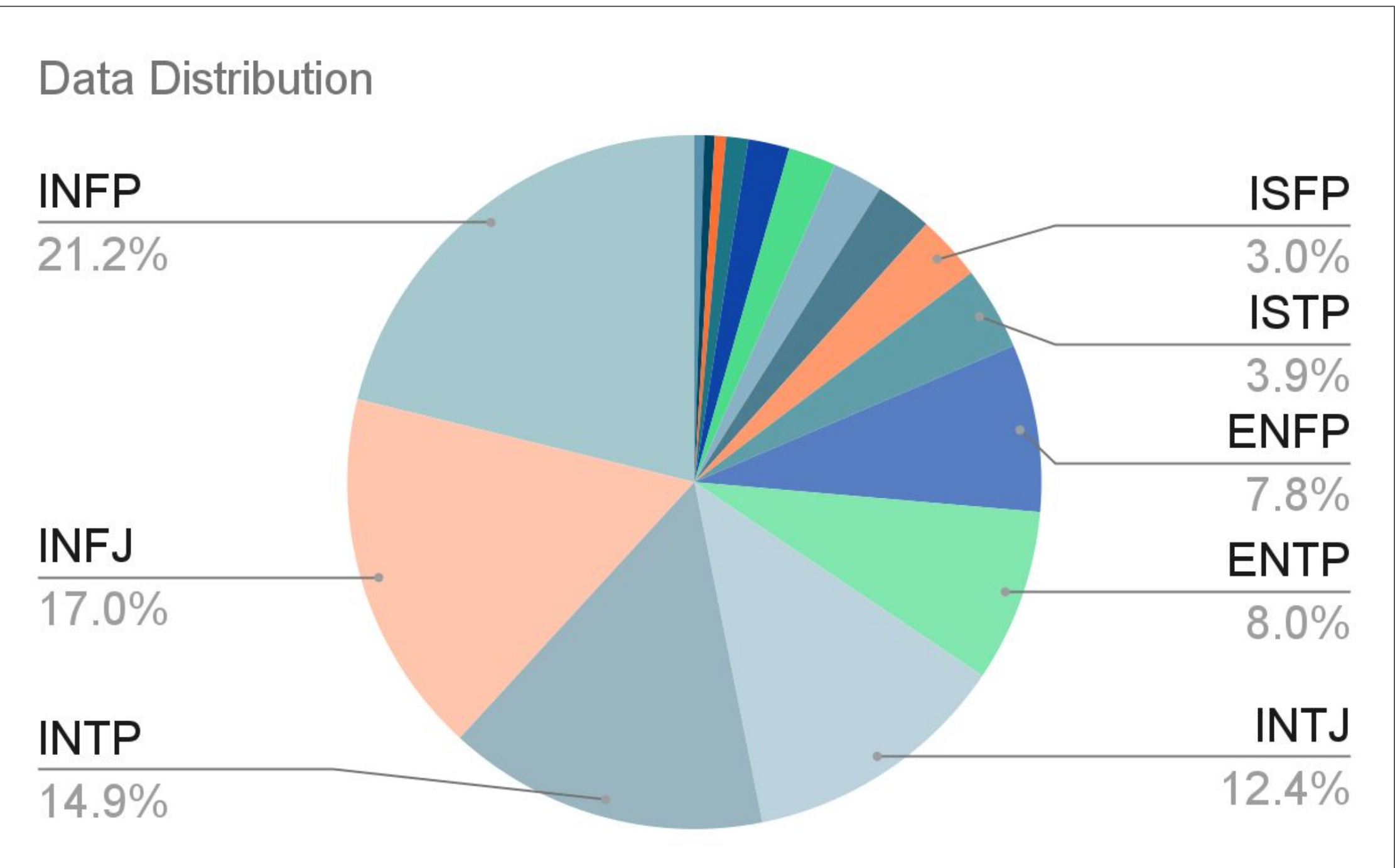
Source of Energy	Processing of Information	Decision Making	Need for Structure
<b>E</b> xtroverts	<b>S</b> ensors	<b>F</b> eelers	<b>P</b> erceivers
<b>I</b> ntroverts	<b>N</b> tuitives	<b>T</b> hinkers	<b>J</b> udgers

- NLP has enhanced our understanding of personality
    - Can evaluate the reliability of psychometric measures
    - No one has studied if Emojis are indicators of personality
- Goal: To determine whether emojis appeared in online communications are reliable predictors of one's personality.**

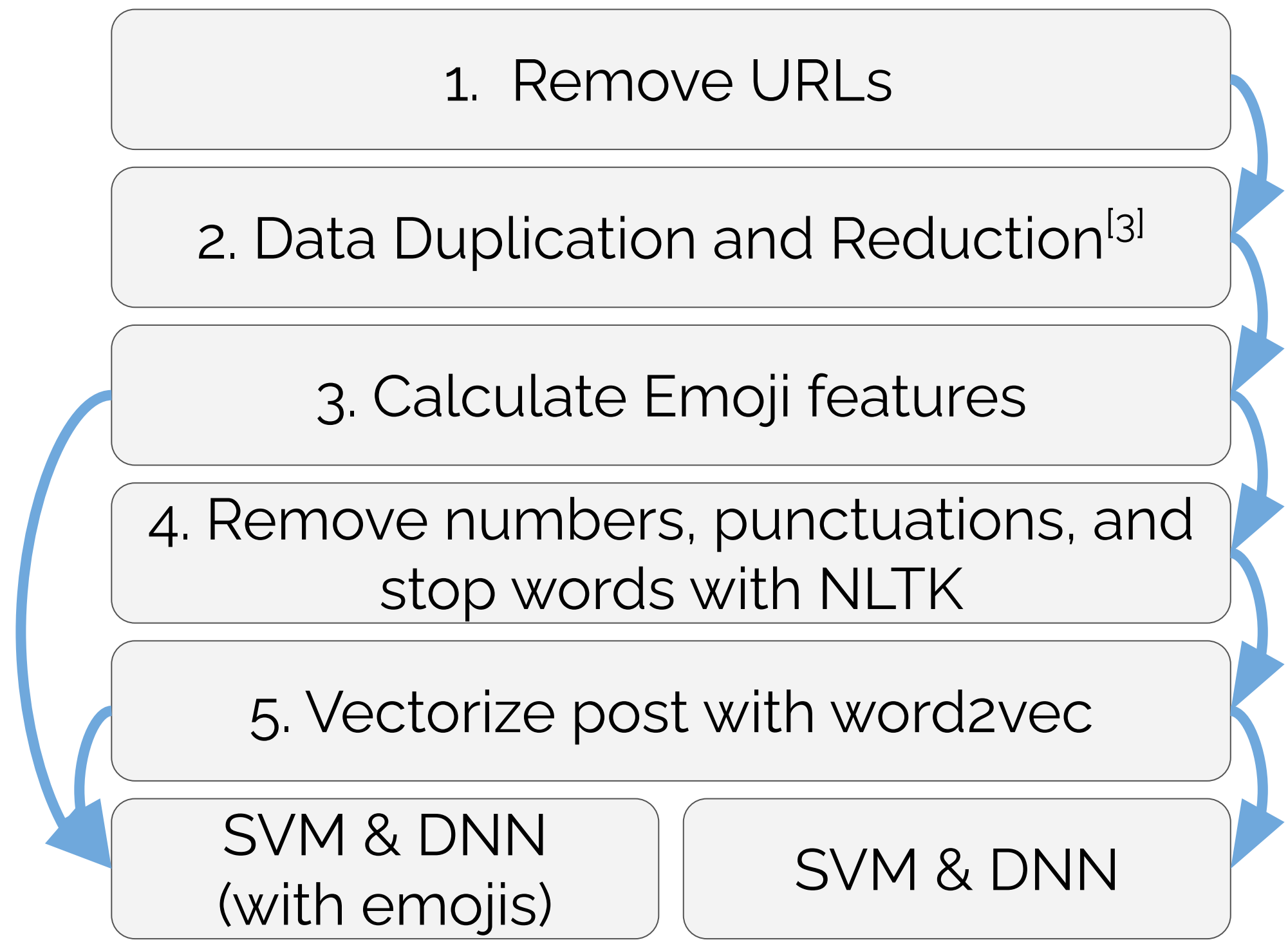
## Data

- MBTI-16 dataset from Kaggle<sup>[1]</sup>
  - Online posts (50 posts from 6800 users, 422,845 total labeled data) with personality labels
  - used data duplication and reduction to balance the dataset for 16 and binary classification
- Emoji dataset from Kralj Novak<sup>[2]</sup>
  - 1000 Emojis with their unicode encoding, positivity, and textual description
- Self-built Emoticon dataset
  - Web scraped data from Wiki, Emoticons, Lifewire
  - 1000 Emoticons with their similar emoji and textual description

\*MBTI dataset distribution



- Data labeling
  - 16 class: 0 to 15 binary classification: 0/1
- Data Preprocessing:
  - SVM and DNN:



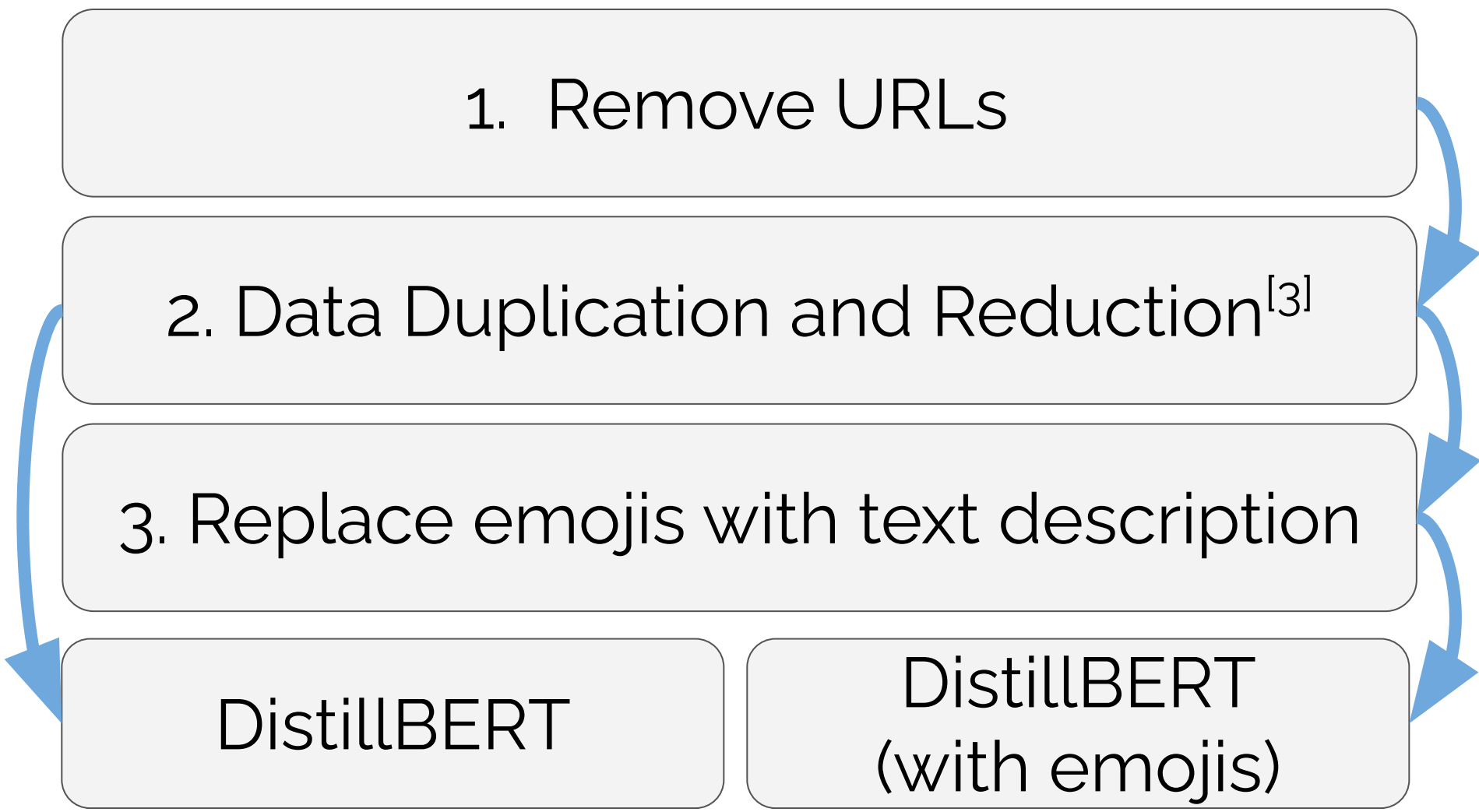
\*Emoji dataset example:

Emoji	Positivity	Unicode name
😊	0.704	SMILING FACE
😭	0.3431	CRYING FACE
👌	0.6574	OK HAND SIGN
💕	0.6742	TWO HEARTS

\*Emoticon dataset example:

Emoticon	Similar Emojis	Unicode name
:~)	😊😊😊😊	Smiley, happy...
-D	😂😂😂😂	Laughing
:-)	😞😞😞😞	Frown, Sad...

- BERT:



[1] [MBTI dataset](#)  
[2] [Emoji Sentiment Dataset from Kralj Kovak](#)  
[3] For both 16 class and binary class datasets, see conclusion for limitation in this method.

## Method

- Feature extraction
    - Emoji
      - extracting count, position (character position) and positivity as 3 features for SVM and DNN
        - the default feature values for the posts that do not have emojis are: 0s
      - generate text description of emojis for BERT
- SVM and DNN emoji features example:

Count	Position	Positivity
1	28/28	0.36

BERT emoji description example:

<CLS> What the hell are you doing! 😡 <SEP>  
<CLS> What the hell are you doing! <negative emoji> <SEP>

- Data Partition:
  - 85% for training 15% for testing
- Proposed Models
  - SVM, DNN, DistilBERT
  - We build two models for each approach: one use emoji info, and one does not use emoji info

## Evaluation

	16 class acc	E vs I	J vs P	S vs N	T vs F
random baseline	6%	50%	50%	50%	50%
SVM	9%	55%	53%	50%	58%
SVM with emojis	9%	55%	54%	52%	57%
DNN	9%	55%	55%	55%	59%
DNN with emojis	9%	55%	55%	55%	58%
DistillBERT	27%	62%	60%	60%	61%
DistillBERT with emojis	27%	63%	60%	61%	63%

## Conclusion

- Limitations:
  - Because of imbalanced data, we have to use data duplication to augment insufficient class examples, which in turns affect our validation.
  - Occurrences of emojis are too few in the dataset that they barely made any impact in prediction
  - Personality labels are self-reported and biased
  - Data are limited to Personality Cafe
  - MBTI remains debatable
- Future works
  - We can use similarly method to test what are important predictors of personality