DOLPHIN: Interpretable Discovery of Exceptional Subgroups in Longitudinal Data

Chinmay Joshi[1,2], David Antony Selby[1,3], Ayush Patnaik[4], Sebastian Vollmer[1,3], and Gerrit Großmann[1]

[1] Data Science and its Applications Research Group, German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany
{chinmay_parag.joshi,david.selby,sebastian.vollmer,gerrit.grossmann}@dfki.de
[2] Saarland University, Saarbrücken, Germany
[3] Department of Computer Science, University of Kaiserslautern–Landau (RPTU), Kaiserslautern, Germany
[4] XKDR Forum, Mumbai, India
ayushpatnaik@gmail.com

Abstract. Longitudinal data capture repeated measurements of the same entities over time and appear across healthcare, economics, and the social sciences. Subgroup discovery finds groups of entities that move differently from the rest and describes what sets them apart, but its standard form targets one value per entity rather than a whole trajectory. We introduce DOLPHIN, a method for interpretable subgroup discovery in longitudinal data. DOLPHIN scores each trajectory by how far it departs from the population baseline. It then fits a pool of random forests to these scores, keeps a compact forest that predicts them well, and reads short Boolean rules from the forest's decision paths. Each rule names a subgroup through its conditions and through the trajectory that separates it from the baseline. We evaluate DOLPHIN on two economic panels, GDP per capita across countries from the World Bank and household income from the CMIE survey. The matched subgroups separate from the baseline by 0.50 to 2.42 standard deviations on the World Bank data and by more than twice the population median deviation for the strongest CMIE groups. Code is available at github.com/chinmayj23/Dolphin.

1 Introduction

Economic outcomes unfold over time. Income per head across countries, earnings across households, and employment across regions are observed repeatedly, and the structure worth studying lies in how these paths differ across the population. Two households can reach the same yearly income through steady wages or through a late move into capital income, and the two routes carry different implications for policy. The analyst wants the groups that move differently from the rest, together with a description of what sets them apart.

Subgroup discovery addresses this task directly. It searches for subsets of the data on which a target behaves unusually and describes each subset with the

covariates [25,43,20,1]. A description takes the form of a rule over the covariates, a conjunction of conditions such as a crude birth rate above 21 per 1,000 people together with internet use below 15 percent of the population.

Standard subgroup discovery targets a single value for each entity [20,28]. In longitudinal data, the target carries temporal dynamics, and a group stands out through the shape of its change over time, such as a steeper climb, a delayed recovery, or a flat path within a growing population. We therefore ask how to find groups whose entire temporal behavior is exceptional while describing each with a Boolean rule that a human analyst can interpret.

We propose DOLPHIN: Discovery Of Longitudinal Patterns with Human-Interpretable Rules. DOLPHIN scores each entity by how far its trajectory departs from a global baseline, regresses this score with a forest chosen from a pool of candidate random forests, and reads Boolean rules over the covariates from the forest's decision paths. It ranks the rules by the trajectory of the entities they select, tests each against random grouping, and returns a small, diverse set. Each rule defines a subgroup, a set of entities with a human-readable description and the trajectory that makes it exceptional.

We evaluate DOLPHIN on two large economic datasets, country-level GDP per capita from the World Bank [42] and household income from the CMIE survey [7], where the CMIE study contributes a large and recent micro-level benchmark for longitudinal subgroup discovery.

2 Related Work

Subgroup discovery. Subgroup discovery searches a dataset for subsets whose target differs from the whole and describes each subset with the covariates [25,43]. A quality measure combines subgroup size with target deviation, and an exhaustive or beam search maximizes it over conjunctions [20,2,18]. Early work used binary and categorical targets, and later work reaches numeric targets, from mean-based quality measures to dispersion-corrected and distributional ones [4,30], with open implementations that make the methods reusable [29]. Exceptional model mining keeps the short-rule description and widens the target from a single value to a local model such as a regression or a correlation structure [28,12]. A single search returns overlapping descriptions, and methods control this redundancy by weighted covering [26,18], by diverse beam search that leaves the data weights untouched [27,23], or by information-theoretic encoding [37]. DOLPHIN follows this line and sets its target to the temporal behavior of a longitudinal outcome.

Temporal targets and trajectories. Two communities study temporal heterogeneity. In pattern mining, subgroup discovery describes units whose temporal behavior stands out in streams and time series [21], for example households with unusual electricity use [22]. Temporal exceptional model mining extends this to dynamic targets, scoring how the dynamic Bayesian network of a subgroup departs from the population [6]. In statistics, group-based trajectory and growth mixture models fit latent classes of similar paths, and functional and longitudinal clustering group entities by curve similarity [34,35,38,11,15]. Rules have

been attached to these temporal targets directly, with SubgroupSEM reading subgroups whose latent growth curves differ from the population [24] and SDLD growing interaction trees that split longitudinal treatment-effect trajectories under targeted maximum likelihood estimation [45]. These methods fit a structural equation model, a dynamic Bayesian network, or a targeted estimator at every candidate, which grounds them in parametric or causal assumptions and limits the scale of the search. DOLPHIN differs from both in how it forms and checks rules, reading them from the decision paths of a random forest and testing each against equal-size random subgroups, where SubgroupSEM searches by beam and SDLD grows a single interaction tree. It scores each trajectory once against a baseline before the search, trading the parametric and causal guarantees of these methods for a descriptive heuristic that scales to many engineered features and large populations.

Rules from tree ensembles and interpretability. A random forest predicts accurately and keeps the learned structure implicit [5]. Post-hoc extraction methods read rules out of a trained ensemble, with RuleFit selecting node paths through a sparse linear model [16], inTrees pruning and ranking the paths of a random forest [10], Node Harvest selecting nodes through a constrained quadratic program [33], and SIRUS keeping the splits that recur across resamples [3]. Direct rule induction builds rules from the data without an ensemble, by sequential covering in RIPPER [9], by minimum-description-length search in Grab [14], and across the foundations of rule learning [17]. DOLPHIN takes the post-hoc route, reading the axis-aligned decision paths of a random forest fit to the trajectory-deviation score, where continuous subgroup discovery optimizes soft differentiable predicates by gradient [44]. Interpretability resists a single definition [31], and the case for models that are interpretable by construction takes the description that defines a group as the object a user inspects [40,41]. Post-hoc model explainers approximate a trained model or attribute its outputs [39,32,19,13]. Reading an exceptional path as a departure from a baseline links the task to anomaly detection [8].

3 Method

3.1 Problem and Notation

A longitudinal dataset observes the same entities at several time points. Let $\mathcal{I} = \{1, \dots, n\}$ index the entities and $\mathcal{T} = \{1, \dots, T\}$ the ordered times. Each entity $i \in \mathcal{I}$ has a target trajectory $\mathbf{y}_i = (y_{i1}, \dots, y_{iT_i})$ over its T_i observed times and explanatory covariates x_{it}^j [11,15]. DOLPHIN receives $\{(\mathbf{y}_i, x_i)\}$ and returns a small set of rules $\mathcal{R}^\star$. Each rule is a conjunction of conditions on the covariates, it selects a subgroup of entities, and the trajectories of that subgroup deviate from the population baseline. The covariates carry the rule language, and the target trajectory decides whether a subgroup is exceptional.

3.2 Feature Engineering

DOLPHIN reads the covariates through a fixed-length descriptor z_i built from each entity's covariate history. For every time-varying covariate it records the lagged value at each lag $\ell \in L$, the value ℓ periods before the entity's last observation. Over each window $w \in W$ it records three summaries, the mean of the covariate across the window, its net change from the start of the window to its end, and its standard deviation within the window, which describe the level, the direction of movement, and the variability of the covariate over a horizon of length w. Static covariates enter as their entity-level value, ordered categories become ordinal codes, and unordered categories become binary indicators. The lags and windows follow a fixed step in the time unit of the data, up to the largest length for which at least 80 percent of entities still hold enough history, so no feature depends on a horizon that few entities support.

3.3 Trajectory Deviation

DOLPHIN measures how far each entity's trajectory deviates from the population, and this measurement is what defines exceptional behavior. Entities differ in length and in sampling, so DOLPHIN places every trajectory on a common grid of size G by linear interpolation over its observed range, and keeps entities with at least $m_{\min}$ observed points and at most a fraction $\rho_{\max}$ missing so the interpolation rests on real data. It then mean-centers each trajectory, $\tilde{\mathbf{v}}_i = \mathbf{v}_i - \bar{v}_i \mathbf{1}$ with $\bar{v}_i = G^{-1} \sum_g v_{ig}$, which removes the absolute level so two entities with the same shape of change align regardless of where they sit. The population baseline is the average centered path $\boldsymbol{\mu} = n^{-1} \sum_i \tilde{\mathbf{v}}_i$, and the deviation score is

$$a_i = d_a(\tilde{\mathbf{v}}_i, \boldsymbol{\mu}),$$

with d_a a Euclidean or Wasserstein distance [36]. The score orders the entities by how unusual their trajectory is, and a threshold on it can already separate the exceptional entities from the rest.

3.4 Rule Generation

DOLPHIN regresses the above-mentioned trajectory deviation score with a random forest [5]. The random forest splits the covariate space along axis-aligned thresholds, each leaf collects entities of similar deviation, and every root-to-leaf path reads as a short conjunction such as $z_{i,p} \le c_1 \wedge z_{i,q} > c_2$ that marks a candidate subgroup [16,10,3]. To avoid a single arbitrary fit, DOLPHIN trains $M = 1000$ random forests under different seeds, each grown under one budget, a bounded depth $d_{\max}$, a minimum leaf size $\ell_{\min}$, a fixed number of trees, and a cap on the features per split, which holds every path short enough to read. For each one it records the held-out predictive performance P, the number of split tests C, and the leaf separation H between high-deviation and low-deviation entities. Among the forests that fit within a tolerance of the best performance,

$P \geq P_{\max} - \epsilon_P$, DOLPHIN prefers the one with the fewest split tests, and when several are equally simple it keeps the one whose leaves separate high- from low-deviation entities most cleanly. DOLPHIN then turns every root-to-leaf path into a rule, merges repeated tests on one covariate into the tightest interval, so $z_j > 2 \wedge z_j > 5$ becomes $z_j > 5$, drops contradictory paths, and keeps the rules with support $\text{supp}(r) = |S_r|/n$ in $[s_{\min}, s_{\max}]$.

3.5 Rule Scoring, Significance, and Selection

DOLPHIN scores each candidate rule by the trajectory of the subgroup it selects. For a rule r with subgroup S_r, it averages the centered trajectories of the matched entities into a subgroup path $\boldsymbol{\mu}_r = |S_r|^{-1} \sum_{i \in S_r} \tilde{\mathbf{v}}_i$ and measures how far that path sits from the population baseline, $D_r = d_r(\boldsymbol{\mu}_r, \boldsymbol{\mu})$. The quality $Q(r) = |S_r|^\alpha D_r$ combines this divergence with subgroup size, and the exponent α sets how much size compensates for a smaller divergence. DOLPHIN discards any rule with $D_r < \lambda \sigma_D$, where σ_D is the standard deviation of the entity deviation scores, so a subgroup whose path differs from the baseline by less than the natural spread is not reported.

A rule on few entities can reach a large divergence by chance, so DOLPHIN tests each surviving rule against random grouping. It draws 5,000 random subgroups of the same size and takes the permutation p-value as the fraction that match or exceed the rule, reporting a rule as strong when it stays significant on both the subgroup path and the entity-level deviation.

Rules read from one random forest cover overlapping entities, so DOLPHIN selects a small diverse set by weighted covering [26,18]. Every entity starts at weight $w_i = 1$, and DOLPHIN repeatedly adds the rule of highest weighted quality $Q_w(r) = (\sum_{i \in S_r} w_i)^\alpha D_r$, then multiplies the weight of every entity that rule covers by a factor $\gamma \in [0, 1]$. The decay makes already-covered entities count for less, so each new rule is rewarded for reaching parts of the population the chosen rules miss, and a small γ pushes toward disjoint subgroups while a γ near one allows overlap. DOLPHIN admits a rule only when it stays far enough from the chosen rules, with Jaccard overlap below τ_J and a subgroup path at distance above τ_D and correlation below τ_C, and stops once it holds k rules. Every returned rule carries its conditions, support, matched entities, subgroup trajectory, effect size, mean absolute change, and permutation p-value.

4 Case Studies

The evaluation uses two longitudinal economic datasets at different scales, summarized in Table 1. The World Bank data covers a few hundred countries at annual resolution, and the CMIE data covers several thousand households at monthly resolution, which tests DOLPHIN across population sizes and sampling frequencies.

Table 1: Summary of the two longitudinal datasets.

Dataset	Target	Entities	Observations	Periods	Frequency
World Bank	GDP per capita	220	13,981	64	annual
CMIE	Household income	9,316	281,393	40	monthly

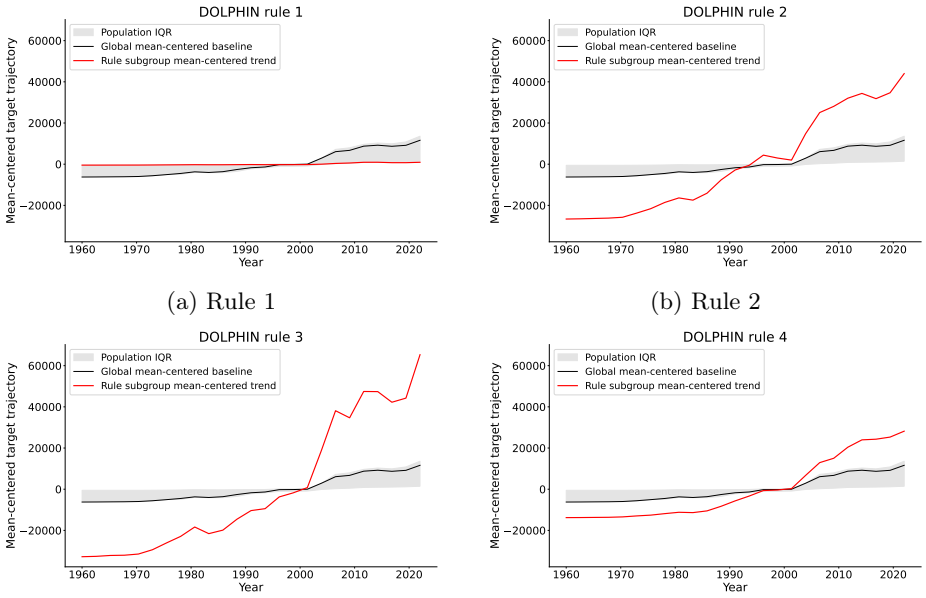
Table 2: DOLPHIN subgroups for World Bank GDP per capita. Conditions are covariate summaries over the stated window. Birth and death rates are per 1,000 people, internet use and trade in services are percentages of population and of GDP, D_r/σ_D is the baseline effect size, and p is the trajectory permutation p-value from 5,000 random subgroups.

	Conditions	Supp.	n	D_r/σ_D	p
R1	birth rate (10y mean) > 21.2 internet use (26y mean) ≤ 15.1 internet use (8y mean) ≤ 97.5 electrification volatility (32y) $> 1.7 \times 10^{-4}$	0.244	52	0.50	< 0.001
R2	death rate change (16y) > 0.33 internet use change (6y) ≤ 11.6 internet use (8y mean) ≤ 97.5 electrification volatility (32y) $\leq 1.7 \times 10^{-4}$	0.066	14	1.67	< 0.001
R3	internet use (10y mean) > 86.0 death rate volatility (6y) ≤ 0.19	0.023	5	2.42	< 0.001
R4	trade in services (12y lag) > 82.6	0.047	10	0.85	0.012

4.1 World Bank GDP per Capita

The first dataset is the World Bank World Development Indicators, a country-year panel of development statistics drawn from officially recognized international sources [42]. We choose GDP per capita as the target and compute it for each country as GDP in current US dollars divided by population. After removing aggregate regions the panel holds 220 countries, and 213 keep enough target history. The covariates cover demography, connectivity, infrastructure, and trade, and DOLPHIN builds lag and rolling-window features on a two-year step up to 32 years, which leaves 491 features after the availability filter. DOLPHIN returns four subgroups that separate countries by long-run development conditions, listed in Table 2 with trajectories in Figure 1.

The four subgroups map to recognizable development regimes. Rule 1 holds 52 countries with high birth rates and low long-run internet use in spite of stable access to electricity, including Ethiopia, Kenya, Mozambique, Nepal, Pakistan, and Uganda. Its panel in Figure 1a stays near zero and below the country baseline, so the group grows more slowly than the world, and its GDP per capita



(a) Rule 1

(b) Rule 2

(c) Rule 3

(d) Rule 4

Fig. 1: Mean-centered GDP-per-capita trajectories of the four DOLPHIN subgroups. Each panel shows the subgroup trend (red), the global country baseline (black), and the population interquartile range (shaded).

changes by about 1,382 US dollars over the grid against about 17,843 dollars for the baseline. Rule 2 holds 14 mature high-income economies whose populations have begun to age, whose connectivity has reached saturation with little recent change, and whose infrastructure has stayed stable for decades, including Austria, Canada, France, Germany, Japan, and Singapore. These economies developed early and grow consistently, and their panel in Figure 1b sits well above the baseline, rising by about 70,621 dollars. Rule 3 holds five small, very high-income economies that combine near-universal internet use with low mortality volatility, namely Iceland, Ireland, Luxembourg, Norway, and Qatar. Its panel shows the steepest climb, about 98,101 dollars, and the largest effect size in Table 2. Rule 4 holds 10 economies where trade in services makes up a high share of GDP, including Aruba, Ireland, Luxembourg, Malta, Singapore, and Tuvalu, and rises by about 41,983 dollars.

Three subgroups survive permutation tests. Rules 1, 2, and 3 differ from random subgroups of the same size on both the trajectory divergence and the entity-level deviation, each at $p < 10^{-4}$. Their mean absolute annual change in GDP per capita confirms the split, about 245 US dollars for the population against about 40 dollars for the flat Rule 1 and about 1,750 and 3,910 dollars for the

Table 3: DOLPHIN subgroups for CMIE total household income, with windows in months. Income shares are fractions of total income, D_r/σ_D is the baseline effect size, and p is the trajectory permutation p-value from 5,000 random subgroups.

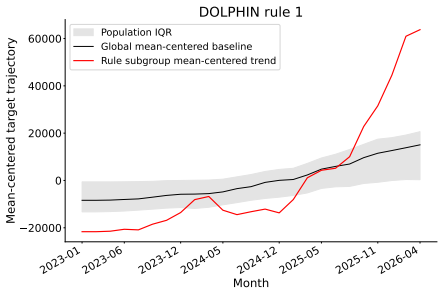
	Conditions	Supp.	n	D_r/σ_D	p
R1	capital share change (18m) $> 1.5 \times 10^{-3}$ transfer share (2m mean) $\leq 3.1 \times 10^{-3}$ wage share volatility (6m) > 0.19	0.016	147	0.57	< 0.001
R2	capital share change (4m) > -0.39 transfer share (6m mean) $\leq 2.2 \times 10^{-3}$ wage share volatility (6m) ≤ 0.32	0.163	1,505	0.36	< 0.001
R3	transfer share change (14m) ≤ -0.38 transfer share (18m lag) $> 7.3 \times 10^{-4}$ transfer share (2m mean) $> 3.1 \times 10^{-3}$	0.011	106	0.17	0.023

steep Rules 2 and 3. Rule 4 passes the trajectory test ($p = 0.012$) and does not separate on baseline deviation ($p = 0.27$), so it asks for caution when used.

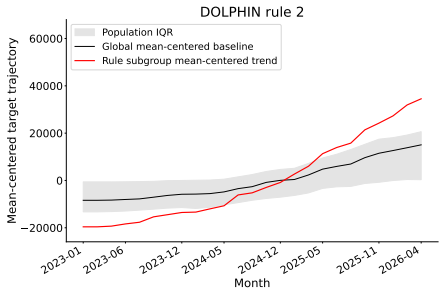
4.2 CMIE Household Income

The second dataset is the CMIE Consumer Pyramids Household Survey, a monthly panel of Indian households [7]. We choose total household income as the target. The panel holds 9,316 households, and 9,230 keep enough history. The covariates describe household composition and the share of income from wages, business, transfers, and capital, together with household size, employed members, education, and the dominant income source. DOLPHIN builds lag and rolling-window features on a two-month step up to 20 months, which leaves 253 features after filtering. The covariates exclude state and region as location descriptors. DOLPHIN returns three subgroups that separate households by income composition, listed in Table 3 with trajectories in Figure 2.

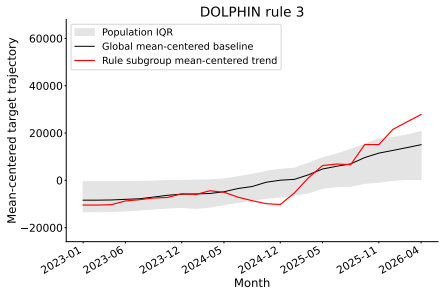
Income composition separates the household trajectories. Rule 1 holds 147 households whose capital-income share rose over 18 months, whose transfer share stays low, and whose wage share is volatile. Its panel in Figure 2a climbs farthest above the household baseline, rising by about 85,401 rupees against about 23,447 rupees for the baseline, the strongest movement among the three. Rule 2 holds 1,505 households with a stable capital-income share, a low transfer share, and steady wage composition, a broad group whose path rises by about 54,124 rupees. Rule 3 holds 106 households whose transfer share fell from an earlier positive level, which marks a move off transfer support. It rises by about 38,279 rupees and tracks the baseline more closely, and its smaller effect size in Table 3 records that.



(a) Rule 1



(b) Rule 2



(c) Rule 3

Fig. 2: Mean-centered total-income trajectories of the three DOLPHIN household subgroups. Each panel shows the subgroup trend (red), the global household baseline (black), and the population interquartile range (shaded).

Two rules are significant, one is not. On this large and recent household panel, Rules 1 and 2 separate from random subgroups of the same size with baseline-deviation $p < 10^{-31}$ for the focused Rule 1 and $p < 10^{-200}$ for the broad Rule 2. Their mean absolute monthly change in income rises from about 1,690 rupees for the population to about 4,440 and 3,790 rupees for Rules 1 and 2. Rule 3 differs only marginally in shape ($p = 0.023$) and does not separate at the entity level ($p = 0.78$), so it reads as a weak pattern.

5 Discussion

The two case studies show DOLPHIN recovering subgroups that hold together on development regimes on the World Bank panel and income-composition groups on the CMIE panel, each with a short rule and a separating trajectory. The design earns this at low cost, but reading the association rules and the temporal behavior from a mechanistic lens requires a domain study.

Two limitations follow from how DOLPHIN scores a trajectory. First, the deviation score summarizes the whole observed path in one number, so it measures whether the entire trajectory is exceptional rather than when the exceptional

behavior occurs. An entity that tracks the baseline for most of its history and departs only in its last few periods earns a moderate score, and that recent departure averages against the calm earlier stretch. DOLPHIN therefore reports entities whose full path is unusual, and it does not isolate a subgroup that turned exceptional only in a specific window such as the last ten years. Second, the score measures distance from the baseline, a magnitude that carries no direction. Two entities equally far from the baseline, one rising and one falling, earn the same score and can reach the same leaf. Hence, in rare cases, rule generation can place opposite trajectories in one subgroup. The post-hoc trajectory divergence catches some of these groupings, as the fourth World Bank rule and the third CMIE rule show, where a rule reads well in its covariates and passes the shape test but separates weakly at the entity level.

A score that localizes the deviation within the path, or that carries its direction, would move both checks into generation rather than leaving them to the post-hoc evaluator.

6 Conclusion

We set out to find the groups that move differently from the rest and to describe in plain terms what sets them apart, and we get both. On the World Bank panel the rules recover recognizable development regimes, and on the CMIE panel they separate households by the composition of their income, each subgroup arriving with the trajectory that makes it exceptional and a rule a reader can check against the covariates. A deviation is only a number until someone can say what it means, and the worth of an exceptional trajectory lies as much in the rule that names it as in the shape itself.

References

1. Atzmueller, M.: Subgroup discovery. WIREs Data Mining and Knowledge Discovery **5**(1), 35–49 (2015)
2. Atzmueller, M., Puppe, F.: SD-Map – a fast algorithm for exhaustive subgroup discovery. In: Knowledge Discovery in Databases (PKDD). LNCS, vol. 4213, pp. 6–17. Springer (2006)
3. Bénard, C., Biau, G., Da Veiga, S., Scornet, E.: Interpretable random forests via rule extraction. In: Proceedings of the 24th International Conference on Artificial Intelligence and Statistics (AISTATS). PMLR, vol. 130, pp. 937–945 (2021)
4. Boley, M., Goldsmith, B.R., Ghiringhelli, L.M., Vreeken, J.: Identifying consistent statements about numerical data with dispersion-corrected subgroup discovery. Data Mining and Knowledge Discovery **31**(5), 1391–1418 (2017)
5. Breiman, L.: Random forests. Machine Learning **45**(1), 5–32 (2001)
6. Bueno, M.L.P., Hommersom, A., Lucas, P.J.F.: Temporal exceptional model mining using dynamic Bayesian networks. In: Advanced Analytics and Learning on Temporal Data (AALTD). Lecture Notes in Computer Science, vol. 12588, pp. 97–112. Springer (2020)

7. Centre for Monitoring Indian Economy: Consumer pyramids household survey. https://consumerpyramidsdx.cmie.com/ (2026)
8. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: A survey. ACM Computing Surveys **41**(3), 1–58 (2009)
9. Cohen, W.W.: Fast effective rule induction. In: Proceedings of the Twelfth International Conference on Machine Learning (ICML). pp. 115–123. Morgan Kaufmann (1995)
10. Deng, H.: Interpreting tree ensembles with inTrees. International Journal of Data Science and Analytics **7**(4), 277–287 (2019)
11. Diggle, P.J., Heagerty, P., Liang, K.Y., Zeger, S.L.: Analysis of Longitudinal Data. Oxford University Press, 2nd edn. (2002)
12. Duivesteijn, W., Feelders, A.J., Knobbe, A.: Exceptional model mining: Supervised descriptive local pattern mining with complex target concepts. Data Mining and Knowledge Discovery **30**(1), 47–98 (2016)
13. Fischer, J., Oláh, A., Vreeken, J.: What's in the box? exploring the inner life of neural networks with robust rules. In: Proceedings of the 38th International Conference on Machine Learning (ICML). PMLR, vol. 139, pp. 3352–3362 (2021)
14. Fischer, J., Vreeken, J.: Sets of robust rules, and how to find them. In: Machine Learning and Knowledge Discovery in Databases (ECML PKDD). LNCS, vol. 11906, pp. 38–54. Springer (2019)
15. Fitzmaurice, G.M., Laird, N.M., Ware, J.H.: Applied Longitudinal Analysis. Wiley, 2nd edn. (2011)
16. Friedman, J.H., Popescu, B.E.: Predictive learning via rule ensembles. The Annals of Applied Statistics **2**(3), 916–954 (2008)
17. Fürnkranz, J., Gamberger, D., Lavrač, N.: Foundations of Rule Learning. Springer (2012)
18. Gamberger, D., Lavrač, N.: Expert-guided subgroup discovery: Methodology and application. Journal of Artificial Intelligence Research **17**, 501–527 (2002)
19. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., Pedreschi, D.: A survey of methods for explaining black box models. ACM Computing Surveys **51**(5), 1–42 (2018)
20. Herrera, F., Carmona, C.J., González, P., del Jesus, M.J.: An overview on subgroup discovery: Foundations and applications. Knowledge and Information Systems **29**(3), 495–525 (2011)
21. Hudson, D., Wiltshire, T.J., Atzmueller, M.: Local exceptionality detection in time series using subgroup discovery: An approach exemplified on team interaction data. In: Discovery Science (DS). LNCS, vol. 12986, pp. 435–445. Springer (2021)
22. Jin, N., Flach, P., Wilcox, T., Sellman, R., Thumim, J., Knobbe, A.: Subgroup discovery in smart electricity meter data. IEEE Transactions on Industrial Informatics **10**(2), 1327–1336 (2014)
23. Kalofolias, J., Boley, M., Vreeken, J.: Efficiently discovering locally exceptional yet globally representative subgroups. In: IEEE International Conference on Data Mining (ICDM). pp. 197–206. IEEE (2017)
24. Kiefer, C., Lemmerich, F., Langenberg, B., Mayer, A.: Subgroup discovery in structural equation models. Psychological Methods **29**(6), 1025–1045 (2024)
25. Klösgen, W.: Explora: A multipattern and multistrategy discovery assistant. In: Advances in Knowledge Discovery and Data Mining, pp. 249–271. AAAI Press (1996)
26. Lavrač, N., Kavšek, B., Flach, P., Todorovski, L.: Subgroup discovery with CN2-SD. Journal of Machine Learning Research **5**, 153–188 (2004)

27. van Leeuwen, M., Knobbe, A.: Diverse subgroup set discovery. Data Mining and Knowledge Discovery **25**(2), 208–242 (2012)
28. Leman, D., Feelders, A., Knobbe, A.: Exceptional model mining. In: Machine Learning and Knowledge Discovery in Databases (ECML PKDD). LNCS, vol. 5212, pp. 1–16. Springer (2008)
29. Lemmerich, F., Becker, M.: pysubgroup: Easy-to-use subgroup discovery in python. In: Machine Learning and Knowledge Discovery in Databases (ECML PKDD). LNCS, vol. 11053, pp. 658–662. Springer (2019)
30. Lijffijt, J., Kang, B., Duivesteijn, W., Puolamäki, K., Oikarinen, E., De Bie, T.: Subjectively interesting subgroup discovery on real-valued targets. In: IEEE 34th International Conference on Data Engineering (ICDE). pp. 1352–1355. IEEE (2018)
31. Lipton, Z.C.: The mythos of model interpretability. Communications of the ACM **61**(10), 36–43 (2018)
32. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems 30 (NeurIPS). pp. 4765–4774 (2017)
33. Meinshausen, N.: Node harvest. The Annals of Applied Statistics **4**(4), 2049–2072 (2010)
34. Nagin, D.S.: Group-Based Modeling of Development. Harvard University Press (2005)
35. Nagin, D.S., Odgers, C.L.: Group-based trajectory modeling in clinical research. Annual Review of Clinical Psychology **6**, 109–138 (2010)
36. Peyré, G., Cuturi, M.: Computational optimal transport. Foundations and Trends in Machine Learning **11**(5–6), 355–607 (2019)
37. Proença, H.M., Grünwald, P., Bäck, T., van Leeuwen, M.: Robust subgroup discovery. Data Mining and Knowledge Discovery **36**(5), 1885–1970 (2022)
38. Ramsay, J.O., Silverman, B.W.: Functional Data Analysis. Springer, 2nd edn. (2005)
39. Ribeiro, M.T., Singh, S., Guestrin, C.: "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 (KDD). pp. 1135–1144. ACM (2016)
40. Rudin, C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence **1**(5), 206–215 (2019)
41. Rudin, C., Chen, C., Chen, Z., Huang, H., Semenova, L., Zhong, C.: Interpretable machine learning: Fundamental principles and 10 grand challenges. Statistics Surveys **16**, 1–85 (2022)
42. World Bank: World development indicators. `https://databank.worldbank.org/source/world-development-indicators` (2024)
43. Wrobel, S.: An algorithm for multi-relational discovery of subgroups. In: Principles of Data Mining and Knowledge Discovery (PKDD). LNCS, vol. 1263, pp. 78–87. Springer (1997)
44. Xu, S., Walter, N.P., Kalofolias, J., Vreeken, J.: Learning exceptional subgroups by end-to-end maximizing KL-divergence. In: Proceedings of the 41st International Conference on Machine Learning (ICML). PMLR, vol. 235, pp. 55267–55285 (2024)
45. Yang, J., Mwangi, A.W., Kantor, R., Dahabreh, I.J., Nyambura, M., Delong, A., Hogan, J.W., Steingrimsson, J.A.: Tree-based subgroup discovery using electronic health record data: heterogeneity of treatment effects for DTG-containing therapies. Biostatistics **25**(2), 323–335 (2024)