Publisher Name: **MKS Publication**
Editorial Book: Artificial Intelligence and Machine Learning Applications
ISBN: 978-81-994361-0-7

Chapter 3

Predicting Employee Turnover with Explainable AI: A Manufacturing Case Study

Dr. Thanakit Ouanhlee[1,2*]

[1]*California Intercontinental University, USA.*
[2]*Thipsamai Research Institute, Thailand.*
[*]*Corresponding Author.*

Abstract

Voluntary turnover is costly for labour-intensive manufacturing in emerging economies, and machine learning is increasingly proposed as a way to anticipate which workers are most at risk. A real, end-to-end explainable machine-learning study of voluntary resignation in a large textile manufacturer in Thailand draws on 1,160 employee records, of which 232 are voluntary leavers. Gradient-boosted decision trees are combined with Shapley Additive explanations (SHAP) to predict and explain resignation, with interpretable baselines for comparison and a group check across gender. The model is genuinely predictive (cross-validated ROC-AUC of 0.78, compared with 0.62 for logistic regression), but its most useful contribution lies in the explanations it reveals. Contrary to the common assumption that low pay drives attrition, the workers who left earned substantially more than those who stayed, and higher pay raised predicted resignation risk; the conventional stress indicators of overtime, absenteeism, and age did not distinguish leavers from stayers at all. The decisive evidence is behavioural: 183 of the 232 leavers (79%) subsequently returned to the firm at their former wage, reporting that the outside offers that drew them away were either too distant or simply untrue. The study's central lesson is that, in this workforce, resignation is driven by external labour-market pull and misinformation that internal HR data capture only indirectly, so the practical value of an explainable model lies less in the score than in redirecting retention effort toward marketable workers in exposed departments and toward countering false external promises with transparent information about total compensation.

Keywords: Explainable AI, Employee Turnover, People Analytics, Gradient Boosting, SHAP, Boomerang Employees, Manufacturing, Emerging Economies.

Introduction

Few problems in manufacturing are as quietly expensive as voluntary turnover. When a line operator resigns, the visible cost is a recruitment advertisement and a supervisor's time; the hidden cost is the lost output while the post is vacant, the slower pace and higher defect rate of a half-trained replacement, and the load placed on the colleagues who absorb the gap. In a labour-intensive plant running continuous shifts, these costs accumulate across hundreds of departures each year. For firms in emerging economies competing largely on cost, a churning workforce can threaten viability.

Machine learning is increasingly offered as a remedy, designed to estimate, for each worker, the probability of leaving, so that attention can be directed to those most at risk before they resign. A large applied literature reports turnover models with very high accuracy, and the most accurate are typically complex, opaque ensembles. Two difficulties attend this prospect. The first is interpretability: an opaque score offers no basis on which a manager can act intelligently and no protection against the model encoding bias. The second, less often acknowledged, is that the headline accuracies in the literature are usually obtained on clean public benchmark datasets, and may not survive contact with the messy operational data of an actual factory.

The work reported here directly confronts both difficulties, using a real organisation's data rather than a demonstration on benchmark data. Working with the complete personnel records of a large Thai textile manufacturer, it builds a gradient-boosted turnover model, explains it with SHAP, compares it against interpretable baselines, and checks its behaviour across groups. The aim is not to claim a record accuracy

How To Cite: T. Ouanhlee. *Predicting Employee Turnover with Explainable AI: A Manufacturing Case Study.* In K. Madhu (Ed). Artificial Intelligence and Machine Learning Applications. Pages 18-28, 2026. MKS Publication, India. ISBN: 978-81-994361-0-7

but to show honestly what such a model achieves on real data, and, more importantly, what its explanations reveal about why workers actually leave. Recovering those explanations — making a complex model state, in human terms, why it predicts as it does — is the task of explainable artificial intelligence, and it is the methodological core of the study.

What they reveal is unexpected, and it is the central finding. The model is genuinely predictive, but the workers who resigned were not the underpaid, overworked, or disengaged. They earned more than those who stayed, worked no more overtime, and were neither older nor younger; the usual internal markers of strain simply did not separate leavers from stayers. The decisive clue lay outside the model: of the 232 workers who voluntarily resigned, 183 came back, returning to the same wage they had left, and reported that the outside offers that had lured them away were either too far from home or had been misrepresented. Resignation in this workforce, in short, is driven by an external pull, frequently based on misinformation, that the firm's internal systems cannot see directly.

The study makes three contributions. First, it provides a transparent, reproducible account of building and evaluating an explainable turnover model on genuine operational data, including the honest, modest performance such data yield. Second, it demonstrates how explanation, rather than the score alone, turns the exercise into a diagnosis: SHAP reveals that pay and department, not overtime or tenure, carry the signal, and that the relationship with pay runs counter to received wisdom. Third, it documents the 'boomerang' phenomenon and draws from it a practical reframing of retention in this setting, away from internal stress reduction and toward countering external misinformation. To respect the firm's data-governance agreement, the organisation is not named and is referred to as 'the Company'; figures are rounded, and minor identifying details are altered without affecting the analysis.

Background and Related Work

People analytics and the turnover problem

People analytics applies statistical and machine-learning methods to workforce data, and turnover is among its most studied targets. Decades of organisational research have catalogued the antecedents of voluntary departure, ranging from job satisfaction and perceived fairness of pay to commuting burden and the availability of alternative employment (Allen, Bryant and Vardaman, 2010). Predictive modelling is intended to convert that knowledge into a forward-looking, employee-level estimate of risk.

In the specific setting studied here, the relevance of pay is not merely assumed. An earlier correlational study of the same organisation, drawing on survey responses from 110 employees analysed in SPSS, found that satisfaction with the wage structure was strongly associated with overall job satisfaction ($r = 0.76$), motivation ($r = 0.69$) and intention to remain ($r = 0.71$; all $p < 0.001$), and that roughly four in five employees regarded transparent communication of pay as important (Ouanhlee, 2025a). These employees also formed part of a larger, industry-wide study by the same author, in which survey responses from 1,110 central-region textile workers showed the same positive associations between wage structure and satisfaction, motivation and retention (all $p < 0.001$), indicating that the firm-level pattern reflects a wider industry reality rather than a peculiarity of one workplace (Ouanhlee, 2025b). Taken together, these studies establish, through stated perceptions, that compensation matters for retention in this workforce and across the sector. The present study asks the complementary, behavioural question of whether the firm's operational data can identify who actually leaves, and finds, as will be seen, that the relationship between pay and leaving is more surprising than a simple reading of the survey would suggest.

Tambe, Cappelli and Yakubovich (2019) caution that human-resources data differ from the settings where machine learning matured: datasets are small, outcomes such as resignation are comparatively rare, the data reflect prior management decisions, and the stakes for individuals are high. These features recur throughout this study.

Why workers leave: push, pull, and shocks

Why employees leave has been studied for more than a century, and the resulting theory distinguishes two broad forces (Hom et al., 2017). The classic statement is that of March and Simon (1993), who separated the desirability of movement — how much a worker wishes to leave, driven chiefly by dissatisfaction with the present job — from the ease of movement, the worker's perception of the alternatives available in the external labour market. Dissatisfaction is an internal push; the attraction of perceived alternatives is an external pull. A contented worker may leave because an appealing opportunity appears, while a dissatisfied one may stay because none does. This simple separation already anticipates the present case, in which the operative force proves to be a pull rather than a push.

Mobley (1977) traced the process linking these forces to the act of quitting, describing a sequence in which dissatisfaction prompts thoughts of leaving, followed by a job search, then a comparison of any alternative with the current role, then an intention to quit, and finally departure. Mobley's lasting contribution was to place job search and the evaluation of perceived alternatives at the centre of turnover, establishing that what a worker believes is available elsewhere matters as much as how they feel about the job they hold. That belief, crucially, may be mistaken — a point the analysis returns to later.

The most direct antecedent of the present findings is the unfolding model of Lee and Mitchell (1994), which holds that many departures are precipitated not by accumulating dissatisfaction but by a 'shock': a jarring event that prompts a worker to reconsider their position. Shocks may be personal or organisational, and one recognised path begins with an unsolicited outside offer — a specific, concrete alternative arrives and prompts a worker who was not otherwise looking to leave. This 'pull' shock describes the workers in this study almost exactly. They were not, for the most part, dissatisfied employees working through Mobley's sequence; they were largely content workers presented with an external offer that triggered a sudden re-evaluation.

A complementary literature asks not why people leave but why they stay. Mitchell et al. (2001) introduced job embeddedness, the web of links, fit, and prospective sacrifice that holds a worker in place. They showed that it predicts staying over and above job satisfaction and perceived alternatives. Embeddedness has an off-the-job, community dimension as well as an on-the-job one: proximity to home, family ties, and roots in a locality can bind a worker as firmly as anything inside the factory gates. This dimension proves central here, for it explains why so many who left came back. A worker drawn away by a distant offer discovers that the new post severs their community embeddedness — the commute is unsustainable, the family routine breaks — and the pull of home reasserts itself.

Taken together, these strands portray turnover as the joint product of internal push, external pull, and precipitating shocks, moderated by how deeply a worker is embedded (Hom et al., 2017). The present case sits firmly at the pull-and-shock end of that spectrum: departures

ISBN: 978-81-994361-0-7

were triggered by outside offers rather than by dissatisfaction, and they were undone by the embeddedness those offers ignored. To this established picture, the analysis adds an element that the classical models treated only lightly — the accuracy of the perceived alternative. Mobley's worker weighs the current job against an alternative, but if the alternative has been misrepresented, the comparison is corrupted. As Section 5.6 shows, that is precisely what occurred here: the offers were frequently untrue, and the resulting departures reversed once reality intervened.

Modelling employee attrition

Most applied attrition studies rely on public benchmark datasets such as the IBM HR Analytics set and report accuracies above ninety per cent. These are useful proofs of concept but should be read with care: benchmark data are clean, static, and easily balanced, and do not exhibit the missingness, drift, and definitional ambiguity of operational records. Reported accuracies are also misleading when the positive class is small. For these reasons, the analysis reports threshold-independent and class-sensitive metrics on genuinely operational data, treating the more modest numbers as a feature of realism.

On the methodological front, tree-based ensembles have become the default for tabular human-resources data. Gradient boosting, formalised by Friedman (2001) and made scalable by implementations such as XGBoost (Chen and Guestrin, 2016) and LightGBM (Ke et al., 2017), typically matches or exceeds neural networks on structured data while remaining cheaper to train. Class imbalance is commonly handled with class weighting or with synthetic oversampling such as SMOTE (Chawla et al., 2002); a key discipline, often neglected, is that any rebalancing must be confined to the training data, since rebalancing the test set inflates reported performance.

The accuracy–explainability tension

The predictive strength of ensembles comes at the cost of interpretability, and the field of explainable artificial intelligence (XAI) has grown to address this gap (Adadi and Berrada, 2018; Barredo et al., 2020). Two techniques dominate practice: LIME approximates a complex model near a single prediction with a readable surrogate (Ribeiro, Singh and Guestrin, 2016), and SHAP assigns each feature a contribution to a prediction using a game-theoretic foundation, with an efficient exact algorithm for tree ensembles (Lundberg and Lee, 2017; Lundberg et al., 2020). SHAP has become the de facto standard because the same values support both a global account of which features matter and a local account of an individual prediction.

Rudin (2019) argues that for high-stakes decisions, one should use models that are interpretable by construction rather than rely on black-box explanations after the fact. The stance taken here is pragmatic: the decisions the model supports are advisory, interpretable baselines are reported alongside the complex model so the cost of opacity is known, and SHAP is used to interrogate and audit the model, not to replace human judgement.

Fairness and the emerging-economy context

Because human-resources data encode past decisions, models trained on them can reproduce disadvantage (Mehrabi et al., 2022). Fairness criteria such as demographic parity and equalised odds (Hardt, Price and Srebro, 2016) can conflict, so the choice of metric must follow from the use case (Barocas, Hardt and Narayanan, 2019; Verma and Rubin, 2018), and fairness is best treated as a property of the sociotechnical system rather than of the model alone (Selbst et al., 2019).

The emerging-economy manufacturing setting sharpens these concerns. The workforce is predominantly hourly and shift-based, turnover is structurally high, data are fragmented across legacy systems so that integration dominates the work, and scrutiny of labour practices is intensifying. A model deployed here must be intelligible to managers with limited statistical training and defensible to external auditors. These realities shaped every methodological choice in this study.

The Company, the Data, and Governance

The Company and the dataset

The Company is a textile manufacturer in Thailand's central industrial belt, employing roughly 1,500 people across spinning, knitting and finishing lines, warehousing, quality control, maintenance and administration. The great majority of the workforce is hourly and works rotating shifts. Turnover is highest among production workers and is sensitive to the local labour market, rising whenever competing employers nearby raise wages or open new lines.

The analysis dataset comprises 1,160 employee records drawn from the Company's own systems, of which 232 represent voluntary resignations, and 928 represent employees still active, a leaver share of twenty per cent that is consistent with the firm's elevated turnover. Each record describes one employee: the demographic, contractual, compensation, and working-time attributes set out below, together with whether and when the worker voluntarily resigned. Involuntary separations — dismissals and contract terminations — were excluded from the leaver group, since the aim is to understand voluntary departure rather than management's own decisions. Because leavers are identified retrospectively, the design is a case-control comparison of the profiles of those who left against those who stayed, rather than a forward-looking forecast; the implications of this choice are discussed in Section 7.

Features and data quality

Table 1 lists the fields collected and their status in the analysis. An important and instructive feature of real operational data is that several intended variables carried no usable signal: performance ratings were recorded as the same middle value for everyone, attendance showed no recorded absence, and the interval since the last pay review was a fixed annual figure for all staff. These constant fields were set aside, as were fields the Company does not record at the employee level. The analysis therefore rests on a compact set of genuinely varying features. Two fields required harmonisation before use: tenure, supplied in days, was converted to months, and overtime, supplied as an annual total, was converted to a monthly average. Reporting these steps openly is part of making the study reproducible.

Table 1: Fields collected and their role in the analysis. That several intended predictors were constant or unrecorded is itself part of the finding: the internal data describe a narrower slice of working life than the drivers of turnover

Field	Description	Status in analysis
Monthly base wage	Basic monthly wage in baht	Used (informative)
Department	Production line / functional area	Used (informative)
Age	Age in years	Used
Tenure	Length of service (converted days → months)	Used
Overtime hours	Monthly average (converted from annual)	Used
Overtime pay	Monthly overtime pay in baht	Used
Role; contract type	Job category; contract type	Used (weak)
Gender	Used only for the group check, not prediction	Audit only
Worker group	Local / migrant (only 24 migrants)	Not usable (too few)
Performance rating	Same middle value for all staff	Set aside (constant)
Absenteeism; lateness	No recorded absence	Set aside (constant)
Months since pay review	Fixed annual figure for all	Set aside (constant)
Commuting distance; transfers; supervisor span	Not recorded at the employee level	Not available

Ethics and data governance

Three commitments framed the work. The system is advisory: a score informs where retention effort is directed, never an automatic adverse action. Protected and sensitive attributes were withheld from the predictive feature set: gender was retained only for the group check of Section 5.5, and the local/migrant distinction was set aside both because only twenty-four migrant workers appeared in the data, too few to support any credible group analysis, and to avoid the model using group membership. Access to scores and explanations was restricted to a small number of trained human-resources staff, with the purpose fixed in advance as retention support. All records were anonymised, with no names or identifiers, consistent with Thailand's data-protection framework and the fair-recruitment principles that govern this sector (Principles, 2017).

Methodology

The pipeline is summarised in Figure 1. Its priorities, in order, were honest evaluation, interpretability, and a fair group check, with predictive accuracy pursued within those constraints.

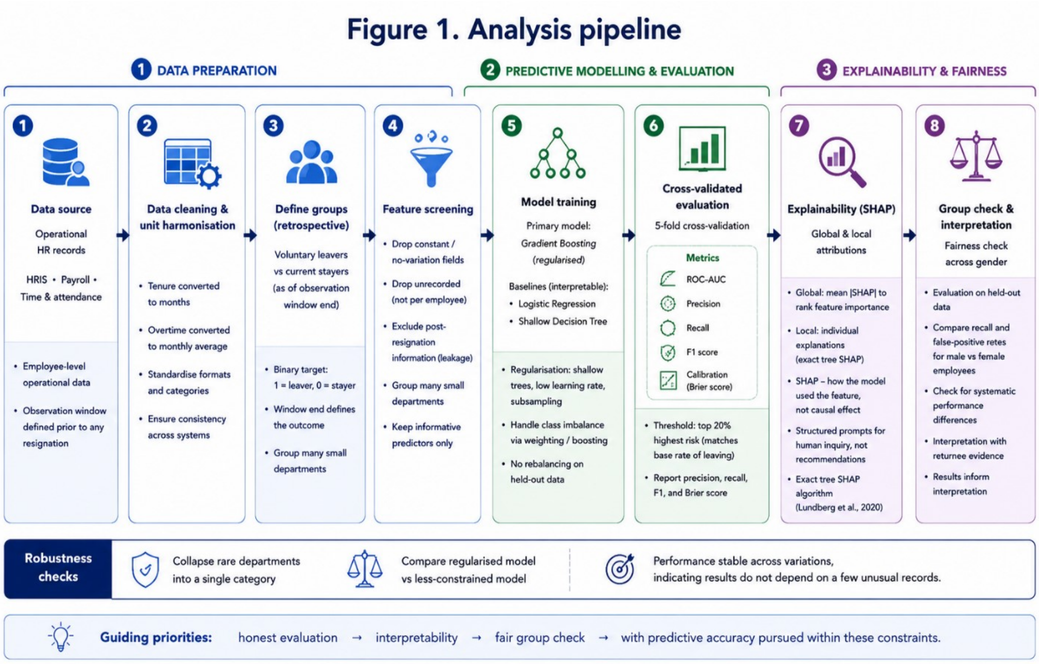

Figure 1: The analysis pipeline, from operational records through explanation and group interpretation

Preprocessing and feature screening

Continuous features were left on their natural scale, since tree-based models are insensitive to monotonic transformations and unscaled units keep SHAP explanations legible. Categorical fields were encoded directly, and the many small departments were grouped so that no rare category could be memorised. Beyond the two unit conversions already described, the main preprocessing step was screening: fields that were constant across all staff, or that the Company does not record per employee, were removed, because a variable with no variation cannot help a model and only clutters its explanation. Care was taken to exclude any field that would not have been known before a resignation, so that the comparison is not contaminated by information that becomes available only on departure.

Models and evaluation

Gradient-boosted decision trees were adopted as the primary model, deliberately regularised — shallow trees, a slow learning rate and subsampling — to guard against overfitting on a dataset of this size. Logistic regression and a shallow decision tree were retained as interpretable baselines, both to quantify the predictive value of the complex model and to offer a transparent comparison. Performance was assessed by five-fold cross-validation using the area under the ROC curve for discrimination, with precision, recall and the F1 score reported at an operating point that flags the riskiest fifth of workers — matching the base rate of leaving — and the Brier score for calibration. Class imbalance was handled by weighting within the baselines and by the boosting procedure itself; no rebalancing was applied to the held-out data, so the reported figures reflect the true class distribution.

Robustness was checked by collapsing rare departments into a single category and by comparing the regularised model against a less constrained one; the reported performance was stable across these variations, indicating that the result does not depend on a few unusual records.

Explainability and group check

Explanations were generated with SHAP using the exact tree algorithm (Lundberg et al., 2020), providing both a global view — averaging the magnitudes of each feature's contributions — and a local view that decomposes an individual's score. As throughout, a SHAP value describes how the model used a feature, not the causal effect of that feature on resignation, and the explanations are read as structured prompts for human inquiry rather than as recommendations. Because the migrant population was too small for a group analysis, the fairness check was conducted across gender, comparing the model's recall and false-positive rates for male and female workers on held-out data.

Results

Results are reported in six parts: predictive performance, the diagnostic comparison of leavers and stayers, the global explanation, an individual case, the gender check, and the boomerang finding that frames the interpretation. All figures are derived from the Company's 1,160 records, with discrimination measured via cross-validation and threshold metrics on a held-out test set.

Predictive performance

Table 2 compares the models. The gradient-boosted model is genuinely predictive, with a cross-validated ROC-AUC of 0.78, well above the logistic baseline at 0.62 and the decision tree at 0.69, and its probabilities are better calibrated, as shown by the lower Brier score. The ROC curve in Figure 2 illustrates this discrimination. The figure is deliberately well short of the near-perfect accuracies reported on clean benchmark data; on real operational records, 0.78 is a solid and honest result, and the gap is the price, and the proof, of evaluating without inflation.

Table 2: Cross-validated discrimination (ROC-AUC) and held-out metrics at an operating point flagging the riskiest 20% of workers. At that point, recall and precision coincide by construction. Lower Brier scores indicate better calibration

Model	ROC-AUC	Recall	Precision	F1	Brier
Logistic regression (baseline)	0.62	0.34	0.34	0.34	0.229
Decision tree (baseline)	0.69	0.43	0.34	0.38	0.213
Gradient boosting (primary)	0.78	0.49	0.49	0.49	0.132

Who leaves? A surprising comparison

Before interpreting the model, it is worth comparing leavers and stayers directly on each feature, as in Table 3. The result overturns the usual expectation. On age, overtime hours, and overtime pay, leavers and stayers are statistically indistinguishable. Tenure differs only slightly. The one large and highly significant difference is pay, and it runs the opposite way to the conventional assumption: the workers who left earned markedly more than those who stayed, with a median of about 15,094 baht compared with 11,340. Figure 3 shows the gap. Far from being underpaid, leavers were among the better-paid members of the workforce.

Table 3: Leavers versus stayers. SMD is the standardised mean difference; p is from a Mann–Whitney test. Only pay shows a substantial difference, and leavers earn more, not less

Feature	Leavers (median)	Stayers (median)	SMD	p
Monthly base wage (THB)	15,094	11,340	0.3.0	$<$0.001
Tenure (months)	17.0	16.2	0.18	0.005
Age (years)	40	41	0.00	0.86
Overtime (hours/month)	9.1	9.3	-0.04	0.97
Overtime pay (THB)	6,216	6,218	-0.02	0.90

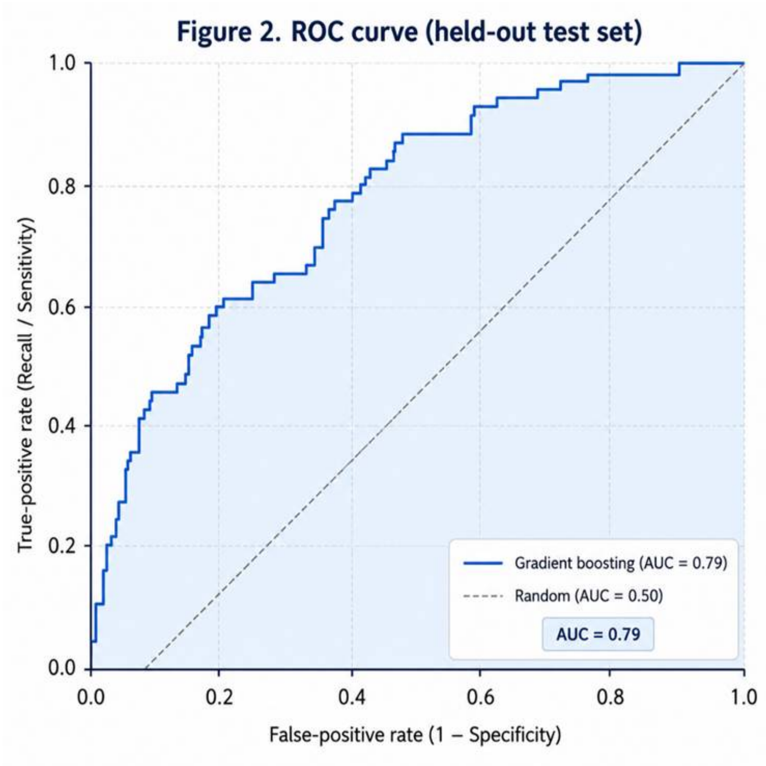

Figure 2: ROC curve for the gradient-boosted model on the held-out test set (AUC = 0.78).

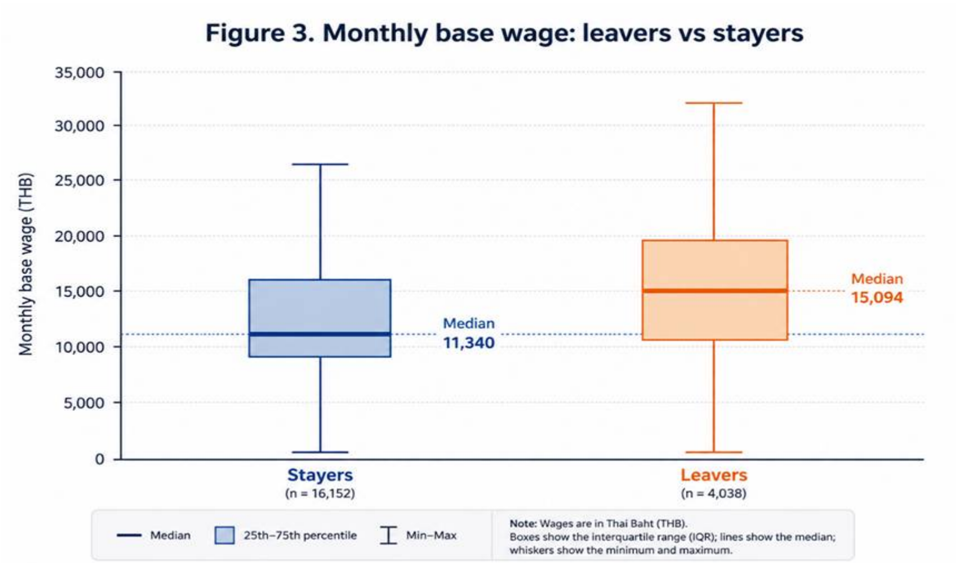

Figure 3: Monthly base wage by group. Leavers (median 15,094 THB) earned more than stayers (median 11,340 THB)

What the model relies on: global explanation

The SHAP analysis in Figure 4 confirms and extends the comparison. The model's predictions rest chiefly on two features: monthly base wage and department. Age, tenure, and overtime contribute modestly; role and contract type barely register. Crucially, the direction of the pay effect is positive — higher wages push the predicted risk of resignation up, not down — which is the model's way of encoding the pattern seen in Table 3. Department matters because resignation is concentrated in particular production lines, whose turnover rates ranged from under one in ten to more than one in three. The picture is coherent: the workers with the model flags are the better-paid ones in the more departure-prone parts of the plant.

 ISBN: 978-81-994361-0-7

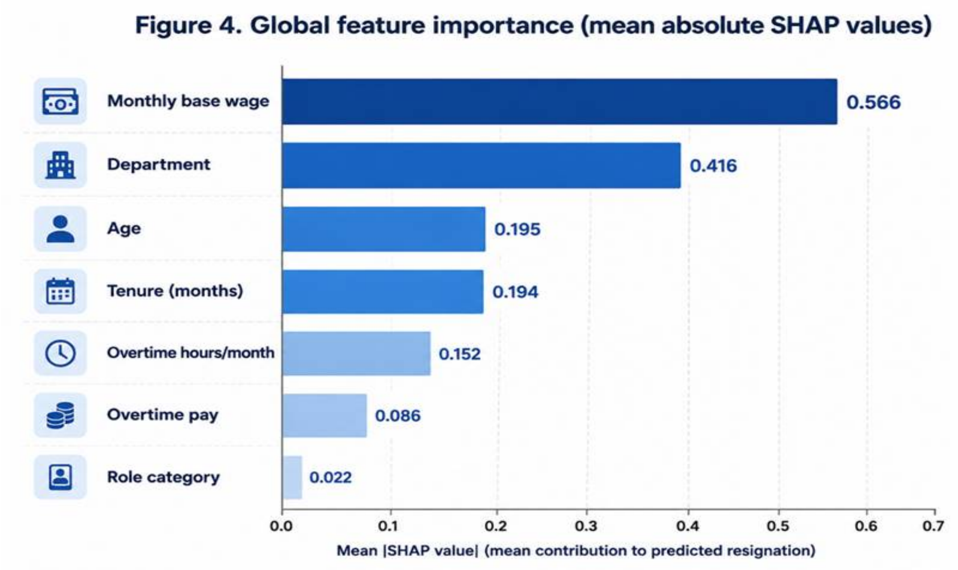

Figure 4: Global feature importance (mean absolute SHAP values) for the gradient-boosted model

Explaining an individual: local explanation

The same method explains a single worker. Consider a real high-risk case from the data: a worker in one of the higher-turnover production departments, aged forty, with eighteen months of service and a monthly wage of around 45,600 baht — several times the workforce median. The model assigned a resignation probability of 0.68, and the worker did in fact leave. The local SHAP explanation attributes that score chiefly to the department and, second, to the high wage, with age and tenure adding smaller upward contributions. The explanation is legible to a manager: this is a well-paid, experienced worker in a part of the plant that competitors target, precisely the profile an outside recruiter would approach.

This case also makes the limits of acting on the score clear. The lever the model implies — that high pay is associated with leaving — cannot sensibly be pulled by cutting pay. The explanation is useful not because it prescribes an action but because it identifies who is exposed and prompts the right question: what is drawing this worker away, and is it real?

Group check across gender

Because the migrant population was too small to analyse, the model was checked across genders on held-out data. The two groups were flagged at almost identical rates, but the model identified female leavers more reliably than male leavers, with a recall of 0.62 for women against 0.41 for men. In contrast, false-positive rates were similar and low (0.11 and 0.14). In equalised-odds terms, there is a recall gap in favor of women, most likely because the compact feature set captures the drivers of male departure less well. The gap is reported transparently rather than smoothed over; in deployment, it would warrant a group-specific operating threshold and, more importantly, better features, since the deeper remedy for an uneven model is more informative data rather than threshold arithmetic alone.

The decisive finding: most leavers came back

The result that reframes everything is not in the model at all. Of the 232 workers who voluntarily resigned, 183 — seventy-nine per cent — subsequently returned to the Company and, on return, received the same wage they had left on. Such a boomerang rate is extraordinary, and the reasons the returners gave are the key to the whole study. They fell into three recurring accounts. Some had taken jobs that proved, once started, to be too far from home to sustain, the daily commute outweighing the nominal gain. Others had been promised fringe benefits that did not materialise once they arrived. Others still had been enticed by offers of large bonuses, of the order of two or three times basic salary, that turned out to be untrue. In each case, the worker had been drawn away by an external offer that was either impractical or misrepresented, and the Company was able to take them back at their previous pay.

This analysis explains the otherwise puzzling results above. Leavers earned more not because pay caused them to leave but because better-paid, more marketable workers are the ones outside recruiters approach; overtime and absenteeism did not distinguish leavers because dissatisfaction inside the factory was not what moved them; and the model could reach only a moderate accuracy because the decisive variable — the existence and credibility of an outside offer — is nowhere in the firm's data. The internal record can identify who is exposed to being approached, but not when an approach will occur or whether the worker will believe it.

Discussion

Turnover as external pull, not internal push

The dominant narrative in retention research is one of internal push: workers leave because they are underpaid, overworked, poorly managed, or disengaged. This case shows a workforce in which that narrative does not hold. The leavers were paid more, not less; they were not working longer hours; they were not disproportionately new or old; and four in five returned. What moved them was external pull — the appearance of a better opportunity elsewhere — frequently amplified by misinformation about pay, benefits or bonuses that did not survive contact with reality.

This reframing matters because it changes what retention means. If departures were driven by internal strain, the response would be to reduce that strain: lighten overtime, improve supervision, raise pay. The evidence here suggests those levers would largely miss, because the workers leaving were not the strained ones. The effective response is different in kind: to anticipate which workers are exposed to external approaches, and to counter the misinformation that makes those approaches persuasive.

Why explanation mattered more than the score

Had this project stopped at a turnover score, it would have produced a list of higher-paid workers in certain departments and little understanding of the problem. It was the explanation — the SHAP analysis showing the counterintuitive direction of the pay effect, read alongside the simple comparison of leavers and stayers — that exposed the external-pull pattern and prompted the question that the returnee evidence then answered. A black-box score would have concealed exactly the feature, the inverted pay relationship, that made the situation intelligible. This episode is a concrete instance of the general argument for explainability in workforce models: the explanation is not decoration on top of a prediction but the means by which a prediction becomes understanding that managers can act on.

It also illustrates the value of honest evaluation. A model reported at an inflated accuracy would have invited misplaced confidence in the score. The moderate, truthful 0.78, set against the finding that the decisive variable is absent from the data, tells managers the right thing: use the model to see who is exposed, but do not expect it to foretell individual departures, because the trigger lives outside the firm's records.

Implications for practice

Three practical conclusions follow. First, retention effort should be directed by exposure rather than by strain: the model usefully identifies well-paid, marketable workers in poaching-prone departments, and these are the workers for whom retention conversations and total-reward reviews are worth the investment. Second, because the offers that lure workers away are so often misrepresented, the single most promising lever is information — being transparent with workers about their true total compensation, including benefits and the real value of overtime and stability, so that an inflated external promise can be weighed against an accurate internal picture. This recommendation connects directly to the earlier finding in this workforce that pay transparency was widely valued (Ouanhlee, 2025a). In the broader industry study (Ouanhlee, 2025b), the clarity of the organisation's communication about pay was the highest-rated aspect of compensation, even as the adequacy of pay against the cost of living was the lowest — evidence that workers prize transparency yet are not placated by it when wages fall short. Transparency here is therefore not only about internal fairness but about inoculating workers against external misinformation. Third, the boomerang rate itself is an asset. That seventy-nine per cent returned, at their former wage, suggests a low-cost, structured re-hiring policy — keeping the door open, tracking returners, and welcoming them back without penalty — may do more for net staffing than expensive efforts to prevent the initial departure.

The migrant dimension, prominent in the sector at large, could not be examined here because the Company's data contained too few migrant workers to support analysis. This limitation is itself worth recording: where a group is small or poorly captured in the data, the honest course is to decline a group claim rather than to manufacture one from a handful of cases.

The economics of a boomerang re-hiring policy

The most distinctive feature of this workforce — that seventy-nine per cent of leavers returned — carries an economic logic worth setting out, because it bears directly on how the Company should respond. Replacing a worker is expensive in ways that are easy to underestimate. The cost is not merely the recruitment advertisement but the sum of several components: the administrative cost of separation, the lost output while the post sits vacant, the expense of sourcing and screening candidates, the induction and training of whoever is hired, and — often the largest element — the period during which a new recruit works below full productivity before reaching the standard of the person who left (Allen, 2008; Hom et al., 2017). For hourly manufacturing roles, these costs are commonly estimated in the region of fifteen to twenty per cent of a worker's annual pay, and they recur with every departure.

Against this background, a worker who has left and is willing to return is an unusually economical source of labour, for three connected reasons. First, the recruitment and search costs are close to zero: the Company need not advertise, source, or extensively screen someone it already employs. Second, the uncertainty that makes external hiring risky is much reduced. Engaging a stranger is a decision under incomplete information, in which the employer cannot fully observe a candidate's true productivity in advance; a returner's past performance is already known, so the firm avoids the adverse-selection problem of paying for a quality it cannot verify. Studying more than thirty thousand hires, Arnold et al. (2021) found that the behaviour of returning employees was largely predictable from their earlier tenure, which is precisely what makes them a lower-risk bet. Third, and most importantly, a returner retains firm-specific human capital — the knowledge of the Company's machines, processes, products, and people that, in Becker's (1975) classic distinction, has value chiefly inside this particular firm. A returner brings that capital back intact and so reaches full productivity far faster than an external recruit, compressing or eliminating the costly ramp-up period; consistent with this, Arnold et al. (2021) found that boomerang employees performed on a par with newly hired and internally promoted staff in their first year.

The Company's position is more favourable still, because its returners come back at their former wage. The broader evidence on boomerang employment indicates that returners usually command a pay premium — industry surveys put the average increase at around a quarter of prior pay, and matched-sample research finds that returners are better compensated than comparable colleagues who never left (Snyder, Stewart, and Shea, 2021). The Company avoids this premium entirely: the same workers who left at a given wage return at that wage, so the firm recovers their firm-specific capital at no incremental cost and without disturbing its internal pay structure.

A simple illustration makes the magnitude concrete. Suppose, purely for illustration, that replacing a production worker through external hiring costs the equivalent of two months' pay once vacancy, recruitment, training and lost productivity are combined, while bringing a returner back — with no advertising, minimal screening and a short re-orientation — costs a small fraction of that, perhaps a fortnight's equivalent. On those assumptions, every returner the Company welcomes back rather than replacing from outside saves on the order of six weeks' pay per worker, and across the 183 who returned, a difference of that kind compounds into a substantial sum — one that accrues

ISBN: 978-81-994361-0-7

without any improvement in the underlying rate of leaving. The figures here are illustrative rather than measured, since the Company does not yet cost its turnover in this way. However, the direction is unambiguous, and quantifying these components with the firm's own data would be a worthwhile next step.

Three honest qualifications temper this optimism. The first is repeat departure: the same evidence that recommends returners also warns that they leave again more readily than other employees, and that a second departure tends to resemble the first (Arnold et al., 2021; Shipp et al., 2014). A worker drawn out once by an external offer may be drawn out again, so a re-hiring policy cannot substitute for addressing the external pull itself. The second is long-run development: Arnold et al. (2021) found that internally and externally hired staff improved more over time than returners, and Snyder, Stewart and Shea (2021) found that returners were no better performers than those who never left. Returners are an efficient way to restore capacity, not a shortcut to a stronger workforce. The third is fairness: a policy that welcomes leavers back risks resentment among those who stayed loyally — a risk the Company's same-wage practice helps to contain, since returners gain no financial advantage over their steadfast colleagues.

The practical implication is to make the boomerang route deliberate rather than incidental. An amicable separation process, a maintained connection with former staff, and a clear, no-penalty path back would formalise what is already happening informally and capture its value reliably. Most valuable of all, the return conversation is a source of intelligence the Company can obtain in no other way: a returner can report exactly what a competitor promised and how it differed from reality. Gathering that information systematically would turn returners into an early-warning system for the very misinformation that, as this study has shown, drives the initial departures — closing the loop between the analysis and the action it implies.

Transferability

Although the specific finding — turnover driven by external pull and misinformation — is particular to this firm, the approach is portable. The pattern of building a model on the data a firm already holds, comparing leavers and stayers directly, explaining the model with SHAP, and then seeking the behavioural evidence (here, the returnees) that the data alone cannot supply, is general. So too is the discipline of reporting honest performance and declining group claims the data cannot support. A practitioner elsewhere might find the opposite pattern — genuine internal push — but would find it the same way.

Limitations and Deployment Considerations

Limitations

Several limitations qualify the findings. The study is of a single organisation, so the specific drivers should not be assumed to generalise; what generalises is the method. The design is a retrospective case-control comparison: leavers' attributes reflect their records around the time of departure and stayers' a current snapshot, so the analysis identifies the profile associated with leaving rather than forecasting future departures, and some difference between the groups could reflect timing rather than behaviour. The feature set is compact because several intended variables were constant or unrecorded; richer data, especially anything capturing exposure to outside offers, might change both the accuracy and the explanation. The pay finding is associational and may partly reflect departmental composition, since better-paid workers cluster in particular lines. The gender check rests on a single held-out split and a small number of metrics. Furthermore, the central interpretation draws on the returnees' own stated reasons, which, while striking, are self-reported and were not independently verified.

From a model to a working practice

Turning this into sustained practice raises several issues beyond the model itself.

- **Monitoring and drift**. Turnover patterns shift with the local labour market; the model should be refreshed periodically and watched for drift in both its inputs and the realised leaving rate.
- **The intervention feedback loop**. If retention effort succeeds, a flagged worker does not leave, and naive retraining would record that as a false positive; interventions must be logged and accounted for so success is not mistaken for error.
- **Capacity and operating point**. The threshold should reflect how many workers the firm can genuinely support; flagging more than supervisors can engage produces alert fatigue.
- **Governance**. The advisory-only rule, the exclusion of protected attributes from prediction, restricted access, and a periodic group check are standing controls, not one-off settings.
- **Acting on exposure, not just scores**. Because the model identifies exposure rather than imminent departure, it is best used to prioritise total-reward conversations and transparent communication, with supervisors trained to read an explanation as a prompt rather than a verdict.

Conclusion

The study set out to predict voluntary turnover in a large textile manufacturer with an explainable machine-learning model, and to do so on real data rather than a clean benchmark. The model is genuinely predictive, with a cross-validated ROC-AUC of 0.78, but its lasting value lies in what its explanations revealed. Contrary to the assumption that low pay and overwork drive attrition, the workers who left this firm earned more than those who stayed and were no more overworked; the model leaned on pay and department, with the pay effect running counter to received wisdom. The decisive evidence came from outside the model: seventy-nine per cent of leavers returned, at their former wage, having been drawn away by outside offers that were too distant or untrue. Resignation here is an external pull amplified by misinformation, a driver the firm's internal data can only see indirectly, which is why an honest model reaches moderate rather than spectacular accuracy.

For practitioners, the lesson is to let explanation and evidence, not the score alone, guide action: direct retention effort by exposure rather than by strain, counter external misinformation with transparency about true total compensation, and treat a high boomerang rate as an asset to be managed rather than a failure to be hidden. For researchers, the case argues for honest evaluation of operational data, for pairing models with the behavioural evidence that explains them, and for the discipline of declining claims — such as a fairness audit across a group of twenty-four — that the data cannot support. The transformative potential of artificial intelligence in workforce management will be realised not by the firms with the highest reported accuracy, but by those willing to understand, honestly, what their models can and cannot see.

References

[1] Adadi, A. and Berrada, M., 2018. Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access, 6, pp. 52138–52160. https://doi.org/10.1109/access.2018.2870052.

[2] Allen, D. G., 2008. Retaining talent: A guide to analyzing and managing employee turnover. WRAP: Warwick Research Archive Portal. https://wrap.warwick.ac.uk/id/eprint/57100/.

[3] Allen, D. G., Bryant, P. C. and Vardaman, J. M., 2010. Retaining talent: Replacing misconceptions with evidence-based strategies. Academy of Management Perspectives, 24(2), pp. 48–64. https://doi.org/10.5465/amp.24.2.48.

[4] Arnold, J. D., Van Iddekinge, C. H., Campion, M. C., and Bauer, T. N., 2021. Welcome back? Job performance and turnover of boomerang employees compared to internal and external hires. Journal of Management, 47(8), pp. 2198–2225. https://doi.org/10.1177/0149206320936335.

[5] Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R. and Herrera, F., 2020. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, pp. 82–115. https://doi.org/10.1016/j.inffus.2019.12.012.

[6] Barocas, S., Hardt, M., and Narayanan, A., 2019. Fairness and Machine Learning: Limitations and Opportunities. Fairmlbook. Retrieved 2026, from https://fairmlbook.org.

[7] Becker, G. S., 1975. Human Capital: A theoretical and empirical analysis, with special reference to education (2nd ed.). National Bureau of Economic Research. https://www.nber.org/system/files/chapters/c3730/c3730.pdf.

[8] Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P., 2002. SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, pp. 321–357. https://doi.org/10.1613/jair.953.

[9] Chen, T., and Guestrin, C., 2016. XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

[10] Friedman, J. H., 2001. Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5). https://doi.org/10.1214/aos/1013203451.

[11] Hardt, M., Price, E., and Srebro, N., 2016. Equality of Opportunity in Supervised Learning. 30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain. NeurIPS Proceedings.

[12] Hom, P. W., Lee, T. W., Shaw, J. D., and Hausknecht, J. P., 2017. One hundred years of employee turnover theory and research. Journal of Applied Psychology, 102(3), pp. 530–545. https://doi.org/10.1037/apl0000103.

[13] Principles, G., 2017. Operational Guidelines for Fair Recruitment. International Labour Organization: Geneva, Switzerland, pp.5-7.

[14] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.Y., 2017. LightGBM: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30.

[15] Lee, T. W., and Mitchell, T. R., 1994. An alternative approach: The unfolding model of voluntary employee turnover. The Academy of Management Review, 19(1), pp. 51. https://doi.org/10.2307/258835.

[16] Lundberg, S.M. and Lee, S.I. 2017. A unified approach to interpreting model predictions, Advances in Neural Information Processing Systems 30, pp. 4765–4774.

[17] Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I., 2020. From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence, 2(1), pp. 56–67. https://doi.org/10.1038/s42256-019-0138-9.

[18] March, J.G. and Simon, H.A., 1993. Organizations. John wiley & sons.

[19] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., and Galstyan, A., 2022. A survey on bias and fairness in machine learning. ACM Computing Surveys, 54(6), pp. 1–35. https://doi.org/10.1145/3457607.

[20] Mitchell, T. R., Holtom, B. C., Lee, T. W., Sablynski, C. J., and Erez, M., 2001. Why people stay: Using job embeddedness to predict voluntary turnover. Academy of Management Journal, 44(6), pp. 1102–1121. https://doi.org/10.2307/3069391.

[21] Mobley, W. H., 1977. Intermediate linkages in the relationship between job satisfaction and employee turnover. Journal of Applied Psychology, 62(2), pp. 237–240. https://doi.org/10.1037/0021-9010.62.2.237.

[22] Ouanhlee, T., 2025a The Impact of Wage Structure on Employment Outcomes and on Employee Satisfaction, Motivation, and Retention in the Thai Textile-Manufacturing Industry. DBA dissertation. California Intercontinental University.

[23] Ouanhlee, T., 2025b. The impact of wage structure on employee satisfaction, motivation, and retention in the Thai textile industry. American Journal of Industrial and Business Management, 15(11), pp. 1726–1770. https://doi.org/10.4236/ajibm.2025.1511090.

[24] Ribeiro, M.T., Singh, S. and Guestrin, C., 2016, August. " Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144). https://doi.org/10.1145/2939672.2939778.

[25] Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), pp. 206–215. https://doi.org/10.1038/s42256-019-0048-x.

[26] Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., and Vertesi, J., 2019. Fairness and abstraction in Sociotechnical systems. Proceedings of the Conference on Fairness, Accountability, and Transparency, pp. 59–68. https://doi.org/10.1145/3287560.3287598.

[27] Shipp, A. J., Furst-Holloway, S., Harris, T. B., and Rosen, B., 2014. Gone today but here tomorrow: Extending the unfolding model of turnover to consider boomerang employees. Personnel Psychology, 67(2), pp. 421–462. https://doi.org/10.1111/peps.12039.

[28] Snyder, D. G., Stewart, V. R., and Shea, C. T., 2021. Hello again: Managing talent with boomerang employees. Human Resource Management, 60(2), pp. 295–312. https://doi.org/10.1002/hrm.22051.

[29] Tambe, P., Cappelli, P., and Yakubovich, V., 2019. Artificial intelligence in Human Resources management: Challenges and a path forward. California Management Review, 61(4), pp. 15–42. https://doi.org/10.1177/0008125619867910.

[30] Verma, S., and Rubin, J., 2018. Fairness definitions explained. Proceedings of the International Workshop on Software Fairness, pp. 1–7. https://doi.org/10.1145/3194770.3194776.