

# SENTIMENT ANALYSIS - CLASSIFICATION OF TEXT BASED ON KEYWORDS

## EXECUTIVE SUMMARY

*Industry - Logistics*

*Customer - Largest logistics company that provides courier delivery services*

*Segment - Sentiment Analysis*

*Services - Clustering text based on keywords with Sentiment Analysis*

*Technologies: Python, NLP, ML - K-means, SG rank, Fuzzy Logic*

*“Analyzed customer reviews, ratings and feedbacks to assist*

*Businesses frequently use sentiment analysis on textual data to track the perception of their brands and products in customer reviews and to better understand their target market.*

*We are helping the logistics company business in extracting data from customer review comments, feedbacks and ratings with populating graphs based on sentiment to reflect the good and bad points about a different categories like Delivery related, Delivery person related, Late delivery, Call/customer care related and product related.*

*For Example:*

*1) Customer Review Text: I liked the way of Delivery*

*Results: Delivery related, positive*

*2) Customer Review Text: Received product at wrong address*

*Results: Address-related, negative*

---

*This usecase relates to the extraction of necessary data from review texts to assist companies in making better business decisions by optimize the processes.*

---

## SOLUTION SUMMARY

*Text processing part is the process of extracting relevant information from unstructured text. The undertaken data is based with customer reviews & feedbacks. The customer reviews are processed using text processing technology.*

*Using the processed text, we have performed two major tasks:*

- Clustering based on specific keywords from the given range of keywords & reviews to identify the keywords belongs to which class of keywords*
- Sentiment Analysis is drawn using algorithm to understand the positive, Negative OR Neutral Sentiments*

*Finally, the outputs are merged & saved in Database for viewing the Sentiment, keywords & further used for building dashboards.*

## Additional Challenges and Solutions:

### Step 1 - Challenges:

- Out-of-Vocabulary & Misspelled words
- Presence of unusual Indian chat shortcuts and Grammatical mistakes
- Presence links in text, digits, whitespaces, links, emails

### Solution:

- Hybrid usage contextual Spell Check & Auto Spell check
- Merging a Jaccard distance & Edit distance Method can give more accuracy
- Categorizing the data using fuzzy logic, fine tuning the data for the better outcome of the model for the larger set of data

### Step 2 - Solution Implementation:

- The text based challenges were handled as a part of the data preparation as explained above.
- Preparing for a hybrid usage with K-Means, SGRank, YAKE & Text Rank to explore the more topics underneath the data.
- Classification of the refined data to groups of keywords like Intent & labels for extracting a sentiment.
- Checking the intensity & polarity while considering the value returned by each approach, a threshold values is generated based on the threshold value the intensity and polarity.

## BENEFITS

- The below figure shows the distribution of good, neutral, and negative reviews by category.
- Each category is extracted directly from the text of the customer reviews.
- This helps the logistic company in identifying their weak points and development opportunities.
- Based on the output graphs, the logistic company was able to recognize the strengths and weaknesses with respect to each category and will utilize this information to grow their business.

Fig: Category wise Total Counts.

