A Multi-Year Machine Learning Analysis of Developer Experience: Influencing Factors, Prediction, and Profile Discovery

Msc Research Project
Data Analytics

Alina Mathew
Student ID: x23412399

School of Computing
National College of Ireland

Supervisor: Dr. Abdul Razzaq

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Alina Mathew
Student ID:	x23412399
Programme:	Data Analytics
Year:	2025
Module:	Msc Research Project
Supervisor:	Dr. Abdul Razzaq
Submission Due Date:	28/01/2026
Project Title:	A Multi-Year Machine Learning Analysis of Developer Experience: Influencing Factors, Prediction, and Profile Discovery
Word Count:	5829
Page Count:	22

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	ALINA MATHEW
Date:	28th January 2026

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

A Multi-Year Machine Learning Analysis of Developer Experience: Influencing Factors, Prediction, and Profile Discovery

Alina Mathew
x23412399

Abstract

It has been established that developer experience (DevX) affects productivity, well-being and tool adoption, but factors that determine its emotional, value-related and perceptual aspects are understudied. Based on the 2022, 2024, and 2025 datasets of the Stack Overflow Annual Developer Survey, this thesis creates an analytical pipeline, end-to-end, to determine the primary determinants of DevX, the predictability and the latent profiles of the developers. The pipeline combines several steps: data harmonisation and MICE imputation; feature selection with the help of FLO that is model-agnostic; joint multi-target prediction with a residual multilayer perceptron; and explainability of the model with the help of SHAP-based feature attributions. Latent Profile Analysis is carried out with Gaussian Mixture Models on calibrated probability vectors and SHAP embeddings as representation space. Bayesian Information Criterion and bootstrap-adjusted Rand Index are used to calculate model selection and profile stability. A similar pattern of behavioural and organisational determinants arises across the three years of the survey with a gradual change to AI-mediated development practices in the recent years. Joint prediction has a high performance (macro-F1: 0.74-0.94), which indicates that DevX is statistically learnable. LPA has a small set of coherent developer profiles that vary in the DevX dimensions, work practices and exposure to the emerging technologies, and are stable in structure over years. These findings show that DevX is empirically structured, predictable and profilable. The results support future DevX-aware interventions, tooling strategies and organisational assessment practices.

1 Introduction

The process of software development has grown to be more complex, collaborative and cognitively challenging. The experience of software developers such as how they feel during the process of work, what they value, and what they perceive of their environment has increased with the growth in technical roles and has become a significant factor in productivity, motivation and retention. Previous research has established that emotions, satisfaction, cognitive load and perceptions of the workplace have an impact on both individual well-being and team performance (Fagerholm and Jürgen (2012); Greiler et al. (2023); Meyer et al. (2021)) ; however, the research on developer experience (DevX) is still disjointed and tends to look at these dimensions individually.

One of the most important issues in the field is how various moderating variables (e.g., individual, behavioural and workplace factors) influence various dimensions of DevX together (Forsgren et al. (2024); Razzaq et al. (2025)). These dimensions are rarely combined and rarely explored in the existing empirical literature as to their common predictors. In addition, although it is speculated that developers can develop identifiable sub-groups that exhibit specific patterns of experience, little evidence exists on whether data-driven clustering or model-based representations can be used to identify those profiles. This paper fills these gaps by examining massive scale responses to the Stack Overflow Annual Developer Survey [1], one of the largest datasets on developer practices and perceptions in the world.

To provide the structure of the investigation, the research is informed by three main questions:

- **RQ1:** Which survey variables are most strongly associated with key developer-experience dimensions such as emotions, values, and perceptions?
- **RQ2:** How well do survey variables predict DevX dimensions using supervised machine-learning models?
- **RQ3:** What latent developer profiles can be identified based on DevX-related factors and the three DevX dimensions (emotions, values, and perceptions)?

The study produces answers to these questions using a data driven pipeline that uses MICE to fill in the missing data, FLO-style permutation features selection, SMOTE balancing, an AutoML tree-based baseline model and model-agnostic explainability techniques(Chawla et al. (2002), van Buuren and Groothuis-Oudshoorn (2011)). The learned model representations are then subjected to latent profile analysis using Gaussian Mixture Models to provide clear profiles of the developers. Like any other study that is based on the survey, there are limitations to this work. The analysis is also based on self-reported perception of a non-random sample of developers and thus is prone to sampling and reporting bias. The research is not causal but merely descriptive of statistical correlations as opposed to conclusive causalness. One principled pipeline out of a variety of alternatives is modelling decisions like particular hyperparameters, imputation strategy, and clustering techniques. Such limitations are recognized in order to interpret the results properly. The rest of this report is organized in the following way: Section 2 is a review of existing literature about developer experience.Section 3 presents the dataset, operationalisation of variables and the entire modelling pipeline. Section 4 describes the design specification of the proposed system. Section 5 is about the models and analysis pipeline implementation. The predictive and clustering findings and their analysis are provided in section 6. Lastly, Section 7 is a conclusion on main contributions, limitations and future research directions.

2 Related Work

Emotions, motivation, satisfaction, productivity perceptions, and cognitive load have been studied as a way of gaining a better understanding of developer experience. Although such studies can be valuable, they are very fragmented and seldom combined

[1] `https://survey.stackoverflow.co/`

with various dimensions of. developer experience. To base our analysis, we outline the major definitions of developer experience and point to empirical evidence that defines its fundamental emotional, value-oriented, and perceptual aspects.

2.1 Developer Experience

Fagerholm and Münch (2012) originally formalised developer experience (DevX), as the way developers think and feel about their work and working environment (Fagerholm and Jürgen (2012)). Their definition is based on the psychological trilogy. of mind, that is, which sets the distinction of cognition (thinking), affect (feeling) and conation (motivation and values) as the fundamental dimensions that create experience. The dimensions are used to describe how developers judge their work context, how they feel about it, and how they interpret their activities or give them a sense of meaning or value. It has since been incorporated in a number of other latter studies such as Greiler et al. (2023) who referred to this definition as the conceptual framework behind building a series of empirically based DX Framework on the basis of interviewing industry developers. Their study determines factors, obstacles and measures that influence developer experience. Razzaq et al. (2024) build on this perspective and present 33 developer-experience factors as a result of a systematic review of the literature, describing the broad spectrum of effects on the perception, feelings, and values of the developers (Razzaq et al. (2025)). They describe a Dev-X factor as a characteristic of the developer (e.g., skills), a characteristic of an activity (e.g., interruptions), an object that they work with (e.g., code), or a process that they follow that has an impact on their experience and which can also impact productivity. Their review covers these factors in a comprehensive manner, but does not measure the relationships between the core experience dimensions. To conclude, previous research makes DevX a multi-dimensional construct that is based on cognitive, affective, and motivational. Based on this premise, the research incorporates **emotions (E)**, **values (V)**, and **perceptions (P)** as the central DevX dimensions and examines the relationship between various Dev-X variables and its results.

2.2 Studies on Developer-Experience Dimensions

2.2.1 Emotional Dimensions

Emotion could be described as a complex process, which includes subjective experience, cognitive, physiological arousal, expressive behaviour, motivational tendencies that work together to enable an individual to respond to important events (Kleinginna and Kleinginna (1981); Parrott (2001)). Empirical studies, including affective-computing and field studies, indicate that the daily work environment factors, including stalling, time constraints, coordination problems, quality of collaboration, and tool stability, have a direct effect on the emotional states (Graziotin et al. (2018); Girardi et al. (2022)). Experimental results also show that development practices (e.g., TDD) have the ability to invoke systematic emotional responses (Baldassarre et al. (2022)), and quantitative syntheses of software-engineer well-being point to larger individual, team and organisational predictors of affective response Godliauskas and Šmite (2024). Other complementary productivity-based studies investigate DevX-proximate variables, such as sentiment-based models of the relationship between emotional tone of artefacts and measures of code-quality (Patnaik and Padhy (2024)).

2.2.2 Developer Values

Values may be conceived as motivationally organised principles that regulate behaviour and assessment and are arranged into types of universal values that mirror common priorities between cultures (Schwartz (1992)). Motivational and value-oriented research findings provide evidence that when developers feel autonomy, recognised, are tasked with meaningful work, and work in teams with a sense of psychological safety, they become more aligned and thus stay engaged(França et al. (2018); Beecham et al. (2008); Wassouf-Jr et al. (2025)). Outside empirical research, a series of conceptual and practitioner-based studies exist to offer cognitive or organisational DevX patterns including infrastructure alignment, sufficient skills, mentorship, and structured knowledge sharing indicating theoretically relevant impacts on perception and cognition, but not modelling it statistically or evaluating it quantitatively (Pinho et al. (2024)). These contributions aid in explaining possible mechanisms but they are non-empirical in nature.

2.2.3 Developer Perception

Perception deals with the way one receives and reacts to sensory stimulation at a particular time (Attneave (1962)). According to perceptual studies, cognitive effort, clarity, ease to use and task difficulty are associated with workload patterns, interruption, and adequacy of tools (Gonçales et al. (2019); Meyer et al. (2021)). Experiments that focus on tools also indicate a tendency of ease of perception and learnability to be more dependent on tooling design than on experience (Guthardt et al. (2024)). Recent outcome oriented work operationalises DevX in flow, feedback loops and cognitive load, visualising the relationship between these perceptual factors and productivity, quality and retention at the level of developer, team and organisation (Forsgren et al. (2024)). The perceived productivity in different work conditions is also studied by Clustering analyses (Meyer et al. (2017)). Although they are collectively used to demonstrate the modern approaches to quantifying the impact of experience related variables on the performance outcomes, they focus more on productivity and software quality rather than on experience per se, which is why they are not similar to the experiential one of the present research.

Table 1: Relevant DevX studies and their covered dimensions

Study	**Dimension**	**Factors**	**Prediction**	**Clustering**	**Productivity**
Razzaq et al. (2025)	E,V,P	✓	×	×	✓
Greiler et al. (2023)	E,V,P	✓	×	×	×
Meyer et al. (2021)	E,P	✓	×	×	✓
Graziotin et al. (2018)	E,P	✓	×	×	✓
Girardi et al. (2022)	E,P	×	✓	×	✓
Meyer et al. (2017)	P	×	×	✓	✓
Forsgren et al. (2024)	E,V,P	×	✓	×	✓
Butler and Jaffe (2021)	E,V,P	✓	×	×	✓
Baldassarre et al. (2022)	E	×	×	×	×
Patnaik and Padhy (2024)	E	×	✓	×	✓
Sampson (2025)	E,V,P	✓	×	×	✓
Wassouf-Jr et al. (2025)	E,V,P	✓	×	×	✓
Godliauskas and Šmite (2024)	E,V	✓	×	×	✓

Summary and motivation for the present study. As outlined in Table 1, previous research has analysed emotional, value oriented, and perceptual experience aspects of developer experience individually and independently, without attempting to model how different DevX sources interact to influence them. This fragmentation is supported by recent holistic Developer experience research that observes that the experience of developers is the result of inter-relations between cognitive, affective, and conative factors that are characterized by workload, tooling, autonomy, and collaboration (Sampson (2025)). Razzaq et al. (2024) also find significant gaps especially in empirical studies on values and urge more multi-dimensional and integrated studies of DevX (Razzaq et al. (2025)). Although certain studies have used clustering to apply to solitary constructs like perceived productivity, joint clustering methods that examine several DevX dimensions along with the factors that affect them have not been extensively studied. These gaps inspire the current study, which researches the joint impact of a wide spectrum of DevX factors on emotional (E), value-oriented (V), and perceptual (P) dimensions of developer experience, and whether the clustering techniques can be used to identify specific developer profiles by this factors and dimensions.

3 Methodology

3.1 Research Questions and Analytical Workflow

The general objective of this paper is to establish the most influential factors in terms of key developer experience (DevX) dimensions like emotions, values, and perceptions basing on existing literature on DevX Forsgren et al. (2024); Greiler et al. (2023); Razzaq et al. (2025). Another objective is to estimate these relationships by supervised learning and to obtain meaningful profile of developers by unsupervised clustering. In order to address the research questions (RQ1-RQ3),To solve the research questions (RQ1-RQ3), the analysis utilizes a structured work flow comprising a number of modelling and evaluation tasks:

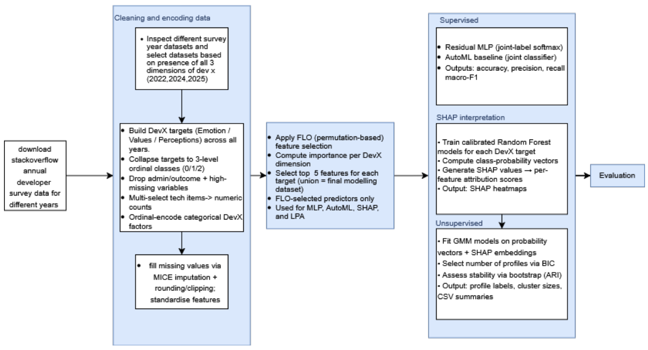


Figure 1: End-to-end analytical workflow from raw multi-year survey data to supervised DevX prediction and latent profile analysis.

1. **Variable construction and harmonisation.** The variables that are to be modelled are the three developer-experience dimension variables (emotional states, personal values, perceptions to base indicators), which are chosen based on the survey. The raw Stack Overflow survey has many items that are not similar in relevance, thus a preliminary variable categorisation step is undertaken. All of the questions are categorized into three groups, i.e. (i) DevX factor variables, (ii) DevX dimen-

sion variables and (iii) outcome variables (Razzaq et al. (2025)). The variables that are retained as targets are the DevX dimension variables and outcome variables are eliminated in this research.

2. **Preprocessing and feature selection.**

   All the retained variables were harmonised, coded into numeric formats and normalised to create a complete model ready feature matrix.

   There were cases of missing information that have been filled in with Multiple Imputation by Chained Equations (MICE) that maintains the multivariate structure and less biased compared to single-worth imputation van Buuren and Groothuis-Oudshoorn (2011). Following imputation, categorical and ordinal predictors were put back into their corresponding discrete scale.

   Model-agnostic feature selection techniques were used to determine which factors have the strongest effect on the three dimensions of developer experience (RQ1). Sought after in particular, were the permutation-based importance and SHAP (SHapley Additive explanations) value ranking, which were employed to derive robust estimates of the contribution each predictor makes across models (Lundberg and Lee (2017)). These complementary methods have the benefit of giving both global and local accounts of the effect of variables, which will guarantee a sound identification of the most significant DevX factors.

3. **Supervised prediction of developer-experience dimensions.** In order to know the extent to which the variables chosen are able to predict each of the DevX dimensions (RQ2), we train different supervised models, such as a feed-forward neural network and baseline with an AutoML. We then compare their performance based on the standard regression and classification metrics including accuracy, precision, recall, and F1 score.

4. **Unsupervised clustering to uncover developer profiles.** In order to answer RQ3, the latent profile analysis (LPA) is done based on Gaussian Mixture Models (GMMs). Two clustering spaces are analyzed (i) model-based predicted probability distribution over the DevX dimensions, and (ii) a composite explanatory space where the predicted probability distributions are combined with SHAP-based feature-attribition vectors. A number of GMMs of varying sizes are also fitted to each space and the best number of profiles identified by the Bayesian Information Criterion (BIC). Cluster stability is determined using bootstrap resampling along with the calculation of Adjusted Rand Index (ARI) of bootstrap clusterings with the original solution.

5. **Interpretation and profiling.** The resulting latent profiles are described by looking at the average predicted probabilities, SHAP attribution patterns and distribution of key factors variables within each the profile. The analyses are done to determine the coherent groups of developers that share common behavioural, perceptual, and emotional patterns with regard to their experience as developers. Figure 1 summarises the end-to-end workflow.

3.2 Dataset Characteristics

This analysis will be based on three publicly available public Stack Overflow Annual Developer Survey datasets in 2022, 2024, and 2025. The survey dataset for 2023 was excluded due to lack of presence of all the three DevX dimenional variables. Every dataset has tens of thousands of replies of developers around the globe, including demographics, work practices, use of technologies, learning behaviour, and even perceptions that are applicable to the experience of developers (DevX).

The sizes of the raw datasets of each of the survey years are presented in Table 2.

Table 2: Dataset sizes before preprocessing.

Year	Raw Rows	Raw Vars	Vars after Feature Selection
2022	73,268	79	16
2024	65,437	114	17
2025	49,123	170	16

3.3 Algorithms and Modelling Pipeline

The modelling pipeline is composed of a series of steps that include variable encoding and feature preparation, imputation and feature selection, supervised prediction, model explainability and latent profile analysis.

Encoding, Variable Filtering, and Feature Preparation: All the survey years were subjected to a harmonised encoding procedure. Column names were normalised and administrative or non-informative fields (`ResponseId`, `SurveyLength`, `SurveyEase`, `CompTotal`, `ConvertedCompYearly`, `Country`, `Currency`) were eliminated.

Multi-select technology fields like `LanguageHaveWorkedWith`, `DatabaseHaveWorkedWith`, `PlatformHaveWorkedWith`, `WebframeHaveWorkedWith` etc. were turned in to count features. The rest of the predictors were turned to ordinal integer coded variables,`OrgSize`, `RemoteWork`, `YearsCode`, to make sure that the final modelling dataset will contain only numeric variables.

The experience variables (`YearsCode`, `YearsCodePro`, `WorkExp`) were transformed into numeric values using explicit rules. Columns that have excessive missingness (>60%) and no variables that represent DevX factors. were dropped (e.g., `LanguageWantToWorkWith_Count`, `DatabaseWantToWorkWith_Count`, `PlatformWantToWorkWith_Count`).

Table 3 summarises the independent variables retained for modeling and their encoding transformations across survey years.

Variable Encoding and Rationale	**2022**	**2024**	**2025**
Binary Indicators (0/1): Survey yes/no items mapped to binary integers to simplify categorical responses and support SHAP interpretability.	ProfessionalTech TrueFalse_1 LearnCodeOnline	–	AIAgentChange NewRole PurchaseInfluence TechEndorse_1 TechOppose_1
Ordinal Scales: Numeric ordinal scoring maintains monotonic semantics for skills/attitudes and avoids information loss for SHAP.	Knowledge_1 Knowledge_3	Knowledge_1 Knowledge_3	AIAgents AIComplex AIHuman SO_Dev_Content
Ordinal Categorical: Categorical constructs (branch, role, organisation) encoded as ranked integers to avoid one-hot dimensionality and reflect hierarchy.	DevType MainBranch TBranch Onboarding OrgSize	OrgSizeOrd	Employment MainBranchOrd
Continuous(Numeric): Unscaled numeric values retained to preserve distribution and support non-linear splits in tree-based models.	Frequency_1 TimeAnswering TimeSearching	YearsCode	WorkExp AIAgent_Uses
Multi-Select Counts: Multi-response survey fields converted into count features to reduce dimensionality and preserve breadth of tooling exposure.	–	Count variables: AIChallenges, AIToolCurrently_Using, CodingActivities, DatabaseHaveWorkedWith, LanguageHaveWorkedWith, LearnCode, OfficeStackSyncHaveWorkedWith, ProfessionalTech, SOHow, ToolsTechHaveWorkedWith	–

Table 3: Unified encoding strategies for independent variables across the 2022, 2024, and 2025 datasets.

Construction of Target Variables: Three targets that matched the DevX dimensions which included emotions, values and perceptions, were identified per survey year and represented with three-level ordinal classes (0, 1, 2) to ensure consistency across years. **Emotions:** In 2024, `Frustration` was the emotions dimension modelled and rebinned 0-5, 0-2 scale, which is a count of items in the workflow, selected to indicate stressor. In 2025, `AIFrustration` (measures responses to the AI problems or frustrations experienced) question was normalised, counted, and collapsed under the same 02 scale; the variable was then standardised under the umbrella term Frustration. The 2022 representation of emotions was in terms of `MentalHealth`, a multi-select item, the valid selections of which were aggregated and summarized into three levels (0 = none, 1 = 12, 2 = 3).

Values: `JobSat` was used to express values across the years 2024-2025, with the 0-10 rating rebinned to 0 (0-4), 1(5-7), and 2(8-10). The equivalent value construct, which was already three-level Likert scale in 2022, was PurchaseInfluence which was directly mapped to 0-2.

Perceptions: `JobPerspectiveClass`, 2024 was modeled on five AINext items, one field was mapped to a 0-2 score of optimism, and the median was used to combine the scores. The procedure was repeated in 2025 on the AITool fields (e.g. AIToolCurrently mostly AI) to produce a harmonised JobPerspectiveClass. In 2022, the perceptions were assessed through BuyNewTool, which is a frequency variable condensed to 0-2.

All the targets were formatted as ordinal variables to ensure uniform treatment across the further modelling phases.

Construct Category / Encoding and Rationale	**2022 Targets (0/1/2)**	**2024 Targets (0/1/2)**	**2025 Targets (0/1/2)**
Emotion: Emotional load / strain measures collapsed into 3-class ordinal targets to capture affective intensity relevant for DevX.	MentalHealth Derived from medical/psychological condition count. 0 = none, 1 = mild (1-2), 2 = high (3).	Frustration Derived from frustration item count (0-5). 0 = low (<=1), 1 = medium (<=3), 2 = high (>3).	Frustration (original: AIFrustration) 0 = none (0), 1 = mild (1-2), 2 = high (>2).
Perception: Cognitive appraisal of influence and future job/tech prospects collapsed into 0/1/2 classes for comparability.	PurchaseInfluence Mapped directly: 0 = little/no influence, 1 = some influence, 2 = great influence.	JobPerspectiveClass Derived via median aggregation of AINEXT future-integration items. Rounded to 0/1/2 (pessimistic/neutral/optimistic).	JobPerspectiveClass Derived via median aggregation of AITOOL adoption levels. Rounded to 0/1/2 (low/mid/high).
Values: Preference, satisfaction and behavioural intention measures collapsed into tertiles for ordinal modeling.	BuyNewTool 0 = no intent, 1 = moderate, 2 = high intent.	JobSat Original 0-10 scale rebinned: 0 = <= 4, 1 = 5-7, 2 = 8.	JobSat (original: AISatisfaction) Same 0/1/2 rebinned rule as 2024.

Table 4: Construction, encoding and rationale of 0/1/2 target variables across the 2022, 2024, and 2025 datasets, categorised into emotion, perception, and values/intent domains.

Missing Data Handling. The values that were missing were filled in with Multiple Imputation by Chained Equations (MICE) which was run on scikit-learn as IterativeImputer with a Bayesian Ridge model. MICE does not bias the multivariate relationships in comparison to single-value imputations van Buuren and Groothuis-Oudshoorn (2011). Since MICE is used to generate continuous estimates, all imputed predictors were coded back to their ordinal or count basis form, and target variables were capped to the valid range of 02. No missing values were left out in the final datasets.

Feature Selection. FLO was employed since permutation based score drop is a stable model agnostic measure of the contribution that each predictor has on performance. Each DevX target was treated independently with LightGBM classifiers and the importance of predictors was determined as the average loss in balanced accuracy under repeat permutation. The top five features were picked out of the total features of each target and the combination of the features was taken as the final modelling feature subset.

Supervised Prediction Models. The supervised models make comparisons between a customized neural architecture and a general purpose AutoML baseline.

1. The main model was chosen as a **Residual MLP** with one joint-label softmax head since it can train a shared structure across all three DevX dimensions simultaneously, and as such, the relationships among the three dimensions are learnt in a single multi-target space. The number of possible combinations of the joint label is up to 27 and some of these combinations are rare, so SMOTE oversampling (Chawla et al. (2002)) and class-weighted loss were applied to counteract severe class imbalance and stabilise learning. Following the joint label SMOTE-based oversampling, the data were (stratified) split into 80% training, 10% validation, and 10% test sets.

2. A baseline of an AutoML was trained in a similar joint-label formulation. we performed split on 80/20 train/test to ensure that all joint classes are proportionally represented as they are multi-targeted. AutoGluon conducted its own model selection and ensembling internally to do validation. Nevertheless, Table 7 indicates that the AutoML models yielded very low performance over all years (macro-F1 <0.18), which may indicate that tree-based pipelines based on the generic model cannot handle the high-cardinality joint output space, as well as the non-linear nature of DevX factor interaction. These constraints also encourage the application of a specialised neural architecture that would make multi-target predictions.

Explainability: SHAP. To be interpretable, a calibrated Random Forest was applied to each DevX target, and SHAP was also implemented in TreeExplainer to calculate the global importances and instance SHAP vectors. These embeddings describe the effect of each of the features on the predicted DevX dimension of an individual.

Latent Profile Analysis (LPA). Gaussian Mixture Models (GMMs) were used to obtain latent profiles of developers based on (i) concatenated calibrated probability vectors of the Random Forest models (ii) combined SHAP attribution embeddings. The Bayesian Information Criterion (BIC) was used to determine model fit and bootstrap adjusted Rand Index (ARI) was used to determine the stability.

4 Design Specification

4.1 System Architecture

The system is structured into a modular pipeline as described below:

- **Input layer:** Harmonised 2022, 2024, and 2025 Stack Overflow datasets with numeric, ordinal-coded predictors and three DevX targets per year (emotions, values, perceptions).

- **Preprocessing layer:** The output of variable filtering, encoding, and MICE imputation is complete numeric design matrices.
- **Feature selection layer:** FLO was trained on per-target with LightGBM and the best-5 features of every target combined to form the final predictor subset.
- **Supervised modelling, explainability & profiling layer:** A joint-label softmax MLP residual is the main model, and a tree-based AutoML model serves as a baseline. Random Forest + SHAP is used to interpret model behaviour in terms of feature attributions, and then Gaussian Mixture Models are used to analyse the latent profile.
- **Evaluation layer:** Assesses predictive metrics, SHAP-stability as well as GMM-fit (ARI/BIC).

4.2 Key Algorithmic Design and Hyperparameters

Component	**Hyperparameters used**
MICE Imputation	IterativeImputer (BayesianRidge); max_iter = 20; random_state = 42.
FLO Feature Selection	Base model: LightGBM (n_estimators = 400, learning_rate = 0.05, subsample = 0.9); Permutation repeats = 3; scoring = balanced accuracy; top–5 features per target.
SMOTE Balancing	SMOTE (k_neighbors = 3; random_state = 42); applied to joint multi-class label.
Residual MLP (Joint Model)	Hidden size = 256; depth = 5; dropout = 0.3/0.2; AdamW (lr = 3e–4); batch size = 1024; epochs = 80.
AutoML Baseline	`AutoGluon TabularPredictor` with `presets="best_quality"` and `eval_metric="balanced_accuracy"`; time_limit = 3600s. Models trained include `NeuralNetFastAI_BAG_L1/L2`, `LightGBMXT_BAG_L1/L2`, and weighted ensembles (`WeightedEnsemble_L2/L3`); best model selected: `WeightedEnsemble_L3`.
SHAP (Calibrated RF)	RandomForest (n_estimators = 200); calibrated with isotonic regression; SHAP via TreeExplainer; sample size 1000.
Gaussian Mixture (LPA)	GaussianMixture (covariance = full; random_state = 42); candidate components k = 2–7; selected via minimum BIC.

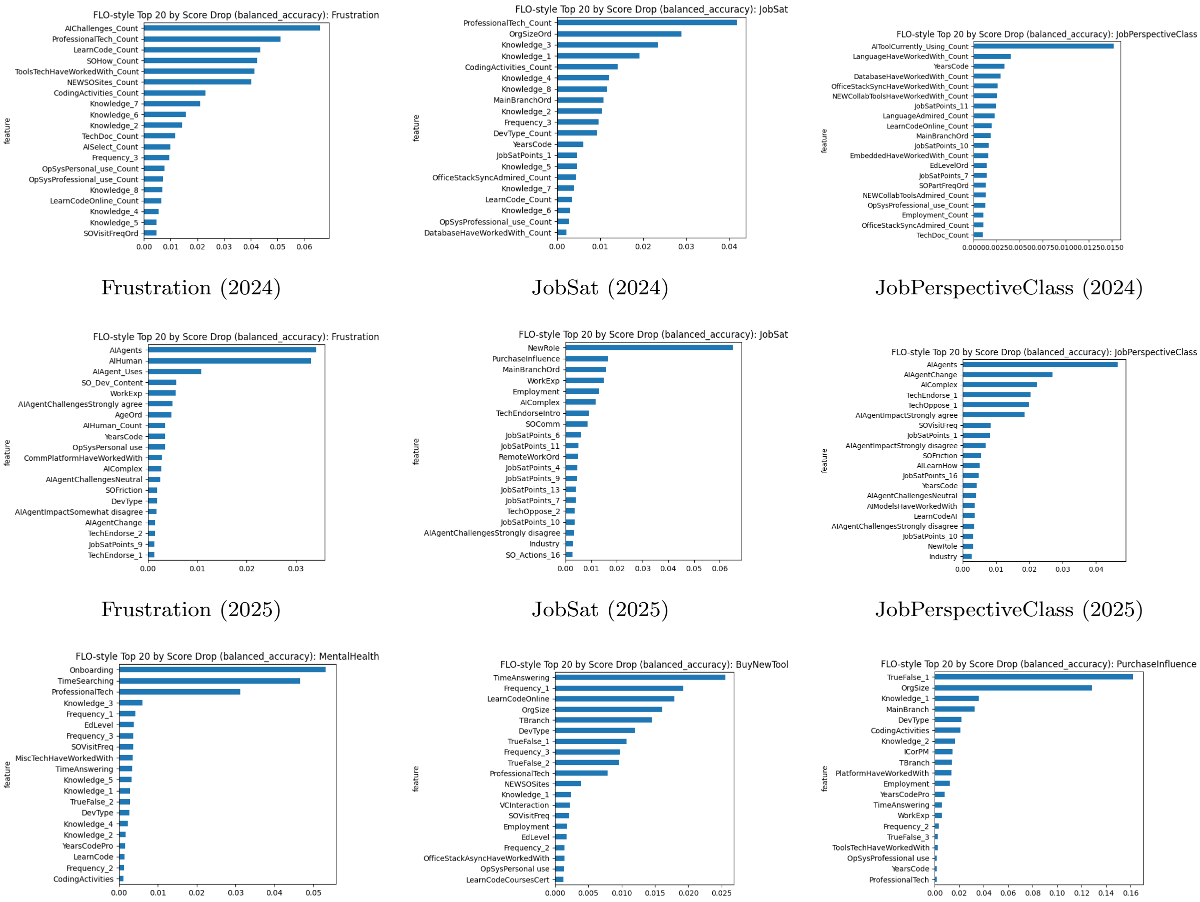
Table 5: Concise summary of key algorithmic components and dominant hyperparameters.

5 Implementation

This pipeline has been coded in Python with the help of `pandas`, `scikit-learn`, `PyTorch`, `LightGBM`, and `shap`, and was created in an entirely reproducible Google Colab environment. All the survey data were standardised into one numerical form and imputed, producing harmonised datasets (2022, 2024, 2025). All the artefacts necessary to carry out the analysis, such as supervised model results, clustering, and visualisations were created by the pipeline.

5.1 Feature Selection Outputs

The FLO-based permutation importance was carried out on a target-based basis individually per year of DevX. The implementation generated visual summaries of the 20 highest features of each target.



Frustration (2024) JobSat (2024) JobPerspectiveClass (2024)

Frustration (2025) JobSat (2025) JobPerspectiveClass (2025)

MentalHealth (2022) BuyNewTool (2022) PurchaseInfluence (2022)

Figure 2: FLO permutation-based feature importances across 2022, 2024, and 2025.

5.2 Supervised Model Outputs

The supervised modelling generated:

- a trained **Residual MLP (joint-label softmax)** which made joint predictions and decomposed per-target predictions on emotion, values, and perceptions;

Table 6: Joint multi-target prediction performance across survey years.

Year	Accuracy	Macro-F1	Precision	Recall	Balanced Acc.
2022	0.756	0.740	0.73	0.756	0.756
2024	0.946	0.943	0.944	0.946	0.946
2025	0.906	0.899	0.899	0.906	0.906

- a tuned **AutoML tree-based baseline** tree-based baseline, which was trained on the entire set of classification measures to compare the results (Table 7).

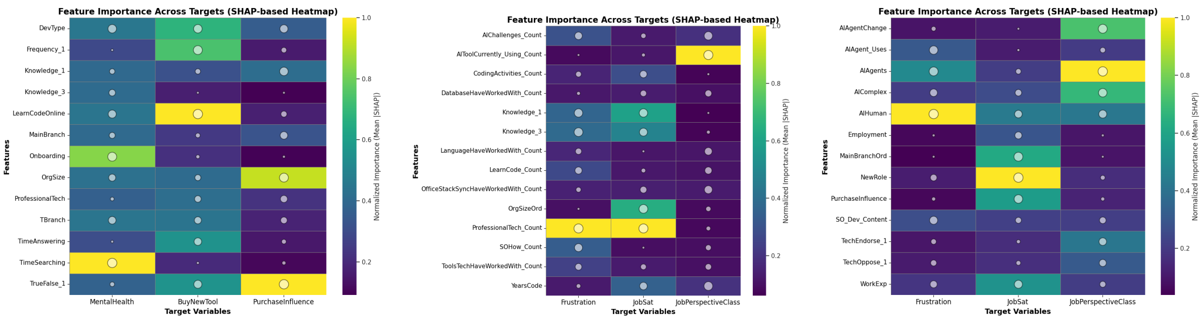
Table 7: AutoML baseline performance on joint multi-target DevX prediction.

Year	Accuracy	Macro-F1	Precision	Recall	Balanced Acc.
2022	0.2657	0.1706	0.1681	0.1976	0.1976
2024	0.3772	0.1185	0.1253	0.1415	0.1415
2025	0.4722	0.1695	0.1725	0.1722	0.1722

5.3 SHAP Explainability Outputs

In order to visualize the contribution of each DevX factor to predictions, we evaluated SHAP values of a TreeExplainer based on calibrated Random Forests. The implementation produced:

- SHAP-generated instance embeddings,which were applied in the clustering workflow
- heatmaps showing the difference in feature importance across years.

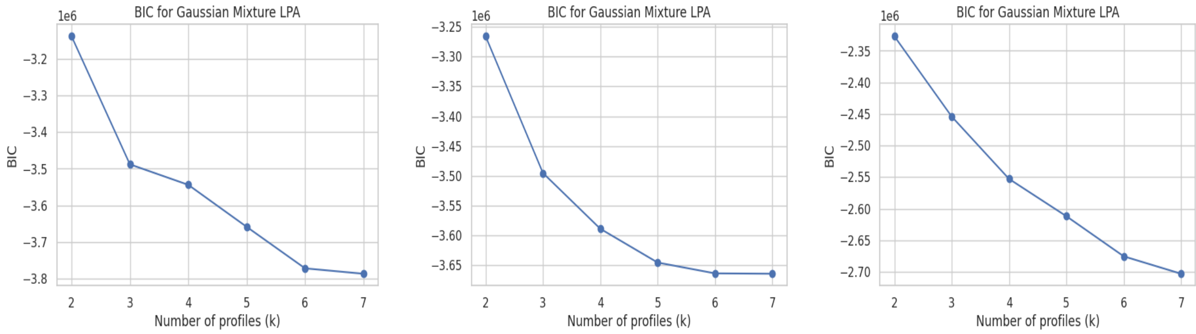


2022 2024 2025

Figure 3: SHAP feature-importance heatmaps across survey years.

5.4 Latent Profile Analysis Outputs

Gaussian Mixture Models were used in the Latent Profile Analysis (LPA) procedure applying to the predicted probability vectors of the supervised models, which gave a smooth, low-dimensional mixture modelling space. The results of every survey year yielded BIC curves to select a profile, ARI bootstrap stability scores, and profile summaries including group sizes, mean feature values and mean DevX classes probability distributions Fig. 4 . The process produced interpretable profiles of moderate stability, and was provided as CSV artefacts. Table 9 indicates profile and distributions of respondents. These outputs respond to RQ3 and establish individual developer subpopulations characterised by DevX factors and dimensions.



2022 2024 2025

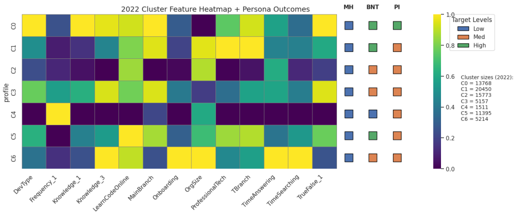
Figure 4: Gaussian Mixture BIC model-selection curves across survey years.

Table 8: Bootstrap stability of Gaussian Mixture LPA across survey years (Adjusted Rand Index).

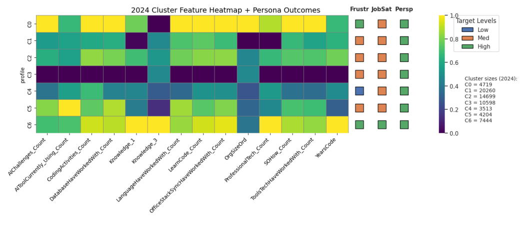
Survey Year	**Bootstrap ARI Scores**	**Mean ARI**	**Std. Dev.**
2022	0.914, 0.691, 0.451, 0.820, 0.602	0.696	0.162
2024	0.738, 0.589, 0.581, 0.830, 0.861	0.720	0.117
2025	0.627, 0.490, 0.963, 0.653, 0.661	0.679	0.155

Table 9: LPA cluster sizes across survey years.

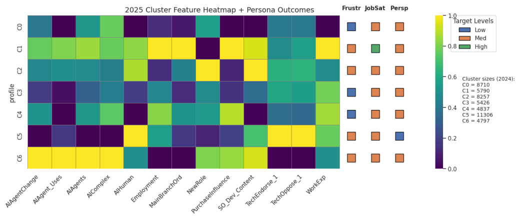
Profile	**2022 Size**	**2024 Size**	**2025 Size**
0	13768	4719	8710
1	20450	20260	5790
2	15773	14699	8257
3	5157	10598	5426
4	1511	3513	4837
5	11395	4204	11306
6	5214	7444	4797



2022



2024



2025

Figure 5: Cluster feature heatmaps and persona outcomes across survey years

6 Evaluation and Discussion

This section provides the findings of the empirical study that will answer RQ1-RQ3. Only the most pertinent results are provided with the use of visual overviews of FLO permutation importance, SHAP explainable, supervised model performance, and Gaussian Mixture latent profiling.

6.1 RQ1: Which variables most influence DevX dimensions?

To answer RQ1, we utilized FLO permutation importance that ranks features according to their predictive accuracy together with SHAP magnitude analysis (Figure 3) to determine which variables have the largest impact on emotions, values, and perceptions. across all survey years. Since SHAP heatmaps are based on absolute attributions, the results show how strong is the influence of each of the dimension variables.

2022 Findings: Both SHAP and FLO point to organisational and workflow factors as the most predictive factors. The main features are: time spent searching to find answers, time spent to answer questions, difficulty onboarding, and interaction with members outside an immediate team. The variables of knowledge-access (e.g. finding up-to-date information) and signs of organisational technical maturity (ProfessionalTech) are also highly important predictors of values and perceptions. The findings indicate that flow of information, collaboration load and organisational support structures are the main determinants of the 2022 DevX dimensions.

2024 Findings: In the case of 2024, the most important features shift toward *workflow breadth* and use of *AI-tool*. FLO and SHAP concur that variables that explain the diversity of coding activities, the volume of tools and languages utilised, organisational technical capabilities (ProfessionalTechCount), knowledge access variables and years of coding experience are some of the most powerful predictors in all of the DevX targets. Also, AI-related variables, like AIToolCurrently Using Count and AIChallenges Count, become salient, which indicates the shift towards AI-assisted development processes.

2025 Findings: The prevailing predictors by 2025 are even more biased towards AI-agent integration and career mobility. The most influential variables are: AIComplex (perceived capability of AI tools), AIAgentChange (changes as a result of AI adoption), AIAgents (agent use) and AIHuman (remaining human-dependence situations). Such features as work experience, role change (NewRole), and attitudes towards AI-enabled technology (TechEndorse 1) are also highly influential, particularly the values and perceptions.

Cross-year Summary: Combined, FLO and SHAP make up a consistent picture of the development of the DevX drivers.In 2022, knowledge flow and time pressure are the predominant sources of influence. By 2024 it is oriented towards tooling breadth and initial AI integration and by 2025, AI-agent ecosystems, role transition, and technology influence are most influential driving DevX. On the whole, these trends indicate that the DevX dimensions are the reflections of larger trends in the software practice and the faster integration of AI into the development processes.

6.2 RQ2: How Well Do Survey Variables Predict DevX Dimensions Using Supervised Machine-Learning Models?

In order to determine the extent to which the survey factors predict the three dimensions of the DevX (emotions, values and perceptions), we trained a joint-label Residual MLP and tested it on the three survey years. Since the model predicts joint DevX state (27 possible label combinations), its performance is based on how well the predictors tell the structure of overall developer experience as opposed to a single outcome.

Residual MLP (Joint Softmax): All the datasets indicate high and consistent performance of the MLP (Table 4), where accuracy varies between 0.76 (2022) and 0.95 (2024) and macro-F1 between 0.74 and 0.94. These findings indicate that the features that FLO selected provide significant signal on DevX, and that the joint modelling methodology is efficient in detecting common patterns among emotions, values and perceptions. The high results of 2024 and 2025 are indicative of the fact that more predictable by AI-rich survey items in the later years make DevX states even more predictable.

AutoML Baseline: In all three datasets, AutoML system (AutoGluon TabularPredictor) chose the best-performing model, which is the WeightedEnsemble L3 based on the balanced accuracy measure. Though the ensemble performed better than all individual AutoML candidates, its combined prediction accuracy was small (26–47%, Table 7). This can be forecasted by the fact that the task involves a separation of 27 potential joint label combinations, where tree-based models do not perform well in expressing cross-target relationships. The comparison of the Residual MLP with the AutoML baseline indicates that classical tabular ones are not capable of learning the shared structure between Emotions, Values, and Perceptions fully to justify the necessity of a joint neural model.

On the whole, the findings indicate that the DevX dimensions can be well predicted by survey variables, and the joint MLP can reliably recapitulate intricate DevX states. These results suggest that the chosen factors measure significant behavioural and organisational determinants of developer experience, that modelling the three dimensions of DevX comprehensively enhances predictive strength and predictability, and that FLO and SHAP accurately reflect variables that offer significant and reliable predictive evidence. Therefore, the following RQ2 is answered: survey variables, in particular, the variables related to the patterns of workflows, access to knowledge, and the use of AI-tools provide high predictive power regarding the emotional well-being, job values, and future perspectives of the developers.

6.3 RQ3: Developer Profiles from Latent Profile Analysis

Each of the calibrated probability vectors of the years (2022, 2024, 2025) were modeled with Gaussian Mixture Models. Figure 5 provides a comparison of the latent profiles across survey years, showing feature means and associated DevX persona outcomes for each profile. The consistent pattern of model selection based on the BIC was solutions with five to seven profiles, and bootstrap stability was moderate to strong with the mean ARI ranging between 0.69 and 0.72. These findings indicate that the specified profiles are based on the actual underlying trends, not merely a pure chance.

The LPA outputs across all three surveys (`LPA_full_profile_summary_2022/2024/2025.csv`) indicate a limited number of recurring developer personalities.

2022 Persona Profiles

Mental health was low on all profiles in 2022, whereas tool adoption and purchasing distinguished groups of clusters into four groups.

High-Adoption, High-Influence: This group demonstrated the most interest in embracing new tools and has great influence on purchase decisions. C0 comprises professionally oriented and collaborative developers in smaller organizations who learn actively and contribute to others.

High-Adoption, Moderate-Influence: High willingness to use new tools was observed in this group accompanied by moderate purchasing power. C1 includes developers of larger organizations that are organized in terms of their technical environment and learn online actively. C5 has structured and collaborative developers, who consider themselves to be developers and who learn online moderately accessing knowledge and interact across teams.

Moderate-Adoption, Moderate-Influence: This population was neutral in terms of new tool adoption and moderate in terms of purchasing power. C2 includes respondents in large organizations that depend on online learning extensively but lack onboarding support and experience low collaboration and access to internal information. C3 has developers who are very collaborative, highly identified in their roles, with high access to knowledge, and often peer assisted. C6 consists of respondents in more structured and larger organisations that have robust onboarding strategies, where access to information is high, and respondents are proactive in pursuing solutions and helping behaviours.

Low-Adoption, Moderate-Influence: This group was the most reluctant to use new tools but still had a moderate purchasing power. C4 represents participants who are low in terms of developer identification, collaborative effort, and low in terms of learning or support action, but have high external help needs, and work in medium-sized organizations with minimal structure in terms of technical support.

2024 Persona Profiles

Perspective in 2024 was consistently high across the clusters, whereas frustration and job satisfaction separated four outcome groups.

Strained Optimists: The group was very frustrated, had moderate job satisfaction and positive long-term perspective. C0 attracts skilled and tool-spread developers of larger organizations who are heavy online and community learners and work and learn through distributed knowledge streams.

Neutral Optimists: This category portrayed moderated frustrations and job satisfaction and looked at the future positively. C1 gets developers in small organisations with a low-level infrastructure who rely on external learning communities, C2 gets developers with broad exposure to tools and workflow and with stable information flow, C3 gets minimally tooled developers in small organisations with little community or internal support and C5 gets highly tooled and externally oriented developers with strong learning behaviour and with weaker internal information access.

Low-Strain Optimists: This group indicated low frustration, moderate job satisfaction and positive long-term outlook. C4 attracts middle-level developers in smaller companies that work on different teams but have relatively small infrastructure and reduced interaction with the community.

Intense Optimists: This group consisted of high frustration, high job satisfaction as well as the optimistic view on the long term. C6 obtains most senior and well-equipped developers with high access to knowledge, cross-team work and high use of AI tools despite working in small organizational settings.

2025 Persona Profiles.

There were four outcome groups in 2025 based on frustration, job satisfaction, and long-term perspective, which reflected the transition to AI-built-in development workflows, selective human-AI delegation, and new hybrid roles identities.

Low-Strain Practitioners: This group was low in frustration, moderate in satisfaction, and moderate in long-term outlook. C0 represents the group of cautious users that utilizes AI tools to a moderate extent and who maintain low developer identity and purchasing power. C3 attracts seasoned evaluators who can selectively use AI-based criteria but have poor developer identity and influence. C4 represents established practitioners, greater buying power, and average AI participation as well as traditional support.

Stable Integrators: This group conveyed moderated degrees of frustration, satisfaction and perspective. C2 represents regular AI users, who use agents as a part of workflows and do not use human assistance selectively, with low levels of purchasing power and a high level of career mobility. C6 is an indicator of profoundly AI-dependence practitioners who heavily utilize different support systems and are more technology decision-makers, albeit with poor developer identity.

Contented Power Users: This category was characterized by moderate frustration and high satisfactions as well as moderate long term perspective. C1 represents AI-intensive, highly self-identifying developers with extensive work experience, high endorsement and rejection sensitivity to AI capabilities, high purchasing power and stable employment.

Pessimistic Traditionalists: This group was moderately frustrated, moderately satisfied and had low long-term outlook. C5 are human-centered practitioners who have not adopted AI widely and prioritize AI features when either approving or denying the technologies but have low purchasing power, weak developer identity, and low mobility.

Cross-Year Observations

The profiles indicated a concise and decipherable developer experience structure across the three years. Although differentiation between the 2022 and 2024 clusters was primarily determined by the organizational context and dynamics of tool adoption, the 2025 profiles were characterized by the change to AI-mediated workflows, role identity and career mobility. The latent personas were under a constant state of experience gradients (e.g., adoption, satisfaction, influence, or outlook) despite these changes indicating that a limited number of archetype developers could be used to describe heterogeneity in DevX despite changes in technologies and practices.

7 Conclusion and Future Work

This thesis aimed to provide answer to the three research questions: (RQ1) which survey variables most strongly influence developer-experience (DevX) dimensions; (RQ2) how well these variables can predict emotions, values, and perceptions; and (RQ3) whether coherent developer profiles can be uncovered using unsupervised methods. The project was directed at creating a reproducible and data-driven pipeline, which could be used to model DevX using the 2022, 2024, and 2025 datasets of Stack Overflow.

The research incorporated imputation (MICE), permutation-based feature selection (FLO), joint multi-target prediction with a Residual MLP, SHAP-based model explainability, and Gaussian Mixture Models to find latent profiles. All the three research questions could be answered with the help of this pipeline. In the case of RQ1, a consistent base of organisational, behavioural, and AI-related variables appeared as the most significant causes of emotions, values and perceptions, a fact that happened to be repeated by previous studies that have suggested developer well-being. is influenced by the socio technical environment and work load pressures Fagerholm and Jürgen (2012); Razzaq et al. (2025); Greiler et al. (2023). In the case of RQ2, these factors facilitated quite precise multi-target predictions of the DevX dimensions using Residual MLP that were significantly better than those of a supervised AutoML baseline. In the case of RQ3, the clustering approach showed patterns of interpretable developer profiles that are AI-intensive practitioners, early career generalists, and strained but engaged developers showing that DevX patterns are not random but structured. These results can be implied in a number of ways. First, they demonstrate that DevX is modelable with minimal supervision and allows the systematic evaluation of the way developers feel, work and perceive modifications in their environment.

Second, the persistence of influential factors in the years suggests that some behavioural and organisational circumstances such as knowledge flow, time pressure, tooling breadth, and AI engagement, are continually involved in developer experience. Third, the specified profiles offer a basis of personalised interventions, benchmarking of organisations, and selective tooling approaches.

The study however has limitations. The pipeline is based on self-reported survey data which can lack a behavioural nuance and can be biased due to reporting.Variations in the structure of the survey between years necessitated harmonisation decisions that can have an effect on comparability. Although stable, the clustering method is restricted by the representational space adopted (probabilities and SHAP embeddings) and can fail to capture behavioural patterns of higher order that cannot be described by survey variables. This study can be expanded in various directions in future work. To begin with, multiple years of data in an actual longitudinal design. would enable the researchers to trace the change in emotions, satisfaction and perceptions of individual developers over time as opposed to handling it on a year by year basis. Second, the identified profiles could be confirmed in real organisations in future studies, analysing their association with outcomes (e.g. productivity, risk of burnout, or onboarding efficacy). This validation would provide enhanced practical relevance of DevX profiling. Third, more detailed data like activity logs of development tools or version control systems might be used to supplement survey results and provide a more efficient overview of developer life in the day-to-day. Lastly, the modelling pipeline herein discussed gives a foundation to practical implementation, which can be customised developer-support tools, enhanced onboarding systems, and specific interventions to at-risk developer groups. These guidelines would enhance the scientific knowledge and practical application of DevX analytics.

References

Attneave, F. (1962). Perception and related areas, *in* S. Koch (ed.), *Psychology: A Study of a Science, Vol. 4. Biologically Oriented Fields: Their Place in Psychology and in Biological Science*, McGraw-Hill, pp. 619–659.

Baldassarre, M. T., Caivano, D., Fucci, D., Romano, S. and Scanniello, G. (2022). Affective reactions and test-driven development: Results from three experiments and a survey, *Journal of Systems and Software* **185**(22).

Beecham, S., Baddoo, N., Hall, T., Robinson, H. and Sharp, H. (2008). Motivation in software engineering: A systematic literature review, *Information and Software Technology* **50**: 860–878.

Butler, J. and Jaffe, S. (2021). Challenges and gratitude: A diary study of software engineers working from home during COVID-19 pandemic, *Proceedings of the 43rd International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP '21)*, IEEE, pp. 362–363.

Chawla, N., Bowyer, K., Hall, L. and Kegelmeyer, W. (2002). Smote: Synthetic minority over-sampling technique, *J. Artif. Intell. Res. (JAIR)* **16**: 321–357.

Fagerholm, F. and Jürgen, M. (2012). Developer experience: Concept and definition, *2012 International Conference on Software and System Process (ICSSP)*, pp. 73–77.

Forsgren, N., Kalliamvakou, E., Noda, A., Greiler, M., Houck, B. and Storey, M.-A. (2024). Devex in action: A study of its tangible impacts, *Queue* **21**(6): 47–77.

França, C. A., Silva, F. and Sharp, H. (2018). Motivation and satisfaction of software engineers: A large-scale study, *IEEE Transactions on Software Engineering* **46**(2): 118–140.

Girardi, D., Lanubile, F., Novielli, N. and Serebrenik, A. (2022). Emotions and perceived productivity of software developers at the workplace, *IEEE Transactions on Software Engineering* **48**(9): 3326–3341.

Godliauskas, P. and Šmite, D. (2024). The well-being of software engineers: a systematic literature review and a theory, *Empirical Software Engineering* **30**(1): 35.

Gonçales, L., Farias, K., da Silva, B. and Fessler, J. (2019). Measuring the cognitive load of software developers: A systematic mapping study, *International Conference on Program Comprehension*, pp. 42–52.

Graziotin, D., Fagerholm, F., Wang, X. and Abrahamsson, P. (2018). What happens when software developers are (un)happy, *Journal of Systems and Software* **140**: 32–47.

Greiler, M., Storey, M.-A. and Noda, A. (2023). An actionable framework for understanding and improving developer experience, *IEEE Transactions on Software Engineering* **49**(4): 1411–1425.

Guthardt, T., Kosiol, J. and Hohlfeld, O. (2024). Low-code vs. the developer: An empirical study on the developer experience and efficiency of a no-code platform, *Proceedings of the ACM/IEEE 27th International Conference on Model Driven Engineering Languages and Systems*, MODELS Companion '24, Association for Computing Machinery, New York, NY, USA, pp. 856–865.

Kleinginna, P. R. and Kleinginna, A. M. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition, *Motivation and Emotion* **5**(4): 345–379.

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions, Vol. 30, Curran Associates, Inc.

Meyer, A. N., Barr, E. T., Bird, C. and Zimmermann, T. (2021). Today was a good day: The daily life of software developers, *IEEE Transactions on Software Engineering* **47**(5): 863–880.

Meyer, A., Zimmermann, T. and Fritz, T. (2017). Characterizing software developers by perceptions of productivity, *International Symposium on Empirical Software Engineering and Measurement*, pp. 105–110.

Parrott, W. G. (2001). *Emotions in Social Psychology: Essential Readings*, Psychology Press.

Patnaik, A. and Padhy, N. (2024). A hybrid approach to enhance software quality by sentiment analysis of developer, *Computing, Communication and Learning*, Springer Nature Switzerland, Cham, pp. 113–125.

Pinho, D., Aguiar, A. and Amaral, V. (2024). Cognitive patterns for developer experience, *Proceedings of the 29th European Conference on Pattern Languages of Programs, People, and Practices*, EuroPLoP '24, Association for Computing Machinery, New York, NY, USA, pp. 1–10.

Razzaq, A., Buckley, J., Lai, Q., Yu, T. and Botterweck, G. (2025). A systematic literature review on the influence of enhanced developer experience on developers' productivity: Factors, practices, and recommendations, *ACM Computing Surveys* **57**(1): 1–46.

Sampson, M. (2025). Enhancing developer experience: Burnout, resilience, and the digital environment in software engineering, *Proceedings of the 38th International BCS Human-Computer Interaction Conference*, BCS HCI '25, BCS Learning & Development Ltd, Swindon, GBR, pp. 369–374.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries, *in* M. P. Zanna (ed.), *Advances in Experimental Social Psychology*, Vol. 25, Academic Press, pp. 1–65.

van Buuren, S. and Groothuis-Oudshoorn, C. (2011). Mice: Multivariate imputation by chained equations in R, *Journal of Statistical Software* **45**(3): 1–67.

Wassouf-Jr, E., Fukuda, P. and Fontão, A. (2025). Investigating the developer experience of lgbtqiapn+ people in agile teams, *Proceedings of the IEEE/ACM 47th International Conference on Software Engineering: Software Engineering in Society*, pp. 43–54.