IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15416

PREDICTIVE ANALYTICS FOR WORKFORCE REDUCTION IN MULTINATIONAL CORPORATIONS: A LOGISTIC REGRESSION APPROACH

Dr. Vimal Kumar D[1], Mr. Arockiya John Prasanna X[2], Ms Hemalatha A[3]

Head of Department, Department of Computer Science, Rathinam College of Arts and Science, Coimbatore – 641021[1]

M.Sc. Data Science and Business Analysis, Rathinam College of Arts and Science, Coimbatore-641021[2]

Assistant Professor, Department of Computer Science, Rathinam College of Arts and Science, Coimbatore-641021[3]

Abstract: Corporate workforce restructuring has become one of the defining business challenges of our era. Over the past several years, multinational corporations across technology, finance, and manufacturing have carried out large-scale layoffs driven by a mix of macroeconomic headwinds, rapid automation, shrinking profit margins, and rising debt loads. Yet despite the frequency of these events, most organisations still lack any reliable early-warning system to flag layoff risk before decisions are locked in. This study introduces a data-driven approach to predicting corporate layoff events. Using a synthetic dataset of 1,500 firm-year observations spanning 2020 to 2026, we built a logistic regression model that draws on financial indicators, operational metrics, and macroeconomic data. The model was trained, validated through stratified cross-validation, and tested against a comprehensive set of classification metrics. We engineered several composite features including revenue decline flags, high-automation indicators, debt-to-equity thresholds, and an overall risk score that meaningfully improved predictive accuracy. We then applied the trained model to estimate layoff probabilities for ten prominent global companies in 2025–2026. The model achieved a ROC-AUC of around 0.82, with well-balanced precision and recall. Our aim is to offer a practical, reproducible methodology that organisations and policymakers can genuinely use for early workforce risk assessment.

Keywords: corporate layoffs, logistic regression, workforce analytics, predictive modelling, MNC risk assessment, automation adoption, financial indicators, ROC-AUC.

INTRODUCTION

Over the past five years, the world has watched multinational corporations shed hundreds of thousands of jobs in ways that felt both sudden and, in hindsight, foreseeable. What started as emergency cost-cutting during the economic disruptions of 2020 gradually shifted into deliberate structural reorganisation. By 2022 and 2023, companies in technology, finance, automotive, and telecoms were announcing redundancies at a pace that left employees, communities, and investors struggling to keep up. Amazon, Meta, Intel, and Boeing alone collectively cut more than 100,000 roles within eighteen months a stark illustration of how quickly workforce restructuring can accelerate.

Yet despite the sheer scale and regularity of these events, organisations rarely have any robust mechanism to anticipate them. Human resources teams, investors, and policymakers tend to react to announcements rather than prepare for them. The costs of this reactive posture are real: employees get little time to make alternative plans, local economies absorb sudden income shocks, and companies suffer lasting damage to their employer brand.

Machine learning can offer a different lens here. By training models on historical patterns of corporate financial health, automation investment levels, and macroeconomic conditions, it becomes possible to assign probabilistic risk scores to companies on a rolling basis. Logistic regression, though a relatively straightforward method, is particularly well suited to this task its coefficients are easy to interpret, it trains quickly, and its outputs naturally map onto a binary outcome: layoff or no layoff.

This study builds a complete, end-to-end prototype of such a system, covering everything from synthetic data generation through feature engineering, model training, evaluation, and forward-looking projections for 2025 and 2026. The code is written in Python using standard scientific libraries, and all results are reproducible from a fixed random seed. While the underlying dataset is synthetic, the feature distributions and outcome-generation logic are designed to reflect patterns well documented in the corporate downsizing literature.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15416

II. PROBLEM STATEMENT

When large corporations carry out mass layoffs, the ripple effects extend far beyond the individuals directly affected. Local economies lose income, communities lose stability, and national labour markets absorb disruptions that take years to fully work through. The data to anticipate these events financial filings, operational reports, market signals has long been available. What has been missing is a structured, data-driven way to actually use it.

Most current approaches rely on qualitative assessments, analyst commentary, or lagging indicators like quarterly earnings releases all of which arrive too late for any meaningful preventive action. The core challenge we address in this paper has three dimensions.

First, can internal financial metrics revenue trajectory, profit margins, debt levels be meaningfully combined with external macroeconomic indicators to create a useful early-warning signal for layoff risk? Second, which features actually carry the most predictive weight in separating financially vulnerable companies from stable ones?

Third, how well does a logistic regression model trained on historical data generalise to new, unseen companies and forward-looking scenarios?

Answering these questions requires a carefully constructed dataset that reflects the real heterogeneity of global corporate environments across industries, geographies, and time periods. The synthetic dataset we developed for this study was built with precisely this in mind, encoding realistic relationships between financial stressors and layoff outcomes while avoiding any reliance on proprietary or sensitive data.

III. PROPOSED WORK

The system we propose is a supervised binary classification pipeline that brings together financial feature engineering, logistic regression modelling, and risk-stratified forecasting into a single coherent workflow. We deliberately chose logistic regression over more complex alternatives. The goal was to build something decision-makers could trust and interrogate a system where the factors driving a risk score are visible and explainable, not hidden inside a black box.

The dataset covers 1,500 firm-year records from 2020 to 2026, drawn from 44 named multinational corporations across 8 industries and 10 countries. Each record includes 12 raw features: company size, revenue change, automation adoption score, profit margin, debt-to-equity ratio, prior layoff history, national unemployment rate, and technology investment as a share of revenue. From these, five engineered features are derived to sharpen the model's ability to detect layoff-prone firms.

The target variable a binary layoff flag is generated through a logistic function applied to a weighted combination of the raw features, with weights calibrated to reflect established findings from the corporate downsizing literature. A stochastic noise term is included to ensure that the relationship between features and outcomes is probabilistic rather than mechanically deterministic, which makes for a more realistic and challenging classification problem.

Fig. 1 - System Architecture Flow

Pipeline Stage
Raw Corporate Dataset (1,500 records, 2020–2026)
Data Pre-processing & Cleaning
Exploratory Data Analysis (EDA)
Feature Engineering (Risk Score, Binary Flags)
Train / Test Split (80% / 20%, Stratified)
Standard Scaler Normalisation
Logistic Regression Training
Model Evaluation (AUC, Accuracy, F1, CV)
2025–2026 Layoff Risk Forecast

Once trained, the model is applied to a curated set of ten real-world companies, using forward-projected feature values for 2025 and 2026. Risk levels are grouped into three bands: Low (below 40%), Medium (40–65%), and High (above 65%) giving end users an output that is actionable rather than merely statistical. The full dataset, enriched with predicted probabilities and risk labels, is exported for integration with tools like Power BI.

IV. METHODOLOGY

The methodology unfolds across six sequential phases, each feeding into the next: data synthesis, exploratory analysis, feature engineering, model development, evaluation, and forward inference. Figure 2 maps this pipeline visually.

Fig. 2 - Research Methodology Pipeline

Step	Phase
Step 1	Data Generation
Step 2	EDA & Visualisation
Step 3	Feature Engineering
Step 4	Model Training
Step 5	Evaluation & Tuning
Step 6	Prediction & Export

Data Synthesis

The synthetic dataset of 1,500 records was generated using NumPy's pseudo-random number generators, seeded at 42 to ensure reproducibility. Companies were sampled with replacement from a list of 44 global MNCs, with industries and countries drawn from predefined categorical sets. Continuous features were drawn from empirically motivated ranges: revenue change from −40% to +30%, profit margin from −20% to +35%, and debt-to-equity from 0.1 to 5.0. Automation adoption was represented as a normalised score between 0 and 1.

The binary layoff outcome was constructed by computing a weighted risk score across features, passing it through a logistic sigmoid to obtain probabilities, and applying a 0.5 threshold. Protective features company size, revenue growth, profit margin, and technology investment received negative coefficients, while amplifying features automation adoption, prior layoff history, and unemployment received positive ones.

Exploratory Data Analysis

EDA covered six dimensions: layoff rates by year, by industry, and by country; revenue change distributions stratified by outcome; automation adoption patterns by layoff status; and a full feature correlation heatmap. As expected from a uniformly sampled dataset, no strong temporal trend emerged from the yearly analysis. The industry breakdown did show somewhat higher average layoff rates in technology and telecommunications consistent with the faster pace of automation in those sectors. The correlation heatmap confirmed that automation adoption and prior layoff history were the strongest positive predictors of the layoff flag, while revenue change and profit margin showed the most pronounced negative correlations.

Feature Engineering

Five engineered features were derived from the raw inputs to help the model capture threshold effects that are difficult to represent with continuous variables alone. Figure 3 summarises the transformation logic

Fig. 3 - Feature Engineering Transformation Map

Raw Condition	Engineered Feature
Revenue_Change_Pct < 0	Revenue_Decline (Binary)
Automation_Adoption > 0.6	High_Automation (Binary)
Debt_To_Equity > 2.5	High_Debt (Binary)

DOI: 10.17148/IJARCCE.2026.15416

Raw Condition	Engineered Feature
Profit_Margin < 0	Negative_Margin (Binary)
Sum of above flags + Prev_Layoff	Risk_Score (0–5)

Revenue Decline flags any firm reporting negative revenue growth. High Automation identifies firms with an automation adoption score above 0.6, roughly corresponding to aggressive technology deployment. High Debt captures firms with a debt-to-equity ratio above 2.5, a commonly used marker of financial stress. Negative Margin flags loss-making firms. The composite Risk Score sums these four binary flags along with prior layoff history, yielding an ordinal variable from 0 to 5 that provides a concise, interpretable snapshot of overall vulnerability.

Model Development and Evaluation

The dataset was split into training (80%) and testing (20%) subsets using stratified sampling to preserve class balance across both partitions. Before fitting, all features were standardised using z-score normalisation a necessary step for logistic regression to produce well-calibrated and comparable coefficients. The logistic regression estimator was trained with a maximum of 1,000 iterations.

Model performance was assessed across five metrics: accuracy, precision, recall, F1-score, and ROC-AUC. Five-fold stratified cross-validation was run on the full dataset to estimate how well the model would generalise to new data. The confusion matrix was examined to check for systematic misclassification patterns, and feature importance was inferred from the magnitude and sign of the standardised coefficients.

V. RESULTS AND ANALYSIS

The logistic regression model delivered consistent, strong performance across every primary evaluation metric. On the held-out test set, classification accuracy reached 81.33%, with precision and recall in close alignment indicating that the model is not systematically biased toward either predicting layoffs or avoiding false alarms. The ROC-AUC of approximately 0.82 confirms that the model's discriminative ability is well above chance.

Table I - Logistic Regression Performance Metrics

Metric	Value
Accuracy	81.33%
Precision	0.8264
Recall	0.7988
F1-Score	0.8124
ROC-AUC	0.8201
5-Fold CV AUC	0.82 ± 0.02

Five-fold cross-validation returned a mean AUC of 0.82 with a standard deviation of just 0.02, indicating that performance is stable and not a fluke of one particular random split. The close alignment between the held-out test AUC and the cross-validated AUC gives us reasonable confidence that the model generalises well to new data.

Examining the standardised coefficients, automation adoption, prior layoff history, and the composite risk score all carried the largest positive weights. Revenue change, profit margin, and technology investment were the strongest negative predictors that is, firms with healthier financials were significantly less likely to be flagged. Company size, interestingly, had a near-zero coefficient. Once other financial indicators are accounted for, firm size alone tells us very little about layoff risk a finding that mirrors real-world patterns where large corporations have restructured just as readily as smaller peers.

The forward-looking predictions for 2025 and 2026 produced a risk profile that aligns well with publicly available reporting on the financial health of the selected firms. Intel, Boeing, and Ford each characterised by negative revenue trends, elevated debt, and a history of prior restructuring received the highest predicted probabilities. Google, Meta, and

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁝ Peer-reviewed & Refereed journal ⁝ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15416

Wipro with strong revenue growth, low debt, and high technology investment were classified as low risk. The full prediction output is shown in Table II.

Table II -Predicted Layoff Risk: 2025–2026

Company	Country	Year	Probability	Risk Level
Intel	USA	2025	82.1%	High
Boeing	USA	2026	79.4%	High
Ford	USA	2025	74.6%	High
Amazon	USA	2025	71.3%	High
Nokia	Finland	2026	68.7%	Medium
IBM	USA	2026	64.2%	Medium
HSBC	UK	2025	58.9%	Medium
Wipro	India	2026	42.1%	Low
Meta	USA	2025	38.6%	Low
Google	USA	2026	31.2%	Low

Looking at the 2025-2026 cohort as a whole, 40% of firms fell into the High-risk category, 30% into Medium, and 30% into Low. This distribution reflects the current macro environment elevated interest rates, slowing revenue growth in several key sectors, and continued automation investment all of which are contributing to a structurally elevated risk to global workforce stability.

VI. CONCLUSION AND FUTURE SCOPE

This paper set out to build a practical, transparent pipeline for predicting corporate layoff risk and the results suggest it is a goal worth pursuing. The logistic regression model achieved a ROC-AUC of approximately 0.82 and demonstrated stable performance across cross-validation folds, making it both methodologically sound and genuinely usable.

The contribution here is not primarily about algorithmic sophistication. It is about coherence: the study shows how a relatively simple statistical model, embedded in a structured workflow from dataset design through feature engineering to forward inference, can produce outputs that are actionable, interpretable, and accessible to non-technical decision-makers. The three-tier risk stratification Low, Medium, and High and the integration pathway with Power BI are designed precisely to bridge the gap between a data model and the people who actually need to act on it.

Several directions stand out for future development. The most impactful near-term step would be replacing the synthetic dataset with real financial filings from sources like SEC EDGAR, Refinitiv, or Bloomberg, which would allow the model to be validated against actual layoff announcements and test whether the feature importance patterns observed here hold in practice. Beyond that, the logistic regression baseline could be benchmarked against gradient-boosted models like XGBoost or LightGBM to assess whether additional predictive accuracy is achievable. Extending the static feature vector into a time-series framework using an LSTM or Transformer architecture would allow the model to capture the temporal evolution of corporate financial health, which the current approach does not.

NLP-based signals could also add significant value. Earnings call transcripts, analyst reports, and even social media sentiment can serve as leading indicators that precede formal restructuring announcements. Finally, any production deployment would require a fairness audit across countries, industries, and firm sizes to ensure that predictions are not systematically skewing outcomes for particular groups of workers.

Taken together, these extensions would move the system from a well-structured proof of concept toward something genuinely capable of supporting both corporate strategic planning and public labour-market policy.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15416

REFERENCES

[1] J. P. Hausknecht and J. A. Trevor, "Collective turnover at the group, unit, and organizational levels: Evidence, issues, and implications," Journal of Management, vol. 37, no. 1, pp. 352–388, 2011.
[2] W. F. Cascio, "Downsizing: What do we know? What have we learned?" Academy of Management Perspectives, vol. 7, no. 1, pp. 95–104, 1993.
[3] A. K. Mishra and K. E. Mishra, "The role of mutual trust in effective downsizing strategies," Human Resource Management, vol. 33, no. 2, pp. 261–279, 1994.
[4] K. S. Cameron, S. J. Freeman, and A. K. Mishra, "Best practices in white-collar downsizing: Managing contradictions," Academy of Management Perspectives, vol. 5, no. 3, pp. 57–73, 1991.
[5] D. Datta, J. Guthrie, D. Basuil, and A. Pandey, "Causes and effects of employee downsizing: A systematic review," Journal of Management, vol. 36, no. 1, pp. 281–348, 2010.
[6] T. H. Davenport and J. Harris, Competing on Analytics: The New Science of Winning. Boston, MA: Harvard Business Press, 2007.
[7] P. Tambe and L. Hitt, "The productivity of information technology investments: New evidence from IT labor data," Information Systems Research, vol. 23, no. 3, pp. 599–617, 2012.
[8] A. Ng, Machine Learning Yearning. DeepLearning.AI, 2018.
[9] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, 2nd ed. New York, NY: Springer, 2009.
[10] C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY: Springer, 2006.
[11] A. Brynjolfsson and A. McAfee, The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. New York, NY: W. W. Norton & Company, 2014.
[12] D. Autor, "Work of the past, work of the future," AEA Papers and Proceedings, vol. 109, pp. 1–32, 2019.
[13] M. Frey and C. Osborne, "The future of employment: How susceptible are jobs to computerisation?" Technological Forecasting and Social Change, vol. 114, pp. 254–280, 2017.
[14] L. Mishel and J. Bivens, "The zombie robot argument lurches on: There is no evidence that automation leads to joblessness or inequality," Economic Policy Institute, Washington, D.C., Tech. Rep., 2017.
[15] G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning. New York, NY: Springer, 2013.
[16] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
[17] W. McKinney, Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, 2nd ed. Sebastopol, CA: O'Reilly Media, 2017.
[18] J. D. Hunter, "Matplotlib: A 2D graphics environment," Computing in Science & Engineering, vol. 9, no. 3, pp. 90–95, 2007.
[19] M. Waskom, "Seaborn: Statistical data visualization," Journal of Open Source Software, vol. 6, no. 60, p. 3021, 2021.
[20] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, San Francisco, CA, 2016, pp. 785–794.
[21] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," The Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.
[22] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[23] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, 2002.
[24] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
[25] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," Nature, vol. 323, pp. 533–536, 1986.

