

Regression Exception Rules

Dawid Macha^{*†}, Łukasz Wróbel^{*} and Marek Sikora^{*}

^{*} Department of Computer Networks and Systems,
Silesian University of Technology
ul. Akademicka 16, 44-100 Gliwice, Poland
Email: dawid.macha@polsl.pl

[†] Łukasiewicz Research Network – Institute of Artificial Intelligence and Cybersecurity,
Leopolda 31, 40-189 Katowice, Poland

Abstract—This paper introduces a novel framework for discovering exception rules in regression problems. While exception rules have been widely studied in classification and association analysis, their application to continuous outcomes has remained largely unexplored. To address this gap, we propose a formal definition of regression exception rule. We also present a rule induction framework based on sequential covering for discovering such patterns. Experiments conducted on 48 benchmark datasets show that the proposed approach is able to identify rare and interpretable exceptions while maintaining predictive performance comparable to other interpretable regression models. The results demonstrate that regression exception rules can effectively support knowledge discovery by revealing non-obvious relationships in continuous data.

I. INTRODUCTION

REGRESSION is one of the fundamental tasks in statistics and machine learning, focusing on modeling relationships between variables and predicting continuous values. Modern machine learning approaches, including ensemble methods and other advanced techniques, often achieve high predictive performance in regression tasks. However, they typically function as black-box models, providing limited insight into the underlying decision-making process [1], [2], [3]. This lack of interpretability can be particularly problematic in domains such as healthcare or finance, where transparency and trust are essential. In response to these limitations, rule-based models offer a more interpretable alternative by representing relationships in the form of human-readable conditional statements with associated numerical predictions [4], [5]. While most rule induction methods are designed to capture dominant patterns in the data, deviations from these patterns may also carry important information. Exception rules address this aspect by identifying statistically significant deviations from general patterns, thereby enabling the discovery of rare yet meaningful phenomena [6], [7], [8].

Although exception rules have been widely explored in the context of classification and association analysis, their application to regression problems has not been previously investigated. To the best of our knowledge, no formal definition of exception rules for regression has been proposed in the literature. To address this gap, we introduce a formal definition of regression exception rules that accounts for the continuous nature of the response variable. Furthermore, we

propose a framework for discovering such rules and evaluate its performance experimentally.

The contributions of this paper can be summarized as follows. We first define exception rules in the context of regression, ensuring consistency with existing formulations while adapting them to continuous outcomes. Next, we present a dedicated algorithm for identifying these rules in data. Finally, we demonstrate through experiments that the proposed approach is capable of uncovering insightful and non-obvious patterns that would be difficult to detect using standard methods.

The remainder of the paper is structured as follows. Section 2 provides the necessary background. Section 3 introduces the formal definition of regression exception rules and presents the proposed induction algorithm. Section 4 presents the results of exception discovery, including a detailed case study. Section 5 provides a comparative analysis with interpretable regression models. Section 6 concludes the paper and outlines directions for future research.

II. BACKGROUND

A. Rule-Based Learning

Rule-based learning constitutes a fundamental approach in knowledge discovery, primarily due to its high level of interpretability. Models expressed in the form of rules closely resemble the way humans describe relationships and make decisions, which makes them particularly useful in applications requiring transparency [4], [9]. A typical example of such a rule is:

*IF income > 4000\$ AND credit history = good THEN
class = approve loan*

Formally, a decision rule can be denoted as $A \rightarrow X$, where A represents the premise (the IF part), and X denotes the consequent (the THEN part).

In the case of regression, the rules take the form:

*IF condition₁ AND condition₂ AND ... AND
condition_n THEN target = {val} [min, max]*

Here, *val* is the mean or median value of the decision attribute for the examples covered by the rule. *Min* and *max* represent the *mean ± standard deviation* of the decision attribute for the examples covered by the rule. [10] Each elementary

condition $condition_i$ is defined on a single attribute. For nominal attributes, it has the form $a_j = v$, where $v \in V_{a_j}$, while for numerical attributes it is expressed as an inequality, i.e., $a_j \leq v$ or $a_j \geq v$, where v is an attribute value.

It is also worth noting that the conclusion of a regression rule may take a more general form, such as a function that relates the values of the dependent variable to the values of independent variables. However, in this work we focus on a simpler form of the regression rule conclusion, represented by a single descriptive statistic of the decision attribute.

B. Definition of Exception Rules

Exception rules aim to capture situations in which the behavior observed in a subset of data deviates from the dominant patterns. Rather than focusing exclusively on the most common relationships, they highlight subgroups where the expected outcome does not hold, thus revealing potentially surprising or informative phenomena.

A standard formulation of exception rules relies on two components: a *commonsense rule*, representing the general trend, and an associated *exception rule*, which describes a deviation from this trend [13]:

$$\begin{aligned} \text{Commonsense rule: } & A \rightarrow X \\ \text{Exception rule: } & A \wedge B \rightarrow Y \end{aligned}$$

The commonsense rule reflects a broadly valid relationship in the data and is typically supported by a large number of examples, yielding high reliability. The exception rule refines this general pattern by introducing additional conditions, thereby focusing on a more specific subgroup. Within this subgroup, the observed outcome differs from the one suggested by the general rule.

Due to their specificity, exception rules usually cover fewer instances and therefore tend to have lower support. As a consequence, many candidate exceptions may correspond to noise or weak patterns, making it challenging to identify those that are truly meaningful. To mitigate this issue, the basic framework can be enriched by introducing a *reference rule (RR)*, which provides additional context for assessing the validity of an exception [7]:

$$\text{Reference rule: } B \rightarrow Y$$

The role of the reference rule is to describe the relationship between the additional conditions and the outcome independently of the commonsense rule. In general, this relationship is expected to be weak (e.g., characterized by low support or confidence). This ensures that the deviation observed in the exception rule cannot be trivially explained by the additional conditions alone.

A particularly interesting situation occurs when the reference rule indicates the same outcome as the commonsense rule, i.e., $B \rightarrow X$. In this case, the additional conditions alone reinforce the general pattern rather than the exceptional one. Therefore, the shift in outcome observed in the rule $A \wedge B \rightarrow Y$ cannot be attributed to B in isolation, but instead arises from the interaction between A and B .

This distinction is essential for identifying non-trivial dependencies. If B alone were sufficient to produce the outcome Y , the exception would not be surprising. However, when B supports the general outcome X , and only in combination with A leads to a different outcome Y , the resulting pattern reflects a genuine interaction effect that would remain hidden if the variables were analyzed independently.

Taking such reference rules into account makes it possible to separate simple effects of additional conditions from true interactions. As a result, the discovered exceptions are more informative and provide deeper insight into the structure of the data.

Although exception rules have been extensively studied in classification and association rule mining [6], their applicability extends beyond these domains. In particular, they can be extended to regression tasks, where deviations are defined with respect to continuous outcomes rather than discrete classes.

C. Methods for Generating Exception Rules

In generating exception rules, two approaches can be distinguished: *direct* and *undirect*.

1) *Direct Approach*: In the direct approach, the user provides commonsense rules (CR) based on their own knowledge, and the algorithm searches for exceptions (ER). An example of such an algorithm might be the one proposed by Silberschatz and Tuzhilin [11] or Padmanabhan and Tuzhilin [12]. With this approach, the user discovers exceptions for their own knowledge; however, the drawback is that the user may not have complete knowledge of the data and could miss important relationships.

2) *Undirect Approach*: In the undirect approach, background knowledge is not provided by the user; the algorithm itself searches for pairs or triplets of rules. Several approaches have been proposed in the literature. Several representative approaches are described below:

- **MEPRO** — In this algorithm, a discovery task is viewed as a search problem, where a node of a search tree represents a rule pair. A depth-first search method is used to traverse this tree. The top K most interesting rules are found according to the GACE measure [13].
- **MEPROUX** — An extension of the MEPRO algorithm that checks additional constraints during evaluation. According to the authors, more interesting rules are generated by considering these constraints [14].
- In [7], the authors propose an algorithm that generates commonsense rules (CR), then combines them and checks whether they meet the criteria to be classified as exception rules (ER). These criteria are defined as support and confidence thresholds.
- The authors of [15] proposed a genetic algorithm to search for general knowledge rules as well as exceptions. Additionally, they defined a new class of exceptions, so-called *intra-class exceptions*. These describe exceptional situations within a single class. For example, the platypus is exceptional within the class of mammals because it lays eggs.

- The authors of [16] proposed the use of an ant colony algorithm to search for exceptions.

All of these methods are limited to classification problems, and no methods applicable to regression tasks have been identified in the existing literature.

III. METHODS

A. Regression Exception Rules

1) *Definition*: The concept of exception rules can be naturally extended from classification to regression by replacing discrete class labels with continuous target values. While in classification the objective is to identify subgroups with different class assignments, in regression the goal is to detect subgroups whose expected values differ from the general pattern.

Analogously to the classification setting, regression exception rules consist of three components: a *commonsense rule* (CR), a *reference rule* (RR), and an *exception rule* (ER):

$$\begin{aligned} \text{CR: } A &\rightarrow \mu_A [l_A, u_A], \\ \text{ER: } A \wedge B &\rightarrow \mu_{AB} [l_{AB}, u_{AB}], \\ \text{RR: } B &\rightarrow \mu_B [l_B, u_B]. \end{aligned}$$

Each rule is associated with a numerical estimate of the target variable together with an interval describing the variability of the target values for the examples covered by the rule. The interval associated with each rule is computed as the mean value $\pm$ one standard deviation. Although not a formal confidence interval, it provides a robust and interpretable measure of variability suitable for detecting deviations in regression exception rules.

For a rule covering a set of examples S , the predicted value and its variability interval are defined as

$$\mu_S = \frac{1}{|S|} \sum_{i \in S} y_i, \quad l_S = \mu_S - \sigma_S, \quad u_S = \mu_S + \sigma_S,$$

where σ_S denotes the standard deviation of the target variable for the examples in S . [17] [10]

The commonsense rule captures the dominant trend in the data, while the reference rule verifies whether the additional conditions alone influence the target variable. The exception rule identifies a subgroup whose expected value deviates from the general pattern.

A regression exception rule is defined by the following conditions:

- Consistency of the commonsense and reference rules:

$$\mu_A \in [l_B, u_B] \quad \text{and} \quad \mu_B \in [l_A, u_A].$$

- Deviation of the exception rule from the commonsense rule:

$$\mu_{AB} \notin [l_A, u_A].$$

- Deviation of the exception rule from the reference rule:

$$\mu_{AB} \notin [l_B, u_B].$$

Algorithm 1 Induction of a Regression Exception Rule Set

Require: D — training dataset, $mincov$ — minimum number of previously uncovered examples that a new rule has to cover

Ensure: R — set of regression exception rules

```

1:  $D_U \leftarrow D$  ▷ set of uncovered examples
2:  $R \leftarrow \emptyset$ 
3: repeat
4:    $(cr, rr, er) \leftarrow \text{GROWCR}(D, D_U, mincov)$ 
5:   if  $cr = \emptyset$  then
6:     break
7:   end if
8:    $R \leftarrow R \cup \{(cr, rr, er)\}$ 
9:    $D_U \leftarrow D_U \setminus \text{Cov}(cr, D_U)$ 
10: until  $D_U = \emptyset$ 
11: return  $R$ 

```

These conditions ensure that the commonsense and reference rules describe consistent behavior, while the exception rule captures a meaningful deviation. In particular, the additional conditions alone do not alter the expected value, but their combination with the conditions of CR leads to a significant change. This indicates an interaction effect between conditions A and B .

To avoid spurious exceptions supported by very small sample sizes, candidate rules are further validated using coverage-based selection and statistical hypothesis testing based on the Mann–Whitney U test. [18]

2) *Algorithm*: The induction of regression exception rules follows a separate-and-conquer strategy. Each rule is obtained through the incremental growth of a commonsense rule (CR). During this process, after each extension of the CR, the algorithm simultaneously searches for a corresponding exception rule (ER) by inducing a reference rule (RR). The conjunction of the CR and RR forms a candidate exception rule, which is evaluated against the interval-based definition of an exception and validated using statistical criteria. The overall procedure is presented in Algorithm 1. The procedure for inducing a single rule is presented in Algorithm 2.

During exception discovery presented in Algorithm 3, candidate rules are first verified using the interval-based definition. If multiple candidates satisfy these constraints, the rule covering the largest number of examples is selected. This strategy reduces the risk of overfitting and improves the stability of the induced model.

To ensure statistical validity, candidate exception rules are further verified using the Mann–Whitney U test. The test is applied to the following disjoint subsets of examples:

- $CR \setminus ER$ versus ER ,
- $RR \setminus ER$ versus ER ,
- $CR \setminus ER$ versus $RR \setminus ER$.

An exception rule is accepted if statistically significant differences are observed between ER and both $CR \setminus ER$ and $RR \setminus ER$, while no statistically significant difference exists between $CR \setminus ER$ and $RR \setminus ER$. The Mann–Whitney U test is

Algorithm 2 Growing a Commonsense Rule

Require: D — training dataset, D_U — set of uncovered examples, $mincov$ — minimum coverage

Ensure: (cr, rr, er) — regression exception rule or $\emptyset$

```

1:  $cr \leftarrow \emptyset$ 
2: repeat
3:    $c_{best} \leftarrow$  best condition extending  $cr$ 
4:   if  $c_{best} = \emptyset$  then
5:     break
6:   end if
7:    $cr \leftarrow cr \wedge c_{best}$ 
8:    $(rr, er) \leftarrow \text{FIND\_EXCEPTION}(cr, D, D_U, mincov)$ 
9:   if  $rr \neq \emptyset$  then
10:    return  $(cr, rr, er)$ 
11:  end if
12: until no further extension of  $cr$  is possible
13: return  $(CR, \emptyset, \emptyset)$ 

```

particularly suitable in this context because it does not assume normality of the target variable and is robust to outliers and small sample sizes.

IV. RESULTS OF EXCEPTION DISCOVERY

The first experiment was designed to evaluate the ability of the proposed method to discover exception rules in regression data. The proposed algorithms were evaluated on 48 benchmark datasets obtained from publicly available repositories, including UCI, OpenML, and Kaggle. The complete collection of datasets is available at: <https://github.com/ruleminer/exception-rules>. These datasets constitute the primary benchmark used in this study and represent diverse application domains, varying in dimensionality, size, and statistical properties, thereby ensuring a comprehensive and reliable assessment of the proposed approach. Furthermore, the repository provides the full implementation of the developed algorithm, along with the source code required to reproduce all experiments, promoting transparency and reproducibility of the reported results.

As the primary objective of this experiment was knowledge discovery, the analysis was conducted on complete datasets rather than focusing exclusively on predictive performance. For each dataset, detailed information about the discovered rule sets was collected, including the number of induced rules, their structure, and coverage. Additionally, qualitative analyses were performed to assess the interpretability and usefulness of the identified exceptions.

To the best of our knowledge, there are currently no dedicated methods for mining exception rules in regression. Consequently, direct comparison with alternative exception rule discovery techniques is not feasible. Instead, this experiment evaluates the capability of the proposed framework to identify meaningful and statistically justified deviations from dominant regression patterns.

In all experiments, the minimum number of previously uncovered examples required to induce a new rule ($mincov$)

Algorithm 3 Finding an Exception for a Given Commonsense Rule

Require: cr — commonsense rule, D — training dataset, D_U — set of uncovered examples, $mincov$ — minimum coverage

Ensure: (rr, er) — reference and exception rules or $(\emptyset, \emptyset)$

```

1:  $rr \leftarrow \emptyset$ 
2:  $best \leftarrow (\emptyset, \emptyset)$ 
3: repeat
4:    $c_{best} \leftarrow \emptyset$ 
5:    $er\_cov_{best} \leftarrow 0$ 
6:   for all conditions  $c$  extending  $rr$  do
7:      $rr' \leftarrow rr \wedge c$ 
8:      $er' \leftarrow cr \wedge rr'$ 
9:     if  $Cov(rr', D_U) \geq mincov$  and  $er'$  satisfies interval-based conditions then
10:       $er\_cov \leftarrow |Cov(er', D)|$ 
11:      if  $er\_cov > er\_cov_{best}$  and  $er'$  satisfies statistical tests then
12:         $c_{best} \leftarrow c$ 
13:         $er\_cov_{best} \leftarrow er\_cov$ 
14:         $best \leftarrow (rr', er')$ 
15:      end if
16:    end if
17:  end for
18:  if  $c_{best} = \emptyset$  then
19:    break
20:  end if
21:   $rr \leftarrow rr \wedge c_{best}$ 
22: until no further extension of  $rr$  is possible
23: return  $best$ 

```

was set to 5, in accordance with the default configuration of the RuleKit library [19]. Rule quality during both the growing and pruning phases was evaluated using C2 measure, a standard criterion in regression rule induction [20]. To prevent overly specific models and preserve interpretability, the maximum number of conditions added during the growing phase was limited to five.

Table I summarizes the number of induced rules and discovered exceptions for datasets in which such patterns were identified (17/48). Exception rules were not detected in every dataset, which is consistent with their definition as rare and non-trivial deviations from dominant trends. The results indicate that the proposed method successfully identifies exceptions across a wide range of regression problems, with the number of exceptions varying depending on the structural complexity of the data. In particular, datasets such as *friedman*, *compressive*, and *plastic* exhibit a relatively large number of exceptions, reflecting their heterogeneous and non-linear characteristics.

The structure of the discovered patterns is presented in Table II, which lists the commonsense (CR) and reference (RR) rules. The exception rule (ER) is defined as the conjunction of CR and RR, i.e., $ER = CR \wedge RR$. For the sake of transparency

TABLE I: Number of rules and exceptions per dataset (Exceptions > 0)

Dataset	Rules	Exceptions
bodyfat	36	1
breasttumor	50	1
cholesterol	44	11
compressive	61	20
concrete	11	7
friedman	68	48
kidney	13	1
laser	97	3
pbc	65	11
plastic	188	24
pollution	9	2
pwlinear	36	1
sensory	91	3
stock	67	9
treasury	79	3
triazines	20	1
veteran	16	1

and readability, the table presents one representative rule from each dataset in which exception patterns were discovered. The identified rules are concise, interpretable, and reflect meaningful relationships between explanatory variables and the target variable across different domains.

A quantitative characterization of the discovered rules is provided in Table III. This table reports the predicted values (conclusions), their confidence intervals, and coverage, as well as statistical comparisons between CR, RR, and ER. The numbering of rules is consistent with Table II, enabling straightforward cross-referencing between the rule definitions and their statistical properties.

The reported results indicate that the predicted values of the commonsense and reference rules are generally similar, whereas the exception rules frequently exhibit substantial deviations from both. These differences confirm that the identified exceptions capture meaningful departures from dominant regression patterns. The confidence intervals further support the statistical distinctiveness of the exception subgroups. Moreover, the reported p -values demonstrate that the differences between ER and both CR and RR are statistically significant in the majority of cases, validating the relevance and reliability of the discovered exceptions.

Coverage values reveal that exception rules typically describe small subsets of data, confirming their rare and specific nature. Despite their limited support, these rules provide valuable insights into atypical behaviors and local deviations that remain undetected by traditional regression models.

Overall, the results confirm that meaningful exception rules can be identified in regression datasets. The proposed approach proves to be selective and robust, avoiding trivial or spurious patterns while uncovering statistically significant and interpretable deviations from general regression trends.

Consequently, it provides an additional layer of transparent and actionable knowledge beyond standard regression modeling techniques.

A. Case study: Concrete dataset

To further illustrate the practical relevance of the proposed method, an additional analysis was conducted using the *Concrete* dataset [21]. This dataset describes the composition of fresh concrete mixtures and their influence on workability, measured by the slump test. It includes variables such as cement, fly ash, slag, water, superplasticizer, and aggregates. The target variable is the concrete slump (in centimeters), which reflects the consistency and flow characteristics of the mixture.

The algorithm identified the following exception structure:

- **CR:** $Cement \leq 162.50$
- **RR:** $Fly\ ash \leq 155.00$
- **ER:** $Cement \leq 162.50$ AND $Fly\ ash \leq 155.00$

The rules are associated with the following predicted values and confidence intervals:

- **CR:** target = 31.69 [26.00, 37.37]
- **RR:** target = 33.18 [25.99, 40.37]
- **ER:** target = 25.16 [20.12, 30.20]

The commonsense rule (CR) indicates that mixtures with a relatively low cement content exhibit an average slump of approximately 31.69 cm. The reference rule (RR) shows that mixtures with limited amounts of fly ash also display comparable workability, with an average slump of 33.18 cm. The substantial overlap between the confidence intervals of CR and RR suggests that both rules represent consistent baseline behavior.

The exception rule (ER) identifies a subgroup characterized by simultaneously low contents of cement and fly ash. Although each of these conditions individually corresponds to moderate workability, their combination results in a noticeably lower average slump of 25.16 cm. This value lies below the predictions of both parent rules, satisfying the formal definition of a regression exception rule.

From an engineering perspective, this finding reveals a negative interaction between the two key binder components of the mixture. Cement serves as the primary binding material, while fly ash enhances particle packing and improves rheological properties due to its fine and spherical particles. When both components are reduced, the volume of the cementitious paste becomes insufficient, leading to decreased cohesion and diminished flowability of the fresh concrete. Consequently, the mixture exhibits lower consistency than would be expected from either condition alone.

The statistical analysis confirms the exceptional nature of the identified subgroup. The difference between CR and RR is not statistically significant ($p = 0.30777$), whereas the ER differs significantly from both CR ($p = 0.00019$) and RR ($p = 0.00006$). These results validate the presence of a meaningful and statistically significant regression exception.

TABLE II: Commonsense (CR) and reference (RR) rules discovered for regression datasets. The exception rule is defined as $ER = CR \wedge RR$.

No.	Dataset	CR	RR
1	bodyfat	Density ≤ 1.04 AND Weight ≤ 221.12 AND Hip ≤ 113.95	Hip ≥ 112.15
2	breasttumor	Age ≤ 64.50 AND Age ≥ 60.50	Age ≤ 61.50
3	cholesterol	Thalach ≥ 150.50 AND Age ≥ 39.50 AND Thalach ≥ 156.50	Age ≥ 66.50
4	compressive	Age ≥ 73.00	Superplasticizer ≥ 8.04
5	concrete	Cement ≤ 162.50	Fly ash ≤ 155.00
6	friedman	Input1 ≤ 0.13	Input4 ≤ 0.05
7	kidney	Age ≤ 45.50 AND Age ≥ 25.00	Age ≤ 29.00
8	laser	Input1 ≥ 44.50 AND Input2 ≤ 44.50 AND Input3 ≤ 153.50 AND Input3 ≥ 8.00 AND Input4 ≤ 95.50	Input2 ≤ 9.50
9	pbc	Z11 ≤ 17.50	Z2 ≤ 46.50
10	plastic	Strength ≤ 24.35 AND Strength ≥ 18.15 AND Temperature ≤ 205.00	Strength ≥ 23.10
11	pollution	NONW ≤ 9.15	JANT ≥ 39.50
12	pwlinear	a1 ≤ 0.00 AND a9 ≤ 0.50 AND a3 ≥ -0.50 AND a7 ≤ 0.50	a2 ≥ 0.50
13	sensory	Rows = {2} AND Interval = {2} AND Judges = {4}	Trellis = {3}
14	stock	Company7 ≥ 80.19	Company8 ≤ 26.06
15	treasury	Currency ≥ 3474.10	CheckableDeposits ≥ 1148.30
16	triazines	p4_branch ≥ 0.40	p4_flex ≤ 0.16
17	veteran	Age ≥ 58.50 AND Karnofsky ≥ 55.00 AND Karnofsky ≤ 77.50	Months ≥ 13.50

TABLE III: Statistical characteristics and coverage of commonsense (CR), reference (RR), and exception (ER) rules for regression datasets.

No.	CR concl.	CR CI	CR cov.	RR concl.	RR CI	RR cov.	ER concl.	ER CI	ER cov.	CR vs RR	ER vs CR	ER vs RR
1	29.14	[25.28, 33.00]	50	32.86	[26.02, 39.71]	13	43.80	[40.10, 47.50]	2	0.067249	0.018653	0.025641
2	24.89	[15.13, 34.66]	28	24.78	[14.59, 34.97]	233	14.50	[9.94, 19.06]	4	0.557321	0.016566	0.020288
3	249.93	[194.20, 305.66]	123	259.96	[186.51, 333.41]	26	340.20	[226.81, 453.59]	5	0.913350	0.013199	0.013378
4	46.97	[34.22, 59.72]	190	43.11	[25.61, 60.62]	400	60.75	[49.14, 72.36]	46	0.078053	4.25×10^{-15}	5.77×10^{-13}
5	31.69	[26.00, 37.37]	42	33.18	[25.99, 40.37]	48	25.16	[20.12, 30.20]	10	0.307769	0.000187	0.000057
6	9.52	[5.89, 13.15]	157	8.87	[4.94, 12.80]	64	4.88	[2.36, 7.41]	14	0.942553	2.81×10^{-6}	1.88×10^{-5}
7	0.96	[0.48, 1.44]	28	1.43	[0.77, 2.10]	12	2.30	[2.30, 2.30]	2	0.091689	0.021774	0.036368
8	115.25	[74.13, 156.36]	166	113.52	[32.42, 194.63]	48	203.44	[168.87, 238.02]	9	0.050321	2.60×10^{-6}	0.000296
9	2644.38	[1944.53, 3344.22]	16	1999.96	[911.88, 3088.05]	156	3487.25	[2969.43, 4005.07]	4	0.050290	0.004396	0.009139
10	13.85	[11.36, 16.34]	80	14.39	[11.05, 17.74]	1319	11.00	[9.75, 12.25]	14	0.683115	1.67×10^{-6}	8.26×10^{-5}
11	905.15	[855.53, 954.76]	29	934.44	[852.57, 1016.32]	16	847.88	[808.62, 887.15]	4	0.053553	0.016336	0.019780
12	3.79	[0.98, 6.59]	26	1.20	[-4.35, 6.75]	55	7.54	[4.74, 10.33]	6	0.099626	0.000556	0.002520
13	15.88	[15.20, 16.55]	16	15.28	[14.48, 16.09]	144	16.75	[16.19, 17.31]	4	0.077285	0.007555	0.002585
14	48.17	[44.51, 51.84]	137	46.20	[39.36, 53.04]	691	53.45	[50.12, 56.78]	33	0.117571	1.34×10^{-12}	3.35×10^{-10}
15	5.46	[5.17, 5.76]	250	5.31	[4.70, 5.92]	35	5.94	[5.83, 6.05]	4	0.054602	0.002003	0.048704
16	0.58	[0.43, 0.73]	37	0.61	[0.40, 0.82]	67	0.10	[0.10, 0.10]	2	0.235890	0.020417	0.017455
17	105.46	[21.27, 189.65]	35	112.90	[17.49, 208.32]	21	208.75	[154.60, 262.90]	4	0.483457	0.005530	0.035257

Overall, this example demonstrates that regression exception rules are capable of uncovering subtle and non-linear interactions in concrete technology. Such insights can support mixture design, quality control, and process optimization, contributing to improved performance and reliability of construction materials.

V. COMPARISON WITH DIFFERENT INTERPRETABLE MODELS

Since there are no established methods for discovering exception rules in regression, the comparison was performed against interpretable regression models that generate general rule-based or rule-equivalent structures. These methods were selected as suitable reference points for evaluating predictive

performance, model compactness, and computational efficiency. The restriction to inherently interpretable approaches was intentional and consistent with the primary objective of this study, which is knowledge discovery through transparent and human-readable patterns. Consequently, black-box models were not included in the comparison, as they do not provide transparent insights into relationships and would not align with the goals of this work. This evaluation protocol ensures methodological fairness and aligns with established practices in interpretable machine learning.

The developed algorithm was considered in two variants: with exception discovery (RegressionExceptionRules) and without exception discovery (RegressionRules). These variants were

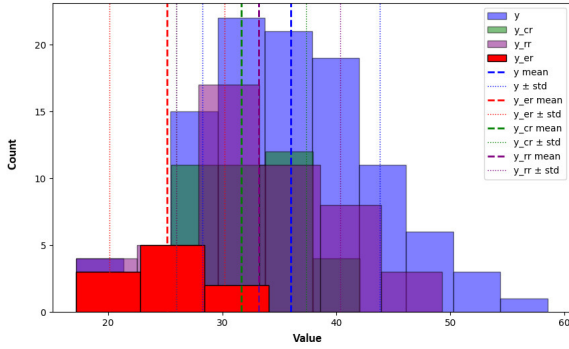


Fig. 1: Distributions of the target variable for the commonsense (CR), reference (RR), and exception (ER) rules in the *Concrete* dataset.

compared with the RuleKit framework [19] and regression trees (DecisionTreeRegressor [22]).

A. Predictive Performance and Model Compactness

For predictive evaluation, each dataset was randomly partitioned into training and test subsets. Performance was assessed using the Root Mean Squared Error (RMSE). In addition, model compactness was evaluated using the number of generated rules. In the case of regression trees, decision rules were extracted from root-to-leaf paths to ensure structural comparability.

The algorithm parameters were set as in the first experiment. To ensure a fair comparison, the maximum depth of regression trees was also limited to 5.

The results are summarized using critical difference (CD) diagrams presented in Figure 2. These diagrams illustrate the average ranks of the compared methods across datasets with respect to RMSE on the test sets and model complexity measured by the number of rules. In these diagrams, lower ranks indicate better performance. The statistical significance of the observed differences was assessed using the Friedman test followed by the Nemenyi post-hoc test at a significance level of $\alpha = 0.05$. The critical difference (CD) value was computed based on the average ranks of the methods and the number of evaluated datasets. Methods whose average ranks differ by at least the CD are considered significantly different, while those connected by a horizontal line in the diagram do not differ significantly [23], [24].

The predictive performance results indicate that there are no statistically significant differences between the compared methods in terms of RMSE. Although DecisionTreeRegressor achieves the best average rank, the CD diagram shows that its advantage over the remaining methods is not statistically significant. A particularly important comparison concerns RegressionExceptionRules and RegressionRules. Their similar rankings confirm that incorporating exception discovery maintains stable

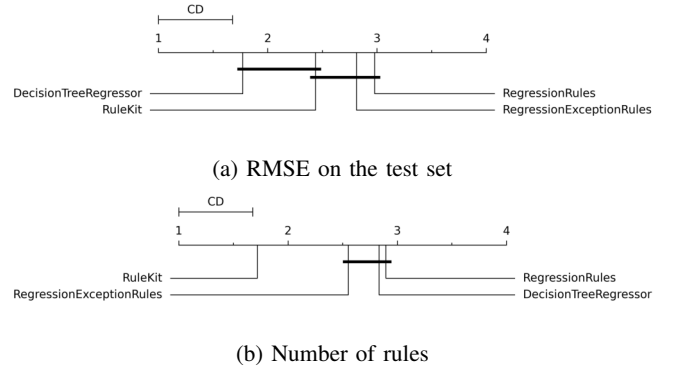


Fig. 2: Critical Difference (CD) diagrams comparing the evaluated methods: (a) RMSE on the test set, (b) number of rules.

and reliable predictive performance without introducing overfitting.

Differences are more pronounced in terms of model complexity, as illustrated by the number of generated rules. RuleKit produces the most compact models, achieving the best average rank. The RegressionExceptionRules approach generates more complex models than RuleKit, yet remains more concise than the remaining methods. The RegressionRules method yields more extensive rule sets, comparable in complexity to those produced by the DecisionTreeRegressor.

Overall, the results demonstrate that the proposed approach maintains predictive performance comparable to state-of-the-art interpretable regression methods while providing the unique capability of discovering exceptional patterns. Together with the results of the first experiment, this confirms its effectiveness as a tool for knowledge discovery in regression analysis.

B. Computational Efficiency Analysis

In addition to predictive performance and model compactness, the compared methods were evaluated with respect to their computational efficiency. The analysis focused on the training time required to induce regression models.

The results presented in Table IV reveal substantial differences in computational cost among the evaluated methods. As expected, DecisionTreeRegressor is the most efficient algorithm, with training times typically below 0.01 seconds. This efficiency results from the relatively low computational complexity of tree induction and its highly optimized implementation. Among the interpretable rule-based approaches, RuleKit demonstrates the highest efficiency.

The RegressionRules method, which constitutes the core of the proposed approach without exception discovery, requires noticeably more computational resources and is slower than both RuleKit and DecisionTreeRegressor. It is important to note that the implementation of the proposed algorithms was developed in Python with an emphasis on

TABLE IV: Summary of training times (in seconds) for the compared methods

Method	Mean	Median	Q1	Q3	Min	Max
DecisionTreeRegressor	0.009	0.008	0.006	0.010	0.004	0.017
RuleKit	0.696	0.182	0.062	1.298	0.010	2.951
RegressionRules	15.414	1.148	0.366	9.028	0.028	166.118
RegressionExceptionRules	77.055	4.863	1.345	41.804	0.080	975.515

clarity, readability, and reproducibility rather than computational optimization. Therefore, the current version prioritizes transparency and ease of understanding. In future work, the code will be optimized to improve efficiency, for example by incorporating parallel processing, optimized data structures, and compiled numerical routines. Consequently, further performance improvements are expected without altering the methodological foundations of the proposed approach.

The most computationally demanding approach is `RegressionExceptionRules`. This is expected, as the method extends standard regression rule induction by discovering statistically validated exception patterns. The additional steps introduce significant computational overhead. Consequently, the median training time increases to approximately 4.86 seconds. As with most machine learning algorithms, the training time depends on the size and complexity of the datasets, increasing with the number of instances and attributes.

Despite the increased computational cost, the observed training times do not constitute a practical limitation. Even in the most demanding cases, model induction requires minutes rather than hours. This level of efficiency is acceptable given that the primary objective of the proposed approach is knowledge discovery rather than real-time work. The additional computational effort is justified by the ability to identify rare, interpretable, and statistically significant exception patterns that remain undetected by conventional regression models.

VI. CONCLUSIONS

This paper introduced regression exception rules as a novel framework for discovering exceptional patterns in regression tasks. The proposed approach extends the concept of exception rules beyond classification and association analysis by providing a formal definition adapted to continuous target variables. In addition, a dedicated induction procedure was developed to identify commonsense, reference, and exception rules within a unified and statistically validated framework.

The experimental results obtained on benchmark datasets demonstrate that the proposed method is capable of identifying rare yet meaningful deviations from dominant regression patterns. The discovered rules are interpretable and provide insights into local interactions and non-obvious relationships that are not directly captured by standard regression models. At the same time, the comparative analysis shows that the incorporation of exception discovery preserves predictive performance at a level comparable to other interpretable regression approaches.

The presented findings confirm that regression exception rules constitute a useful tool for knowledge discovery in continuous data. By combining interpretability with the ability to detect statistically significant deviations, the proposed framework enriches regression analysis with an additional descriptive layer focused on exceptional subgroups.

Future research should focus on investigating alternative measures of variability and statistical significance, examining the integration of regression exception rules with interpretable and hybrid learning models, and optimizing computational implementations to improve efficiency and scalability.

REFERENCES

- [1] Q. An, S. Rahman, J. Zhou, J.J. Kang. A Comprehensive Review on Machine Learning in Healthcare Industry: Classification, Restrictions, Opportunities and Challenges. *Sensors*, 23(9):4178, 2023. <https://doi.org/10.3390/s23094178>
- [2] R. Pugliese, S. Regondi, R. Marini. Machine learning-based approach: global trends, research directions, and regulatory standpoints. *Data Science and Management*, 4:19–29, 2021. <https://doi.org/10.1016/j.dsm.2021.12.002>
- [3] Maszczyk, C., Kozielski, M., Sikora, M. (2022). Rule-Based Approximation of Black-Box Classifiers for Tabular Data to Generate Global and Local Explanations. *Annals of Computer Science and Information Systems*, vol. 30, pp. 89–92. <https://doi.org/10.15439/2022F258>
- [4] J. Fürnkranz, D. Gamberger, N. Lavrač. *Foundations of Rule Learning*. 2012. 10.1007/978-3-540-75197-7.
- [5] B. Letham, C. Rudin, T. H. McCormick, and D. Madigan. Interpretable classifiers using rules and Bayesian analysis: Building a better stroke prediction model. *Annals of Applied Statistics*, 9:1350–1371, 2015.
- [6] B. Duval, A. Salles, C. Vrain. On the Discovery of Exception Rules: A Survey. In Guillet, F.J., Hamilton, H.J. (eds) *Quality Measures in Data Mining. Studies in Computational Intelligence*, vol 43. Springer, Berlin, Heidelberg, 2007. https://doi.org/10.1007/978-3-540-44918-8_4
- [7] F. Hussain, H. Liu, E. Suzuki, H. Lu. Exception Rule Mining with a Relative Interestingness Measure. In *Proceedings of the 4th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 86–97, 2000. 10.1007/3-540-45571-X_11.
- [8] M. Calvo-Flores, M. Ruiz, D. Sánchez. New Approaches for Discovering Exception and Anomalous Rules. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 19:361–399, 2011. doi:10.1142/S0218488511007039
- [9] Zielosko, B., Stańczyk, U., Jabłoński, K. (2023). Filtering Decision Rules Driven by Sequential Forward and Backward Selection of Attributes: An Illustrative Example in Stylometric Domain. *Annals of Computer Science and Information Systems*, vol. 35, pp. 833–842. <https://doi.org/10.15439/2023F7295>
- [10] Sikora, M., Skowron, A., Wróbel, Ł. (2012). Rule Quality Measure-Based Induction of Unordered Sets of Regression Rules. In: Ramsay, A., Agre, G. (eds) *Artificial Intelligence: Methodology, Systems, and Applications. AIMSA 2012. Lecture Notes in Computer Science*(), vol 7557. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-33185-5_18
- [11] A. Silberschatz and A. Tuzhilin. On subjective measure of interestingness in knowledge discovery. In: *Proceedings of the First International Conference on Knowledge Discovery and Data Mining*, Montreal, August 20–21, 1995, pp. 275–281.

- [12] B. Padmanabhan and A. Tuzhilin. A belief-driven method for discovering unexpected patterns. In: *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, New York, August 27–31, 1998, pp. 94–100.
- [13] E. Suzuki and M. Shimura. Exceptional knowledge discovery in databases based on information theory. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD'96)*, AAAI Press, pages 275–278, 1996.
- [14] E. Suzuki. Discovering unexpected exceptions: A stochastic approach. In: *Proceedings of the Fourth International Workshop on Rough Sets, Fuzzy Sets, and Machine Discovery (RSFD)*, 1996.
- [15] J. Vashishtha, D. Kumar, S. Ratnoo. An evolutionary approach to discover intra-and inter-class exceptions in databases. *International Journal of Intelligent Systems Technologies and Applications*, 12:283–300, 2013. 10.1504/IJISTA.2013.056535.
- [16] S. Ratnoo, A. Pathak, J. Ahuja, J. Vashishtha. Exception Discovery Using Ant Colony Optimization. *International Journal of Computational Systems Engineering*, 4, 2018. 10.1504/IJCSYSE.2018.10011809.
- [17] F. Janssen, J. Fürnkranz. Heuristic rule-based regression via dynamic reduction to classification, in: *IJCAI'11*, AAAI Press, 2011, pp. 1330–1335.
- [18] H. B. Mann and D. R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18(1):50–60, 1947.
- [19] A. Gudyś, M. Sikora, Ł. Wróbel. RuleKit: A comprehensive suite for rule-based learning. *Knowledge-Based Systems*, 194, 2020, 105480, ISSN 0950-7051. <https://doi.org/10.1016/j.knsys.2020.105480>.
- [20] Ł. Wróbel, M. Sikora, M. Michalak. Rule Quality Measures Settings in Classification, Regression and Survival Rule Induction — an Empirical Approach. *Fundamenta Informaticae*, 149(4):419–449, 2016. doi:10.3233/FI-2016-1455
- [21] Yeh, I. (1998). Concrete Compressive Strength [Dataset]. UCI Machine Learning Repository. <https://doi.org/10.24432/C5PK67>.
- [22] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. VanderPlas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2011, 2825–2830. <https://jmlr.org/papers/v12/pedregosa11a.html>.
- [23] P. Nemenyi. Distribution-Free Multiple Comparisons. *PhD Thesis*, Princeton University, Princeton, NJ, 1963.
- [24] J. Demšar. Statistical Comparisons of Classifiers over Multiple Data Sets. *Journal of Machine Learning Research*, 7, 2006, 1–30. <https://jmlr.org/papers/v7/demsar06a.html>.