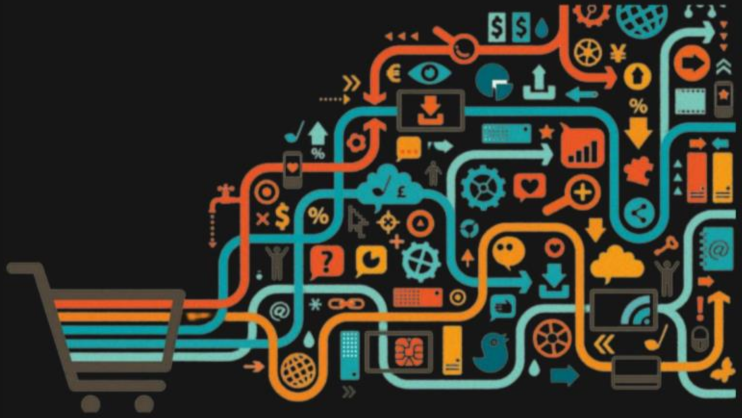
Corso di Laurea Magistrale in Data Science and Economics

UNIVERSITÀ DEGLI STUDI DI MILANO
FACOLTÀ DI SCIENZE POLITICHE, ECONOMICHE E SOCIALI



E-commerce Insights: Analyzing Customer Reviews through LDA Topic Modeling and Association Rules

Luca Bertoletti

Anno Accademico 2022 / 2023

TABLE OF CONTENTS

01 INTRODUCTION

02 OBJECTIVE

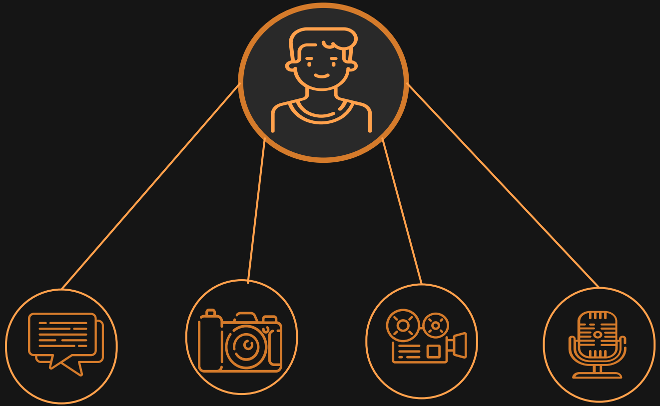
03 METHODOLOGY

04 RESULTS

05 CONCLUSIONS

01 INTRODUCTION

Nowadays, the diffusion of user-generated content has become a defining characteristic of online interactions.



Users can easily share their opinions and experiences, reaching global audiences instantaneously.

App stores like Google Play or Apple AppStore have emerged as important sources of user reviews

02 OBJECTIVE

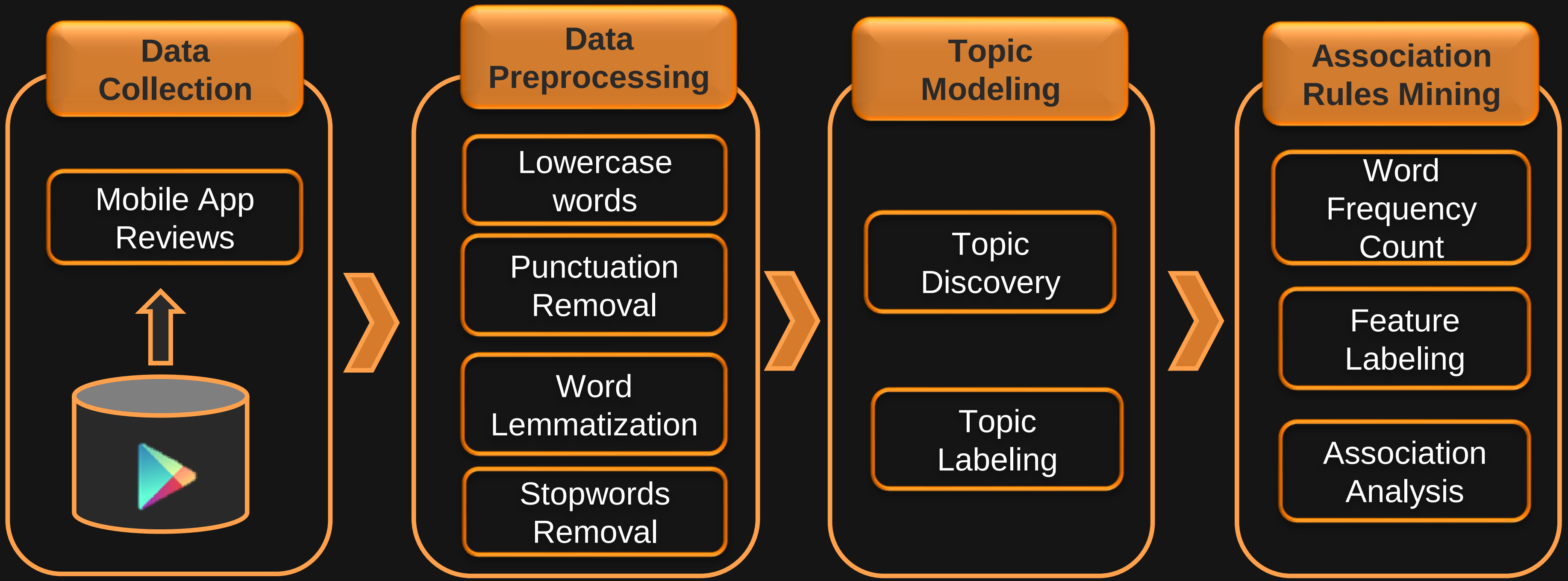
Uncover valuable insights hidden within user reviews, with a particular focus on those from the Chinese e-commerce app Temu.

TEMU

Two main questions are addressed:

- What are the main topics discussed in the user reviews?
- Which features do users view positively and which negatively?

03 METHODOLOGY



DATA COLLECTION

A total of 174,554 reviews from the Temu US market, covering the period from September 2022 to November 2023, are scraped.

- **Review ID**, the unique identifier of the review.
- **Username**, the name of the user who has written the review.
- **User Image**, the profile pic of the user.
- **Content**, the text of the review.
- **Score**, the score (1 to 5) of the review.
- **Thumbs Up Count**, number of thumbs up received by the review
- **Review Created Version**, app version when the review was written.
- **At**, date of the review.
- **Reply Content**, the text written by Temu in response to a review.
- **Replied At**, date of the Temu's reply.
- **App Version**, version of the app.

DATA PREPROCESSING

FROM THIS

[‘Temu is a mostly good and intuitive app. Prices and items are surprisingly good. It's been great so far with the many orders that I've placed and the shipping has always been within the estimates for arrival! However, I am deducting a star because they got rid of the wishlist without explaining why, so the only way to save items is to add them to your cart. That's disappointing.’]

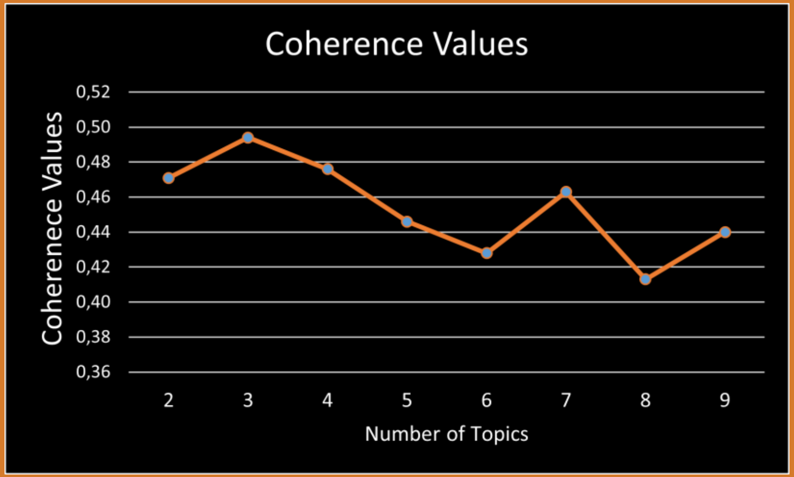
TO THIS

['temu', 'good', 'intuitive', 'app', 'price','item','surprisingly', 'good', 'great', 'far', 'many', 'order', 'place', 'shipping', 'within', 'estimate', 'arrival','however', 'deduct', 'star','get','rid', 'wishlist', 'without', 'explain', 'way', 'save', 'item','add','cart', 'disappointing']

TOPIC MODELING

Topics in the reviews are identified via Latent Dirichlet Allocation.
LDA can find the relationships between words, then allocate them inside the corresponding topics.

However, LDA cannot automatically suggest the number of topics within the reviews.



Coherence is a metric used to measure how well-defined and meaningful the topics are.

ASSOCIATION RULES MINING

REVIEW EVALUATION	PROCESSED CONTENT	FEATURES
positive	[temu, good, intuitive, app, price, item, surprise, …]	[Product, Shopping, Advertising]
negative	[pretty, app, sale, tactic, illegal, …]	[Product, Shopping]
…	…	…

Reviews are labeled with the corresponding e-commerce feature and association rules, through the A-Priori algorithm, are generated to uncover the relationship between these features and the reviews' evaluation.

04 LDA RESULTS

The chosen LDA model has 4 topics with the second highest coherence score of 0.4765.
Each topic is then named by looking at what keywords it is composed of.

TOPIC NUMBER	TOPIC NAME	KEYWORDS
1	Security concerns	spyware, malware, intrusive, data, steal, fraud...
2	Product quality	quality, merchandise, item, fit, size...
3	Shopping Experience	app, deal, service, fast, delivery...
4	Advertising	advertising, pyramid, real, false, scammer...

04

ASSOCIATION RULES RESULTS

Only the association rules with a lift greater than 1 were considered.
A rule with a lift greater than 1 indicates that the antecedent has a significant positive impact on the occurrence of the consequent.

LHS	RHS	SUPPORT	CONFIDENCE	LIFT
Security, Advertising	Negative	0.002	0.883	4.599
Security, Shopping, Advertising	Negative	0.013	0.839	4.370
Security	Negative	0.010	0.834	4.347
Security, Shopping	Negative	0.019	0.726	3.784
Product, Security, Shopping, Advertising	Negative	0.023	0.686	3.576
Product, Security	Negative	0.001	0.618	3.220
Product, Shopping	Positive	0.321	0.918	1.136
Product	Positive	0.059	0.900	1.114
Shopping	Positive	0.277	0.886	1.096

05 CONCLUSIONS

What are the main topics discussed in the user reviews?

- Security Concerns
- Product Quality
- Shopping Experience
- Advertising

Which features do users view positively and which negatively?

- Reviews revolving around Product Quality and Shopping Experience tend to evoke positive feedback from users.
- Security-related topics tend to evoke negative sentiments among users

THANK YOU
FOR THE
ATTENTION!