MSc International Consulting and Management
Research Topics

Predicting Job Flourishing through Algorithmic Management: a study of Machine Learning techniques

Atis Kazaferi

Supervisor:
DR. JEROEN G. MEIJERINK MSC
DR. LARA CARMINATI

July, 2024

UNIVERSITY OF TWENTE.

Faculty of Behavioural and Management Sciences,
Business Administration,
University of Twente

Contents

1 Introduction 1

2 Theoretical Framework 4
- 2.1 Employee flourishing and Organizations . . . 4
- 2.2 Antecedents of Flourishing and Relation to Work Engagement . . . 5
  - 2.2.1 Flourishing and Work Engagement . . . 6
  - 2.2.2 Resources and Demand: JD-R theory . . . 7
  - 2.2.3 Variables . . . 8
- 2.3 Algorithmic Management . . . 11
  - 2.3.1 Prediction Models . . . 12
  - 2.3.2 Hyperparameters and Tuning Techniques . . . 15
  - 2.3.3 Measuring performance . . . 18

3 Research Methodology 21
- 3.1 Research Design . . . 21
- 3.2 Data Collection . . . 21
- 3.3 Measures and Operationalization . . . 22
- 3.4 Model Building Process . . . 24
  - 3.4.1 Data Preparation . . . 24
  - 3.4.2 Data Analysis . . . 25
  - 3.4.3 Training, Testing, and Validation . . . 25
  - 3.4.4 Models and Optimization . . . 25
  - 3.4.5 Evaluating and Reporting . . . 26

4 Results 27
- 4.1 Data Preparation . . . 27
  - 4.1.1 Missing Values . . . 27
  - 4.1.2 Job Flourishing Labeling . . . 28
- 4.2 Data Analysis . . . 29
  - 4.2.1 Correlation Analysis . . . 29
  - 4.2.2 Factor Analysis . . . 29
  - 4.2.3 Iterative Feature Engineering . . . 30
- 4.3 The Built Models . . . 33
  - 4.3.1 Training, Testing, and Validation . . . 33
  - 4.3.2 Model Description and Optimization . . . 33
  - 4.3.3 First Iteration Evaluation Report . . . 33
  - 4.3.4 Second Iteration Evaluation Report . . . 34
  - 4.3.5 Third Iteration Evaluation Report . . . 35
- 4.4 Final Evaluation . . . 37

5 Discussion 40
- 5.1 Theoretical Implications . . . 40
  - 5.1.1 Regarding Features and Predictors . . . 40

5.1.2 Regarding the Usage Map 41
5.2 Practical Implications 42

6 Limitations and Future Research **44**
6.1 Utilizing Work Engagement 44
6.2 Sample Size and Data Preparation 44
6.3 Techniques and Algorithms 44

7 Conclusions **46**

8 Appendix **47**
8.1 Appendix A: Factor Analysis 47
8.2 Appendix B: K-means clustering information 50
8.3 Appendix C: Configurations of GA and BOA per iteration 51
8.4 Appendix D: Iteration 3 Feature Distribution 53
8.5 Appendix E: Best Logistic Regression Coefficients 54

Bibliography **57**
Articles 63

Abstract

Nurturing a positive state of mind and engaging employees is vital for organizations to ensure, among others, employee retention and organizational performance. Drawing upon insights from Algorithmic Management (AM), this thesis aimed to define the best methods and criteria to predict job flourishing in employees of an educational institution in the Netherlands, utilizing secondary data. Job flourishing was defined as a broader construct of work engagement, and we based a binary classification on highly engaged employees. A few challenges were found in terms of defining the best criteria in the case of job flourishing, depending on organizational HR goals. Additionally, it was challenging to optimize the defined models due to the hidden underlying structure of the dataset. Results indicated the importance in flourishing prediction of item level constructs of structural job resources, job characteristics, and supervisor relations, supporting theoretical evidence. Moreover, of interest was the contribution of tenure and contract type, possibly due to organizational context, which led to a negative correlation of these socio-demographics to job flourishing prediction. Interestingly, the feature engineered logistic Regression model with Genetic Algorithm application was the best at predicting flourishing employees, while multiple models performed well at prediction of non-flourishing employees. We summarize our findings in a usage map for practitioners and researchers that want to bridge the gap between well-being and artificial intelligence tools, by developing AM tools contributing to the HR and well-being domain.

Keywords: Algorithmic Management; Job Flourishing; Work Engagement; Machine Learning

Chapter 1

Introduction

Nurturing a positive state of mind and engaging employees is vital for organizations to ensure, among others, employee retention (Ford, Newman, and Ford, 2023). Human Resource Management (HRM) has long been interested in employee experiences and management related to well-being, stress, and burnout, with a recent surge in popularity (e.g., Yang et al., 2023; Sutton and Atkinson, 2023; Cavazotte, Mello, and Oliveira, 2021; Ren et al., 2015). Thus, across the years, flourishing as a golden rule for well-being has received research attention as a construct that can aid employees in several domains, including engagement, proactive behavior, and work performance (A'yuninnisa, Carminati, and Wilderom, 2023). According to Keyes and Annas (2009), flourishing is a state of high emotional well-being and functioning, whereby employees experience high levels of emotional, psychological, and social well-being. Job flourishing is an extension of Keyes and Annas(2009)'s definition applied in a work context with job-related antecedents.

Currently, the job flourishing state-of-the-art focuses on measuring job flourishing and finding and understanding its antecedents. Various scales have been developed, including specific work scales and more general flourishing measures (eg. the flourishing scale from Diener et al. (2010); the mental health continuum from Keyes (2005); the PERMA profiler by Butler and Kern (2016)). Through different foundational theories (eg. Job Demands and Resources Model (JD-R), Broaden and Build Theory (BBT), Conservation of Resources Theory (COR), etc.), literature connects job flourishing to job satisfaction, employee mental health, and even turnover (Rothmann, 2014; Keyes, 2005). For example, Rothmann (2014) links flourishing to advancement and remuneration. Advancement includes training, career opportunities, and promotion, while remuneration focuses on expected and matching compensation to the work needed. Both affect flourishing through work engagement and performance and lead to a need for fair appraisal and evaluation of employees (Latham et al., 2005). This stream of research aims to find predictor features so practitioners and organizations can enhance, for instance, HR practices that lead to employee flourishing. Hence, the focus is on analyzing and tracing features that can help an employee flourish. This way, researchers often hope to break down a multidimensional construct into its constituent dimensions to identify patterns and relationships that can be changed at a practical (organizational) level (A'yuninnisa, Carminati, and Wilderom, 2023).

To better understand the relationship between job flourishing and its antecedents, both researchers and practitioners can rely on insights from algorithmic management (AM). AM practices have recently been interesting in HRM research (Meijerink and Bondarouk, 2023; Tambe, Cappelli, and Yakubovich, 2019). AM is often defined as *a system that automatically processes machine-readable data input to support or execute decision-making* (Duggan et al., 2019; Zalmanson, 2017). These systems involve the usage of diverse technological tools and reliance on data collection techniques and artificial intelligence (AI), which allow for the automation of decisions in HR (Parent-rocheleau et al., 2024; Piasna, 2024). AM can aid decision-making in HR practices, including practices related to job flourishing and its outcomes, for example in relation to employee engagement, training and development, or employee turnover (Garg, Sinha, and Kar, 2019). These concepts sit at the border of job flourishing and HRM practices (Ford, Newman, and Ford, 2023; A'yuninnisa, Carminati, and Wilderom, 2023). Garg, Sinha, and Kar (2019) highlights that Machine Learning (ML) in HRM is evolving fast, and integration is inevitable in the foreseeable future. 22% of firms have already adopted analytics in HR across different functionalities (mostly recruitment), yet there is still room for development when it comes to algorithms related to well-being (Tambe, Cappelli, and Yakubovich, 2019).

Importantly, Deobald et al. (2019) highlights three types of algorithms used in AM: **descriptive, predictive, and prescriptive**. In this context, **descriptive** analysis describes relationships between variables based on historical data, mainly to visualize patterns, relationships, and trends. **Predictive** analysis utilizes advanced machine learning (ML), data mining (DM), historical data, and statistical algorithms to forecast and evaluate the likelihood of an event. Lastly, **prescriptive** provides scenario-based decisions focusing on recommendations and taking actions. This thesis is an instrumental pathway to move from descriptive to predictive AM in job flourishing.

As observed in A'yuninnisa, Carminati, and Wilderom (2023) systematic review, most flourishing research focuses on descriptive analysis through data processing and means statistics. This research aims to describe and find relationships between antecedents and predictors, often using regression models. Although regression is part of inferential statistics, it is often used in job flourishing to describe relationships by computing correlations based on historical or contextual data. The next step in prediction—algorithms that forecast the likelihood of an employee flourishing has yet to be taken in the field. This is the case for most research using ML in HRM areas. Garg, Sinha, and Kar (2019)'s review reveals that ML applications are the strongest in areas such as recruitment and performance management due to their measurability and data availability. Meanwhile, ML is still in the introductory phase of concepts that require an understanding of psychological phenomena. To fully understand the complex nature of ML needed for intricate cognitive understanding, this thesis draws upon the benefits and limitations of different algorithms and their implementation. Garg, Sinha, and Kar (2019) calls for sophisticated approaches to implemented algorithms for decisions in HRM, with various heuristics and metaheuristics approaches to predicting. Heuristics (according to Kumar and Yadav, 2022: techniques that try to solve a problem fast) such as neural networks, decision trees, and k-means clustering, are often used in HRM for classification and prediction.

Some recent literature in health and computational sciences utilizes advanced prediction ML techniques for well-being. For example, Pap et al. (2022) utilizes neural networks and a type of social evolutionary algorithm (imperialist competitive algorithm) to evaluate performance and well-being through correlation analysis. Also present in well-being and HR research are gradient-boosting trees (Nurajijah, Wijaya, and Sari, 2022; Pourkhodabakhsh, 2023; Meraliyev, Alibekova, and Bekturganova, n.d.) and random forests with regression analysis (Vera Cruz et al., 2023; Hoekstra et al., 2023). Pourkhodabakhsh (2023) conducted various analyses focusing on employee turnover with a primary focus on feature elimination, although there was a limitation in using real-life data for the study . Garg, Sinha, and Kar (2019) attributes the lack of use of neural networks, support vector machines, and genetic algorithms in HRM to the limited understanding of how the algorithms operate. Indeed, these types of complex heuristics are rarely found in the recent HR and organization literature, despite their ability to predict.

Due to this limited investigation of complex heuristics in HR, the aim of this research is to explore what ML models and techniques can be used to predict job flourishing, given the current data science literature. Prediction can be achieved through ML methodologies within supervised or unsupervised learning. AM faces challenges in data authenticity, data availability, data management, and the explainability and fairness of algorithms (Garg, Sinha, and Kar, 2019; Tambe, Cappelli, and Yakubovich, 2019). These challenges have yet to be connected and explored in job flourishing prediction. Hence, our exploration is limited to supervised learning techniques that can be used for job flourishing prediction. Supervised learning offers explainability and works well with limited data (from availability or lack of management). As job flourishing relates to potentially sensitive employee information, ethical algorithms need to be developed and explored. Currently, there is no benchmark for what "the right" algorithm implies, so this thesis attempts to set the stage for these questions.

So, we ponder what defines the best models to predict and how they can be used and recommended.

Mean squared error (MSE) is a criterion often used for comparisons (see Vera Cruz et al., 2023; Pap et al., 2022). However, the use of MSE implies that accuracy (correct predictions of the model) is the most important performance indicator for an algorithm. Similarly, the F1 score is sometimes used as a comparison criterion, representing a balance of recall and precision (see Meraliyev, Alibekova, and Bekturganova, n.d.); other researchers focus solely on recall (as the cost of false negative classified instances) or a presentation of the performance across various measures (Pourkhodabakhsh, 2023; Nurajijah, Wijaya, and Sari, 2022). Recall is often used to ensure that each class can be properly identified in ML. This begs the question, what is the most important aspect to evaluate and improve algorithms in? Hence, whether researchers and practitioners should, for instance, focus on improving the identification of flourishing employees at the cost of precision should be considered to advance the field. What does each criterion mean for HRM practitioners, and how should research proceed in developing AM for job flourishing?

Considering this need to explore various methods and techniques and understand the different measures and criteria in AM concerning job flourishing, we pose the overarching research question: **What are the best methods for predicting job flourishing, considering different criteria?** To answer this, the thesis draws upon job resources and demand and, subsequently the strong connection between flourishing and work engagement (see Ariza-Montes et al., 2018; Demerouti, Bakker, and Gevers, 2015; Imran et al., 2020). This is done, as research shows that employees flourish with increased work engagement and when actively seeking resources. The systematic literature review and the meta-model from A'yuninnisa, Carminati, and Wilderom (2023) further highlight these relationships.

The ability to predict provides great insights for extending the knowledge and capabilities of AM in HRM. This contribution can be seen as a direct response to Tambe, Cappelli, and Yakubovich (2019)'s suggestions on the next steps for ML algorithms and data science (calling for more ML algorithms on topics such as employee advice for training programs, new jobs, well-being, and retirement planning advice). Additionally, this research contributes to Garg, Sinha, and Kar (2019)'s call for a nuanced understanding of how ML could support the effectiveness of HR functions through ML and HR practitioners' cooperation. Therefore, our aim is to merge data science knowledge with an in-depth theoretical background to predict job flourishing through its antecedents. This is done with two primary objectives:

1. To create the stepping stones for studying AM in the context of flourishing and well-being as an unexplored area of HRM, creating the possibility for a future research agenda.
2. To map the methods and algorithms across various criteria and measurements to create an accessible overview for practitioners and developers of intricate AM algorithms.

Our analysis and investigation revolve around data collected in two waves in 2022 in the context of higher education employees at the University of Twente. The primary data measured well-being as "employees' perceptions of work engagement and work pressure (strain)" (Leede and Meijerink, 2019, p. 7). Although a different definition of well-being than a job flourishing scale, this research utilizes work engagement as a proxy for job flourishing. It argues that employees with high levels of work engagement have a high likelihood of flourishing at work. Therefore, the study attempts to define the boundary of work flourishing for this organization.

The upcoming chapter will detail the theoretical background behind flourishing, its antecedents, and measures, including the work engagement approach undertaken in this research. Additionally, the chapter dives into the context of AM and its relevance in HR. Following is a description of the ML models used and their applications in AM.

Chapter 2

Theoretical Framework

Firstly, this section defines job flourishing as the golden rule of well-being and describes its predictors and antecedents in the workplace. The current body of knowledge is grounded in popular management theories (e.g. Conservation of Resources; Job Demand and Resources) and consists of various antecedents and indicators. Hence, the research aims to understand the current state of the art and its needs. By formulating propositions about what it means to study flourishing at work, this section is a segue toward the chosen proxy of job flourishing (work engagement). Secondly, the progress and expectations of the AM field are investigated to provide insights into job flourishing prediction. Additionally, this section outlines potential algorithms that are often used to predict complex issues, particularly in the fields of computer science and business information management. Through this, the thesis establishes a baseline and benefits for the next steps relevant to predicting job flourishing in the context of organizational HR demands.

2.1 Employee flourishing and Organizations

Flourishing is an extended definition of mental health often described as a state of successful performance of cognitive functioning (Keyes, 2005). To understand flourishing and to be able to predict it, it is essential to distinguish between mentally healthy adults and those who are flourishing or languishing in the workplace (the absence of mental health) (Keyes, 2005). This distinction allows for a differentiation between well-being baselines and the ability given by different antecedents to highly function. Hence, not all employees who are mentally healthy can flourish. Flourishing is a dynamic optimal state of psychological functioning arising from multiple psychological domains and dimensions, which have been investigated in the past years through different models. (Butler and Kern, 2016).

This multi-dimensional concept comes from the combination of emotional vitality and positive feelings towards life (hedonia) and symptoms of positive functioning (eudemonia) (Keyes, 2005). **Hedonic well-being** links to subjective or emotional well-being, consisting of life satisfaction, happiness, and evaluation of one's situation (Schotanus-dijkstra and Pieterse, 2016). **Eudemonic well-being**, on the other hand, focuses on self-development and autonomy as higher-order needs in Maslow's hierarchy. From a capability approach, it links to the ability to choose as well as the actual opportunities taken (Vanhoutte, 2015). Hence, rather than studying hedonic and eudemonic well-being alone, it is essential to investigate in a multifaceted manner both what brings happiness to one's life and the ability to further develop (Schotanus-dijkstra and Pieterse, 2016).

Rothmann (2014) was one of the first to investigate the concept of flourishing at work in a multifaceted manner, with results showing that job characteristics and co-worker relations make a significant contribution to one's flourishing. In fact, flourishing at work has a 54% variance to flourishing in life (Rautenbach, 2015). This implies a strong correlation between flourishing in life and job flourishing, although they are different concepts that have different attributes and factors that contribute to their measurement. Making use of these correlations, it can be stated that job flourishing relates to the study of job characteristics and the affiliated job environment. According to Rautenbach(2015) this job environment is also linked to the social context. Moreover, Redelinghuys, Rothmann, and Botha (2019) highlights the need for organizations to select and retain productive employees, which can be achieved through the study of job flourishing. Hence, this relationship between flourishing, as a golden rule for well-being, and organizational performance (as measured by retention and selection of highly performing employees) is highly significant in researching flourishing at work.

Regarding the relation of flourishing studies to job characteristics and work environment, the studied corpus suggests some contradictory findings, mostly due to dimensionality analysis (Peccei and Van De Voorde, 2019). It is well established that flourishing is a multifaceted construct (A'yuninnisa, Carminati, and Wilderom, 2023; Rothmann, 2014; Keyes and Annas, 2009). Being such a complex construct, different dimensions and operationalizations are adopted across the scientific corpus. A'yuninnisa, Carminati, and Wilderom (2023) highlights the utilization of generic flourishing scales in addition to those scales specialized for workplace flourishing. Despite this, the current state of the art has developed a series of well-grounded theories and antecedents that play a role in job flourishing and how job flourishing affects performance or organizations. The corpus establishes a clear link between flourishing, work engagement, job crafting, extra-role behaviour, job satisfaction, work strain, etc utilizing various scales (Rothmann, 2014, Ariza-Montes et al., 2018, Demerouti, Bakker, and Gevers, 2015; Bakker and Sanz-Vergel, 2013).

Based on Guerci, Hauff, and Gilardi (2022) and the previously made link between job flourishing and performance, well-being is considered a key outcome within HR, as it is attributed to positive impacts such as employee retention (organizational level) or individual performance (individual level). Multiple studies relate to this impact within HR and organizations, both on individual and organizational levels. On the other side, according to Peccei and Van De Voorde (2019) HRM practices also have a direct influence on well-being, with contradictory research results. Linking holistic well-being to flourishing, it is important to look into HRM contexts when treating well-being, leading to HRM decisions. Hence, job flourishing is linked to HR practices and organizational context for organizational decision-making.

From this integral relationship of job flourishing with the job environment, organizational performance, and HR decision-making, it can be concluded that to further study job flourishing, researchers need a contextual business understanding of the organization's HR practices. In other words, they should consider the organizational stimuli when studying job flourishing. Hence, it is important to contemplate the role of job flourishing with employee satisfaction with HR practices such as training, appraisal, and feedback, as well as job-creating characteristics.

2.2 Antecedents of Flourishing and Relation to Work Engagement

An overview of the concept's operationalization and a good understanding of the job-flourishing body of knowledge are needed to discuss and explore its prediction through AM. Flourishing at work consists of emotional well-being (job satisfaction, positive affect, and low negative affect), psychological well-being (autonomy, competence, relatedness, meaning, engagement, and learning), and social well-being (social contribution, integration, actualization, acceptance, and coherence).

Since flourishing was first introduced, multiple studies have evolved the research stream by investigating empirically or through meta-analysis the antecedents and predictors of flourishing based on different measurement scales. Rothmann (2014)'s results show that work role fit, job characteristics, co-worker relationships, and remuneration affected flourishing positively, while overload, supervisor relations, and advancement did not necessarily predict it. The most significant contributor, according to his research, was job characteristics, implying a higher degree of job flourishing given challenging tasks that allow for variety and autonomy. Additionally, when it comes to flourishing in life and at work, personality traits describe most of the variance, characterizing flourishers by extraversion, less neuroticism, and conscientiousness, while socio-demographics, such as gender, age, and education, least explain or predict well-being (Schotanus-dijkstra and Pieterse, 2016). Despite this, longer tenure in roles or organizations can be seen as an antecedent of flourishing due to a feeling of competence (related

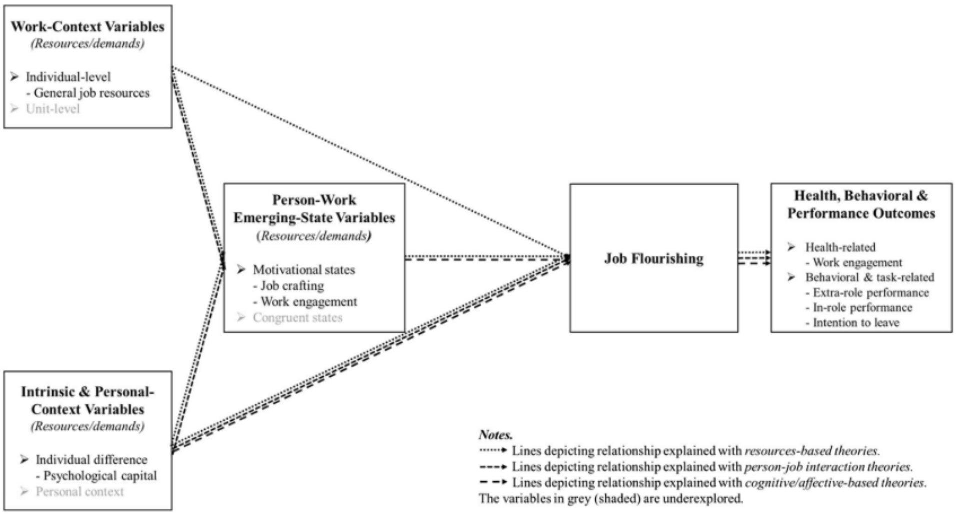


FIGURE 2.1: Visualization of the dynamic relationship between the most frequently found variables associated with job flourishing and the utilized theoretical framework (A'yuninnisa, Carminati, and Wilderom, 2023)

to self-affect) and belonging (Rautenbach, 2015).

Drawing upon A'yuninnisa, Carminati, and Wilderom (2023)'s systematic literature review, there is a link defined across the flourishing literature between different types of variables through resource-based theories (such as JD-R). As seen in figure 2.1, variables like general job resources, psychological capital (efficacy, optimism, etc.), job crafting (employee's proactive action for better alignment), and work engagement are well-established in flourishing. Among these, the most studied variable is work engagement, which is seen as an outcome and a predictor of job flourishing. Hence, in the upcoming chapters, this lineage is investigated.

2.2.1 Flourishing and Work Engagement

Work engagement is shown to link with different job characteristics through motivation reasons (intrinsic or extrinsic) (Bakker and Sanz-Vergel, 2013). Similarly, Demerouti, Bakker, and Gevers (2015) elaborates on the impact of job crafting on flourishing, arriving at similar statistical results about flourishing and work engagement. Meanwhile, in areas of calling (such as religious relations), work engagement seems to have a very strong correlation to flourishing (Ariza-Montes et al., 2018). In fact, according to the literature body, work engagement can be seen as a direct predictor as well as an outcome of flourishing at work, depending on the research conducted, with more than 7 studies exploring the relationship (see A'yuninnisa, Carminati, and Wilderom, 2023). Work engagement is treated as the assumed opposite of burnout, within the continuum of work-related mental health, although this comes with limitations (Schaufeli and Bakker, 2004). Due to the complex and intricate relationship regarding mental health, flourishing, and work engagement, it is valid to state that flourishing at work is highly correlated to work engagement. Moreover, highly engaged employees have a higher possibility to flourish, based on the relationship work engagement has with other job flourishing antecedents, such as job crafting, self-efficacy, and autonomy (Bakker and Oerlemans, 2019).

Based on these high correlations, in the context of this thesis, high work engagement is characterized as a narrow conceptualization of job flourishing. This choice is made as there are three major complications in taking job flourishing as a broad concept into AM:

1. In line with A'yuninnisa, Carminati, and Wilderom (2023)'s research, job flourishing consists of multiple variables (see figure 2.3). Considering all these variables can lead to extensive and unexplainable variation across different data points. As this is a starting point for AM in job flourishing, the research requires model simplification, aiming for reliability.

2. The systematic review also reveals challenges in alignment with the operationalization of the broad concept of job flourishing across literature. Researchers in job flourishing only recently have adopted an aligned measuring instrument. Correctness in the used measurement scale is important for ML algorithms to create real-life validity. Hence, narrowing down to a surrogate concept, which is well-researched and operationalized similarly across different research, provides a validated playground for the initiation of AM in job flourishing, given the scope of this thesis.

3. Lastly, work engagement is seen across the body of knowledge as a highly direct predictor and as an outcome of job flourishing. This complex and reciprocal correlation makes it the most well-suited concept for predicting job flourishing. Additionally, it is the one that has the most academic support (figure 2.3).

Schaufeli and Bakker (2004) developed the Utrecht work engagement scale (UWES) which is widely used to validate the relationship between work engagement and flourishing. This scale measures vigor (high levels of energy), dedication (a sense of significance coming from work), and absorption (totally and happily immersed in work). These elements can be directly linked with some of Rautenbach (2015)'s flourishing at work scale, such as engagement, meaningfulness, positive affect, etc.

This narrowed surrogate conceptualization of job flourishing attempts to create a simplified model of a multidimensional construct. Meanwhile, job crafting and job resources are some of the most investigated variables as predictors of job flourishing, as well as work engagement. Due to this and the often-utilized resource-based theories, our exploration is extended to job characteristics, resources, and demands (covered in the next chapter).

2.2.2 Resources and Demand: JD-R theory

In literature, the most commonly used measures of flourishing at life and work are the three generic flourishing in life measures: the flourishing scale (Diener et al., 2010), the mental health continuum (Keyes, 2005), and the PERMA profiler (Butler and Kern, 2016). The flourishing at work scale from Rautenbach (2015) is also utilized by academics. Particularly difficult is the fact that life flourishing and work flourishing are often evaluated in terms of cohesion, although there is a significant difference between the two.

Flourishing research, as well as work engagement research, has often focused on identifying job demands and their relationship with well-being. More importantly, research often attempts to relate flourishing measures or scales with conceptual theories. Three theories are commonly used to analyze flourishing predictors: Broaden and Build theory (BBT), Conservation of Resources theory (COR), and Job Demands and Resources (JD-R) (A'yuninnisa, Carminati, and Wilderom, 2023). Among these we focus on the vastly used and most supported with work engagement: the JD-R model (Bakker and Sanz-Vergel, 2013).

Job Demands and Resources

In this model, employee well-being consists of job demands and job resources (Bakker and Demerouti, 2007). Job demands are physical, psychological, social, or organizational aspects of the job that require certain physiological and/or psychological costs. Job resources are physical, psychological, social, or organizational aspects of the job that are functional to achieve goals, reduce negative demands, or stimulate growth and learning.

When it comes to the JD-R, workload, and job insecurities are the demands that decrease flourishing. Job resources such as advancement and authentic leadership explain most of the variance of job flourishing, while compensation is important but not a predictor (Rautenbach, 2015; Rothmann, 2014). Additionally, positive personality characteristics play a vital role in emotional adeptness, leading to optimal functioning (flourishing) (Stanley and Schutte, 2023).

Due to the strong usage in recent literature of job demands and job resources towards flourishing, the JD-R scale was used for data collection, specifically the variables in figure 2.2. This data, which aids in measuring job flourishing, was based on the validated scales of work engagement (Schaufeli

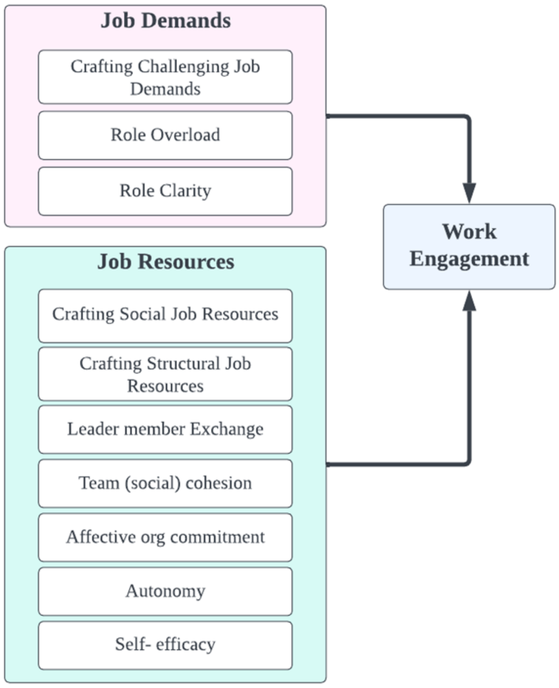


FIGURE 2.2: Conceptual Model on Employee Well-being Survey based on Leede and Meijerink (2019)

and Bakker, 2004), the definitions of job crafting within resource and demand (Tims, Bakker, and Derks, 2013; Tims, Bakker, and Derks, 2012), and the psycho-social workload scale (Van Veldhoven and Meijman, 1994). Additionally, autonomy and self-efficacy are evaluated as adjacent constructs to psychological capital and proactivity.

2.2.3 Variables

Here we provide the key definitions used in the primary data collection. Based on the systematic literature review of A'yuninnisa, Carminati, and Wilderom (2023) (figure 2.1, these definitions are part of the results of the most investigated relationships in the field.

Job Demands

- **Crafting Challenging Job Demands** A component of job crafting, where the employee experiences an adequate level of stimulating challenges in their work.
- **Role overload:** The employee feels that there are too many responsibilities expected of them.
- **Role clarity:** The employee has sufficient and available information regarding what is expected of them.

Job Resources

- **Crafting Social Job Resources** The employee has access to a variety of resources, opportunities for development, and access to autonomy
- **Crafting Structural Job Resources** The employee has a support system in the form of social support, supervisory coaching, and feedback (Tims, Bakker, and Derks, 2013)
- **Leader Member Exchange:** The experienced relationship between the supervisor or manager and the employee

- **Team (social) cohesion:** A feeling of belongingness with their colleagues
- **Affective organizational commitment:** The employee is positively and emotionally attached to the organization, identifying with its goals.
- **Autonomy:** The perceived degree the employee has the opportunity to decide regarding job demands and characteristics.
- **Self-efficacy:** The employee's beliefs about their knowledge

Considering the discussed antecedents, the research accounts for most of the elements of Rautenbach, 2015's flourishing at work scale (positive and negative affect; job satisfaction, autonomy, competence, relatedness, engagement, and social well-being), except a direct scale link to life meaning and learning. Arguing based on A'yuninnisa, Carminati, and Wilderom (2023)'s findings, the elements of the study are all strong antecedents of job flourishing. Hence, a high overall score over these antecedents implies a predicted job flourishing (refer to figure 2.3. As this is an exploration within AM to predict, this theoretical background is taken to a more practical level through ML. Hence, this correlation is sufficient and valid for this exploration.

Table 4 Group of significant predictors and outcomes of job flourishing

PREDICTORS			**OUTCOMES**
i. Intrinsic & personal-context variables	**ii. Work-context variables**	**iii. Person-work emerging-state variables**	**iv. Health, behavioral, & performance outcomes**
Individual difference	**Individual-level**	**Motivational states**	**Health-related**
Psychological capital (3/+)	***Job resources (3/+)***	***Job crafting (4/+)***	***Work engagement (3/+)***
Core self-evaluation (2/+)	Autonomy (1/+)	***Work engagement (4/+)***	Job satisfaction (1/+)
Occupational calling (2/+)	Responsibilities (1/+)	*Job meaningfulness (2/+)*	Subjective well-being (1/+)
Growth mindset (2/+)	Learning demands (1/+)	*Work-related stress (2/−)*	Burnout (1/−)
Conscientiousness (1/+)	Career advancement (1/+)	Organizational-based self-esteem (1/+)	
Proactive personality (1/+)	Job demands (1/−)	Positive work reflection (1/+)	**Behavioral & task-related**
Prosocial motivation (1/+)	Emotional demand (1/−)	Problem solving pondering (1/+)	***Extra-role performance (4/+)***
Energy (1/+)	Work pressure (1/−)	Psychological detachment (1/+)	***In-role performance (3/+)***
Optimism (1/+)	(Quantitative) job insecurity (1/−)	Rumination (1/−)	***Intention to leave (3/−)***
Self-efficacy (1/+)	*Job insecurity (1/−)**	Segmentation behavior family-to-work (1/+)	*Counterproductive behavior (2/−)*
Self-esteem (1/+)	*Work overload (1/−)**	Workaholism (1/−)	Creativity (1/+)
Valued living (1/+)	*Workload (1/−)**		Knowledge sharing (1/+)
Mastery-approach goal orientation (1/+)		**Congruent states**	Proactive behavior (1/+)
Mastery-avoidance goal orientation (1/+)	**Unit-level**	*Work-family balance (2/+)*	Community citizenship behavior (1/+)
(Primary) challenge appraisal (1/+)	*Perceived organizational support (2/+)*	*Person-environment fit (2/+)*	Organizational commitment (1/+)
Personal values (1/+)	*Authentic leadership (2/+)*	Person-job fit (1/+)	
Curiosity (1/+)	*Ethical leadership (2/+)*	Values fit (1/+)	
Gender (1)	Empowering leadership (1/+)	Basic needs satisfaction (1/+)	
Pride (1/+)	Participative path goal leadership (1/+)	Organizational embeddedness (1/+)	
(Primary) hindrance appraisal (1/−)	Coworker exchange (1/+)	Boundary violations family-to-work (1/−)	
Performance-avoidance goal orientation (1/−)	Peer support (1/+)		
Hostile attribution bias (1/−)	Staff consultation (1/+)		
	Leadership status (1/+)		
Personal-context	Ethical climate (1/+)		
Community embeddedness (1/+)	Organizational spirituality (1/+)		
Social support from family (1/+)	Corporate social responsibilities (1/+)		
Work-family enrichment (1/+)	Linguistic ostracism (1/−)		
Family hassles (1/−)	Social burden (1/−)		
	Bullying (1/−)		

The parentheses contain information about the number of studies that examined the given variables and the results of the hypothesis testing:

(+) = positively related, (−) = negatively related

The variables in italics were studied frequently; those in bold provided strong evidence of an association with job flourishing, the magnitude range of the relationship is given

*) the results should be considered cautiously due to the presence of non-significant findings

FIGURE 2.3: Group of significant predictors and outcomes of job flourishing (A'yuninnisa, Carminati, and Wilderom, 2023)

In conclusion, job flourishing can be narrowed through work engagement, the main component to be studied in this thesis. Referring to figure 2.3, this thesis includes variables across the different predictor areas, as job flourishing is a multidimensional measure. Particularly, A'yuninnisa, Carminati, and Wilderom (2023) encourages the validation and investigation of these variables in further research involving job flourishing. Additionally, job resources and demands are the main antecedents of work engagement. Hence utilizing the JD-R model is imperative when investigating work engagement.

2.3 Algorithmic Management

Algorithmic Management (AM) is any system that automatically processes machine-readable data input to support or execute decision-making (Duggan et al., 2019; Zalmanson, 2017). AM is currently being applied in various forms throughout the HR life cycle (recruiting, performance, reimbursement, etc.). It encompasses a powerful research stream that has become popular in organizations due to the potential of increased efficiency in HR tasks. This efficiency comes through the advanced processing of employee performance data by exploiting novel data collection methods and increased computational abilities. Three types of AM in HR should be distinguished: descriptive, predictive, and prescriptive (Deobald et al., 2019).

1. **Descriptive**

   These are algorithms that focus on the past and influence the present. The main instrument is descriptive analysis (means, standard deviation, etc.) attempting to describe relationships between variables from historical data. For example, *in the context of job flourishing, descriptive algorithms compile employee well-being data and link it to certain practices within the organization based on historical data.*

2. **Predictive**

   Predictive algorithms focus on forecasting the future as a result of past or real-time data. This is achieved through advanced regression techniques, machine learning (ML), deep learning (DL), or data mining (DM). So far, such predictions have more interestingly been applied in recruitment, when organizations skip interviews and predict suitable candidates (Deobald et al., 2019). *In the context of flourishing, predictive tools would be able to predict in advance whether employees are flourishing or not (or to what extent they are flourishing), based on different features of an employee (eg. tenure, social resources, etc.).* These types of algorithms are the focus of this thesis.

3. **Prescriptive**

   As a step further from forecasting, prescriptive algorithms attempt to provide scenario-based solutions and decisions, aiding the decision process. This is often achieved through simulations, as the field emerges from operation research. *In our context, these algorithms would aim to provide scenarios and practices to award or aid depending on the prediction of job flourishing, for example, by offering training schemes or different scenarios with different types of managerial and organizational interventions to an individual who might be flourishing or not.*

Tambe, Cappelli, and Yakubovich (2019) suggests that the next step for ML algorithms and data science is working in environments that evaluate employees and the current processes/activities of the organization. Moreover, there are a few challenges for AM. Firstly, the complex environment and variables that play a role in HR that lead to decision-making are hard to grasp in computerized models. Additionally, there are a lot of data limitations, particularly in small to medium-sized organizations. Lastly, there is an inherent ethical question on decision-making, which also leads to legalities and autonomy issues with AM (see Tambe, Cappelli, and Yakubovich, 2019; Deobald et al., 2019; Meijerink and Bondarouk, 2023). However, one of the benefits of ML is the ability to predict without the goal of advancing a particular theory. Also, keeping humans in the loop is important with even the most automated and autonomous algorithmic systems (Meijerink and Bondarouk, 2023).

Seeing the practical importance of this research stream, it is natural to link the benefit of well-being prediction. Firstly, so far in the context of employee well-being and flourishing, there has been an interest in causal discovery. Through AM, the output of this casual discovery can become more predictive due to the algorithm's interest in seeing a more complex picture rather than detailing a particular model (Tambe, Cappelli, and Yakubovich, 2019). Secondly, we refer to predictive algorithm usage, classified by research as particularly helpful in recruitment and performance appraisal (including on-boarding and advancement) (Sharma, 2021; Tambe, Cappelli, and Yakubovich, 2019). Predicting employee wellbeing and addressing challenges through HR practices fall under these categories, for which literature reviews establish a huge benefit in the field of HRM (Sharma, 2021; Garg, Sinha, and Kar, 2019).

Intelligent systems learning from specific problems to predict new data, without human intervention are categorized as machine learning (ML) models (Belattar, Abdoun, and Haimoudi, 2023). ML has various applications across different fields, such as educational data mining, pricing predictions, weather forecasting, student performance, and wellbeing (Chaka, 2022; Savadatti et al., 2022; Zhuang et al., 2022). Progress has been made in the field of ML and HRM regarding recruitment (see Meraliyev, Alibekova, and Bekturganova, n.d.; Shet and Nair, 2022; Nurajijah, Wijaya, and Sari, 2022). Additionally, ML is often used with a main focus on employee retention and performance (Pourkhodabakhsh, 2023; Rama et al., 2023). Despite the importance of proactive and predictive HRM across all functions, ML applications are the strongest in the areas of recruitment and performance management, mostly due to their measurability (Garg, Sinha, and Kar, 2019). Due to this measurability, it is implied that the maturity of ML and AM depends on the complexity of the HR problem, depending on well established domain knowledge. Well-being and job flourishing are classified as complex social and psychological contexts.

Tambe, Cappelli, and Yakubovich calls for ML algorithms on topics that "do not involve HR outcomes subject to fairness consideration," such as employee advice for "training programs that make sense for them, new jobs for which they might apply, well-being, and retirement planning advice" (2019). On the contrary, current research through artificial intelligence in HRM is focused on recruitment and turnover (eg. Rama et al., 2023; Shet and Nair, 2022; Meraliyev, Alibekova, and Bekturganova, n.d.; Pourkhodabakhsh, 2023; Jin et al., 2020). Nevertheless, Garg, Sinha, and Kar (2019) emphasizes the need for future research and integration of HRM and ML, where practitioners work together in the high-complex nature of HRM problems. AI for HR is still in its infancy when it comes to complex concepts. Filling this gap implies a need to investigate flourishing through predictive AM, in an attempt to further the knowledge of AM applications in HR practices. To achieve predictive AM, ML models could be used to analyze the data of employees flourishing and not flourishing and predict with new data across work engagement dimensions whether the employee is flourishing. This is a classic binary classification problem (figure 2.4). The following section dives into the background behind prediction models and measurements to best predict binary classification.

2.3.1 Prediction Models

To accomplish a binary classification for job flourishing, the following supervised learning prediction models will be used: Logistic Regression (LR), Random Forest (RF), and Support Vector Machines (SVM). In this section covers major choices made regarding prediction. Firstly, the decision on supervised learning models is outlined based on recent developments in the field of AM. Then, the major trade-off in ML between bias and variance is highlighted. It is this tradeoff that leads to the comparison of the models, improvement techniques, and performance measures. Lastly, each of the chosen models is discussed in terms of its benefits for flourishing classification.

Supervised Learning

Supervised learning models are chosen due to their interpretability. In contrast to unsupervised learning, which focuses on pattern detection and data mining, supervised learning has a baseline understanding of the output values (labeled data). HRM faces a lot of challenges in data authenticity, data availability, data management, and the explainability and fairness of algorithms (Garg, Sinha, and Kar, 2019; Tambe, Cappelli, and Yakubovich, 2019). Supervised learning allows for modeling relationships and dependencies between input features and output, leading to more interpretability. This form of machine learning also allows us to utilize advanced domain knowledge within a limited population size. High interpretability can lead to algorithmic fairness, explainability of results, and control over data management. Within supervised learning, the choice of LR, RF, and SVM comes due to various theoretical and practical reasons.

Bias vs Variance

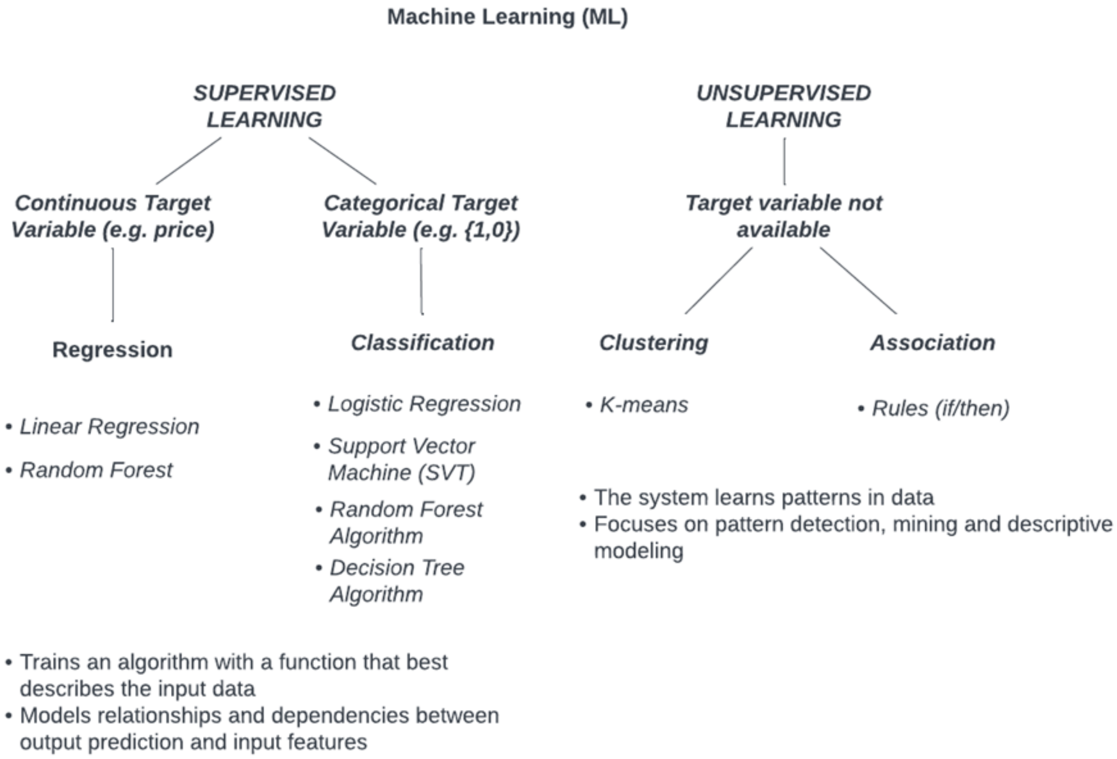


FIGURE 2.4: A simplified overview of supervised and unsupervised learning in ML

Each of the chosen models performs uniquely in terms of bias and variance in prediction. A model's bias is its inability to capture or match the true relationship in the data, leading to an error introduced by simplifying a real-world situation. For example, linear regression can imitate the relationship of the input features, resulting in decent accuracy, but the actual relationship could be exponential.

On the other hand, variance describes the difference in fit of the model across different data sets. To illustrate, a model is trained with a subset of half of the population size, and it is concluded that a linear relationship fits well. However, when testing with the rest of the population, it is observed that this linear relationship leads to a high error, and the actual relationship for this second half is exponential. This implies that the model has high variance across the dataset.

So, each of the constructed models should be evaluated to some extent for the risk of high variance and high bias. By considering these limitations, we recognize potential improvements of the models via optimization methods, converted in the following subsection.

Logistic Regression (LR)

Regression analysis is the most utilized model in explaining the relationship between the antecedents of job flourishing and work engagement (such as Ariza-Montes et al., 2018; Dam, Noordzij, and Born, 2020; Rautenbach, 2015). Binary classification studies often use Logistic Regression (LR). In the early 2000s, articles on LR were education-related on adjustment and performance (Peng, Lee, and Ingersoll, 2002). Kullolli, Trebicka, and Fortuzi (2024) also note down the benefits of utilizing LR for customer satisfaction, particularly due to the capacity to model the probability of a specific outcome based on predictor variables .

As one of the objectives of predicting job flourishing is to create opportunities (supported by figure 2.5) and increase retention possibilities, the findings can be extended in the learning and devel-

HRM function	Objectives for ML application
Recruitment	Assessing suitability of candidates for job positions; extracting information from resumes and analysing applicants' profiles
Selection	Identifying decision attributes for selection and developing selection models
Employee engagement	Understanding brand engagement of employees; current sentiment of employees and factors boosting job stress
Training and development	Identifying training needs; recommending relevant courses and measuring training effectiveness
Performance management	Performance evaluation; performance prediction; detection of bias in appraisal process; estimation of expertise level of employees and development of customized incentives for employees
Employee turnover	Predicting employees' turnover using their personal and work-related factors
Team dynamics	Recommending team members; assessing team performance; understanding teams' sentiments and understanding teams' interaction patterns
Human resource allocation	Allocating people to different categories

FIGURE 2.5: Objectives identified for ML applications in HRM Functions (Garg, Sinha, and Kar, 2019)

opment domain to utilizing LR for HRM purposes. LR is a powerful tool to predict work engagement due to its simplicity, interpretability, and efficiency with small to medium-sized datasets (Pampel, 2022). Additionally, the function can handle both categorical and continuous variables, which makes it interesting for practitioners depending on the data available (Jawa, 2022). However, LR has a risk of overfitting the training dataset and can have difficulty handling imbalanced data, leading to large variances. Nonetheless, LR provides an efficient model with the possibility for good results in binary classification and it creates a baseline for flourishing prediction.

Random Forests (RF)

Similarly, Random Forests (RF) are often used within well-being literature (e.g. Vera Cruz et al., 2023; Hoekstra et al., 2023). RF is an ensemble of decision trees (DT) voting for the most popular classification chosen for various reasons. Firstly, DTs are attractive for high dimensional spaces, as they simplify and approximate the decision with high efficiency and accuracy (Safavian and Landgrebe, 1991). Additionally, they tend to be able to adapt to complex datasets well (leading to low bias), but the accumulated errors from level to level can lead to high variance. Consequently, designing an optimal DT for flourishing can be challenging. For this reason, RF is utilized across the literature to produce better results due to generating multiple possibly optimal DTs.

The forest results depend largely on the strength of individual trees and the correlation between them (Breiman, 2001). So, predicting with an RF can be an efficient and accurate path, but RFs highly depend on the methodology chosen to build the individual trees and evaluate the data.

From a computational and statistical standpoint, they are appealing due to their trainability, few tuning parameters, capability to handle regression and classification, capability for high dimensionality problems, and ability to detect outliers (Cha and Yunqian, 2012). Hence, RF is a viable model solution for predicting flourishing.

Support Vector Machines (SVM)

Garg, Sinha, and Kar (2019) attribute the lack of use of complex algorithms like support vector machines and genetic algorithms in HRM to the limited understanding of how the algorithms operate. Indeed, these complex models are rarely found in the recent HR and organization literature, despite their ability to predict.

SVMs utilize a hyperplane to separate classes based on the input feature and a kernel function (Ezzati et al., 2019). The base of this model is to transform the data (through the kernel function) and determine boundaries between data points. For a domain such as wellbeing, identifying the threshold for which an employee is flourishing can be quite challenging, and dependent on environmental, family, and individual characteristics (A'yuninnisa, Carminati, and Wilderom, 2023).

SVMs work on the interplay of finding an adjustable threshold (decision boundary) to classify the data. An SVM that classifies flourishing employees can be quite accurate, despite the countless choices on a decision boundary (flourishing or not flourishing). The most difficulty with this prediction model is its complex nature and the tuning of the parameters for better results. This process gives SVMs limitations with new data (leading to high variance). Nonetheless, it is an interesting prediction model for flourishing classification due to its ability to handle the complexity of well-being.

Conclusions on Prediction Models

In order to achieve predictive algorithms to further the AM domain within job flourishing, we study **logistic regression, random forests, and support vector machines** in the context of binary classification of flourishing employees (flourishing or not flourishing). These supervised machine learning models tend to have a low bias and, consequently, a high risk of variance. Due to their low bias and high variance nature, the chosen models can capture complex relationships, such as the well-being case. To construct models tailored to job flourishing prediction, data scientists adjust the right parameters. This is the topic of the following subsection.

2.3.2 Hyperparameters and Tuning Techniques

Every ML model has a set of parameters and hyperparameters. The main difference is that hyperparameters are the configuration of parameters set by the practitioner that govern the model's learning process, compared to the parameters that are defined by the given data (Sun et al., 2021). Finding an optimal set of hyperparameters depends on the fine-tuning process, which happens through experimentation and optimization techniques (Siri, Afroz, and Rani, 2024). This section covers the configurable hyperparameters of LR, RF, and SVM, and elaborates on the chosen optimization techniques and how to measure their optimality.

Hyperparameters

Hyperparameters are vital in governing the complexity and behavior of the ML model. To respond to Tambe, Cappelli, and Yakubovich (2019)'s and Garg, Sinha, and Kar (2019)'s call on domain knowledge interacting with HRM in complex ML models, research should consider the importance of hyperparameters for each model. Optimizing these parameters tailors the models to the case of job flourishing.

1 **Logistic Regression (LR)**

As defined in the sklearn python library, LR has a multitude of hyperparameters to enhance its capability. This thesis' scope is limited on the most commonly adjusted parameters according to various data science studies (eg. Sun et al., 2021; Siri, Afroz, and Rani, 2024). The data defines the LR coefficients. However, LR can require regularization, which prevents data overfitting by decreasing model variance. Regularization works by adding a penalty term to the cost function that reduces the prediction error. Table 2.1 presents hyperparameters related to this regularization process, which can lead to improvement in the model's results.

Regarding the penalty choice, it is noteworthy to mention that Elasticnet is a combination of Lasso and Ridge Regularization. Both regression types reduce the LR coefficients in an attempt to regulate based on a penalty added to the cost function (error calculation function). Lasso adds a penalty equal to the absolute value of the magnitude of the coefficient($+\lambda \sum_{j=0}^{p} |\beta_j|$), while Ridge adds the squared magnitude of the coefficients ($+\lambda \sum_{j=0}^{p} {\beta_j}^2$). In both cases, lambda is the regularization strength. So, Lasso coefficients can become 0. Hence, Lasso LR is quite good at dealing with high dimensionality of the dataset. Lastly, the difference between the solvers is beyond the scope of this thesis.

Table 2.1: The Main Hyperparameters of Logistic Regression

Hyperparameter	Description (Sklearn Python library)
Penalty	Choice of regularization methods: None; Ridge (l2); Lasso (l1); Elasticnet
Regularization Strength	The intensity for the regularization (penalty)
Solver	Algorithms that help optimize the minimization of the loss function: lbfgs, liblinear, newton-cg, newton-cholesky, sag, saga
Maximum Iterations	Maximum iteration number to take for converging the solver
Tolerance	Tolerance to stop the iteration

2 **Random Forests (RF)**

Breiman introduced RF with the premise of improved outcomes of decision trees while maintaining a low number of hyperparameters. As mentioned, this is one of the most beneficial aspects of RFs. Table 2.2 displays the chosen parameters to be studied, following similar recent research in other fields (eg. Sun et al., 2021; Moeini, 2024; Hridoy et al., 2024).

Table 2.2: The Main Hyperparameters of Random Forest (Classifier)

Hyperparameter	Description (Sklearn Python library)
Number of estimators	The number of decision trees in the forest
Maximum Features	The maximum number of features to be considered to split a node
Maximum Depth	The maximum depth of the tree (length of the longest path)
Maximum Samples Split	The minimum number of data points placed in a node before the node is split
Minimum Samples Leaf	The minimum number of samples required to be at a leaf node (the end of a decision path)

The dataset defines the DTs and their structure. In RF, the number of trees and their depth are adjustable. An important hyperparameter is the number of features that a node should consider to make a decision (Breiman, 2001). Additionally, the maximum number of samples that need to be placed at a node before it continues to another level can be set, as well as the minimum number of samples that need to be at the end leaf (the final decision). Choosing the right parameters depends on the data and the features.

3 **Support Vector Machines (SVM)**

Similar to LR, SVMs have a regularization aspect with Ridge penalty and a strength regularization parameter. Differently from the LR and RF, the focus of hyperparameters for SVMs is on the kernel function and its coefficients (Lessmann, Stahlbock, and Crone, 2005) (see table 2.3). The choice of the kernel function plays a crucial role in determining the decision boundary, by looking at the data in a linearly separable way. The choice of the kernel function is between linear, polynomial, radial basis, and sigmoid (Pedregosa et al., 2011).Please refer to the scikit learn documentation for SVM Kernels (Buitinck et al., 2013)

Tuning Techniques

The introduced hyperparameters need tuning to avoid underfitting and overfitting the model to the data. In ML, hyperparameter optimization enhances a model's performance (Hridoy et al., 2024). This optimization can be done through metaheuristics (techniques). Multiple studies have been conducted on hyperparameter tuning, including grid search, random search, etc. Here insights can be drawn from Lessmann, Stahlbock, and Crone (2005)'s SVM optimization with Genetic Algorithms (GA), Moeini (2024)'s Bayesian optimization algorithm (BOA) for RF, and Sun et al. (2021)'s LR hyperparameter tuning. According to Garg, Sinha, and Kar (2019) GA and BAO applications are in their infancy in complex processes related to socio-psychological phenomena . These algorithms are well-studied and

Table 2.3: The Main Hyperparameters of Support Vector Machines (SVM)

Hyperparameter	Description (Sklearn Python library)
Regularization parameter	Defines the strength of the regularization. Penalty used is a squared Ridge regularization
Kernel	The kernel function used A kernel is a function that looks into a non-linear surface and transforms it into a linear equation in a higher number of dimension spaces (for more information, refer to Lessmann, Stahlbock, and Crone, 2005)
Gamma/Degree	Kernel coefficient related to the chosen kernel function - gamma (scale, auto): used for rbf, poly, sigmoid - degree: used for poly

popular hyperparameter tuning metaheuristics that result in high-performance enhancement (Mori, Takeda, and Matsumoto, 2005). This section discusses the structure and general benefits of each.

1 **Evolutionary Computing: Genetic Algorithms (GA)**

Evolutionary computing is a set of nature-inspired AI algorithms that solve optimization problems through an iterative process. GA is the most popular algorithm in evolutionary computing. It is a search and optimization technique inspired by the mutation of genes and the idea of "survival of the fittest"(Kumar and Yadav, 2022).

GA utilizes three operations- mutation, crossover, and selection, - to provide an optimal solution (a set of fine-tuned hyperparameters). New solutions are generated by crossing over previous solutions or by mutating them. Crossover combines two solutions found in the population. Mutation allows for diversity through randomization. Each solution is evaluated by a fitness function. For example, Lessmann, Stahlbock, and Crone (2005) uses an algorithmic mean of the accuracy of the model generated from the training and validation set to optimize the kernel parameters in their SVM model . Overall these algorithms tend to be quite powerful in searching for the optimal solution. However, when there is a complex solution structure and the crossover does not go well, there are significant defects (e.g. the crossover destroys the high-performing parts of the solution) (Mori, Takeda, and Matsumoto, 2005).

2 **Estimation of Distribution Algorithms: Bayesian Optimiser Algorithm (BOA)**

Estimation of Distribution Algorithms utilize global information to make a new population, compared to GAs that construct new solutions based on individual ones (Mori, Takeda, and Matsumoto, 2005). BOA is one of the most popular of these algorithms to fine-tune hyperparameters (see Mori, Takeda, and Matsumoto, 2005; Moeini, 2024; Sun et al., 2021). It is robust, fast, and efficient, as it uses past performances (Bergstra et al., 2011). The algorithm models the probability distribution of the observed values and it focuses the search on areas that are expected to provide the right information to find a solution (find the right parameter configuration). It differs from GA as it does not have a risk of crossover malfunctions. BOA attempts to preserve the building blocks of a good solution by "learning" the distribution of the hyperparameters.

Conclusions on Hyperparameters and Tuning

To create tailored prediction models, researchers optimize their hyperparameters. Drawing from Mori, Takeda, and Matsumoto (2005) and Garg, Sinha, and Kar (2019), who note the utilization of metaheuristics, our study focuses on the application of GA and BOA to fine-tune LR, RF, and SVM models. In LR we work on fine-tuning the regularization choice. In RF, GA and BOA will be studied to build the forest (i.e., the number of trees, the depth, etc.). Lastly, for SVMs, we focus on finding the optimal regularization strength and kernel function for the job-flourishing case. To measure the performance of the model and evaluate it, common metrics for ML models are explained in the following chapter. These performance measures relate to the second research question regarding what defines a good model for job-flourishing prediction.

2.3.3 Measuring performance

In the previous chapter, we noted the important trade-off between bias and variance in the prediction models. As LR, RF, and SVM can suffer from moderate to high variance, hyperparameter tuning is important to be applied. This leads to an improved MSE, which in turn can improve the accuracy of the model. This is the model evaluation approach that some previous research in flourishing has taken (Vera Cruz et al., 2023; Pap et al., 2022).

However, accuracy is not the only way to evaluate and measure the prediction ability of the model. This section evaluated other metrics to measure performance and set the scene for comparisons of our models and tuning techniques. The tuning will optimize a fitness (objective) function that is based on improved accuracy and F1-score. Lastly, we aim to record and present all available metrics.

It is important to note the trade-offs between recall and precision (represented in F1 score), as there is no acceptable benchmark yet within flourishing or well-being prediction. Benchmarks are there to support practitioners and researchers and guide discussions regarding the objectives behind the predictions.

TABLE 2.4: Confusion Matrix.

Predicted	**Actual**	
	Positive	Negative
Positive	True Positive	False Positive TYPE I Error
Negative	False Negative TYPE II Error	True Negative

Confusion Matrix: A matrix showing the counts of true positive, true negative, false positive, and false negative predictions (see table 2.4). These examples are described in terms of flourishing and not flourishing (binary), and this thesis refrains from the usage and classification of languishing or the flourishing continuum.

True Positives (TP): The model has correctly predicted that an individual is flourishing.

True Negatives (TN): The model has correctly predicted that an individual is not flourishing.

False Positives (FP): The model has predicted that an individual is flourishing, but in reality, the individual is not (according to labeled data).

False Negatives (FN): The model has predicted that an individual is not flourishing, but in reality, the individual is flourishing.

Accuracy (2.1): Indicates the general correctness of the model based on the labeled data. In the context of flourishing, it would represent the cumulatively correctly identified flourishing and not flourishing individuals (TP and TN).

$$Accuracy = \frac{TruePositives + TrueNegatives}{TotalNumberofObservations} \tag{2.1}$$

Accuracy can be misleading as a standalone metric if there is an imbalance in the data. For example, if most of the training data instances are classified as not flourishing, then the model will be well-trained in this classification, leading to high accuracy. However, it might still contain a low TP percentage.

Precision (2.2): It indicates the proportion of correctly predicted positive observations to the total predicted positives. Hence, precision represents the proportion between correctly identified employees flourishing (according to the labeled data) and the predictions that say that an employee is flourishing.

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \tag{2.2}$$

Often, high precision is important when the cost of false negatives is high. For example, in the context of manufacturing, highly precise models are important, as otherwise non-defective products would be thrown away. The question we ask is: What are the negative consequences of incorrectly identifying someone as flourishing when they are not? Are these consequences critical enough to require a high level of precision at the cost of sensitivity (as a direct trade-off)?

Recall/Sensitivity (2.3): Indicates the proportion of true predicted positives over all the true positives and false negatives. This is a direct trade-off with precision. Recall is the proportion of correctly predicted flourishing individuals out of the total number of individuals who are genuinely flourishing.

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} \tag{2.3}$$

Recall is important when the cost of failing to identify someone as flourishing is high. For example, models with high sensitivity are often used in cancer treatment, as the cost of misdiagnoses can be fatal for the patient. Similarly to precision, we ponder the consequences of not having a highly sensitive model for predicting flourishing individuals. How sensitive should a prediction tool be in this domain?

F1- Score (2.4): The F1 score is the harmonic mean of precision and recall. A high F1 score usually indicates a good balance between precision and recall. However, the interpretation of "high" varies across domains (Kavyashree, Vimala, and Shreyas, 2024)

$$F1 = 2 * \frac{Recall x Precision}{Recall + Precision} \tag{2.4}$$

Error Rate (2.5): The proportion of misclassified observations.

$$ErrorRate = \frac{FalsePositives + FalseNegatives}{TotalNumberofObservations} = 1 - Accuracy \tag{2.5}$$

The error rate does not capture the difference between false positives and false negatives. So we pose a question for discussion over the difference in consequences for organizations and practitioners if an AM tool incorrectly identifies someone as flourishing when they are not and if it identifies someone as not flourishing when they are.

Sometimes researchers utilize the MSE as a measure for the error rate of a model (Vera Cruz et al., 2023; Kavyashree, Vimala, and Shreyas, 2024)

Conclusions on Performance Measures

This section outlines possible performance indicators that can be used by practitioners. Predicting in the context of job flourishing is quite novel. The techniques used are still novel within HRM and quite complex.(Garg, Sinha, and Kar, 2019). For these reasons, we choose to initially optimize towards an acceptable accuracy and F1 score while measuring error rate, precision, and recall, and reporting on the confusion matrix per studied case. The choice to maintain a balanced F1 comes at the cost of not putting importance on recall or precision, as there is a lack of understanding of whether there is a cost of not knowing if someone is flourishing or if there is a cost of mislabeling someone as flourishing or not. This is an important question to be considered during the results and discussion. Importantly, cross-validation techniques need to be implemented to achieve true benchmarks and understand the quality of predicting job flourishing.

Chapter 3

Research Methodology

3.1 Research Design

The research design is primarily quantitative, designed to answer the two main objectives of this thesis as per the research question: **What are the best methods for predicting job flourishing, considering different criteria?** This quantitative design includes descriptive statistics, inferential statistics (such as correlation analysis, ANOVA, and regression), and factor analysis.

The main reason behind this design is the needed quantitative comparison among different methods, to determine the most optimal achievements. Additionally, through the quantitative results, it is possible to generalize the theoretical and practical results with the context of similar organizations. Hence, the research design aligns with the two parts of the research questions: best methods and different criteria.

The first part of the research question relates to predictive AM algorithms that can be used in the context of job flourishing. To answer this, three predictive ML models for binary classification are applied. Two algorithms are considered to tune (adapt) these models to the case of job flourishing in a higher education institute.

The second query in this research question is what makes a model successful in the context of job flourishing. Here, various performance measures are considered. The main aim was to map the methods and algorithms across these measures and provide an overview for practitioners and future researchers. With this overview, we create a first instrument to discuss what benchmarks organizations can use to successfully predict job flourishing through AM.

3.2 Data Collection

We utilized primary data from the survey conducted on academic and non-academic staff at higher-level education institutions in the Netherlands. The dataset is selected due to the versatility of the data, based on various scales, according to recent flourishing at work studies. Data collection was done in 2022 in two waves.

- **May - June 2022**: Collection of 706 instances In this data collection period, there were 706 responses, conducted in English and Dutch. However, 127 of these instances had missing data in demographics. Hence we can process 579 instances from this wave, in terms of demographic analysis.
- **November - December 2022**: Collection of 1076 instances In this data collection period, there were 1076 responses, conducted in English and Dutch. However, 548 of these instances were incomplete. Hence, we can process 528 instances from this wave in terms of demographic analysis.

We conducted an ANOVA analysis between the two waves and the average mean of work engagement (WE) on the dimensions of vigor and dedication (please refer to 3.3 for the measurement items). Across these two scores, there was not enough evidence to indicate a difference in means between WE scores in wave 1 and wave 2 for p = 0.05 (**WE_Vigor**: F = 0.638, p = 0.425; **WE_Dedication**: F = 1.434, p = 0.231).

With this result, we merge the dataset between the two waves, resulting in 1451 instances with sufficient data. These observations will be used for training, testing, and validation. Tables 3.1 and 3.2 give an overview of the demographic distribution of the data across the two waves. Rautenbach, 2015 found a correlation between the potential for flourishing and tenure and job level, mainly due to the employees' job competence level. No such relation is found between gender and place of birth regarding job flourishing or WE. Hence, we analyzed tenure and level of education as moderating variables, and their significance will be reported in the results.

Additionally, an initial analysis of demographics indicates an insignificant difference between male/female participants in the merged dataset. One-way ANOVA results of Gender on Tenure, indicate that there's no evidence to reject the null hypothesis of equal means across groups (**Wave 1:** $F = 0.42, p = 0.519$; **Wave 2:** $F = 3.59, p = 0.587$). However, when it comes to the gender means on the level of education, the hypothesis is rejected (**Wave 1:** $F = 16.78, p = 4.8e^{-05}$). Hence, there might be a relational factor in terms of gender and level of education, which can influence job flourishing scores.

TABLE 3.1: Wave 1 Demographic Frequencies.

Gender		**Level of Education**		**Tenure**		**Place of Birth**	
Male	259	Msc, MA or LLM	227	1 - 5 years	220	Netherlands	403
Female	259	PhD	185	20+ Years	89	EU/ UK	80
Unknown	57	BA	112	<1 year	85	Outside EU/UK	57
Other	4	Senior Secondary Vocational Education	36	10 - 20 years	78	Unknown	39
		BSc	11	5 - 10 years	68		
		Secondary Education	7	Unknown	39		
		Primary School	1				

TABLE 3.2: Wave 2 Demographic Frequencies.

Gender		**Level of Education**		**Tenure**		**Place of Birth**	
Female	247	Msc, MA or LLM	187	1 - 5 years	148	Netherlands	483
Male	232	BA	154	20+ Years	116	Unknown	23
Unknown	48	PhD	107	10 - 20 years	91	EU/ UK	13
Other	1	Senior Secondary Vocational Education	61	<1 year	78	Outside EU/UK	9
		BSc	12	5 - 10 years	59		
		Secondary Education	6	Unknown	36		
		Primary School	1				

3.3 Measures and Operationalization

In the context of the survey, work engagement (WE) is measured according to the Utrecht Work Engagement Scale, in the dimension of Vigor and Dedication (table 3.3)(Schaufeli and Bakker, 2004). The Absorption dimension is not part of this study. The data collection includes the variables introduced in figure 2.1. Table 3.4 presents an overview of all the constructs and the items used for measurement, as well as their internal reliability. Cronbach's alphas indicate good internal reliability for most constructs. An exception is made for crafting challenging job demands, which has a lower Cronbach's alpha value. We chose to include the two items in the study due to their potential effect on WE and role overload (Tims, Bakker, and Derks, 2013).

TABLE 3.3: Measuring Work Engagement

Vigor	At my work, I feel full of energy My job gives me energy When I get up in the morning, I feel like going to work
Dedication	I am enthusiastic about my job I am proud of the work that I do My job inspires me

We highlight here design choices regarding adapting the usual measurements of some of the variables. Role clarity and role overload have been based on the Rizzo, House, and Lirtzman's early scale development (1970). However, role overload has been adapted to include psychological demands and workload measures. Additionally, we measure autonomy and self-efficacy as described by Spreitzer based on self-determination theory (1995). We also introduce a measure of wellbeing in terms of self-reported mental and physical health in line with the inclusion of hedonic wellbeing in job flourishing antecedents. All items have been measured in a 7-point Likert scale unless otherwise mentioned in table 3.4. This thesis argues with the following rating of this 7-point Likert scales: low item values (1 - 2), moderate item values (3 - 5), high item values (6 - 7).

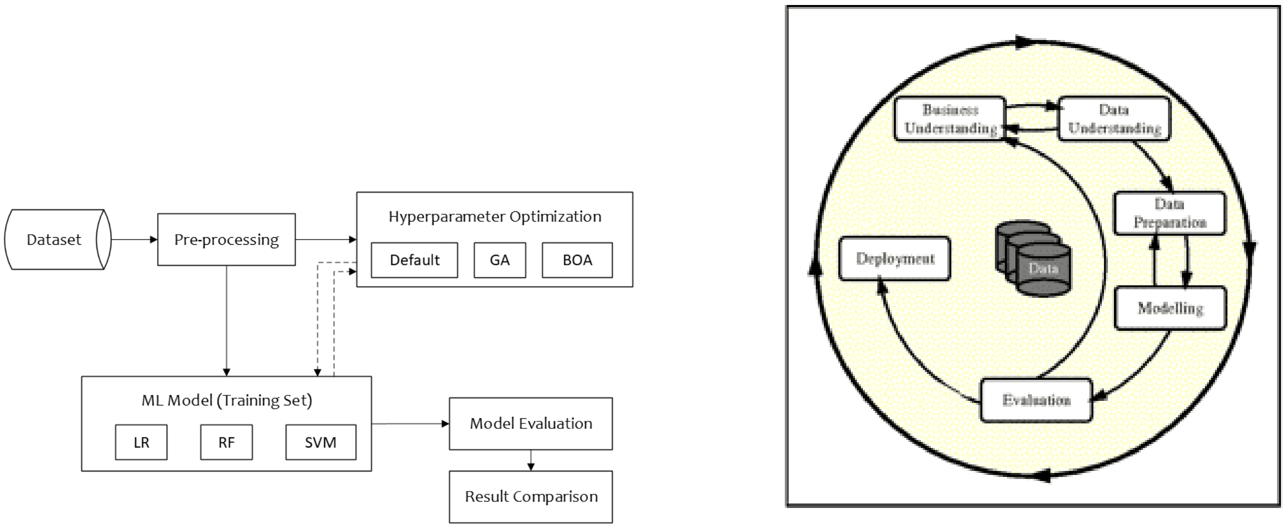
TABLE 3.4: Overview of Constructs and Items (Features)

CONSTRUCT	ITEMS	CRONBACH'S ALPHA	BUILT UPON
Affective Organizational Commitment	I really feel as if the organization's challenges are my own I feel like 'a part of the community' at the organization I feel 'emotionally attached' to the organization I feel a 'strong' sense of belonging to the organization	0.878	Allen and Meyer, 1990 Originally: 8 Items Scale
Autonomy	I have autonomy in determining how I do my job I can decide on my own how I do my work I have considerable opportunity for independence and freedom in how I do my work	0.901	Spreitzer, 1995 Part of Self-determination
Crafting Challenging Job Demands	I regularly take on extra tasks, even though I do not receive extra salary for them I start new projects at work	0.610	Tims, Bakker, and Derks, 2012 Originally: 5 Items Scale
Leader Members Exchange (LMX)	I know how satisfied my supervisor is with what I do My supervisor understands my needs well My supervisor recognizes my qualities The probability that my supervisor uses their influence to advance my interest at work is high I have enough confidence in my supervisor My working relationship with my supervisor is good	0.912	Graen and Uhl-Bien, 1995 Originally: 7 Items Scale
Role Clarity	I know what my responsibilities are I know what my supervisor expects of me It is clear to me what I need to do in my job	0.840	Adapted from Rizzo, House, and Lirtzman, 1970 and Beauchamp et al., 2002 Originally: 15 Items scale
Role Overload	I have difficulties relaxing at work Problems at work stay on my mind when I am not at work Problems at work occupy my thoughts even during my vacation My workload is ...	0.800	Adapted from Rizzo, House, and Lirtzman, 1970 and Beauchamp et al., 2002 Originally: 15 Items scale over 9 dimensions Workload is measured in a 5 items scale "Very low to Very High"
Self-Efficacy	I am confident about my ability to do my job I am self-assured about my knowledge and skills necessary for doing my job I have mastered the knowledge and skills necessary for my job	0.869	Spreitzer, 1995 Part of Competence
Structural Job Resources	I develop my knowledge and professional skills I try to learn new things at work I make sure that I use my capacities to the fullest I decide on my own how I do things	0.806	Tims, Bakker, and Derks, 2012 Originally: 5 Items scale
Social Job Resources	I ask my supervisor to coach me I ask whether my supervisor is satisfied with my work I ask others for feedback on my job performance I ask colleagues for advice	0.745	Tims, Bakker, and Derks, 2012 Originally: 5 Items scale
Team (Social) Cohesion	I feel a sense of belonging with my colleagues I get along with my colleagues I like my colleagues	0.837	Sargent and Sue-Chan, 2001 Originally: 4 Items scale
Wellbeing (Mental & Physical)	In general, my mental health is ... In general, my physical health is ...	0.711	5 Point Likert scale

3.4 Model Building Process

In this section, we present the procedure for building a prediction model. As displayed in figure 3.1a, we started with data preprocessing and continued model building while fine-tuning the parameters. This empirical framework is in line with methodologies such as CRISP-DM (Wirth and Hipp, 2000). In CRISP-DM, the standard for data mining and data science methodological approaches, model building is done in iterative steps. This iteration comes due to the high relationship between data and business (contextual) understanding as well as the "trial and error" nature of data preparation and modeling.

In this section, we summarize the main methodological design choices, the results of which are presented in the subsequent section (Results) together with the final model predictions. Data preparation and analysis were conducted in a combination of SPSS and Python libraries such as sklearn.



(A) Research Design Framework for Model Building

(B) CRISP-DM Data Science Modeling (Nahm, 2022)

FIGURE 3.1: Research Framework and Methodological Approach

3.4.1 Data Preparation

In data preparation, we focused on dealing with missing data and labeling job flourishing according to WE scoring. At first, we decided on the number of instances that could be utilized by looking at how much information each instance withheld with minimal assumptions. An exploratory descriptive analysis of the features indicated minimal outliers. Particularly, tenure and job title had 87 and 91 cases (respectively) of the category 'Prefer to not say'. For the prediction model, this value was treated as a separate category, without imputation to preserve the original intent of the data and reduce additional assumptions. The decision was particularly taken on the basis of the previous findings of Rothmann (2014) regarding the potential relation of tenure and job title to job flourishing. For the Age variable, we chose a simple imputation of the mean value of all missing and 'Prefer not to say' categories. All data was checked for inconsistencies and syntax errors before the data analysis.

Lastly, in data preparation, we labeled according to the WE-score measurement flourishing employees in the scale (languish, moderate well-being, and flourish). This was done to imitate the natural three categories in the job flourishing continuum (Keyes, 2005). Then the data was recorded into the dichotomy flourishing (1) and not flourishing (0) indicating employees who show job flourishing through high WE-scores and employees who have moderate well-being or are languishing.

We operationalized the dichotomous job-flourishing label as the work engagement score, an equal weighting of Vigor and Dedication. According to Schotanus-dijkstra and Pieterse (2016), about 36.5% of Dutch people are flourishing, while 1.6% are languishing. Taking this statistic we split the data on the 36.5 percentile of the WE score as flourishing. To validate this division we conducted a Kcluster analysis with SPSS to identify any discrepancies. This unsupervised learning method identified 3 cluster patterns (languish, moderate well-being, flourishing) on the WE items. This new label was used for robustness validation in the three models.

Lastly, we note that all categorical data was one-hot encoded to prevent ordinal relationships in categorical data. One-hot encoded transforms all the categories per feature in (1,0) format. For continuous features, we normalized the features through standard scaling, aiming for comparable means and higher model performance.

3.4.2 Data Analysis

For better data understanding we utilized a correlation analysis based on the correlation matrix, focusing on variables that are closely related to the label job flourishing. Additionally, we confirmed the dimensionality with a factor (principal component) analysis with WE as the dependent variable. The aim was to identify model simplifications and understand patterns in the data. The outcomes are presented in the Results chapter and Appendix A.

Additionally, following the iterative CRISP-DM methodology, we conducted three iterations of scaling and feature engineering on the basis of correlation analysis and factor analysis. Through these two analyses, we explored the creation of meaningful features and enhanced interpretability. Hence, we interpreted their results and utilized them in each iteration. Specifically:

1 **First Iteration:** Keeping most constructs in a variable (item) scale. The interpretation of the analysis of the correlation of each item to the job flourishing label was used to create partial composite variables, which were added to the dataset. There is no dimensionality reduction and clear model simplification.

2 **Second Iteration:** Based on the results of the factor analysis and correlation analysis, we create composite scores of higher constructs (such as LMX and Structural job resources). The model is simplified by removing previous partial composite variables and item-level features (such as LMX1, LMX 2, etc.).

3 **Third Iteration:** Based on the factor analysis results, we integrated moderating variables. These come from features that were cross-loading. Additionally, we aim at the creation of new meaningful variables based on factor loading. We also implemented some feature scaling. The minimal feature scaling was pursued due to the potential loss of interpretability if further transformations were undertaken. Instead of dimensionality reduction, we aimed to create new higher-order constructs and moderators.

3.4.3 Training, Testing, and Validation

Due to the moderate sample size and soft boundaries, we need to be wary of misclassified instances. Reliability can be assured through cross-validation in training, as well as training with an additional validation set. Training, Testing, and Validation were done in 80%, 10%, and 10% random split. Validation accuracy was recorded to see how the models performed on different datasets.

3.4.4 Models and Optimization

We construct a separate LR, RF, and SVM model. For each of these models, we compared the results of the default parameter settings with tuning with GA and BOA, as described in section 2.3.2. We utilized the sklearn_genetic and skopt library for GASearchCV and BayesSearchCV, grid searching algorithms based on BOA and GA focused on hyperparameter tuning and utilizing cross-validation of 5 folds. The fitness scoring was done by maximizing the Area under the Curve (AUC) of the Receiver operating characteristic curve (ROC) (Nahm, 2022). ROC is a plot of the true positive rate (sensitivity/ recall) and the false positive rate (1 − specificity), measuring the ability to differentiate between classes while capturing the balance between sensitivity and specificity (Kullolli, Trebicka, and Fortuzi, 2024). The choice of optimizing with a different metric from the recorded criteria was made to reduce potential bias in evaluating the models. Importantly, to optimize for recall or precision, we can tune the hyperparameters to focus on this.

3.4.5 Evaluating and Reporting

We report and evaluate the models based on recall, precision, F1-score, and Accuracy. Additionally, we present the confusion matrix to highlight areas of improvement. This evaluation is done per iteration and then in an overall comparison across the different feature engineering levels. We believe such a comparison provides answers to our main research question regarding best prediction models.

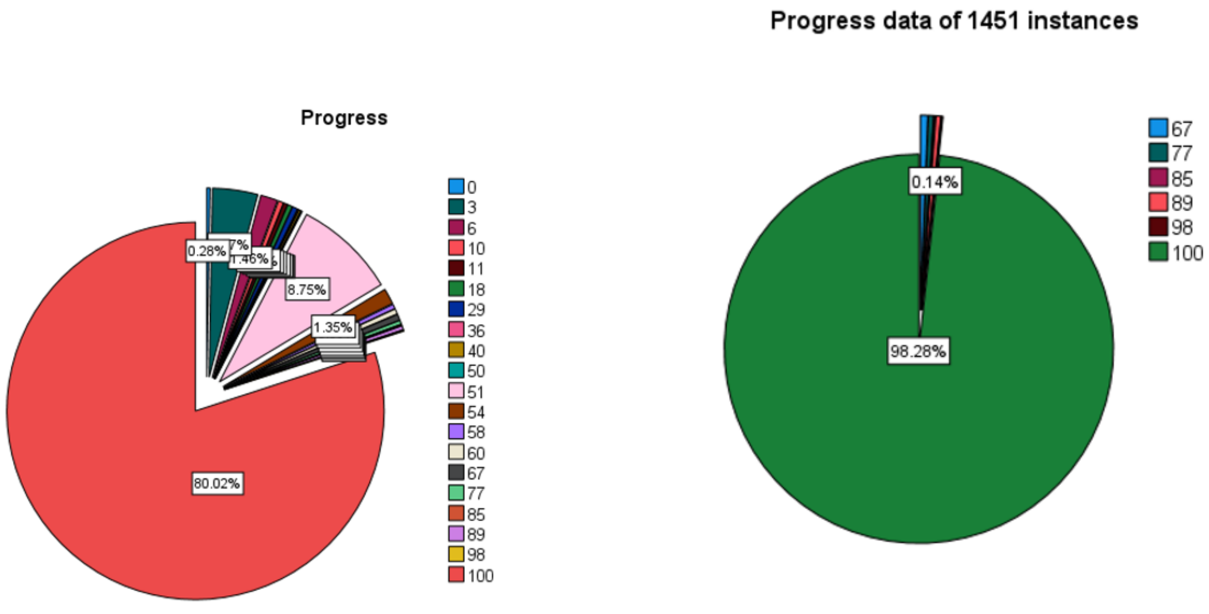
Chapter 4

Results

The main goal of the research was to explore three predictive AM algorithms that can be used in the context of job flourishing, for binary classification. These algorithms are compared and evaluated through different criteria and iteratively applied to three different feature sets. Here, we want to provide results to further discuss the practical implications of job flourishing prediction and the theoretical backbone behind such a task. Hence, we display results in regards to the ML algorithms and on the features (items and constructs) that led to their predictions.

4.1 Data Preparation

4.1.1 Missing Values



(A) Dataset completion progress

(B) Final 1451 instances completion progress (Nahm, 2022)

FIGURE 4.1: Missing Data based on survey completion progress

The survey consisted of 1782 instances, each with different completion progress, as indicated in Figure 4.1a. Out of this set, 229 lacked information regarding WE measurement, and they were excluded from the study. A total of 1451 instances can be utilized for supervised learning prediction. These instances had a completion progress above 67%. From the selected data, only 1.74% had progress less than 100%. The missing information for these instances was related to demographics and the mental and physical well-being score (for those with 67% completion progress). For these instances, we imputed the mode of all instances with the same WE score. For all demographics, the 'Prefer not to say' option was categorized as a value (80–100 instances have these imputations). This way, we maintained the original answer and avoided unnecessary assumptions. Regarding the age of the participants, all missing instances and the 'prefer not to say' option (a total of 294 instances) were imputed to the mean age (39.94).

4.1.2 Job Flourishing Labeling

The WE score was calculated as an average of the Dedication and Vigor scales, based on three items each. The values follow a seemingly normal curve (Leguina, 2015), with Kurtosis 1.075 and Skewness -0.904, and a peak around the value of 6.00 (7-point Linkert scale as shown in figure 4.2). According to Schotanus-dijkstra and Pieterse (2016), 1.6% of the Dutch population languishes, and 36.5% flourishes. In terms of WE distribution, based on these findings, languishers correspond to a value of 2.67 or below (1.7%), while flourishes have a score of 6.00 and up (38.1%). Hence, we choose a binary classification of all employees that score 6.00 and higher as flourishes.

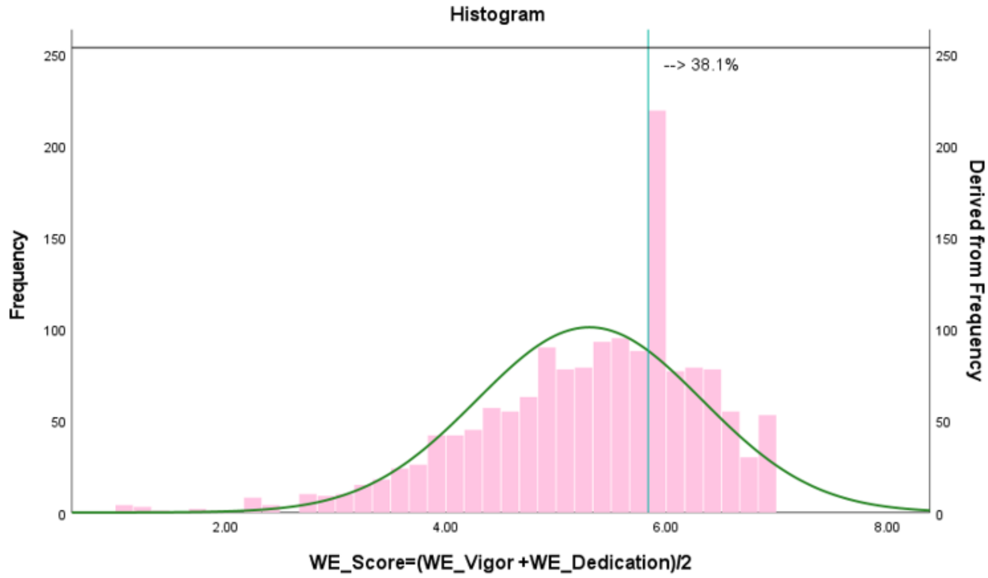


FIGURE 4.2: Distirbution of WE scoring across 1451 instances

For validation, we check this labeling with a preliminary K-means classification. The SPSS K-means algorithm indicates 50% of the population in the flourishing cluster, compared to the selected 38.1%. This would create a more balanced classification; however, it contradicts the found literature that less than half the population flourishes. For further analysis, we continued utilizing the 38.1% flourishing division (binary classification) and used the K-means categories to check the robustness of one of the best-performing models. Figure 4.3 shows the frequency results for the cluster of cases (K-Means) and flourishing label, based on the WE items. The centroids can be found in Appendix B.

Flourishing

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	904	62.3	62.3	62.3
	1.00	547	37.7	37.7	100.0
	Total	1451	100.0	100.0	

Clustering_Flourishing

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	726	50.0	50.0	50.0
	1.00	725	50.0	50.0	100.0
	Total	1451	100.0	100.0	

FIGURE 4.3: Flourishing classification and labeling

4.2 Data Analysis

In this section, we describe the findings from the conducted correlation analysis (based on the correlation matrix) and the principal component analysis. These findings are then integrated into the iterative feature engineering. So, we report the included features per iteration.

4.2.1 Correlation Analysis

Firstly, we conducted a series of correlation-based analyses for features that relate to the flourishing label. The most correlated items to job flourishing (p >= 0.3) were three items from structural job resources ("I make sure I use my capabilities to the fullest" (p = 0.46); "I try to learn new things at work" (p = 0.38); "I develop my knowledge and professional skills" (p = 0.38)), two items from organizational commitment ("I feel a strong sense of belonging" (p = 0.35); "I feel like a part of the community at the organization" (p = 0.3)), and one from role clarity ("It is clear to me what I need to do in my job" (p = 0.31)). Structural job resource items related to the usage of capabilities and learning and development were the most correlated to job flourishing (high WE scoring). The correlation matrix also revealed strong correlations between these three items of structural job resources (p > 0.5) and three items of organizational commitment (p > 0.5).

For further investigation, we looked at the values (1 -7) of these correlated features and their correlation to a positive job flourishing classification (figure 4.4). This revealed that flourishers are most correlated to high values of the three discussed structural job resources and high values of organizational commitment item 4 ("I feel a strong sense of belonging to the organization"). Noteworthy is the correlation with high values of role clarity item 1 ("I know what my responsibilities are") and the high correlation of not flourishing employees to average values of the discussed structural job resources items (values 3 and 4).

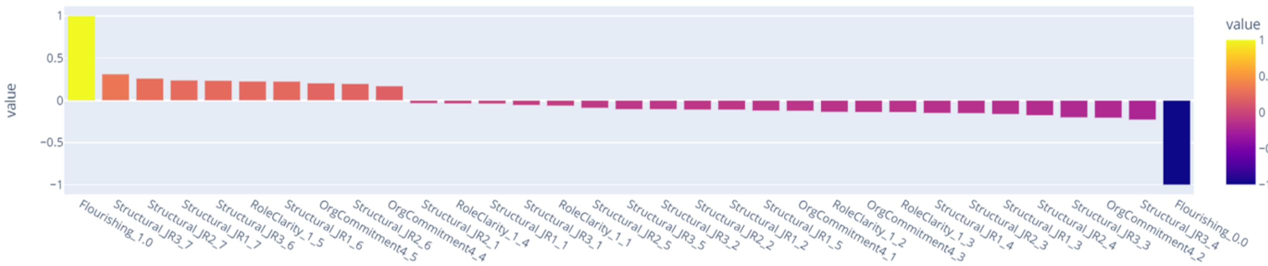


FIGURE 4.4: Correlation between Feature Values and the Binary Flourishing Classification

4.2.2 Factor Analysis

To enrich our analysis, as mentioned in the Methodology, we conducted a factor analysis. The analysis was conducted with a total of 46 variables, which included demographic information. The KMO for these variables was 0.873, implying good sampling adequacy and benefits in conducting a factor analysis. Additionally, Barlett's test for sphericity indicates that there are significant relationships in the dataset (test statistic = 37162.995, df = 1081, $p < 0.001$).

The factor analysis, conducted via SPSS, was a principal component analysis with direct oblimin rotation (delta = 0) due to the possible correlation among the variables. 12 factors were extracted based on scree plots and eigenvalues higher than 1, which retain 71.022% of the variance. From these, component one explains about 19.8% of the variance. This component is loaded mainly by the LMX items, and partially from role clarity item 2 ("I know what my supervisor expects of me"), and social job resources item 1 ("I ask my supervisor to coach me"). Hence, the component is a representation of supervisor relationships.

Additionally, we identified cross-loadings in components 5 (5.72% of the variance) and 12 (2.25% of the variance) from physical wellbeing, mental wellbeing, and education. Component 5 mainly loads the three items from Role Overload, implying that there is a relationship between how employees perceive their workload and their self-perceived well-being, possibly related to their attained education. Similarly, components 2 (9.1% of the variance) and 3 (6.7% of the variance) are cross-loaded by social job resources item 1, – “I ask my supervisor to coach me”,- social job resources item 2, – “I ask whether my supervisor is satisfied with my work”, - and challenging job demands item 2, – “I start new projects at work”. Component 2 is a composition of these three factors age, education, and job title, while Component 3 is a combination of social job resources and challenging job demands. These factor loadings indicate a moderating effect of education and job title towards social job resources and challenging job demands. These variables moderate an employee’s feedback loop with their supervisor and their willingness to pick up new projects.

Moreover, we investigated the structure matrix for underlying relationships. In addition to the previously cross-loading variables, we found multiple high indirect relationships of structural job resources item 3 (“I make sure that I use my resources to the fullest”) and LMX item 6 ("My working relationship with my supervisor is good"). To emphasize, mental well-being had moderate to high indirect relationships with multiple components. Particularly unique was its indirect influence on the component loading from structural job resources and on the one loading from organizational commitment.

The main implications of this factor analysis are on feature scaling and engineering. For this reason, we relate back to the correlation analysis, where items from structural job resources, organizational commitment, and role clarity are the most correlated to the flourishing label. Structural job resources 3 and 4 also hold moderate correlation among components. Meanwhile, structural job resources 1 and 2 are cross-loading features, with direct contributions to multiple extracted components. The first component explains almost a quarter of the variance and is mainly loaded from the LMX items. Lastly, we note the moderating nature with moderate loading and moderate correlation of mental well-being, education, and crafting challenging job demands item 2. In other words, the complex relationship between supervisor and employee, as explained by LMX and social job resources items, can be key in predicting flourishing, moderated by learning and development through crafting new projects.

4.2.3 Iterative Feature Engineering

Drawing from the correlation and factor analysis, along with the literature supporting the various items utilized, we go through an iterative process of feature scaling and engineering. This section describes the chosen dataset per iteration and highlights the main constructs.

1. **Iteration 1: Item Level Descriptions**

   In this iteration we aimed to explore model performance for item level features. We aimed at reducing multicollinearity of the variables, which can lead to suboptimal model performance, by combining the highly correlated variables based on the correlation matrix. Subsequently, composite components were created and added to the dataset: Structural job resource (items 1, 2 3) (p = 0.46 with flourishing), Organizational Commitment (items 2, 3 4) (p = 0.33 with flourishing), Role Clarity (items 1,2 3) (p = 0.33 with flourishing), Leader Member Exchange (items 1, 2 3) (p = 0.29 with flourishing). These features were added to the dataset, with no dimensionality reduction (Table 4.1).

2. **Iteration 2: High Order Constructs**

   In this iteration we draw upon insights from the Factor Analysis. As component 1 which explained the most variance on the dataset was mostly based on LMX items, we created an LMX composite score (average score of 6 item)(p = 0.29 with flourishing). Similarly, we combine all items of structural job resources (p = 0.46) as they show high correlations and cross-loadings in the factor analysis. Hence, from the previous dataset we remove individual LMX items and Structural Job Resources (items 1 and 2) and their previous composite score, to reduce dimensionality and simplify the model. We keep structural job resources item 3 and 4 due to the associated high

Table 4.1: Features used in iteration 1

Item Level Features (Categorical)				
- Structural JR 1	- Role Overload 1	- LMX 1	- Self-Efficacy 1	- Age
- Structural JR 2	- Role Overload 2	- LMX 2	- Self- Efficacy 2	- Education
- Structural JR 3	- Role Overload 3	- LMX 3	- Self- Efficacy 3	- Job Title
- Structural JR 4	- Role Overload 4	- LMX 4	- Org. Commimtnet 1	- Gender
- Social JR 1	- Role Clarity 1	- LMX 5	- Org. Commitment 2	- Family Status
- Social JR 2	- Role Clarity 2	- LMX 6	- Org. Commitment 3	- Home Situation
- Social JR 3	- Role Clarity 3	- TSC 1	- Org. Commitment 4	- Tenure
- Social JR 4	- Autonomy 1	- TSC 2	- Physical Wellbeing	- Contract Type
- Challenging JD 1	- Autonomy 2	- TSC 3	- Mental Wellbeing	
- Challenging JD 2	- Autonomy 3			
Average Scores Based on Correlation (Continuous)				
- Structural JR (1,2,3)				
- Org. Commitment (2,3,4)				
- Role Clarity Composite				
- LMX (1,2,3)				

values that predict flourishing. Removing these items, might lead to lost information in averaging. In fact, values of 6 and 7 in Linkert scale of structural job resource item 3 ("I make sure that I use my capabilities to the fullest") have a positive correlation of 0.24 and 0.31 with flourishing. Compared to the fact that the other stronger correlations are form composite scores, this variable is significant in predicting flourishing.

Moreover, due to the high loadings (around 0.9 in all items), we create composite scores (the average score of items) for Autonomy, Self-Efficacy, and TSC. Lastly, we combine the two items social job resources which had a direct contribution to multiple components of the factor analysis (item 1 2). Please refer to Table 4.2 for the full list of items.

Table 4.2: Features used in iteration 2

Item Level Features (Categorical)		
- Structural JR 3	- Role Overload 1	- Age
- Structural JR 4	- Role Overload 2	- Education
- Social JR 3	- Role Overload 3	- Job Title
- Social JR 4	- Role Overload 4	- Gender
- Challenging JD 1	- Org. Commitment 1	- Family Status
- Challenging JD 2	- Physical Wellbeing	- Home Situation
	- Mental Wellbeing	- Tenure
		- Contract Type
Average Scores Based on Correlation and Factor Analysis (Continuous)		
- Structural JR Composite	- Autonomy Composite	
- Org. Commitment (2,3,4)	- TSC Composite	
- Role Clarity Composite	- Self-Efficacy Composite	
- LMX Composite	- Social JR (1,2)	

3. **Iteration 3: Engineered Constructs and Moderators**

In this third iteration, we keep all the previous adjustments and focus on interaction variables (Table 4.3). Based on the factor analysis, education, mental wellbeing and physical wellbeing have cross-loadings as well as high indirect correlation with other variables. For this reason, we create a standardized interaction term for education and mental wellbeing, relating to the interaction of self-perceived wellbeing and attained education. Due to the hedonic nature of the self-perceived wellbeing and the contribution of physical wellbeing to flourishing (A'yuninnisa, Carminati, and Wilderom, 2023), we create an interaction term named perceived Well-being (p = 0.35 with job flourishing). These two variables were moderately related with a correlation coefficient of 0.55. Similar interaction terms were created for mental wellbeing and role overload item 4, structural job resources composite (p = 0.51 with flourishing), structural job resource item 4 ("I decide on my own how I do things") (p = 0.4).

In line with the structure matrix values for component 1 of the factor analysis, we create two features that capture the indirect interactions within this component. Firstly, we interpret the supervisor relationship interaction (p = 0.32), - an interaction term of an employee's average LMX score, self-rated role clarity item "I know what my supervisor expects of me" and the average score of two social job resources items: "I develop my knowledge and professional skills" and "I try to learn new things at work".Secondly, we developed the feature Job Nature (p = 0.46 with job flourishing), which captures the interaction of "I have considerable opportunity for independence and freedom in how I do my work", "It is clear to me what I need to do in my job", and "I decide on my own how to do things". The interaction of these three items indicates the extent to which an employee knows their tasks and is able to take decisions on how to complete them, independently. Lastly, inspired by component 3 of the factor analysis, we created an interaction term for social job resources and crafting challenging job demands. All variables were distributed with a simingly normal distribution, with some skewness to the left for Supervisor relationship and JD x Social JR. LMX was slightly skewed to the right. For better interpretation of the results, we chose to not pursue further scaling or transformations of these variables. However, for validation of iteration 3 model, we scaled 5 of the features in order to normalize their skeweness. This is reported in Appendix D and iteration 3 Evaluation.

Table 4.3: Features used in iteration 3

Item Level Features (Categorical)	
- Structural JR 3	- Education
- Role Overload 4	- Gender
- Physical Wellbeing	- Family Status
- Mental Wellbeing	- Home Situation
- Age	- Tenure
- Job Title	- Contract Type
Average Scores Based on Correlation and Factor Analysis (Continuous)	
- Structural JR Composite	
- Org. Commitment Composite	- Perceived Wellbeing
- Role Clarity Composite	- Education x Wellbeing
- LMX Composite	- Overload x Wellbeing
- Autonomy Composite	- Structural JR x Wellbeing
- Self-Efficacy Composite	- Cabaility x Wellbeing
- Social JR Composite	- Supervisor Relations
- TSC Composite	- Job Nature
- Role Overload Composite	- JD x Social JR
- Challenging JD Composite	

Considering these three iterations, we display the results of the built models in the following section. Firstly through a description of the models, secondly per iteration, and thirdly through a cross-comparison, which we hope will provide insights for future decision-making in algorithmic management.

4.3 The Built Models

4.3.1 Training, Testing, and Validation

As mentioned in the methodology, training, validation and testing was done in a 80%, 10% (291 instances), 10% (290 instances) randomized split. Validation accuracy was recorded per model to check performance with a new dataset. 5-fold cross-validation was used in all the optimized models. This form of validation randomly divides the training dataset into 5 folds and evaluates the performance multiple times during training. The validation was done only the optimized models, and for the dafault settings we explored with no cross-validation to allow for a basic comparison.

Additionally, we validated the Iteration 3 models with the K-means labels of job flourishing (Appendix B). The highest obtained accuracy was 68% with random forests. For this interpretation of the WE scores into job flourishing, the models indicated difficulty in identifying non-flourishers, leading to high misclassifications. Hence, the classification of job flourishing is a significant determinator of the success of the model.

This indicates that organizations need to have a proper understanding and make decisions on which employee is classified as a flourisher prior to training the model. Soft boundaries lead to poor performance, emphasized by the misclassifications and low sensitivity for non-flourishing employees in the validation set. Additionally, this validation shows that our labeling based on the WE score is a closer representation of the hidden patterns in the data. This means that the study of Schotanus-dijkstra2016 (2016) is still valid for the 2022 employees in this Dutch academic institution.

4.3.2 Model Description and Optimization

We utilize sklearn_genetic and skopt library for BayesSearchCV and GASearchCV, grid searching algorithms based on BOA and GA and focused on hyperparameter tuning with cross-validation. As mentioned in the methodology we optimize based on ROC, to reduce potential bias from evaluating the models, but capturing a series of the criteria to compare. During iterations we performed one BOA and GA optimization with recall as the objective, leading to suboptimal results (60% accuracy).

In Appendix C we showcase the list of the final configurations of each optimization model per iteration. These models were run multiple times and trying multiple parameter grids to avoid being stuck at local optima, and only the highest results of these "trial and errors" were recorded. In the following section we highlight per iteration these results

4.3.3 First Iteration Evaluation Report

The models in this iteration are characterised by overall moderate to high accuracy, setting a benchmark for future models (lowest: 76.29%; highest: 80.41%). Overall, the models struggle with identifying flourishing employees (recall). The less than 0.7 recall values indicate that within the dataset there are instances that are not easy to distinguish if they are flourishing (recall (1) range is 0.59 - 0.67). As seen in figure 4.5, the random forest models are the most accurate. Genetic Algorithm performs the best at optimizing, and SVMs are the ones with the lowest overall performance. However, these complex algorithms function well in the validation set, implying the need for a larger dataset and more tuning. Regarding prediction of non-flourishers (0), the iteration has a low variation across models in how often models correctly predict non-flourishers (precision: 0.75 - 0.78). These moderately high values indicate a good prediction ability of non-flourishing employees). RF performs quite well at finding all instances of non-flourishers (recall: 0.93 - 0.95).

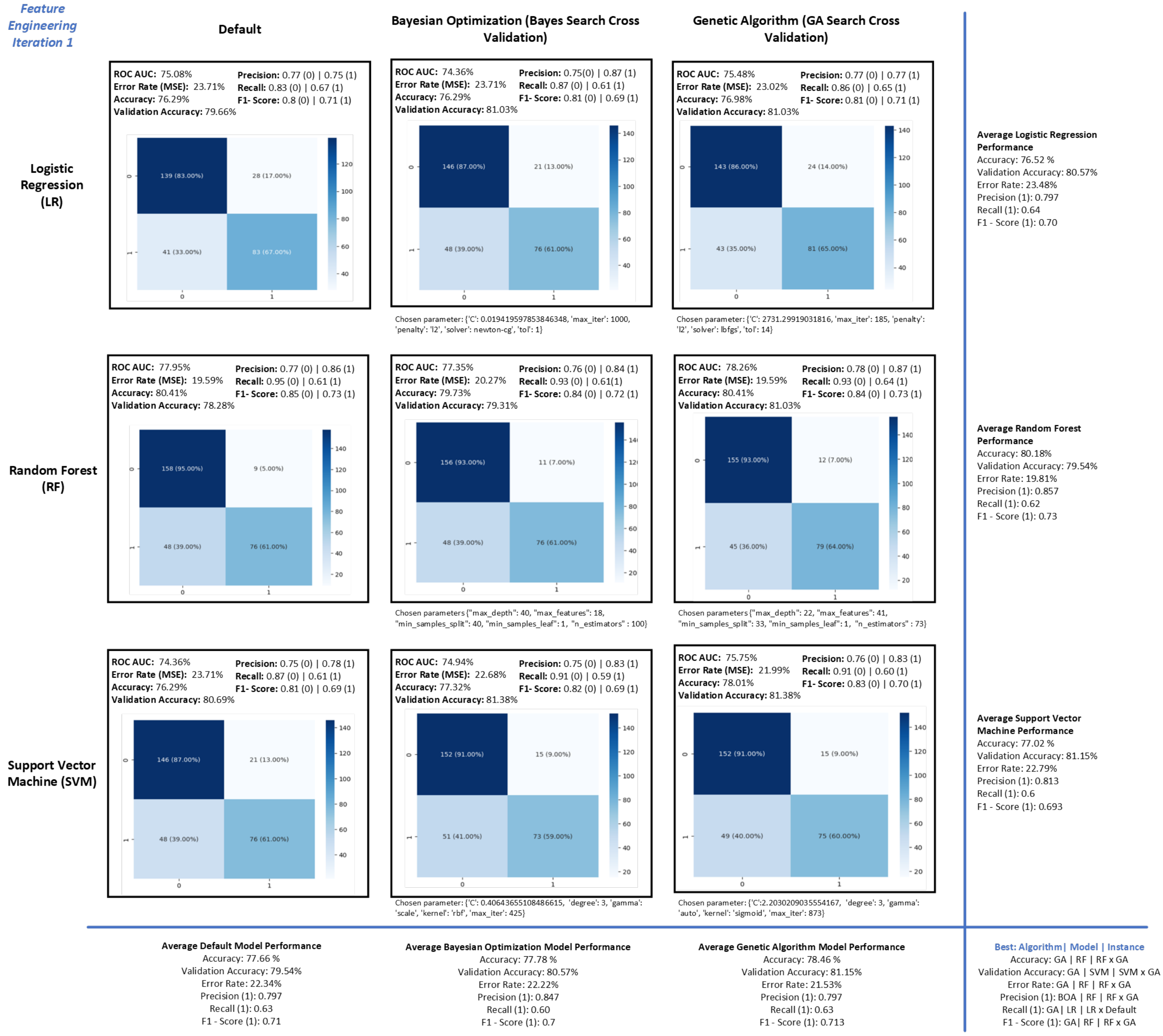


FIGURE 4.5: Iteration 1: Item Level Criteria Evaluation

4.3.4 Second Iteration Evaluation Report

Models are characterized by moderate accuracy (lowest: 76.98% ; highest: 79.38%). Recall for flourishing prediction takes on lower ranges than in the first iteration (0.53 - 0.67), and RF does not perform up to par. This is most likely due to loss in individual score information, which was used in RF node splitting. Overall models have a very high flourishing precision (0.76 - 0.88); they avoid classifying non-flourishers as flourishers. RF struggles the most with the precision/ recall trade-off, resulting in low F1 scores. Noteworthy is that SVM is the most consistent across optimizations and works best with new datasets, making it the best tool within this iteration if hyperparameter tuning is too costly. Interestingly, LR with GA optimization outperforms the other models, indicating that the features are better captured by linear relationships.

When it come to prediction of non-flourishing employees (0), this second iteration has the lowest variation for non-flourishing sensitivity(0.84 - 0.95). This means that with high order constructs, the models perform best at identifying non-flourishers. RFs perform similar to the item level iteration when it comes to recall. However, LR with GA is best at precision of non-flourishing employees (0.78).

Overall, in this iteration the models have a higher ability to predict non-flourishing employees.

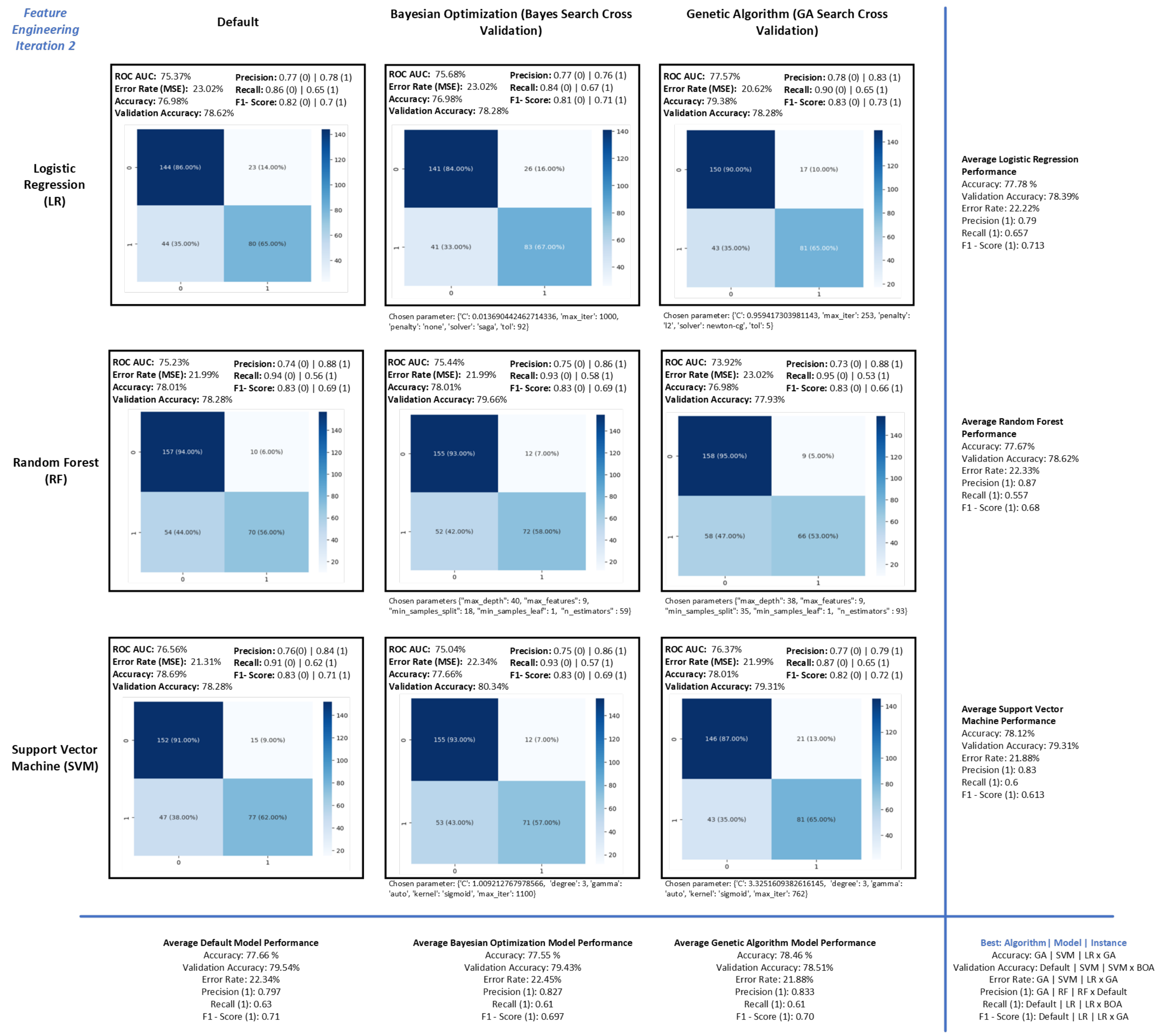


FIGURE 4.6: Iteration 2: High Order Constructs Criteria Evaluation

4.3.5 Third Iteration Evaluation Report

This iteration is characterized by high accuracy and high validation accuracy on all studied models. Important to note is the fact that LR with optimizations performs significantly better than other models, and it benefits significantly from the GA tuning. Meanwhile no improvements were found in RF after optimizations, and the highest accuracy model suffers significantly in recall and has a significantly lower validation accuracy (76.90%). Hence it suffers from data overfitting. In addition SVM struggles to find optimized parameters, making the algorithm challenging to implement when dealing with moderators. Regarding prediciton of non-flourishing employees (0), noteworthy is that LRs perform best for precision, while SVMs are best for sensitivity of predicting non-flourishing employees.

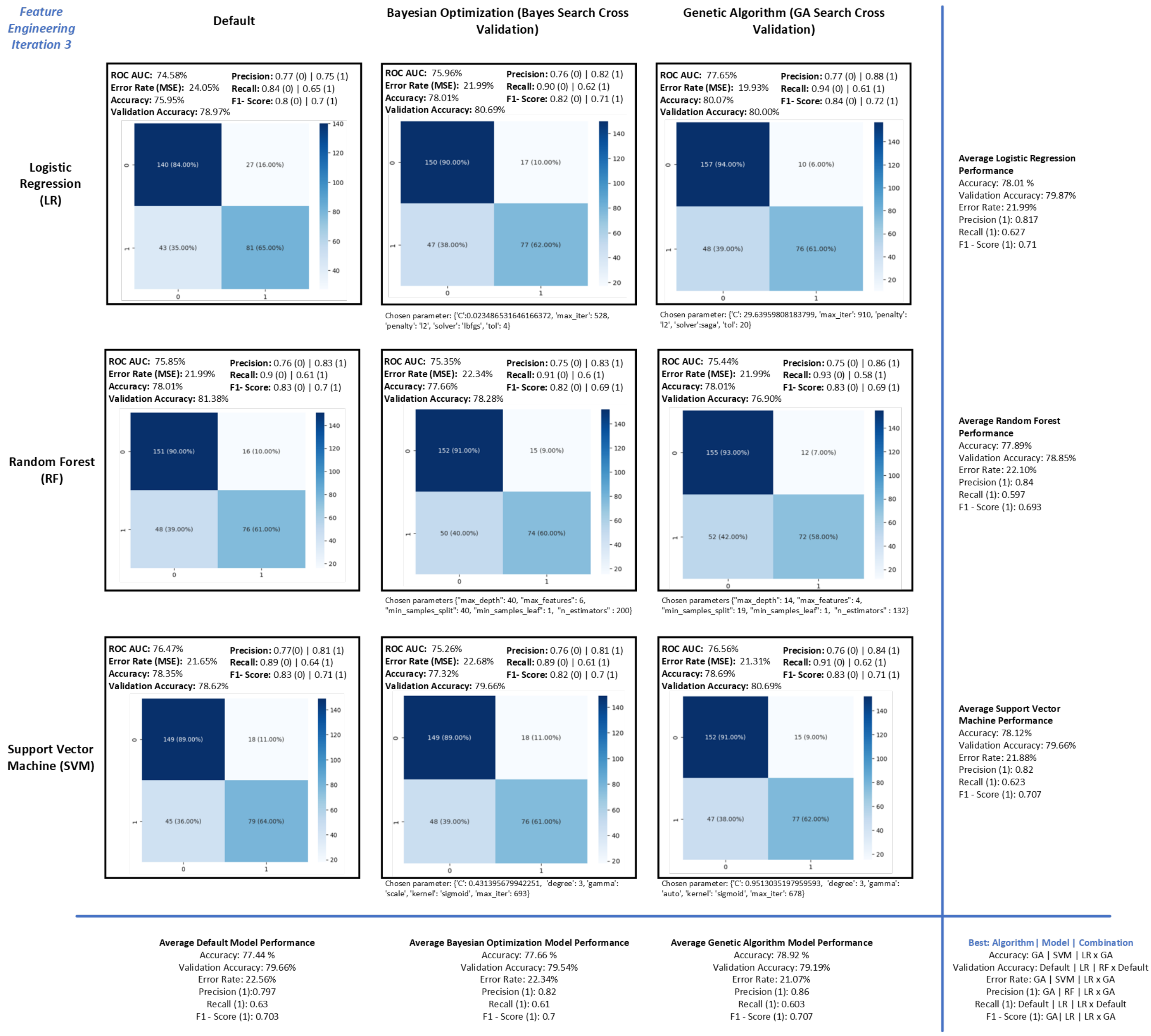


Figure 4.7: Iteration 3: Engineered Constructs and Moderators

For validation, we ran the optimizations with a few transformation based on the feature distributions (Appendix D). We squared the significantly left skewed features (LMX Composite score, Autonomy composite score, and Self-efficacy composite score) and took the logarithmic of the significantly right skewed ones (Supervisor Relations and Job Demands x Social Job Resources). This led to a fairly more symmetrical distribution of these features, which was utilized to obtain the results presented in figure 8.5. These results showcase the high performance of LR optimized with GA in iteration 3, with the highest recall achieved in the study (0.69) and a maintained high accuracy (80.07). These models however, suffer from a slightly lower overall precision (in range 0.78 - 0.85).

<table>
<tr><th>Feature Engineering with transformations in Iteration 3</th><th>Default</th><th>Bayesian Optimization (Bayes Search Cross Validation)</th><th>Genetic Algorithm (GA Search Cross Validation)</th></tr>
<tr><td>Logistic Regression (LR)</td><td>ROC AUC: 75.07%
Error Rate (MSE): 23.37%
Accuracy: 76.63%
Validation Accuracy: 78.97%
Precision: 0.76 (0) | 0.86(1)
Recall: 0.86 (0) | 0.65 (1)
F1- Score: 0.81 (0) | 0.7 (1)</td><td>ROC AUC: 77.06%
Error Rate (MSE): 20.96%
Accuracy: 79.04%
Validation Accuracy: 80.69%
Precision: 0.77 (0) | 0.83 (1)
Recall: 0.90 (0) | 0.64 (1)
F1- Score: 0.83 (0) | 0.72 (1)</td><td>ROC AUC: 78.69%
Error Rate (MSE): 19.93%
Accuracy: 80.07%
Validation Accuracy: 80.00%
Precision: 0.79 (0) | 0.81 (1)
Recall: 0.88 (0) | 0.69 (1)
F1- Score: 0.84 (0) | 0.75 (1)</td></tr>
<tr><td>Random Forest (RF)</td><td>ROC AUC: 76.15%
Error Rate (MSE): 21.65%
Accuracy: 78.35%
Validation Accuracy: 78.97%
Precision: 0.76 (0) | 0.84 (1)
Recall: 0.91 (0) | 0.61 (1)
F1- Score: 0.83 (0) | 0.71 (1)</td><td>ROC AUC: 75.54%
Error Rate (MSE): 21.99%
Accuracy: 78.01%
Validation Accuracy: 79.31%
Precision: 0.75 (0) | 0.85 (1)
Recall: 0.92 (0) | 0.59 (1)
F1- Score: 0.83 (0) | 0.7 (1)</td><td>ROC AUC: 75.35%
Error Rate (MSE): 21.99%
Accuracy: 77.66%
Validation Accuracy: 76.90%
Precision: 0.75 (0) | 0.83 (1)
Recall: 0.91 (0) | 0.60 (1)
F1- Score: 0.82 (0) | 0.69 (1)</td></tr>
<tr><td>Support Vector Machine (SVM)</td><td>ROC AUC: 76.37%
Error Rate (MSE): 21.99%
Accuracy: 78.01%
Validation Accuracy: 80.00%
Precision: 0.77(0) | 0.79 (1)
Recall: 0.89 (0) | 0.65 (1)
F1- Score: 0.82 (0) | 0.72 (1)</td><td>ROC AUC: 76.66%
Error Rate (MSE): 21.31%
Accuracy: 78.69%
Validation Accuracy: 80.34%
Precision: 0.77 (0) | 0.83 (1)
Recall: 0.83 (0) | 0.63 (1)
F1- Score: 0.82 (0) | 0.72 (1)</td><td>ROC AUC: 75.27%
Error Rate (MSE): 23.02%
Accuracy: 76.98%
Validation Accuracy: 79.66%
Precision: 0.76 (0) | 0.78 (1)
Recall: 0.87 (0) | 0.64 (1)
F1- Score: 0.81 (0) | 0.70 (1)</td></tr>
</table>

FIGURE 4.8: Iteration 3: Results with feature transformations

To enrich our analysis, we looked into the best model's performance (Iteration 3 with transformations LR x GA). The analysis and coefficients can be found in Appendix E. What is noteworthy to mention is the degree to which structural job resources influence the likelihood of flourishing, although these outcomes are hard to interpret due to scaling and transformation. Another interesting finding is the potential relative risk factor of males with permanent contracts and 1 - 5 years of working in the company, who have a 60% decrease in odds of being identified as flourishing by the model. From all the missclassified instances there were no instances that had this characteristics. Descriptive analysis revealed that out of the 120 instances with this profile (8.27% of the dataset), only 34 were flourishing. So only 28.3% of these individual flourish, which is significantly less than the taken baseline (Schotanus-dijkstra and Pieterse, 2016). Overall these instances were characterised by average scores across Autonomy items (4.13 average), moderate perceived mental well-being (3.73), and a moderately low sense of belonging to the organization (organizational commitment item 4 mean = 3.36). Moreover, gender, tenure and type of contract were found to have a significant association to flourishing.

Lastly, iteration 3 missclassification across different models were investigated. There were 35 similar instances that were classified as non-flourishing when we classified them as flourishing. These are the instances that caused the low recall, making them the most challenging to classify. Due to standardization and transformation, directly interpreting the continuous variables was challenging. However, the mean for these instances in the continuous variables was close to the mean of the population, indicating an average score of these employees. Noteworthy is that the model struggled with employees that scored overall average, but high (5-7) in structural job resources item 3 "I make sure that I use my capabilities to the fullest" and Role Overload item 4 score "Good" and "Too High".

4.4 Final Evaluation

Overall, noteworthy for this dataset and job flourishing prediction is the struggle to predict flourishers with high sensitivity (recall). The algorithms perform well in precision and are fairly accurate, mostly in identifying non-flourishers. It might be tempting to optimize further on recall, however, organizations need to decide themselves what is the cost of not identifying someone as flourishing, and if that is worth the risk of missclasifying non-flourishing employees. This links to deciding on a clear "performance" boundary within the flourishing continuum of when organizations should intervene. For a binary classification, the recall of 0.67 (highest) and precision of 0.89 (highest) can be set a benchmark for future models in predicting which employees have high well-being.

Regarding optimization, there are significant benefits in optimizing in item level features. These benefits deminish slightly for Random Forests and SVMs as we move on to higher order constructs and moderates. This is mainly due to the fact that these more complex models are challenging to tune and GA specifically is easily stuck in local optima. Additionally, as the models move to utilizing higher construct features there is a more visible linear relationship between the flourishing label and constructs such as structural job resources, which is better captured after optimization.

Lastly, our findings are mapped in the developed guideline presented in figure 4.9 for practitioners and future research in bridging well-being and algorithmic management in HRM. Here the best performing algorithms are recommended, depending if the data is in item level, higher constructs, or further engineered. The choice for the level of analysis should depend on the organization size and the willingness and time of participants to answer surveys, as well as the ability and time of the data scientists to engineer the presented features. Additionally, LR with GA optimization with engineered data and moderators is the best performing model for prioritizing the identification of job flourishers. This means that the model will perform best at prioritizing the correct identification of flourishing employees, with some degree of avoiding missed classifications. The opposite is true for the item level RF with GA optimization, which practitioners can develop for prioritizing the avoidance of missed classifications of flourishing, at the cost of not all employees that can be flourishing will be classified as such. Moreover, The LR with GA model performs well for identification of non-flourishing employees, but the features are engineered towards flourishing prediction. Hence, item level or construct level RF (with GA) is recommended for identification of non-job flourishers, but at the cost of not identifying all flourishers.

We note here that the guideline presented in figure 4.9 needs validation across similar companies and institutions, and across different cultures. This investigation was beyond the scope of this research, which aimed to set the foundational work towards such a project. The following chapter discusses these findings and recommendations and sets the stage for implications in research and practice.

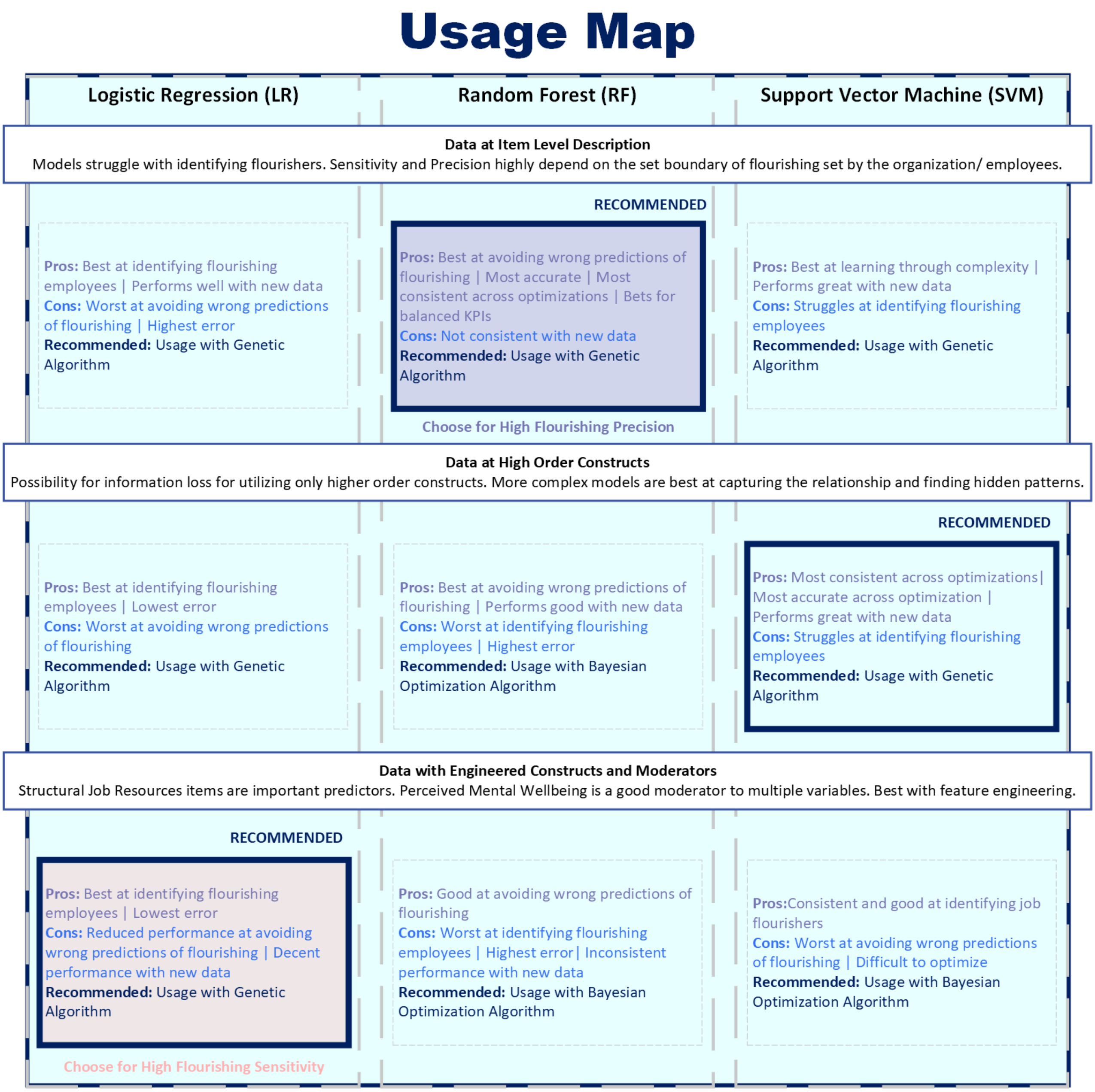


Figure 4.9: Usage Map: a guideline for practitioners and researchers to predict flourishing in the workplace

Chapter 5

Discussion

This research aimed to create the foundations for predictive AM in the field of job flourishing. This foundation helps in moving a step closer to full algorithmic predictions and moving away from descriptive analysis focused on relationships based on historical data (Deobald et al., 2019). In previous chapters, this thesis established the interest in casual discovery in the field of well-being and the need to move towards capturing these relationships in predictive models. Developing such models is useful to practitioners and organizations for tasks such as performance appraisals and recruitment (Sharma, 2021). Particularly, relating to Tambe, Cappelli, and Yakubovich (2019)'s identification of AI prediction tasks per HR operation, job flourishing prediction can contribute to: (1) Training by predicting if there should be an intervention to improve performance; (2) Performance Management by predicting if practices in HR have been improving job performance; (3) Retention by predicting likelihood of performance in new roles.

In this discussion, we firstly cover the research's theoretical contribution and discuss in terms of two research streams, job flourishing and AM. Here, we outline what the studied ML models relied upon and discuss the three levels of features. Moreover, we discuss how to link the presented usage map (figure 4.9) to organizational needs and AM implementations, as a guide to answer our main research question regarding best methods to predict job flourishing and highlight this research's practical contributions.

5.1 Theoretical Implications

We argue that the contribution of job flourishing to HR operations comes due to the relationship of flourishing with concepts such as work engagement, under the theoretical underpinning of JD-R. So, employees who have the right resources and challenges in their work, are more likely to display high work engagement and flourish. This leads to retention possibilities, and overall better performance in the organization. If they are not flourishing, organizations can intervene.

5.1.1 Regarding Features and Predictors

There are a few studies that support job crafting, job resources and work engagement as antecedents of job flourishing (see Ariza-Montes et al., 2018; Demerouti, Bakker, and Gevers, 2015; Imran et al., 2020). This thesis is based on the likelihood of flourishing of highly work engaged employees, particularly under the dimensions of vigor and dedication.

First up for discussion are the findings of labelling based on WE score as contrasted by the K-means classifications. Based on our validation (section 4.3.1), the WE score classification performs best when considering constructs such as autonomy, LMX, organizational commitment and subjective well-being, which have been linked to job flourishing. This is further supported by the similarity that the WE score distribution in this studied case has to the results in the Dutch population of Schotanus-dijkstra and Pieterse (2016). These results correspond to 36.5% flourishing employees, despite the overall skewness of the case population towards high values. As our studied population consisted primarily of employees born in the Netherlands, the Schotanus-dijkstra and Pieterse (2016)' results are validated within the Dutch population. However, such generalizations cannot be made for other countries due to limited literature found on country level work-place flourishing outside the Netherlands.

On the other hand, we have to highlight some potential contradictions. Schotanus-dijkstra and Pieterse (2016)'s study indicates that socio-demographics such as gender, age, and education least explained well-being. Our highest performing model (logistic regression model with genetic algorithm optimization) was able to predict the studied population of more than 1400 employees by giving high negative coefficients to some categories: tenure (1-5 years with the organization), contract type (permanent) and gender (males). As suggested by Peccei and Van De Voorde (2019), these findings could be credited on the organizational setting and HR performance, on which we have limited information.

Tenure is supported by Rothmann(2014) as an antecedent of job flourishing. The profile of males working on an organization for a sustained amount of time and with permanent contracts could be considered a relevant risk factor to job flourishing. Females of a similar profile were proportionally classified as flourishing, although tenure and contract type had a general significant negative relation to job flourishing prediction. This is a relative finding that needs further validation in the higher education industry and across countries. While, Padkapayeva et al. (2018) indicate that supervisor support is beneficial to reduce stress in women compared to men, and our dataset shows LMX is significant in measuring job flourishing, this socio-demographic risk cannot solely be attributed to gender differences in supervisor support. Hence, organizations should not dismiss the possibility that socio-demographics can have an impact on non-flourishing employees, preventing them from building upon the antecedents necessary to flourish in the workplace. With this finding, we encourage further research in job security (e.g. permanent contracts with a familiar work environment of 1 - 5 years) and gender differences, as research in the field has not been conducted and older research indicates variety in gendered job security perception based on age and year of the recorded data (Tolbert and Moen, 1998).

In addition to socio-demographic literature discrepancies, this research's findings showcase a lack of significant contribution of job crafting to predicting job flourishing. While Robledo, Zappala, and Topa (2019) attribute high engagement to job crafting behaviour, and these behaviour leading to high social-psychological well-being, the studied case did not in particularly showcase this mediation. Our results indicate support to the study of job characteristics, job environment, and positive mental health and their relation to job flourishing. Particularly of interest is the influence that high values in the structural job resource item 3 (I make sure that I use my capabilities to the fullest) and organizational commitment item 4 (I feel a strong sense of belonging to the organization) have on flourishing prediction.

Moreover, it was found that positive moderation of perceived mental (subjective) well-being on structural job resources was highly affecting flourishing prediction. Aligned with this positive relationship towards flourishing prediction was the developed variable Job Nature. This variable indicated the extent to which an employee knows their tasks and is able to take decisions on how to complete them, independently. Hence, structural resources in relation to subjective well-being are important features in prediction, but not necessary all aspect of job crafting as related to JD-R have a direct or indirect influence on predicting flourishing employees. Bakker and Demerouti (2007) emphasize that organizations should be attentive to job characteristics such as autonomy and workload when designing jobs. In addition to their emphasis, our study shows that more task oriented clarity and autonomy is needed for high likelihood of flourishing.

5.1.2 Regarding the Usage Map

Touching upon potential AM developments in job flourishing, we need to discuss the recommendations of this study as set in the Usage Map in figure (4.9). With these recommendations, we have answered the research call for algorithms that would be able to support performance measurement and training, at least on a foundational level (Garg, Sinha, and Kar, 2019; Tambe, Cappelli, and Yakubovich, 2019). Additionally, by studying complex unsupervised algorithms, we have contributed to Garg, Sinha, and Kar (2019)'s concern on under-investigated complex algorithms. However, our findings indicate that complex models such as SVMs are not as effective when measuring on an item level, due to their dependence on hyperparameter tuning. On the other hand, they perform well and learn fast in capturing underlying structures where other models like RFs struggle, particularly for high order constructs, without moderators.

Additionally, our foundational work highlights the importance of defining prediction tasks in organizations. These definitions are at the core of prediction criteria, and greatly impact the choice of optimization. Organizations could train algorithms by focusing on the prediction of non-flourishing employees. This could lead to the development of training interventions and related HR practices. Otherwise, organizations could prioritise predicting flourishing employees to understand retention rates and build profiles for effective recruitment. Nevertheless, as Tambe, Cappelli, and Yakubovich (2019) argue, algorithms should not focus on either one or the other, but examine both groups to build stronger predictions. As our most optimal model only achieved 80% accuracy it can be concluded that prediction of well-being should include input features that directly measure lack of well-being (or languishing). There are only few found studied variables that are negative antecedents to job flourishing. For example, Robledo, Zappala, and Topa (2019) identified emotional exhaustion as a possible negative variable through the JD-R model. Hence, theoretical identification of both antecedents and detractors is important for a complete prediction in job-flourishing to be used within AM. Utilizing solely positive descriptors of job flourishing can lead to sub optimal algorithms, having ethical implications when applied in organizational settings due to their reliability, which Tambe, Cappelli, and Yakubovich (2019) highlight as one of the biggest challenges in HRM ML for small datasets.

5.2 Practical Implications

Evident from this study is the need to find the "minimal amount" of features needed to predict flourishers. Focusing on identifying employees who have the most likelihood to flourish, positive antecedents were used. Diedericks and Rothmann (2013) found that social exchanges with supervisors impact work engagement and job satisfaction of employees, and subsequently flourishing. While the variance in the dataset was majorly captured by the relation to supervisors and the nature of the job, other variables, including socio-demographics, were important for high prediction results. Moreover, higher level constructs were not directly contributing to high prediction without moderators as much as item level information. These findings showcase the importance of item level self-reported measures, which should further be explored for personal context and cognitive variables. This way, researchers and practitioners developing well-being tools can move towards minimal attainable information which organizations can record and use for algorithm developments in job flourishing, such as through deep learning or language process modelling. Additionally, we have showcased how researchers in job flourishing can utilize a data science methodology to evaluate their studied antecedents, looking at performance. Prediction and the measurement of job flourishing can be significantly different and need to be researched hand-in-hand to develop correct features and apply it in organizations.

Lastly, we have set the stage to ponder upon the purpose of such algorithms when it comes to well-being. By being able to know the most relevant tools for their purpose, practitioners are able to choose. However, should practitioners opt for a decision making algorithm that is able to correctly identify all flourishing employees or one that aims at avoiding wrong predictions in this class? Is the cost of wrongly predicting someone as flourishing higher than the cost of not identifying someone as a flourisher? These are questions regarding aims within AM that need to be well-defined prior to implementation of algorithms. These questions have been well answered in the field of recruitment, but are still a mystery in training and appraisal (Garg, Sinha, and Kar, 2019). This thesis aimed to propose and to create a basis for job flourishing prediction as a route for training and appraisal algorithms. From here, data science practitioners and AM researchers can look upon the fairness and ethical constraints that come with predicting well-being by evaluating minimal information needed and related item level information, as supported by Deobald et al. (2019). Hence, with this research we have contributed with a foundation to the development of flourishing prediction algorithms. It is now imperative that organizations and data scientists ask the right questions to extend ethically the prediction of well-being towards employee retention and job satisfaction.

Chapter 6

Limitations and Future Research

It is important to note that this is a first step into a novel approach to job flourishing, as it attempts to create a foundation for further predictions and ML applications. This leads to a variety of limitations in the choices of the research design and the data collection.

6.1 Utilizing Work Engagement

The data collected utilizes only the Vigor and Dedication components of the Utrecht Work Engagement Scale. As absorption data is not collected, the work engagement score is incomplete, which might have changed the labeling process and the pattern recognition of the model. Additionally, utilizing work engagement as a narrow concept of flourishing limits the understanding of the prediction results to other dimensions of job flourishing (that include hedonic and eudemonic wellbeing). This limitation particularly holds as Self-Perceived Mental Well-being has a Pearson correlation of 0.157 with the calculated Work Engagement Score.

6.2 Sample Size and Data Preparation

The sample size is characterized by an imbalanced category but not considered "rare" cases. Although beyond the scope of this thesis, an imbalanced data set implies challenges in guaranteeing generalization of the techniques for more than binary classification. Additionally, it makes validating with new data challenging due to the small size of the validating set. This is emphasized by the HR and organizational context of the data, which could have been a major influence on the underlying relationships. Hence, similar studies should be replicated across different industries and cultures, extending Schotanus-dijkstra and Pieterse (2016)'s findings in the Netherlands. Particularly, as tenure and contract type played a big role on the flourishing classification in this dataset, it would be interesting if further studies investigated job characteristics, job resources, and organizational commitment in cooperation with contract type and tenure.

6.3 Techniques and Algorithms

In this thesis, we choose to work with supervised learning techniques, as a result of the vast casual analysis which can be used from the domain knowledge. This choice leaves room for exploring further prediction techniques in areas such as Artificial Neural Networks and Deep Learning. Although more complex and focused on computerized patterns, these techniques might be helpful to advance job flourishing predictions and recognize patterns outside of validated and researched scales. Similarly, diving into other hyperparameter tuning techniques, such as grid search and randomized search might yield different results.

Moreover, the choice of optimization through ROC (AUC) could have been significant for the performance on recall and precision. Hence, for furthering recommendation on which criteria to pursue, future studies could focus on specific parameter tuning per criteria. Additionally, for future implemented predicting tools Genetic algorithms and Bayesian optimization algorithms need to be developed that have tailored parameter grids to the organizational structure. In this study we focused mainly on multiple exploratory testing. These types of algorithms are particularly interesting for SVMs with soft boundaries, which can allow for better performance on recall.

Lastly, the higher order composites were made by averaging the items, in line with the scales. However, investigating the influence of individual item level components on job flourishing can result in tailored features with weighted average. For example, structure job resources item 3 had a high influence on flourishing, hence, an increase in weight for structural job resource composite could have resulted in less information loss for the high order construct level. The exploration of these item level influences of individual level work-context variables and emerging states could be extremely beneficial to the field of job flourishing and its AM prediction.

Chapter 7

Conclusions

In conclusion, the best method for job flourishing prediction is dependent on the level of the feature (item or statement level, high order constructs, or engineered with moderators). Overall the models struggled with identifying flourishing employees and performed well at avoiding 'false' flourishers. Looking within the data structure a strong relation of structural job resources to job flourishing was found. This was particulalry evident when it comes to employees utilizing their capability to the fullest. Moreover, supervisor relations and job characteristics (nature) had a large impact on the variance of the employees. Meanwhile, some socio-demographics like tenure, gender, and contract type, linked to job security, had a negative influence on the odds of flourishing, which can be attributed to the organizational context, but needs further validation.

While this research contributed to the research calls of Garg, Sinha, and Kar(2019), Tambe, Cappelli, and Yakubovich (2019), and A'yuninnisa, Carminati, and Wilderom (2023), further exploration is necessary for result validation and generalization. We encourage future practitioners to look into the development of complex, fair and ethical algorithms for well-being predictions, particularly when it comes to the purpose of developing prediction tools in this complex domain. Additionally, the study contributes sufficiently for a foundational work towards job flourishing prediction in AM, and finds it imperative to deepen the investigation on the relationship between antecedents' individual items and job flourishing, as well as the identification of job flourishing detractors.

Chapter 8

Appendix

8.1 Appendix A: Factor Analysis

Description: After rotation, 12 factors were loaded, which explain about 71% of the variance in the dataset. Component 1 explains about 19.8% of the variance, indicating its importance. The pattern matrix showcase high loadings for items with high Cronbach's alpha, particularly LMX, Autonomy, Organizational Commitment, Self-Efficacy, Team Social Cohesion, and Role Clarity. Other constructs which had a slightly lower Cronbach's alpha correspond to cross-loadings in various components. The structure matrix reveals the indirect relationships of Mental Wellbeing across multiple components, making it a significant mdoerator.

Total Variance Explained

Component	Initial Eigenvalues Total	Initial Eigenvalues % of Variance	Initial Eigenvalues Cumulative %	Rotation Sums of Squared Loadings[a] Total
1	9.105	**19.794**	19.794	6.101
2	4.170	9.064	28.859	2.691
3	3.099	6.737	35.596	2.754
4	2.714	5.899	41.495	3.507
5	2.632	5.722	47.217	3.550
6	2.222	4.831	52.048	4.639
7	2.041	4.438	56.486	3.297
8	1.634	3.552	60.038	4.186
9	1.552	3.374	63.412	3.478
10	1.244	2.703	66.115	3.921
11	1.220	2.653	68.769	4.733
12	1.037	2.253	71.022	1.491
13	.895	1.946	72.968	
14	.710	1.543	74.511	
15	.669	1.454	75.965	
16	.656	1.426	77.390	
17	.624	1.358	78.748	
18	.601	1.306	80.054	
19	.551	1.197	81.251	
20	.525	1.141	82.392	
21	.517	1.125	83.517	
22	.479	1.041	84.558	
23	.467	1.014	85.572	
24	.445	.968	86.540	
25	.431	.937	87.476	
26	.421	.915	88.391	
27	.397	.864	89.255	
28	.380	.827	90.082	
29	.375	.815	90.896	
30	.363	.789	91.685	
31	.353	.768	92.453	
32	.338	.734	93.187	
33	.299	.650	93.838	
34	.298	.647	94.485	
35	.283	.615	95.100	
36	.278	.605	95.705	
37	.262	.569	96.274	
38	.238	.518	96.792	
39	.232	.504	97.297	
40	.224	.488	97.784	
41	.203	.440	98.225	
42	.186	.405	98.630	
43	.168	.366	98.996	
44	.161	.349	99.345	
45	.156	.338	99.683	
46	.146	.317	100.000	

Extraction Method: Principal Component Analysis.

a. When components are correlated, sums of squared loadings cannot be added to obtain a total variance.

FIGURE 8.1: Factor Analysis: Components and their explainable variance on the dataset

And more weird text, Maybe I should add more text? I am not sure how to do this

Pattern Matrix[a,b]

	Component											
	1	2	3	4	5	6	7	8	9	10	11	12
Structural_JR1	.014	-.048	.038	.006	-.008	.036	-.022	.876	-.005	-.009	-.014	.016
Structural_JR2	.026	-.054	.050	.005	.026	.020	-.052	.886	-.018	.017	-.018	.005
Structural_JR3	.075	.049	-.070	-.010	.032	.070	.121	.724	-.022	-.120	-.058	-.061
Structural_JR4	.013	.060	-.013	-.041	-.020	.764	-.008	.188	.048	.016	.002	-.022
Social_JR1	.306	-.387	.434	-.006	-.040	-.036	-.137	.102	.051	.031	-.047	-.062
Social_JR2	.202	-.353	.649	-.006	-.048	-.038	-.025	-.046	.101	.009	-.069	-.084
Social_JR3	-.114	-.114	.809	-.017	-.034	.000	-.028	.014	-.013	-.101	-.057	-.061
Social_JR4	-.089	.027	.641	-.053	-.083	-.042	-.099	.154	-.258	-.020	.040	-.041
Challenging_JD1	.031	.263	.587	.037	.111	.142	.149	.072	.098	.141	-.070	.205
Challenging_JD2	.003	.416	.416	-.017	.300	-.021	.213	-.018	-.015	.086	-.056	.171
RoleOverload1	-.006	-.099	-.002	.034	.784	-.022	-.005	-.004	.068	.044	.020	-.081
RoleOverload2	-.052	.027	-.027	.025	.792	.040	-.110	-.012	-.012	-.060	-.002	.038
RoleOverload3	-.077	-.052	-.018	.020	.813	-.018	-.051	.003	.071	-.058	-.056	.057
RoleClarity_1	-.049	.009	.039	.006	.053	.031	-.006	.010	-.035	-.898	-.033	.045
RoleClarity_2	.392	.003	.062	-.028	.019	-.010	.046	-.032	.021	-.668	-.014	-.048
RoleClarity_3	.008	.014	-.015	-.005	.042	.063	.087	.104	-.001	-.819	.003	.003
LMX1	.726	.104	.061	.006	-.009	.001	.130	-.102	-.016	-.126	-.012	-.020
LMX2	.794	-.038	-.002	-.021	-.065	-.050	.051	.052	.038	-.114	-.044	-.031
LMX3	.839	.052	-.061	-.034	.015	.054	.006	.039	-.045	.004	.008	.015
LMX4	.794	.024	-.047	-.016	-.001	.013	-.073	.112	-.039	.112	-.045	.008
LMX5	.819	-.043	-.015	-.008	-.032	.040	-.072	.053	-.045	-.012	-.013	.046
LMX6	.798	-.008	-.028	-.007	-.015	.097	-.062	-.009	-.063	-.050	.020	.030
TSC1	.063	.047	.065	.002	-.044	.006	.016	.011	-.725	-.039	-.154	.028
TSC2	.023	-.007	-.007	.017	.011	.063	.035	.009	-.885	.006	-.002	-.005
TSC3	.062	-.136	-.027	-.017	.037	.030	.038	.010	-.850	.002	-.057	-.013
Autonomy1	.003	.014	.017	.008	-.012	.900	.009	-.022	-.049	-.040	.010	-.020
Autonomy2	-.028	-.027	.024	-.006	.025	.937	.032	-.056	-.025	.002	-.004	.001
Autonomy3	.065	-.048	-.024	.016	-.025	.820	-.007	-.012	-.038	-.027	-.031	.004
Selfefficacy1	-.008	-.133	-.027	.019	-.051	-.003	.918	.035	-.031	-.034	-.001	.006
Selfefficacy2	-.024	-.096	-.011	-.010	-.063	-.025	.936	.012	-.018	-.020	-.015	.004
Selfefficacy3	-.001	.070	-.024	-.008	-.001	.083	.790	-.028	-.027	-.031	.002	-.066
OrgCommitment1	.025	-.070	.019	-.018	.054	-.004	.045	.084	.081	.045	-.745	-.018
OrgCommitment2	.008	-.014	-.008	.007	-.002	-.008	-.014	-.033	-.077	-.015	-.883	.023
OrgCommitment3	-.016	.077	-.020	-.009	-.016	.031	-.031	-.025	-.080	-.011	-.857	-.023
OrgCommitment4	-.022	.033	-.005	-.008	-.023	.009	-.016	-.008	-.081	-.040	-.868	.012
RoleOverload4	.002	.262	.027	.036	.611	-.136	.091	.051	-.137	.052	.094	.017
PhysicalWellbeing	.004	.093	.040	.004	-.413	.023	-.036	.071	-.054	-.071	-.021	.631
MentalWellbeing	.046	.140	.009	.005	-.515	-.003	.113	.132	-.006	-.143	-.097	.467
Age	-.011	.691	-.160	-.004	-.013	-.059	.038	.025	.074	-.055	-.157	-.061
Education	.046	-.325	-.072	.004	.305	.035	-.008	-.043	.009	.063	.005	.626
Job_Title	.005	.431	.104	.431	-.163	-.019	.051	-.067	.003	.034	-.065	-.309
Gender	-.020	-.087	-.015	.819	.064	-.015	-.013	-.034	-.048	-.001	-.011	.021
FamilyStatus	.027	-.083	.010	.904	.007	.013	.008	-.001	.009	.048	.039	.023
HomeSituation	.019	-.090	-.004	.872	-.020	-.007	.028	.068	.034	.004	.021	.002
Tenure	-.072	.178	-.042	.724	.031	-.017	-.041	-.018	.002	-.061	-.007	.037
ContractType	-.056	-.714	.011	-.027	.001	-.075	.024	.196	.023	.035	-.021	.050

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 17 iterations.

b. Loadings legend: Low (white) < 0.3; Moderate (yellow) 0.3 - 0.5; High (orange) 0.5 - 0.7; Very High (red) > 0.7

Figure 8.2: Factor Analysis: Pattern Matrix indicating factor loadings

Structure Matrix[a]

	Component 1	2	3	4	5	6	7	8	9	10	11	12
Structural_JR1	.264	-.166	.227	-.150	-.112	.281	.022	.911	-.182	-.162	-.280	.160
Structural_JR2	.263	-.183	.244	-.151	-.075	.258	-.017	.916	-.186	-.126	-.276	.148
Structural_JR3	.315	-.002	.095	-.129	-.102	.341	.210	.776	-.227	-.324	-.339	.078
Structural_JR4	.263	.037	.018	-.156	-.118	.808	.125	.380	-.153	-.186	-.189	.099
Social_JR1	.416	-.494	.512	-.141	-.096	.036	-.227	.298	-.013	-.009	-.132	-.032
Social_JR2	.302	-.412	.683	-.092	-.073	-.021	-.105	.167	.038	.016	-.135	-.065
Social_JR3	.055	-.145	.807	-.051	-.042	-.004	.002	.200	-.092	-.027	-.185	-.029
Social_JR4	.070	-.038	.671	-.092	-.099	.019	-.048	.296	-.308	-.003	-.166	.002
Challenging_JD1	.068	.254	.616	.041	.118	.168	.257	.208	-.032	.104	-.219	.259
Challenging_JD2	-.059	.447	.422	.068	.315	.004	.335	.023	-.085	.091	-.165	.202
RoleOverload1	-.175	-.099	-.005	.102	.804	-.150	-.070	-.128	.157	.229	.168	-.072
RoleOverload2	-.169	.017	-.040	.080	.790	-.059	-.089	-.100	.048	.126	.103	.045
RoleOverload3	-.199	-.048	-.021	.077	.819	-.119	-.058	-.091	.127	.143	.085	.067
RoleClarity_1	.260	.059	-.012	-.068	-.122	.228	.221	.177	-.206	-.890	-.244	.057
RoleClarity_2	.604	.000	.062	-.127	-.178	.239	.196	.184	-.161	-.792	-.256	-.034
RoleClarity_3	.306	.062	-.041	-.084	-.138	.282	.291	.253	-.185	-.865	-.234	.039
LMX1	.744	.076	.121	-.077	-.151	.233	.192	.112	-.162	-.381	-.235	-.006
LMX2	.853	-.116	.105	-.159	-.230	.226	.068	.275	-.125	-.395	-.266	-.003
LMX3	.857	-.044	.049	-.168	-.143	.315	.036	.258	-.191	-.298	-.226	.050
LMX4	.800	-.095	.086	-.155	-.138	.251	-.069	.305	-.171	-.171	-.241	.043
LMX5	.868	-.155	.101	-.170	-.189	.298	-.054	.295	-.190	-.294	-.237	.076
LMX6	.846	-.106	.063	-.156	-.173	.336	-.026	.233	-.203	-.324	-.206	.056
TSC1	.241	.108	.161	-.058	-.134	.231	.149	.231	-.806	-.235	-.428	.080
TSC2	.175	.063	.076	-.035	-.062	.261	.137	.191	-.904	-.178	-.289	.052
TSC3	.226	-.069	.075	-.090	-.048	.243	.105	.217	-.875	-.187	-.320	.049
Autonomy1	.270	.041	.009	-.105	-.113	.912	.170	.234	-.244	-.244	-.185	.088
Autonomy2	.229	.009	.012	-.114	-.063	.922	.177	.204	-.214	-.189	-.171	.113
Autonomy3	.320	-.029	-.012	-.112	-.129	.854	.130	.249	-.231	-.240	-.208	.105
Selfefficacy1	.041	.109	.024	.059	-.073	.168	.897	.102	-.131	-.255	-.171	.094
Selfefficacy2	.015	.150	.034	.045	-.079	.141	.917	.074	-.116	-.239	-.177	.090
Selfefficacy3	.026	.284	-.005	.062	-.026	.210	.822	.018	-.128	-.234	-.158	.011
OrgCommitment1	.201	.000	.171	-.077	-.034	.141	.138	.291	-.152	-.124	-.737	.045
OrgCommitment2	.228	.090	.147	-.056	-.110	.173	.147	.241	-.337	-.218	-.896	.072
OrgCommitment3	.202	.176	.123	-.055	-.123	.199	.150	.227	-.342	-.216	-.880	.027
OrgCommitment4	.215	.135	.145	-.066	-.134	.195	.163	.261	-.351	-.245	-.899	.065
RoleOverload4	-.178	.278	.028	.143	.633	-.168	.122	-.088	-.065	.160	.098	.024
PhysicalWellbeing	.145	.067	.071	-.104	-.428	.189	.085	.245	-.160	-.179	-.176	.635
MentalWellbeing	.235	.153	.061	-.098	-.566	.227	.256	.314	-.175	-.335	-.305	.490
Age	-.072	.722	-.178	.108	-.032	-.016	.228	-.081	-.008	-.118	-.193	-.077
Education	.012	-.339	-.033	-.080	.326	.057	-.055	.056	.034	.120	.064	.635
Job_Title	-.094	.524	.060	.526	-.127	-.090	.172	-.187	-.013	.013	-.082	-.366
Gender	-.151	.046	-.034	.815	.134	-.129	.032	-.153	.014	.089	.066	-.062
FamilyStatus	-.125	.047	-.009	.891	.089	-.112	.043	-.135	.070	.124	.107	-.063
HomeSituation	-.099	.035	-.010	.850	.045	-.100	.068	-.062	.072	.063	.069	-.071
Tenure	-.202	.293	-.086	.758	.095	-.121	.074	-.171	.039	.031	.048	-.048
ContractType	.049	-.735	.092	-.149	.000	-.047	-.168	.264	.053	.063	.033	.092

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.
a. Correlation legend: Low (white) < 0.3; Moderate (yellow) 0.3 - 0.5; High (orange) 0.5 - 0.7; Very High (red) > 0.7

FIGURE 8.3: Factor Analysis: Structure Matrix indicating indirect and direct relationships

8.2 Appendix B: K-means clustering information

Final Cluster Centers

	Cluster Languish	Flourish	Moderate
WE_Vigor1	3	6	5
WE_Vigor2	3	6	5
WE_Vigor3	3	6	5
WE_Dedication1	3	6	6
WE_Dedication2	3	6	6
WE_Dedication3	3	6	5

Distances between Final Cluster Centers

Cluster	Languish	Flourish	Moderate
1		7.217	5.861
2	7.217		1.887
3	5.861	1.887	

Number of Cases in each Cluster

Cluster	Languish	73.000
	Flourish	725.000
	Moderate	653.000
Valid		1451.000
Missing		.000

FIGURE 8.4: Centroids and clustering information for validation

Description: The K-means algorithm identifies 725 individuals as flourishing and 726 individuals as not flourishing. The centroid for flourishing individuals is at WE average score 6.00, while that for languishing is at WE average score 3.00. The average scoring for moderate wellbeing is 5.33. The distance between clusters indicates a significantly lower distance for mdoerate wellbeing and flourishing, corresponding to a skewed to higher values of WE score for the population.

8.3 Appendix C: Configurations of GA and BOA per iteration

Table 8.1: Configurations for Iteration 1

		Iteration 1
GA	*RF*	10 generations Population size = 50 Grid: "max_depth": Integer (1, 40), "max_features": Integer (1, 47), "min_samples_split": Integer (2, 40), "min_samples_leaf": Integer (1, 40), "n_estimators" : Integer (1, 100) 10-fold Cross Validation Parallel jobs: 4 Scoring: roc_auc
	LR	10 generation Population size = 50 Grid: "C": Real(1e-4, 1e+4, prior=’log-uniform’), "solver": ["newton-cg", "lbfgs", "liblinear", "sag", "saga"], "max_iter": Integer (100, 1000), "tol": Integer (1, 100), Second Grid: ”penalty": ["l1","l2", "none", "elasticnet"] 5-fold Cross Validation Parallel jobs: 4 Scoring: roc_auc
	SVM	10 generations Population size = 50 Grid: "C": Real(1e-2, 1e+2, prior=’log-uniform’), "kernel": Categorical(["poly", "rbf", "sigmoid"]), "gamma": Categorical(["scale", "auto"]), "max_iter": Integer (100, 2000), Second Grid: ”degree": Integer (1,5) 5-fold Cross Validation Parallel jobs: 4 Scoring: roc_auc
BOA	*RF*	50 iterations Grid: "max_depth": Integer (1, 40), "max_features": Integer (1, 47), "min_samples_split": Integer (2, 40), "min_samples_leaf": Integer (1, 40), "n_estimators" : Integer (1, 100) 10-fold Cross Validation Parallel jobs: 4 Scoring: roc_auc
	LR	50 iterations Grid: "C": Real(1e-6, 1e+6, prior=’log-uniform’), "solver": ["newton-cg", "lbfgs", "liblinear", "sag", "saga"], "max_iter": Integer (500, 1000), "tol": Integer (1, 100), Second Grid: ”penalty": ["l1","l2", "none", "elasticnet"] 5-fold Cross Validation Parallel jobs: 4 Scoring: roc_auc
	SVM	50 iterations Grid: "C": Real(1e-6, 1e+6, prior=’log-uniform’), "kernel": Categorical(["poly", "rbf", "sigmoid"]), "gamma": Categorical(["scale", "auto"]), "max_iter": Integer (100, 1000), Second Grid: ”degree": Integer (1,5) 5-fold Cross Validation Parallel jobs: 4 Scoring: roc_auc

Table 8.2: Configurations for Iteration 2

		Iteration 2
GA	*RF*	10 generations Population size = 100 Grid: "max_depth": Integer (1, 40), "max_features": Integer (1, 50), "min_samples_split": Integer (2, 40), "min_samples_leaf": Integer (1, 20), "n_estimators" : Integer (50, 120) 5-fold Cross Validation Parallel jobs: 4 Scoring: roc_auc Mutation Probability = 0.1 Crossover Probability = 0.9
	LR	20 generation Population size = 100 Grid: "C": Real(1e-2, 1e+2, prior=’log-uniform’), "solver": ["newton-cg", "lbfgs", "liblinear", "sag", "saga"], "max_iter": Integer (100, 1000), "tol": Integer (1, 100), Second Grid: ”penalty": ["l1","l2", "none", "elasticnet"] 5-fold Cross Validation Parallel jobs: 4 Scoring: roc_auc Mutation Probability = 0.1 Crossover probability = 0.9
	SVM	15 generations Population size = 100 Grid: "C": Real(1e-3, 1e+3, prior=’log-uniform’), "kernel": Categorical(["poly", "rbf", "sigmoid"]), "gamma": Categorical(["scale", "auto"]), "max_iter": Integer (300, 1000), Second Grid: ”degree": Integer (1,5) 7-fold Cross Validation Parallel jobs: 4 Scoring: roc_auc Mutation probability = 0.1 Crossover probability = 0.9
BOA	*RF*	50 iterations Grid: "max_depth": Integer (1, 40), "max_features": Integer (1, 50), "min_samples_split": Integer (2, 40), "min_samples_leaf": Integer (1, 40), "n_estimators" : Integer (1, 110) 5-fold Cross Validation Parallel jobs: 4 Scoring: roc_auc
	LR	50 iterations Grid: "C": Real(1e-4, 1e+4, prior=’log-uniform’), "solver": ["newton-cg", "lbfgs", "liblinear", "sag", "saga"], "max_iter": Integer (500, 1000), "tol": Integer (1, 100), Second Grid: ”penalty": ["l1","l2", "none", "elasticnet"] 5-fold Cross Validation Parallel jobs: 4 Scoring: roc_auc
	SVM	100 iterations Grid: "C": Real(1e-4, 1e+4, prior=’log-uniform’), "kernel": Categorical(["poly", "rbf", "sigmoid"]), "gamma": Categorical(["scale", "auto"]), "max_iter": Integer (100, 1500), Second Grid: ”degree": Integer (1,5) 5-fold Cross Validation Parallel jobs: 5 Scoring: roc_auc

Table 8.3: Configurations for Iteration 3

		Iteration 2
GA	*RF*	12 generations Population size = 50 Grid: "max_depth": Integer (1, 40), "max_features": Integer (1, 20), "min_samples_split": Integer (2, 40), "min_samples_leaf": Integer (1, 40), "n_estimators" : Integer (1, 200) 5-fold Cross Validation Parallel jobs: 5 Scoring: roc_auc Mutation Probability = 0.15 Crossover Probability = 0.85
	LR	20 generation Population size = 50 Grid: "C": Real(1e-2, 1e+2, prior='log-uniform'), "solver": ["newton-cg", "lbfgs", "liblinear", "sag", "saga"], "max_iter": Integer (100, 1000), "tol": Integer (1, 100), Second Grid: ”penalty": ["l1","l2", "none", "elasticnet"] 5-fold Cross Validation Parallel jobs: 5 Scoring: roc_auc Mutation Probability = 0.15 Crossover probability = 0.85
	SVM	20 generations Population size = 50 Grid: "C": Real(1e-2, 1e+2, prior='log-uniform'), "kernel": Categorical(["poly", "rbf", "sigmoid"]), "gamma": Categorical(["scale", "auto"]), "max_iter": Integer (100, 1000), Second Grid: ”degree": Integer (1,5) 75fold Cross Validation Parallel jobs: 5 Scoring: roc_auc Mutation probability = 0.15 Crossover probability = 0.85
BOA	*RF*	50 iterations Grid: "max_depth": Integer (1, 40), "max_features": Integer (1, 20), "min_samples_split": Integer (2, 40), "min_samples_leaf": Integer (1, 40), "n_estimators" : Integer (1, 200) 5-fold Cross Validation Parallel jobs: 4 Scoring: roc_auc
	LR	50 iterations Grid: "C": Real(1e-4, 1e+4, prior='log-uniform'), "solver": ["newton-cg", "lbfgs", "liblinear", "sag", "saga"], "max_iter": Integer (500, 1000), "tol": Integer (1, 100), Second Grid: ”penalty": ["l1","l2", "none", "elasticnet"] 5-fold Cross Validation Parallel jobs: 4 Scoring: roc_auc
	SVM	100 iterations Grid: "C": Real(1e-3, 1e+3, prior='log-uniform'), "kernel": Categorical(["poly", "rbf", "sigmoid"]), "gamma": Categorical(["scale", "auto"]), "max_iter": Integer (100, 1000), Second Grid: ”degree": Integer (1,5) 5-fold Cross Validation Parallel jobs: 5 Scoring: roc_auc

8.4 Appendix D: Iteration 3 Feature Distribution

Description: The composite scores of the items were mostly left skewed, indicating a concentration on higher values for these scores. So overall we can conclude that the population is characterised by higher autonomy, leader member exchange, role clarity, and self-efficacy. This explains why moderate values of Structural Job Resources are related to non-flourishing employees. Interestingly, the interacting variables that represent the moderating effect of mental well-being on education, overload, structural job resources, and capability are quite normally distributed. The constructed Supervisor relationships interaction and the interaction of social job resources with job demands shows right skewness.

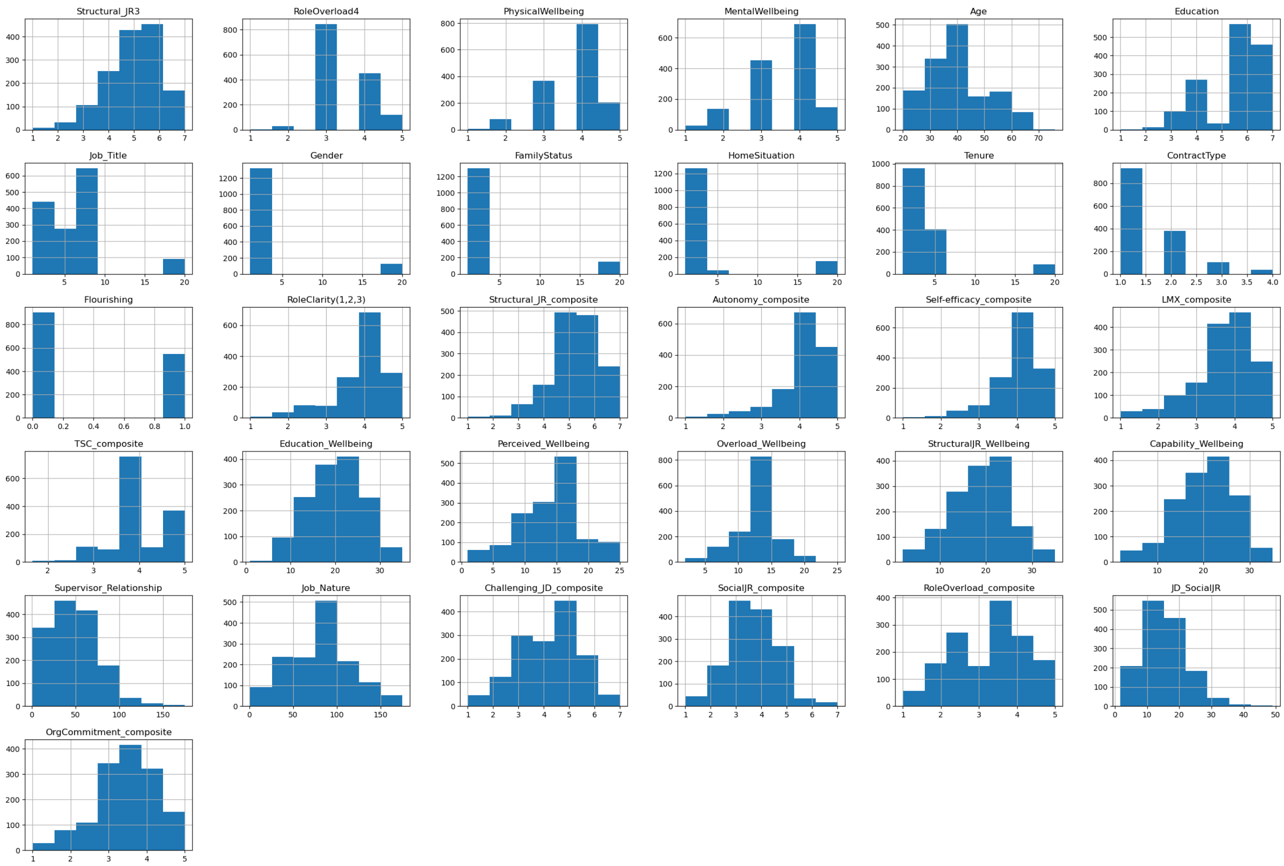


Figure 8.5: Iteration 3: Distribution of all features

8.5 Appendix E: Best Logistic Regression Coefficients

Description: Model can benefit from dimensionality reduction of the less influential variables. Interestingly, an increase in LMX composite does not increase the odds of being classified as flourishing. The most influential variable is structural JR composite, with a one standard deviation increase resulting in 52% increase in odds of flourishing (substituting for e^{β}). This is followed by Organizational Commitment composite with a 37.9% increase per standard unit. Being male counts for a 33% decrease in likelihood of flourishing on its own. Additionally, employees that have worked for 1 year at the university account for 19% increase in likelihood of flourishing, while employees who have worked for 1 -5 years have a 29.8% decrease in odds of flourishing. Permanent contract types also account for a similar decrease (33%) in flourishing odds, although other contract types do not account for a significant increase.

With these combinations we notice that structural JR and organizational commitment are good predictor of flourishing. However, there are no demographics that strongly support a flourishing employee. In addition, the model emphasizes a need to balance the score towards 0 values. This is supproted by the negative intercept and the low positive coefficients. This can be interpreted as an inability to see a direct link of these features to an employee flourishing.

Intercept:	*- 0.12772357*	Education_7	+ 0.017312
Age:	+ 0.118945	Job_Title_1	- 0.170038
TSC_composite	+ 0.176595	Job_Title_2	- 0.046011
SocialJR_composite	+ 0.131071	Job_Title_3	+ 0.065750
LMX_composite	+ 0.003057	Job_Title_4	- 0.066777
Self-efficacy_composite	**+ 0.231453**	Job_Title_5	- 0.051901
Autonomy_composite	**- 0.301368**	Job_Title_6	+ 0.021657
OrgCommitment_composite	**+ 0.448795**	Job_Title_7	+ 0.084240
RoleClarity_composite	+ 0.166664	Job_Title_8	+ 0.005442
Structural_JR_composite	**+ 0.547791**	Job_Title_9	- 0.114298
RoleOverload_composite	- 0.003263	Job_Title_20	+ 0.144687
Challenging_JD_composite	- 0.046700	**Gender_1**	**- 0.271314**
Job_Nature	**+ 0.318646**	Gender_2	+ 0.127870
Supervisor_Relationship	+ 0.082458	Gender_3	- 0.008160
JD_SocialJR	- 0.040109	Gender_20	+ 0.024356
Education_Wellbeing	+ 0.029951	FamilyStatus_1	- 0.060872
Perceived_Wellbeing	+ 0.077521	FamilyStatus_2	- 0.077642
Overload_Wellbeing	+ 0.113881	FamilyStatus_3	+ 0.065746
StructuralJR_Wellbeing	**+ 0.363533**	FamilyStatus_20	- 0.054480
Capability_Wellbeing	+ 0.135528	HomeSituation_1	- 0.139064
Structural_JR3_1	- 0.003402	HomeSituation_2	+ 0.022250
Structural_JR3_2	- 0.016687	HomeSituation_3	- 0.047762
Structural_JR3_3	- 0.123472	HomeSituation_4	- 0.008202
Structural_JR3_4	- 0.152722	HomeSituation_20	+ 0.045530
Structural_JR3_5	- 0.108170	**Tenure_1**	**+ 0.305268**
Structural_JR3_6	+ 0.108468	**Tenure_2**	**- 0.224728**
Structural_JR3_7	+ 0.168737	Tenure_3	+ 0.027619
RoleOverload4_1	- 0.018945	Tenure_4	- 0.118395
RoleOverload4_2	+ 0.016747	Tenure_5	- 0.082239
RoleOverload4_3	- 0.013065	Tenure_20	- 0.034774
RoleOverload4_4	- 0.167788	**ContractType_1**	**- 0.275096**
RoleOverload4_5	+ 0.055801	ContractType_2	+ 0.039399
PhysicalWellbeing_1	- 0.009300	ContractType_3	+ 0.101221
PhysicalWellbeing_2	+ 0.001610	ContractType_4	+ 0.007227
PhysicalWellbeing_3	- 0.026296	Self-efficay_composite_1.0	- 0.000502
PhysicalWellbeing_4	- 0.069188	Self-efficay_composite_1.77	+ 0.000000
PhysicalWellbeing_5	- 0.024076	Self-efficay_composite_2.77	- 0.003561
MentalWellbeing_1	- 0.016181	Self-efficay_composite_4.0	- 0.013774
MentalWellbeing_2	- 0.020278	Self-efficay_composite_5.44	+ 0.025380
MentalWellbeing_3	- 0.018952	Self-efficay_composite_7.11	- 0.016179
MentalWellbeing_4	- 0.132560	Self-efficay_composite_9.0	- 0.024956
MentalWellbeing_5	+ 0.060722	Self-efficay_composite_11.11	- 0.089797
Education_1	+ 0.013058	Self-efficay_composite_13.44	- 0.076288
Education_2	+ 0.007326	Self-efficay_composite_16.0	+ 0.128064
Education_3	+ 0.052468	Self-efficay_composite_18.77	- 0.088362
Education_4	- 0.072169	Self-efficay_composite_21.77	- 0.070043
Education_5	+ 0.018672	Self-efficay_composite_25.0	+ 0.1027
Education_6	- 0.163916		

FIGURE 8.6: Coefficients from Iteration 3 with transformations of LR x GA

Regarding Tenure, Gender and Contract Type, following the model's strong negative likelihood for a particular case, a Chi-square test was conducted with flourishing (figure 8.7). This emphasized for $p = 0.05$ that there is a significant association between Tenure, Gender and Contract Type towards flourishing (high levels of work engagement).

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Tenure * Flourishing	1428	98.4%	23	1.6%	1451	100.0%
Gender * Flourishing	1428	98.4%	23	1.6%	1451	100.0%
ContractType * Flourishing	1428	98.4%	23	1.6%	1451	100.0%

Crosstab

Count

		Flourishing		Total
		.00	1.00	
Tenure	1	118	164	282
	2	336	162	498
	3	107	51	158
	4	127	60	187
	5	135	82	217
	20	65	21	86
Total		888	540	1428

Chi-Square Tests (Tenure)

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	66.767[a]	5	<.001
Likelihood Ratio	65.491	5	<.001
Linear-by-Linear Association	13.805	1	<.001
N of Valid Cases	1428		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 32.52.

Crosstab

Count

		Flourishing		Total
		.00	1.00	
Gender	1	404	267	671
	2	386	237	623
	3	6	0	6
	20	92	36	128
Total		888	540	1428

Chi-Square Tests (Gender)

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	9.888[a]	3	.020
Likelihood Ratio	12.152	3	.007
Linear-by-Linear Association	6.018	1	.014
N of Valid Cases	1428		

a. 2 cells (25.0%) have expected count less than 5. The minimum expected count is 2.27.

Crosstab

Count

		Flourishing		Total
		.00	1.00	
ContractType	1	604	306	910
	2	219	160	379
	3	44	57	101
	4	21	17	38
Total		888	540	1428

Chi-Square Tests (Contract Type)

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	25.578[a]	3	<.001
Likelihood Ratio	25.066	3	<.001
Linear-by-Linear Association	20.887	1	<.001
N of Valid Cases	1428		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 14.37.

FIGURE 8.7: Chi square test results on Gender, Contract Type and Tenure

Bibliography

Allen, Natalie J. and John P. Meyer (1990). "The measurement and antecedents of affective, continuance and normative commitment to the organization". In: *Journal of Occupational Psychology* 63, pp. 1–18. DOI: 10.4324/9781351121149-6.

Ariza-Montes, Antonio et al. (2018). "Work engagement and flourishing at work among nuns: The moderating role of human values". In: *Frontiers in Psychology* 9.SEP, pp. 1–11. ISSN: 16641078. DOI: 10.3389/fpsyg.2018.01874.

A'yuninnisa, Rizqi N., Lara Carminati, and Celeste P.M. Wilderom (2023). "Job flourishing research: A systematic literature review". In: *Current Psychology* March. ISSN: 19364733. DOI: 10.1007/s12144-023-04618-w. URL: https://doi.org/10.1007/s12144-023-04618-w.

Bakker, Arnold B. and Evangelia Demerouti (2007). "The Job Demands-Resources model: State of the art". In: *Journal of Managerial Psychology* 22.3, pp. 309–328. ISSN: 02683946. DOI: 10.1108/02683940710733115.

Bakker, Arnold B and Wido G M Oerlemans (2019). "Daily job crafting and momentary work engagement : A self- determination and self-regulation perspective". In: *Journal of Vocational Behavior* 112.December 2018, pp. 417–430. ISSN: 0001-8791. DOI: 10.1016/j.jvb.2018.12.005. URL: https://doi.org/10.1016/j.jvb.2018.12.005.

Bakker, Arnold B. and Ana Isabel Sanz-Vergel (2013). "Weekly work engagement and flourishing: The role of hindrance and challenge job demands". In: *Journal of Vocational Behavior* 83.3, pp. 397–409. ISSN: 00018791. DOI: 10.1016/j.jvb.2013.06.008. URL: http://dx.doi.org/10.1016/j.jvb.2013.06.008.

Beauchamp, Mark R. et al. (2002). "Role ambiguity, role efficacy, and role performance: Multidimensional and mediational relationships within interdependent sport teams". In: *Group Dynamics* 6.3, pp. 229–242. ISSN: 10892699. DOI: 10.1037/1089-2699.6.3.229.

Belattar, Sara, Otman Abdoun, and El Khatir Haimoudi (2023). *Overview of Artificial Intelligence in Agriculture: An Impact of Artificial Intelligence Techniques on the Agricultural Productivity.* Vol. 772 LNNS. Springer Nature Switzerland, pp. 447–461. ISBN: 9783031435195. DOI: 10.1007/978-3-031-43520-1_38. URL: http://dx.doi.org/10.1007/978-3-031-43520-1_38.

Bergstra, James et al. (2011). "Algorithms for hyper-parameter optimization". In: *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011, NIPS 2011*, pp. 1–9.

Breiman, Leo (2001). "Random Forests". In: *Machine Learning, 45, 5-32*. Vol. 12343 LNCS, pp. 503–515. ISBN: 9783030620073. DOI: 10.1007/978-3-030-62008-0_35.

Buitinck, Lars et al. (2013). "API design for machine learning software: experiences from the scikit-learn project". In: *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pp. 108–122.

Butler, Julie and Margaret L. Kern (2016). "The PERMA-Profiler: A brief multidimensional measure of flourishing". In: *International Journal of Wellbeing* 6.3, pp. 1–48. DOI: 10.5502/ijw.v6i3.526.

Cavazotte, Flavia, Sylvia Freitas Mello, and Lucia B. Oliveira (2021). "Expatriate's engagement and burnout: the role of purpose-oriented leadership and cultural intelligence". In: *Journal of Global Mobility* 9.1, pp. 90–106. ISSN: 20498802. DOI: 10.1108/JGM-05-2020-0031.

Cha, Zhang and Ma Yunqian (2012). "Ensemble Machine Learning". In: *Ensemble Machine Learning.* Chap. 5, pp. 157–175. ISBN: 9781441993250. DOI: 10.1007/978-1-4419-9326-7.

Chaka, Chaka (2022). "Educational data mining, student academic performance prediction, prediction methods, algorithms and tools: an overview of reviews". In: *Journal of E-Learning and Knowledge Society* 18.2, pp. 58–69. ISSN: 19718829. DOI: 10.20368/1971-8829/1135578.

Dam, Arjan van, Gera Noordzij, and Marise Born (2020). "Thriving Under Uncertainty: The Effect of Achievement Goal Orientation on Job Insecurity and Flourishing". In: *Social Indicators Research* 150.2, pp. 659–678. ISSN: 15730921. DOI: 10.1007/s11205-020-02337-4. URL: https://doi.org/10.1007/s11205-020-02337-4.

Demerouti, Evangelia, Arnold B. Bakker, and Josette M.P. Gevers (2015). "Job crafting and extra-role behavior: The role of work engagement and flourishing". In: *Journal of Vocational Behavior* 91, pp. 87–96. ISSN: 00018791. DOI: 10.1016/j.jvb.2015.09.001. URL: http://dx.doi.org/10.1016/j.jvb.2015.09.001.

Deobald, Ulrich Leicht et al. (2019). "The Challenges of Algorithm - Based HR Decision - Making for Personal Integrity". In: *Journal of Business Ethics* 160.2, pp. 377–392. ISSN: 1573-0697. DOI: 10.1007/s10551-019-04204-w. URL: https://doi.org/10.1007/s10551-019-04204-w.

Diedericks, Elsabé and Sebastiaan Rothmann (2013). "Flourishing of information technology professionals: The role of work engagement and job satisfaction". In: *Journal of Psychology in Africa* 23.2, pp. 225–233. ISSN: 14330237. DOI: 10.1080/14330237.2013.10820618.

Diener, Ed et al. (2010). "New well-being measures: Short scales to assess flourishing and positive and negative feelings". In: *Social Indicators Research* 97.2, pp. 143–156. ISSN: 03038300. DOI: 10.1007/s11205-009-9493-y.

Duggan, James et al. (2019). "Algorithmic management and app-work in the gig economy : A research agenda for employment relations and HRM". In: June 2017, pp. 114–132. DOI: 10.1111/1748-8583.12258.

Ezzati, A et al. (2019). "Optimizing Machine Learning Methods to Improve Predictive Models of Alzheimer's Disease". In: *Journal of Alzheimer's Disease* 71.3, pp. 1027–1036. ISSN: 1875-8908. DOI: 10.3233/JAD-190262.Optimizing. URL: http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L629572252%0Ahttp://dx.doi.org/10.3233/JAD-190262.

Ford, Robert C., Sean A. Newman, and Loren R. Ford (2023). "Giving to get loyalty: How organizations signal their loyalty to employees". In: *Organizational Dynamics* 52.1, p. 100956. ISSN: 00902616. DOI: 10.1016/j.orgdyn.2022.100956. URL: https://doi.org/10.1016/j.orgdyn.2022.100956.

Garg, Swati, Shuchi Sinha, and Arpan Kumar Kar (2019). "A review of machine learning applications in human resource management". In: Ml. DOI: 10.1108/IJPPM-08-2020-0427.

Graen, George B. and Mary Uhl-Bien (1995). "Relationship-Based Approach to Leadership: Development of Leader-Member Exchange (LMX) Theory of Leadership over 25 Years: Applying a Multi-Level Multi-Domain Perspective George". In: *Management Department Faculty Publications* 57. DOI: 10.9744/jmk.19.2.112-117. URL: https://digitalcommons.unl.edu/managementfacpub/57%0AThis.

Guerci, Marco, Sven Hauff, and Silvia Gilardi (2022). "High performance work practices and their associations with health, happiness and relational well-being: are there any tradeoffs?" In: *International Journal of Human Resource Management* 33.2, pp. 329–359. ISSN: 14664399. DOI: 10.1080/09585192.2019.1695647. URL: https://doi.org/10.1080/09585192.2019.1695647.

Hoekstra, Jurriaan et al. (2023). "Predicting self-perceived general health status using machine learning: an external exposome study". In: *BMC Public Health* 23.1, pp. 1–17. ISSN: 14712458. DOI: 10.1186/s12889-023-15962-8. URL: https://doi.org/10.1186/s12889-023-15962-8.

Hridoy, Rashidul Hasan et al. (2024). "Performance enhancement of machine learning algorithm for breast cancer diagnosis using hyperparameter optimization". In: *International Journal of Electrical and Computer Engineering* 14.2, pp. 2181–2190. ISSN: 20888708. DOI: 10.11591/ijece.v14i2.pp2181-2190.

Imran, Muhammad Yasir et al. (2020). "Impact of perceived organizational support on work engagement: Mediating mechanism of thriving and flourishing". In: *Journal of Open Innovation: Technology, Market, and Complexity* 6.3. ISSN: 21998531. DOI: 10.3390/JOITMC6030082.

Jawa, Taghreed M. (2022). “Logistic regression analysis for studying the impact of home quarantine on psychological health during COVID-19 in Saudi Arabia”. In: *Alexandria Engineering Journal* 61.10, pp. 7995–8005. ISSN: 11100168. DOI: 10.1016/j.aej.2022.01.047. URL: https://doi.org/10.1016/j.aej.2022.01.047.

Jin, Ziwei et al. (2020). “RFRSF: Employee Turnover Prediction Based on Random Forests and Survival Analysis”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 12343 LNCS, pp. 503–515. ISSN: 16113349. DOI: 10.1007/978-3-030-62008-0_35.

Kavyashree, C., H. S. Vimala, and J. Shreyas (2024). “A systematic review of artificial intelligence techniques for oral cancer detection”. In: *Healthcare Analytics* 5.January, p. 100304. ISSN: 27724425. DOI: 10.1016/j.health.2024.100304. URL: https://doi.org/10.1016/j.health.2024.100304.

Keyes, Corey L.M. (2005). “Mental illness and/or mental health? Investigating axioms of the complete state model of health”. In: *Journal of Consulting and Clinical Psychology* 73.3, pp. 539–548. ISSN: 0022006X. DOI: 10.1037/0022-006X.73.3.539.

Keyes, Corey L.M. and Julia Annas (2009). “Feeling good and functioning well: Distinctive concepts in ancient philosophy and contemporary science”. In: *Journal of Positive Psychology* 4.3, pp. 197–201. ISSN: 17439760. DOI: 10.1080/17439760902844228.

Kullolli, Taulant, Brunela Trebicka, and Shkelqim Fortuzi (2024). “Understanding Customer Satisfaction Factors: A Logistic Regression Analysis”. In: *Journal of Educational and Social Research* 14.2, p. 218. ISSN: 2239-978X. DOI: 10.36941/jesr-2024-0038.

Kumar, Vijendra and S. M. Yadav (2022). “A state-of-the-Art review of heuristic and metaheuristic optimization techniques for the management of water resources”. In: *Water Supply* 22.4, pp. 3702–3728. ISSN: 16070798. DOI: 10.2166/ws.2022.010.

Latham, Gary et al. (2005). “New developments in performance management”. In: *Organizational Dynamics* 34.1, pp. 77–87. ISSN: 00902616. DOI: 10.1016/j.orgdyn.2004.11.001.

Leede, Jan de and Jeroen Meijerink (2019). *Work engagement and work pressure: still in balance? A well being study among UT employees.* Tech. rep. Enschede, Netherlands: HR department, University of Twente.

Leguina, Adrian (2015). *A primer on partial least squares structural equation modeling (PLS-SEM).*

Lessmann, Stefan, Robert Stahlbock, and Sven F. Crone (2005). “Optimizing hyperparameters of support vector machines by genetic algorithms”. In: *Proceedings of the 2005 International Conference on Artificial Intelligence, ICAI’05* 1, pp. 74–80.

Meijerink, Jeroen and Tanya Bondarouk (2023). “The duality of algorithmic management: Toward a research agenda on HRM algorithms, autonomy and value creation”. In: *Human Resource Management Review* 33.1. ISSN: 10534822. DOI: 10.1016/j.hrmr.2021.100876.

Meraliyev, Bakhtiyor, Balzhan Alibekova, and Islana Bekturganova (n.d.). “Machine Learning as an effective tool for Human Resource management in recruiting process in the higher educational field”. In: *2023 17th International Conference on Electronics Computer and Computation (ICECCO)* (), pp. 1–5. DOI: 10.1109/ICECCO58239.2023.10147133.

Moeini, Mohammadreza (2024). “Hyperparameter tuning of supervised bagging ensemble machine learning model using Bayesian optimization for estimating stormwater quality”. In: *Sustainable Water Resources Management* 10.2, pp. 1–12. ISSN: 23635045. DOI: 10.1007/s40899-024-01064-9. URL: https://doi.org/10.1007/s40899-024-01064-9.

Mori, Naoki, Masayuki Takeda, and Keinosuke Matsumoto (2005). “A comparison study between genetic algorithms and Bayesian Optimize Algorithms by Novel Indices”. In: *GECCO 2005 - Genetic and Evolutionary Computation Conference* July, pp. 1485–1492. DOI: 10.1145/1068009.1068244.

Nahm, Francis Sahngun (2022). “Receiver operating characteristic curve: overview and practical use for clinicians”. In: *Korean Journal of Anesthesiology* 75.1, pp. 25–36. ISSN: 20057563. DOI: 10.4097/kja.21209.

Nurajijah, Siti, Dedy Rahman Wijaya, and Siska Komala Sari (2022). "Gradient Tree Boosting for HR Talent Management Application". In: *2022 10th International Conference on Information and Communication Technology (ICoICT)*, pp. 264–269. DOI: 10.1109/ICoICT55009.2022.9914833.

Padkapayeva, Kathy et al. (2018). "Gender/sex differences in the relationship between psychosocial work exposures and work and life stress". In: *Annals of Work Exposures and Health* 62.4, pp. 416–425. ISSN: 23987316. DOI: 10.1093/annweh/wxy014.

Pampel, Fred C (2022). "Sage Research Methods Logistic Regression: A Primer". In: URL: https://doi.org/10.4135/9781071878729.

Pap, Jozsef et al. (2022). "Correlation Analysis of Factors Affecting Firm Performance and Employees Wellbeing: Application of Advanced Machine Learning Analysis". In: *Algorithms* 15.9. ISSN: 19994893. DOI: 10.3390/a15090300.

Parent-rocheleau, Xavier et al. (2024). "Creation of the algorithmic management questionnaire : A six-phase scale development process". In: pp. 25–44. DOI: 10.1002/hrm.22185.

Peccei, Riccardo and Karina Van De Voorde (2019). "Human resource management–well-being–performance research revisited: Past, present, and future". In: *Human Resource Management Journal* 29.4, pp. 539–563. ISSN: 17488583. DOI: 10.1111/1748-8583.12254.

Pedregosa, Fabian et al. (2011). "Scikit-learn: Machine learning in Python". In: *Journal of machine learning research* 12.Oct, pp. 2825–2830.

Peng, Chao Ying Joanne, Kuk Lida Lee, and Gary M. Ingersoll (2002). "An introduction to logistic regression analysis and reporting". In: *Journal of Educational Research* 96.1, pp. 3–14. ISSN: 19400675. DOI: 10.1080/00220670209598786.

Piasna, Agnieszka (2024). "Algorithms of time : how algorithmic management changes the temporalities of work and prospects for working time reduction". In: May 2023, pp. 115–132.

Pourkhodabakhsh, Nima (2023). "Effective machine learning , Meta-heuristic algorithms and multi-criteria decision making to minimizing human resource turnover". In.

Rama, Sai et al. (2023). "Journal of High Technology Management Research Exploring human resource management intelligence practices using machine learning models". In: *Journal of High Technology Management Research* 34.2, p. 100466. ISSN: 1047-8310. DOI: 10.1016/j.hitech.2023.100466. URL: https://doi.org/10.1016/j.hitech.2023.100466.

Rautenbach, C (2015). "Whole dissertation - includes various studies incl FAWS-SF". In: November.

Redelinghuys, Kleinjan, Sebastiaan Rothmann, and Elrie Botha (2019). "SA journal of industrial psychology = SA tydskrif vir bedryfsielkunde." In: *SA Journal of Industrial Psychology* 45.0, p. 11. ISSN: 2071-0763. URL: https://sajip.co.za/index.php/sajip/article/view/1549/2334.

Ren, Hong et al. (2015). "Expatriate success and thriving: The influence of job deprivation and emotional stability". In: *Journal of World Business* 50.1, pp. 69–78. ISSN: 10909516. DOI: 10.1016/j.jwb.2014.01.007.

Rizzo, John R, Robert J House, and Sidney I Lirtzman (1970). "Role Conflict and Ambiguity in Complex Organizations". In: *Administrative Science Quarterly* 15.2, pp. 150–163. URL: https://www.jstor.org/stable/2391486.

Robledo, Enrique, Salvatore Zappala, and Gabriela Topa (2019). "Job Crafting as a Mediator between Work Engagement and Wellbeing Outcomes : A Time-Lagged Study". In: *Int. J. Environ. Res. Public Health* 16, p. 1376. DOI: doi:10.3390/ijerph16081376.

Rothmann, Ian (2014). "Flourishing in work and careers". In: *Psycho-Social Career Meta-Capacities: Dynamics of Contemporary Career Development*. Springer International Publishing Switzerland 2014. Chap. 11, pp. 203–220. ISBN: 9788578110796. URL: 10.1007/978-3-319-00645-1_11.

Safavian, S. Rasoul and David Landgrebe (1991). "A Survey of Decision Tree Classifier Methodology". In: *IEEE Transactions on Systems, Man and Cybernetics* 21.3, pp. 660–674. ISSN: 21682909. DOI: 10.1109/21.97458.

Sargent, Leisa D. and Christina Sue-Chan (2001). "Does diversity affect group efficacy? The intervening role of cohesion and task interdependence". In: *Small Group Research* 32.4, pp. 426–450. ISSN: 10464964. DOI: 10.1177/104649640103200403.

Savadatti, Mamta B. et al. (2022). "An Overview of Predictive Analysis based on Machine learning Techniques". In: *Proceedings - IEEE International Conference on Advances in Computing, Com-*

munication and Applied Informatics, ACCAI 2022, pp. 1–6. DOI: 10.1109/ACCAI53970.2022.9752630.

Schaufeli, W. B. and A. B. Bakker (2004). "Utrecht work engagement scale Preliminary Manual Version 1.1". In: *Occupational Health Psychology Unit Utrecht University* December, pp. 1–60. ISSN: 03685810. DOI: 10.1037/t01350-000. URL: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Utrecht+work+engagement+scale#0.

Schotanus-dijkstra, Marijke and M E Pieterse (2016). "What Factors are Associated with Flourishing ? Results from a Large Representative National Sample". In: *Journal of Happiness Studies* 17.4, pp. 1351–1370. ISSN: 1573-7780. DOI: 10.1007/s10902-015-9647-3.

Sharma, Gaurav (2021). "A literature review on application of Artificial Intelligence in Human Resource Management and its practices in current organizational scenario". In: *2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pp. 594–600. DOI: 10.1109/I-SMAC52330.2021.9640954.

Shet, Sateesh and Binesh Nair (2022). "Quality of hire : expanding the multi-level fi t employee selection using machine learning". In: DOI: 10.1108/IJOA-06-2021-2843.

Siri, Yellu, Suhail Afroz, and Rella Usha Rani (2024). "Enhancing Sentiment Analysis Accuracy by Optimizing Hyperparameters of SVM and Logistic Regression Models". In: *E3S Web of Conferences* 472. ISSN: 22671242. DOI: 10.1051/e3sconf/202447201017.

Spreitzer, Gretchen M. (1995). "Psychological Empowerment in the Workplace: Dimensions , Measurement , and Validation Author (s): Gretchen M . Spreitzer Source : The Academy of Management Journal , Vol . 38 , No . 5 (Oct ., 1995), pp . 1442-1465 Published by : Academy of Management". In: *Academy of Management Journal* 38.5, pp. 1442–1465. ISSN: 0001-4273. arXiv: 0803973233.

Stanley, Peter J and Nicola S Schutte (2023). "New Ideas in Psychology Merging the Self-Determination Theory and the Broaden and Build Theory through the nexus of positive affect : A macro theory of positive functioning". In: *New Ideas in Psychology* 68.August 2022, p. 100979. ISSN: 0732-118X. DOI: 10.1016/j.newideapsych.2022.100979. URL: https://doi.org/10.1016/j.newideapsych.2022.100979.

Sun, Deliang et al. (2021). "Assessment of landslide susceptibility mapping based on Bayesian hyperparameter optimization: A comparison between logistic regression and random forest". In: *Engineering Geology* 281.May 2020, p. 105972. ISSN: 00137952. DOI: 10.1016/j.enggeo.2020.105972. URL: https://doi.org/10.1016/j.enggeo.2020.105972.

Sutton, Anna and Carol Atkinson (2023). "Performance at the cost of well-being? Testing the multi-level effects of HR practices on organisational performance via employee experiences and well-being". In: *Evidence-based HRM*. ISSN: 20493991. DOI: 10.1108/EBHRM-12-2022-0299.

Tambe, Prasanna, Peter Cappelli, and Valery Yakubovich (2019). "Artificial intelligence in human resources management: Challenges and A path forward". In: *California Management Review* 61.4, pp. 15–42. ISSN: 21628564. DOI: 10.1177/0008125619867910.

Tims, Maria, Arnold B. Bakker, and Daantje Derks (2012). "Development and validation of the job crafting scale". In: *Journal of Vocational Behavior* 80.1, pp. 173–186. ISSN: 00018791. DOI: 10.1016/j.jvb.2011.05.009. URL: http://dx.doi.org/10.1016/j.jvb.2011.05.009.

— (2013). "The impact of job crafting on job demands, job resources, and well-being". In: *Journal of Occupational Health Psychology* 18.2, pp. 230–240. ISSN: 10768998. DOI: 10.1037/a0032141.

Tolbert, Pamela S. and Phyllis Moen (1998). "Men's and women's definitions of "good" jobs: Similarities and differences by age and across time". In: *Work and Occupations* 25.2, pp. 168–194. ISSN: 07308884. DOI: 10.1177/0730888498025002003.

Van Veldhoven, M and T. Meijman (1994). "Het meten van psychosociale arbeids-belasting met een vragenlijst: De Vragenlijst Beleving en Beoordeling van de Arbeid (VBBA) [Measuring psychosocial workload with a questionnaire]". In: *Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde* 2, pp. 229–230.

Vanhoutte, Bram (2015). "Hedonic and eudemonic wellbeing". In: *ESS ERIC (2015) Measuring and reporting on Europeans' well-being: Findings from the European Social Survey*, pp. 4–7. URL: http://www.esswellbeingmatters.org/data/index.html.

Vera Cruz, Germano et al. (2023). "Using artificial intelligence to identify the top 50 independent predictors of subjective well-being in a multinational sample of 37,991 older European Israeli adults". In: *Scientific reports* 13.1, p. 11352. ISSN: 20452322. DOI: 10.1038/s41598-023-38337-w. URL: https://doi.org/10.1038/s41598-023-38337-w.

Wirth, Rüdiger and Jochen Hipp (2000). "CRISP-DM: towards a standard process model for data mining. Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining, 29-39". In: *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining* 24959, pp. 29–39. URL: https://www.researchgate.net/publication/239585378_CRISP-DM_Towards_a_standard_process_model_for_data_mining.

Yang, Duanyi et al. (2023). "Working from Home and Worker Well-being: New Evidence from Germany". In: *ILR Review* 76.3, pp. 504–531. ISSN: 2162271X. DOI: 10.1177/00197939221148716.

Zalmanson, Lior (2017). "Hands on the Wheel : Navigating Algorithmic Management and Uber Drivers ' Autonomy". In: December.

Zhuang, Wen Long et al. (2022). "Moderating Effect of Paternalistic Leadership on the Relationship between Thriving at Work and Expatriate Performance". In: *Sustainability (Switzerland)* 14.13, pp. 1–20. ISSN: 20711050. DOI: 10.3390/su14138119.

Articles

Allen, Natalie J. and John P. Meyer (1990). "The measurement and antecedents of affective, continuance and normative commitment to the organization". In: *Journal of Occupational Psychology* 63, pp. 1–18. DOI: 10.4324/9781351121149-6.

Ariza-Montes, Antonio et al. (2018). "Work engagement and flourishing at work among nuns: The moderating role of human values". In: *Frontiers in Psychology* 9.SEP, pp. 1–11. ISSN: 16641078. DOI: 10.3389/fpsyg.2018.01874.

A'yuninnisa, Rizqi N., Lara Carminati, and Celeste P.M. Wilderom (2023). "Job flourishing research: A systematic literature review". In: *Current Psychology* March. ISSN: 19364733. DOI: 10.1007/s12144-023-04618-w. URL: https://doi.org/10.1007/s12144-023-04618-w.

Bakker, Arnold B. and Evangelia Demerouti (2007). "The Job Demands-Resources model: State of the art". In: *Journal of Managerial Psychology* 22.3, pp. 309–328. ISSN: 02683946. DOI: 10.1108/02683940710733115.

Bakker, Arnold B and Wido G M Oerlemans (2019). "Daily job crafting and momentary work engagement : A self- determination and self-regulation perspective". In: *Journal of Vocational Behavior* 112.December 2018, pp. 417–430. ISSN: 0001-8791. DOI: 10.1016/j.jvb.2018.12.005. URL: https://doi.org/10.1016/j.jvb.2018.12.005.

Bakker, Arnold B. and Ana Isabel Sanz-Vergel (2013). "Weekly work engagement and flourishing: The role of hindrance and challenge job demands". In: *Journal of Vocational Behavior* 83.3, pp. 397–409. ISSN: 00018791. DOI: 10.1016/j.jvb.2013.06.008. URL: http://dx.doi.org/10.1016/j.jvb.2013.06.008.

Beauchamp, Mark R. et al. (2002). "Role ambiguity, role efficacy, and role performance: Multidimensional and mediational relationships within interdependent sport teams". In: *Group Dynamics* 6.3, pp. 229–242. ISSN: 10892699. DOI: 10.1037/1089-2699.6.3.229.

Bergstra, James et al. (2011). "Algorithms for hyper-parameter optimization". In: *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011, NIPS 2011*, pp. 1–9.

Butler, Julie and Margaret L. Kern (2016). "The PERMA-Profiler: A brief multidimensional measure of flourishing". In: *International Journal of Wellbeing* 6.3, pp. 1–48. DOI: 10.5502/ijw.v6i3.526.

Cavazotte, Flavia, Sylvia Freitas Mello, and Lucia B. Oliveira (2021). "Expatriate's engagement and burnout: the role of purpose-oriented leadership and cultural intelligence". In: *Journal of Global Mobility* 9.1, pp. 90–106. ISSN: 20498802. DOI: 10.1108/JGM-05-2020-0031.

Chaka, Chaka (2022). "Educational data mining, student academic performance prediction, prediction methods, algorithms and tools: an overview of reviews". In: *Journal of E-Learning and Knowledge Society* 18.2, pp. 58–69. ISSN: 19718829. DOI: 10.20368/1971-8829/1135578.

Dam, Arjan van, Gera Noordzij, and Marise Born (2020). "Thriving Under Uncertainty: The Effect of Achievement Goal Orientation on Job Insecurity and Flourishing". In: *Social Indicators Research* 150.2, pp. 659–678. ISSN: 15730921. DOI: 10.1007/s11205-020-02337-4. URL: https://doi.org/10.1007/s11205-020-02337-4.

Demerouti, Evangelia, Arnold B. Bakker, and Josette M.P. Gevers (2015). "Job crafting and extra-role behavior: The role of work engagement and flourishing". In: *Journal of Vocational Behavior* 91, pp. 87–96. ISSN: 00018791. DOI: 10.1016/j.jvb.2015.09.001. URL: http://dx.doi.org/10.1016/j.jvb.2015.09.001.

Deobald, Ulrich Leicht et al. (2019). "The Challenges of Algorithm - Based HR Decision - Making for Personal Integrity". In: *Journal of Business Ethics* 160.2, pp. 377–392. ISSN: 1573-0697. DOI: 10.1007/s10551-019-04204-w. URL: https://doi.org/10.1007/s10551-019-04204-w.

Diedericks, Elsabé and Sebastiaan Rothmann (2013). "Flourishing of information technology professionals: The role of work engagement and job satisfaction". In: *Journal of Psychology in Africa* 23.2, pp. 225–233. ISSN: 14330237. DOI: 10.1080/14330237.2013.10820618.

Diener, Ed et al. (2010). "New well-being measures: Short scales to assess flourishing and positive and negative feelings". In: *Social Indicators Research* 97.2, pp. 143–156. ISSN: 03038300. DOI: 10.1007/s11205-009-9493-y.

Duggan, James et al. (2019). "Algorithmic management and app-work in the gig economy : A research agenda for employment relations and HRM". In: June 2017, pp. 114–132. DOI: 10.1111/1748-8583.12258.

Ezzati, A et al. (2019). "Optimizing Machine Learning Methods to Improve Predictive Models of Alzheimer's Disease". In: *Journal of Alzheimer's Disease* 71.3, pp. 1027–1036. ISSN: 1875-8908. DOI: 10.3233/JAD-190262.Optimizing. URL: http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L629572252%0Ahttp://dx.doi.org/10.3233/JAD-190262.

Ford, Robert C., Sean A. Newman, and Loren R. Ford (2023). "Giving to get loyalty: How organizations signal their loyalty to employees". In: *Organizational Dynamics* 52.1, p. 100956. ISSN: 00902616. DOI: 10.1016/j.orgdyn.2022.100956. URL: https://doi.org/10.1016/j.orgdyn.2022.100956.

Garg, Swati, Shuchi Sinha, and Arpan Kumar Kar (2019). "A review of machine learning applications in human resource management". In: Ml. DOI: 10.1108/IJPPM-08-2020-0427.

Graen, George B. and Mary Uhl-Bien (1995). "Relationship-Based Approach to Leadership: Development of Leader-Member Exchange (LMX) Theory of Leadership over 25 Years: Applying a Multi-Level Multi-Domain Perspective George". In: *Management Department Faculty Publications* 57. DOI: 10.9744/jmk.19.2.112-117. URL: https://digitalcommons.unl.edu/managementfacpub/57%0AThis.

Guerci, Marco, Sven Hauff, and Silvia Gilardi (2022). "High performance work practices and their associations with health, happiness and relational well-being: are there any tradeoffs?" In: *International Journal of Human Resource Management* 33.2, pp. 329–359. ISSN: 14664399. DOI: 10.1080/09585192.2019.1695647. URL: https://doi.org/10.1080/09585192.2019.1695647.

Hoekstra, Jurriaan et al. (2023). "Predicting self-perceived general health status using machine learning: an external exposome study". In: *BMC Public Health* 23.1, pp. 1–17. ISSN: 14712458. DOI: 10.1186/s12889-023-15962-8. URL: https://doi.org/10.1186/s12889-023-15962-8.

Hridoy, Rashidul Hasan et al. (2024). "Performance enhancement of machine learning algorithm for breast cancer diagnosis using hyperparameter optimization". In: *International Journal of Electrical and Computer Engineering* 14.2, pp. 2181–2190. ISSN: 20888708. DOI: 10.11591/ijece.v14i2.pp2181-2190.

Imran, Muhammad Yasir et al. (2020). "Impact of perceived organizational support on work engagement: Mediating mechanism of thriving and flourishing". In: *Journal of Open Innovation: Technology, Market, and Complexity* 6.3. ISSN: 21998531. DOI: 10.3390/JOITMC6030082.

Jawa, Taghreed M. (2022). "Logistic regression analysis for studying the impact of home quarantine on psychological health during COVID-19 in Saudi Arabia". In: *Alexandria Engineering Journal* 61.10, pp. 7995–8005. ISSN: 11100168. DOI: 10.1016/j.aej.2022.01.047. URL: https://doi.org/10.1016/j.aej.2022.01.047.

Jin, Ziwei et al. (2020). "RFRSF: Employee Turnover Prediction Based on Random Forests and Survival Analysis". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 12343 LNCS, pp. 503–515. ISSN: 16113349. DOI: 10.1007/978-3-030-62008-0_35.

Kavyashree, C., H. S. Vimala, and J. Shreyas (2024). "A systematic review of artificial intelligence techniques for oral cancer detection". In: *Healthcare Analytics* 5.January, p. 100304. ISSN: 27724425. DOI: 10.1016/j.health.2024.100304. URL: https://doi.org/10.1016/j.health.2024.100304.

Keyes, Corey L.M. (2005). "Mental illness and/or mental health? Investigating axioms of the complete state model of health". In: *Journal of Consulting and Clinical Psychology* 73.3, pp. 539–548. ISSN: 0022006X. DOI: 10.1037/0022-006X.73.3.539.

Keyes, Corey L.M. and Julia Annas (2009). "Feeling good and functioning well: Distinctive concepts in ancient philosophy and contemporary science". In: *Journal of Positive Psychology* 4.3, pp. 197–201. ISSN: 17439760. DOI: 10.1080/17439760902844228.

Kullolli, Taulant, Brunela Trebicka, and Shkelqim Fortuzi (2024). "Understanding Customer Satisfaction Factors: A Logistic Regression Analysis". In: *Journal of Educational and Social Research* 14.2, p. 218. ISSN: 2239-978X. DOI: 10.36941/jesr-2024-0038.

Kumar, Vijendra and S. M. Yadav (2022). “A state-of-the-Art review of heuristic and metaheuristic optimization techniques for the management of water resources”. In: *Water Supply* 22.4, pp. 3702–3728. ISSN: 16070798. DOI: 10.2166/ws.2022.010.

Latham, Gary et al. (2005). “New developments in performance management”. In: *Organizational Dynamics* 34.1, pp. 77–87. ISSN: 00902616. DOI: 10.1016/j.orgdyn.2004.11.001.

Lessmann, Stefan, Robert Stahlbock, and Sven F. Crone (2005). “Optimizing hyperparameters of support vector machines by genetic algorithms”. In: *Proceedings of the 2005 International Conference on Artificial Intelligence, ICAI'05* 1, pp. 74–80.

Meijerink, Jeroen and Tanya Bondarouk (2023). “The duality of algorithmic management: Toward a research agenda on HRM algorithms, autonomy and value creation”. In: *Human Resource Management Review* 33.1. ISSN: 10534822. DOI: 10.1016/j.hrmr.2021.100876.

Meraliyev, Bakhtiyor, Balzhan Alibekova, and Islana Bekturganova (n.d.). “Machine Learning as an effective tool for Human Resource management in recruiting process in the higher educational field”. In: *2023 17th International Conference on Electronics Computer and Computation (ICECCO)* (), pp. 1–5. DOI: 10.1109/ICECCO58239.2023.10147133.

Moeini, Mohammadreza (2024). “Hyperparameter tuning of supervised bagging ensemble machine learning model using Bayesian optimization for estimating stormwater quality”. In: *Sustainable Water Resources Management* 10.2, pp. 1–12. ISSN: 23635045. DOI: 10.1007/s40899-024-01064-9. URL: https://doi.org/10.1007/s40899-024-01064-9.

Mori, Naoki, Masayuki Takeda, and Keinosuke Matsumoto (2005). “A comparison study between genetic algorithms and Bayesian Optimize Algorithms by Novel Indices”. In: *GECCO 2005 - Genetic and Evolutionary Computation Conference* July, pp. 1485–1492. DOI: 10.1145/1068009.1068244.

Nahm, Francis Sahngun (2022). “Receiver operating characteristic curve: overview and practical use for clinicians”. In: *Korean Journal of Anesthesiology* 75.1, pp. 25–36. ISSN: 20057563. DOI: 10.4097/kja.21209.

Nurajijah, Siti, Dedy Rahman Wijaya, and Siska Komala Sari (2022). “Gradient Tree Boosting for HR Talent Management Application”. In: *2022 10th International Conference on Information and Communication Technology (ICoICT)*, pp. 264–269. DOI: 10.1109/ICoICT55009.2022.9914833.

Padkapayeva, Kathy et al. (2018). “Gender/sex differences in the relationship between psychosocial work exposures and work and life stress”. In: *Annals of Work Exposures and Health* 62.4, pp. 416–425. ISSN: 23987316. DOI: 10.1093/annweh/wxy014.

Pampel, Fred C (2022). “Sage Research Methods Logistic Regression: A Primer”. In: URL: https://doi.org/10.4135/9781071878729.

Pap, Jozsef et al. (2022). “Correlation Analysis of Factors Affecting Firm Performance and Employees Wellbeing: Application of Advanced Machine Learning Analysis”. In: *Algorithms* 15.9. ISSN: 19994893. DOI: 10.3390/a15090300.

Parent-rocheleau, Xavier et al. (2024). “Creation of the algorithmic management questionnaire : A six-phase scale development process”. In: pp. 25–44. DOI: 10.1002/hrm.22185.

Peccei, Riccardo and Karina Van De Voorde (2019). “Human resource management–well-being–performance research revisited: Past, present, and future”. In: *Human Resource Management Journal* 29.4, pp. 539–563. ISSN: 17488583. DOI: 10.1111/1748-8583.12254.

Pedregosa, Fabian et al. (2011). “Scikit-learn: Machine learning in Python”. In: *Journal of machine learning research* 12.Oct, pp. 2825–2830.

Peng, Chao Ying Joanne, Kuk Lida Lee, and Gary M. Ingersoll (2002). “An introduction to logistic regression analysis and reporting”. In: *Journal of Educational Research* 96.1, pp. 3–14. ISSN: 19400675. DOI: 10.1080/00220670209598786.

Piasna, Agnieszka (2024). “Algorithms of time : how algorithmic management changes the temporalities of work and prospects for working time reduction”. In: May 2023, pp. 115–132.

Pourkhodabakhsh, Nima (2023). “Effective machine learning , Meta-heuristic algorithms and multi-criteria decision making to minimizing human resource turnover”. In.

Rama, Sai et al. (2023). “Journal of High Technology Management Research Exploring human resource management intelligence practices using machine learning models”. In: *Journal of High Technology*

Management Research 34.2, p. 100466. ISSN: 1047-8310. DOI: 10.1016/j.hitech.2023.100466. URL: https://doi.org/10.1016/j.hitech.2023.100466.

Rautenbach, C (2015). "Whole dissertation - includes various studies incl FAWS-SF". In: November.

Redelinghuys, Kleinjan, Sebastiaan Rothmann, and Elrie Botha (2019). "SA journal of industrial psychology = SA tydskrif vir bedryfsielkunde." In: *SA Journal of Industrial Psychology* 45.0, p. 11. ISSN: 2071-0763. URL: https://sajip.co.za/index.php/sajip/article/view/1549/2334.

Ren, Hong et al. (2015). "Expatriate success and thriving: The influence of job deprivation and emotional stability". In: *Journal of World Business* 50.1, pp. 69–78. ISSN: 10909516. DOI: 10.1016/j.jwb.2014.01.007.

Rizzo, John R, Robert J House, and Sidney I Lirtzman (1970). "Role Conflict and Ambiguity in Complex Organizations". In: *Administrative Science Quarterly* 15.2, pp. 150–163. URL: https://www.jstor.org/stable/2391486.

Robledo, Enrique, Salvatore Zappala, and Gabriela Topa (2019). "Job Crafting as a Mediator between Work Engagement and Wellbeing Outcomes : A Time-Lagged Study". In: *Int. J. Environ. Res. Public Health* 16, p. 1376. DOI: doi:10.3390/ijerph16081376.

Safavian, S. Rasoul and David Landgrebe (1991). "A Survey of Decision Tree Classifier Methodology". In: *IEEE Transactions on Systems, Man and Cybernetics* 21.3, pp. 660–674. ISSN: 21682909. DOI: 10.1109/21.97458.

Sargent, Leisa D. and Christina Sue-Chan (2001). "Does diversity affect group efficacy? The intervening role of cohesion and task interdependence". In: *Small Group Research* 32.4, pp. 426–450. ISSN: 10464964. DOI: 10.1177/104649640103200403.

Savadatti, Mamta B. et al. (2022). "An Overview of Predictive Analysis based on Machine learning Techniques". In: *Proceedings - IEEE International Conference on Advances in Computing, Communication and Applied Informatics, ACCAI 2022*, pp. 1–6. DOI: 10.1109/ACCAI53970.2022.9752630.

Schaufeli, W. B. and A. B. Bakker (2004). "Utrecht work engagement scale Preliminary Manual Version 1.1". In: *Occupational Health Psychology Unit Utrecht University* December, pp. 1–60. ISSN: 03685810. DOI: 10.1037/t01350-000. URL: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Utrecht+work+engagement+scale#0.

Schotanus-dijkstra, Marijke and M E Pieterse (2016). "What Factors are Associated with Flourishing ? Results from a Large Representative National Sample". In: *Journal of Happiness Studies* 17.4, pp. 1351–1370. ISSN: 1573-7780. DOI: 10.1007/s10902-015-9647-3.

Sharma, Gaurav (2021). "A literature review on application of Artificial Intelligence in Human Resource Management and its practices in current organizational scenario". In: *2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pp. 594–600. DOI: 10.1109/I-SMAC52330.2021.9640954.

Shet, Sateesh and Binesh Nair (2022). "Quality of hire : expanding the multi-level fi t employee selection using machine learning". In: DOI: 10.1108/IJOA-06-2021-2843.

Siri, Yellu, Suhail Afroz, and Rella Usha Rani (2024). "Enhancing Sentiment Analysis Accuracy by Optimizing Hyperparameters of SVM and Logistic Regression Models". In: *E3S Web of Conferences* 472. ISSN: 22671242. DOI: 10.1051/e3sconf/202447201017.

Spreitzer, Gretchen M. (1995). "Psychological Empowerment in the Workplace: Dimensions , Measurement , and Validation Author (s): Gretchen M . Spreitzer Source : The Academy of Management Journal , Vol . 38 , No . 5 (Oct ., 1995), pp . 1442-1465 Published by : Academy of Management". In: *Academy of Management Journal* 38.5, pp. 1442–1465. ISSN: 0001-4273. arXiv: 0803973233.

Stanley, Peter J and Nicola S Schutte (2023). "New Ideas in Psychology Merging the Self-Determination Theory and the Broaden and Build Theory through the nexus of positive affect : A macro theory of positive functioning". In: *New Ideas in Psychology* 68.August 2022, p. 100979. ISSN: 0732-118X. DOI: 10.1016/j.newideapsych.2022.100979. URL: https://doi.org/10.1016/j.newideapsych.2022.100979.

Sun, Deliang et al. (2021). "Assessment of landslide susceptibility mapping based on Bayesian hyperparameter optimization: A comparison between logistic regression and random forest". In: *Engi-*