­­av

**Levenshtein**

**and Soundex**

**on original word**

Not

Stop-words Removal

If found

If found

Discard word

Not

If not found

Add initial word to corpus

To Classification Model

If found

Discard word

Discard word

If found

Not

If found

Stemmed word

Stemming

Not

Not

If found

Tokenized word

Tokenization

Polarity Assignment

Text transliteration

Data Collection