

Lazada

Product Title

Quality Challenge

Data Analytics Squad

Akshay Sharma

Ashish Shah

Ayush Dattagupta

Nikhil Thakur

Rahul Dhavalikar





Problem Statement

- In this **Product Title Quality Challenge**, we are provided with a set of **product titles**, description, and attributes like price, country etc, together with the associated title quality scores - **clarity** and **conciseness**.
- Our task is to build a product title quality model that can automatically grade the **clarity** and the **conciseness** of a product title.
- **Clarity**: The product title is clear if within five seconds one can understand the title, what the product is, and quickly figure out the key attributes (color, size, model, ...).
- **Conciseness**: The product title is concise if it is short enough to contain all the necessary information. Otherwise, i.e., the title is too long with many unnecessary words, Or it is too short such that it is unsure what the product is.



Examples

- hot red clutch rug sack travel backpack unisex cheap with free gift
 - Red unisex travel rucksack backpack
- 1 Pair of Unisex Touch Screen Sensitive Gloves Knitted Winter Warm Christmas Glove Red
 - Touchscreen sensitive knitted gloves red
- Hot Tom Clovers Womens Mens Classy Look Cool Simple Style Casual Canvas Crossbody Messenger Bag Handbag Fashion Bag Tote Handbag Gray
 - Tom Clovers canvas crossbody messenger bag gray



Challenges

- Clarity and conciseness are subjective
- Limited available dataset for training
- No labels for validation and testing datasets
- Imbalanced dataset with 94% positive examples and 6% negative examples for Clarity



Our Approach

- Different approaches for Clarity & Conciseness models
- Preprocessing, formatting of data
- Feature extraction using NLP techniques
- Deep learning for conciseness
- Undersampling method for skewed datasets
- Boosting, random forests for clarity
- Grouping of data
- Evaluation and optimization on different performance measures - Precision, Recall, F1 score, Area under ROC, PR curve
- K-fold cross validation



Highlights

- Analysis and comparison of different models and methods
- Analysis of feature engineering
- Tackling skewed datasets



Thank You

Data Analytics Squad