The American Journal of Applied Sciences
ISSN 2689-0992 | Open Access

TYPE Original Research
PAGE NO. 31-39



# Big Data and Text Analytics on 833K Glassdoor Reviews: Aspect-Level Employee Sentiment as a Short-Horizon Early-Warning Signal of Firm Distress

OPEN ACCESS

SUBMITED 18 November 2025
ACCEPTED 20 December 2024
PUBLISHED 30 January 2025
VOLUME Vol.07 Issue 01 2025

CITATION
Singh, N. (2025). Big Data and Text Analytics on 833K Glassdoor Reviews: Aspect-Level Employee Sentiment as a Short-Horizon Early-Warning Signal of Firm Distress. The American Journal of Applied Sciences, 7(01), 31–39. Retrieved from https://theamericanjournals.com/index.php/tajas/article/view/8016



**Nidhi Singh**

Senior Data Analyst, State of Alabama, AL USA

**Abstract:** We test whether the tone of free-text employee reviews, decomposed into specific human-resource (HR) aspects, anticipates near-term deterioration in a firm's standing on the same review platform. Using 833,760 Glassdoor reviews of 332 large employers (2008–2021), we measure sentiment on five aspects—management, work–life balance, compensation, culture, and career growth—with a transparent, reproducible lexicon-based aspect sentiment procedure, and we validate this text signal against the platform's own human-assigned sub-ratings (Pearson r between 0.42 and 0.52). In a firm–month panel with two-way (firm and calendar-month) fixed effects and firm-clustered standard errors, more negative management, culture, and career-growth sentiment in month t−1 is associated with a significantly higher probability of a firm-level rating decline in month t. The association is statistically robust but economically modest (within-$R^2$ ≈ 0.027), and—importantly—it is short-horizon: at a three-month lead the predictive content essentially vanishes (within-$R^2$ ≈ 0.000; joint F not significant). An event study around 703 distress onsets shows aspect sentiment dipping roughly one to two months before the rating trough, and Granger-precedence tests support a lead for a majority (14 of 25) of long-series firms. We deliberately frame the outcome as an internal platform signal rather than a hard external event (e.g., layoffs): the canonical external layoffs dataset does not overlap the review window in time, precluding external validation here. Results should therefore be read as evidence that

aspect sentiment leads a soft, platform-internal distress proxy at a short horizon, not as a validated predictor of corporate restructuring.

**Keywords:** aspect-based sentiment analysis; employee reviews; Glassdoor; early-warning indicators; panel fixed effects; organizational distress.

## I. INTRODUCTION

Crowdsourced employer-review platforms such as Glassdoor have become a large, continuously updated record of how employees evaluate their workplaces. A growing literature shows that aggregate employee satisfaction carries information about subsequent firm performance [1], [2], and practitioner work links cultural deterioration to elevated attrition [3]. A natural managerial question follows: do the specific dimensions of employee discontent—poor management, eroding culture, stalled career growth—move before broader, more visible signs of organizational trouble, and can they therefore serve as an early-warning indicator? We make this concrete with two research questions. (RQ1) Does aspect-level employee sentiment in a given month lead a near-term decline in the firm's standing on the same platform, net of firm and time effects? (RQ2) Over what horizon does any such leading-indicator content persist—at a one-month lead, or as far ahead as three months?

This paper provides a careful, deliberately conservative test of that question on a single large corpus. We make three design choices intended to keep the claims well-calibrated. First, rather than relying on document-level polarity, we measure sentiment at the level of five HR aspects using an aspect-segmented lexicon procedure that is fully reproducible and requires no proprietary model or manual labeling. Second, we exploit a feature of the data that prior work has not used for validation: the platform records human per-aspect sub-ratings alongside the free text, giving us a built-in gold standard against which to benchmark the text-derived signal. Third, because the most natural hard outcome—mass layoffs—cannot be linked to this corpus for a simple but consequential reason (the reviews end in mid-2021, before widely used layoff datasets begin), we define the outcome as an internal, platform-based distress event and are explicit that this is a softer target than an external restructuring event.

Our contribution is therefore evidential and methodological rather than a claim of a deployable layoff predictor. We document that (i) a transparent lexicon aspect measure agrees moderately with human sub-ratings; (ii) management, culture, and career-growth sentiment are statistically significant short-horizon leading indicators of a within-platform rating decline, net of firm and time effects; (iii) the lead is short (about one month), with no measurable three-month-ahead content; and (iv) the pattern survives changes in the distress threshold and an alternative, rating-independent outcome. We also report, as a substantive finding for anyone attempting the layoffs design, that the commonly cited review and layoff datasets do not overlap temporally.

The remainder of the paper is organized as follows. Section 2 reviews related work on employee sentiment, aspect-based sentiment analysis, and panel inference. Section 3 describes the data and documents the temporal non-overlap with the layoffs source. Section 4 sets out the aspect-sentiment measurement, the firm–month panel, the distress outcome, and the econometric specification. Section 5 presents the validation and the main leading-indicator results, including the event study and Granger-precedence tests. Section 6 reports robustness checks, Section 7 discusses the findings, Section 8 states the limitations and threats to validity, and Section 9 concludes.

## II. LITERATURE REVIEW

A substantial body of work establishes that employee-generated assessments of their employers carry information about the firm. Edmans [3] shows that employee satisfaction, proxied by the "Best Companies to Work For" list, is associated with superior long-run equity returns, consistent with intangible human capital being imperfectly priced. Working directly with crowdsourced reviews, Green et al. [5] find that improving Glassdoor employer reviews predict positive abnormal returns and earnings surprises. Together these studies motivate treating employee-generated text as informative about a firm's prospects, while also cautioning that such information may be partially priced and that review platforms are subject to selection.

Methodologically, sentiment analysis has progressed

from document-level polarity classification [10], [11] toward aspect-level extraction, which attributes sentiment to specific facets of an entity and was formalized in shared tasks such as SemEval ABSA [13], and more recently in transformer encoders [2], [7]. Transformer ABSA is the current state of the art, but it requires substantial compute and, for a domain-specific gold standard, manual annotation. Lexicon methods such as VADER [8] are weaker in raw accuracy but transparent, deterministic, and free of training data, which makes them attractive when reproducibility and auditability are priorities.

For describing the latent structure of free-text reviews, non-negative matrix factorization [9] over TF–IDF features remains a classic and interpretable alternative to Latent Dirichlet Allocation [1] and to embedding-based methods such as BERTopic [6]. On the inferential side, firm–month panels are typically handled with two-way fixed-effects models and firm-clustered standard errors [12], while temporal precedence between series is commonly assessed using Granger [4] tests, which are properly interpreted as predictive precedence rather than causation.

***Research Gap***

Although this literature establishes that aggregate employee sentiment is informative about firm outcomes, it leaves several questions open. Prior work has largely relied on document-level or firm-level aggregate polarity rather than decomposing sentiment into the specific human-resource dimensions—management, work–life balance, compensation, culture, and career growth—through which employee discontent is actually experienced. Whether these specific dimensions move ahead of broader, more visible signs of organizational trouble, and could therefore serve as early-warning indicators, remains unresolved.

A second gap concerns validation. Establishing that a text-derived sentiment measure is trustworthy normally requires costly manual annotation to build a domain-specific gold standard, which limits reproducibility. The Glassdoor corpus, however, records human per-aspect sub-ratings alongside the free text, providing a built-in benchmark that prior work has not exploited for validation. This makes it possible to assess a transparent, reproducible lexicon procedure directly against human judgments without bespoke labeling—an approach that has been underused relative to compute-intensive transformer models.

A third gap concerns the outcome itself. The most natural hard outcome for an early-warning study, mass layoffs, cannot be linked to this corpus, because the reviews end in mid-2021, before the widely used layoffs datasets begin, so an inner join on firm and month is empty by construction. This temporal non-overlap means the relevant question is not whether sentiment predicts external restructuring but whether aspect-level sentiment leads a softer, platform-internal distress signal, and over what horizon. The present study addresses these gaps by measuring sentiment at the aspect level with a transparent lexicon procedure, validating it against the platform's human sub-ratings, and testing its short-horizon leading-indicator content within a firm–month panel.

*Table 1. Sample construction and key panel quantities.*

| Quantity | Value |
|---|---|
| Raw reviews | 838,566 |
| Firms (raw) | 428 |
| Review window | 2008-01 to 2021-06 |
| Reviews after cleaning (≥8 words, firm ≥100 reviews) | 833,760 |
| Firms retained | 332 |
| Firm–months after activity filters (≥3 reviews/mo, | 21,819 |

| | |
|---|---|
| ≥24 months) | |
| Firms in panel | 277 |
| Distress firm–months (rating-decline definition) | 1,599 (7.3%) |
| Distress onsets (events) | 703 |

***The layoffs outcome does not overlap the reviews***

The original motivation for this design was to merge employee sentiment with an external hard outcome—mass layoffs—using a widely cited tech-layoffs dataset covering 2022–2024. We report a blocking fact: the review corpus ends in **June 2021**, before that layoff window begins, so an inner join on firm and month is empty by construction. Any layoffs-based test on this corpus would require either a layoffs source covering 2008–2021 or an extension of the review scrape into 2022–2024. We therefore proceed with a within-platform outcome and return to this point in Section 8.

## III. METHODOLOGY

### *A. Aspect-based sentiment measurement*

We score each review on five aspects $a \in$ {management, work–life balance, compensation, culture, career growth}. For a review we split the concatenated text into sentences, retain the sentences whose tokens match a hand-curated keyword set for aspect a, and assign the aspect score as the mean VADER (Hutto & Gilbert, 2014) compound polarity over those sentences:

$$s\_a(\text{review}) = (1 / |S\_a|) \cdot \Sigma\_(j \in S\_a)\ \text{compound}(j), \text{ where } S\_a = \{\text{sentences mentioning aspect } a\} \quad (1)$$

If no sentence mentions aspect a, the score is missing (we never impute a neutral zero). Averaging only aspect-relevant sentences is what makes the measure aspect-based rather than document-level. The share of reviews with at least one aspect-relevant sentence ("coverage") ranges from 32% (management) to 53% (culture). This lexicon procedure is the transparent baseline; a fine-tuned transformer ABSA model is the natural production replacement and is interchangeable because all downstream steps consume only the s_a columns.

### *B. Validation against human sub-ratings*

Because the data carry human sub-ratings $r\_a \in \{1,\ldots,5\}$ that map one-to-one onto the five aspects, we validate s_a directly by correlating it with r_a at the review level. This is the gold-standard check that comparable studies usually defer to costly manual annotation.

### *C Firm–month panel and distress outcome*

We aggregate to firm × calendar month, taking the mean of each aspect sentiment, review volume, mean overall rating, and the share of negative-outlook reviews. We keep firm–months with at least three reviews and firms with at least 24 monthly observations. We define a *distress* state for firm i in month t when the three-month rolling mean of the overall rating falls at least 0.35 points below the firm's own trailing baseline (its lagged expanding median):

$$Distress\_it = 1[\ rating^{(3)}\_it \le baseline\_it - 0.35\ ] \quad (2)$$

An *onset* is the first month of a new distress spell. Crucially, the outcome is built from the *structured* rating field while the predictors are *free-text* aspect sentiment, so the two draw on different signals within the reviews; we discuss the residual shared-source concern in Section 8.

### *D Econometric specification*

We estimate a linear probability model with two-way fixed effects,

$$Distress\_it = \Sigma\_a\ \beta\_a \cdot s\_(a,i,t-\ell) + \alpha\_i + \gamma\_t + \varepsilon\_it \quad (3)$$

where $\alpha\_i$ are firm effects, $\gamma\_t$ are calendar-month effects, $\ell \in \{1,3\}$ is the lead, and standard errors are clustered by firm (Petersen, 2009). In Eq. (3), firm effects absorb persistent differences in how firms are rated; month effects absorb common macro shocks. We complement the regression with (i) an event study averaging aspect sentiment in a window around onsets and (ii) per-firm Granger-precedence tests (Granger,

1969).

## IV. RESULTS

### *A. The lexicon measure is a valid, if imperfect, proxy*

Table 2 and Figure 1 report the validation correlations. All five aspects show positive, statistically meaningful agreement with the human sub-ratings, with Pearson r from 0.42 (culture, career growth) to 0.52 (compensation). These are *moderate* correlations: the text signal clearly tracks human judgments but is far from a substitute for them, consistent with the known limitations of general-purpose lexicons in a specialized domain.

***Table 2. Validation of text-derived aspect sentiment against human sub-ratings (review level).***

| Aspect | Pearson r | Spearman ρ | n |
|---|---|---|---|
| Management | 0.456 | 0.446 | 237,752 |
| Work–life balance | 0.510 | 0.508 | 260,137 |
| Compensation | 0.517 | 0.518 | 307,396 |
| Culture | 0.422 | 0.375 | 353,539 |
| Career growth | 0.420 | 0.394 | 316,496 |

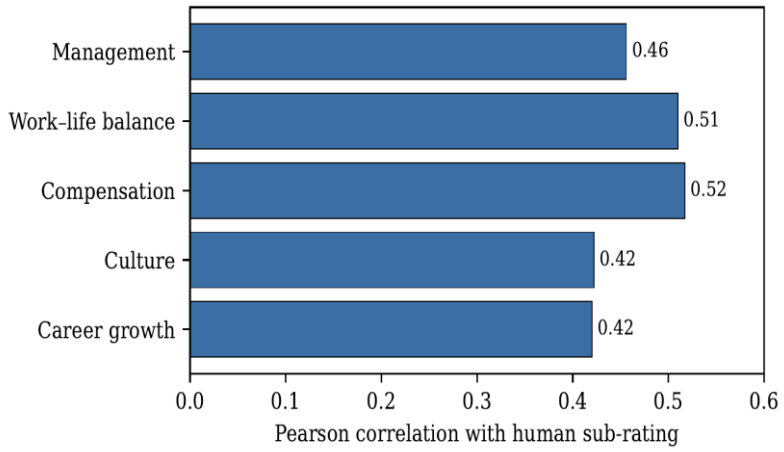


*Figure 1. Pearson correlation between the lexicon aspect sentiment and the platform's human sub-rating, by aspect. All correlations are positive and moderate (0.42–0.52).*

### *B. Descriptive topic structure*

A TF–IDF/NMF decomposition of the cons text recovers interpretable themes that align with the five aspects: work–life balance and "long hours"; "poor/senior management"; "low pay/salary"; and "working environment/pressure." This is descriptive context only and is not used in the inferential analysis.

### *C. Aspect sentiment leads distress at a one-month horizon*

Table 3 reports the panel estimates. At a one-month lead (Specification A), management (β=−0.063, p<0.001), culture (−0.102, p<0.001), and career-growth (−0.070, p<0.001) sentiment are significant negative predictors of next-month distress: when these aspects sour, distress becomes more likely. Compensation is marginal (−0.018, p=0.059) and work–life balance is insignificant. The signs are uniformly negative and the model is jointly significant (F=85.2), but the explained variation is modest (within-$R^2$ = 0.027), as expected for a noisy month-ahead binary outcome.

The three-month-lead specification (B) tells an important cautionary story: the coefficients collapse toward zero, none of the five is significant at conventional levels, two even change sign, and the joint fit is negligible (within-$R^2$ = 0.0001, F=2.9). The leading-indicator content is therefore **short-horizon**—roughly one month—and we do *not* find evidence of a usable three-month-ahead signal. Figure 2 visualizes this contrast.

***Table 3. Linear probability model of firm-level distress on lagged aspect sentiment (two-way fixed effects; firm-clustered standard errors in parentheses).***

| Dependent variable: | (A) One-month lead | (B) Three-month lead |
|---|---|---|
| Distress_it | s_a,t−1 | s_a,t−3 |
| Management | −0.063*** (0.010) | −0.011 (0.010) |
| Work–life balance | −0.009 (0.011) | −0.016 (0.010) |
| Compensation | −0.018† (0.010) | +0.018† (0.011) |
| Culture | −0.102*** (0.014) | +0.022† (0.012) |
| Career growth | −0.070*** (0.012) | −0.005 (0.010) |
| Firm FE / Month FE | Yes / Yes | Yes / Yes |
| Observations | 17,439 | 16,938 |
| Firms | 277 | 277 |
| Within $R^2$ | 0.027 | 0.0001 |
| Joint F | 85.2 | 2.9 |

p<0.01, **p<0.05, *p<0.1, †p<0.1 marginal. Coefficients are changes in the probability of a distress month per unit of aspect sentiment (range −1 to +1).

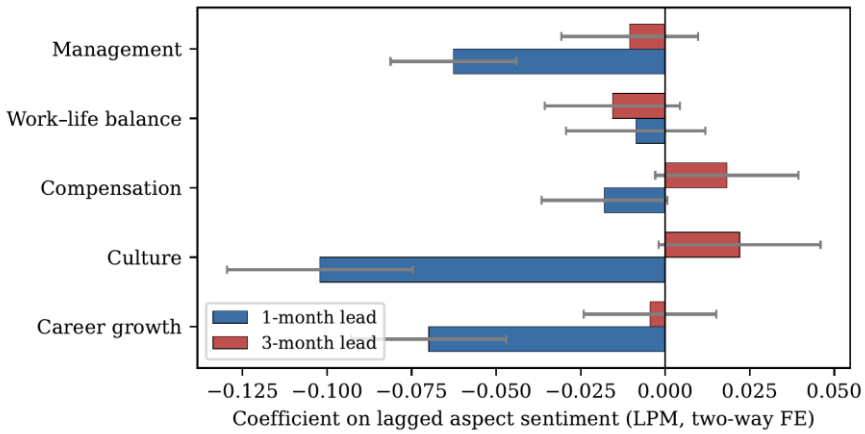


*Figure 2. Coefficients on lagged aspect sentiment with 95% confidence intervals. The one-month-lead effects (blue) are negative and precise for management, culture, and career growth; the three-month-lead effects (red) are indistinguishable from zero.*

### *D Event study around distress onset*

Aligning all 703 onsets at month 0 and averaging across events (Figure 3) shows the same short lead. Management sentiment is stable from month −6 to −3 (around 0.08–0.12), then falls at month −2 (0.03) and −1 (0.00) before reaching its trough at month 0 (−0.04); culture and career growth follow the same shape. The overall rating itself drops from 3.71 at month −3 to 2.67 at onset and then partially recovers. The dip in aspect sentiment thus *precedes* the rating trough by about one to two months, which is the visual analogue of the regression result and consistent with a short, not long, warning window.

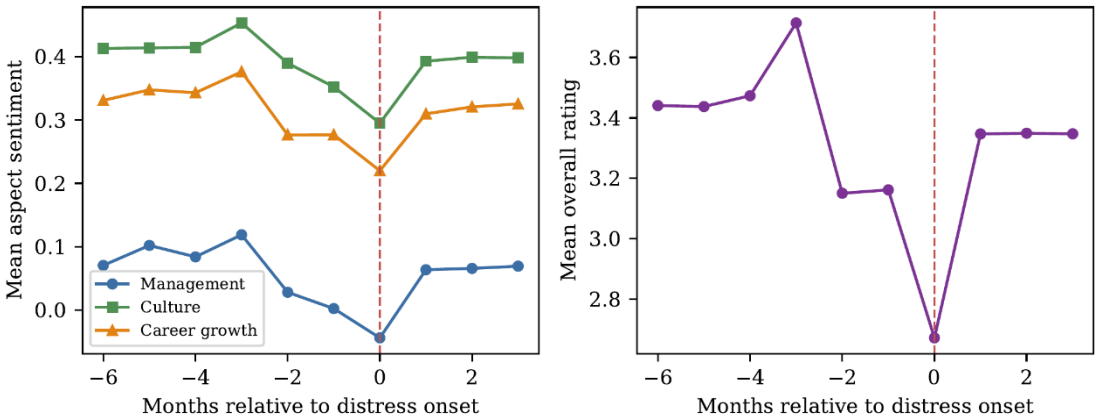


***Figure 3.*** *Event study around distress onset (month 0; dashed line), averaged over 703 onsets. Left: aspect sentiment; right: overall rating. Aspect sentiment begins falling at months −2/−1, ahead of the rating trough.*

### *E. Granger-precedence*

Across the 25 firms with the longest series, lagged management sentiment Granger-precedes the overall rating at $p<0.05$ for 14 firms (56%), including IBM, McDonald's, HSBC, Google, and American Express (minimum p-values ≤ 0.001). We interpret this as predictive precedence consistent with the panel result, not as evidence of causation.

## V Robustness

**Distress threshold.** Re-deriving the outcome at thresholds of 0.25, 0.35, and 0.50 rating points (Table 4, top) leaves management, culture, and career growth negative and significant at the 1% level throughout; magnitudes attenuate at stricter thresholds, as expected when events become rarer and more extreme.

**Alternative outcome.** We replace the rating-based outcome with an independent structured signal: a surge in the share of negative business-outlook reviews (three-month rolling share rising at least 0.10 above the firm baseline; 4,798 events). This outcome does not use the overall rating at all. The one-month-lead estimates (Table 4, bottom) remain negative and significant for management (−0.057), compensation (−0.041), culture (−0.046), and career growth (−0.051), with work–life balance again insignificant. The within-$R^2$ is smaller (0.006), but the corroboration across two differently constructed outcomes reduces the concern that the main result is an artifact of the rating-decline definition.

***Table 4. Robustness of the one-month-lead estimates.***

| *Panel 1: distress-threshold sensitivity (key coefficients)* | | | | | |
|---|---|---|---|---|---|
| Threshold | # events | Management | Culture | Career growth | |
| 0.25 | 2,385 | −0.075*** | −0.121*** | −0.078*** | |
| 0.35 | 1,599 | −0.062*** | −0.102*** | −0.070*** | |
| 0.50 | 833 | −0.044*** | −0.057*** | −0.027*** | |
| | | | | | |

| *Panel 2: alternative outcome — negative-outlook surge (n=17,439, within $R^2$=0.006)* | | | | | |
|---|---|---|---|---|---|
| | Management | Work–life | Compensation | Culture | Career growth |
| Coefficient | −0.057*** | −0.019 | −0.041*** | −0.046*** | −0.052*** |

Two-way fixed effects, firm-clustered SEs. ***p<0.01.

## VI. Conclusion

Several constraints bound our claims. First, the outcome is soft and internal: "distress" is a decline in the firm's own Glassdoor rating rather than an external hard event such as a layoff, bankruptcy, or earnings miss, and we make no claim that aspect sentiment predicts layoffs—the external test is infeasible here because the review window, ending in 2021-06, does not overlap the commonly used layoffs data from 2022 onward. Second, there is a shared-source association, since the predictors and the rating-based outcome both come from reviews; we mitigate this with a one-period lag, firm and month fixed effects, and an alternative outcome built from a different field (business outlook), but a mechanical component cannot be fully ruled out, and although the short-lead-only pattern is consistent with genuine dynamics rather than a static mechanical link, it is not dispositive. Third, there are measurement limitations: VADER is a general-purpose lexicon that is not tuned to employee-review language such as sarcasm, jargon, and domain idioms; aspect assignment is keyword-based; and aspect coverage is only 32–53% of reviews, so each aspect score is computed on a self-selected subset of text. Fourth, selection and generalizability are concerns, because we study firms with at least 100 reviews—large, visible employers—and reviews are written by a self-selecting set of (often departing) employees, so the results need not extend to small firms or to the population of all employers. Fifth, our inference is associational: Granger tests establish temporal precedence rather than causation, and we describe them as such throughout, while the distress-onset definition involves analyst choices of window, threshold, and baseline, of which we probe the threshold but do not exhaustively sweep all design choices. Taken together, these constraints frame our overall contribution: on 0.8 million Glassdoor reviews of 332 large firms, aspect-level employee sentiment—measured with a transparent, reproducible lexicon procedure and validated against human sub-ratings at moderate correlations—acts as a statistically robust but economically modest, short-horizon leading indicator of a within-platform distress signal, with management, culture, and career-growth sentiment as the informative aspects, predictive content concentrated at a one-month lead and disappearing by three months, and a pattern that is stable across distress thresholds and across an independent outlook-based outcome; we remain explicit about what the data do not allow, namely a validated link to external restructuring events, which would require a layoffs source overlapping the review window, so that replacing the lexicon with a fine-tuned transformer, obtaining a temporally aligned hard outcome, and adding placebo and employee-status splits are the priorities for turning this evidential study into a deployable early-warning system.

## VII. REFERENCES


1. Blei, D.M., Ng, A.Y., Jordan, M.I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993–1022. https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf

2. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *NAACL-HLT*. https://aclanthology.org/N19-1423/

3. Edmans, A. (2011). Does the stock market fully value intangibles? Employee satisfaction and equity prices. *Journal of Financial Economics*, 101(3), 621–640. https://doi.org/10.1016/j.jfineco.2011.03.021

4. Granger, C.W.J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, 37(3), 424–438. https://www.jstor.org/stable/1912791

5. Green, T.C., Huang, R., Wen, Q., Zhou, D. (2019). Crowdsourced employer reviews and stock returns. *Journal of Financial Economics*, 134(1), 236–251. https://doi.org/10.1016/j.jfineco.2019.03.012

**6.** Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv:2203.05794*. https://arxiv.org/abs/2203.05794

**7.** He, P., Liu, X., Gao, J., Chen, W. (2021). DeBERTa: Decoding-enhanced BERT with Disentangled Attention. *ICLR*. https://openreview.net/forum?id=XPZIaotutsD (arXiv: https://arxiv.org/abs/2006.03654)

**8.** Hutto, C.J., Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *ICWSM*. https://ojs.aaai.org/index.php/ICWSM/article/view/14550

**9.** Lee, D.D., Seung, H.S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401, 788–791. https://www.nature.com/articles/44565

**10.** Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool. https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf

**11.** Pang, B., Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135. https://doi.org/10.1561/1500000011

**12.** Petersen, M.A. (2009). Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches. *Review of Financial Studies*, 22(1), 435–480. https://doi.org/10.1093/rfs/hhn053

**13.** Pontiki, M., Galanis, D., Pavlopoulos, J., et al. (2014). SemEval-2014 Task 4: Aspect Based Sentiment Analysis. *Proceedings of SemEval-2014*. https://aclanthology.org/S14-2004/

**14.** Sull, D., Sull, C., Zweig, B. (2022). Toxic Culture Is Driving the Great Resignation. *MIT Sloan Management Review*, 63(2). https://sloanreview.mit.edu/article/toxic-culture-is-driving-the-great-resignation/