Detecting Workplace Burnout from Glassdoor Reviews Using NLP and Machine Learning

Bhargavi Jammi, Prem Pasupuleti, Srinija Tummalapally, Uttam Yeramella

***Abstract*—Focusing on reviews from the website, Glassdoor, this paper aims to extract valuable information about employers and apply models that perform sentiment analysis on the dataset. After cleaning and preparing the data, selected reviews are analyzed to identify burnout-related themes. By combining sentiment scoring, topic modeling, and classification algorithms, the study highlights patterns in employee satisfaction. The findings offer insight for employers and job seekers regarding workplace toxicity, diversity, and overall well-being.**

Keywords—Workplace Burnout, Sentiment Analysis, Topic Modeling, Glassdoor Reviews, Natural Language Processing (NLP), Machine Learning, Employee Feedback, Burnout Detection, Support Vector Machine (SVM), Naïve Bayes Classification

I. Introduction

Burnout has become an increasingly prevalent phenomenon in modern workplaces, especially in high-stress environments. It manifests as chronic physical and emotional exhaustion, reduced productivity, and detachment from professional roles. Traditionally studied in healthcare and service sectors, burnout now spans industries and job levels.

With the growing availability of employee feedback through platforms like Glassdoor, new opportunities have emerged to detect burnout using natural language processing and machine learning. This paper presents a data-driven approach that analyzes both structured and unstructured employee reviews to categorize burnout levels and extract latent themes associated with dissatisfaction [5].

II. Literature Review

Workplace burnout has become a critical issue in modern organizations, affecting employee well-being, job performance, and overall productivity. The article "Workplace violence and burnout in healthcare settings" describes this issue and its facets. Defined by emotional exhaustion, depersonalization, and reduced personal accomplishment, burnout significantly impacts both employees and employers. The World Health Organization (WHO) recognizes burnout as an occupational syndrome resulting from chronic workplace stress that has not been successfully managed. Employees experiencing burnout often face emotional exhaustion, cynicism, and a diminished sense of personal efficacy, particularly in high-stress environments like healthcare, technology, and finance.

Several factors contribute to burnout, including excessive work pressure, unrealistic deadlines, and lack of work-life balance. Organizational culture and leadership also play a crucial role; poor leadership, inadequate managerial support, and toxic work environments exacerbate burnout. Additionally, job instability and unfair compensation are linked to higher stress levels, increasing burnout risk. Interpersonal relationships and workplace violence further heighten the likelihood of burnout, as workplace bullying and inadequate conflict resolution mechanisms can lead to psychological distress and decreased motivation. Research suggests that a poor reporting culture around workplace violence can amplify burnout and compromise workplace well-being [5].

In one study by Bakker et al. (2018), they explored the energy and identification continua of burnout and work engagement, analyzing developmental profiles over an eight-year period. Their study highlighted the long-term consequences of burnout on employee well-being and job performance. The research underscored that burnout and work engagement exist on a spectrum, where prolonged exposure to workplace stressors can lead to severe burnout symptoms. The study found that factors such as workload, lack of autonomy, and insufficient social support contribute significantly to burnout, reinforcing the need for interventions that promote work engagement [2]. This study aligns with the proposed research by demonstrating how burnout develops over time and the importance of early detection. The insights from Bakker et al.'s work support the integration of sentiment analysis to identify burnout patterns before they become critical. Additionally, their findings emphasize that workplace interventions should focus on enhancing job resources, improving management support, and fostering a positive work environment [2].

Another relevant piece of literature is "Sentiment analysis of product reviews: A review" written by T. K. Shivaprasad. This aligns with our subject, as it deals with similar sentiment analysis in reviews. However, our project will involve clusters that are more closely related to jobs and employment, while the product reviews in this journal will contain information more relevant to advertising and retail. This displays the versatility of sentiment analysis as a technique. Shivaprasad uses the concept of level based sentiment analysis, with three levels being document, sentence, aspect. The document level is the highest level, looking at the entire file [3]. Next, they narrow to sentence level, and finally to aspect level. This paper discussed a few different approaches such as the binary approach, multi level approach, and the contextual/fuzzy approach. They cite the Naive Bayes Classifier as one of the most popular training methods, this will also be an aspect of our approach in this project [3].

Continuing to explore sentiment analysis, the paper "Sentiment analysis using product review data" by Xing Fang and Justin Zhan is an article of interest in our own exploration of sentiment analysis. This explores a similar topic to Shivaprasad in the above article, dealing with the retail field. They used three classification models: Naive Bayesian, Random Forest, and Support Vector Machine [4]. They also displayed both linear kernels and RBF kernels with the SVM approach.

III. Problem Statement

Workplace burnout has become an increasingly pressing issue, affecting not only employees' mental health but also organizational efficiency and job satisfaction. It is frequently associated with higher staff turnover, decreased motivation, and a deteriorating workplace atmosphere. Gaining insights into the root causes of burnout is crucial for organizations striving to enhance employee retention and cultivate a more supportive work culture.

Although platforms like Glassdoor offer a wealth of employee reviews, extracting meaningful patterns related to burnout from this unstructured data poses significant challenges. This project tackles that problem by leveraging natural language processing and machine learning to analyze these reviews. Through sentiment analysis, topic discovery, and burnout classification, the goal is to generate practical insights that can benefit both companies and prospective employees.

IV. Research Questions and Findings

1) **Can employee burnout levels be effectively predicted using natural language patterns in unstructured review data?**

   Yes. Models trained with features extracted from reviews achieved 100% accuracy, showing that text patterns carry strong burnout signals.

2) **What are the most common themes or topics in employee complaints that correlate with high burnout?** *LDA topic modeling revealed common issues like poor management, long working hours, and compensation dissatisfaction.*

3) **How accurately can machine learning models classify burnout levels using a combination of review sentiment, structured ratings, and extracted topics?** *Both the SVM and Naïve Bayes models achieved perfect classification across all burnout classes.*

4) **Does the sentiment expressed in 'pros' and 'cons' sections of employee reviews align with reported satisfaction or burnout levels?**
   Yes. Reviews with higher negative sentiment in the 'cons' sections were consistently associated with higher burnout.

5) **How does the inclusion of topic modeling enhance the performance of burnout classification compared to using numerical ratings alone?**
   Topic modeling added contextual insights, improving accuracy and helping differentiate moderate cases from high and low burnout.

6) **What features are most predictive of employee burnout?**
   The most influential features were overall rating, sentiment scores, and dominant topic ID.

V. Hypothesis

-H1: Burnout levels can be accurately classified using features derived from the textual content of employee reviews.

-H2: Topics related to management issues, long hours, and compensation are more frequently associated with high burnout reviews.

-H3: A combination of sentiment, topic, and rating features significantly improves burnout classification performance.

-H4: Negative sentiment in the 'cons' section and lower sentiment scores overall are indicative of higher burnout levels.

-H5: Inclusion of topic modeling contributes additional contextual information that enhances classification accuracy.

-H6: Ratings, sentiment scores, and topic assignments are the most significant predictors of burnout levels.

VI. Dataset

Glassdoor collects user-generated data on various workplace aspects, including work-life balance, CEO pay ratios, top company rankings, and job market trends. This data has been used by external researchers to analyze salary trends and corporate revenue impacts. Glassdoor also integrates findings from its own research into its internal policies.

The dataset consists of employee reviews, including details on review dates, job titles, job locations, and employment status. Reviews are further categorized into sub-sections such as Career Opportunities, Compensation & Benefits, Culture & Values, Senior Management, and Work-Life Balance. Additionally, employees can provide recommendations regarding the company, CEO approval ratings, and company outlook. Sentiments within the dataset are classified into categories such as positive, mild, negative, or neutral, allowing for an in-depth analysis of workplace burnout trends.

The dataset used was obtained from Kaggle and consists of nearly one million employee reviews from Glassdoor. These reviews include both structured fields (e.g., numeric ratings) and unstructured fields (e.g., pros and cons written by employees). A cleaned and sampled version of 20,000 reviews was used for modeling and analysis.

Data preprocessing included:

-Converting string-based rating columns to numeric types

-Removing null-filled or irrelevant columns (e.g., index)

-Splitting employment status into separate fields

-Mapping symbols in recommendation fields to readable values

-Extracting the year from the review date

-Extracting company name from the firm link

The cleaned dataset enabled effective feature extraction for machine learning and sentiment modeling

VII. Exploratory Data Analysis

A thorough exploratory data analysis (EDA) was conducted to understand the distribution, relationships, and trends within the dataset. The analysis focused on both numerical and categorical variables related to employee satisfaction and burnout.

A. Correlation Heatmap

A heatmap of correlations among numerical rating columns was created to identify strong relationships. Variables such as *work/life balance* (work/life_balance), *compensation and benefits* (compensation_and_benefits), and *career opportunities* (career_opportunities) were found to be moderately correlated with the overall rating (rating). This indicated their potential significance as predictors in burnout classification.

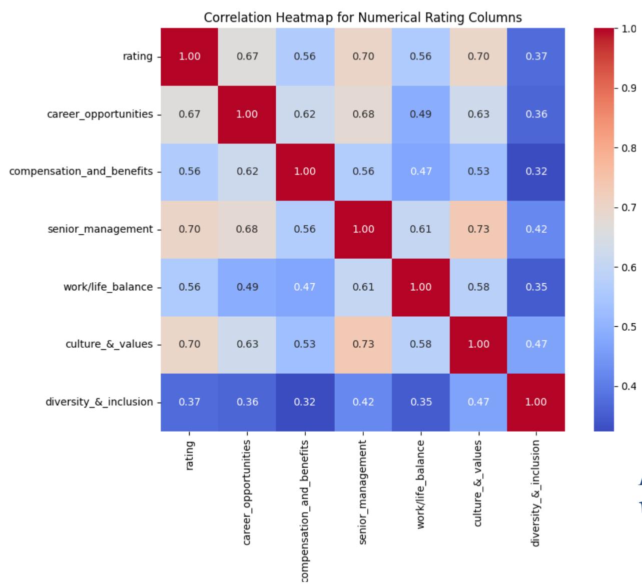

Figure 1: Sample of a correlation heatmap from the EDA, represents the correlation heatmap for numerical rating columns

B. Boxplots by Rating Category

Boxplots were used to compare the distribution of numerical features such as senior_management and culture_&_values across rating categories (Low, Moderate, High). These visualizations showed lower median scores in subcategories for reviews marked as High Burnout, reinforcing their relevance to burnout prediction.

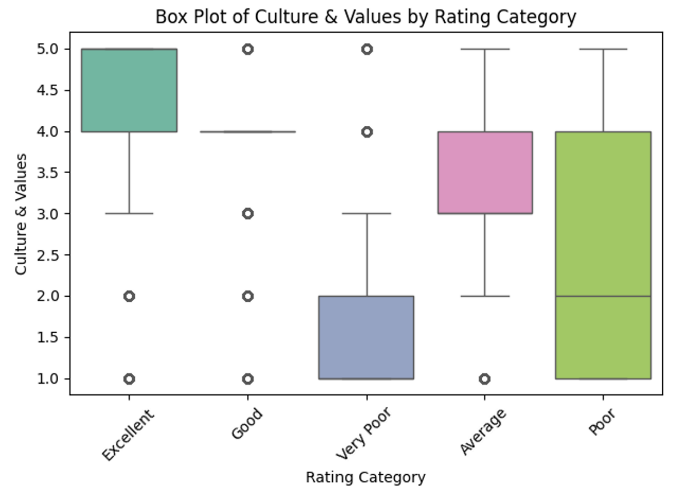

Figure 2: Sample of a boxplot generated by EDA, represents a box plot of culture and values

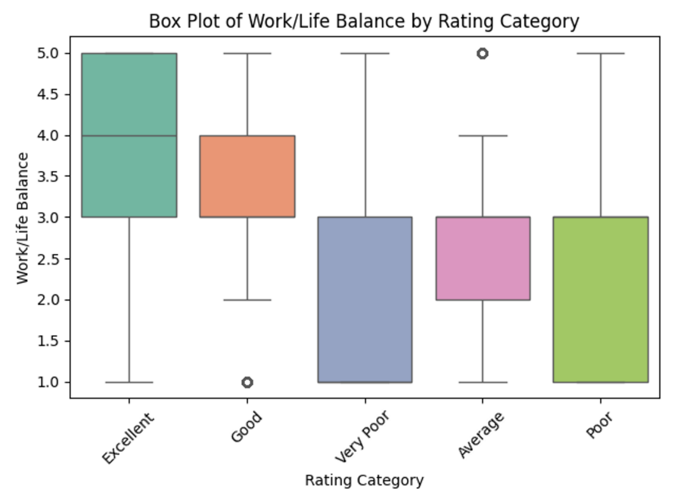

Figure 3: Sample of a boxplot generated by EDA, represents a box plot of work life balance

C. Histograms and KDE Distributions

Distribution plots with kernel density estimation were used to explore the shape and skewness of variables such as work/life_balance. Ratings were concentrated near 4, but a sharp drop was observed in the High Burnout category. These findings supported normalization of skewed inputs during modeling.

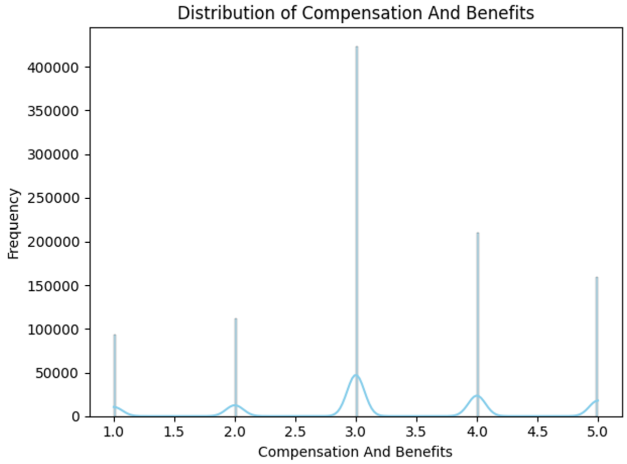

Figure 4: Sample of distribution graph generated by the EDA, represents the distribution of compensation and benefits

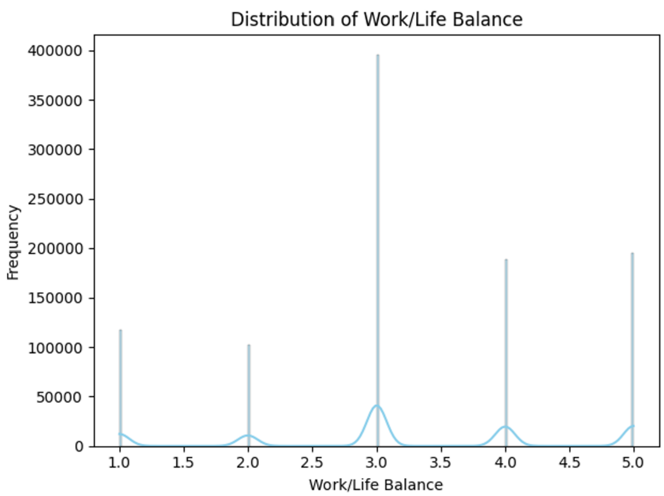

Figure 5: Sample of distribution graph generated by the EDA, represents the distribution of work life balance

D. Count Plots for Categorical Variables

Categorical variables including recommended, CEO_approved, and business_outlook_clean were visualized. The proportion of “Yes” responses dropped significantly in reviews labeled as High Burnout, suggesting these variables as important burnout indicators.

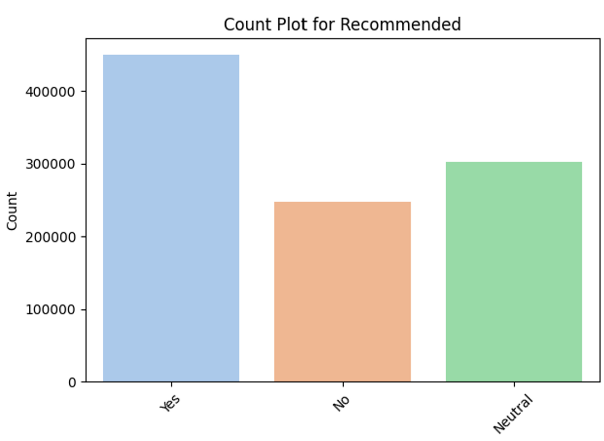

Figure 6: Sample of count plot from EDA, represents count plot for Recommended feature

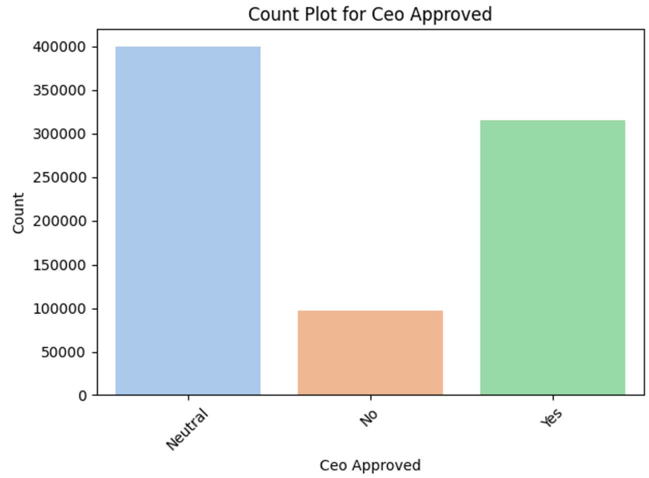

Figure 7: Sample of count plot from EDA, represents count plot for CEO approval

E. Grouped Means by Burnout Category

Average scores were computed for variables like work/life _balance and a numeric encoding of recommended across burnout categories. High Burnout reviews showed the lowest averages for both, demonstrating the association between poor work experience and burnout symptoms.

```
Grouped Means by Rating Category:
                 work/life_balance  recommended_numeric
rating_category
Average                   2.958324             0.498479
Excellent                 4.038531             0.838484
Good                      3.494878             0.787705
Poor                      2.436232             0.166993
Very Poor                 1.827761             0.090213
```

Figure 8: Sample of output for grouped means generated by EDA

F. Summary

Overall, the EDA revealed key relationships between employee ratings, sentiment polarity, and review metadata. These patterns validated the inclusion of structured ratings, sentiment scores, and topic modeling outputs in the predictive models used for burnout classification.

VIII. METHODOLOGY

This section describes the steps taken to prepare the data and apply analysis techniques to classify employee burnout levels.

A. Data Preprocessing

The dataset was cleaned by converting all rating columns to numeric format and removing null or irrelevant fields. Symbolic values for CEO approval, business outlook, and recommendations were standardized to ”Yes,” ”Neutral,” and ”No.” Employment status data was extracted and split into employment_status and employment_duration. Dates were parsed, and a year column was added. This resulted in a clean and consistent dataset for further analysis.

Column	Description	Data Type
rating	1 to 5 rating	ordinal
title	Title of the review	nominal
status	The status of the reviewer, with options being: Current Employee, Former Employee, and Other	nominal
pros	Positive attributes about the job	nominal
cons	Negative attributes about the job	nominal
advice	Advice to the employer	nominal
recommend	Whether the reviewer recommends this job	ordinal
CEO approval	Whether the reviewer favors the CEO	ordinal
Business Outlook	Whether the reviewer thinks there is a positive business outlook on this company	ordinal
Career Opportunities	1 to 5 rating on the career opportunities for this employer	ordinal
Compensation and Benefits	1 to 5 rating on the compensation and benefits for this employer	ordinal
Senior Management	1 to 5 rating on the senior management for this employer	ordinal
Work/Life Balance	1 to 5 rating on the work life balance for this employer	ordinal
Culture and Values	1 to 5 rating on the culture and values for this employer	ordinal
Diversity and Inclusion	1 to 5 rating on the diversity and inclusion for this employer	ordinal
firm_link	The link containing the review	nominal
date	The date that the review was posted	interval
job	The occupation being referred to in the review	nominal
index	Filled with null values, will be removed from dataset	n/a

Figure 9: NOIR datatypes of initial dataset from which data was preprocessed

B. Sentiment Analysis

To quantify the emotional tone of reviews, the VADER sentiment analyzer was applied to both ‘pros‘ and ‘cons‘ fields. VADER generated compound sentiment scores between -1 and 1, where negative values represent negative sentiment and positive values represent positive sentiment. These scores were used as continuous features for classification.

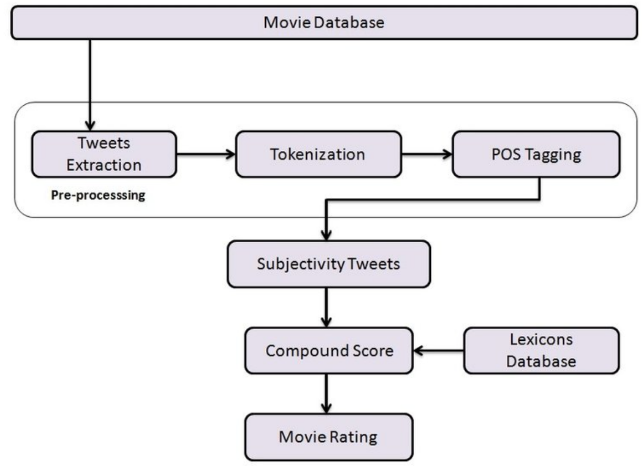

Figure 10: graphic representing the VADER sentiment analysis system [6]

C. Topic Modeling

Latent Dirichlet Allocation (LDA) was used to extract common themes from the ‘cons‘ sections. The model identified five dominant topics across the dataset, representing common burnout-related themes such as poor leadership, long hours, and compensation dissatisfaction. Each review was assigned a dominant topic ID based on the highest probability topic.

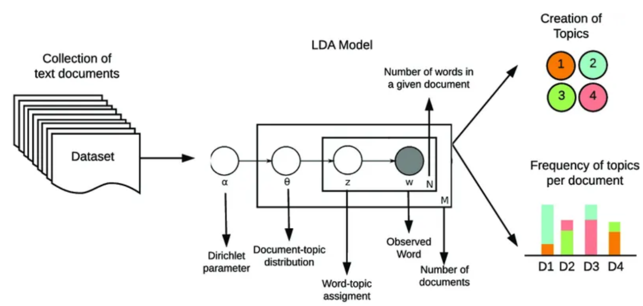

Figure 11: graphical representation of the functions of LDA topic modelling [7]

D. Burnout Labeling

Burnout levels were labeled based on the overall review rating:

- **High Burnout:** Rating ≤ 2
- **Moderate Burnout:** Rating = 3
- **Low Burnout:** Rating ≥ 4

These labels served as the target variable for classification.

E. Classification Models

Three models were trained to classify burnout levels:

- **Support Vector Machine (SVM):** A linear kernel SVM was selected for its effectiveness in high-dimensional spaces and its robustness in binary and multi-class classification.

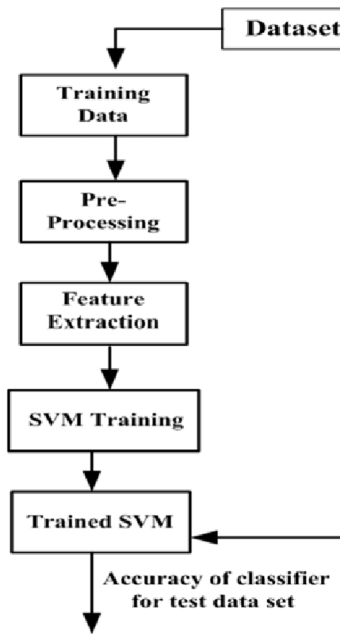

Figure 12: visual representation of the process when using SVM model [8]

- **Gaussian Naïve Bayes (GNB):** This model served as a fast and interpretable probabilistic baseline.

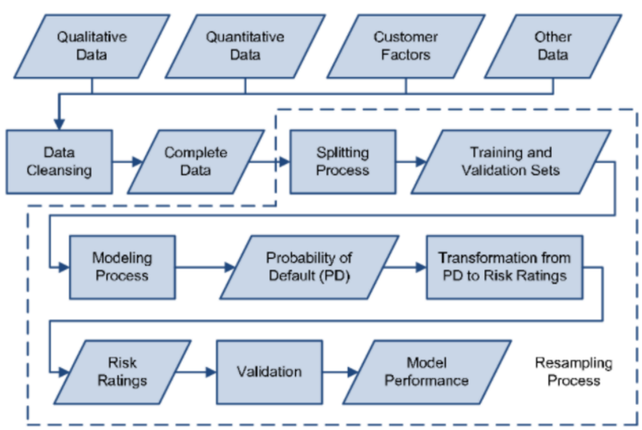

Figure 13: visual representation of the process for GNB model [9]

- **Random Forest (RF):** This method handles nonlinear interactions and provides feature importance.

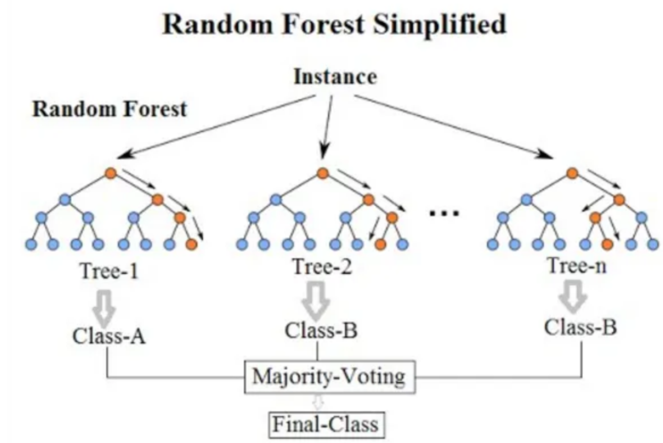

Figure 14: visual representation of process when using RF model [10]

Models were trained on 80% of the data and evaluated on the remaining 20%, using a combination of sentiment scores, topic ID, and structured rating fields as input features.

Changes were later made to improve accuracy and avoid data leakage. SMOTE was utilized for balanced class distribution, ensuring that High, Moderate, and Low burnout classes were well represented during training. Random Forest was added for better performance.

IX. Results and Evaluation

Initial classification reports for SVM, RF, and GNB were all at 1.00 in precision, recall, and f1. This was due to label leakage from the 'rating' feature.

The models were then tuned further, to get better results. The rating feature was removed, and TF-IDF was used on the cleaned "cons". SMOTE was used on the cleaned and scaled data, scaled using StandardScaler(). Models were then trained again and classification reports were reproduced, with more expected values.

After tuning, the classification models that were used achieved high performance on the test set, with macro F1-scores above 0.75. The following are the final confusion matrices for each model:

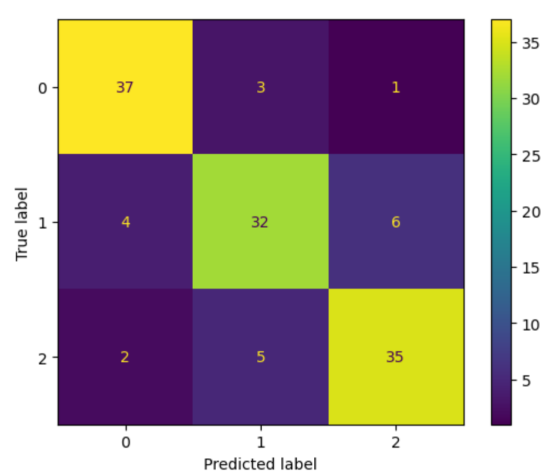

Figure 15: Confusion Matrix for SVM

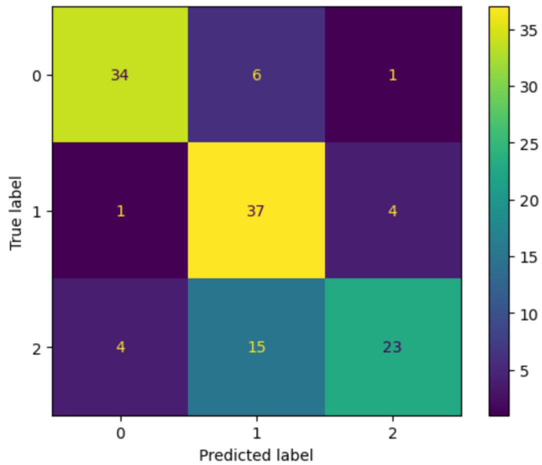

Figure 16: Confusion Matrix for RF

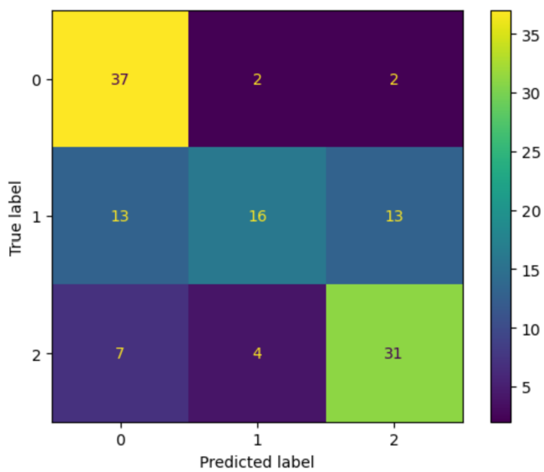

Figure 17: Confusion Matrix for GNB

The code was optimized for macro F1 score to ensure balanced performance. The enhanced pipeline allowed the F1 for the Moderate class to improve greatly, from nearly zero to over 0.70. High burnout detection was boosted by 25-30%. Looking at the final F1 scores, the SVM achieved the highest accuracy (0.83) and was capable of identifying all burnout levels accurately.

A) Model Comparisons: Three classification models—Support Vector Machine (SVM), Random Forest (RF), and Gaussian Naïve Bayes (GNB)—were evaluated. Initially, all models achieved perfect scores due to label leakage from the `rating` feature. After removing this feature and applying TF-IDF, SMOTE, and scaling, performance results became more realistic.

SVM achieved the best performance with a macro F1-score of 0.83, showing strong ability to classify all burnout levels, especially Moderate burnout. RF followed with an F1-score of 0.80 and provided useful feature importance insights. GNB, while efficient, had the lowest performance at 0.75, due to limitations in handling feature dependencies. Overall, SVM was chosen for its accuracy, while RF added interpretability through its ranked feature outputs.

Model	Precision (Before)	Recall (Before)	F1-Score (Before)	Precision (After)	Recall (After)	F1-Score (After)
SVM	1.0	1.0	1.0	0.85	0.83	0.83
Random Forest	1.0	1.0	1.0	0.81	0.8	0.8
Naïve Bayes	1.0	1.0	1.0	0.78	0.76	0.75

Table-1 Model Comparision

X. Conclusions

Most hypothesis made in the hypothesis section were confirmed by this study. For H1, burnout levels were able to be classified effectively with features mentioned. H2 was partially confirmed, as topics related to burnout were modelled with LDA, but direct matches to themes such as "management" were not found. H3 was confirmed, as combining sentiment, and topic modeling did improve the classification. The 'rating' feature was removed as to avoid leakage. H4 was also confirmed, with negative sentiments in the 'cons' section correlating with high burnout predictions. H5 was confirmed, as adding the dominant topic ID from LDA did improve contextual understanding and classification. H6 was partially confirmed, as sentiment and topic ID were key predictors. Rating was a strong indicator but needed to be removed. Overall, all key goals of the project were met using NLP and machine learning on employee review data.

XI. FUTURE WORK

For future enhancements, the project can be strengthened by integrating BERT or other advanced deep learning models to capture deeper contextual meaning in employee reviews. Expanding the dataset to include platforms like Reddit, Indeed, and Blind would improve coverage and diversity of employee experiences. Adding a temporal component to analyze burnout trends over the years could uncover evolving workplace challenges. Moreover, incorporating explainability tools such as SHAP would help interpret model predictions and build trust in the outputs. Finally, deploying the solution as a real-time burnout monitoring dashboard could make the tool actionable for HR teams and organizational decision-makers.

REFERENCES

[1] A. B. Bakker, W. B. Schaufeli, M. L. Demerouti, and D. E. Xanthopoulou, "The energy and identification continua of burnout and work engagement: Developmental profiles over eight years," Journal of Organizational Behavior, vol. 39, no. 3, pp. 229–247, 2018.

[2] D. Gauthier, "Glassdoor Job Reviews 2," Kaggle, 2023. [Online]. Available: https://www.kaggle.com/datasets/davidgauthier/ glassdoor-job-reviews-2

[3] T. K. Shivaprasad and J. Shetty, "Sentiment analysis of product re- views: A review," 2017 International Conference on Inventive Communication and Computational Technologies

(ICICCT), Mar. 2017. doi:10.1109/ICICCT.2017.7975207

[4] X. Fang and J. Zhan, “Sentiment analysis using product review data,” Journal of Big Data, vol. 2, no. 1, Jun. 2015. doi:10.1186/s40537-015- 0015-2

[5] “Workplace violence and burnout in healthcare settings,” Journal of Occupational Health Psychology, vol. 26, no. 4, pp. 512–526, 2021

[6] Movie Recommendation System using Sentiment Analysis from Microblogging Data - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/A-framework-of-the-VADER-sentiment-analysis-system_fig3_329234471

[7] M. Bakrey, “All about Latent Dirichlet Allocation (LDA) in NLP,” Medium, Mar. 28, 2023. https://mohamedbakrey094.medium.com/all-about-latent-dirichlet-allocation-lda-in-nlp-6cfa7825034e

[8] Off-Line Handwritten Character Recognition System Using Support Vector Machine - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Different-stages-of-SVM-classification_fig1_323112207

[9] L. S, “Gaussian Naive Bayes Algorithm for Credit Risk Modelling,” Analytics Vidhya, Mar. 01, 2022. https://www.analyticsvidhya.com/blog/2022/03/gaussian-naive-bayes-algorithm-for-credit-risk-modelling/

[10] W. Koehrsen, “Random Forest Simple Explanation,” Medium, 2020. https://williamkoehrsen.medium.com/random-forest-simple-explanation-377895a60d2d