CARMA 2026

*8th Int. Conf. on Advanced Research Methods and Analytics*
1-3 July 2026, Valencia, Spain
DOI: https://doi.org/10.4995/CARMA2026.2026.21985

# Content is King, Context is Not: Predicting Social Media Engagement in the Italian Telecom Sector

**Giuseppe Di Renzo[1], Giuliano Resce[1], Santiago Nunez Rimedio[2]**
[1]Department of Economics, University of Molise, Italy, [2]Dimensione s.r.l., Italy

How to cite: Di Renzo, G.; Resce, G.; Nunez Rimedio, S. (2026). Content is King, Context is Not: Predicting Social Media Engagement in the Italian Telecom Sector. In: 8th International Conference on Advanced Research Methods and Analytics (CARMA 2026). Valencia, 26-28 June 2026. https://doi.org/10.4995/CARMA2026.2026.21985

***Abstract***

*We investigate whether the textual content of social media posts drives user engagement in the Italian telecommunications sector, or whether contextual factors such as brand identity and posting time are the primary determinants. Using 2,588 Facebook posts from 12 Italian telecom operators (2023-2024), we compare three XGBoost models: one using only contextual features (company identity, temporal variables), one using only content features (384-dimensional multilingual sentence embeddings), and one combining both. Results show that context alone has no predictive power, while content alone explains 17.5%. A complementary Lasso analysis on TF-IDF features reveals that technical, service-oriented vocabulary outperforms generic promotional and entertainment language. Additionally, we integrate corporate financial reporting data and textual characteristics of corporate disclosures (narrative length, tone, thematic emphasis) using descriptive and regression analyses. Results indicate that firms' strategic communication patterns in reporting are associated with social media content effectiveness, highlighting the role of managerial signaling and textual strategy in driving engagement. Per-company analysis shows that content is especially critical for smaller operators: Dimensione, a regional fiber provider, shows the highest content dependence among mid-sized operators. These findings challenge the assumption that brand size determines social media success and suggest that smaller operators can compete effectively through content quality.*

***Keywords:*** *NLP, Financial Disclosure, Text Mining, social media engagement, XGBoost, sentence embeddings*

Article in Press

## 1. Introduction

Corporate financial reporting functions not only as a mechanism for communicating economic performance but also as a strategic channel through which firms shape stakeholder perceptions

This work is licensed under a Creative Commons License CC BY-NC-SA 4.0
Editorial Universitat Politècnica de València

and influence external audiences (Healy & Palepu, 2001). In recent years, research has increasingly emphasized the qualitative dimensions of corporate disclosure, showing how narrative components of annual reports complement numerical accounting data by revealing managerial intentions, strategic priorities, and organizational positioning (Li, 2008; Loughran & McDonald, 2011). Features of corporate narratives, including linguistic tone, textual length, and thematic orientation, have been found to influence investor beliefs, market reactions, and the interpretation of firm performance (Kothari et al., 2009; Dikolli et al., 2014). From an industrial organization perspective, such communication can also be interpreted as a signaling mechanism through which firms convey financial strength, technological capability, and long-term strategic commitments, thereby shaping expectations and competitive dynamics in capital-intensive industries.

In parallel, digital communication has emerged as a key interface between firms and stakeholders. Social media platforms enable companies to distribute content rapidly while receiving immediate feedback through engagement metrics such as reactions, comments, and shares. A central unresolved question is whether engagement is primarily driven by contextual elements—such as brand identity, firm size, or posting time—or by the semantic substance of the content itself. This issue is particularly relevant in the telecommunications sector, where firms with heterogeneous scale and visibility compete for attention in highly saturated information environments. Smaller operators, such as Dimensione, often face structural disadvantages in brand recognition, making content strategy a potentially critical lever for visibility and engagement. This study connects these two strands by jointly analyzing corporate disclosure practices and social media communication strategies. Using a dataset of 2,588 Facebook posts from 12 Italian telecom operators over the 2023–2024 period, we combine financial and organizational indicators derived from corporate reports with textual features extracted from both annual disclosures and social media content. We develop and compare predictive models based on XGBoost to evaluate the relative importance of contextual variables versus semantic representations in explaining engagement outcomes. By integrating narrative characteristics from annual reports with embedding-based representations of social media text, we aim to identify systematic patterns in firm communication strategies and assess how textual content shapes stakeholder interaction across firms of different scale and market position.

This study makes three main contributions. First, it provides novel empirical evidence on the relative importance of semantic content versus contextual factors in predicting social media engagement within a competitive telecommunications market, showing that semantic information substantially outperforms structural and temporal determinants. Second, it bridges two largely separate strands of literature—social media engagement modeling and textual analysis of corporate disclosures—by integrating unstructured text from both communication channels, thereby framing corporate communication as a unified strategic capability. Third, it introduces a multi-method design combining machine learning prediction (XGBoost),

Article in Press

interpretable feature selection (Lasso with TF-IDF), and panel regression with firm fixed effects, improving both predictive robustness and economic interpretability. Overall, the results suggest that content quality is a key driver of both digital engagement and firm performance, while contextual factors such as brand identity and posting timing play a limited role in this setting.

## 2. Data and Methods

### 2.1 Social Media Data

We collected public Facebook posts from 12 Italian telecommunications operators for 2023–2024: CoopVoce, Dimensione, Fastweb, Ho., Iliad, Kena, Poste Mobile, Sky Mobile, Tiscali Mobile, Very Mobile, Vodafone, and WINDTRE. The final dataset comprises 2,588 posts after filtering for text availability and excluding operators with insufficient temporal coverage. Engagement is measured as the sum of comments, shares, and reactions per post. To enable cross-company comparison, engagement is self-normalized by dividing each post's total by its company's annual mean, producing a scale where 1.0 represents an average post for that company in that year.

### 2.2 Financial and Corporate Communication Data

Financial statement data were obtained from the **AIDA (Computerized Analysis of Italian Companies)** database by Bureau van Dijk, containing balance sheets, personal data, and product information for all active Italian capital companies. Corporate communication data were obtained through web scraping of annual reports and social media posts using Python and the Beautiful Soup package. The reference period is 2022–2024, and the sample includes 12 firms selected based on: operating in the same industry sector for comparability, being direct competitors to allow benchmarking, similar size (total assets and revenues) to minimize scale distortions.

### 2.3 Feature Sets

For the analysis, we considered two main types of features: **contextual** and **content-based**. The **contextual features** include 23 variables capturing structural and temporal aspects of each post. These consist of company identity indicators (11 dummy variables), cyclical encodings of the posting hour, day of the week, and month (6 variables), as well as additional indicators for weekends and year. We also included structural characteristics of the text itself, such as character count, word count, the presence of URLs, and the use of emojis. These features allow us to account for timing, company-specific effects, and general textual structure, independent of the semantic content of the posts. The **content features** are 384-dimensional sentence embeddings generated using the multilingual transformer model *paraphrase-multilingual-*

Article in Press

*MiniLM-L12-v2*. Each post's text is encoded into a normalized vector that captures semantic and linguistic meaning. This allows us to quantify the textual content of the posts in a way that is suitable for machine learning models.

### 2.4 Engagement Models

To evaluate the drivers of social media engagement, we estimated three predictive models using XGBoost gradient boosting, each evaluated via 5-fold cross-validation.

1. **Context Only:** using only the 23 contextual features described above;
2. **Content Only:** using only the 384 embedding dimensions;
3. **Combined:** incorporating all 407 features.

**Table 1. Per-company $R^2$ and Content Lift (Combined $R^2$ minus Context Only $R^2$).**

| Company | Context Only | Content Only | Combined | Content Lift |
|---|---|---|---|---|
| Vodafone | 0.466 | 0.914 | 0.905 | + 0.439 |
| Very Mobile | 0.061 | 0.358 | 0.433 | + 0.372 |
| Sky mobile | -0.290 | -0.042 | -0.025 | + 0.265 |
| Fastweb | -0.210 | -0.072 | 0.040 | + 0.250 |
| WINDTRE | 0.482 | 0.678 | 0.711 | + 0.229 |
| Dimensione | -0.171 | 0.019 | 0.021 | + 0.192 |
| Tiscali Mobile | -0.101 | 0.057 | 0.069 | + 0.170 |
| Kena mobile | -0.278 | -0.110 | -0.137 | + 0.141 |
| CoopVoice | 0.084 | 0.118 | 0.171 | + 0.087 |
| Ho. | 0.080 | 0.119 | 0.117 | + 0.037 |
| Poste mobile | -0.010 | -0.012 | 0.023 | + 0.033 |
| Iliad | 0.523 | 0.531 | 0.537 | + 0.014 |

All models were estimated with fixed hyperparameters - 300 estimators, a maximum tree depth of 6, learning rate of 0.05, and L1/L2 regularization - to ensure fair comparison of feature importance across models. In addition, we performed a complementary analysis using **Lasso regression** on TF-IDF features extracted from the text (both unigrams and bigrams). This approach allowed us to identify specific words or phrases associated with higher or lower engagement, providing interpretable insights alongside the embedding-based models.

**Table 2. Top Lasso coefficients: words associated with higher (+) and lower (−) engagement.**

| Engagement Boosters | Coef. | Engagement Killers | Coef. |
|---|---|---|---|
| comunicazione (communication) | + 3.98 | gratis (free) | -2.14 |
| ufficiale (official) | + 8.02 | ultraveloce (ultra-fast) | -1.87 |
| rame (copper network) | + 6.11 | sorriso (smile) | -1.92 |

Article in Press

| | | | |
|---|---|---|---|
| stabilità jitter (stability) | + 7.40 | cambiamento (change) | -1.69 |
| protezione (protection) | + 2.58 | sociale (social) | -1.54 |
| servizio (service) | + 1.51 | digitale (digital) | -1.10 |
| comune (municipal) | + 1.93 | calendario (calendar) | -0.96 |
| compromesso (compromise) | + 3.50 | capire (to understand) | -0.92 |
| supporto (support) | + 1.84 | auguri (greetings) | -0.71 |
| prezzo (price) | + 1.96 | app | -0.35 |

### 2.5 Financial Performance Analysis

To explore how corporate communication relates to firm performance, we constructed a **Strategic Index** for each firm-year. This index aggregates textual features extracted from both Facebook posts and annual reports, including measures such as post length, linguistic tone, and thematic emphasis on technology and infrastructure.

We then estimated a **fixed-effects (within) panel regression** of operating margin on the Strategic Communication Index. In these models, we controlled for major cost components - services, personnel, and materials - to isolate the effect of strategic communication on financial performance. By using a within-firm approach, we account for time-invariant firm-specific factors, focusing on how changes in communication strategy are associated with changes in operating margins over time.

The fixed-effects model is defined as follows:

$$Op.\ Margin_{it} = \alpha_i + \beta_1\ Strategic\ Communication\ Index_{it} + \beta_2\ Cost\ of\ Services_{it} + \beta_3\ Cost\ of\ Personnel_{it} + \beta_4\ Cost\ of\ Materials_{it} + \epsilon_{it}$$

Where:

- $i = 1, \dots, N$ denotes the firm, and $t = 1, \dots, T$ denotes the year;
- $\alpha_i$ represents **firm-specific fixed effects**, capturing time-invariant unobserved heterogeneity across firms;
- Operating Margin$_{it}$ is the ratio of operating income to total revenues for firm $i$ in year $t$;
- Strategic Communication Index$_{it}$ is a composite index of corporate communication for firm $i$ in year $t$, calculated by combining textual features such as Post length, Linguistic tone (positive/negative/neutral), Thematic emphasis on technology, infrastructure, and service quality. The index is normalized across firms to allow comparability;

Article in Press

- Cost of Services$_{it}$, Cost of Personnel$_{it}$, Cost of Materials$_{it}$ are ratios of the respective costs to total revenues, controlling for differences in cost structure that could influence profitability;
- $\epsilon_{it}$ is the idiosyncratic error term.

**Table 3. Fixed-Effects Regression on Operating Margin.**

| Variable | Coef. | Std. Error | t-value | p-value | Significance |
|---|---|---|---|---|---|
| Strategic Index | 0.1835 | 0.0594 | 3.089 | 0.021 | * |
| Cost Services | -0.5987 | 0.3912 | -1.531 | 0.177 | |
| Cost Personnel | -0.3445 | 0.5968 | -0.577 | 0.585 | |
| Cost Materials | -1.4433 | 1.0709 | -1.348 | 0.226 | |

## 3. Conclusions

The analyses conducted in this study provide a comprehensive view of how corporate communication influences both social media engagement and financial performance in the Italian telecommunications sector. Our findings indicate that the **semantic content of posts** is the primary driver of user interaction, while contextual factors such as brand identity, posting time, or day of the week contribute negligibly. The XGBoost models show that content features alone explain approximately **17.5–19% of engagement variance**, whereas context alone has effectively **zero predictive power**. This demonstrates that high-quality, informative, and service-oriented messaging matters far more than superficial posting patterns or brand recognition, particularly for smaller operators like Dimensione, where content strategy is a critical competitive lever. Complementing this, the **fixed-effects panel regression** linking corporate communication to operating margins reveals that improvements in the **Strategic Communication Index** are positively and significantly associated with higher operating profitability ($\beta = 0.184$, $p < 0.05$). This indicates that firms that actively invest in clear, strategic, and technically-oriented communication not only generate higher engagement online but also experience tangible benefits in operational performance. Cost variables—services, personnel, and materials—did not show statistically significant effects in this panel, suggesting that communication quality can influence profitability independently of cost structure variations over the observed period.

Taken together, these results underscore a **dual impact of communication quality**:

Article in Press

1. **Digital engagement:** Content-rich, technically informative posts outperform generic promotional content, driving interaction even for smaller firms with limited brand awareness;
2. **Financial performance:** Firms with more strategic and focused communication see higher operating margins, highlighting a measurable economic benefit from well-managed disclosure and messaging strategies.

Several limitations should be acknowledged. First, engagement measures capture only active interactions (likes, comments, shares) and do not account for passive consumption, such as views or reach. Second, the text-based analysis does not include visual or multimedia content, which may further influence engagement. Third, the social media panel covers a relatively short period (2023–2024), limiting the ability to generalize over time. Finally, while the FE regression controls for unobserved time-invariant firm heterogeneity, the small panel size reduces statistical power and precludes firm-specific causal inferences. Despite these constraints, the findings provide strong evidence that **strategic, service-oriented communication is a key determinant of both engagement and profitability**. For managers, this implies that investment in technical, informative, and targeted messaging can yield measurable benefits in user interaction and operational outcomes. For researchers, the integration of text-based social media features with financial performance metrics demonstrates the value of combining qualitative disclosure analysis with quantitative financial modeling.

## References

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

Cormier, D., & Magnan, M. (2015). The economics of disclosure and the role of textual analysis. *Accounting, Auditing & Accountability Journal, 28*(5), 713–740.

Dikolli, S. S., Kulp, S., & Sedatole, K. L. (2014). The value of narrative in management disclosures: Evidence from earnings conference calls. *The Accounting Review, 89*(3), 993–1025.

García-Sánchez, I. M., Cuadrado-Ballesteros, B., & Frias-Aceituno, J. V. (2019). Corporate disclosure strategies: A literature review and future research avenues. *Spanish Journal of Finance and Accounting, 48*(2), 123–151.

Healy, P. M., & Palepu, K. G. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics, 31*(1–3), 405–440.

Kothari, S. P., Li, X., & Short, J. E. (2009). The effect of disclosures by management, analysts, and financial press on the stock price: Evidence from textual analysis. *The Accounting Review, 84*(5), 1639–1670.

Article in press

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics, 45*(2–3), 221–247.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance, 66*(1), 35–65.

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing* (pp. 3982–3992). Association for Computational Linguistics.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological), 58*(1), 267–288.

Article in Press