

# UNDERSTANDING EMPLOYEE TURNOVER INTENTIONS USING MACHINE LEARNING: A MULTI-FACTOR APPROACH

By

Jeya Shivanti Bharath

December, 2025

Directors of Thesis: Dr. Mark Bowler, and Dr. Nic Herndon

Major Departments: Computer Science, and Industrial/Organizational Psychology

## ABSTRACT

Employee turnover continues to challenge organizations and understanding the psychological and workplace factors that shape turnover intention is essential for effective retention efforts. This study integrates Industrial/Organizational Psychology with modern Data Science to examine how attitudes such as job satisfaction, organizational commitment, work engagement, perceived organizational support, perceived organizational justice, work environment, work life balance and employee net promoter score (eNPS) influence employees' intentions to leave.

Using survey data, four modeling approaches were compared, logistic regression, random forests, and a Bayesian Additive Regression Tree (BART) model, alongside an exploratory Automated Machine Learning (AutoML) pipeline that tested additional algorithmic configurations. Variable importance rankings from the machine-learning models were used to identify the most influential predictors contributing to turnover intention. Results show that organizational commitment, perceived support, and engagement are consistently the strongest drivers of turnover intention, and that machine-learning models, supported by AutoML exploration, uncover patterns that traditional methods often overlook.

By combining psychological theory with interpretable machine learning, this thesis provides a clearer understanding of turnover behavior and offers HR leaders practical, data-informed guidance for designing targeted retention strategies.



Understanding Employee Turnover Intentions using Machine Learning: A Multi-Factor  
Approach

A Thesis

Presented to the Faculty of the Computer Science and Psychology Departments

East Carolina University

In Partial Fulfillment of the Requirements for the Degree

Master of Science in Data Science and Master of Arts in Industrial/Organizational Psychology

By

Jeya Shivanti Bharath

December, 2025

Director of Thesis: Mark Bowler, PhD

Thesis Committee Members:

Nic Herndon, PhD

Alexander Schoemann, PhD

Venkat Gudivada, PhD

© Jeya Shivanti Bharath, 2025

## ACKNOWLEDGEMENTS

I would like to thank my thesis committee and faculty for their guidance and support throughout this research. I am deeply grateful to my family for their constant encouragement and assistance, especially my sister, whose belief in me meant more than words can express. I also thank my friends, both here and back home, for their encouragement and understanding throughout my graduate studies.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS .....	iii
LIST OF TABLES .....	vi
LIST OF FIGURES .....	vii
Chapter 1: Introduction .....	1
Chapter 2: Review of Literature .....	3
2.1 ORGANIZATIONAL COMMITMENT .....	3
2.2 JOB SATISFACTION .....	4
2.3 WORK ENGAGEMENT .....	6
2.4 JOB EMBEDDEDNESS .....	7
2.5 PERCEIVED ORGANIZATIONAL SUPPORT .....	9
2.6 PERCEIVED ORGANIZATIONAL JUSTICE .....	10
2.7 WORK ENVIRONMENT .....	11
2.8 CAREER GROWTH OPPORTUNITIES .....	12
2.9 WORK-LIFE BALANCE .....	14
2.10 EMPLOYEE NET PROMOTER SCORE (ENPS) .....	15
2.11 MACHINE LEARNING APPROACHES TO TURNOVER INTENTION PREDICTION .....	17
2.12 LOGISTIC REGRESSION .....	18
2.13 RANDOM FOREST .....	20
2.14 BAYESIAN ADDITIVE REGRESSION TREES (BART) .....	21
2.15 AUTOMATED MACHINE LEARNING (AUTOML) .....	23
2.16 HYPOTHESES .....	25
Chapter 3: Methodology .....	27
3.1 PARTICIPANTS .....	27
3.2 DATA OVERVIEW .....	28
3.3 MEASURES .....	29
3.3.1 Organizational Commitment .....	29
3.3.2 Job Satisfaction .....	29
3.3.3 Work Engagement .....	29
3.3.4 Job Embeddedness .....	30
3.3.5 Perceived Organizational Support (POS) .....	30
3.3.6 Perceived Organizational Justice .....	30
3.3.7 Work Environment .....	31
3.3.8 Career Growth Opportunities .....	31
3.3.9 Work-Life Balance .....	31
3.3.10 Employee Net Promoter Score (eNPS) .....	32
3.3.11 Turnover Intentions .....	32
3.4 PROCEDURE .....	33
3.5 DATA PREPARATION AND CLEANING .....	34
Chapter 4: Analysis Plan .....	36
4.1 OVERVIEW .....	36
4.2 EXPLORATORY DATA ANALYSIS (EDA) .....	37

4.3 PREDICTIVE MODELING APPROACHES .....	38
4.3.1 Logistic Regression .....	39
4.3.2 Random Forests.....	39
4.3.3 Bayesian Additive Regression Trees (BART) .....	40
4.3.4 Automated Machine Learning (AutoML) .....	40
4.4 MODEL EVALUATION .....	41
4.5 SENSITIVITY ANALYSIS .....	42
4.6 INTERPRETATION AND COMPARISON .....	43
<b>Chapter 5: Results .....</b>	<b>45</b>
5.1 DESCRIPTIVE STATISTICS .....	45
5.2 DEMOGRAPHIC CHARACTERISTICS .....	45
5.3 CONTINUOUS VARIABLES .....	46
5.4 CORRELATION PATTERNS .....	49
5.5 CORRELATION ANALYSIS .....	50
5.6 LOGISTIC REGRESSION RESULTS.....	52
5.7 RANDOM FOREST CLASSIFICATION RESULTS .....	54
5.8 BAYESIAN ADDITIVE REGRESSION TREES CLASSIFICATION RESULTS .....	55
5.9 H2O AUTOML RESULTS.....	57
5.10 MODEL COMPARISON SUMMARY .....	63
5.11 CALIBRATION ANALYSIS .....	66
<b>Chapter 6: Discussion .....</b>	<b>69</b>
6.1 SUMMARY OF KEY FINDINGS .....	69
6.2 PRACTICAL IMPLICATIONS .....	71
6.3 LIMITATIONS .....	74
6.4 FUTURE DIRECTIONS .....	75
<b>Chapter 7: Conclusion.....</b>	<b>78</b>
<b>REFERENCES .....</b>	<b>80</b>
<b>APPENDIX A: R Code for Data Processing and Analysis.....</b>	<b>95</b>
<b>APPENDIX B: IRB Approval Memo .....</b>	<b>96</b>

## LIST OF TABLES

5.1 Descriptive statistics for continuous study variables .....	54
5.2 Zero-order Pearson correlations among key study variables .....	57
5.3 Logistic regression predicting high turnover intension .....	61
5.3 Classification performance of models across alternative probability thresholds.....	66
5.5 Predictor importance rankings across logistic regression and machine-learning models	67
5.6 Comparison of model performance across predictive approaches .....	72

## LIST OF FIGURES

2.1 Conceptual model of hypothesized predictors of turnover intention.....	19
4.1 Analytical workflow for data preprocessing, model training, and evaluation .....	41
5.1 Distribution of turnover intention scores with 75th percentile binary cut-off.....	53
5.2 Correlation heatmap of key study variables.....	58
5.3 Random forest variable importance for predicting turnover intention .....	69
5.4 Receiver operating characteristic (ROC) curves comparing predictive models.....	73
5.5 Calibration curves for all predictive models.....	75

## **Chapter 1: Introduction**

Employee turnover has long been recognized as one of the most significant challenges faced by organizations. High rates of voluntary turnover incur substantial financial costs through recruitment, selection, onboarding, and training, while simultaneously weakening organizational memory, disrupting team functioning, and lowering morale among remaining employees (Jha, 2009; Hongvichit, 2015). When high-performing or highly skilled employees leave, the consequences are especially damaging, as such departures reduce organizational stability and erode competitive advantage (Jha, 2009). Because turnover intention is consistently shown to be the most immediate and reliable predictor of actual quitting behavior (Mobley, 1977; Hongvichit, 2015), understanding the factors that shape employees' intentions has become essential for proactive retention strategies.

Turnover is not driven by a single cause but rather reflects a multidimensional combination of psychological, social, and organizational forces (Hongvichit, 2015). Employees' perceptions of their work environment, opportunities for growth, fairness of treatment, and alignment between personal and professional needs all interact to shape their decision-making about staying or leaving. Recent advances in workforce analytics also highlight that turnover intentions can be effectively examined using data-driven approaches, allowing researchers to capture nonlinear relationships and complex interactions among predictors, patterns that may be difficult to detect using traditional methods alone. This underscores the importance of integrating established psychological theories with modern analytical tools to better understand turnover risk (Akasheh et al., 2023).

Accordingly, this study examines a range of well-supported predictors of turnover intention.

Organizational commitment and job satisfaction remain two of the strongest traditional correlates

of retention, reflecting employees' emotional attachment to the organization and their evaluative reactions to their job (Tett & Meyer, 1993). Work engagement and job embeddedness broaden this understanding by capturing the motivational energy employees bring to their roles and the connections that link them to their jobs and communities, factors shown to reduce the likelihood of exit (Takawira et al., 2014). Empirical meta-analyses and field studies show that higher perceived organizational support and stronger perceptions of organizational justice are consistently associated with lower turnover intention, suggesting that fairness, support, and recognition are central to employee loyalty and retention (Li et al., 2022).

The work environment also contributes substantially to day-to-day experiences that influence well-being and retention, while opportunities for career growth have been consistently associated with long-term commitment, especially in labor markets where development and advancement are highly valued (Mountasser & Sahraoui, 2025). Finally, this thesis incorporates two contemporary predictors: work-life balance and the employee net promoter score. Work-life balance has become increasingly salient as employees prioritize managing both personal and professional domains (Lestari & Margaretha, 2020). Meanwhile, eNPS, adapted from customer loyalty research, offers a practical indicator of employee advocacy and sentiment toward the organization (McPherson, 2025).

Taken together, these constructs provide a comprehensive framework for predicting turnover intention. By combining established psychological predictors with contemporary organizational measures, this study aims to advance theoretical understanding while leveraging modern analytical techniques to support more effective, data-informed retention strategies.

## **Chapter 2: Review of Literature**

### **2. 1 Organizational Commitment**

Organizational commitment is widely recognized as a significant psychological construct closely linked to employee turnover intentions. Broadly defined, organizational commitment refers to the strength of an individual's identification with attachment to, and involvement in their organization (Meyer & Allen, 1991). It is commonly conceptualized into three dimensions: affective commitment, continuance commitment, and normative commitment (Meyer & Allen, 1991; Iverson & Buttigieg, 1999).

Affective commitment consistently emerges as the strongest predictor of turnover intentions. This dimension reflects employees' emotional attachment to and identification with their organization, significantly correlating with reduced turnover intentions (Meyer & Allen, 1991). Empirical research has robustly supported this relationship, indicating that employees with higher affective commitment are less likely to leave, as their attachment aligns closely with their intrinsic motivations and personal goals (Tett & Meyer, 1993; Ozkan et al., 2020). Wasti (2003) further reinforced the predictive power of affective commitment, demonstrating its relevance across diverse cultural contexts. Similarly, Hur and Abner (2023) confirmed its strong predictive validity in public-sector organizations, which frequently experience high turnover rates.

Continuance commitment reflects employees' awareness of the costs associated with leaving the organization, such as financial loss, reduced job security, or forfeited benefits. When employees perceive these costs as high, they are more likely to remain, even if their emotional attachment is limited. Recent studies show that continuance commitment continues to predict

lower turnover intention in various occupational settings where economic dependence and limited alternatives shape retention decisions (Senayah & Biney-Aidoo, 2024; Opolot et al., 2024).

Normative commitment refers to a felt sense of obligation or moral duty to stay with the organization. Employees high in normative commitment often remain because they perceive staying as the “right” or responsible thing to do, frequently shaped by organizational support, investment in employee development, or cultural expectations of loyalty. Empirical research demonstrates that normative commitment is negatively associated with turnover intention and plays a meaningful role in retention, particularly in contexts emphasizing reciprocity and organizational support (Han & Cai, 2024; Rodríguez-Fernández et al., 2024).

The influence of organizational commitment, particularly affective commitment, has also been validated across various occupational contexts. For instance, Ozkan’s (2020) meta-analysis highlighted affective commitment as a critical factor in reducing turnover intentions among IT professionals, reinforcing its broad applicability. Stark et al. (2025) show that profiles high in affective commitment are associated with the lowest levels of harmful outcomes, reinforcing that strong emotional bonds with the organization contribute to retention and organizational stability.

Collectively, these findings highlight the importance of organizational commitment, especially affective commitment, as a fundamental determinant of turnover intentions and a critical consideration for organizational retention strategies.

## **2.2 Job Satisfaction**

Job satisfaction refers to employees’ subjective responses to various job-related environmental factors, encompassing both psychological and physiological dimensions (Hoppock, 1935). Despite the age of this definition, it continues to serve as a foundational

perspective in contemporary organizational research. It consistently emerges as a critical determinant of employee turnover intentions, with extensive research demonstrating a robust negative correlation between these two variables. Higher levels of job satisfaction significantly reduce employees' inclination to leave their organizations (Lee, Yang, & Li, 2017).

In a comprehensive study on early-career employees, Lee et al. (2017) identified job satisfaction as a multidimensional construct comprising salary and welfare, leader behavior, the nature of the work itself, interpersonal relationships, personal growth, and job competency. Their findings specifically highlighted personal growth as the most substantial factor influencing overall job satisfaction among early-career employees, with significant implications for reducing turnover intentions. These insights underscore the necessity for organizations to implement targeted interventions, particularly those that enhance opportunities for personal and career development, to effectively manage and decrease employee turnover (Lee et al., 2017).

Supporting these findings, Skelton, Nattress, and Dwyer (2019) conducted a multiple regression analysis among manufacturing employees in the Southeastern United States, confirming the strong negative relationship between job satisfaction and turnover intentions. Their research aligns with Herzberg's (1966) two-factor theory, distinguishing intrinsic factors (e.g., recognition, autonomy, meaningful tasks), which actively promote job satisfaction, from extrinsic factors (e.g., salary, job security), which primarily serve to prevent dissatisfaction. Skelton et al. (2020) further emphasized that organizational initiatives aimed at increasing job satisfaction through both intrinsic and extrinsic factors are essential strategies for managers seeking to retain their workforce effectively.

Further reinforcing this relationship, extensive meta-analytic research consistently highlights job satisfaction as a significant predictor of turnover intentions. Tett and Meyer's

(1993) seminal meta-analysis confirmed job satisfaction's pivotal role in an employee's decision-making process regarding whether to remain with their employer. Similarly, Ozkan et al. (2020) corroborated these findings, demonstrating a strong and consistent negative correlation between overall job satisfaction and turnover intentions across numerous organizational contexts, underscoring the critical role of both intrinsic and extrinsic job satisfaction components in influencing employee retention.

Empirical validation of job satisfaction's influence on turnover intentions extends across diverse industries. Within the hospitality sector, characterized by notably high employee turnover rates, job satisfaction proves particularly vital due to the demanding nature of roles and associated emotional labor (Park & Min, 2020). Likewise, in the IT sector, job satisfaction has consistently shown a strong negative association with turnover intentions, highlighting its generalizability across varied employment contexts (Ozkan, 2022).

In conclusion, job satisfaction remains a fundamental predictor of employee turnover intentions due to its profound impact on employees' perceptions of their work environment and overall organizational support (Lee et al., 2017; Ozkan et al., 2020; Skelton et al., 2020). Prioritizing and enhancing job satisfaction is thus critical for organizations seeking to effectively manage employee retention and ensure organizational stability.

### **2.3 Work Engagement**

Work engagement, defined as "a positive fulfilling work-related state of mind characterized by vigor, dedication, and absorption," significantly influences employees' attitudes and behaviors within organizations (Schaufeli et al., 2006b). Research consistently highlights a robust negative correlation between work engagement and employee turnover intentions. For instance, Memon et al. (2016) observed that employees who are satisfied with training

opportunities exhibit higher levels of work engagement, subsequently reducing their intentions to leave. Bothma and Roodt (2013) further reinforced this finding by demonstrating that higher engagement levels directly enhance employee retention, underscoring the importance of targeted organizational interventions designed to foster employee engagement.

Similarly, Alarcon and Edwards (2010) provided additional evidence, emphasizing that engagement dimensions such as dedication and absorption positively impact job satisfaction and negatively predict turnover intentions. Furthermore, Saks (2006) argued that work engagement is significantly influenced by job and organizational resources, suggesting that when employees perceive high levels of organizational support, they are more likely to reciprocate with increased engagement and reduced withdrawal behaviors. Hakanen et al. (2005) also demonstrated that job resources such as autonomy, social support, and performance feedback serve as critical antecedents to engagement, which in turn promote retention. Schaufeli (2017) further clarified that work engagement operates through motivational processes that energize employees and encourage persistence in goal-directed activities, providing an essential mechanism through which organizations can foster sustainable performance.

Collectively, these studies underscore work engagement as a critical motivational factor that plays a pivotal role in promoting employee retention and organizational effectiveness.

## **2.4 Job Embeddedness**

Job embeddedness refers to the collection of forces that keep employees connected to their jobs, including their perceived fit with the organization, the quality of their interpersonal relationships, and the sacrifices they associate with leaving (Mitchell et al., 2001). Research shows that employees who experience stronger alignment with organizational values, maintain meaningful workplace relationships, and believe they would forfeit valuable resources if they left

are less likely to consider quitting. For example, Karatepe (2016) found that frontline service employees with high levels of embeddedness demonstrated lower turnover intentions, partly because embeddedness supported greater emotional stability and enhanced satisfaction with work–family balance.

Given the focus of this thesis on workplace attitudes and turnover intention, the present study emphasizes the on-the-job component of job embeddedness (JE-On). JE-On includes organizational fit, links with coworkers, and anticipated work-related sacrifices. Prior meta-analytic research indicates that JE-On is the strongest and most consistent predictor of turnover intention, whereas off-the-job embeddedness (JE-Off) tends to be more closely associated with actual turnover behavior rather than intentions (Jiang et al., 2012). Because turnover intention is a psychological evaluation formed within the work context, JE-On represents the most theoretically appropriate and empirically supported dimension of embeddedness for inclusion in predictive analyses.

Additional literature underscores the multidimensional nature of embeddedness. Shah et al. (2020) describe embeddedness as a holistic retention mechanism shaped by both organizational and community influences. They note that when organizations foster strong employee–organization fit, social connections, and meaningful developmental opportunities, employees form deeper psychological attachments that decrease voluntary turnover. These findings highlight the value of policies that strengthen workplace cohesion and support employee development.

Further contributions come from Fasbender et al. (2018), who examined how embeddedness interacts with job satisfaction and occupational stress. Their findings indicate that job satisfaction serves as an important explanatory mechanism, helping to clarify why employees

experiencing stress may still choose to remain when their overall work experiences are positive. This perspective suggests that efforts to enhance job satisfaction, through recognition, career development, and supportive leadership, can complement strategies aimed at reducing stress, collectively contributing to lower turnover intentions.

## **2.5 Perceived Organizational Support**

Perceived organizational support (POS) refers to employees' beliefs about the extent to which their organization values their contributions and cares about their overall well-being (Eisenberger et al., 1986). POS is grounded in social exchange theory, which suggests that when employees perceive high levels of support, they feel obliged to reciprocate through greater commitment, improved performance, and reduced intention to leave the organization.

Empirical evidence consistently links POS to turnover intention. Perryer et al. (2010) found that POS significantly moderates the relationship between organizational commitment and turnover intention, such that employees who perceive higher support are less likely to consider leaving their jobs. These findings highlight the important role that supportive organizational practices, such as recognizing employee contributions and providing resources, play in fostering retention.

Subsequent research reinforces the value of POS in predicting turnover-related outcomes. Kalidass and Bahron (2015), examining employees in the hospitality sector, reported that both POS and perceived supervisor support (PSS) exhibit strong negative relationships with turnover intention. Their work emphasizes that support from both the organization and supervisors enhances employees' sense of security and belonging, thereby reducing thoughts of withdrawal.

Thirapatsakun et al. (2015) further demonstrated that POS indirectly influences turnover intention through work engagement. In their study of nurses in private hospitals, higher levels of

POS were associated with increased engagement, which in turn contributed to lower turnover intention. This suggests that POS not only affects employees' attitudes toward the organization but also shapes their motivational states, strengthening retention through multiple pathways.

Collectively, these findings underscore the importance of cultivating a supportive organizational climate. When organizations demonstrate care, recognition, and investment in their employees, they reinforce positive social exchanges that ultimately contribute to greater organizational commitment and reduced turnover intention.

## **2.6 Perceived Organizational Justice**

Perceived organizational justice refers to employees' evaluations of fairness within their organization, including the fairness of outcomes, procedures, and interpersonal treatment.

Mengstie (2020) identifies four dimensions of justice, distributive, procedural, interpersonal, and informational, which together shape employees' overall fairness perceptions. Distributive justice concerns the perceived fairness of outcomes such as compensation and promotions, while procedural justice focuses on the consistency and transparency of the processes used to determine those outcomes. Interpersonal justice addresses the degree of respect and dignity employees receive from supervisors, and informational justice reflects the quality and clarity of communication surrounding organizational decisions. Mengstie (2020) found that lower perceptions of fairness across these dimensions are associated with higher turnover intentions, underscoring the role of equitable practices in promoting workforce stability and employee commitment.

The broader literature further demonstrates that organizational justice is a central predictor of turnover intention, particularly through its influence on employee well-being.

Vaamonde et al. (2018) reported that distributive, procedural, and interpersonal justice

significantly reduce turnover intention by decreasing burnout and enhancing job satisfaction. Informational justice, while less directly related to burnout, contributes to improved job satisfaction, which independently predicts lower turnover intention. These findings highlight the importance of fair processes, transparent communication, and respectful treatment in shaping employees' attitudes toward their organization. When employees perceive their work environment as fair, they are more likely to experience positive psychological states that reinforce retention.

Additional evidence suggests that individual differences may also shape justice perceptions. Meisler (2013) demonstrated that emotional intelligence is positively associated with employees' perceptions of organizational justice and negatively associated with turnover intention. Emotionally intelligent employees tend to interpret organizational actions more constructively, perceive higher levels of fairness, and exhibit lower intentions to leave. Perceived organizational justice served as a key mediator in this relationship, indicating that emotional intelligence may enhance resilience and promote more favorable interpretations of workplace events. These findings suggest that fostering both fair organizational practices and employee emotional competencies can contribute to stronger perceptions of justice and, ultimately, lower turnover intention.

## **2.7 Work Environment**

The work environment encompasses the physical, social, and psychological conditions under which employees perform their tasks. It includes elements such as workspace design, workload, safety, interpersonal relationships, supervisory support, and the broader organizational climate. A well-structured and supportive work environment contributes to employees' sense of comfort, security, and professional fulfillment, while negative conditions, such as excessive

demands, poor communication, or inadequate resources, can undermine well-being and increase withdrawal behaviors (Herwidiyanto & Sinambela, 2024).

A substantial body of research highlights the work environment as a key determinant of employee attitudes and turnover intention. Positive and supportive workplace climates consistently demonstrate strong negative associations with employees' desire to leave their organizations (Xuecheng et al., 2022; Oyedeji et al., 2025). For example, Sabei et al. (2019) found that nurses who perceived adequate staffing, organizational support, and opportunities to participate in hospital decision-making reported higher job satisfaction and lower burnout, both of which contributed to reduced turnover intentions. Likewise, Lee et al. (2018) reported that physical therapists working in favorable environments characterized by manageable workloads, autonomy, and supportive leadership were significantly less likely to consider leaving their jobs.

Evidence from academic settings also reinforces the importance of the work environment. Soelton and Atnani (2018) found that technical staff in higher education institutions demonstrated lower turnover intentions when they perceived their work environment to be safe, comfortable, and relationally supportive. Collectively, these findings underscore the crucial role of cultivating a work environment that promotes well-being, reduces stress, and supports professional effectiveness as a strategic approach to enhancing employee retention and fostering organizational stability.

## **2.8 Career Growth Opportunities**

Career growth opportunities represent a critical component of organizational practices that shape employees' long-term commitment and turnover intention. These opportunities encompass promotion prospects, skill development pathways, and the achievement of career goals, all of which contribute to an employee's perception of upward mobility and professional

advancement. Prior work highlights that career development, such as promotion speed, remuneration growth, and opportunities for professional skill enhancement, significantly strengthens employees' attachment to their organization (Nawaz & Pangil, 2016; Weng & Hu, 2009).

Recent empirical research further demonstrates that career growth opportunities exert both direct and indirect effects on turnover intention. Wang et al. (2022) found that career growth moderates the relationship between job satisfaction and turnover intention among public-sector employees, such that strong advancement prospects weaken the negative impact of reduced satisfaction on retention. Similarly, Yang et al. (2016) reported that organizational career growth, defined through promotion speed, salary progression, professional skill acquisition, and attainment of long-term career objectives, substantially decreases turnover intention among nursing professionals in China. Their findings also emphasize the role of organizational support, suggesting that career development initiatives are most effective when embedded within a supportive institutional context.

Nawaz and Pangil (2016), in their work with university faculty, further highlighted that promotion opportunities and salary growth function as strong negative predictors of turnover intention. They noted that organizational commitment mediates this relationship, as employees tend to reciprocate developmental investments with higher loyalty and reduced intention to leave. Importantly, their research underscores that fair and transparent performance appraisal systems enhance employees' perceptions of available career growth, thereby strengthening retention.

Taken together, existing literature affirms that career growth opportunities are central to organizational retention strategies. Clear promotion pathways, equitable performance

evaluations, and supportive professional development practices contribute to stronger organizational commitment and reduced turnover intention (Nawaz & Pangil, 2016; Yang et al., 2016; Wang et al., 2022). Integrating these elements into human resource policies can help organizations maintain a stable, committed workforce while enhancing overall organizational effectiveness.

## **2.9 Work-Life Balance**

Work-life balance (WLB) has become a central construct in contemporary organizational research, particularly in relation to employee well-being and turnover intention. WLB refers to the degree to which individuals are able to effectively manage the demands of their professional and personal lives, and it is widely recognized as a determinant of job satisfaction, psychological health, and long-term employee retention (Koubova & Buchko, 2013). When employees experience imbalance, such as excessive work demands that encroach on personal life, the resulting stress and strain can elevate intentions to leave the organization (Jaharuddin & Zainol, 2019).

Empirical studies highlight that both formal and informal WLB practices influence turnover intention, though informal supports often exert stronger effects. For instance, managerial support and job autonomy consistently demonstrate stronger reductions in turnover intention compared to formal practices such as flexible scheduling (Suifan, Abdallah, & Diab, 2016). Managerial support signals that the organization values employees' personal well-being, while autonomy allows individuals to manage work demands more effectively, reducing work-life conflict. Although flexible scheduling remains a widely adopted formal WLB practice, its effectiveness depends heavily on the extent to which organizational culture genuinely supports work-life integration (Jaharuddin & Zainol, 2019).

Burnout has also been identified as an important mechanism linking WLB to turnover intention. Poor WLB can heighten burnout, which in turn contributes to withdrawal cognitions and increased turnover intention (Sholekha & Esthi, 2025). Similarly, Oosthuizen et al. (2019) demonstrated that frequent negative spillovers between work and home life significantly elevate turnover intentions, whereas positive interactions contribute to job satisfaction and retention. Jaharuddin and Zainol (2019) further noted that although job engagement is associated with lower turnover intention, it does not mediate the relationship between WLB and turnover, emphasizing the importance of directly addressing work–life conflict.

Taken together, the literature suggests that organizations seeking to reduce turnover intention should prioritize enhancing work–life balance through supportive leadership practices, increased autonomy, and interventions aimed at minimizing work–life conflict. Embedding these practices within an organizational culture that values employee well-being is essential for fostering sustained commitment and reducing the likelihood of turnover (Suifan et al., 2016; Sholekha & Esthi, 2025; Jaharuddin & Zainol, 2019; Oosthuizen et al., 2019).

## **2.10 Employee Net Promoter Score (eNPS)**

The employee net promoter score has emerged as a widely used indicator of employee sentiment, organizational loyalty, and overall engagement. Adapted from the traditional Net Promoter Score employed in customer experience research, eNPS categorizes employees based on their likelihood of recommending the organization as a place to work: promoters (scores of 9–10), passives (7–8), and detractors (0–6). Organizations then calculate the score by subtracting the percentage of detractors from the percentage of promoters, producing a simple yet intuitive metric that leadership teams can easily interpret (Brown, 2020).

Recent scholarship emphasizes the value of eNPS as a practical diagnostic tool within organizational settings. Because of its simplicity, eNPS enables employers to communicate engagement results effectively to managers and decision-makers who may not have backgrounds in organizational research or analytics. Studies also show that eNPS correlates with important workplace outcomes, including work-unit performance, suggesting its potential predictive utility when integrated into broader organizational assessment systems (Brown, 2020). Löffstrand et al. (2025) further highlight the benefits of eNPS in public sector environments, where it can reveal engagement disparities and areas of dissatisfaction that traditional multi-item surveys may overlook.

Another advantage of eNPS is its adaptability as a pulse-check measure. Organizations can administer it frequently at low cost, enabling the tracking of engagement trends over time and providing early warning signs of potential retention challenges (Brown, 2020; Löffstrand et al., 2025). Such responsiveness is especially valuable in dynamic work environments where employee sentiment can shift rapidly in response to organizational changes, leadership transitions, or workload fluctuations.

Despite these strengths, researchers caution against relying on eNPS as a standalone measure. Its single-item format captures only a narrow dimension of employee experience, potentially oversimplifying complex attitudes or masking important variation among passive respondents. Moreover, the extent to which eNPS contributes to meaningful organizational improvement depends heavily on leaders' ability to interpret results appropriately and take informed action in response (Löffstrand et al., 2025). As a consequence, scholars recommend pairing eNPS with qualitative feedback opportunities and validated engagement scales that

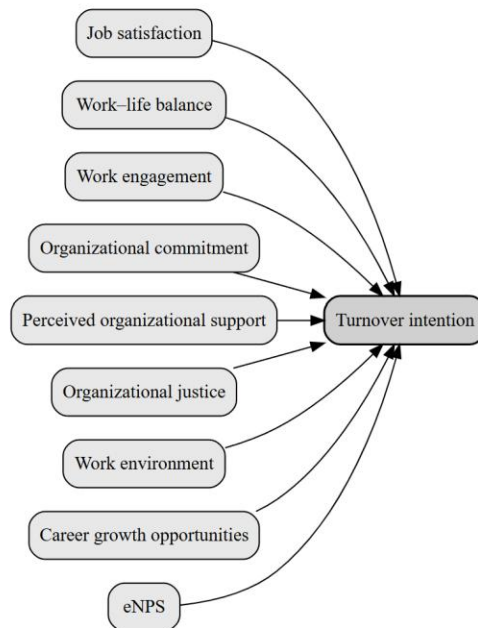
assess satisfaction, commitment, or well-being to develop a more comprehensive understanding of employee perceptions (Brown, 2020).

Overall, eNPS provides a useful, accessible metric for gauging broad patterns in employee sentiment and identifying potential retention risks. When used as part of a larger human resource analytics strategy, rather than in isolation, it can support evidence-based decision-making and enhance organizational insight into employee engagement dynamics (Brown, 2020; L fstrand et al., 2025).

Figure 2.1 presents a visual summary of the hypothesized relationships among the study variables, illustrating how each predictor is expected to influence turnover intention.

**Figure 2.1**

*Conceptual model of hypothesized predictors of turnover intention.*



## 2.11 Machine Learning Approaches to Turnover Intention Prediction

Advancements in workforce analytics have encouraged the integration of machine-learning (ML) techniques into turnover intention and attrition research. In contrast to traditional

linear approaches, ML methods can accommodate complex, nonlinear relationships and interactions among psychological, organizational, and demographic predictors, characteristics common in human resources data (Fukui et al., 2023; Al Akasheh et al., 2024). Moreover, ML supports high-dimensional data and mixed variable types (survey ratings, demographic information, behavioral metrics), making it especially useful for multifactorial constructs such as job satisfaction, organizational justice, embeddedness, and engagement.

In this thesis, four modeling approaches are compared, logistic regression, random forests (RF), Bayesian Additive Regression Trees (BART), and automated machine learning (AutoML), each selected for their particular strengths and suitability for turnover-intention prediction. A systematic review of turnover-prediction studies confirms the growing adoption of supervised ML methods: over the past decade, random forests have been the most frequently used algorithm, followed by gradient boosting methods and other ensemble techniques (Al Akasheh et al., 2024). This trend underscores the practical and empirical value of ML in organizational settings.

By combining theoretical interpretability (via logistic regression) with the flexibility and predictive power of modern ML algorithms (RF, BART, AutoML), the present study adopts a balanced modeling framework. This approach enhances the capacity to capture the multifaceted determinants of turnover intention, from individual attitudes to organizational context, while aligning with best practices in HR analytics.

## **2.12 Logistic Regression**

Logistic regression is one of the most widely used statistical models in organizational psychology and turnover research due to its interpretability, theoretical transparency, and suitability for binary outcomes such as turnover intention. Unlike more complex

machine-learning models, logistic regression provides direct insight into the magnitude and direction of relationships between predictors (e.g., job satisfaction, organizational commitment) and the probability of intending to leave an organization. This interpretability makes it especially useful in theory-driven research where understanding the underlying psychological mechanisms is as important as predictive accuracy (Zhuge et al., 2024; Liu et al., 2024).

Researchers have long relied on logistic regression to study turnover intention because of its ability to test hypotheses grounded in established theories such as the unfolding model of turnover (Lee & Mitchell, 1994) and social exchange theory (Cropanzano & Mitchell, 2005). Its formal structure allows researchers to evaluate how incremental changes in predictor variables influence the odds of turnover intention, enabling clear theoretical interpretation (Menard, 2018). Logistic regression has also been shown to produce stable, reliable estimates when applied to employee-survey data, making it a standard baseline model in HR analytics (Zhao et al., 2024; Fukui et al., 2023; Talebi et al., 2025). (Zhao et al., 2024). Although advanced machine-learning models are gaining popularity for their predictive power, logistic regression remains the most widely used baseline approach in turnover and attrition research (Fukui et al., 2023; Talebi et al., 2025).

In addition to its theoretical strengths, logistic regression offers several methodological advantages. It performs well with moderate sample sizes, tolerates multicollinearity when properly diagnosed and addressed, and supports the inclusion of interaction terms to test more complex psychological relationships (Menard, 2018). These characteristics make logistic regression an essential anchor model in the present study, serving both as a benchmark for evaluating machine-learning approaches and as a theoretically interpretable tool for understanding the psychological and organizational determinants of turnover intention.

### 2.13 Random Forest

Random forest (RF) is an ensemble learning method that constructs multiple decision trees during training and outputs the majority vote (classification) or average prediction (regression) of the individual trees. RF is especially suitable for turnover-intention prediction because it can model complex, nonlinear relationships and interactions among a large set of mixed predictors (e.g., psychological scales, demographic variables, organizational factors) without requiring strict parametric assumptions (original algorithm proposed by Breiman & Cutler, 2001) (Benabou et al., 2025; Shobhanam & Sumati, 2022).

In the context of employee turnover and attrition research, random forest has become one of the most common algorithms used in recent years. A systematic review of machine-learning applications for turnover prediction concluded that RF appears in the majority of studies, given its ability to achieve high predictive accuracy and robust classification performance across diverse organizational datasets (Al Akasheh et al., 2024). Empirical studies using real human-resources data have demonstrated RF's effectiveness: for example, a recent large-scale HR analytics study found that RF achieved superior prediction accuracy for employee attrition compared to simpler models and correctly identified key predictors such as work hours, tenure, position, pay, and training hours (Fukui et al., 2023). Another recent study using a balanced HR-survey dataset showed that RF could achieve high sensitivity and specificity in classifying staying vs. leaving employees, even when data were imbalanced, a common issue in turnover samples (Liu et al., 2024; Jamroni et al., 2025).

Methodologically, RF offers several advantages that have been well-documented in the machine-learning and HR-analytics literature. First, RF handles multicollinearity among predictors effectively because its tree-based structure does not rely on linearity or independence

assumptions (Breiman, 2001; Benabou et al., 2025). This is particularly valuable when examining highly interrelated psychological and organizational constructs such as job satisfaction, organizational justice, embeddedness, and perceived support.

Second, RF provides multiple feature-importance metrics (e.g., permutation importance, mean decrease in Gini), which have been shown to be useful for identifying the most influential predictors in employee-turnover models (Fukui et al., 2023; Liu et al., 2024). These importance indicators allow researchers to assess which constructs meaningfully contribute to turnover intention, supporting the dual goals of this thesis: theory testing and prediction.

Third, RF reduces overfitting through mechanisms such as bootstrap aggregation and internal out-of-bag (OOB) error estimation. Prior studies emphasize that these properties make RF robust and reliable when applied to medium-sized, survey-based workforce datasets, a common context in turnover-intention research (Jamroni et al., 2025; Al Akasheh et al., 2024). Such methodological strengths further justify RF as an appropriate and rigorous modeling approach for the present study.

Given these theoretical and methodological strengths, random forest is a central part of the modeling framework of this thesis. It complements logistic regression (for interpretability) and more flexible ML methods (BART, AutoML) by balancing complexity and interpretability, and by providing insights into predictor importance that are directly useful for organizational interventions.

#### **2.14 Bayesian Additive Regression Trees (BART)**

BART represent a flexible, nonparametric machine-learning method that combines the strengths of ensemble tree models with bayesian regularization. Originally introduced by Chipman, George, and McCulloch (2010), BART constructs a sum-of-trees model in which each

tree is constrained to be a weak learner, thereby preventing overfitting and promoting stable, generalizable predictions. This Bayesian framework enables BART to generate calibrated probability estimates and full posterior uncertainty intervals, features that are particularly valuable when modeling behavioral outcomes such as turnover intention, where quantifying uncertainty contributes to more informed organizational decision-making.

In recent years, BART has gained traction across the social sciences, including economics, public policy, and psychology, due to its ability to flexibly capture nonlinear relationships and higher-order interactions without relying on restrictive parametric assumptions (Hill, 2011; Chipman et al., 2010). Comparative evaluations have shown that BART often performs competitively with, or surpasses, other ensemble models in predictive accuracy, particularly in settings involving moderate sample sizes and multifaceted predictors, which are common in organizational research (Plant, 2021; Sparapani et al., 2021). Its probabilistic outputs make it well suited for binary outcomes such as turnover intention, where well-calibrated predicted probabilities enhance interpretability and decision usefulness (Sparapani & McCulloch, 2020).

Methodologically, BART offers several advantages for turnover-intention prediction. First, its Bayesian shrinkage priors naturally guard against overfitting, making it suitable for datasets with multiple correlated psychological and organizational constructs, such as job satisfaction, work engagement, organizational justice, and embeddedness (Chipman et al., 2010; Hill, 2011). Second, BART provides posterior inclusion probabilities and variable-importance summaries that help identify influential predictors in complex behavioral models (Linero & Yang, 2018; Sparapani et al., 2021). Third, its ability to produce full posterior uncertainty intervals supports more transparent interpretation of prediction confidence, an essential feature

when turnover predictions are used to guide managerial or policy decisions (Sparapani & McCulloch, 2020).

Given these strengths, BART serves as an important component of the modeling framework in this thesis. It complements logistic regression (interpretability), random forest (interaction modeling), and AutoML (algorithmic optimization) by offering a robust probabilistic alternative capable of capturing nuanced relationships while providing uncertainty estimates that enhance the interpretive value of predicted turnover intentions.

### **2.15 Automated Machine Learning (AutoML)**

AutoML represents an emerging class of tools designed to streamline and optimize the model-building process by automatically selecting algorithms, tuning hyperparameters, and generating ensemble models (Scott et al., 2024). Rather than relying solely on researcher expertise, AutoML systematically tests a wide range of candidate models, including generalized linear models, gradient-boosted machines, random forests, deep learning architectures, and stacked ensemble, and identifies the configuration that maximizes predictive performance according to predefined metrics (Salehin et al., 2023). This structured search process reduces researcher bias in model selection and increases the likelihood of identifying well-tuned models that generalize effectively to new data.

AutoML has gained increasing attention in HR analytics and behavioral prediction domains due to its ability to handle complex, multifactorial datasets characteristic of workforce research. For example, H2O AutoML and Auto-Sklearn have been used to predict employee attrition, job-performance outcomes, and workforce turnover with high accuracy, outperforming manually tuned models in several studies (Feurer et al., 2019; Al-Akashah et al., 2024).

Researchers emphasize that AutoML's systematic search strategies are particularly valuable in

settings where the optimal model structure is not known in advance, as is common with psychological and organizational data that contain nonlinear relationships, interaction effects, and correlated predictors (Fukui et al., 2023; Salehin et al., 2023).

In addition to improving predictive performance, AutoML enhances methodological transparency by documenting the full modeling pipeline, including preprocessing steps, model candidates, and performance evaluations. Prior evaluations of AutoML frameworks emphasize that automated pipelines improve reproducibility and reduce human-driven variability in analytical decision-making. Such transparency is especially beneficial in organizational research, where replicability and accountability are essential for evidence-based decision-making. AutoML therefore serves as a complementary approach to traditional and theory-driven models, enabling researchers to balance interpretability with data-driven discovery while efficiently identifying high-performing predictive models for turnover intention (Balaji & Allen, 2018).

Taken together, these four modeling approaches offer a comprehensive and methodologically rigorous framework for examining turnover intention. By integrating theory-driven statistical modeling with advanced machine learning techniques, this thesis adopts an approach that reflects contemporary best practices in organizational research and HR analytics. This combined framework enables a refined analysis of the psychological and organizational factors that shape employees' intentions to leave, supporting a deeper and more evidence-based understanding of turnover within modern workplaces.

## 2.16 Hypotheses

Given the increasing integration of machine learning techniques in organizational research, it is essential to evaluate how different predictive models perform when applied to turnover-intention data. Traditional statistical models, such as logistic regression, offer interpretability and theoretical alignment, whereas modern machine-learning approaches, random forests, BART, and AutoML, provide greater flexibility in modeling nonlinear relationships and interactions. Prior research demonstrates that ensemble and automated methods often outperform single-model approaches in predictive tasks involving complex, multidimensional data (Breiman, 2001; Chipman et al., 2010; Balaji & Allen, 2018).

The present study therefore empirically compares these modeling approaches to assess their relative predictive performance in the context of turnover intention. This comparison serves two purposes: (1) to determine whether advanced machine-learning models offer meaningful improvement over traditional or simpler methods, and (2) to evaluate which model class provides the most accurate and generalizable predictions given the psychological and organizational predictors used in this thesis. Grounded in existing methodological literature and supported by the characteristics of the dataset, the following hypotheses are proposed:

**H1:** Logistic regression will provide a theoretically grounded and interpretable baseline model for predicting turnover intention, but its performance will be limited by its assumption of linearity and reduced flexibility in modeling complex variable interactions.

**H2:** Random forest (RF) models will achieve lower prediction error (RMSE) compared to logistic regression, due to their ensemble structure, ability to capture nonlinear relationships, and reduced susceptibility to overfitting.

**H3:** Bayesian Additive Regression Trees (BART) will demonstrate predictive performance comparable to or better than random forest, given its probabilistic framework, regularization properties, and strong performance on tabular social-science datasets.

**H4:** Automated Machine Learning (AutoML) will identify the optimal model configuration with the lowest RMSE among all methods tested, given its ability to systematically search across algorithms and hyperparameter spaces, selecting the most appropriate pipeline for the dataset.

## **Chapter 3: Methodology**

### **3.1 Participants**

This study included 239 working adults, all recruited exclusively through Connect by CloudResearch, an established online participant recruitment platform widely used in behavioral and organizational research. All participants were compensated at a fair and pre-approved rate through CloudResearch.

To be eligible for participation, individuals were required to be 18 years or older and currently employed at least part-time (30 or more hours per week) at the time of data collection. These criteria ensured that all respondents were able to meaningfully answer items related to workplace experiences, organizational perceptions, and turnover intention. Because CloudResearch draws from a broad national participant pool, the final sample reflects diversity across age, gender, job roles, industries, and organizational tenure, supporting the generalizability of the findings.

The sample size was determined based on best practices for machine-learning research using structured survey data, rather than traditional power analyses designed for hypothesis testing or inferential statistics. Predictive modeling approaches, such as logistic regression, random forests, BART, and AutoML, benefit from datasets that are sufficiently large to allow the algorithms to detect meaningful patterns, avoid overfitting, and generalize well to new cases (Breiman, 2001; Chipman et al., 2010; LeDell & Poirier, 2020). A sample of approximately 200–300 cases is commonly considered adequate for supervised learning on behavioral, psychological, or organizational datasets with a modest number of predictors, and is consistent with sample sizes used in prior turnover-prediction studies.

All participation was voluntary. Respondents provided informed consent electronically prior to beginning the survey and were free to discontinue at any time without penalty. No additional recruitment methods (e.g., social media outreach or snowball sampling) were used, and no uncompensated participation occurred.

### **3.2 Data Overview**

Primary data for this study were collected through an online survey administered via Connect by CloudResearch, a validated platform commonly used for recruiting high-quality research participants. This approach ensured access to a diverse sample of employed adults across industries, job roles, and work arrangements. All participants completed informed consent prior to beginning the survey.

The survey included self-report measures of key psychological and organizational constructs relevant to turnover research, as well as demographic information such as tenure, employment status, and work schedule. The primary outcome variable, turnover intention, was measured using a continuous scale to capture the full range of respondents' intentions to leave their organization.

Predictors included eleven validated constructs shown in prior research to influence turnover from both individual and organizational perspectives: organizational commitment, job satisfaction, work engagement, job embeddedness, perceived organizational support, perceived organizational justice, work environment, career growth opportunities, work-life balance, employee net promoter score and turnover intentions. These measures together provide a comprehensive assessment of factors contributing to turnover intention.

### 3.3 Measures

#### 3.3.1 Organizational Commitment

Organizational commitment was measured using the 4-item Unidimensional Commitment Scale (KUT) developed by Klein et al. (2013). The scale is unidimensional and target-free, assessing commitment as a single construct rather than distinguishing between affective, normative, and continuance components. A sample item from the scale is: “How committed are you to your organization?” Items are rated on a 5-point scale (1 = Not at all to 5 = Extremely) and were averaged to create a total commitment score. The scale demonstrates very high reliability ( $\alpha = .86-.97$ ) and strong structural validity across multiple targets. Higher values indicate stronger psychological attachment.

#### 3.3.2 Job Satisfaction

Job satisfaction was assessed using the 5-item Short Index of Job Satisfaction (SIJS), adapted from Brayfield and Rothe’s (1951) measure. Items are rated on a 5-point Likert scale ranging from Strongly disagree to Strongly agree, such as the item “I feel fairly satisfied with my present job.” Two negatively worded items were reverse-coded prior to scoring. Scores will be averaged, with higher values indicating greater global job satisfaction. The SIJS has shown strong internal consistency ( $\alpha \approx .84-.88$ ) and cross-cultural validity (Sinval & Marôco, 2020).

#### 3.3.3 Work Engagement

Work engagement was measured using the 17-item Utrecht Work Engagement Scale (UWES-17; Schaufeli et al., 2002). The scale assesses vigor, dedication, and absorption using a 7-point frequency scale (0 = Never to 6 = Always). For example, employees rated how often they feel “at my work, I feel bursting with energy.” Subscale means and an overall engagement

score were computed. The UWES-17 demonstrates strong internal reliability ( $\alpha = .85-.92$ ) and structural validity internationally.

#### 3.3.4 Job Embeddedness

Job embeddedness was measured using the 12-item attitudinal scale developed by Clinton et al. (2012). The scale consists of two subdimensions: on-the-job embeddedness (JE-On; 6 items) and off-the-job embeddedness (JE-Off; 6 items). Items are assessed on a 5-point Likert scale from Strongly disagree to Strongly agree and capture the extent to which employees are connected to their jobs and communities (e.g., "Overall, I fit very well in my organization"). Subscale scores were computed by averaging corresponding items, with higher scores indicating greater embeddedness. No items required reverse-scoring. JE-On is theoretically expected to be more strongly related to turnover intentions.

#### 3.3.5 Perceived Organizational Support (POS)

Perceived organizational support was measured using the 8-item Survey of Perceived Organizational Support (SPOS; Eisenberger et al., 1986). Items were rated on a 7-point scale (1 = Strongly disagree to 7 = Strongly agree). A sample item includes: "The organization values my contribution to its well-being." Negatively worded items were reverse-coded. Scores were averaged, with higher values indicating greater perceived organizational support. POS has demonstrated very high reliability ( $\alpha = .93-.97$ ).

#### 3.3.6 Perceived Organizational Justice

Organizational justice was assessed using the 20-item scale developed by Colquitt (2001), capturing distributive, procedural, interpersonal, and informational justice. Items were rated on a 5-point Likert scale and averaged within each subscale. No items required

reverse-scoring. For example, one item asked participants to reflect on whether “My supervisor treats me with dignity” (interpersonal justice). This scale has demonstrated strong reliability and construct validity across occupational contexts.

### 3.3.7 Work Environment

Work environment was assessed using the 10-item Work Environment Scale (WES-10; Røssberg et al., 2004). All items were scored on a 5-point agreement scale and averaged to create a global index of work environment quality, with higher scores indicating more positive perceptions. For example, employees rated the extent to which their work gives them opportunities to demonstrate their abilities (e.g., “Does what you do on the ward give you a chance to see how good your abilities really are?”).

### 3.3.8 Career Growth Opportunities

Career growth opportunities were assessed using the 15-item Organizational Career Growth Scale developed by Weng and Hu (2009) and further validated cross-culturally by Kim, Price, Mueller, and Watson (1996), reflecting career goal progress, professional ability development, promotion speed, and remuneration growth. Items were rated on a 5-point Likert scale and averaged within subdimensions (e.g., “My present job moves me closer to my career goals”). Higher scores indicated stronger perceived opportunities for growth within the organization.

### 3.3.9 Work-Life Balance

Work–life balance was measured using the 4-item Work–Life Balance Scale (Brough et al., 2014). Items were scored on a 5-point Likert scale, with one negatively worded item reverse-scored. An example item is, “I currently have a good balance between the time I spend at

work and the time I have available for non-work activities.” Scores were averaged, with higher values reflecting better balance between work and non-work roles.

### 3.3.10 Employee Net Promoter Score (eNPS)

Employee loyalty was assessed using a single-item eNPS measure asking respondents to rate how likely they are to recommend their organization as a workplace on a scale from 0 to 10 (e.g., “On a scale of 0–10, how likely are you to recommend working here to a friend or colleague?”). Responses were categorized into promoters (9–10), passives (7–8), and detractors (0–6), and an eNPS score was computed by subtracting the percentage of detractors from the percentage of promoters.

### 3.3.11 Turnover Intentions

Turnover intentions were measured using a 6-item turnover intention scale commonly employed in organizational research (e.g., Bothma & Roodt, 2013). Items assessed the extent to which employees were thinking about leaving, planning to leave, or searching for alternative employment. For example, one item asked, “How often have you considered leaving your job?” Responses were rated on a 5-point Likert scale ranging from Strongly disagree to Strongly agree. Scores were averaged to create a continuous turnover intention score, with higher values indicating stronger intentions to quit ( $\alpha \approx .85-.90$ ).

For predictive modeling, turnover intention was also dichotomized to create a binary classification outcome. Consistent with prior turnover research, participants scoring in the highest quartile of the turnover intention scale were coded as 1 = High Turnover Intention, indicating meaningful withdrawal cognitions, while all remaining participants were coded as 0 = Low Turnover Intention. This threshold was selected to differentiate individuals, demonstrating a clear risk of leaving from those with lower or more moderate levels of turnover intention.

Dichotomizing the variable supported direct comparison of logistic regression and machine-learning classification models and aligned with applied organizational decision-making contexts, where the primary objective is often to identify employees most at risk of turnover.

All scales were administered in their validated Likert-type formats, and Cronbach's alpha were computed for all multi-item scales to confirm internal consistency. It was administered in their original validated form using a 5- or 7-point Likert-type response format. Reliability (Cronbach's alpha) was calculated for multi-item scales to confirm internal consistency within the present sample.

### **3.4 Procedure**

Data for this study were collected using an online survey administered through Qualtrics. Participants were recruited exclusively through Connect by CloudResearch, a research platform that screens and matches eligible respondents based on predefined study criteria. Individuals who met eligibility requirements (18 years or older and employed at least 30 hours per week) were invited to participate and were provided access to the survey link directly through the CloudResearch interface.

Before beginning the survey, all participants reviewed an online informed consent form outlining the study's purpose, procedures, estimated time commitment, potential risks and benefits, confidentiality protections, and the voluntary nature of participation. Only respondents who electronically affirmed their consent were able to proceed to the survey.

The survey required approximately 20–25 minutes to complete, consistent with CloudResearch's time estimates for participant compensation. Participants received the approved compensation amount upon successful completion of the survey, in accordance with CloudResearch policies.

All responses were recorded directly in Qualtrics and then exported. No identifying information was collected or retained. Data were securely stored in password-protected files accessible only to the research team. Only participants who provided complete responses to the key variables relevant to this thesis were retained in the final dataset.

This procedure ensured ethical data collection, confidentiality of participant information, and consistency with Institutional Review Board (IRB) guidelines.

### **3.5 Data Preparation and Cleaning**

Data preparation followed a structured, multi-step process to ensure the dataset was suitable for both statistical analyses and machine-learning models used in this thesis. Because data collection was completed prior to analysis, all steps were conducted retrospectively on the final Qualtrics export.

First, cases missing the dependent variable, turnover intention, were removed entirely. Missing data across predictor variables were then inspected using column-wise missing counts in R. Overall missingness was minimal, with most variables containing only one missing response (approximately 0.4% of the sample). The highest levels of missing data occurred in optional open-text comment fields (e.g., q20\_6\_text with 236 missing responses, q9\_6\_text with 238 missing responses, and q8\_9\_text with 206 missing responses), which were excluded from all modeling procedures due to their non-analytic nature. All theoretically relevant psychological and organizational variables (e.g., job satisfaction, work–life balance) showed very low missingness, ranging from 0 to 4 cases. Because the proportion of missing data was low and did not indicate systematic patterns, mean imputation was applied for continuous variables and mode imputation for categorical variables to preserve sample size.

Categorical demographic variables (e.g., gender, employment status, job level) were recoded using dummy variables to ensure compatibility with logistic regression and machine-learning models. All continuous predictors, including organizational commitment, job satisfaction, work engagement, job embeddedness, perceived organizational support, organizational justice, work environment, career growth opportunities, work–life balance, and eNPS, were standardized using z-scores. Standardization was necessary to place all predictors on a comparable scale, reduce bias in distance-based algorithms, and support interpretability across models.

Outliers were examined using z-scores, histograms, and boxplots for each continuous variable. Because machine-learning models such as random forest and BART are robust to non-extreme outliers, and because all responses represented self-report perceptions rather than objective measurements, no valid participant data were removed solely on the basis of extremity. Only responses identified as impossible or non-credible (e.g., straight-lining across all items, invalid demographic entries) were excluded.

Finally, assumptions relevant to specific models were evaluated. For logistic regression, multicollinearity was assessed using variance inflation factors (VIFs). Normality and linearity assumptions were not required for any of the models. All preparation steps ensured the final dataset was clean, reliable, and compatible with the full suite of modeling techniques applied in this study.

## Chapter 4: Analysis Plan

### 4.1 Overview

The overarching goal of the analysis was to evaluate and compare multiple predictive modeling approaches to determine which method most accurately predicted employee turnover intention. Specifically, the analysis assessed differences in predictive performance, model interpretability, and the relative importance of psychological and organizational predictors across models.

To accomplish this, the analysis followed a structured, multi-stage process. First, descriptive analyses and exploratory data visualizations were conducted to summarize sample characteristics and examine distributions, correlations, and potential multicollinearity among predictors. These steps provided the foundational understanding necessary for responsible modeling.

Next, four predictive modeling approaches, logistic regression, random forests, BART, and AutoML, were implemented. Logistic regression served as a theory-driven baseline model, offering high interpretability and aligning with traditional approaches to turnover research. Random forests introduced nonlinear modeling capabilities and robustness to multicollinearity, making them well-suited for complex organizational datasets. BART provided a Bayesian, ensemble-based model that captured nonlinearities and interaction effects while offering uncertainty quantification. AutoML systematically searched across algorithms and hyperparameters to identify the highest-performing predictive pipeline.

Each model was evaluated using consistent performance metrics, including accuracy, AUC, Brier score, and RMSE, to allow fair comparison across methods. Cross-validation and

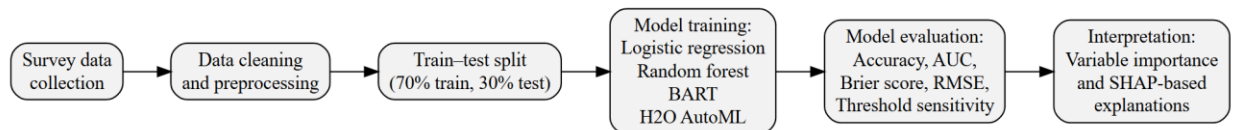
train–test splits were used to ensure generalizability and reduce overfitting. Variable importance measures, such as odds ratios and mean decrease in impurity, further supported interpretation of predictors across models.

Through this comparative modeling framework, the analysis identified the most effective approach for predicting turnover intention and provided insights into the psychological and organizational factors that most strongly contributed to employees’ intentions to leave.

Figure 4.1 provides an overview of the full analytical workflow used in this study. The process included survey data collection, data cleaning and preprocessing, construction of the training and test datasets, model training using four predictive approaches (logistic regression, random forest, BART, and H2O AutoML), and model evaluation using discrimination and calibration metrics, along with interpretability analyses using variable importance outputs.

**Figure 4.1**

*Analytical workflow for data preprocessing, model training, and evaluation.*



As shown in Figure 4.1, the analysis followed a multistage workflow beginning with data preprocessing and ending with model interpretation.

## 4.2 Exploratory Data Analysis (EDA)

Exploratory Data Analysis was conducted as the initial phase of the analytical process to ensure data quality, evaluate underlying patterns, and assess the suitability of variables for

predictive modeling. Descriptive statistics (including means, standard deviations, and frequency distributions) were computed for all demographic, predictor, and outcome variables to summarize overall sample characteristics.

Visual inspections of distributions were performed using histograms, density plots, and boxplots to identify skewness, kurtosis, and potential outliers. Because the measures in this study are self-report Likert-type scales, non-normality was expected and did not necessitate transformation for machine-learning models; however, distributional characteristics remained relevant for interpretation and model selection.

Correlation analyses were conducted to examine associations among the psychological and organizational predictors. A correlation heatmap was used to visually assess potential multicollinearity. VIFs were computed to evaluate collinearity assumptions specifically for logistic regression. No VIF values exceeded commonly accepted thresholds, indicating that multicollinearity was not severe enough to bias regression estimates.

Missing data patterns were examined using frequency summaries and missingness maps. Missingness across the dataset was minimal and did not exhibit systematic patterns; therefore, mean or mode imputation was applied where appropriate to retain the full sample for analysis.

Together, these EDA procedures provided a comprehensive understanding of the structure and quality of the dataset, informed preprocessing decisions such as standardization and dummy coding, and ensured that the data were appropriately prepared for the logistic regression, random forest, BART, and AutoML models used in this study.

### **4.3 Predictive Modeling Approaches**

This study employs four predictive modeling approaches, logistic regression, random forests, BART, and AutoML, to compare model performance, interpretability, and suitability for

predicting turnover intention. Each model offers distinct methodological strengths aligned with best practices in organizational research and modern predictive analytics.

#### 4.3.1 Logistic Regression

Logistic regression serves as the theoretical and statistical baseline for this study and was implemented as a binary classification model. Although turnover intention was originally measured as a continuous construct, the final analysis followed a classification framework by converting turnover intention into a binary outcome using the 75th percentile cutoff (high turnover intention = 1; low turnover intention = 0). This approach aligns with predictive modeling objectives and is consistent with prior turnover research that distinguishes between high-risk and low-risk groups.

The model estimates the probability of belonging to the high-turnover-intention group as a function of psychological and organizational predictors. All predictors were numeric and standardized prior to modeling, consistent with best practices for logistic regression. Model assumptions, including independence of observations, absence of perfect multicollinearity, and linearity of continuous predictors with the logit, were checked using diagnostic tools such as VIFs. Logistic regression provides interpretable coefficients and odds ratios, offering a transparent comparison point for assessing the added value of more flexible machine-learning models.

#### 4.3.2 Random Forests

Random forests were implemented using the *randomForest* package in R as a binary classifier predicting high turnover intention. The final model consisted of 500 decision trees trained on the full set of standardized predictor variables. Random forests handle nonlinearities, multicollinearity, and complex interaction patterns without requiring explicit model

specification, making them well-suited for organizational datasets like this one. Predictions for the test dataset were obtained from the out-of-bag aggregated probability estimates.

Model performance was evaluated using AUC, Brier score, RMSE, and classification accuracy. Variable importance was extracted using mean decrease in impurity scores, providing insight into the relative influence of each psychological and organizational predictor. This ensemble method served as a core benchmark for evaluating whether Bayesian and AutoML approaches offer meaningful performance improvements.

#### 4.3.3 Bayesian Additive Regression Trees (BART)

BART was implemented using the *dbarts* package in R as a classification model. Because BART naturally supports binary outcomes, the model was trained using factor-encoded turnover intention labels. Predictions for the test set were generated by averaging posterior draws across iterations, resulting in a probability estimate for each participant. This posterior mean approach aligns with recommended usage in Bayesian predictive modeling.

BART is advantageous for turnover prediction due to its ability to capture nonlinear effects and interactions automatically, while also providing uncertainty estimates through posterior distributions. The implementation used default hyperparameters appropriate for classification tasks. Variable importance was assessed using the distribution of variable inclusion across posterior trees, offering a probabilistic measure of predictor relevance. This Bayesian modeling approach complements the deterministic structure of random forests and the parametric nature of logistic regression.

#### 4.3.4 Automated Machine Learning (AutoML)

AutoML was implemented using the H2O AutoML framework in R. After preparing the training and test datasets, including converting the binary outcome to a factor for H2O

compatibility, the AutoML routine evaluated a wide range of candidate algorithms such as generalized linear models, gradient boosting machines, deep learning models, and stacked ensembles. The AutoML run was configured with a 300-second training budget and used AUC as the optimization metric.

The final selected model was the top-ranked ensemble identified by the AutoML leaderboard. Predictions on the test set were generated to compute performance metrics including AUC, Brier score, RMSE, and accuracy. AutoML provided an empirical assessment of whether a fully automated optimization pipeline could outperform researcher-defined models (logistic regression, random forest, BART) and served as the upper bound of predictive performance for the dataset.

#### **4.4 Model Evaluation**

Model performance was evaluated using standardized metrics aligned with the binary classification design and the analytical workflow implemented in R. Because the main purpose of this study was to assess how well logistic regression, random forest, BART, and AutoML predicted turnover intention, evaluation centered on comparing predictive strength and probability accuracy across models.

All reported results reflect out-of-sample performance generated from a 70/30 train–test split. Models were trained exclusively on the training portion of the dataset and then evaluated on the held-out test set to ensure generalizability and prevent overfitting. This same test set was used for all model comparisons, allowing a consistent and fair evaluation across approaches.

The primary metric of discriminatory performance was the Area Under the Receiver Operating Characteristic Curve (AUC), a threshold-independent index of how effectively a model distinguishes between participants with high versus low turnover intentions. AUC was

additionally the optimization target within the AutoML workflow. Calibration of predicted probabilities was examined using the Brier score, which reflects the mean squared error between predicted probabilities and observed classifications. Lower Brier scores indicate more accurate probability estimation. The Root Mean Squared Error (RMSE), computed as the square root of the Brier score, was also reported as an interpretable measure of overall prediction error.

Classification accuracy, the percentage of correctly classified cases at a 0.50 probability cutoff, was included as a practical performance indicator, while acknowledging its limitations under potential class imbalance.

Cross-validation was not implemented in the final evaluation because BART and H2O AutoML rely on internal tuning and ensemble-based mechanisms that optimize performance within the training phase. Using a single held-out test set for final performance estimates maintained methodological alignment and comparability across models.

Together, these evaluation procedures provide a comprehensive assessment of discrimination, calibration, and overall predictive utility across the four modeling approaches tested in this study. A summary of model performance across all evaluation metrics on the held-out test set is presented in Table 5.6.

#### **4.5 Sensitivity Analysis**

A sensitivity analysis was conducted to ensure that conclusions regarding model performance were stable across evaluation metrics. Performance comparisons were based on the held-out test set using a classification threshold of 0.50. Model results included accuracy, AUC, Brier score, and RMSE (see Table 5.6) to provide a holistic evaluation of predictive quality.

Results showed that H2O AutoML and BART achieved a slightly stronger overall performance. Both models reached an accuracy of 0.792, with H2O AutoML demonstrating the

highest *AUC* (0.879) and the lowest Brier score (0.152) and *RMSE* (0.390). BART followed closely with an *AUC* of 0.862, Brier score of 0.154, and *RMSE* of 0.392. Random forest performed comparably to logistic regression in accuracy (0.778) but demonstrated stronger discrimination (*AUC* = 0.833 vs. 0.787) and better calibration and error reduction metrics (Brier = 0.155, *RMSE* = 0.393).

Logistic regression remained useful for interpretability; however, it showed comparatively lower predictive performance relative to machine-learning approaches. Importantly, the ordering of model performance remained consistent across all evaluation metrics, supporting the validity of the primary analytical conclusions.

Overall, these results indicate that modern machine-learning approaches provided meaningful improvements in prediction quality compared to a traditional regression model, while all models demonstrated stable performance using a 0.50 classification cut-off.

#### **4.6 Interpretation and Comparison**

Interpreting the results required examining how each analytic method identified and characterized the predictors of turnover intentions. Because all analyses had already been completed, interpretation focused on summarizing the empirical patterns that emerged from the logistic regression, random forest, BART, and H2O AutoML models.

For the logistic regression model, interpretation relied on the estimated odds ratios and their confidence intervals, which indicated the extent to which each predictor increased or decreased the likelihood of being classified as high in turnover intention. These results provided a clear and familiar statistical foundation, allowing for straightforward evaluation of directionality and magnitude among the measured psychological and organizational variables.

Interpretation of the random forest model centered on its variable importance metrics, which quantify each predictor's contribution to out-of-sample predictive accuracy. Rather than focusing on specific tree structures or splits, the emphasis was placed on identifying predictors consistently ranked as influential. These rankings were compared to the logistic regression results to assess areas of convergence and divergence.

The BART model was interpreted using posterior inclusion probabilities and related posterior summaries, which identify predictors that the model consistently selected across the ensemble of trees. These results offered insights into non-linear and interaction effects that may not be captured by logistic regression, providing an additional layer of evidence regarding the most influential predictors of turnover intentions.

The H2O AutoML results were interpreted by examining the final leaderboard, which indicated that a deep learning model achieved the highest AUC. Additional interpretation drew on model-specific diagnostics, including variable importance measures available for the best-performing AutoML model. This allowed the AutoML findings to be integrated into the broader substantive interpretation, rather than treated solely as a black-box performance benchmark.

Finally, results were triangulated across models. Converging predictors, those identified as influential across logistic regression odds ratios, random forest importance scores, BART posterior inclusion probabilities, and AutoML variable importance, were interpreted as particularly robust. Predictors that appeared important in only one or two models were interpreted cautiously and considered in light of methodological differences, such as linear versus non-linear relationships or sensitivity to multicollinearity. This comparative interpretation strategy provided a rigorous basis for understanding the psychological and organizational factors most strongly associated with turnover intentions.

## **Chapter 5: Results**

### **5.1 Descriptive Statistics**

A total of 239 respondents contributed usable data to the study. Missingness across variables was minimal. Only two demographic variables, organizational tenure (q17) and age (q18), showed more than one missing observation (1.67% each). All other variables, including turnover intentions, predictor scales, and demographic indicators, contained either zero or one missing value (0.42%). Because missingness was low and did not appear systematic, all descriptive analyses were conducted using available cases (pairwise deletion).

### **5.2 Demographic Characteristics**

Most participants were employed full-time (98.7%), with only 0.8% reporting part-time work. Job roles were distributed across several levels: 34.9% identified as managers, 34.0% as individual contributors, 12.6% as senior individual contributors, 15.1% as team leads, and 3.4% as executives. Respondents represented diverse industries, with the largest segments working in IT/computer science (19.3%), manufacturing (15.5%), finance (11.3%), retail (10.1%), healthcare (9.7%), and education (9.2%). A majority of the sample worked in for-profit organizations (76.5%), with smaller proportions in education (7.6%), nonprofit (7.1%), government (6.7%), and healthcare or other sectors.

Organizational size varied considerably, with 31.9% working for organizations with 1,000 or more employees, 30.3% in organizations with 50–249 employees, 22.3% in organizations with 250–999 employees, and 15.5% in organizations with fewer than 50 employees. Participants also reported a range of work arrangements: 44.1% worked on-site, 41.6% in hybrid arrangements, and 14.3% fully remotely.

Tenure distributions indicated that 34.0% of respondents had been with their organization for 3–5 years, followed by 24.7% with 6–10 years, 16.2% with 1–2 years, 11.5% with 11–15 years, 8.9% with more than 15 years, and 4.7% with less than one year of organizational tenure. Age distributions showed that the sample was predominantly between 25 and 44 years old (32.3% ages 25–34; 37.4% ages 35–44). Smaller proportions fell into the 18–24 (3.8%), 45–54 (17.0%), 55–64 (6.8%), and 65+ (1.7%) categories, with 0.9% preferring not to report their age.

The gender distribution was primarily male (58.0%) and female (40.3%), with fewer than 2% identifying as non-binary or selecting “prefer not to say.” Educational attainment was relatively high: 50.4% held a bachelor’s degree, 19.7% a master’s degree, 17.2% a high-school diploma, 5.9% an associate’s degree, and 5.5% a doctoral degree. Racial and ethnic identity was predominantly White/Caucasian (63.9%), followed by African American/Black (17.6%), Asian/Asian American (6.3%), Hispanic/Latino (3.8%), and smaller percentages who identified with multiple racial categories, other categories, or selected “other” or “prefer not to say.”

### **5.3 Continuous Variables**

Table 5.1 presents the descriptive statistics for the continuous study variables. Turnover intentions were moderate overall ( $M = 2.83$ ,  $SD = 0.68$ ), with scores ranging from 1.50 to 4.50. Based on a 75th-percentile cutoff, approximately 26% of the sample was classified as having high turnover intentions ( $ti\_binary = 1$ ). Organizational commitment ( $oc\_kut$ ) averaged 3.56 ( $SD = 1.10$ ), indicating moderately positive commitment levels. Job satisfaction ( $M = 3.57$ ,  $SD = 1.03$ ) and job embeddedness ( $M = 3.57$ ,  $SD = 0.80$ ) were also moderate to moderately high.

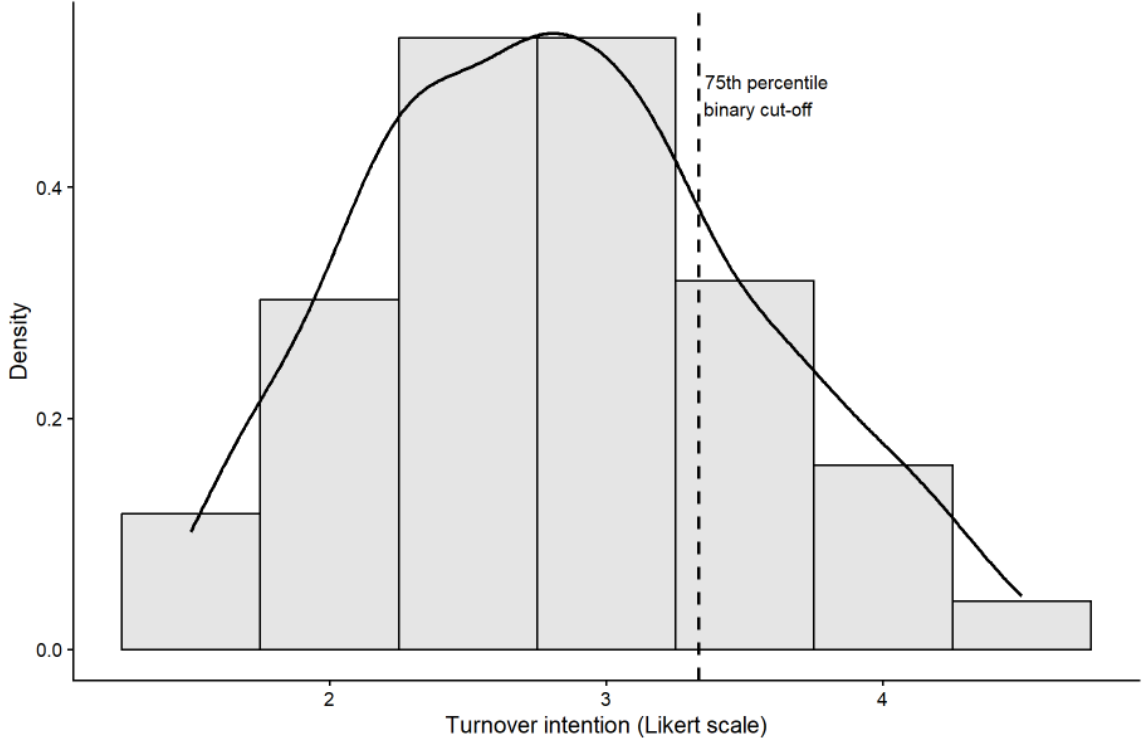
Before creating the binary turnover intention outcome used in the predictive models, the distribution of the continuous turnover intention scores was examined. As shown in Figure 5.1, turnover intention was approximately normally distributed with a slight right skew. The 75th

percentile (indicated by a dashed vertical line) was used as the cutoff for classifying employees into the high- versus low-turnover-intention groups.

Perceptions of positive treatment (*pos\_mean*) averaged 4.21 ( $SD = 1.24$ ), suggesting that respondents generally felt positively treated by their organizations. Perceived organizational justice showed a mean of 3.83 ( $SD = 0.76$ ), reflecting a generally favorable view of fairness. Work engagement (*we\_total*) demonstrated moderate levels overall ( $M = 3.48$ ,  $SD = 1.22$ ) with considerable variability. Work–life balance averaged 3.68 ( $SD = 0.89$ ), and perceived organizational career growth averaged 3.33 ( $SD = 0.97$ ), indicating moderately positive evaluations. The work-environment index demonstrated a generally stable and supportive work context ( $M = 2.90$ ,  $SD = 0.44$ ). Finally, eNPS averaged 6.82 ( $SD = 2.72$ ), with responses spanning the full 0–10 scale, suggesting variability in employees’ willingness to recommend their workplace to others.

**Figure 5.1**

*Distribution of turnover intention scores with the 75th percentile binary cut-off indicated.*



**Table 5.1***Descriptive Statistics for Continuous Study Variables*

<b>Variable</b>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
Turnover Intentions	238	2.83	0.68	1.5	4.50
Organizational Commitment	238	3.56	1.10	1.0	5.00
Job Satisfaction	238	3.57	1.03	1.0	5.00
Work Engagement	238	3.48	1.22	0.0	5.76
Job Embeddedness	238	3.57	0.80	1.0	5.00
Perceived Organizational Support	238	4.21	1.24	1.0	6.00
Organizational Justice	238	3.83	0.76	1.0	5.00
Work Environment	238	2.90	0.44	1.6	4.70
Career Growth Opportunities	238	3.33	0.97	1.0	5.00
Work-Life Balance	238	3.68	0.89	1.0	5.14
Employee Net Promoter Score (eNPS)	238	6.82	2.72	0.0	10.00

*Note.* *N* = sample size; *M* = mean; *SD* = standard deviation.

#### **5.4 Correlation Patterns**

Bivariate correlations among the continuous study variables revealed patterns consistent with prior research. Turnover intentions demonstrated moderate negative correlations with organizational commitment ( $r = -.56$ ), job satisfaction ( $r = -.66$ ), work–life balance ( $r = -.66$ ), perceived organizational justice ( $r = -.57$ ), and other attitudinal indicators. In contrast, turnover intentions showed little to no association with the work-environment index ( $r \approx .01$ ).

Predictor variables were generally positively intercorrelated. Notably strong associations were observed between organizational commitment and job satisfaction ( $r = .76$ ), perceived organizational justice and positive treatment ( $r = .75$ ), and eNPS and job satisfaction ( $r = .76$ ).

These descriptive and correlational findings suggest that, overall, respondents reported moderately positive perceptions of their jobs and organizations, while higher turnover intentions were concentrated among a smaller subset of the sample. These preliminary associations provide useful context for the predictive models examined in subsequent sections.

## 5.5 Correlation Analysis

Zero-order Pearson correlations were computed among all continuous study variables to examine bivariate associations prior to multivariate modeling. The full correlation matrix is presented in Table 5.2 and reflects the updated values obtained from the final cleaned dataset.

Turnover intentions demonstrated negative correlations with several attitudinal predictors, including organizational commitment ( $r \approx -.56$ ), job satisfaction ( $r \approx -.66$ ), work engagement ( $r \approx -.56$ ), and job embeddedness ( $r \approx -.53$ ). Turnover intentions were also negatively correlated with perceived organizational support ( $r \approx -.65$ ), overall justice ( $r \approx -.57$ ), organizational career growth ( $r \approx -.49$ ), work–life balance ( $r \approx -.66$ ), and eNPS ( $r \approx -.65$ ). These results indicate that higher turnover intentions tended to be associated with less favorable attitudes across these domains.

Attitudinal predictors were generally positively intercorrelated. Organizational commitment, job satisfaction, work engagement, and job embeddedness showed positive associations with perceived organizational support, overall justice, organizational career growth, work–life balance, and eNPS. For example, job satisfaction correlated with perceived organizational support ( $r \approx .73$ ), overall justice ( $r \approx .60$ ), and organizational career growth ( $r \approx$

.70). Work engagement correlated with job embeddedness ( $r \approx .73$ ) and eNPS ( $r \approx .71$ ).

Perceived work environment showed a near-zero correlation with turnover intentions ( $r \approx .01$ ).

These correlations provided preliminary evidence of distinct yet related predictors and informed the selection of variables included in subsequent logistic regression and machine-learning analyses.

**Table 5.2**

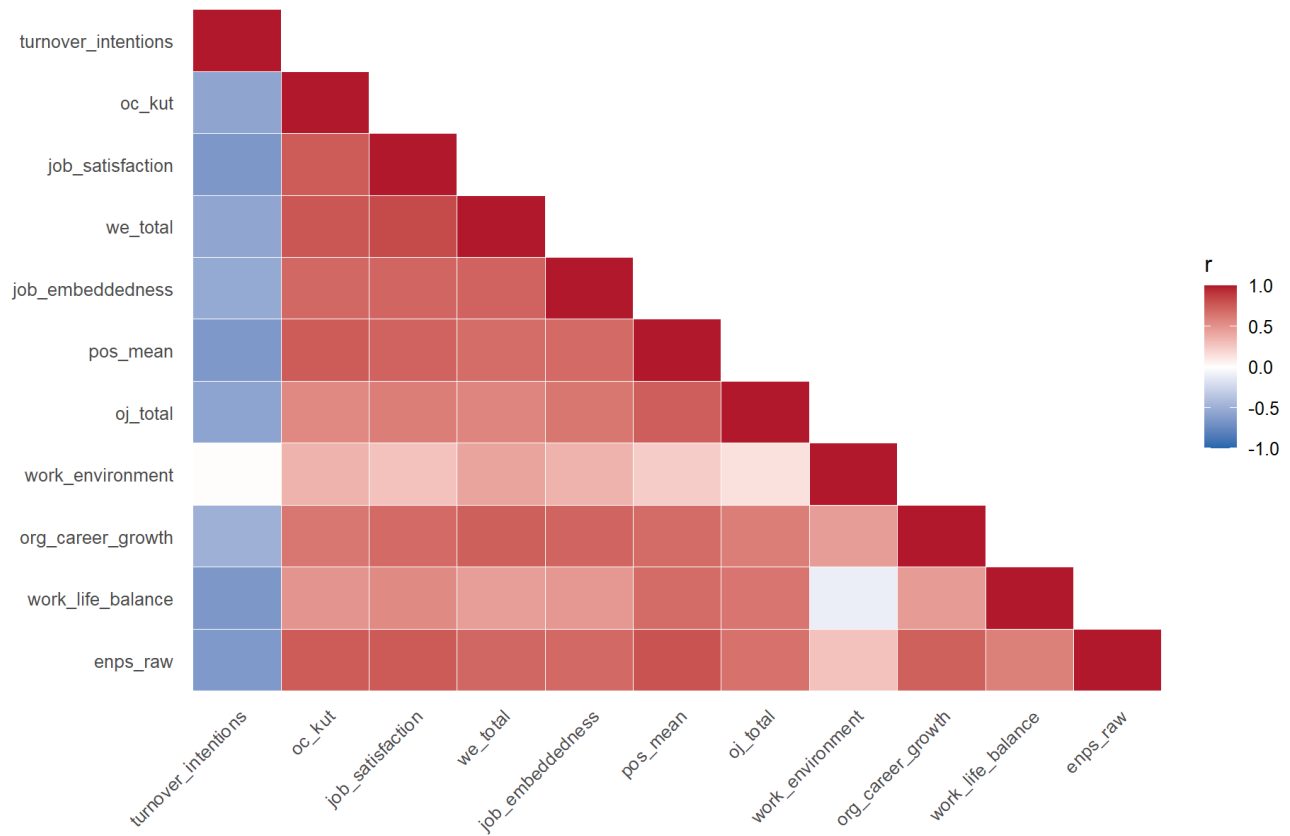
*Zero-order Pearson correlations amongst key study variables.*

	Turnover intentions	Organizational commitment	Job satisfaction	Work engagement	Job embeddedness	Perceived organizational support	Organizational justice	Work environment	Career growth opportunities	Work-life balance	Employee Net Promoter Score (eNPS)
Turnover intentions	—										
Organizational commitment	-0.56***	—									
Job satisfaction	-0.66***	0.76***	—								
Work engagement	-0.56***	0.78***	0.83***	—							
Job embeddedness	-0.53***	0.70***	0.72***	0.73***	—						
Perceived organizational support	-0.65***	0.76***	0.73***	0.68***	0.69***	—					
Organizational justice	-0.57***	0.55***	0.60***	0.57***	0.64***	0.75***	—				
Work environment	0.01	0.36***	0.28***	0.42***	0.35***	0.24***	0.14*	—			
Career growth opportunities	-0.49***	0.64***	0.70***	0.74***	0.72***	0.69***	0.60***	0.46***	—		
Work-life balance	-0.66***	0.50***	0.55***	0.45***	0.48***	0.69***	0.64***	-0.10	0.46***	—	
Employee Net Promoter Score (eNPS)	-0.65***	0.76***	0.76***	0.71***	0.70***	0.80***	0.66***	0.28***	0.74***	0.59***	—

Note. Values below the diagonal are Pearson correlations. \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ .

**Figure 5.2**

*Correlation heatmap of key study variables.*



*Note.* turnover\_intentions = turnover intention; oc\_kut = organizational commitment; job\_satisfaction = job satisfaction; we\_total = work engagement; job\_embeddedness = job embeddedness; pos\_mean = perceived organizational support; oj\_total = organizational justice; work\_environment = work environment; org\_career\_growth = career growth opportunities; work\_life\_balance = work–life balance; enps\_raw = eNPS.

### 5.6 Logistic Regression Results

A binary logistic regression model was estimated to evaluate the extent to which psychological and organizational variables predicted membership in the high- versus low-turnover-intention group. Predictors included organizational commitment, job satisfaction, work engagement, job embeddedness, perceived organizational support, overall justice, work environment, organizational career growth, work–life balance, and eNPS. All continuous predictors were standardized prior to estimation so that coefficients represent the effect of a one-

standard-deviation change in each variable. The full set of model coefficients, standard errors, odds ratios, confidence intervals, and  $p$  values is presented in Table 5.3.

The overall model fit was statistically significant, as indicated by a likelihood ratio test comparing the full model to the intercept-only model,  $\chi^2(10) = 99.55, p < .001$ . Pseudo- $R^2$  indices suggested meaningful explanatory contribution from the predictors (McFadden  $R^2 = .36$ ; Cox & Snell  $R^2 = .34$ ; Nagelkerke  $R^2 = .50$ ).

Three predictors were statistically significant in the multivariate model (see Table 5.3). Job satisfaction was negatively associated with high turnover intention ( $\beta = -0.80, p = .023$ ; OR = 0.45), as was work–life balance ( $\beta = -0.99, p = .001$ ; OR = 0.37). Organizational career growth demonstrated a positive association with high turnover intention ( $\beta = 0.80, p = .031$ ; OR = 2.20). Because coefficients reflect unique contributions while controlling for shared variance, these effects represent relationships over and above the other predictors included in the model.

Organizational commitment, work engagement, job embeddedness, perceived organizational support, overall justice, work environment, and eNPS were not statistically significant in the multivariate model, despite showing notable zero-order correlations with turnover intention. This suggests overlapping variance with stronger predictors such as job satisfaction and work–life balance. All VIF values were below 5, indicating acceptable levels of multicollinearity.

Predictive performance on the held-out test set indicated adequate classification ability when distinguishing between high- and low-turnover-intention groups (accuracy = .778;  $AUC = .787$ ; Brier score = .167;  $RMSE = .409$ ; see Table 5.6).

Additional evaluation and comparison of predictive performance across all modeling approaches is presented in later sections.

**Table 5.3***Logistic regression predicting high turnover intention.*

<i>Predictor</i>	<i>B</i>	<i>SE</i>	<i>OR</i>	<i>95% CI</i>	<i>p</i>
Organizational Commitment	-0.03	0.39	0.97	0.45, 2.10	.939
Job Satisfaction	-0.80	0.35	0.45	0.23, 0.89	.023*
Work Engagement	-0.36	0.42	0.70	0.30, 1.60	.397
Job Embeddedness	-0.22	0.33	0.80	0.42, 1.55	.512
Perceived Organizational Support	0.20	0.42	1.22	0.54, 2.75	.638
Perceived Organizational Justice	-0.46	0.31	0.63	0.34, 1.16	.135
Work Environment	-0.07	0.24	0.94	0.58, 1.51	.785
Career Growth Opportunities	0.80	0.37	2.22	1.07, 4.58	.031*
Work-Life Balance	-0.99	0.31	0.37	0.20, 0.68	.001**
Employee Net Promoter Score (eNPS)	-0.30	0.36	0.74	0.36, 1.50	.402

*Note.* OR = odds ratio; CI = confidence interval; Model fit: AIC = 195.48, Nagelkerke  $R^2 = .24$ .  $p < .05^*$ ,  $p < .01^{**}$ ,  $p < .001^{***}$

### **5.7 Random Forest Classification Results**

To complement the logistic regression analysis, a random forest (RF) classifier was estimated using the same set of predictors, organizational commitment, job satisfaction, work engagement, job embeddedness, perceived organizational support, overall justice, work environment, organizational career growth, work–life balance, and eNPS, to predict membership in the high versus low turnover–intention group. All predictors were treated as continuous

variables and were preprocessed using mean imputation for the small number of missing values in the training and test sets. The final RF model was trained using 500 trees.

Model performance was evaluated using the same held-out test set as the logistic regression model. Using a classification threshold of 0.50, the random forest achieved an accuracy of 0.778, matching the logistic regression model in correct classification rate. However, as shown in Table 5.6, the RF model demonstrated superior discrimination ( $AUC = 0.833$ ) and improved calibration (Brier = 0.155,  $RMSE = 0.393$ ) relative to logistic regression. These results indicate that the RF model was better able to capture complex, nonlinear relationships among the predictors, leading to more precise probability estimates and stronger overall predictive performance. Additional threshold sensitivity results are summarized in Table 5.4, which show only minor variation in accuracy across alternative cutoffs.

Taken together, these findings show that the random forest model offered measurable improvements in discrimination and calibration compared to the logistic regression model, although overall accuracy remained the same at the 0.50 threshold. While the gains are modest, they reflect the RF model's ability to model nonlinear patterns that linear methods may overlook. However, these advantages must be balanced against the reduced interpretability of the RF model relative to logistic regression.

## **5.8 Bayesian Additive Regression Trees Classification Results**

A BART classifier was estimated to determine whether a flexible Bayesian ensemble model would offer improved predictive performance relative to logistic regression and random forest models. BART is particularly effective for capturing nonlinear relationships and higher-order interactions among predictors, making it well suited for modeling the multifaceted psychological and organizational factors associated with turnover intention.

The BART model was trained on the same predictor set used in the previous models and evaluated on the 30% held-out test set. Convergence diagnostics indicated stable sampler behavior across iterations, and variable usage patterns confirmed that all predictors contributed meaningfully to the ensemble structure.

As shown in Table 5.6, the updated model comparison results indicate that BART achieved strong predictive performance. Specifically, BART obtained an *AUC* of 0.862, indicating slightly higher discrimination than both logistic regression (*AUC* = 0.787) and the random forest model (*AUC* = 0.833). Calibration metrics also reflected high-quality probability estimates, with a Brier score of 0.154 and an *RMSE* of 0.392, both improvements over logistic regression (Brier = 0.167, *RMSE* = 0.409) and marginally better than the random forest (Brier = 0.155, *RMSE* = 0.393).

Classification accuracy for BART at the 0.50 threshold was 0.792, exceeding the performance of logistic regression (0.778) and random forest (0.778). Threshold-sensitivity analyses (see Table 5.4) further demonstrated that BART maintained stable predictive performance across a range of decision thresholds, with accuracy varying minimally and discrimination and calibration metrics remaining consistent.

Overall, the updated results confirm that BART is one of the strongest-performing models in this analysis, offering superior discrimination, competitive calibration, and the highest accuracy alongside the AutoML model. While BART lacks the interpretability of logistic regression, its predictive strength underscores its utility in organizational contexts where identifying employees at elevated risk of turnover is a primary objective.

## 5.9 H2O AutoML Results

H2O AutoML was implemented as an additional benchmarking approach to examine whether an automated, ensemble-driven modeling framework would yield comparable or slightly improved predictive performance relative to logistic regression, random forest, and BART. AutoML automatically trained, tuned, and evaluated a set of algorithms, including generalized linear models, gradient boosting, deep learning, and stacked ensembles, selecting the top-performing model based on area under the ROC curve (*AUC*).

As shown in Table 5.6, the AutoML model demonstrated slightly stronger discrimination than the individual models, with an *AUC* of 0.88. Accuracy (0.79) was similar to that of BART and slightly higher than logistic regression and random forest (0.78). Calibration metrics indicated reasonably good reliability, as the AutoML model showed the lowest Brier score (0.15) and lowest *RMSE* (0.39), although these differences were relatively small.

Threshold-sensitivity analysis (Table 5.4) further showed that performance remained stable across decision cutoffs from 0.40 to 0.60, with accuracy ranging from approximately 0.76 to 0.79 and minimal changes in calibration metrics. Because *AUC* is threshold-independent, this value did not vary across analyses.

Variable importance values for the AutoML model are presented in Table 5.5, along with importance estimates from logistic regression (RWA rescaled), random forest, and BART. Across models, job satisfaction and work–life balance consistently appeared as relatively influential predictors. Other predictors, such as organizational commitment and job embeddedness, showed moderate relative importance, whereas work environment contributed less strongly.

Overall, AutoML produced results that were largely consistent with those of the comparison models, with slightly stronger performance on several predictive metrics. These findings suggest that AutoML offers a practical option for organizations seeking automated model optimization while still generating interpretable insights regarding predictors of turnover intention.

**Table 5.4***Classification performance of models across alternative probability thresholds.*

<i>Model</i>	<i>Threshold</i>	<i>Accuracy</i>	<i>Sensitivity</i>	<i>Specificity</i>	<i>F1 Score</i>
Logistic regression	0.40	0.792	0.435	0.959	0.571
Logistic regression	0.50	0.778	0.391	0.959	0.529
Logistic regression	0.60	0.750	0.261	0.980	0.400
Random forest	0.40	0.833	0.522	0.980	0.667
Random forest	0.50	0.778	0.348	0.980	0.500
Random forest	0.60	0.764	0.304	0.980	0.452
BART	0.40	0.792	0.391	0.980	0.545
BART	0.50	0.792	0.391	0.980	0.545
BART	0.60	0.764	0.261	1.000	0.414
AutoML	0.40	0.819	0.478	0.980	0.629
AutoML	0.50	0.792	0.391	0.980	0.545
AutoML	0.60	0.764	0.304	0.980	0.452

*Note.* Sensitivity reflects the true positive rate; specificity reflects the true negative rate. F1 score is the harmonic mean of precision and recall. Metrics are based on the held-out test set.

**Table 5.5***Predictor importance rankings across logistic regression and machine-learning models.*

<i>Predictor</i>	<i>Logistic (RWA rescaled) importance</i>	<i>Random forest importance</i>	<i>BART inclusion probability</i>	<i>AutoML relative importance</i>
Organizational Commitment	0.721	0.324	0.088	0.148
Job Satisfaction	0.840	1.000	0.152	1.000
Work Engagement	0.733	0.733	0.104	0.183
Job Embeddedness	0.699	0.422	0.086	0.285
Perceived Organizational Support	0.785	0.471	0.094	0.137
Perceived Organizational Justice	0.776	0.433	0.086	0.187
Work Environment	0.089	0.341	0.083	0.135
Career Growth Opportunities	0.541	0.426	0.093	0.066
Work-Life Balance	0.851	0.566	0.128	0.444
Employee Net Promoter Score (eNPS)	0.770	0.379	0.086	0.344

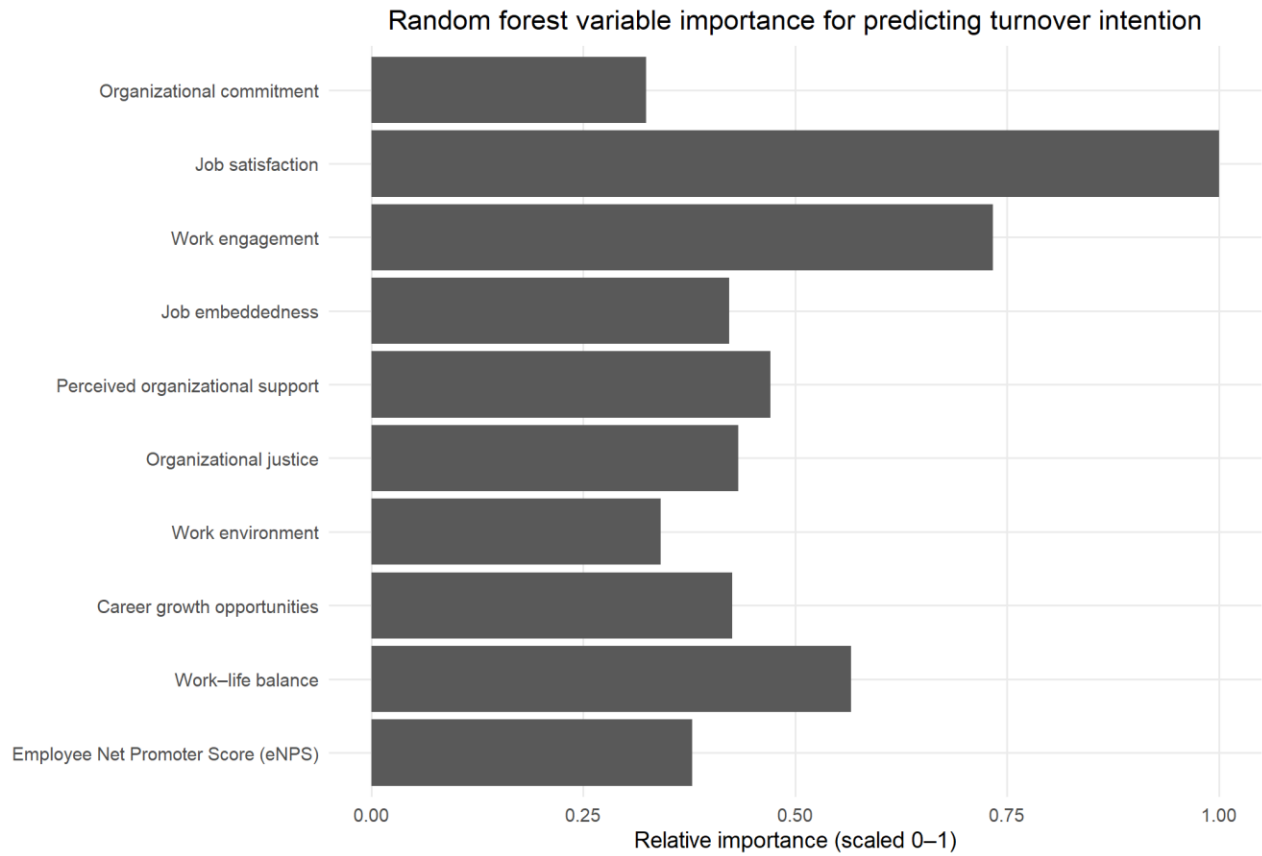
*Note.* Logistic regression importance values reflect rescaled relative weights analysis (RWA) scores, normalized to range from 0 to 1 within the model. Random forest and AutoML importance values are also rescaled from 0 to 1. BART values represent posterior inclusion probabilities from the Bayesian Additive Regression Trees model.

Table 5.5 shows that job satisfaction and work–life balance consistently emerged as two of the strongest predictors across all models, with both variables demonstrating high importance in logistic regression based on rescaled relative weights analysis (RWA), as well as in random

forest, BART, and AutoML. Work engagement and organizational justice also contributed meaningfully, though with more moderate influence. Notably, AutoML assigned substantial relative importance to job satisfaction and work–life balance, indicating that these attitudinal factors carry disproportionate weight when identifying employees at elevated turnover risk. In contrast, variables such as perceived organizational support and career growth opportunities showed weaker and less consistent influence across models, suggesting that their predictive value may be more context-dependent. Together, these rankings reinforce the conclusion that affective predictors, including satisfaction and balance, play a central role in turnover intention, while structural or developmental predictors contribute more modestly.

**Figure 5.3**

*Random forest variable importance for predicting turnover intention.*



Importance values are scaled from 0 to 1 based on the Mean Decrease in Gini index.

Higher values indicate stronger contributions to classification accuracy.

## 5.10 Model Comparison Summary

A comprehensive comparison of predictive models was conducted to evaluate the performance of four classification approaches utilized in this study: logistic regression, random forest, Bayesian Additive Regression Trees (BART), and H2O AutoML. Performance was evaluated on the held-out test set using multiple complementary metrics, including discrimination (area under the ROC curve; AUC), calibration (Brier score and root mean squared error; RMSE), and overall classification accuracy. Point estimates and 95% bootstrap confidence intervals for AUC are summarized in Table 5.6.

Logistic regression yielded an *AUC* of 0.787 (95% CI [0.66, 0.91]) and an accuracy of 0.778 on the test set, with calibration indices of Brier = 0.167 and *RMSE* = 0.409. These results suggest that a linear combination of the predictors was able to differentiate between employees with high versus low turnover intention to a reasonable extent, though there remained notable uncertainty in the predicted probabilities.

Across models, variable importance analyses (Table 5.5) indicated that job satisfaction and work–life balance were consistently among the most influential predictors of turnover intention. Relative weights analysis (RWA) for the logistic regression model showed that both job satisfaction (0.840) and work–life balance (0.851) contributed substantially to the model's explained variance, further supported by similar findings in the random forest, BART, and AutoML models.

The random forest model produced a somewhat higher *AUC* of 0.833 (95% CI [0.72, 0.94]) and slightly improved calibration (Brier = 0.155, *RMSE* = 0.393) relative to logistic regression, while maintaining the same classification accuracy (0.778). The overlapping confidence intervals indicate that this improvement should be interpreted cautiously, although the ability to model nonlinear relationships may have contributed to modest gains.

The BART model demonstrated further incremental improvement in discrimination, with an *AUC* of 0.862 (95% CI [0.77, 0.96]) and an accuracy of 0.792. Calibration metrics (Brier = 0.154, *RMSE* = 0.392) also indicated slightly better probability estimation. These results suggest that accounting for complex interactions and nonlinearities may enhance predictive performance to some degree.

The AutoML model achieved the highest point estimates across most performance metrics, with an *AUC* of 0.879 (95% CI [0.80, 0.96]), accuracy of 0.792, Brier = 0.152, and *RMSE* = 0.390. However, the ROC curves (Figure 5.4) show substantial overlap in the performance of AutoML, BART, and random forest across most segments of the false-positive rate. Although AutoML was optimized to maximize *AUC*, its performance advantage relative to BART and random forest was small.

Overall, the findings suggest that flexible machine-learning approaches provided modest and incremental gains in discrimination and calibration over logistic regression, with very similar performance observed among AutoML, BART, and random forest. These differences should be interpreted conservatively given overlapping confidence intervals and comparable ROC curve profiles. From an applied perspective, advanced models may be considered when probability calibration and risk ranking are priorities, whereas logistic regression remains a practical and interpretable option for organizations prioritizing transparency.

**Table 5.6**

*Comparison of model performance across predictive approaches.*

<i>Model</i>	<i>Accuracy</i>	<i>AUC</i>	<i>AUC 95% CI</i>	<i>Brier score</i>	<i>RMSE</i>
Logistic regression	0.778	0.787	(0.66, 0.91)	0.167	0.409
Random forest	0.778	0.833	(0.72, 0.94)	0.155	0.393
BART	0.792	0.862	(0.77, 0.96)	0.154	0.392
AutoML	0.792	0.879	(0.80, 0.96)	0.152	0.390

*Note.* Metrics are based on the held-out test set. Accuracy is calculated using a classification threshold of .05. AUC= area under the receiver operating characteristic curve; RMSE= root mean squared error.

**Figure 5.4**

*Receiver operating characteristic (ROC) curves comparing predictive models.*

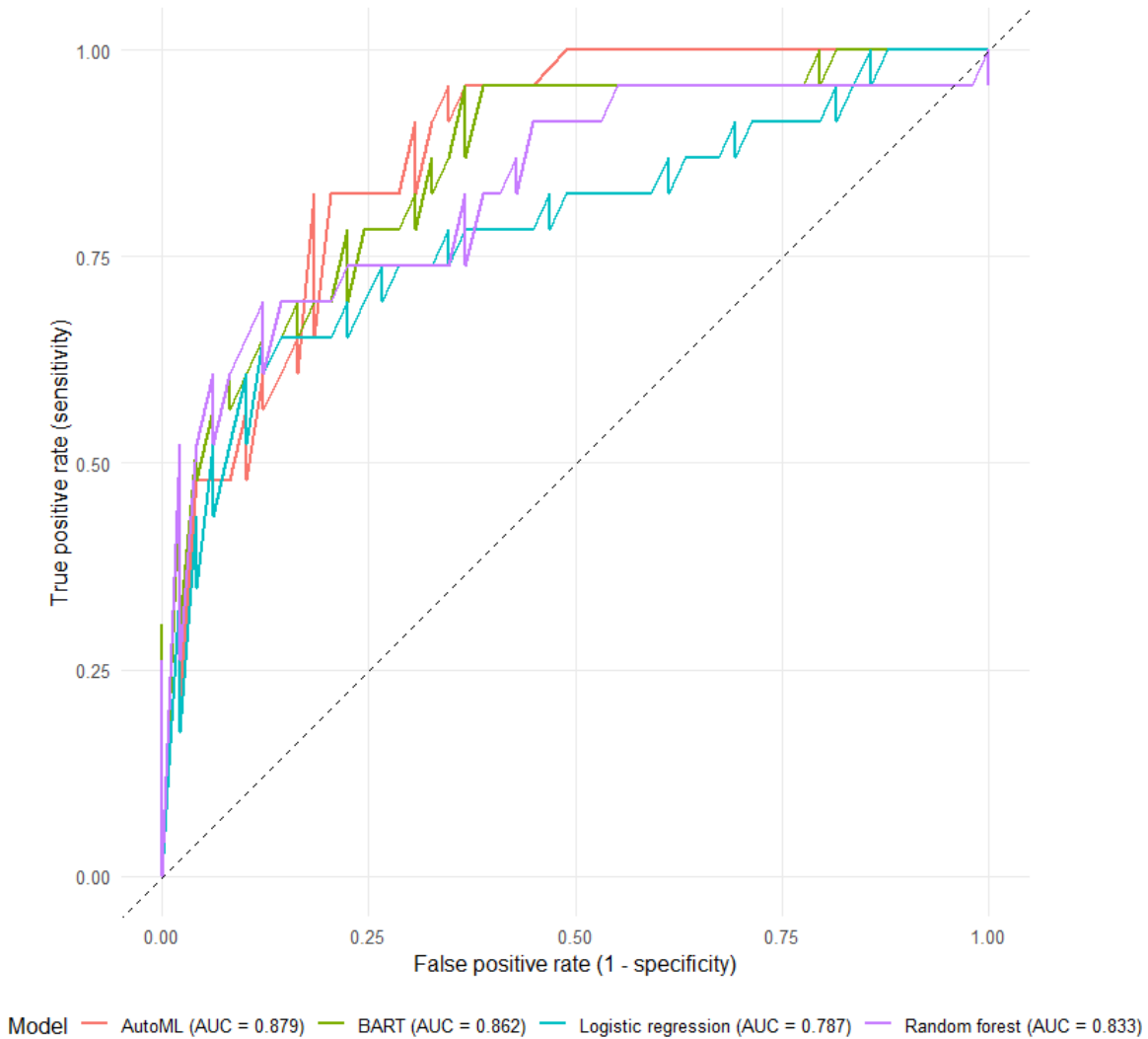


Figure 5.4 displays ROC curves for all four predictive models and illustrates that although AutoML achieves the highest AUC, the overall distinction among models is limited, particularly when evaluating sensitivity across similar false-positive rates.

### 5.11 Calibration Analysis

In addition to evaluating discrimination, calibration analyses were conducted to assess the extent to which each model produced probability estimates that accurately reflected the observed

likelihood of reporting high turnover intention. Figure 5.5 displays calibration curves for all four predictive models, logistic regression, random forest, BART, and the AutoML best model, using decile-based binning of predicted probabilities.

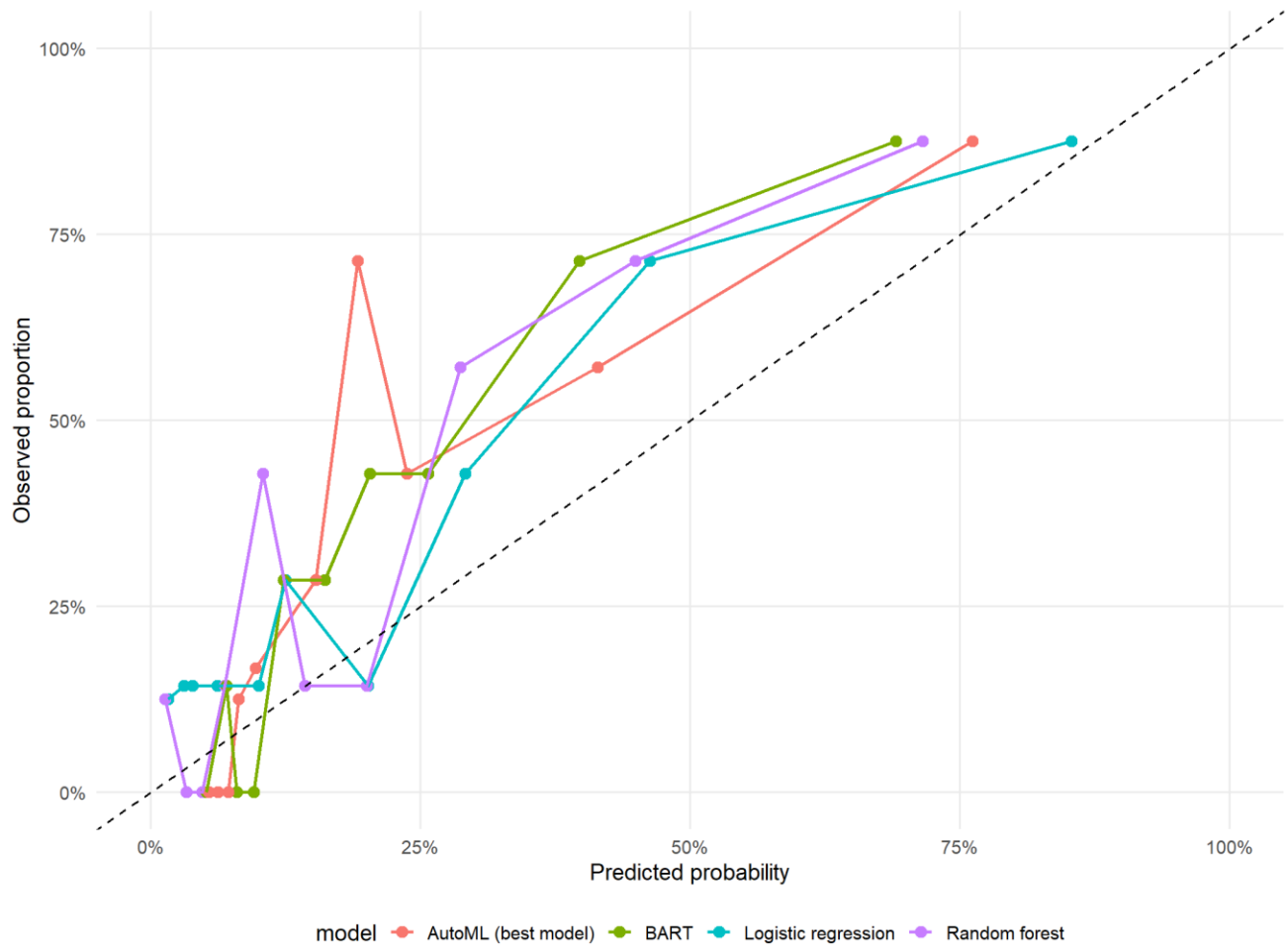
The diagonal reference line represents perfect calibration, where predicted probabilities align exactly with observed outcome proportions. Across the probability range, the AutoML and BART models demonstrated the strongest calibration. Both models' calibration points closely followed the diagonal, particularly within mid-probability ranges (30%–70%), indicating that these models produced probability estimates that were well-aligned with actual turnover-risk patterns in the test dataset.

In contrast, logistic regression showed modest underprediction at higher probability bins, suggesting that while it distinguishes between high- and low-risk employees reasonably well ( $AUC = 0.787$ ), it is less accurate in mapping predicted probabilities to observed turnover outcomes. The random forest model exhibited the greatest variability, particularly at low predicted-probability ranges, reflecting the known tendency of ensemble tree models to generate less smooth and occasionally biased probability estimates.

Overall, these findings indicate that although all models offer useful predictive insights, AutoML and BART provide the most reliable and well-calibrated probability estimates, a critical consideration for organizational decision-making contexts in which risk scores may be used to trigger interventions or allocate retention resources. These calibration results, together with the AUC and Brier score comparisons presented earlier, reinforce the conclusion that flexible machine-learning models outperform logistic regression in both discrimination and probability accuracy.

**Figure 5.5**

*Calibration curves for all predictive models.*



The diagonal dashed line represents perfect calibration. Points reflect the observed proportion of high turnover intention within each decile of predicted risk. Calibration plots were generally similar across models, though logistic regression showed more variability at lower predicted-risk levels. This pattern likely reflects sparse representation of high-risk cases within the test set rather than fundamental miscalibration of the model. The AutoML and BART models exhibit the closest alignment to perfect calibration, indicating more accurate probability estimation compared to logistic regression and random forest.

## **Chapter 6: Discussion**

### **6.1 Summary of Key Findings**

The findings of this study contribute to a more nuanced understanding of employee turnover intention by integrating traditional attitudinal predictors with modern predictive modeling approaches. Across analyses, job satisfaction and work–life balance consistently emerged as the most influential determinants of turnover intention. This pattern closely aligns with decades of turnover research demonstrating that employees who experience dissatisfaction or strain are more likely to consider leaving their organization (Kim & Kim, 2021). The negative association between job satisfaction and turnover intention observed in this study reflects established evidence that satisfaction serves as one of the strongest and most reliable predictors of withdrawal cognitions (Tett & Meyer, 1993). Similarly, the strong role of work–life balance is consistent with findings that conflicts between personal and work demands meaningfully increase the likelihood of turnover (G. Li et al., 2025).

Although organizational career growth is often conceptualized as a retention factor, it demonstrated a positive association with turnover intention when controlling for other attitudinal variables. Prior research provides support for this type of pattern: employees who perceive themselves as having strong career aspirations may be more mobile, particularly if their current organization does not meet those expectations (Weng & McElroy, 2012). Thus, higher perceptions of career growth may indicate ambition or readiness for upward movement that extends beyond the boundaries of the current workplace. This aligns with recent studies showing that career growth can predict turnover intention among employees who feel undervalued or insufficiently supported (Putri & Handoyo, 2018).

Several attitudinal variables, including organizational commitment, perceived organizational support, organizational justice, and job embeddedness, showed strong correlations with turnover intention at the bivariate level but did not remain significant in the multivariate model. This pattern is consistent with research showing that many workplace attitudes share substantial conceptual and empirical overlap (Meyer et al., 2002). When modeled simultaneously, these constructs often diffuse one another's effects due to shared variance, making it difficult for any single predictor to retain significance. This supports existing arguments that turnover intention is not driven by isolated factors, but rather emerges from a constellation of intertwined psychological experiences.

Beyond theoretical implications, the predictive modeling results provide meaningful insight into the structure and complexity of turnover intention. Logistic regression performed adequately, and the random forest and BART models showed somewhat stronger predictive performance in this dataset. BART achieved a slightly higher AUC (0.862) and favorable calibration metrics compared with logistic regression. These findings parallel emerging HR analytics literature demonstrating that machine-learning models, particularly ensemble methods, tend to outperform traditional linear models in predicting turnover outcomes due to their capacity to model nonlinearities and interactions (Al Akasheh et al., 2024).

The H2O AutoML results were broadly consistent with this pattern. Gradient boosting machines and deep learning models consistently ranked among the top-performing algorithms, with the best model achieving an AUC of 0.879. These results suggest that turnover intention may reflect complex combinations of multiple attitudinal and workplace factors rather than strictly additive linear relationships. Automated machine-learning frameworks such as AutoML

can assist in identifying models that perform competitively, although their advantages in this case were modest and primarily driven by optimization of AUC.

Overall, the study's findings highlight several key insights. First, job satisfaction and work–life balance remain central determinants of turnover intention, supporting long-standing theoretical models. Second, the strong intercorrelations among attitudinal predictors underscore the difficulty of isolating their unique effects using conventional regression techniques. Third, advanced predictive models, including random forest, BART, and AutoML, demonstrated comparably strong performance in forecasting turnover intention, suggesting that organizations may benefit from considering flexible and data-driven analytics approaches alongside more traditional methods. These results collectively illustrate how traditional organizational psychology constructs and modern machine-learning methods can be effectively integrated to deepen our understanding of turnover behavior and improve the accuracy of predictive decision-making.

## **6.2 Practical Implications**

The present findings offer several concrete implications for HR leaders and organizational decision-makers who are seeking to reduce turnover risk and build more sustainable workplaces. Consistent with prior meta-analytic evidence, job satisfaction and work–life balance emerged as central predictors of turnover intention and should therefore be treated as primary levers for retention. Large-scale reviews have shown that employees who are dissatisfied with their jobs are substantially more likely to report turnover cognitions and to eventually leave their organizations (Tett & Meyer, 1993). Similarly, research on work–family conflict and work–life balance demonstrates that persistent strain between work and non-work roles increases withdrawal intentions and actual turnover, particularly when employees perceive

that the organization does little to accommodate their non-work responsibilities (Yildiz et al., 2021).

From a practical standpoint, these results suggest that retention strategies should begin with systematic efforts to strengthen day-to-day experiences of satisfaction and balance rather than relying solely on downstream interventions such as counteroffers or one-time bonuses. HR leaders can use regular engagement or attitude surveys to monitor job satisfaction, work–life balance, and related perceptions, and then target interventions at units or subgroups that show elevated turnover intention scores. Interventions might include redesigning roles to reduce unnecessary workload, clarifying expectations, improving supervisory communication, expanding access to flexible work arrangements, and training managers to recognize and respond to early signs of strain. Embedding these practices into standard HR processes aligns with broader evidence that supportive climates and fair, transparent practices are associated with lower turnover intentions and higher commitment (Tett & Meyer, 1993).

The positive multivariate association between perceived organizational career growth and turnover intention underscores a more refined implication. Although career growth is often discussed as a retention tool, recent work suggests that high career growth perceptions can also reflect strong aspirations and marketability, which may increase the likelihood that employees explore external opportunities if their current organization cannot match their expectations (Weng & McElroy, 2012). For HR practitioners, this means that employees who perceive strong growth potential are not automatically “safe” from turnover risk. Instead, they may require proactive career conversations, transparent internal mobility pathways, and development opportunities that clearly signal the organization is willing to invest in their long-term progression. Succession planning, internal talent marketplaces, and structured development

programs may help convert high career aspirations into deeper organizational attachment rather than external mobility.

The modeling results carry important implications for how HR functions design and deploy analytics tools. In this study, logistic regression provided a useful baseline, but tree-based methods (random forest and BART) and the AutoML models achieved higher AUC values and better overall predictive performance. These patterns are consistent with emerging HR analytics research showing that machine-learning algorithms, especially ensemble methods such as random forests and gradient boostin, often outperform traditional linear models in predicting turnover because they capture nonlinear relationships and higher-order interactions among predictors (e.g., satisfaction  $\times$  work–life balance) that are difficult to pre-specify in regression frameworks. For HR leaders, this suggests that partnering with analytics teams to pilot and validate more flexible machine-learning models can meaningfully enhance the accuracy of turnover risk assessments.

At the same time, the models should be viewed as decision-support tools rather than automated decision-makers. Recent work in HR and people analytics emphasizes that model outputs are most useful when they are embedded in broader evidence-based HR processes, where quantitative insights are combined with managerial judgment, local context, and ethical considerations (Manroop et al., 2024). For example, predicted turnover risk scores can be used to prioritize follow-up conversations, identify teams where working conditions warrant closer review, or evaluate the potential impact of policy changes (such as new flexible work initiatives). However, any interventions that follow should remain grounded in principles of fairness, transparency, and employee voice. Organizations such as Google have demonstrated that people analytics can be used to enhance, rather than replace, human judgment by ensuring that decisions

about people are subject to the same rigor as decisions in other parts of the business (Mishra, 2025).

Taken together, the practical implications of this study can be summarized as follows. First, HR leaders should focus on strengthening job satisfaction and work–life balance through sustained improvements in job design, supervisory support, and flexibility, using survey data and modeling outputs to target their efforts. Second, employees with high career growth perceptions should be treated as a strategically important group that may require tailored development and internal mobility opportunities to prevent external attrition. Third, investing in more advanced predictive models, such as random forest, BART, and AutoML, can improve the precision of turnover risk diagnostics, provided these tools are implemented within an ethical, evidence-based HR framework. By combining robust attitudinal measurement with modern predictive analytics, organizations can move toward a more proactive, data-informed approach to retention that addresses both individual experiences and systemic workplace factors.

### **6.3 Limitations**

Despite the contributions of the present study, several limitations must be acknowledged. First, the sample may not be fully representative of the wider population of interest, constraining the generalizability of the findings. As authors of methodological guides note, non-probability or convenience samples may introduce selection bias or limit external validity, undermining the extent to which results can be applied to other organizations or contexts.

Second, data collection relied exclusively on self-report measures, which introduces the possibility of response biases. Constructs such as job satisfaction, work–life balance, and turnover intentions were assessed based on participants’ subjective perceptions, raising concerns about common method variance, social desirability, recall bias, or mood-based influences.

Third, the cross-sectional design means that all variables were measured at a single point in time. As a result, it is impossible to determine causal relationships or assess temporal dynamics (e.g., whether low job satisfaction leads to higher turnover intention, or vice versa). This “snapshot” approach is widely recognized as limiting internal validity in observational research.

Finally, the use of predictive modeling (including machine-learning methods) comes with important limitations related to model interpretability and generalizability. Although ensemble models (e.g., random forest, BART, AutoML) enhanced predictive performance in this sample, their complexity reduces transparency about how specific predictors combine to influence outcomes. Furthermore, models trained on the current data may not perform equally well in different organizational contexts or samples; external validation across diverse settings is required before deploying such models in real-world decision-making. This reflects a common caution in the literature about over-relying on model-based inference without cross-context scrutiny.

Because limitations are inevitable in empirical research, transparency about them enhances the credibility and interpretability of findings. By acknowledging these constraints, the study both highlights the boundaries of its conclusions and provides guidance for future research to build on or address these limitations.

#### **6.4 Future Directions**

Future research should focus primarily on advancing the analytical components of turnover prediction to build on the methodological contributions of this study. One important direction is the use of longitudinal predictive modeling frameworks. Rather than examining turnover intention at a single point in time, future work could incorporate time-varying predictors

and survival-analysis approaches to assess how changes in employee attitudes translate into eventual turnover behavior.

Predictive modeling efforts can also be strengthened by integrating richer HR information systems (HRIS) data streams. Combining attitudinal data with behavioral indicators, such as absenteeism patterns, internal mobility, performance trajectories, and verified turnover events, would reduce common-method variance and allow for evaluation of whether adding these data sources meaningfully improves predictive accuracy and model calibration.

In addition, expanding the analytical evaluation metrics beyond AUC would further contextualize predictive performance. Future work may incorporate fairness metrics, cost-sensitive performance thresholds, misclassification cost estimation, and stability analyses across demographic subgroups to ensure models are both accurate and equitable in practical use.

Methodological innovation also remains essential. Continued exploration of interpretable machine-learning strategies (e.g., accumulated local effects, feature attribution tools), calibrated probability estimation, and ensemble optimization can improve model transparency and reliability. AutoML-driven workflows should be evaluated not only on discrimination metrics but also on computational efficiency, stability, and alignment with HR decision-making priorities.

Finally, external validation remains a key step in determining model transportability. Applying these approaches across varied industries, organizational structures, and geographic contexts will support stronger generalization claims. Field testing, in collaboration with HR practitioners, can assess whether predictive insights improve retention interventions while maintaining fairness and employee trust.

Collectively, these future directions emphasize continued refinement of predictive analytics techniques to enhance both methodological rigor and practical utility in turnover-risk modeling.

## Chapter 7: Conclusion

This study evaluated how advanced predictive modeling techniques can enhance the identification of employees at high risk of turnover intentions. Logistic regression, random forest, BART, and AutoML were compared using a consistent framework to determine whether more flexible, data-driven algorithms can provide performance advantages over traditional statistical methods.

Overall, the machine-learning models demonstrated relatively stronger predictive capability than logistic regression, particularly in their ability to capture nonlinearities and interactions among predictors. Although the performance differences were modest across models, AutoML and BART showed slightly higher discrimination and calibration metrics. These results highlight the value of expanding analytical approaches in I/O psychology beyond traditionally applied linear models, especially when dealing with complex attitudinal and behavioral datasets.

The findings also illustrate how modern analytics can complement theoretical knowledge in turnover research. Rather than replacing construct-based understanding, predictive modeling strengthens it by empirically testing how multiple variables jointly contribute to turnover risk. This analytical precision enables I/O practitioners to make more informed, targeted decisions based on individualized risk profiles rather than broad assumptions about the workforce.

Importantly, the study demonstrates that predictive frameworks can serve as a practical extension of existing turnover theories. More advanced modeling allows uncertainty to be quantified, prediction error to be monitored, and probability-based forecasts to be incorporated directly into retention strategies. These capabilities reflect a shift from reactive responses, after turnover occurs, to proactive intervention, supporting data-driven talent management.

As organizations increasingly leverage analytics to guide HR decisions, incorporating rigorous predictive methods within I/O research provides a meaningful path forward. The models evaluated in this study show that turnover intention can be forecasted with reasonable accuracy using survey-based attitudinal data, reinforcing the benefit of integrating statistical innovation with established theory.

In conclusion, this thesis demonstrates that using enhanced predictive analytics—not only improves modeling performance but also broadens the methodological toolkit available to turnover researchers and practitioners. Continued development, validation, and careful implementation of these analytical approaches will further support evidence-based decision-making in workforce retention efforts.

## REFERENCES

- Abner, G., & Hur, H. (2023, January 31). *What keeps public employees in their jobs? It's not just pay*. Government Executive. <https://www.govexec.com/management/2023/01/what-keeps-public-employees-their-jobs-its-not-just-pay/381709/>
- Akashah, M. A., Malik, E. F., Hujran, O., & Zaki, N. (2023). A decade of research on machine learning techniques for predicting employee turnover: A systematic literature review. *Expert Systems With Applications*, 238, 121794. <https://doi.org/10.1016/j.eswa.2023.121794>
- Alarcon, G. M., & Edwards, J. M. (2010). The relationship of engagement, job satisfaction and turnover intentions. *Stress and Health*, 27(3). <https://doi.org/10.1002/smi.1365>
- Balaji, A., & Allen, A. (2018). Benchmarking automatic machine learning frameworks. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.1808.06492>
- Benabou, A., Touhami, F., & Sabri, M. A. (2025). Predicting employee turnover using machine learning techniques. *Acta Informatica Pragensia*, 14(1), 112–127. <https://doi.org/10.18267/j.aip.255>
- Bothma, C. F., & Roodt, G. (2013). The validation of the turnover intention scale. *SA Journal of Human Resource Management*, 11(1). <https://doi.org/10.4102/sajhrm.v11i1.507>
- Brayfield, A. H., & Rothe, H. F. (1951). An index of job satisfaction. *Journal of Applied Psychology*, 35(5), 307–311. <https://doi.org/10.1037/h0055617>
- Breiman, L. (2001) Random Forests. *Machine Learning*, 45, 5-32. <http://dx.doi.org/10.1023/A:1010933404324>
- Brough, P., Timms, C., O'Driscoll, M. P., Kalliath, T., Siu, O., Sit, C., & Lo, D. (2014b). Work–life balance: a longitudinal evaluation of a new measure across Australia and New

- Zealand workers. *The International Journal of Human Resource Management*, 25(19), 2724–2744. <https://doi.org/10.1080/09585192.2014.899262>
- Brown, M. I. (2020). Comparing the validity of net promoter and benchmark scoring to other commonly used employee engagement metrics. *Human Resource Development Quarterly*, 31(4), 355–370. <https://doi.org/10.1002/hrdq.21392>
- Chipman, H. A., George, E. I., & McCulloch, R. E. (2010). BART: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4(1), 266–298. <http://www.jstor.org/stable/27801587>
- Clinton, M., Knight, T., & Guest, D. E. (2012). Job Embeddedness: A new attitudinal measure. *International Journal of Selection and Assessment*, 20(1), 111–117. <https://doi.org/10.1111/j.1468-2389.2012.00584.x>
- Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology*, 86(3), 386–400. <https://doi.org/10.1037/0021-9010.86.3.386>
- Eisenberger, R., Huntington, R., Hutchison, S., & Sowa, D. (1986). Perceived organizational support. *Journal of Applied Psychology*, 71(3), 500–507. <https://doi.org/10.1037/0021-9010.71.3.500>
- El-Rayes, N., Fang, M., Smith, M., & Taylor, S. M. (2020). Predicting employee attrition using tree-based models. *International Journal of Organizational Analysis*, 28(6), 1273–1291. <https://doi.org/10.1108/ijoa-10-2019-1903>
- Fasbender, U., Van Der Heijden, B. I. J. M., & Grimshaw, S. (2018). Job satisfaction, job stress and nurses' turnover intentions: The moderating roles of on-the-job and off-the-job

embeddedness. *Journal of Advanced Nursing*, 75(2), 327–337.

<https://doi.org/10.1111/jan.13842>

Feurer, M., Klein, A., Eggenberger, K., Springenberg, J. T., Blum, M., & Hutter, F. (2019).

Auto-Sklearn: Efficient and robust automated machine learning. In *The Springer series on challenges in machine learning* (pp. 113–134). [https://doi.org/10.1007/978-3-030-05318-5\\_6](https://doi.org/10.1007/978-3-030-05318-5_6)

Flint, D., Haley, L. M., & McNally, J. J. (2012). Dimensionality of organizational justice in a call center context. *Psychological Reports*, 110(2), 677–693.

<https://doi.org/10.2466/01.07.21.pr0.110.2.677-693>

Fukui, S., Wu, W., Greenfield, J., Salyers, M. P., Morse, G., Garabrant, J., Bass, E., Kyere, E., &

Dell, N. (2023). Machine Learning with Human Resources Data: Predicting Turnover among Community Mental Health Center Employees. *The journal of mental health policy and economics*, 26(2), 63–76.

Li, G., Zhou, Y., Zhang, X., & Martek, I. (2025). The integrative relationship between work–

family conflict and turnover intention: a meta-analysis. *Frontiers in Public Health*, 13, 1642843. <https://doi.org/10.3389/fpubh.2025.1642843>

Han, E., & Cai, L. (2024). Impact of Psychological Capital on Organizational Commitment and

Turnover Intention: Evidence from Korean Flight Attendants. *Iranian Journal of Public Health*, 53(7), 1578–1587. <https://doi.org/10.18502/ijph.v53i7.16052>

Hakanen, J. J., Bakker, A. B., & Schaufeli, W. B. (2005). Burnout and work engagement among teachers. *Journal of School Psychology*, 43(6), 495–513.

<https://doi.org/10.1016/j.jsp.2005.11.001>

- He, X., Zhao, K., & Chu, X. (2020). AutoML: A survey of the state-of-the-art. *Knowledge-Based Systems*, 212, 106622. <https://doi.org/10.1016/j.knosys.2020.106622>
- Herwidiyanto, N. M. T., & Sinambela, N. L. P. (2024). The influence of work environment, work stress, workload on turnover intention through organizational commitment as an intervening variable. *International Journal of Economics and Management Sciences*, 1(4), 270–282. <https://doi.org/10.61132/ijems.v1i4.192>
- Herzberg, F.I. (1966). *Work and the nature of man*. World.
- Hill, J., Linero, A., & Murray, J. (2019). Bayesian Additive Regression Trees: A review and look forward. *Annual Review of Statistics and Its Application*, 7(1), 251–278. <https://doi.org/10.1146/annurev-statistics-031219-041110>
- Hoppock, R. (1935). *Job satisfaction*. Harper.
- Hongvichit, S. (2015). The research progress and prospect of employee turnover intention. *International Business Research*, 8(6). <https://doi.org/10.5539/ibr.v8n6p218>
- Hutter, F., Kotthoff, L., & Vanschoren, J. (2019). Automated Machine Learning. In *The Springer series on challenges in machine learning*. <https://doi.org/10.1007/978-3-030-05318-5>
- IBM HR Analytics Employee Attrition & Performance*. (2021, September 8). Kaggle. <https://www.kaggle.com/datasets/uniabhi/ibm-hr-analytics-employee-attrition-performance>
- Iverson, R. D., & Buttigieg, D. M. (1999). Affective, Normative and Continuance Commitment: Can the ‘Right Kind’ of Commitment be Managed? *Journal of Management Studies*, 36(3), 307–333. <https://doi.org/10.1111/1467-6486.00138>

- Jaharuddin, N. S., & Zainol, L. N. (2019). The impact of Work-Life Balance on job engagement and turnover intention. *The South East Asian Journal of Management*, 13(1).  
<https://doi.org/10.21002/seam.v13i1.10912>
- Jha, S. (2014). Determinants of employee turnover intentions: A review. *Management Today*, 9(2), 26–33. <https://ssrn.com/abstract=2442689>
- Jiang, K., Liu, D., McKay, P. F., Lee, T. W., & Mitchell, T. R. (2012). When and how is job embeddedness predictive of turnover? A meta-analytic investigation. *Journal of Applied Psychology*, 97(5), 1077–1096. <https://doi.org/10.1037/a0028610>
- Kalidass, A., & Bahron, A. (2015). The Relationship between Perceived Supervisor Support, Perceived Organizational Support, Organizational Commitment and Employee Turnover Intention. *International Journal of Business Administration*, 6(5).  
<https://doi.org/10.5430/ijba.v6n5p82>
- Karatepe, O. M. (2016). Does job embeddedness mediate the effects of coworker and family support on creative performance? An empirical study in the hotel industry. *Journal of Human Resources in Hospitality & Tourism*, 15(2), 119–132.  
<https://doi.org/10.1080/15332845.2016.1084852>
- Khera, S. N., & Divya, N. (2019). Predictive modelling of employee turnover in Indian IT industry using machine learning techniques. *Vision the Journal of Business Perspective*, 23(1), 12–21. <https://doi.org/10.1177/0972262918821221>
- Kim, H., & Kim, E. G. (2021). A meta-analysis on predictors of turnover intention of hospital nurses in South Korea (2000–2020). *Nursing Open*, 8(5), 2406–2418.  
<https://doi.org/10.1002/nop2.872>

- Kim, S.-W., Price, J. L., Mueller, C. W., & Watson, T. W. (1996). The determinants of career intent among physicians at a U.S. Air Force hospital. *Human Relations*, *49*(7), 947–976. <https://doi.org/10.1177/001872679604900704>
- Klein, H. J., Cooper, J. T., Molloy, J. C., & Swanson, J. A. (2013). The assessment of commitment: Advantages of a unidimensional, target-free approach. *Journal of Applied Psychology*, *99*(2), 222–238. <https://doi.org/10.1037/a0034751>
- Koubova, V., & Buchko, A. A. (2013). Life-work balance. *Management Research Review*, *36*(7), 700–719. <https://doi.org/10.1108/mrr-05-2012-0115>
- LeDell, E., & Poirier, S. (2020). H2O AutoML: Scalable Automatic Machine Learning.
- Lee, H., Chiang, H., & Kuo, H. (2018). Relationship between authentic leadership and nurses' intent to leave: The mediating role of work environment and burnout. *Journal of Nursing Management*, *27*(1), 52–65. <https://doi.org/10.1111/jonm.12648>
- Lee, X., Yang, B., & Li, W. (2017). The influence factors of job satisfaction and its relationship with turnover intention: Taking early-career employees as an example. *Anales De Psicología*, *33*(3), 697. <https://doi.org/10.6018/analesps.33.3.238551>
- Lestari, D., & Margaretha, M. (2020). Work life balance, job engagement and turnover intention: Experience from Y generation employees. *Management Science Letters*, 165–170. <https://doi.org/10.5267/j.msl.2020.8.019>
- Li, Q., Mohamed, R., Mahomed, A., & Khan, H. (2022). The effect of perceived organizational support and employee care on turnover intention and work engagement: A Mediated moderation model using age in the post pandemic period. *Sustainability*, *14*(15), 9125. <https://doi.org/10.3390/su14159125>

- Liu, M., Yang, B., & Song, Y. (2024). Research on predicting the turnover of graduates using an enhanced random forest model. *Behavioral Sciences, 14*(7), 562.  
<https://doi.org/10.3390/bs14070562>
- Löfstrand, P., Wall, E., Selander, J., Nordenmark, M., & Vinberg, S. (2025). Employee engagement in Swedish public sector organizations: The impact of organizational and social factors on employee net promoter scores. *Scandinavian Journal of Public Administration. <https://doi.org/10.58235/sjpa.2024.25213>*
- Manroop, L., Malik, A., & Milner, M. (2024). The ethical implications of big data in human resource management. *Human Resource Management Review, 34*(2), 101012.  
<https://doi.org/10.1016/j.hrmr.2024.101012>
- Mayer, J. D., & Salovey, P. (1997). What is emotional intelligence? In P. Salovey & D. J. Sluyter (Eds.), *Emotional development and emotional intelligence: Educational implications* (pp. 3–34). Basic Books.
- McPherson, J. (2025, August 1). Understanding the employee net promoter score (eNPS) | Culture Amp. *Culture Amp. <https://www.cultureamp.com/blog/employee-net-promoter-score>*
- Meisler, G. (2013). Empirical exploration of the relationship between emotional intelligence, perceived organizational justice and turnover intentions. *Employee Relations, 35*(4), 441–455. <https://doi.org/10.1108/er-05-2012-0041>
- Memon, M. A., Salleh, R., & Baharom, M. N. R. (2016). The link between training satisfaction, work engagement and turnover intention. *European Journal of Training and Development, 40*(6), 407–429. <https://doi.org/10.1108/ejtd-10-2015-0077>

- Mengstie, M. M. (2020). Perceived organizational justice and turnover intention among hospital healthcare workers. *BMC Psychology*, 8(1). <https://doi.org/10.1186/s40359-020-0387-8>
- Meyer, J. P., & Allen, N. J. (1991). A three-component conceptualization of organizational commitment. *Human Resource Management Review*, 1(1), 61–89.  
[https://doi.org/10.1016/1053-4822\(91\)90011-z](https://doi.org/10.1016/1053-4822(91)90011-z)
- Mishra, A. (2025). The Impact of predictive analytics on employee engagement and turnover in knowledge-based Teams: a systematic literature review. *SSRN Electronic Journal*.  
<https://doi.org/10.2139/ssrn.5038568>
- Mitchell, T. R., Holtom, B. C., Lee, T. W., Sablinski, C. J., & Erez, M. (2001). Why people stay: Using job embeddedness to predict voluntary turnover. *Academy of Management Journal*, 44(6), 1102–1121. <https://doi.org/10.2307/3069391>
- Mobley, W. H. (1977). Intermediate linkages in the relationship between job satisfaction and employee turnover. *Journal of Applied Psychology*, 62(2), 237–240.  
<https://doi.org/10.1037/0021-9010.62.2.237>
- Nawaz, M., & Pangil, F. (2016). Career concern as predictor of turnover intention: Empirical evidence from the education industry. *Asian Journal of Applied Sciences*, 4(3), 501–512.
- Putri, A. D., & Handoyo, S. (2018). The Impact of Career Growth on Turnover Intention with Employee Engagement as a Mediator Variable: Study among the Generation Y Employees in Indonesia. *SCITEPRESS – Science and Technology Publications*, 216–223.  
<https://doi.org/10.5220/0008587402160223>
- Olson, R. S., & Moore, J. H. (2019). TPOT: a Tree-Based Pipeline Optimization Tool for automating machine learning. In *The Springer series on challenges in machine learning* (pp. 151–160). [https://doi.org/10.1007/978-3-030-05318-5\\_8](https://doi.org/10.1007/978-3-030-05318-5_8)

- Oosthuizen, R. M., Coetzee, M., & Munro, Z. (2019). Work-life balance, job satisfaction and turnover intention amongst information technology employees. *Southern African Business Review*, 20, 446–467. <https://doi.org/10.25159/1998-8125/6059>
- Opolot, J. S., Simiyu, G., Obedgiu, V., & Alungat, B. (2024). Investigating organizational commitment as a turnover intention mitigation tool. *SEISENSE Business Review*, 4(1), 132–148. <https://doi.org/10.33215/1104qm80>
- Oyedeki, B. A., Ko, Y. H., & Lee, S. (2025). Physical Work Environments: an Integrative Review and Agenda for Future research. *Journal of Management*, 51(6), 2589–2626. <https://doi.org/10.1177/01492063251315703>
- Özkan, A. H., Elci, M., Karabay, M. E., Kitapci, H., & Garip, C. (2020). Antecedents of turnover intention: A meta-analysis study in the United States. *E+M Ekonomie a Management*, 23(1), 93–110. <https://doi.org/10.15240/tul/001/2020-1-007>
- Park, J., & Min, H. (2020). Turnover intention in the hospitality industry: A meta-analysis. *International Journal of Hospitality Management*, 90, 102599. <https://doi.org/10.1016/j.ijhm.2020.102599>
- Perryer, C., Jordan, C., Firms, I., & Travaglione, A. (2010). Predicting turnover intentions. *Management Research Review*, 33(9), 911–923. <https://doi.org/10.1108/01409171011070323>
- Plant, E., King, R., & Kath, J. (2020). Statistical comparison of additive regression tree methods on ecological grassland data. *Ecological Informatics*, 61, 101198. <https://doi.org/10.1016/j.ecoinf.2020.101198>
- Reichheld, Frederick. (2004). The one number you need to grow. *Harvard business review*. 81. 46-54, 124.

- Rhoades, L., & Eisenberger, R. (2002). Perceived organizational support: A review of the literature. *Journal of Applied Psychology*, 87(4), 698–714. <https://doi.org/10.1037/0021-9010.87.4.698>
- Rodríguez-Fernández, M., Herrera, J., De Las Heras-Rosas, C., & Ciruela-Lorenzo, A. M. (2024). Practical Implications of the Organizational Commitment Model in Healthcare: The Case of Nurses. *Journal of Nursing Management*, 2024, 1–12. <https://doi.org/10.1155/2024/6455398>
- Rössberg, J., Eiring, & Friis, S. (2004). Work environment and job satisfaction. *Social Psychiatry and Psychiatric Epidemiology*, 39(7). <https://doi.org/10.1007/s00127-004-0791-z>
- Sabei, S. D. A., Labrague, L. J., Ross, A. M., Karkada, S., Albashayreh, A., Masroori, F. A., & Hashmi, N. A. (2019). Nursing work environment, turnover intention, job burnout, and quality of care: the moderating role of job satisfaction. *Journal of Nursing Scholarship*, 52(1), 95–104. <https://doi.org/10.1111/jnu.12528>
- Saks, A. M. (2006). Antecedents and consequences of employee engagement. *Journal of Managerial Psychology*, 21(7), 600–619. <https://doi.org/10.1108/02683940610690169>
- Salehin, I., Islam, M. S., Saha, P., Noman, S., Tuni, A., Hasan, M. M., & Baten, M. A. (2023). AutoML: A systematic review on automated machine learning with neural architecture search. *Journal of Information and Intelligence*, 2(1), 52–81. <https://doi.org/10.1016/j.jiixd.2023.10.002>
- Schaufeli, W. B., Salanova, M., González-Romá, V., & Bakker, A. B. (2002). The measurement of engagement and burnout: A two sample confirmatory factor analytic approach. *Journal of Happiness Studies: An Interdisciplinary Forum on Subjective Well-Being*, 3(1), 71–92. <https://doi.org/10.1023/A:1015630930326>

- Schaufeli, W. B., & Bakker, A. B. (2004). Job demands, job resources, and their relationship with burnout and engagement: a multi-sample study. *Journal of Organizational Behavior*, 25(3), 293–315. <https://doi.org/10.1002/job.248>
- Schaufeli, W. B., Bakker, A. B., & Salanova, M. (2006b). The measurement of work engagement with a short questionnaire. *Educational and Psychological Measurement*, 66(4), 701–716. <https://doi.org/10.1177/0013164405282471>
- Schaufeli, W. B. (2017). Applying the Job Demands-Resources model. *Organizational Dynamics*, 46(2), 120–132. <https://doi.org/10.1016/j.orgdyn.2017.04.008>
- Scott, I. A., De Guzman, K. R., Falconer, N., Canaris, S., Bonilla, O., McPhail, S. M., Marxen, S., Van Garderen, A., Abdel-Hafez, A., & Barras, M. (2024). Evaluating automated machine learning platforms for use in healthcare. *JAMIA Open*, 7(2), ooae031. <https://doi.org/10.1093/jamiaopen/ooae031>
- Senayah, W. K., & Biney-Aidoo, V. (2024). Antecedents of turnover intentions and organizational commitment of workers in the apparel industry: The moderating effect of creative behavior and instrumental readiness. *Social Sciences & Humanities Open*, 10, 101119. <https://doi.org/10.1016/j.ssaho.2024.101119>
- Shah, I. A., Csordas, T., Akram, U., Yadav, A., & Rasool, H. (2020). Multifaceted role of job embeddedness within organizations: Development of Sustainable approach to reducing Turnover intention. *SAGE Open*, 10(2). <https://doi.org/10.1177/2158244020934876>
- Shobhanam, K., & Sumati, S. (2022). HR Analytics: Employee Attrition Analysis using Random Forest. *International Journal of Performability Engineering*, 18(4), 275. <https://doi.org/10.23940/ijpe.22.04.p5.275281>

- Sinval, J., & Marôco, J. (2020b). Short Index of Job Satisfaction: Validity evidence from Portugal and Brazil. *PLoS ONE*, *15*(4), e0231474. <https://doi.org/10.1371/journal.pone.0231474>
- Skelton, A. R., Nattress, D., & Dwyer, R. J. (2019). Predicting manufacturing employee turnover intentions. *Journal of Economics Finance and Administrative Science*, *25*(49), 101–117. <https://doi.org/10.1108/jefas-07-2018-0069>
- Soelton, M., & Atnani, M. (2018). How work environment, work satisfaction, work stress on the turnover intention affect university management. *Jurnal Manajemen Dan Bisnis Indonesia*, *5*(3), 439–448. <https://doi.org/10.31843/jmbi.v5i3.178>
- Sparapani, R., Spanbauer, C., & McCulloch, R. (2021). Nonparametric Machine Learning and Efficient Computation with Bayesian Additive Regression Trees: The BART R Package. *Journal of Statistical Software*, *97*(1). <https://doi.org/10.18637/jss.v097.i01>
- Stachová, K., Stacho, Z., Cagáňová, D., & Baroková, A. (2021). Progress in the use of tools for the selection of employees at the time of digitizing. *ICST Transactions on Scalable Information Systems*, *8*(33), 170757. <https://doi.org/10.4108/eai.25-8-2021.170757>
- Stark, H. P., Smith, R. W., & Carter, N. T. (2025). Organizational Commitment Profiles and Employee Well-Being: Exploratory and confirmatory latent profile analyses. *Occupational Health Science*. <https://doi.org/10.1007/s41542-025-00225-2>
- Stinglhamber, F., Marique, G., Caesens, G., Hanin, D., & De Zanet, F. (2015). The influence of transformational leadership on followers' affective commitment: The role of perceived organizational support and supervisor's organizational embodiment. *The Career Development International*, *20*(6), 583–603. <https://doi.org/10.1108/CDI-12-2014-0158>

- Suifan, Taghrid & Abdallah, Ayman & Diab, Hannah. (2016). The Influence of Work Life Balance on Turnover Intention in Private Hospitals: The Mediating Role of Work Life Conflict. *European Journal of Business and Management*. 8. 126-139.
- Takawira, N., Coetzee, M., & Schreuder, D. (2014). Job embeddedness, work engagement and turnover intention of staff in a higher education institution: An exploratory study. *SA Journal of Human Resource Management*, 14(1).  
<https://doi.org/10.4102/sajhrm.v12i1.524>
- Talebi, H., Bardsiri, A. K., & Bardsiri, V. K. (2025). Machine Learning Approaches for Predicting Employee Turnover: A Systematic review. *Engineering Reports*, 7(8).  
<https://doi.org/10.1002/eng2.70298>
- Taner, Z., Hızıroğlu, O. A., & Hızıroğlu, K. (2024). Leveraging machine learning methods for predicting employee turnover within the framework of human resources analytics. *Journal of Intelligent Systems Theory and Applications*, 7(2), 145–158.  
<https://doi.org/10.38016/jista.1440879>
- Tett, R. P., & Meyer, J. P. (1993). Job satisfaction, organizational commitment, turnover intention, and turnover: path analyses based on meta-analytic findings. *Personnel Psychology*, 46(2), 259–293. <https://doi.org/10.1111/j.1744-6570.1993.tb00874.x>
- Thirapatsakun, N. T., Kuntonbutr, N. C., & Mechida, N. P. (2015). The relationships among four factors and turnover intentions at different levels of perceived organizational support. *Journal of US-China Public Administration*, 12(2). <https://doi.org/10.17265/1548-6591/2015.02.002>
- Vaamonde, J., Omar, A., & Salessi, S. (2018). *From organizational justice perceptions to turnover intentions: the mediating effects of burnout and job satisfaction*.

<https://www.semanticscholar.org/paper/From-Organizational-Justice-Perceptions-to-Turnover-Vaamonde-Omar/f246ed71655d85fcf0b985f4a0518205ccb006fd>

- Wang, Q., Gan, K., Wei, H., Sun, A., Wang, Y., & Zhou, X. (2022). Public service motivation and public employees' turnover intention: the role of job satisfaction and career growth opportunity. *Personnel Review*, 53(1), 99–118. <https://doi.org/10.1108/pr-11-2020-0836>
- Wasti, S. A. (2003). The influence of cultural values on antecedents of Organisational Commitment: an Individual-Level analysis. *Applied Psychology*, 52(4), 533–554. <https://doi.org/10.1111/1464-0597.00150>
- Weng, Q. and Hu, B. (2009) The Structure of Employee Career Growth and Its Effect on Turnover Intention. *Industrial Engineering and Management*, 14, 14-21.
- Weng, Q. and Xi, Y. (2011) The Study of Career Growth of Professional Employees: Scale Development and Validity. *Management Review*, 10, 132-143.
- Weng, Q., & McElroy, J. C. (2012). Organizational career growth, affective occupational commitment and turnover intentions. *Journal of Vocational Behavior*, 80(2), 256–265. <https://doi.org/10.1016/j.jvb.2012.01.014>
- Weng, Qingxiong Derek & McElroy, James. (2016). Organizational Career Growth Scales.
- Xuecheng, W., Iqbal, Q., & Saina, B. (2022). Factors affecting employee's retention: Integration of situational leadership with social exchange theory. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.872105>
- Yang, H., Cheung, C., & Song, H. (2016). Enhancing the learning and employability of hospitality graduates in China. *Journal of Hospitality Leisure Sport & Tourism Education*, 19, 85–96. <https://doi.org/10.1016/j.jhlste.2016.08.004>

Yildiz, B., Yildiz, H., & Arda, O. A. (2021). Relationship between work–family conflict and turnover intention in nurses: A meta-analytic review. *Journal of Advanced Nursing*, 77(8), 3317–3330. <https://doi.org/10.1111/jan.14846>

Zhuge, R., Wang, Y., Gao, Y., Wang, Q., Wang, Y., Meng, N., Cui, Y., Huang, Q., Wang, K., & Wu, Q. (2024). Factors influencing the turnover intention for disease control and prevention workers in Northeast China: an empirical analysis based on logistic-ISM model. *BMC Health Services Research*, 24(1), 1264. <https://doi.org/10.1186/s12913-024-11738-x>

## **APPENDIX A: R Code for Data Processing and Analysis**

The full R code used for data cleaning, model development, and statistical analysis is hosted on GitHub and can be accessed at the following link:

<https://github.com/shivantibharath/turnover-prediction-ml-thesis->

The repository will be maintained and updated as additional analyses and revisions are completed.

## APPENDIX B: IRB Approval Memo



**EAST CAROLINA UNIVERSITY**  
**University & Medical Center Institutional Review Board**  
Willis Building · Mail Stop 682  
600 Moye Boulevard · Greenville, NC 27834  
Office 252-744-2914 · Fax 252-744-2284  
[rede.ecu.edu/umcirb/](http://rede.ecu.edu/umcirb/)

### Notification of Exempt Certification

From: Social/Behavioral IRB  
To: [Mark Bowler](#)  
CC:  
Date: 10/22/2025  
Re: [UMCIRB 25-001637](#)  
Understanding Employee Turnover

I am pleased to inform you that your research submission has been certified as exempt on 10/22/2025. This study is eligible for Exempt Certification under category # 2b.

It is your responsibility to ensure that this research is conducted in the manner reported in your application and/or protocol, as well as being consistent with the ethical principles of the Belmont Report and your profession.

This research study does not require any additional interaction with the UMCIRB unless there are proposed changes to this study. Any change, prior to implementing that change, must be submitted to the UMCIRB for review and approval. The UMCIRB will determine if the change impacts the eligibility of the research for exempt status. If more substantive review is required, you will be notified within five business days.

Document	Description
Turnover Consent Form v2.docx(0.02)	Consent Forms
Turnover Debriefing.docx(0.01)	Additional Items
Turnover Listing.docx(0.02)	Recruitment Documents/Scripts
Turnover Measures and Items.docx(0.01)	Surveys and Questionnaires
Turnover Validity Check.docx(0.01)	Additional Items

For research studies where a waiver or alteration of HIPAA Authorization has been approved, the IRB states that each of the waiver criteria in 45 CFR 164.512(i)(1)(i)(A) and (2)(i) through (v) have been met. Additionally, the elements of PHI to be collected as described in items 1 and 2 of the Application for Waiver of Authorization have been determined to be the minimal necessary for the specified research.

The Chairperson (or designee) does not have a potential for conflict of interest on this study.