Causal Drivers of Employee Burnout: A Bayesian Network Analysis

Angela Speranza, Ali Asadi Mohammadi
Master's Degree in Artificial Intelligence, University of Bologna
{ angela.speranza, ali.asadimohammadi }@studio.unibo.it

December 5, 2025

Abstract

This project aims to identify the primary causal drivers of employee burnout using Discrete Bayesian Networks. We utilized the `pgmpy` library to perform structural learning algorithms (Hill-Climb Search and PC), constrained by expert domain knowledge. We found that non-monetary factors—specifically Physical Activity and Work-Life Balance—exhibit a stronger causal influence on burnout risk than financial compensation, suggesting that lifestyle interventions may be more effective than salary increases in mitigating burnout.

Introduction

Domain

Workplace burnout is a pervasive occupational phenomenon characterized by exhaustion, cynicism, and reduced professional efficacy. Understanding the causal mechanisms behind burnout is crucial for developing effective HR strategies. This work models the probabilistic dependencies between an employee's working environment (e.g., Remote Work, Hours), demographics, and psychological outcomes. We approach this using Probabilistic Graphical Models (PGMs) to move beyond simple correlation and explore conditional dependencies and causal relationships in employee data.

Aim

The purpose of this project is to construct a Bayesian Network to:

- Discover the structural dependencies between job factors and mental health states using both score-based and constraint-based learning.
- Quantify the impact of specific variables on `BurnoutRisk` through sensitivity analysis.
- Perform causal inference to test the hypothesis that lifestyle factors (e.g., Sleep, Activity) are more critical determinants of burnout than fixed job attributes like Salary or Job Role.

Method

We analyzed the "Mental Health & Burnout in the Workplace" dataset (Kaggle), preprocessing it by discretizing continuous variables. Due to unbalanced labels, we utilized Decision Tree classifiers to optimize binning thresholds, prioritizing ROC AUC and Balanced Accuracy over standard accuracy. To ensure statistical validity and prevent overfitting, we enforced a constraint requiring each resulting bin to contain at least 5% of the total sample size. Prior to structural learning, we conducted a rigorous dependency analysis using Cramér's V (nominal), Spearman's Rho (ordinal), and Mutual Information (non-linear) to validate relationships and support our expert constraints. To define the network structure, we employed a hybrid approach using the `pgmpy` library. We compared the `HillClimbSearch` (using the AIC score) and the `PC` algorithm (using Chi-square independence tests). Crucially, we injected **Expert Knowledge** by defining a "blacklist" of forbidden edges to ensure temporal and logical consistency (e.g., *Stress* cannot cause *Age*; *Burnout* cannot cause *JobRole*). Parameters (CPTs) were learned using Maximum Likelihood Estimation (MLE) and Bayesian Estimators. Inference was performed using Variable Elimination.

Results

Our sensitivity analysis revealed that `PhysicalActivity` and `WorkLifeBalance` are the top drivers of burnout variance. "What-if" scenarios simulated via inference demonstrated that a shift from a sedentary to an active lifestyle reduces burnout risk significantly more than a transition to a higher salary tier, highlighting the importance of well-being over financial incentives.

Model

The final network consists of 12 nodes representing categorical random variables:

- **Demographics:** `Age` (4 bins: Young to Elder).
- **Job Context:** `JobRole`, `WorkHours` (Standard/Overworked), `RemoteWork`, `SalaryRange`.
- **Health & Psych:** `SleepHours`, `Stress` (Low/Mod/High), `PhysicalActivity`, `WorkLifeBalance`, `JobSatisfaction`, and `HasMentalHealthHelp` (aggregated).
- **Target:** `BurnoutRisk` (Binary: Yes/No).

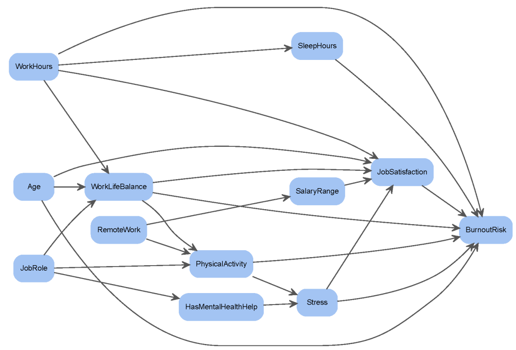

Figure 1: The learned Bayesian Network structure using HillClimb Search with Expert Constraints.

The structure was learned using the Hill-Climb algorithm with AIC scoring, which provided the best balance between model complexity and fit compared to the PC algorithm. We imposed strict constraints to prevent reverse causality: demographic nodes were forced to be root nodes or ancestors, while mental health states (Stress, Burnout) were restricted from influencing job characteristics. The CPTs were learned from the dataset, using a Bayesian Estimator to handle state combinations with sparse data points.

Analysis

Experimental setup

We utilized the network to perform Sensitivity Analysis and Causal Inference. For sensitivity, we calculated the "probability swing" for each feature: the difference in $P(BurnoutRisk = Yes)$ between the feature's best and worst states, holding other factors constant. For causal inference, we simulated intervention scenarios:

1. **The "Targeted Raise":** setting `SalaryRange` to the highest tier.
2. **The "Lifestyle Change":** setting `PhysicalActivity` to 'Active'. We evaluated the posterior probability of Burnout in each scenario to determine the most effective intervention.

Results

The analysis yielded the following key observations:

- **Top Drivers:** The sensitivity analysis ranked `PhysicalActivity` (**14.2%** swing) and `WorkLifeBalance` (**13.4%** swing) as the most influential features. Surprisingly, `SalaryRange` and `RemoteWork` showed significantly lower impact swings.
- **Scenario Comparison:** Increasing salary provided a negligible reduction in burnout risk (approx. 0.2% change). In contrast, increasing physical activity levels resulted in a massive variation (over 10%) in the predicted risk profile.
- **Tipping Points:** We observed a non-linear relationship with `Stress`, where shifting from 'Moderate' to 'High' stress caused a sharp increase in burnout probability, confirming stress as a direct parent node in the learned graph.
- **Vulnerability by Age**: Inference revealed a compounding biological effect. 'Elder' employees exhibited a heightened susceptibility to burnout (44.6% risk) compared to 'Young' employees (40.4%) when facing identical stressors like sleep deprivation.
- **The "Quiet Quitter" Effect**: Inference on dissatisfied employees revealed that remote workers demonstrated significantly higher probability of physical activity compared to on-site peers. This suggests remote work acts as a "lifestyle buffer," allowing employees to maintain health despite professional dissatisfaction.

Conclusion

We successfully modeled employee burnout using Bayesian Networks, demonstrating that the phenomenon is driven more by lifestyle and balance than by structural job attributes like salary or role. The model indicates that while high stress is a precursor, protective factors like physical activity and sleep are the most effective levers for mitigation. Limitations of this study include the discretization of continuous variables which may lose information, and the assumption of causal sufficiency in the provided dataset.

Links to external resources

- Burnout Dataset: `https://www.kaggle.com/datasets/khushikyad001/mental-health-and-burnout-in-the-workplace/data`
- Our GitHub repo: `https://github.com/speranzaaaa/Employee-Burnout-Bayesian-Network`