2. Research Proposal

2a. A description of the proposed research

2a1. Overall aim and key objectives

Mining for Minorities (M4M) automatically discovers underserved minorities in the data that is currently being amassed from controlled experiments. Such experiments are a cornerstone of science. The typical controlled experiment, in which we have several variants between which we must choose, ends in selecting the best variant and discarding the others. This final step is unnecessarily wasteful. Minorities may be better served by a discarded variant; M4M avoids throwing such minorities under the bus.

The simplest version of a controlled experiment is A/B testing [KohLSH2009, SirK2013, KohL2016], encompassing two variants: variant A is currently in use, while variant B is newly developed. Further ingredients of A/B testing are a pool of test subjects, and a measurement of success. Each test subject is randomly assigned to either variant A or B, and the degree of success is measured. Aggregating the success measurements per variant, we can assess whether variant A or B works best. The winning variant is kept, and the losing variant is discarded.

In the internet era, A/B testing has become extremely popular. Since many companies, website owners, pharmaceutical developers, etcetera run A/B tests all the time, much test data is available. Typically, we have more data than just the A/B variant and the positive/negative effect: of a website visitor clicking a button on a website, we also learn their OS and browser version, country, browser screen height is, etc.; of a patient treated with a medicine, we survey all kinds of demographic information. This additional information can potentially be used to identify subgroups that are better served by the variant that loses in the overall population. If the general population is better served by medication A, but a sizeable minority is better served by medication B, then we owe it to the general public to identify these minorities.

Data varying over time

Traditional data mining methods assume data in *Cross-Sectional* (CS) form: at one point in time, we make a number of independent observations, and this is our dataset (see Table 1 for an example). Modern controlled experiments often result in data varying over time, as illustrated in Figure 1; time proceeds along the x-axis, while the y-axis is an unspecified measurement. Here, CS represents data such as in Table 1: a set of people measured exactly once. Three forms of data collection over time can be distinguished.

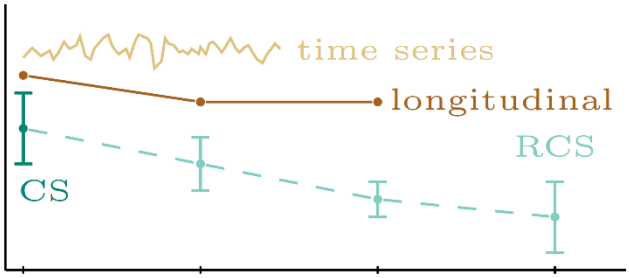


Figure 1: datasets of various types

In *time series*, multiple measurements are sampled per case with very short intervals: think of patients wearing blood pressure sensors. In *longitudinal* data, the same occurs with long intervals in a relatively long period of time: think of post-operative patients who repeatedly fill out surveys on their own well-being. *Repeated Cross-Sectional* (RCS) data is collected from new samples at each measurement occasion, resulting in varying sample sizes: think of yearly survey of cohorts of students of the same age (so across measurements, the age remains the same, but a different number of different students are surveyed). To appropriately serve minorities, M4M should make statements about group behavior over time; if a group is better served with an alternative variant in a certain period of time, we ought to know.

I propose an Exceptional Model Mining (EMM) generalization that automatically finds coherent subgroups displaying exceptional behavior in controlled experiment data, which allows to appropriately serve minorities. To do so, in the Mining for Minorities project, I will generalize EMM to data that varies over time. EMM requires efficient search strategies: the main scientific challenge of M4M is to provide proofs and derive bounds necessary to enable efficient search strategies on time-varying data.

Running example: A/B testing on a toy dataset

Suppose we collected the following data on a patient population, in an A/B test with two medication variants for the same disease, with either a positive or a negative effect:

Table 1: example A/B test dataset

ID	Age	Smokes?	Gender	Country	Region	…	Variant	Effect
01	22	no	M	NLD	Benelux		A	+
02	46	no	F	BEL	Benelux		B	+
03	24	yes	M	LUX	Benelux		A	-
04	25	no	M	ESP	Mediterranean		B	-
05	29	yes	F	ITA	Mediterranean		A	-
06	45	yes	F	FIN	Nordic		B	+
07	63	yes	M	SWE	Nordic		A	-
08	36	yes	M	NOR	Nordic		B	+
09	23	no	F	ESP	Mediterranean		A	+
10	50	yes	F	LUX	Benelux		B	-

The final decision derives from the cross table gathering all data on the right-hand side of the bold line, as illustrated in Table 2. The traditional A/B test concludes the following: variant A has success rate 40%, while variant B has success rate 60%. Since variant B performs better, we all use variant B from now on. In a traditional A/B test, what is best for the majority will then be prescribed to the entire population. This is unnecessarily wasteful, and runs the risk of harming minorities.

Table 2: cross table for entire dataset

		Effect	
		-	+
Variant	**A**	3	2
	B	2	3

Table 3: cross table for subgroup "Smokes?=no"

		Effect	
		-	+
Variant	**A**	0	2
	B	1	1

Table 4: cross table for subgroup "Region=Nordic"

		Effect	
		-	+
Variant	**A**	1	0
	B	0	2

In almost all controlled experiments, we have more information than what is available on the right-hand side of the bold line in Table 1. We can use such information, for instance the demographics in Table 1, to investigate the cross-table for subgroups of the dataset. Tables 3 and 4 are the cross tables for the subgroups "Smokes?=no" (covering the rows 01, 02, 04, and 09) and "Region=Nordic" (covering the rows 06, 07, and 08).

Genetic and environmental circumstances influence how people react to medication, which makes it unlikely that the full population reacts homogeneously. Table 3 displays an example of a subgroup that is better served by medication A, even though the general population is better served by medication B. By contrast, Table 4 displays an example of a subgroup that is substantially better served by medication B; the effect is stronger than in the overall population.

EMM [LemFK2008, Dui2013, DuiFK2016] is a data mining framework seeking subgroups in a dataset. Subgroups are interesting if they satisfy two constraints. On the one hand, subgroups must be interpretable: we must be able to define them as a conjunction of a few conditions on attributes of the dataset. In Table 1, we are not interested in the subgroup consisting of rows 01, 02, 04, and 08, but we might be interested in the subgroup defined as smokers under the age of 30 (encompassing rows 03 and 05). This ensures that found subgroups are actionable: if we have a concise interpretable definition of a subgroup, we can base a policy on it. On the other hand, subgroups must be exceptional: they must display some kind of unusual interaction between columns preselected as targets. Behavior between the targets "Variant" and "Effect" in Table 1 can be gauged in terms of their association [DuiFK2016, DuiFPPWAFP2017] which can be derived from the cross table in Table 2, and subgroups displaying exceptional association are illustrated in Tables 3 and 4.

State-of-the-art and open problems: efficient EMM algorithms for CS data

Even on Cross-Sectional data, the EMM search space is vast. Depending on the prevalence of attributes with many distinct values and expressiveness of the subgroup definition language, the search space can be exponential in the number of both attributes and observations. On top of this expensive generation of candidate subgroups, comes the computational cost of evaluating the exceptionality of each subgroup.

There are two approaches to traversing this vast search space. Heuristic search algorithms presume that it makes no sense to find the absolutely highest possible exceptionality value in a subgroup if its definition becomes unwieldy: who understands a conjunction of 65 conditions? Hence, heuristic search actively restricts exploration to part of the search space. Optimality may be lost, but interpretability is guaranteed. Also, heuristic search can cope with a mixture of attribute types and any desired kind of exceptionality. Exhaustive search algorithms accept only the optimal answer. Some structure must be found allowing to discard parts of the search space while guaranteeing that this discarded space contains nothing of interest; hence, we passively restrict exploration to only the relevant part of the search space. This can be done, but typically only under a restricted problem definition: numeric attributes are disallowed, or a limitation is imposed on the types of exceptional behavior. Within this restricted setting, the algorithm guarantees finding the best answer.

The standard heuristic search algorithm is *Beam Search* [DuiFK2016], a multi-pronged greedy approach: it restricts the number of conditions allowed to conjoin, and on each level, it refines a select number *w* of promising subgroups to generate the next-level candidates. Beam search is a baseline that works with any EMM dataset and any chosen type of exceptional behavior; if all else fails, we can still find minorities with these algorithms. An alternative is *Monte-Carlo Tree Search* (MCTS) [BosBRK2017, MatNBK2021]. Its applicability depends on the derivation of Upper Confidence Bounds in the "Select" phase, which works for the WRAcc quality measure (and hence for a single, binary target variable); its generalization to other modes of exceptionality in CS data is an open problem, as is generalization beyond CS data. Finally, one can employ *Pattern Sampling* (PS) methods [BolLPG2011, BolMG2012, MoeB2014, GiaS2018]. By shifting sampling weights towards observations that are more likely to contribute to exceptional models, CDPS promises to deliver interesting subgroups quickly with high probability. It does so quite successfully for generic pattern mining methods, but within EMM appropriate sampling weights have only been derived for two kinds of exceptional behavior (*model classes*) specifically invented for this purpose [MoeB2014]; generalization to other model classes on CS data, and generalization beyond CS data, are both open problems.

In exhaustive search, most work focuses on *Optimistic Estimates* [GroRW2008]: given a kind of exceptional interaction, can we define an upper bound on the exceptionality of all refinements of the current candidate subgroup? If so, and if this bound does not surpass the quality of the best subgroups already found, we can prune the search tree. This can dramatically speed up the algorithm, but for each new type of exceptional target interaction, deriving these bounds is an open problem. For EMM, *GP-Growth* [LemBA2012] defines a condensed representation of subgroup exceptionality, stored in a tree. By propagating this information through the tree, exceptionality bounds can be derived without traversing the original dataset, improving algorithm efficiency. When the bounds do not surpass a certain quality threshold achieved in earlier evaluated subgroups, the search tree can be pruned. For each new type of exceptional target interaction, deriving the condensed representation and the bounds is an open problem. GP-Growth works only if the condensed representation can be computed with a linear-time algorithm with sublinear memory demands, ruling out more complex types of exceptional target interaction including Bayesian networks [DuiFK2016]. Alternatively, optimistic estimates can be used in *Significant Pattern Mining* [Web2007], to efficiently find subgroups in EMM concerning voting data [BelDPCL2019]. This enables application of EMM on datasets where many values are missing, which is often the case in real-world scenarios.

Recent work delivers the best of both worlds: *Refine&Mine* [BelBK2018] is an anytime algorithm with guarantees in Subgroup Discovery on numeric data. When this algorithm gets infinite time, it provides the optimal subgroup. When pressed for time, we can interrupt the algorithm: it provides the best answer found so far, plus a bound on how far removed from the optimum this answer can possibly be. This very promising (and ECMLPKDD 2018 Best Data Mining Paper Award winning!) work is theoretically wonderful, but the limitation to a single target and to numeric-only data disallows its deployment on many real-life datasets: generalization it to datasets with a mixture of attribute types, more complex types of exceptional target interaction, and beyond CS data, are all open problems.

Why do alternative state-of-the-art methods not suffice?

If we would have a specific single minority to investigate –for instance: the effect of smoking on the efficacy of medications A and B– traditional statistical methods would suffice: a single hypothesis is to be tested, and an old-fashioned t-test would do the trick. Generally, we do not know beforehand which minorities would be better served by an alternative variant. Note that the dataset from Table 1 is a toy example; real-life datasets easily record tens to thousands of properties (columns) over a quarter million of observations (rows). We cannot afford to manually formulate and test hypotheses on datasets of such dimensions, especially since the number of candidate subgroups explodes exponentially with the number of columns: we can define not only the subgroup "smokes?=no", but also "smokes?=no AND age<30", "smokes?=no AND age>30", "smokes?=no AND age<40", "smokes?=no AND age<40 AND gender=M", etcetera. Exploring this exponential search space to find minority subgroups best served by an alternative variant can only be done by EMM [DuiFK2016].

In a medical setting, the polar opposite to the simplistic final decision as made in an A/B test would be to move to personalized medicine [ManCS2000]. Based on an expensive analysis of a single patient's DNA, one could tailor a suggested treatment to the specific patient at hand. This works well for individuals but doesn't scale to populations.

In the example of Table 1, traditional A/B testing would conclude: variant B performs better, so everyone will from now on get variant B. Personalized medicine would draw the conclusion: "based on the DNA of the person with ID 01, we recommend that he is prescribed variant A. Based on the DNA of the person with ID 02...". M4M would draw the conclusion: for the overall population, we prescribe variant B (since Yule's Q [Yul1912] for Table 2 is 0.3846; positive values for Q imply B works better). For the exceptional subgroup "Smokes?=no", however, we prescribe variant A (since Yule's Q for Table 3 is -1). The exceptional subgroup "Region=Nordic" must definitely be prescribed variant B, since Yule's Q for Table 4 is +1; the effect is extra strong in this subgroup. I published a first proof-of-concept paper that incorporates Yule's Q for finding exceptional minorities in A/B testing data with a binary reward function under the name A&B Testing [DuiFPPWAFP2017]. It only works for binary target columns and CS data; generalization over either of these axes is an open problem.

This last example also illustrates how traditional statistical methods do have a place in this project: there is ample potential of synergy between statistics and M4M. When we have fixed a particular kind of target interaction (called *model class* in EMM), and when we are running our search algorithm to explore the exponential search space of subgroups, it makes sense to base the quality measure evaluating these subgroups on solid statistics. For instance, for the correlation [LemFK2008] and rank correlation [DowD2017] model classes, quality measures exist based on the Fisher z transform [NetKNW1996]; for the regression model class [DuiFK2012], parts of the search space can be pruned based on bounds on Cook's distance [Coo1977, CooW1980, CooW1982]; the aforementioned Yule's Q [Yul1912] has been employed for A&B testing [DuiFPPWAFP2017]; the multiple comparisons problem [HocT1987] is relevant for any model class, and tackled in the context of significant pattern mining [Web2007] by a permutation test [DuiK2011] based on the Central Limit Theorem [Lya1901].

2a2. Research plan

Fundamental Challenges [Shared by all]

We will generalize EMM to data that varies over time. We will focus on the three kinds of time-varying data illustrated in Figure 1: time series, longitudinal data, and RCS data. Each of these data types comes with its own challenges, which means that each of the subsequently described challenges will need to be solved for each time-varying data type.

Immediate impact in adjacent scientific fields and society can be had by deploying M4M on data from controlled experiments. Therefore, it stands to reason to investigate the design of more modern controlled experiments, beyond the simple A/B test of our Running Example. We will develop EMM algorithms for time-varying A/B test data with two variants and binary rewards (**FC01**), with more variants in an A/B/C/D/E test (**FC02**), with non-binary values in the "Effect" column in Table 1 (**FC03**), incorporating significance when assessing the difference in success between variants (A/B/"too close to call") (**FC04**), comparing the observed differences with the results of a baseline A/A test [Pet2004] (**FC05**). All these experiments test variants in only a single column of the dataset; combining multiple

columns of variants in Multi-Variate Testing (MVT) [Kos1996] allows us to learn interaction effects. We will develop EMM algorithms incorporating a conscious choice in which subset of variant combinations should take part in an MVT test, through Plackett and Burman designs [PlaB1946] (**FC06**), fractional factorial designs [DavH1950] (**FC07**), and Latin Hypercube Sampling [Ye1998] (**FC08**). A sound pick of MVT design allows to estimate single-factor effects (which an A/B test can also do) and interaction effects (which an A/B test cannot); we will define an informative way to report subgroups, some of which have unusual individual effects, while other feature unusual interaction effects (**FC09**). Finally, we will develop EMM algorithms that can properly analyze Contextual Multi-Armed Bandits (cMAB) [LanZ2007] (**FC10**): cMABs allow to change over time the proportion of visitors assigned to variants, to reduce revenue loss due to unsuccessful variants. Care must be taken to not draw premature conclusions.

All these changes in the A/B testing setup can be combined to generate more interesting versions, and all resulting tests require development of new EMM model classes; we will define these new forms of exceptionality, and subsequently derive new bounds and formulate new proofs.

Existing EMM methods typically assume CS data: observations are independent transactions, and the entire dataset can be overseen. In modern complex systems these assumptions are not accurate: website visitors perform multiple actions in sequence, which influence each other; visitors may never cease interacting with the system, and we never know for sure when a visitor has left. This raises fundamental questions. We will redefine what a subgroup can be when the data contains dependence between observations (**FC11**). We will define over what scale of interactions what kind of exceptional behavior must occur, for a subgroup to be deemed exceptional (**FC12**). We will define statistically sound quality measures for existing model classes when independence between observations does not hold (**FC13**).

Challenge 1: heuristic EMM algorithms for time-varying data [Executed by PhD 1 and PI]

PhD 1 will focus on heuristic algorithms. They will generalize beam search to tackle the A/B test expansions of FC 01-05 (**C1a**), and the cMABs of FC 10 (**C1b**). They will derive new Upper Confidence Bounds for MCTS in time-varying data (**C1c**). They will derive appropriate sampling weights for pattern sampling in time-varying data (**C1d**).

PhD 1 can be inspired by a recent paper [GiaS2021] that samples unsupervised patterns in streaming data under a damping measure of support. Damping measures can inspire our research, but this work needs generalization to longitudinal and RCS data, to non-sampling based algorithms, and to data that is supervised (the paper [GiaS2021] does not Variants or Effects; it merely counts occurences of subgroups).

Challenge 2: exhaustive EMM algorithms for time-varying data [Executed by PhD 2 and PI]

PhD 2 will focus on exhaustive algorithms. They will expand A&B testing with the MVT expansion of FC 06-09 (**C2a**). They will prove new Optimistic Estimates for branch-and-bound algorithms in time-varying data (**C2b**). They will create condensed representations for the GP-Growth algorithm in time-varying data (**C2c**). They will define measures of significance that allow for significant pattern mining in time-varying data (**C2d**).

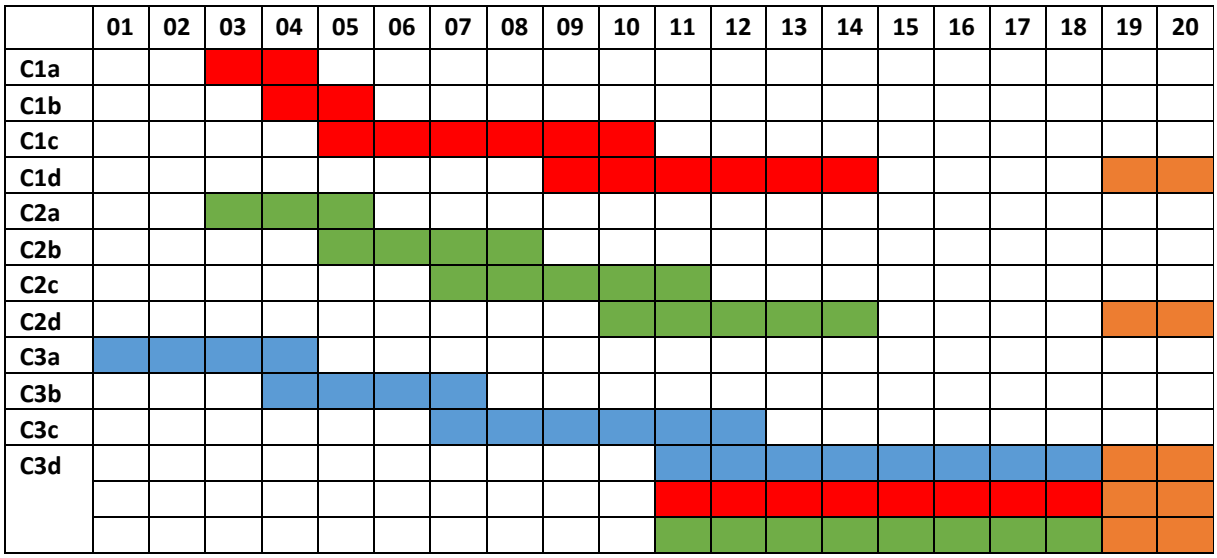
PhD 2 can be inspired by recent work on mining periodic patterns [GalCTTC2018], where repetitive cycles are extracted from event logs. Cycles live in a timescale hierarchy (weekly on top of daily), which can be represented as a graph, enabling finding informative patterns in time-varying data by tricks from graph pattern mining [KurK2001].

Challenge 3: anytime EMM algorithms with guarantees for time-varying data [Executed by PI; contributions by PhDs]

I will focus on anytime algorithms with guarantees. I will answer the foundational questions of FC 11-13 (**C3a**). I will generalize Refine&Mine [BelBK2018] to datasets with mixed attribute types (**C3b**), interaction between multiple targets (**C3c**), and time-varying data (**C3d**). Each generalization requires proofs of bounds on the distance of our current solution to the optimum, whenever the algorithm is interrupted.

Challenge 3 requires an extensive collaboration with the data mining group in Lyon. One of the Refine&Mine paper [BelBK2018] authors, A. Belfodil, visited my group at the TU/e, resulting in a joint paper [BelDPCL2019]; a recurring pattern of mutual research visits between my group and the group of M. Plantevit is desirable.

Timetable



This Gantt chart displays the challenge distribution across team members over the 20 quartiles in the project. The total project duration is five years; labels correspond to the bold text in the Challenge text. Primary responsiblity for tasks in red lies with PhD 1, for tasks in green with PhD 2, and for tasks in blue with the PI. Several tasks will need to be wrapped up by the PI after the PhDs have graduated: those are displayed in orange.

The more fundamental challenges C1a, C1b, C2a, and C3a are set up such that all academic personnel hired on this project, including the relatively inexperienced PhD students, can hit the ground running: several accessible projects await them to be resolved immediately. Having cut our teeth on these initial subtasks, we are to spend more time on the riskier subchallenges (C1c, C1d, C2b, C2c, C2d, C3b, and C3c). I expect PhD 1 to intially focus on time series data, as inspired by the sampling algorithm on streaming data provided by [GiaS2021]. I expect PhD 2 to initially focus on longitudinal data, as inspired by the exhaustive algorithm on event log data provided by [GalCTTC2018]; by elimination, I will initially focus on RCS data myself. Finally, combining all new research into an anytime algorithm with guarantees for time-varying data (subchallenge C3d) requires substantial efforts of the entire team.

2a3. Motivation for choice of host institute

The Data and Artificial Intelligence (DAI) cluster at Technische Universiteit Eindhoven currently encompasses sixteen faculty members, half of which in the Data Mining group. These cluster members all bring complementary expertises to the table: a wide variety of data mining and AI subfields are covered. This ensures a very fruitful environment for academics in all stages of their career (including both faculty members and PhD students) to thrive: on the one hand, we can concentrate on our main research focus without being in direct competition with fellow cluster members, on the other hand, if our research comes into contact with a neighboring data mining subfield, we can make good use of the expertise of other members of our cluster. For instance, if we would deploy M4M on patient data, and we have knowledge of the social network between the patients, it might strengthen M4M to incorporate techniques from Graph Mining; an expert on the topic of Graph Mining happens to occupy the office next to the PI (cf. Challenge 2, last sentence, for an example of the usefulness of such synergies). Similarly, experts on Natural Language Processing, Deep Learning, Matrix Factorization, Metalearning, Decision Support Systems, and so on and so forth are all sharing the same hallway as the to-be-hired PhDs, to share their expertise if necessary.

Beyond the cluster, TU/e has recently invested heavily in EAISI, the Eindhoven Artificial Intelligence Systems Institute. EAISI offers a platform to interact with TU/e researchers who are interested in Artificial Intelligence but stationed outside of the DAI cluster, with representatives from industry, and with student teams.

2b. Scientific and/or societal impact of the proposed project (Knowledge utilisation)

Specify which kind of impact the proposal focusses on:

Scientific and societal impact are of comparable focus

Please elaborate on the scientific and/or societal impact of the proposed project:

The Exceptional Model Mining generalization to be developed in M4M will be presented in open-access academic papers at top-ranked conferences and journals in the field. As a consequence, the fruits of M4M will be available for use to all kinds of stakeholders. Controlled experiments range far beyond computer science. Analyzing student success can make M4M applicable to education [WilLKWMPCH2014]. Medical trials often involve what essentially amounts to an A/B test [JasJ2012]: does medicine variant A or B cure more people, or have fewer side effects, or lead to a better quality of life? These tests are instrumental in determining whether health insurers will cover certain forms of medication, but that is a very blunt one-size-fits-all instrument for such an important domain! The margin of victory is of paramount importance here: if 55% of the population is better served with medicine A, the standard A/B test would dictate that we prescribe medicine A to everyone and shred medicine B. Thereby, 45% of the population suffers. Identifying subgroups for whom the outcome of controlled experiments is exceptional, has the potential to improve the outcome for a substantial number of patients: it moves from one-size-fits-all medicine towards personalized medicine, landing somewhere which might be best described as automatically stratified medicine: basing the medication on the subgroup(s) to which the patient belongs.

Initial rollout of M4M beyond the university grounds is planned in collaboration with two specific partners: [details scrambled to protect their anonymity].

2d. Literature references

[BelBK2018] A. Belfodil, A. Belfodil, M. Kaytoue: Anytime Subgroup Discovery in Numerical Domains with Guarantees. ECML PKDD 2018: 500-516

[BelDPCL2019] A. Belfodil, W. Duivesteijn, M. Plantevit, S. Cazalens, P. Lamarre: DEvIANT: Discovering Significant Exceptional (Dis-)Agreement Within Groups. ECML PKDD 2019: 3-20

[BolLPG2011] M. Boley, C. Lucchese, D. Paurat, T. Gärtner: Direct local pattern sampling by efficient two-step random procedures. KDD 2011: 582-590

[BolMG2012] M. Boley, S. Moens, T. Gärtner: Linear space direct pattern sampling using coupling from the past. KDD 2012: 69-77

[BosBRK2017] G. Bosc, J.-F. Boulicaut, C. Raïssi, M. Kaytoue: Anytime Discovery of a Diverse Set of Patterns with Monte Carlo Tree Search. Data Min. Knowl. Discov. 32(3):604-650 (2017)

[Coo1977] R. D. Cook: Detection of Influential Observation in Linear Regression. Technometrics 19(1):15-18 (1977)

[CooW1980] R. D. Cook, S. Weisberg: Characterizations of an Empirical Influence Function for Detecting Influential Cases in Regression. Technometrics 22(4):495-508 (1980)

[CooW1982] R. D. Cook, S. Weisberg: Residuals and Influence in Regression. Chapman & Hall, London, 1982

[DavH1950] O. L. Davies, W. A. Hay: The Construction and Uses of Fractional Factorial Designs in Industrial Research. Biometrika 6(3):233-249 (1950)

[DowD2017] L. Downar, W. Duivesteijn: Exceptionally Monotone Models - the Rank Correlation Model Class for Exceptional Model Mining. Knowledge and Information Systems 51(2):369-394 (2017)

[Dui2013] W. Duivesteijn: Exceptional Model Mining. PhD thesis, Leiden University, 2013

[DuiFPPWAFP2017] W. Duivesteijn, T. Farzami, T. Putman, E. Peer, H. J. P. Weerts, J. N. Adegeest, G. Foks, M. Pechenizkiy: Have It Both Ways - From A/B Testing to A&B Testing with Exceptional Model Mining. ECML/PKDD (3) 2017: 114-126

[DuiFK2012] W. Duivesteijn, A. Feelders, A. J. Knobbe: Different slopes for different folks: mining for exceptional regression models with cook's distance. KDD 2012: 868-876

[DuiFK2016] W. Duivesteijn, A. Feelders, A. J. Knobbe: Exceptional Model Mining - Supervised descriptive local pattern mining with complex target concepts. Data Min. Knowl. Discov. 30(1): 47-98 (2016)

[DuiK2011] W. Duivesteijn, A. Knobbe: Exploiting False Discoveries - Statistical Validation of Patterns and Quality Measures in Subgroup Discovery. ICDM 2011: 151-160

[GalCTTC2018] E. Galbrun, P. Cellier, N. Tatti, A. Termier, B. Cremilleux: Mining Periodic Patterns with a MDL Criterion. ECML PKDD 2018: 535-551

[GiaS2018] A. Giacometti, A. Soulet: Dense Neighborhood Pattern Sampling in Numeric Data. SDM 2018: 756-764

[GiaS2021] A. Giacometti, A. Soulet: Reservoir Pattern Sampling in Data Streams. ECMLPKDD 2021: 337-352

[GroRW2008] H. Grosskreutz, S. Rüping, S. Wrobel: Tight Optimistic Estimates for Fast Subgroup Discovery. ECML PKDD 2008: 440-456

[HocT1987] Y. Hochberg, A. C. Tamhane: Multiple Comparison Procedures. Wiley, New York, 1987

[JasJ2012] M. Jaskowski, S. Jaroszewicz: Uplift modeling for clinical trial data. Proc. ICML 2012 Workshop on Machine Learning for Clinical Data Analysis (2012)

[KohL2016] R. Kohavi, R. Longbotham: Online Controlled Experiments and A/B Tests. In: C. Sammut, G. I. Webb: Encyclopedia of Machine Learning and Data Mining. Springer, 2016

[KohLSH2009] R. Kohavi, R. Longbotham, D. Sommerfield, R. M. Henne: Controlled experiments on the web: survey and practical guide. Data Min. Knowl. Discov. 18(1): 140-181 (2009)

[Kos1996] R. Koselka: The new Mantra: MVT. Forbes, March 11, pp. 114-118 (1996)

[KurK2001] M. Kuramochi, G. Karypis: Frequent Subgraph Discovery. ICDM 2001: 313-320

[LanZ2007] J. Langford, T. Zhang: The Epoch-Greedy Algorithm for Contextual Multi-armed Bandits. NIPS 2007: 817-824

[LemFK2008] D. Leman, A. Feelders, A. J. Knobbe: Exceptional Model Mining. ECML/PKDD (2) 2008: 1-16

[LemBA2012] F. Lemmerich, M. Becker, M. Atzmueller: Generic Pattern Trees for Exhaustive Exceptional Model Mining. ECML PKDD 2012: 277-292

[Lya1901] A. M. Lyapunov, Nouvelle forme du théorème sur la limite de probabilité, St. Petersburg, 1901

[ManCS2000] L. Mancinelli, M. Cronin, W. Sadée: Pharmacogenomics: The Promise of Personalized Medicine. AAPS PharmSci. 2(1): 29-41 (2000)

[MatNBK2021] R. Mathonat, D. Nurbakova, J.F. Boulicaut, M. Kaytoue: Anytime mining of sequential discriminative patterns in labeled sequences. Knowledge and Information Systems 63(2): 439-476 (2021)

[MoeB2014] S. Moens, M. Boley: Instant exceptional model mining using weighted controlled pattern sampling. IDA 2014: 203-214

[NetKNW1996] J. Neter, M. Kutner, C. J. Nachtsheim, W. Wasserman: Applied Linear Statistical Models. WCB McGraw-Hill, 1996

[Pet2004] E. T. Peterson: Web Analytics Demystified: a Marketer's Guide to Understanding How Your Web Site Affects Your Business. Celilo Group Media and CafePress, 2004

[PlaB1946] R. L. Plackett, J. P. Burman: The Design of Optimum Multifactorial Experiments. Biometrika 33(4):305-325 (1946)

[SirK2013] D. Siroker, P. Koomen: A/B Testing: The Most Powerful Way to Turn Clicks Into Customers. Wiley, 2013

[Web2007] G. I. Webb: Discovering Significant Patterns. Machine Learning 68(1): 1-33 (2007)

[WilLKWMPCH2014] J. J. Williams, N. Li, J. Kim, J. Whitehill, S. Maldonado, M. Pechenizkiy, L. Chu, N. Heffernan: MOOClets: A Framework for Improving Online Education through Experimental Comparison and Personalization of Modules. Working Paper No. 2523265, http://tiny.cc/moocletpdf (2014)

[Ye1998] K. Q. Ye: Orthogonal Column Latin Hypercubes and Their Application in Computer Experiments. Journal of the American Statistical Association 93(444): 1430-1439 (1998)

[Yul1912] G. U. Yule: On the Methods of Measuring Association between Two Attributes. Journal of the Royal Statistical Society. 49(6): 579-652 (1912)

2e. Data management section

1. Will data be collected or generated that are suitable for reuse?

☑ Yes: Then answer questions 2 to 4.

2. *Where will the data be stored during the research?*

All digital data will be stored on departmental servers, which are backed-up automatically and daily. Some data may also be stored on departmentally-owned computers, which are also backed-up daily.

3. *After the project has been completed, how will the data be stored for the long-term and made available for the use by third parties? To whom will the data be accessible?*

Relevant final/milestone research data can be made available to members of the research community or others as long as the data is not subject to confidentiality restrictions, patenting or any other forms of commercial exploitation. For long-term preservation of (processed) data and data dissemination we typically use a combination of in-house storage/archiving with controlled access on secure network drives available at the department (minimum 10years), discipline-specific data repositories and multidisciplinary/generic repositories, such as the Dutch data repository for technical scientific data, 4TU.Centre for Research Data. Research data generated on datasets belonging to potential external project partners may not be made available for reasons of confidentiality; in such cases, anonymized versions of the research data will be made available instead. Expertise on preserving privacy while mining data is available in the TU/e data mining group.

4. *Which facilities (ICT, (secure) archive, refrigerators or legal expertise) do you expect will be needed for the storage of data during the research and after the research? Are these available?*

ICT facilities for storage and access to data are available at our department for this project. Long-term storage and/or archiving may require a combination of in-house facilities and external data repositories, as described above.