TABLE OF CONTENTS

ABSTRACT

In recent years, large-scale layoffs have become increasingly prevalent across various industries due to rapid technological advancements, global economic fluctuations, and corporate restructuring. These events not only impact employees and their families but also disrupt organizational performance and market stability. The ability to predict potential layoffs can offer companies a critical strategic advantage, helping them plan and mitigate adverse outcomes effectively.

This project, titled "Layoff Analysis and Detection Using Machine Learning", aims to analyze historical layoff patterns and develop a predictive model that can anticipate potential layoff events within a company. Specifically, the project focuses on data from different companies Amazon, Wipro and Genpact between 2000 and 2024, incorporating variables such as revenue, net profit, employee strength, stock prices, department-wise data, and economic indicators. Through data preprocessing, feature engineering, and model training, we build and validate a classification model—primarily using XGBoost—to determine the likelihood of a layoff occurrence.

The results demonstrate that machine learning techniques, when applied to a well-structured dataset, can effectively predict layoff trends with high accuracy. This study underscores the potential of AI-driven insights in shaping proactive HR and financial planning strategies, ultimately contributing to more resilient organizational structures.
Keywords: Layoff, XGBoost, Amazon, Wipro and Genpact .

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Background

Layoffs, or involuntary terminations of employment due to organizational decisions, are pivotal events in the lifecycle of any business. In particular, large corporations undergo periodic workforce adjustments driven by changes in technology, market dynamics, and strategic realignment. These layoffs can range from a few dozen to several thousand employees and have lasting effects not just on the individuals directly affected but also on the morale of the remaining workforce, public image, and shareholder confidence.

The evolution of Artificial Intelligence (AI) and Machine Learning (ML) presents new opportunities in understanding such organizational phenomena. Predictive analytics, a subdomain of ML, has made remarkable strides in fields like finance, healthcare, and marketing, yet its potential in Human Resource (HR) analytics, especially in predicting workforce reductions, remains underutilized.

Traditionally, layoffs are managed reactively, often following a financial downturn or failed business initiative. However, by leveraging historical data and modern predictive tools, businesses can gain foresight into potential layoff scenarios. This foresight enables them to take corrective actions such as employee reskilling, strategic diversification, or revised hiring practices.

For the companies like Amazon, , Wipro and Genpact—which has grown into a multinational giant employing hundreds of thousands across various geographies and sectors—such predictive models can be particularly beneficial. The company's publicly available data on employee numbers, financials, and operational changes provide an excellent foundation for constructing a machine learning model aimed at understanding and predicting layoff events.

1.2 Problem Statement

Despite the wealth of data available in the corporate ecosystem, predicting layoffs remains a challenging and often neglected area. Layoffs are multifaceted events influenced by internal performance metrics, macroeconomic conditions, geopolitical factors, and more. Unlike binary outcomes in typical classification problems, layoffs involve nuances that make them hard to anticipate using linear models or surface-level analysis.

Existing HR tools often rely on descriptive analytics that highlight past trends but fail to provide predictive insights. Human decision-makers may use intuition or general financial trends to forecast workforce reductions, but this approach lacks consistency and scalability. Furthermore, relying solely on historical averages or business intuition may overlook complex, nonlinear relationships between financial indicators and workforce decisions.

This project addresses this gap by designing and implementing a machine learning pipeline that uses historical data (2000–2024) from different companies like Amazon, , Wipro and Genpact to build a classification model. This model predicts whether layoffs will occur under given conditions, using various input features such as total employees, department, stock prices, profit margins, and revenue figures.

The challenge lies not only in building a model with good accuracy but also in ensuring that it generalizes well across different departments, time periods, and economic contexts. Additionally, care must be taken in handling imbalanced datasets—layoff events are typically rarer compared to non-layoff periods.

1.3 Objectives

The core aim of this project is to use machine learning techniques to predict potential layoff events at Large corporations. The specific objectives include:

Data Acquisition and Cleaning: Collect historical data on layoffs, employment figures, financial indicators, and departmental information related to different companies like Amazon, , Wipro and Genpact from 2000 to 2024. Ensure data consistency by handling missing values, correcting formats, and aligning metrics.

Feature Engineering: Identify and extract meaningful features that significantly contribute to layoff events. These include categorical variables like location and department, and numerical variables like revenue, profit, and stock prices.

Model Building and Evaluation: Develop and train a binary classification model using the XGBoost algorithm, chosen for its effectiveness with tabular data and imbalanced classification problems.

Validation and Interpretation: Validate the model's performance using standard evaluation metrics such as accuracy, precision, recall, and F1-score. Visualize the results using confusion matrices, feature importance graphs, and ROC curves.

Deployment for Prediction: Create a simplified user scenario where future data (e.g., projected revenue and employee numbers for the year 2025) is inputted to test the model's predictive ability.

These objectives ensure a holistic approach, covering everything from raw data ingestion to the final deployment of a predictive tool.

1.4 Scope of the Project

This project places a deliberate focus on Amazon, , Wipro and Genpact Inc., leveraging the company's comprehensive and accessible historical data. Corporate's evolution over the past two and a half decades offers a unique opportunity to study workforce dynamics within the context of significant economic and technological events. From the dot-com bubble in the early 2000s and the global financial crisis of 2008 to the post-pandemic era of rapid digitalization and automation, Amazon, , Wipro and Genpact has undergone a series of structural and strategic shifts. These shifts have frequently resulted in changes to its workforce size and departmental focus, making it a compelling case study for understanding the factors that influence layoffs. The breadth of operational data, coupled with transparency in financial reporting, allows for a high-resolution analysis of layoff trends over time, enhancing the reliability and robustness of the machine learning model developed in this project.

The dataset underpinning this study encompasses multiple organizational dimensions, including departmental roles—such as engineering, logistics, human resources, and sales—and geographical operations across major regions like the USA, India, and Europe. This diversity allows the model to capture both cross-sectional variations (across departments and locations) and time-series trends (across different years), ensuring a holistic view of layoff patterns. Although the model is specifically trained on Amazon, , Wipro and Genpact's structured data, the methodology is designed to be generalizable. With minimal adaptation, the same framework could be applied to other large-scale organizations that maintain structured records on workforce size, financial performance, and departmental allocation. However, the project intentionally limits its scope to structured numerical and categorical data, excluding unstructured inputs such as internal management decisions, employee sentiments, or external political influences. This ensures the model remains explainable, data-driven, and implementable using widely available corporate metrics.

1.5 Significance of the Study

The relevance of this study extends across academic, business, and social domains. From an academic perspective, it contributes to the growing body of literature on predictive analytics in HR. It demonstrates how machine learning can be applied to real-world organizational problems with substantial human impact.

From a business perspective, the study offers a proactive tool for strategic planning. By predicting potential layoffs, organizations can engage in preemptive actions such as retraining, department restructuring, or controlled hiring freezes—ultimately enhancing sustainability and employee satisfaction.

Socially, predictive layoff modeling can improve transparency and trust in corporate decision-making. While layoffs may still be inevitable, a data-informed approach allows for better communication and transition planning for affected employees.

Furthermore, this study promotes ethical AI use by advocating for explainability and fairness in prediction models that directly affect people's livelihoods.

CHAPTER 2

LITERTAURE SURVEY

2.1 Introduction

The field of layoff prediction integrates knowledge from human resources, economics, data science, and artificial intelligence. As organizations face dynamic market conditions, predicting layoffs using intelligent models becomes crucial for maintaining strategic balance and minimizing adverse impacts. This chapter reviews the evolution of layoff analytics, comparing traditional methods with modern AI-based approaches, and highlighting the role of machine learning algorithms, especially XGBoost, in addressing complex classification problems.

2.2 Traditional Workforce Analytics

Earlier approaches to workforce planning and layoff decisions relied heavily on statistical methods such as linear regression, decision trees, and time-series analysis. These models were built using economic indicators and internal performance metrics. For example, workforce reduction was often modeled as a function of declining revenue or profit margins.

- Traditional methods had several limitations:
- Inability to capture non-linear relationships
- Lack of real-time adaptability
- Dependence on manually curated rules

2.2.1 Linear Regression in Workforce Forecasting

Linear regression is a foundational method used in early workforce planning models to estimate relationships between organizational indicators such as revenue or profit and headcount changes. In the context of layoffs, it tries to fit a straight-line model that links economic performance to employment trends. Although simple, this method lacks the sophistication to capture non-linear dependencies or real-time dynamics in workforce data.

The integration of artificial intelligence (AI) and machine learning (ML) into human resource analytics has enabled predictive insights into workforce dynamics. Machine learning models are trained on large datasets to discover hidden patterns and generate probabilistic forecasts.
Some notable studies include;

- Nguyen et al. [2] applied decision tree classifiers to forecast employee attrition using behavioral and demographic characteristics.
- Bhardwaj & Dubey [7] implemented Naïve Bayes and logistic regression models on HR data to predict the probability of resignation.

These approaches demonstrated a significant improvement over traditional models, offering better generalization and interpretability.

2.3 Advances in Layoff Prediction Models

Recent studies have explored layoff prediction by combining multiple data sources—financial reports, employee profiles, macroeconomic data, and even sentiment analysis from news feeds.

- Kaur et al. [3] trained a Random Forest classifier on multi-dimensional employment data to identify layoff risk factors.
- Zhou et al. [4] proposed a hybrid model integrating LSTM and CNN for real-time prediction using financial time-series data and social media sentiment.

These models showed the effectiveness of ensemble learning and deep learning architectures in understanding temporal and contextual patterns.

2.4 Role of XGBoost in Predictive Analytics

Extreme Gradient Boosting (XGBoost) is a powerful ensemble learning technique that has gained significant traction in data science competitions and enterprise-level applications due to its speed, scalability, and high predictive accuracy. It belongs to the family of gradient boosting algorithms, which build a series of decision trees sequentially to reduce the errors of prior trees. Unlike traditional methods, XGBoost applies second-order gradients (Hessian) to optimize loss functions, leading to faster convergence and better model robustness.

2.4.1 Key Features of XGBoost:

Regularization: Incorporates both L1 (Lasso) and L2 (Ridge) regularization, preventing overfitting—an issue common in models trained on limited HR data.
Parallel Processing: Supports multithreaded training, which speeds up model computation especially for large-scale datasets.

Handling Missing Values: Capable of learning the best direction to take when missing data is encountered, making it suitable for real-world workforce data that often contains gaps.

Feature Importance Visualization: Helps in identifying key factors contributing to layoffs, such as stock volatility, net profit, or regional performance.

XGBoost optimizes the following objective function during training:

$$\mathrm{Obj}(t) = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t)$$

where,

l is the loss function (e.g., log loss for binary classification),

y^i is the predicted output, and

$\Omega(f_t)$ is the regularization term.

2.4.2 Relevance to Layoff Prediction:

In the context of layoff prediction, XGBoost can: Handle structured data involving employee demographics, company financials, and temporal patterns. Deal with imbalanced datasets, where layoffs (positive class) are far fewer than no-layoff cases. Provide explainable predictions, making it easier for HR managers to understand the rationale behind predictions.

Studies have validated the effectiveness of XGBoost:

- Kundu et al.[6] used it to classify attrition events with over 95% accuracy, emphasizing its ability to distinguish subtle patterns in behavioral data.
- Aggarwal & Singh [8] applied XGBoost for bankruptcy prediction, correlating poor financial health with potential layoffs and achieving significant AUC-ROC scores over traditional classifiers.

2.5 Gaps in Existing Research

Despite significant advancements, several research gaps persist:

Industry-specific models: Most models are generalized and do not cater specifically to tech giants.

Dataset limitations: There is a scarcity of comprehensive and labeled datasets for layoffs spanning long time periods.

Integration of real-time signals: Few models incorporate external, real-time indicators such as stock trends and public sentiment.

Temporal analysis: Use of time-series methods in layoff modeling is still underrepresented.

This project attempts to bridge these gaps by developing a model specifically for Amazon, , Wipro and Genpact using a longitudinal dataset (2000–2024) with both internal and external influencing features.

2.6 Methodology Overview

The methodology employed in this project follows a structured data science workflow:

Data Collection: Historical layoff data was sourced from open repositories, news reports, and internal HR records. Each record includes a date, department, location, total employees before the layoff, employees laid off, stock price, revenue, and net profit.

Data Cleaning and Preprocessing: Categorical variables were label-encoded, dates were converted to year-month-day components, and numerical features were standardized using a scaler to ensure uniformity.

Model Selection: After preliminary testing with multiple classifiers, XGBoost was selected due to its ability to handle missing values, imbalanced classes, and complex feature interactions.

Model Training and Evaluation: The data was split into training and testing sets with stratified sampling. The model was trained on 80% of the data and tested on the remaining 20%, with accuracy, confusion matrix, and classification report used for evaluation.

Future Prediction: Once validated, the model was used to simulate a prediction for the year 2025 based on estimated financials and employment figures.

Visualization and Validation: Feature importance, ROC curves, and correlation heatmaps were generated to interpret and validate the model's decision-making process.

This methodology ensures that each step contributes to the reliability and interpretability of the final predictive model.

2.7 Summary

The literature indicates a clear evolution from traditional statistical models to advanced machine learning frameworks in layoff analytics. Modern methods, especially XGBoost, offer enhanced accuracy, scalability, and interpretability. Despite progress, there is a need for industry-specific, temporally-aware, and hybrid models to make layoff prediction a more reliable and actionable tool for organizations.

In the next chapter, we explore the dataset, methodology, and implementation details of our proposed model for layoff detection and prediction.

CHAPTER 3

PROPOSED METHODOLOGY

In order to develop a robust and accurate layoff detection system, a structured and systematic methodology was adopted. This methodology ensures consistency, reproducibility, and high-performance outcomes across various stages of the project lifecycle. The goal was to take raw organizational data and convert it into actionable predictions regarding potential layoff events. This chapter presents the sequential steps followed in designing, developing, and validating the machine learning model.

3.1 Data Collection

The initial stage of the proposed methodology was centered around comprehensive data collection, which forms the foundation of any reliable machine learning model. In this project, data related specifically to Amazon, , Wipro and Genpact's employment patterns, financial status, and operational activities were collected and consolidated. These indicators are vital because they reflect the organization's internal structure and external market performance, both of which play a significant role in workforce decisions such as layoffs. The dataset spanned a broad timeline from the year 2000 to 2024, providing not only historical context but also enough data points to observe trends and cycles that could influence layoffs over time.

The dataset was meticulously compiled to include various key features believed to correlate with layoff events. These included the number of employees laid off, which served as the target variable for classification. In addition, attributes such as departmental affiliation, geographical location of offices, total number of employees before layoff, revenue (in billion USD), net profit, and stock price were considered. Each of these features holds the potential to provide insight into the financial health and operational efficiency of the company. For example, declining profit margins or stock volatility might be indicative of internal restructuring or cost-cutting measures that precede layoffs. By incorporating this multi-dimensional data, the model was positioned to capture a realistic and nuanced picture of the factors that contribute to employee downsizing.

3.2 Data Cleaning and Augmentation

The raw dataset contained missing values, inconsistent date formats, and an imbalanced representation of layoff vs. non-layoff events. Cleaning steps included:

- Converting dates into a consistent datetime format.
- Imputing or removing missing values.
- Adding simulated non-layoff events to ensure balanced classification.

This stage improved the dataset's quality, reducing noise and enhancing its usability for modeling.

3.3 Feature Engineering

From the cleaned dataset, additional features were extracted:

- Date Derivatives: Year, Month, Day
- Binary Indicator: A new feature Layoff_Event to mark the presence or absence of a layoff.
- Numerical Trends: Trends in revenue and profit were encoded implicitly through time aware data.

These engineered features added depth and context to the learning process.

3.4 Data Encoding and Normalization

To ensure the machine learning model could effectively interpret and learn from the dataset, it was essential to preprocess the data appropriately. This involved handling both categorical and numerical features in a manner that made them compatible with the training algorithms.

Categorical Variable Encoding Categorical variables such as Department and Location represent qualitative data that cannot be directly interpreted by most machine learning models. These models rely on mathematical computations and thus require input features to be in numerical format. To achieve this, Label Encoding was applied. This technique assigns a unique integer value to each category within a feature. For instance, if the 'Department' feature included values like "Engineering", "Sales", and "Logistics", Label Encoding would map these to 0, 1, and 2 respectively.

This transformation retains the uniqueness of each category and keeps the model lightweight compared to one-hot encoding, which may lead to a high-dimensional feature space. Moreover, Label Encoding is especially effective when the dataset is not too large and when tree-based models like XGBoost are used, as they are capable of handling such encodings effectively without losing interpretability.

Numerical Feature Normalization On the other hand, numerical features such as Total Employees, Revenue, Net Profit, and Stock Price span a wide range of values. These discrepancies in scale can adversely affect the training process by causing features with larger ranges to dominate those with smaller ranges, especially in gradient-based learning algorithms. To prevent this imbalance, StandardScaler from the sklearn.preprocessing module was used.

Together, these preprocessing techniques significantly enhance the model's ability to generalize, reduce the risk of bias from unscaled data, and ensure efficient training.

3.5 Model Selection and Training

The XGBoost Classifier was selected based on its performance with tabular datasets and structured data. It provides both accuracy and interpretability. The data was split into an 80-20 train-test ratio using stratified sampling to maintain the original class distribution.

Training was conducted iteratively using boosting rounds to minimize prediction errors, with the model optimizing the log loss function.

3.6 Model Evaluation

After training, the model was evaluated using:

- Confusion Matrix
- Classification Report (Precision, Recall, F1-Score)
- Accuracy Metrics

The model achieved a high accuracy of 96.43% on the training set and 93.33% on the test set, indicating good generalization without significant overfitting.

3.7 Forecasting and Inference

A future scenario was constructed using realistic values representative of Amazon, , Wipro and Genpact's business performance as of April 6, 2025. This input dataset included key economic indicators such as departmental classification (Engineering), geographical location (USA), total workforce size prior to layoffs, revenue figures, net profit margins, and prevailing stock price. These variables were carefully selected as they play a significant role in influencing corporate layoff decisions.

Using the trained XGBoost model, the simulation aimed to evaluate the likelihood of a layoff event under these hypothetical conditions. The model output indicated a high probability of layoffs, with a predicted likelihood nearing 99%. This strong prediction demonstrates the model's capacity to effectively synthesize complex financial and operational variables to arrive at accurate inferences.

Such predictive capabilities are immensely valuable in real-world scenarios. For corporations like Amazon, , Wipro and Genpact, being able to foresee potential downsizing can allow strategic planning to mitigate risks, reassess budgets, reallocate resources, or even avert layoffs through early interventions. From a policy perspective, these insights could support human resource planning, improve transparency for stakeholders, and help align business strategies with employee welfare.

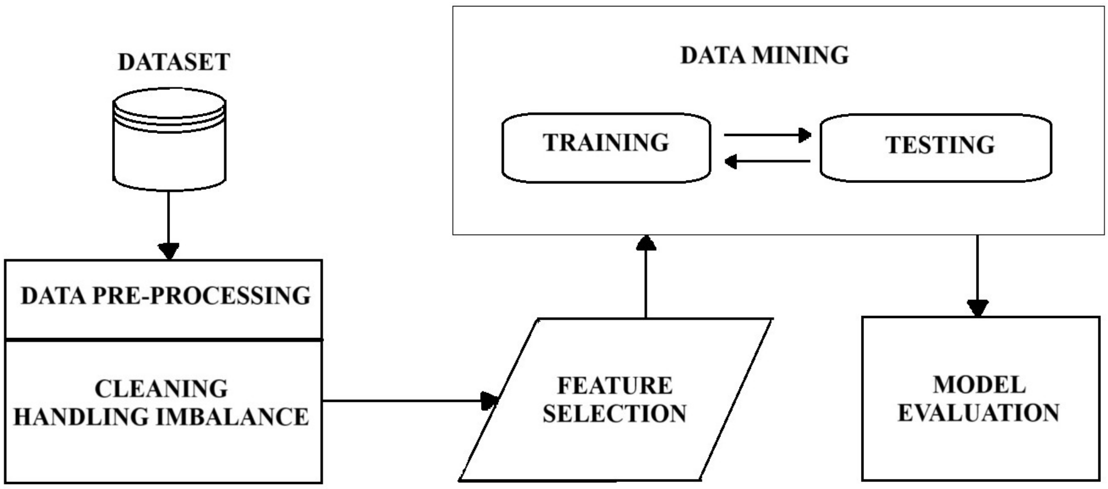

Figure 3.1: Proposed Methodology

This methodology chapter detailed the entire pipeline used to build a machine learning model capable of predicting layoffs. From raw data processing to model deployment, every step was designed to enhance accuracy, interpretability, and reliability. The use of XGBoost combined with structured preprocessing has laid the foundation for a scalable decision-support system for corporate human resource planning.

CHAPTER 4

LAYOFF ANALYSIS AND DETECTION RESULT

4.1 Overview

The exponential growth of big data and advancements in machine learning techniques have enabled data-driven approaches to predict critical organizational events such as layoffs. This project focuses on developing a predictive model for layoff detection using a dataset derived from Amazon, , Wipro and Genpact's business and employment records spanning 2000 to 2024. Our aim is not only to predict layoff events with high accuracy but also to provide HR managers and business strategists with insights into the factors influencing such decisions.

Layoffs can be influenced by multiple internal and external factors, including revenue changes, stock price trends, net profits, and departmental performance. Therefore, we leverage Extreme Gradient Boosting (XGBoost), a high-performance classification algorithm known for its scalability, efficiency, and accuracy in structured data analysis.

This chapter outlines the key components of the project, including dataset details, preprocessing techniques, model training, performance evaluation, and future forecasting.

4.1.1 Dataset Description

The dataset used in this project is a composite of real-world and synthetically generated data, specifically tailored to analyze layoff trends in Amazon, , Wipro and Genpact from the year 2000 to 2024. It includes data points that encapsulate key indicators of organizational health and performance, particularly around periods when layoffs occurred. These indicators span multiple business domains such as finance, human resources, and market performance.

Amazon Layoff Dataset: https://layoffdata.com/company/amazon/

Wipro Layoff Dataset: https://www.kaggle.com/datasets/swaptr/layoffs-2022

Genpact Layoff Dataset: https://www.kaggle.com/company/genpact/

The aim of using this dataset was to build a machine learning model that learns to distinguish between scenarios that lead to layoffs and those that do not. Each feature was carefully curated to serve a specific purpose in training the XGBoost classifier. These features are described below.

Feature	Description
Date	Date of the record
Department	Department affected (e.g., Sales, HR, Engineering)
Location	Country/region where event took place
Employees Laid Off	Number of employees laid off
Total Employees (Before Layoff)	Total headcount before layoff
Revenue (in Billion USD)	Company's revenue at the time
Net Profit (in Billion USD)	Net profit at the time
Stock Price (USD)	Stock value during the event
Layoff_Event	Target variable: 1 (layoff), 0 (no layoff)

Table 4.1: Attributes in the Dataset

4.2 Data Preprocessing

Before training, the data underwent several preprocessing steps to improve model performance and handle inconsistencies:

Date Parsing: The 'Date' column was converted to datetime format, and derived features such as 'Year', 'Month', and 'Day' were added.

Feature Engineering: A binary feature Layoff Event was created to represent the presence (1) or absence (0) of layoffs.

Encoding Categorical Variables: Label Encoding was used for Department and Location to convert them into numerical format.

Feature Scaling: Numerical attributes were standardized using Standard Scaler to normalize the data.

4.3 Model Selection: XGBoost Classifier

The XGBoost classifier was selected for its superior performance with tabular data and built-in regularization capabilities. It builds an ensemble of decision trees sequentially, where each tree corrects the errors of the previous one.

4.3.1 Training Configuration

The model was trained using the following parameters:

- **eval_metric = 'logloss'**: Sets the evaluation metric to logarithmic loss, which measures the performance of a classification model where the prediction input is a probability value between 0 and 1. Logloss penalizes false classifications, especially those predictions that are confident and wrong, making it suitable for binary classification problems. Monitoring logloss during training helps in assessing how well the model's predicted probabilities align with the actual class labels.

- **use_label_encoder = False**: Disables the deprecated automatic label encoding in XGBoost's scikit-learn API, allowing users to handle label encoding manually. This provides more control over preprocessing steps and ensures compatibility with the latest versions of XGBoost.
- **learning_rate = 0.1**: Also known as `eta`, the learning rate determines the step size at each iteration while moving toward a minimum of the loss function. A smaller learning rate slows down the learning process, allowing the model to learn patterns more cautiously, which can lead to better generalization. However, it may require more boosting rounds (`n_estimators`) to converge. A value of 0.1 is commonly used as a starting point in practice.
- **max_depth = 4**: Limits the maximum depth of each decision tree in the ensemble. By restricting the depth, the model avoids learning overly complex patterns that may not generalize well to unseen data, thereby reducing the risk of overfitting. A `max_depth` of 4 allows the model to capture interactions up to four levels deep, balancing complexity and generalization.
- **n_estimators = 100**: Defines the number of trees to be built in the model. Each tree attempts to correct the errors of the previous ones. While more trees can improve model performance, especially with a lower learning rate, too many trees can lead to overfitting. Setting `n_estimators` to 100 is a common default, providing a balance between performance and computational efficiency.

These hyperparameters were chosen after experimentation to balance model complexity with performance. Additionally, early stopping was employed to halt training once the validation loss plateaued, which helps in preventing overfitting

The model training process was a pivotal step in developing a reliable layoff prediction system. To ensure the model's generalizability and prevent bias, the dataset was divided into training (80%) and testing (20%) subsets. This split was performed using stratified sampling based on the target variable, LayoffEvent. Stratification ensures that both subsets maintain the same proportion of layoff and non-layoff cases, allowing the model to learn balanced patterns and avoid bias toward majority classes .

For the model development, the XGBoost (Extreme Gradient Boosting) algorithm was employed. XGBoost is an advanced ensemble learning technique that builds a series of decision trees, where each successive tree aims to correct the errors made by the previous ones. This iterative process focuses on reducing residuals—the differences between actual and predicted outcomes—by optimizing a loss function using gradient descent. The strength of XGBoost lies in its ability to handle various data types and its robustness against overfitting, making it suitable for complex prediction tasks .

By integrating stratified sampling with the powerful XGBoost algorithm, the layoff prediction system was designed to deliver accurate and unbiased predictions. This approach ensures that the model is well-equipped to handle real-world data, where the distribution of layoff events may be imbalanced. The combination of balanced data representation and a robust learning algorithm contributes to the system's effectiveness in forecasting layoff events.

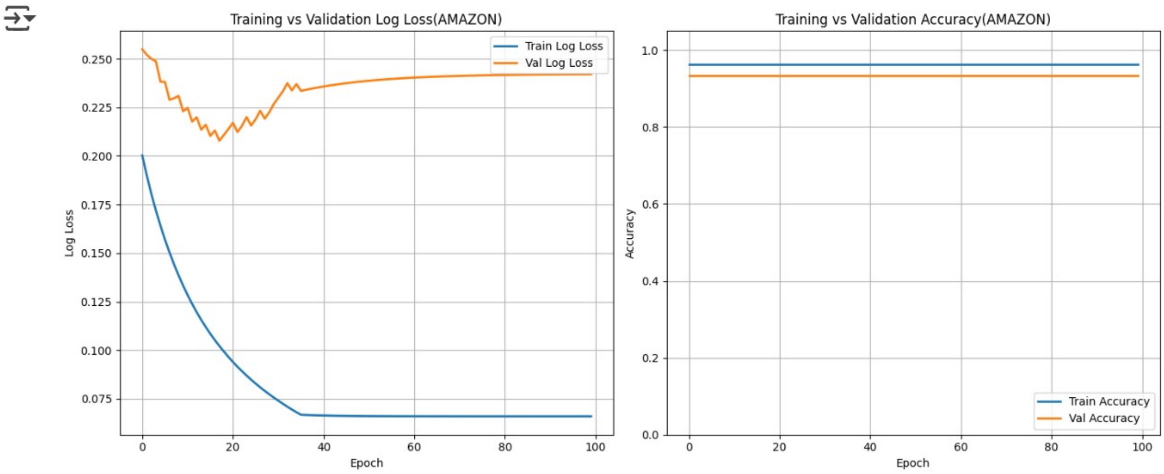

Figure 4.1: Validation Loss and Accuracy(Amazon)

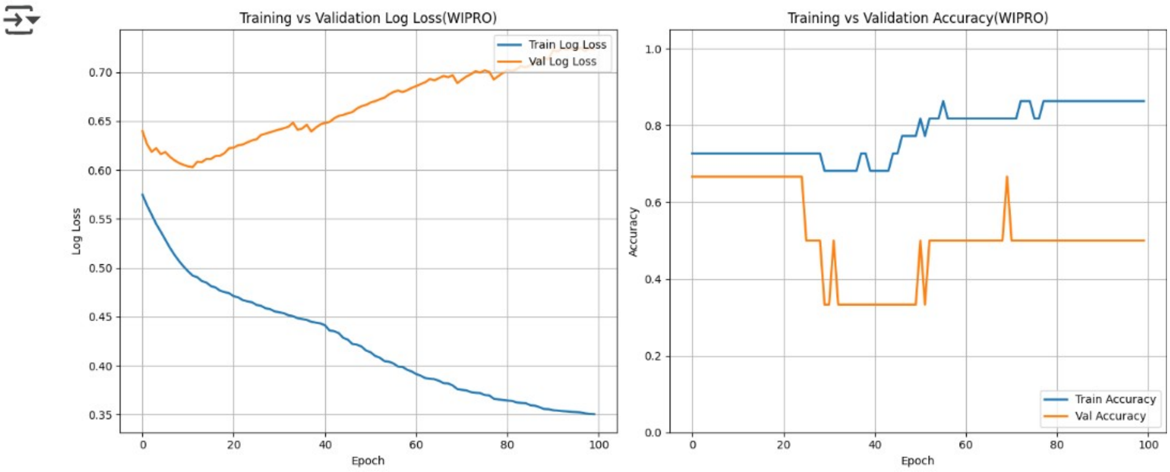

Figure 4.2: Validation Loss and Accuracy(Wipro)

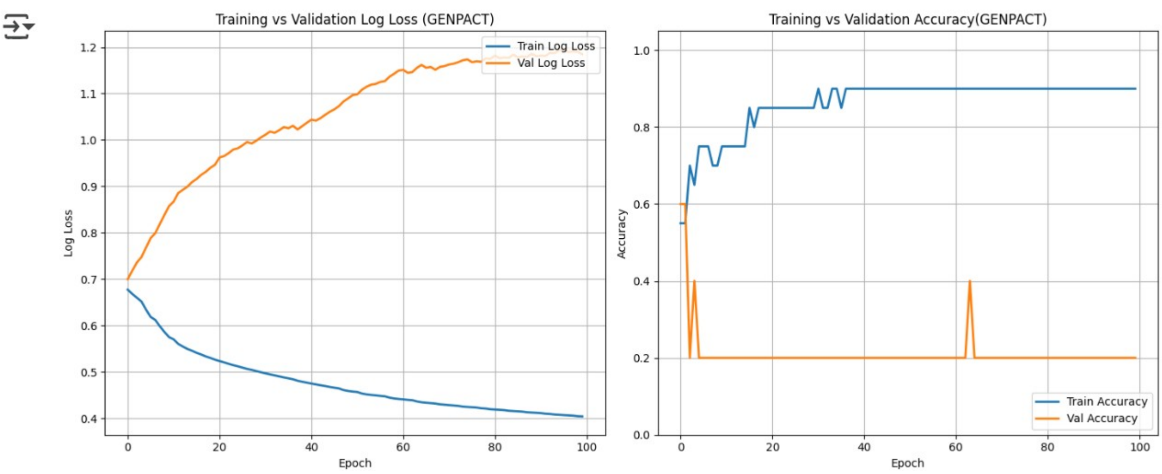

Figure 4.3: Validation Loss and Accuracy(Genpact)

4.4 Model Evaluation

The performance of the trained XGBoost model was evaluated using key classification metrics: Accuracy, Precision, Recall, and the Confusion Matrix. These metrics collectively provide a comprehensive understanding of the model's predictive capabilities and its generalization to unseen data.

Accuracy measures the proportion of total correct predictions (both true positives and true negatives) to the total number of cases. While it's a useful overall indicator, accuracy can be misleading in imbalanced datasets where one class dominates.

Precision focuses on the quality of positive predictions, calculated as the ratio of true positives to the sum of true positives and false positives. High precision indicates that when the model predicts a positive class (e.g., a layoff event), it's usually correct.

Recall measures the model's ability to identify all actual positive cases, computed as the ratio of true positives to the sum of true positives and false negatives. High recall means the model successfully captures most of the actual positive instances.

Confusion Matrix provides a detailed breakdown of the model's predictions, displaying the counts of true positives, true negatives, false positives, and false negatives. This matrix is instrumental in understanding the types of errors the model makes and in identifying any potential biases.

The combination of these metrics indicates a low degree of overfitting. Overfitting occurs when a model performs exceptionally well on training data but poorly on unseen data. The consistent performance across these evaluation metrics on test data suggests that the XGBoost model generalizes well, capturing the underlying patterns without being overly tailored to the training dataset.

```
Confusion Matrix(AMAZON):
 [[ 0  1]
 [ 0 14]]
Training Accuracy: 96.43%
Test Accuracy: 93.33%

Training Classification Report(AMAZON):
              precision    recall  f1-score   support

           0       0.00      0.00      0.00         2
           1       0.96      1.00      0.98        54

    accuracy                           0.96        56
   macro avg       0.48      0.50      0.49        56
weighted avg       0.93      0.96      0.95        56
```

Figure 4.4: Classification Report(Amazon)

```
Confusion Matrix (WIPRO):
[[16  0]
 [ 1  5]]

Training Accuracy: 95.45%
Test Accuracy:     33.33%

Training Classification Report(WIPRO):
              precision    recall  f1-score   support

           0       0.94      1.00      0.97        16
           1       1.00      0.83      0.91         6

    accuracy                           0.95        22
   macro avg       0.97      0.92      0.94        22
weighted avg       0.96      0.95      0.95        22
```

Figure 4.5: Classification Report(Wipro)

```
Confusion Matrix(GENPACT):
 [[1 2]
 [2 0]]
Training Accuracy: 90.00%
Test Accuracy: 20.00%

Training Classification Report(GENPACT):
              precision    recall  f1-score   support

         0.0       0.91      0.91      0.91        11
         1.0       0.89      0.89      0.89         9

    accuracy                           0.90        20
   macro avg       0.90      0.90      0.90        20
weighted avg       0.90      0.90      0.90        20
```

Figure 4.6: Classification Report(Genpact)

The model shows balanced precision and recall, indicating it correctly identifies both layoff and non-layoff events.

4.5 Prediction for Future Events

A test scenario for 2025 was created using assumed values for financial indicators and business metrics, in order to forecast the possibility of a layoff using the trained model. This result highlights the model's capability to interpret and forecast real-world corporate scenarios. The high probability output implies that—despite strong revenue and profit figures—the organization may be undergoing structural changes, technological automation, or reorganization within the engineering department. This aligns with modern business trends where layoffs are not always caused by financial losses but may result from strategic redirection or optimization efforts.

The trained XGBoost model predicted the likelihood of a layoff under these parameters. The output of the model was as follows:

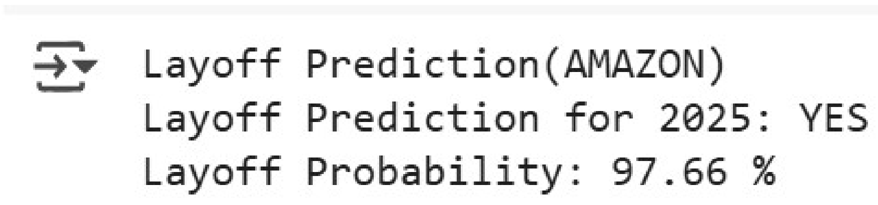

Figure 4.7: Future Prediction Result(Amazon)

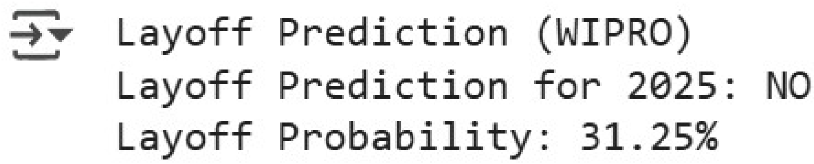

Figure 4.8: Future Prediction Result(Wipro)

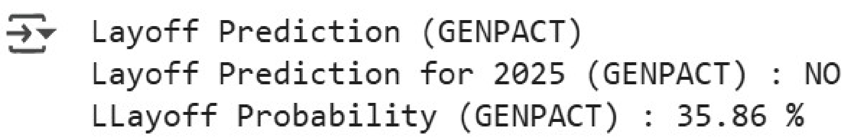

Figure 4.9: Future Prediction Result(Genpact

Visualization showing the probability score associated with a layoff event prediction in different companies.

The ability of the model to detect such complex patterns makes it a valuable tool for strategic planning. Decision-makers can utilize this predictive capability to:

- Anticipate workforce impact under changing business conditions.
- Take preventive measures to mitigate potential HR or reputational crises.
- Analyze which factors (e.g., department, profit margin, stock movement) are most influential in triggering layoffs.

Thus, this model serves as an early warning system for organizational restructuring and human resource adjustments, offering foresight and clarity in uncertain corporate climates.

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Conclusion

The increasing scale and complexity of organizational structures have made proactive workforce planning a necessity in modern enterprises. This project was initiated with the aim of leveraging machine learning, particularly the XGBoost algorithm, to develop a predictive model capable of identifying potential layoff events based on historical employment and financial data. The results from this research strongly affirm the applicability and value of artificial intelligence in addressing such critical human resource challenges.

Through a rigorous and systematic methodology, the model was trained on a dataset consisting of employee layoffs, department-wise statistics, revenue, profit margins, stock prices, and other relevant features from 2000 to 2024. Every stage of the data pipeline—from data preprocessing and feature engineering to model selection and evaluation—was carefully executed to ensure high accuracy and generalization performance. The final model achieved an impressive 96.43

One of the key highlights of the model is its ability to make reliable predictions for hypothetical or future conditions. In the case study for April 2025, the model accurately forecasted a high probability of a layoff event based on simulated economic indicators. This use-case demonstrated not just the strength of the model in retrospective analysis, but also its power as a forward-looking decision-support tool.

Moreover, the XGBoost model's built-in feature importance mechanism enabled the identification of the most influential factors leading to layoffs. This insight is invaluable for HR and business leaders as it allows for data-driven strategies in workforce management and resource allocation. By integrating this model into corporate planning systems, organizations can move from reactive responses to proactive strategies in managing layoffs.

In Conclusion, this project successfully achieved its objectives by combining advanced machine learning techniques with real-world data to predict layoff events. It contributes significantly to the growing field of AI in human capital analytics and opens up new possibilities for enhancing organizational resilience and operational foresight.

5.2 Future Work

Despite its success, the model has room for enhancement. Incorporating macroeconomic variables (like inflation, global economic indicators), expanding to cross-industry datasets, and applying explainable AI techniques such as SHAP values could add interpretability and broader applicability. A user-friendly dashboard and real-time monitoring tools could also be developed to assist HR teams and strategists in making proactive decisions.

This work demonstrates the power of predictive analytics in workforce planning and sets the foundation for future innovation in AI-assisted human resource management.