



Communication Systems Filtering using NLP

Hafsa Traore, Madison Lewis, Ahmed Abdelgalil



What is Natural Language Processing (NLP)?

- Branch of Artificial Intelligence (AI) that uses machine learning to process and interpret text and data.
- Allows computers to understand, interpret, and manipulate human language.
- Examples :
 - Smart assistants (Siri, Alexa, etc.)
 - Spell check
 - Voice text messaging
 - Autocomplete
- Use cases:
 - Improve user experience
 - Ex: autocorrect
 - Automate support
 - Ex: chatbots
 - Monitor & analyze feedback
 - Ex: Analyze user reviews



Communication System Prioritization

- Categorize **priority** of the following
 - Email
 - Text
- Current communication systems:
 - Normally filtered by sender
- NLP Systems provides:
 - Efficient labeling for High Priority/ Spam Emails & Text
- Outcomes:
 - High Priority → Quick Response time
 - Quick Response time → Satisfied Users

The Importance of Email Response Times

In Google consumer research conducted by Arise (2019), consumers were asked how quickly they expected a response to their email:



However, research carried out by Drift (2017) showed:



Companies contacting potential customers within 5 minutes yielded 900% more interest than those taking 10 minutes (Insidesales, 2016)

Priority Filtering

Use filtering process & NLP for determination of level of priority:

- **Important** (High priority)
- **Neutral**
- **Spam**

High Priority

- Text requiring immediate attention from actual people/peers
 - Domain names
 - Signature

Spam

- This is taken into consideration by using NLP for model and training

Neutral Priority

- Considered Neutral if not classified as high priority or spam

Implementation (Overview)

Step 1: Data Preprocessing



Step 2: Removal of “Stop Words”



Step 3: Tokenization



Step 4: Lemmatization

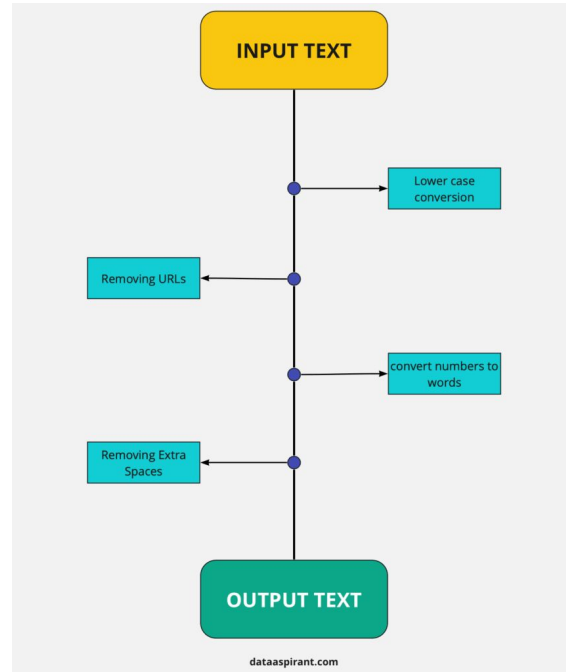


Step 5: Classification

Step 1: Data Preprocessing

- After the text is received via email/SMS, in this stage it is prepared for model building. Some examples of this preprocessing include removing:

- **Whitespace between text**
- **Hyperlinks**



Step 2: Removal of “Stop Words”

- After the text received is broken down into its smallest unit, unnecessary data is once again removed.
 - helps with efficiency of the model by reducing time spent and program size.

- Unwanted Characters:

- And
- Though
- So
- What

Sample text with Stop Words	Without Stop Words
GeeksforGeeks – A Computer Science Portal for Geeks	GeeksforGeeks , Computer Science, Portal ,Geeks
Can listening be exhausting?	Listening, Exhausting
I like reading, so I read	Like, Reading, read

Step 3: Tokenization

- This is the process of taking the processed data (text) and breaking it into its smallest unit.
 - **Ex: sentences are broken down into words. Words are broken down into prefixes and suffixes.**

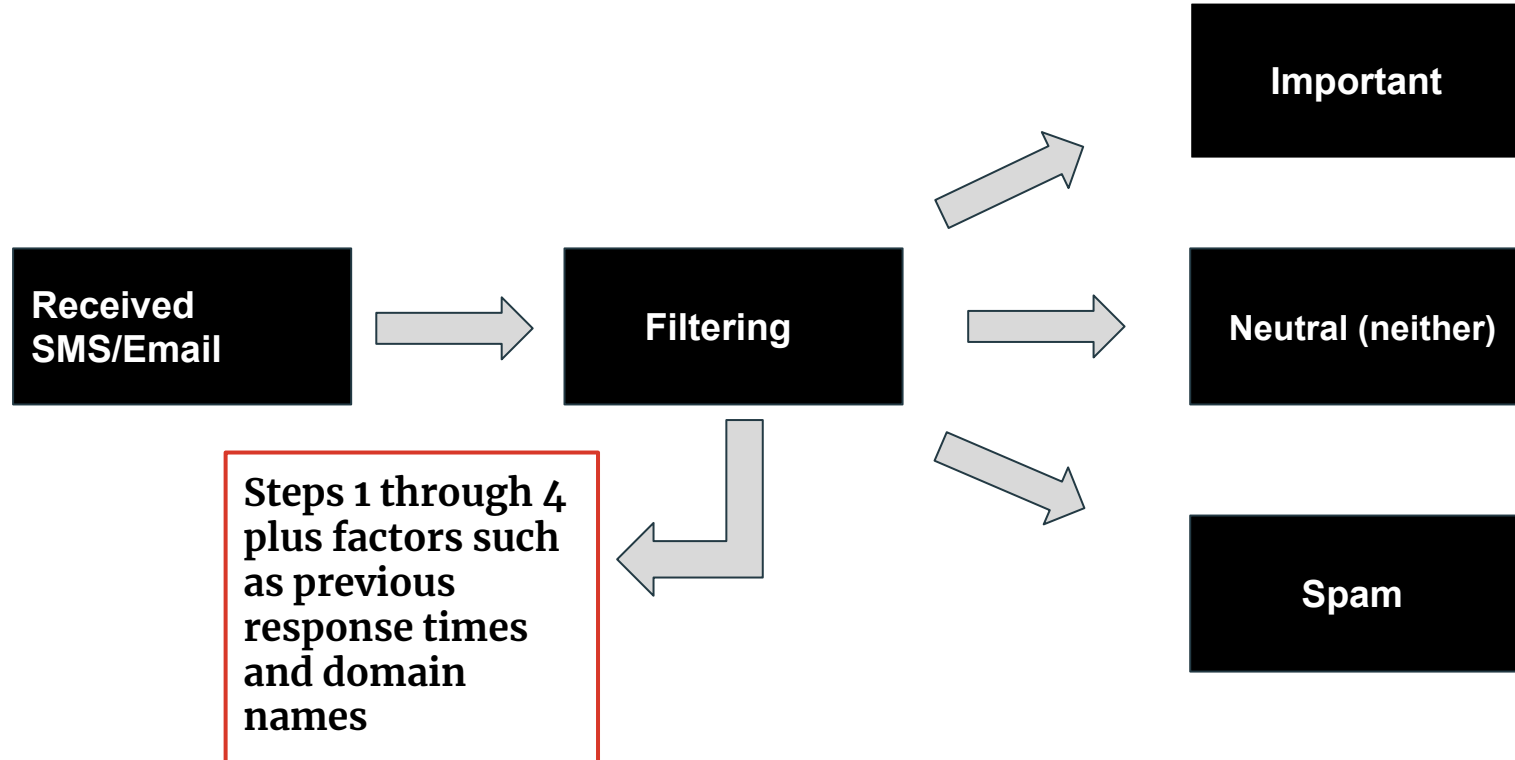
Q Search this file...			
1	Form	Suffix	Stem
2	running	-ing	run
3	runs	-s	run
4	consolidate	-ate	consolid
5	consolidated	-ated	consolid

Step 4: Lemmatization

- In this stage, the mode performs a text-preprocessing technique. This technique allows the NLP model to sort words by grouping variant forms of the same word.
 - **Ex: gone, going, and went are all mapped to go**

Q Search this file...			
1	Form	Morphological Information	Lemma
2	studies	Present tense of the word study	study
3	ran	Past tense of the word run	run

Step 5: Classification



Questions?