2026 경영정보 관련 학회 춘계통합학술대회

Predicting and Explaining Review Helpfulness in E-Commerce Platforms: A Multimodal Approach to Text and Visual Characteristics

Camila Estefania Painefilo Hermosilla [a], *YoungKi Park* [b], *and Taeho Hong* [c]

[a] *School of Business, Pusan National University*
[b] *School of Business, George Washington University*
[c] *School of Business, Pusan National University*

June 4 (Thu.) – June 5 (Fri.), 2026

Content

1. Introduction – Research Background

Rise of Multimodal Reviews:
- Online reviews increasingly combine written comments, ratings, and user-generated images (Jiang & Benbasat, 2007)
- Images can show appearance, defects, size, or usage context
- This makes review helpfulness a multimodal evaluation problem.

Visual Information Beyond Image Presence:
- Prior research has shown that review images can increase perceived diagnosticity and helpfulness (Ma et al., 2018; Kubler et al., 2024)
- However, visual information has often been represented only as **image presence**
- Recent work suggests that **visual quality** matters for review helpfulness (Han et al., 2025)
    - **Aesthetic quality** (MUSIQ): perceptual appeal, composition (Ke et al., 2021)
    - **Technical quality** (BRISQUE): blur, noise, sharpness distortions (Mittal et al., 2012)

Product Type Differences:
- The value of review cues may differ depending on product type (Nelson, 1974)
    - **Search goods**: can be evaluated before purchase through objective attributes
    - **Experience goods**: require use or consumption before quality can be fully assessed ➢ Visual quality may provide more value when consumers need help imagining product use

Methodological Gap:
- Predictive models identify whether features improve classification
- But they do not explain whether features become part of the **sufficient combinations** that lead to high helpfulness (Ragin, 2008; Fiss, 2011)

I. Introduction

Empirical Context:

- MercadoLibre Chile: **major e-commerce platform in Chile** and Latin America
- 80,145 reviews from 593 products (search and experience goods)
- Supports multimodal reviews combining textual and visual information since 2023

Research Questions:

- **RQ1: How does the predictive contribution of visual quality differ between search and experience goods?**
  - ➢ Comparing how MUSIQ and BRISQUE improve classification performance across product types
  - ➢ Using Logistic Regression, Random Forest, and Multilayer Perceptron (MLP) with 10-fold cross-validation
- **RQ2: How do textual and visual conditions combine in pathways leading to high review helpfulness, and does visual quality become core or a peripheral condition?**
  - ➢ Using fuzzy-set Qualitative Comparative Analysis (fsQCA) to identify sufficient configurations
  - ➢ Comparing whether visual quality appears as core, peripheral, or irrelevant conditions

2. Literature Review

2.1. Review Helpfulness and Multimodal Review Information

- Review helpfulness refers to the extent to which consumers perceive a review as useful for making purchase decisions (Mudambi & Schuff, 2010).
  - ➢ Helpful reviews reduce uncertainty by providing diagnostic information
  - ➢ Shaped by textual cues: review length, sentiment, rating information (Hong et al., 2017)
- Reviews are increasingly multimodal, combining text, ratings, and user-generated images
  - ➢ Images provide evidence about product appearance, size, defects, usage context (Ma et al., 2018)
  - ➢ Consumers evaluate reviews by integrating multiple cues from text and images

2.2. Visual Quality in Review Helpfulness

- Prior work has measured visual information mainly as **image presence** (included vs. not included)
  - ➢ This does not capture differences in the quality of images
- Recent research examines **image aesthetics** and their role in review helpfulness (Han et al., 2025)
  - ➢ **MUSIQ**: aesthetic and perceptual image quality (Ke et al., 2021)
  - ➢ **BRISQUE**: technical distortions such as blur, noise, loss of sharpness (Mittal et al., 2012)
- Visual quality may contribute differently depending on the product being evaluated

2. Literature Review

2.3. Product Type Differences: Search and Experience Goods

- **Search goods** can be evaluated before purchase through objective attributes (Nelson, 1974)
  - ➢ Consumers can rely on specifications and descriptions
- **Experience goods** require use or consumption before quality can be fully assessed (Huang et al., 2009)
  - ➢ Consumers depend more on cues that help imagine product use
- The same review feature may not carry the same value across product types
  - ➢ Visual quality may be more useful when consumers need help imagining product performance
  - ➢ When objective attributes are available, visual quality may provide less additional value

2.4. Predictive and Configurational Approaches

- **Predictive modeling** evaluates whether specific variables improve classification performance (variance-oriented view)
  - ➢ Useful for examining whether MUSIQ and BRISQUE add information beyond image presence
- **Configurational analysis (fsQCA)** identifies combinations of conditions that are sufficient for an outcome (systems-oriented view; Ragin, 2008; Fiss, 2011)
  - ➢ Examines whether features become core, peripheral, or irrelevant in pathways to high helpfulness
- **Key distinction**: Predictive usefulness ≠ configurational centrality
  - ➢ A feature may improve classification without becoming a defining condition in sufficient combinations

3. Research Framework

This study develops a **two-phase analytical framework** applied to reviews collected from **MercadoLibre Chile**:

Phase 1: Data Collection & Variable Operationalization

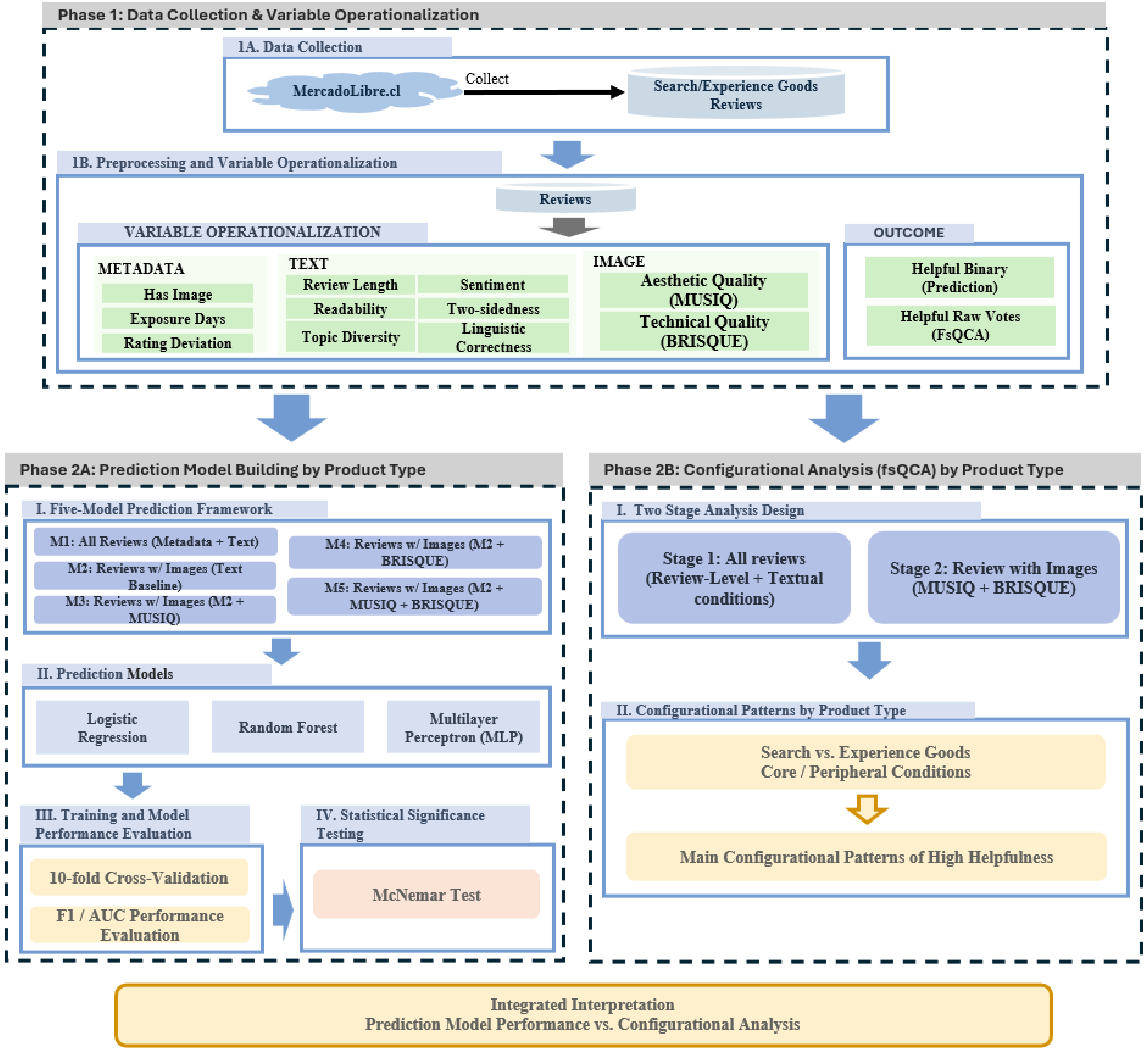
- Reviews classified as search or experience goods
- Variables operationalized across metadata, textual features, and visual quality (MUSIQ, BRISQUE)

Phase 2A: Review Helpfulness Prediction

- Three classifiers (Logistic Regression, Random Forest, MLP) under a five-model progression (M1–M5) isolating visual quality contributions
- Evaluated separately by product type using 10-fold cross-validation

Phase 2B: Review Helpfulness Configurational Analysis (fsQCA)

- Two-stage design: Stage 1 (all reviews, text+metadata); Stage 2 (image reviews, +MUSIQ+BRISQUE)
- Identifies product-type-specific pathways to high helpfulness



<Figure 1> Research Framework

4. Methodology – Data Collection

Source: MercadoLibre Chile (MercadoLibre.cl)
Method: Automated Web Crawling by using Selenium library

Product Selection

- Based on **search vs. experience goods classification** (Nelson, 1974; Huang et al., 2009)

Data Extraction:

- **80,145 reviews** collected across 593 products
- Review metadata (rating, dates, helpfulness votes)
- Review text
- Image-related fields (URLs/files)

Filtering Criteria

- Reviews from **2023 onwards**
- Reviews with **at least one helpful vote** and reviews with **valid text and product type** (Mudambi & Schuff, 2010; Sun et al., 2019)

Final Dataset

- Final analytical dataset**: 14,603 reviews from 579 products**

<Figure 2> Example of a review on MercadoLibre platform.

4. Methodology – Variable Operationalization

Variables organized into **three modalities:**

<Table 1> Operationalization of Multimodal Review Variables

Type	Variable	Operationalization	Key References
Metadata	Image presence	Binary indicator: 1 if the review includes at least one user-generated image	Sun et al., 2019; Kübler et al., 2024
Metadata	Exposure days	Days between review posting date and data collection date	Mudambi & Schuff, 2010; Siering et al., 2018
Metadata	Rating deviation	Absolute deviation between review rating and product average rating	Yin et al., 2016; Kuan et al., 2015
Textual	Review length	Total word count of the review text	Mudambi & Schuff, 2010; Liu & Park, 2015
Textual	Sentiment score	Positive minus negative sentiment probability using a Spanish-language sentiment model	Pérez et al., 2024; Hutto & Gilbert, 2014
Textual	Readability	Spanish-adapted readability score based on the IFSZ formula	Szigriszt Pazos, 1993
Textual	Topic diversity	Proportion of Garvin's product-quality dimensions detected in the review	Garvin, 1984; Sun et al., 2019
Textual	Linguistic correctness	Normalized spelling accuracy based on misspelled words	Ghose & Ipeirotis, 2011
Textual	Two-sidedness	Co-occurrence of positive and negative evaluative content in the review	Mudambi & Schuff, 2010; Hong et al., 2017
Visual	Aesthetic quality	Mean MUSIQ score across review images, capturing perceptual image quality	Ke et al., 2021; Han et al., 2025
Visual	Technical quality	Mean BRISQUE score across review images, capturing blur, noise, and visual distortions	Mittal et al., 2012
Outcome	Review helpfulness	Helpful binary for prediction; fuzzy-calibrated helpfulness for fsQCA	Mudambi & Schuff, 2010; Sun et al., 2019

4. Methodology – Variable Operationalization

<Table 2> Summary Statistics

Group	Variable	N	Missing (n)	Missing (%)	Mean	Std. Dev.	Min	25%	Median	75%	Max	0 (n)	1 (n)	0 (%)	1 (%)
Outcome	Helpfulness Votes	14603	0	0	6.476	35.304	1	1	1	4	2762	NaN	NaN	NaN	NaN
Outcome	Review Helpfulness (Binary)	14603	0	0	0.416	0.493	0	NaN	0	NaN	1	8529	6074	58.41	41.59
Textual	Aspect Diversity	14603	0	0	0.183	0.132	0	0.111	0.111	0.222	0.889	NaN	NaN	NaN	NaN
Textual	Readability	14603	0	0	58.314	31.614	0	45.085	67.696	80.925	100	NaN	NaN	NaN	NaN
Textual	Review Length	14603	0	0	19.195	19.927	1	6	13	25	241	NaN	NaN	NaN	NaN
Textual	Sentiment	14603	0	0	0.59	0.549	-0.984	0.338	0.892	0.96	0.984	NaN	NaN	NaN	NaN
Textual	Two-sidedness	14603	0	0	0.136	1.03	0	0	0	0	55.97	NaN	NaN	NaN	NaN
Textual	Linguistic Correctness	14603	0	0	0.981	0.018	0.8	0.97	0.98	1	1	NaN	NaN	NaN	NaN
Visual	Aesthetic Image Quality (MUSIQ)	4847	9756	66.81	67.505	7.564	21.109	63.556	69.125	73.242	79.084	NaN	NaN	NaN	NaN
Visual	Technical Image Quality (BRISQUE)	4847	9756	66.81	38.369	17.502	-22.633	26.315	37.583	49.047	206.688	NaN	NaN	NaN	NaN
Metadata	Exposure Days	14603	0	0	456.913	304.729	2	169	428	707	1176	NaN	NaN	NaN	NaN
Metadata	Image Presence	14603	0	0	0.332	0.471	0	NaN	0	NaN	1	9756	4847	66.81	33.19
Metadata	Rating Deviation	14603	0	0	0.232	0.471	0	0.015	0.054	0.127	3.821	NaN	NaN	NaN	NaN

4. Methodology – Two Phase Analytical Approach

This study employs **complementary analytical methods** to examine both predictive utility and configurational centrality of visual quality:

- ➢ **Variance-oriented approach**: Evaluates whether variables improve classification (predictive modeling)
- ➢ **Set-theoretic approach**: Identifies sufficient combinations of conditions (fsQCA) ➢ Addresses the question: *Does predictive importance correspond to configurational centrality?* (Ragin, 2008; Fiss, 2011)

Phase 1: Predictive Modeling (Variance-Oriented)

- **Models**: Logistic Regression, Random Forest, Multilayer Perceptron
- **Five-model progression** systematically isolates visual quality contribution:

<Table 3> Five-model Prediction Design

Model	Dataset	Feature set
M1	All reviews	Metadata + Textual
M2	Image-inclusive reviews	Metadata + Textual
M3	Image-inclusive reviews	M2 + MUSIQ (Aesthetic image quality)
M4	Image-inclusive reviews	M2 + BRISQUE (Technical image quality)
M5	Image-inclusive reviews	M2 + MUSIQ + BRISQUE (Full visual quality)

- **Evaluation**: 10-fold cross-validation (F1, AUC) by product type.

4. Methodology – Two Phase Analytical Approach

Phase 2: Configurational Analysis (Set-Theoretic)

- **Method**: Fuzzy-set Qualitative Comparative Analysis (fsQCA)
  - Examines **sufficient combinations** of conditions that lead to high review helpfulness
  - Identifies whether visual quality appears as **core, peripheral, or irrelevant** conditions in pathways to high helpfulness
  - Complements Phase 1 by revealing whether predictive importance translates to configurational centrality

Analysis Approach

- **Necessity analysis**: Tests whether any single condition is necessary for high helpfulness
- **Sufficiency analysis**: Identifies **configurations** (combinations of conditions) sufficient for high helpfulness
  - **Frequency cutoff:** 90th percentile of cumulative cases per analytical subset
  - **Raw consistency:** Stage 1 ≥ 0.80 / Stage 2 ≥ 0.90 (Ragin, 2008)
  - **PRI consistency:** Stage 1 ≥ 0.60 / Stage 2 ≥ 0.65 (Schneider & Wagemann, 2012)
  - **Intermediate solutions excluded:** limited theoretical development on visual quality mechanisms justifies complex and parsimonious solutions only
- **Conducted separately** for search goods and experience goods
  - Allows identification of product-type-specific pathways
  - Addresses whether visual quality contribution varies by product type

<Table 4> Two-Stage fsQCA Analytical Design

Stage	Sample	Conditions Included	Analytical Purpose
Stage 1	All reviews	**Metadata:** image presence, rating deviation **Textual:** sentiment, review length, readability, topic diversity, linguistic correctness	Identify baseline configurations of review helpfulness before incorporating visual quality dimensions
Stage 2	Image-inclusive reviews	**Metadata:** rating deviation **Textual:** sentiment, review length, readability, topic diversity, linguistic correctness **Visual quality:** MUSIQ, BRISQUE	Examine whether aesthetic and technical image quality become core, peripheral, or absent conditions in sufficient pathways

5. Results – Phase 2A: Visual Quality Contribution

Visual quality provides **modest and selective gains** beyond the image-review baseline. The contribution is more visible for **experience goods**, but the gains are not uniformly additive across M3–M5.

- Compared with **M2 (Image baseline)**, visual-quality models show **more consistent gains for experience goods** than for search goods.
- For **experience goods**, RF F1 increases from **82.3% in M2** to **83.0% in M3**, suggesting that aesthetic quality adds a small predictive contribution.
- For **search goods**, RF F1 increases only slightly from **62.2% in M2** to **62.6% in M3**, while M5 does not improve F1.
- AUC provides a secondary check: experience-goods AUC increases from **82.4% in M2** to **83.4% in M4/M5**, but the overall pattern remains modest and non-uniform.
- **McNemar test:** for experience goods, **M5 differs from M3 and M4** (p = 0.001; p = 0.013), suggesting that combining MUSIQ and BRISQUE changes some helpfulness predictions despite small F1/AUC gains.

<Table 5> Predictive Performance across Visual-Quality Model Progressions

Product Type	Metric	M2 (image baseline)	M3 (+MUSIQ)	M4 (+BRISQUE)	M5 (+Both)
Search goods	F1	62.2%	**62.6%**	62.4%	61.8%
	AUC	75.0%	76.4%	75.9%	**76.7%**
Experience goods	F1	82.3%	**83.0%**	82.6%	81.5%
	AUC	82.4%	83.3%	**83.4%**	**83.4%**

<Table 6> ΔF1 relative to M2 baseline

Product type	M3 (+ MUSIQ)	M4 (+ BRISQUE)	M5 (+Both)
Search	+0.4 pp	+0.2 pp	-0.4 pp
Experience	+0.7 pp	+0.3 pp	-0.8 pp

5. Results – Phase 2B: fsQCA Stage 1 Configurations

Absent linguistic correctness, reflecting authentic, colloquial consumer writing, is a core condition for high helpfulness across both product types

- **Absent linguistic correctness** is the only core condition shared across both product types; reviews written in a more natural, informal style consistently form part of sufficient pathways to high helpfulness (Srivastava & Kalro, 2019)

- For search goods, **image presence, review length, and topic diversity** are also core. helpful reviews are informationally rich and multimodal

- For experience goods, **absent linguistic correctness is the only core condition.** Unlike search goods, no condition reaches core status in its present form, suggesting that what defines helpful experience goods reviews is more about writing style than informational richness

- Solution consistency: Search S1a = **0.852** / Experience E1a = **0.819**

Condition	Search Goods (S1a)	Experience Goods (E1a)
Has Image	●	•
Sentiment	•	•
Review Length	●	•
Readability	•	⊗
Rating Deviation	⊗	⊗
Linguistic Correctness	⨂	⨂
Topic Diversity	●	•
Consistency	0.852	0.819
Unique Coverage	0.059	0.076
Overall Consistency	0.852	0.819
Overall Coverage	0.059	0.076

●/⨂ = core present/absent; •/⊗ = peripheral present/absent

<Figure 3> Sufficient configurations for high helpfulness - Stage 1

5. Results – Phase 2B: fsQCA Stage 2 Configurations

Absent high positive sentiment is the sole core condition across all configurations and both product types; once images are present, evaluative tone becomes the defining feature of helpful reviews.

- **Absent sentiment and sufficient review length** are the only two core conditions shared across all 5 search goods configurations, when images are present, reviews with restrained sentiment and adequate length consistently achieve high helpfulness

- For experience goods, **absent sentiment is the sole core condition,** once visual information is available, evaluative tone dominates over writing style

- Visual quality conditions appear **exclusively in peripheral and absent roles** across all 7 configurations, image quality raises overall informativeness but does not define the configurational pathways to helpfulness

- Solution consistency: Search = **0.912** / Experience = **0.965**

	Search goods					Experience goods	
Condition	**S2a**	**S2b**	**S2c**	**S2d**	**S2e**	**E2a**	**E2b**
Sentiment	⨂	⨂	⨂	⨂	⨂	⨂	⨂
Review Length	●	●	●	●	●	•	•
Readability		⊗	•	⊗	•		⊗
Rating Deviation	⊗	⊗	⊗	•	•	⊗	⊗
Linguistic Correctness	•	⊗	•	•	⊗	•	•
Topic Diversity	•	⊗	⊗	⊗	⊗	•	•
Aesthetic Quality	⊗	⊗	⊗	⊗		⊗	
Technical Quality	•	⊗	⊗	⊗	⊗	⊗	⊗
Consistency	**0.972**	**0.963**	**0.966**	**0.969**	**0.988**	**0.973**	**0.969**
Coverage	**0.108**	**0.109**	**0.109**	**0.110**	**0.102**	**0.110**	**0.110**
Overall Consistency	**0.912**					**0.965**	
Overall Coverage	**0.134**					**0.115**	

●/⨂ = core present/absent; •/⊗ = peripheral present/absent

<Figure 4> Sufficient configurations for high helpfulness - Stage 2

6. Discussion – Key Findings

Key Finding: Visual quality supports prediction, especially for experience goods, but does not define configurational pathways.

RQ1: How does the predictive contribution of visual quality differ between search and experience goods?

- Visual quality shows a **stronger predictive role for experience goods** than for search goods.
- Under Random Forest, experience-goods AUC increased from **82.4% in M2** to **83.4% in M4/M5**, while search-goods gains were smaller and less stable.
- The contribution of MUSIQ and BRISQUE is **modest and selective**, suggesting a product-type-sensitive rather than universal effect.

RQ2: How do textual and visual conditions combine in pathways leading to high review helpfulness, and does visual quality become core or a peripheral condition?

- Stage 1: **Absent high linguistic correctness** appears as a core condition across product types, , suggesting that helpful reviews may sound more natural and consumer-generated rather than formally polished.
- Stage 2: **Absent strongly positive sentiment** becomes core once images are present, suggesting the value of balanced evaluative tone.
- MUSIQ and BRISQUE appear mainly as **peripheral or absent conditions**, meaning that image quality may support some pathways but does not define the main configurations leading to high helpfulness.

Visual quality contributes to prediction, but does not become configurationally central: it supports helpfulness selectively, while high-helpfulness pathways remain anchored in textual authenticity and balanced evaluation.

6. Discussion – Contributions

Theoretical Contribution

- Extends review helpfulness literature by showing that **predictive usefulness and configurational centrality do not necessarily align**: visual quality can support prediction without becoming a core condition in high-helpfulness pathways.
- Shows that visual quality has a **product-type-sensitive predictive contribution**: it is more pronounced for experience goods, but remains modest and non-universal across product types.
- Shows that helpful reviews are not necessarily the most polished or highly positive ones; instead, **authentic writing** and a **balanced evaluative tone** are more consistently associated with high helpfulness.

Methodological Contribution

- Develops an **integrated two-phase framework** combining predictive modeling and fsQCA within a unified design.
- Demonstrates that the two methods answer **complementary questions**: predictive modeling evaluates whether visual quality adds incremental usefulness, while fsQCA shows whether it becomes **core, peripheral, or absent** in sufficient pathways

Empirical Contribution

- Suggests that platforms should avoid ranking reviews based only on **image presence**, since the usefulness of visual information depends on product type and image quality.
- Shows that, for **experience goods**, visual-quality indicators may help identify reviews with richer diagnostic information.
- Recommends **product-type-sensitive review ranking** that combines visual-quality indicators with textual signals of authenticity and balanced evaluation.

6. Discussion - Limitations & Future Research

Limitations

- **Helpfulness vote filter constrains the sample:** the analysis includes only reviews with at least one helpful vote; zero-vote reviews may reflect limited exposure rather than genuine low helpfulness.
- **Positive helpful votes only:** MercadoLibre does not provide explicit “unhelpful” votes, limiting the outcome to positive evaluation signals.
- **Single platform context:** findings are based on MercadoLibre Chile and may not fully generalize to other platforms or cultural contexts.
- **Visual quality, not image meaning:** MUSIQ and BRISQUE capture image quality, but not product defects, usage context, emotional cues, or image-text congruence.
- **Limited fsQCA solution coverage:** the identified configurations are theoretically meaningful, but their coverage values are relatively low, meaning that they explain only part of the high-helpfulness cases. This suggests that additional pathways may exist but were not captured due to the current sample distribution and platform-specific voting patterns.

Future Research

- Test the framework across platforms and countries.
- Add richer visual-semantic features.
- Examine both helpful and unhelpful pathways using platforms with symmetric voting mechanisms.

Thank You for your attention!