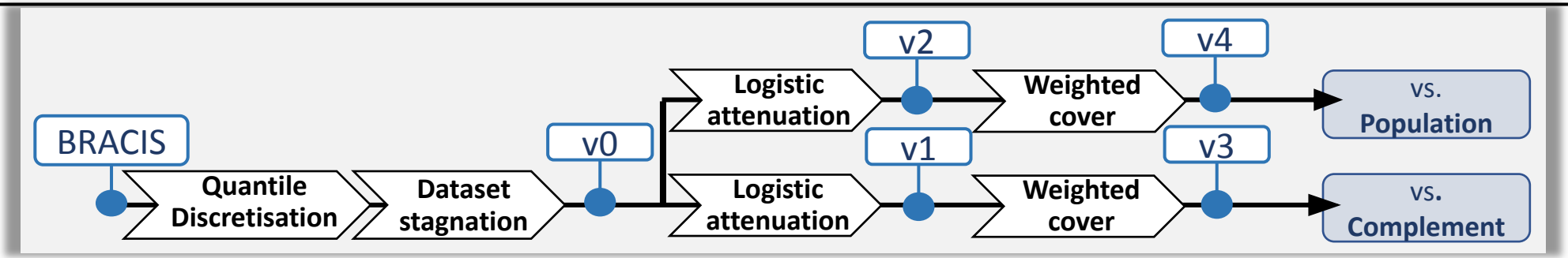
ESMAM ALGORITHM

FINAL VERSION

IMPLEMENTED ALGORITHMS



RESULTS AND POST-PROCESSING CODES OF THE ABOVE DEVELOPING ARE IN D:\GoogleDrive\Mestrado-UFPE_Pesquisa-Mestrado__ESM-AM_algorithm_improvement_results

TRANSITION RULE

$$p(c_{ij}) = \frac{\tau_{ij} * \eta_{ij}}{\sum_{\forall i,j}(\tau_{ij} * \eta_{ij})}$$


Springer Link

Brazilian Conference on Intelligent Systems
BRACIS 2020: Intelligent Systems pp 307-321 | Cite as

Exceptional Survival Model Mining

Authors | Authors and affiliations

Juliana Barcellos Mattos, Eraylson G. Silva, Paulo S. G. de Mattos Neto, Renato Vimieiro

Conference paper
First Online: 13 October 2020

187

Part of the Lecture Notes in Computer Scie

BRACIS 2020

Approached degrees of redundancy:

$$\delta_{ij} = 1 - \frac{1}{1 + e^{-(x_{ij}-5)}}$$

Logistic attenuation δ_{ij}: (DESCRIPTION)

- x_{ij} the counting of the term c_{ij} presence in the final rule model.

$$\psi_{ij} = \frac{1}{|c_{ij}|} \sum_{e \in c_{ij}} 0.9^{x(e,R)}$$

The cover-based attenuation: (COVERAGE)

- $|c_{ij}|$ is the size of the term c_{ij}, i.e. the number of examples e it covers; and
- $x(e,R)$ is how many times the example e is covered by any rule in the final rule model R.

$$\eta_{ij} = \frac{\log_2 2 - H(W|a_i = v_{ij})}{\sum_{\forall i,j} \log_2 2 - H(W|a_i = v_{ij})} = \frac{\zeta_{ij}}{\sum_{\forall i,j} \zeta_{ij}}$$

$$\eta Desc_{ij} = \frac{\delta_{ij} * \zeta_{ij}}{\sum_{\forall i,j}(\delta_{ij} * \zeta_{ij})}$$

$$\eta Cover_{ij} = \frac{\psi_{ij} * \delta_{ij} * \zeta_{ij}}{\sum_{\forall i,j}(\psi_{ij} * \delta_{ij} * \zeta_{ij})}$$

HEURISTIC FUNCTIONS

PROBLEM

MANY RULES WITH SIMILAR MODELS

```
R0: ('X204015_s_at', '[7.00,8.00)')
R1: ('X204015_s_at', '[7.00,8.00)') & ('X217815_at', '[10.00,10.00]')
```

Keep the **general** one

TERM INTERSECTION

```
R2: ('size', '[1.00,2.00)')
R5: ('X202240_at', '[6.00,7.00)')
```

Add **both** rules

NO INTERSECTION

```
R5: ('X202240_at', '[6.00,7.00)')
R6: ('X202240_at', '[4.00,6.00)')
```

Try to **merge**

ATTRIBUTE INTERSECTION

PROPOSAL

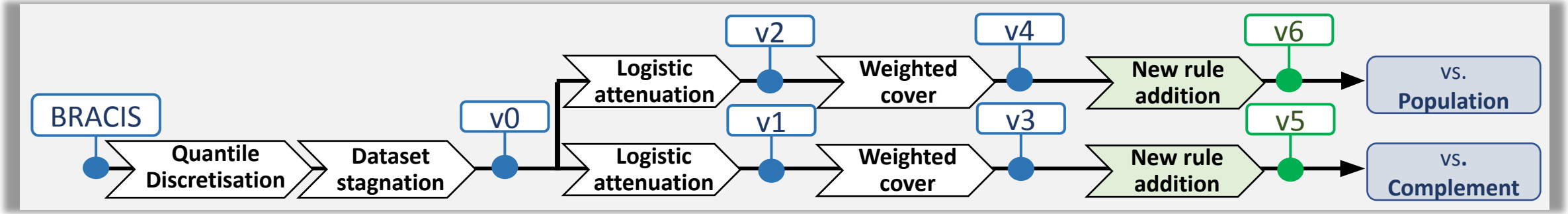
```
R0: ('X204015_s_at', '[7.00,8.00)')
R1: ('X204015_s_at', '[7.00,8.00)') & ('X217815_at', '[10.00,10.00]')

R5: ('X202240_at', '[6.00,7.00)')
R6: ('X202240_at', '[4.00,6.00)')
```

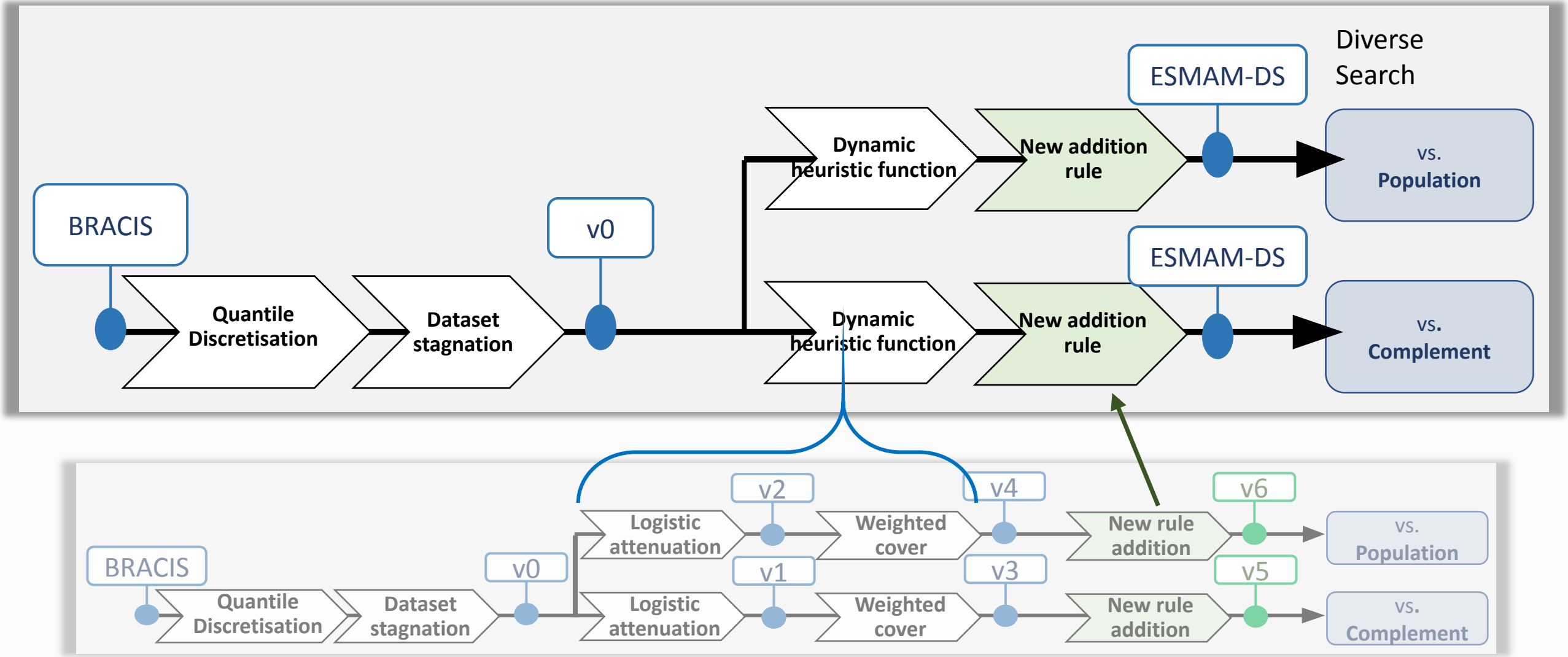
SIMILAR **MODELS**
&
SIMILARITIES IN **DESCRIPTION**

→ KEEP **GENERALIZATION**

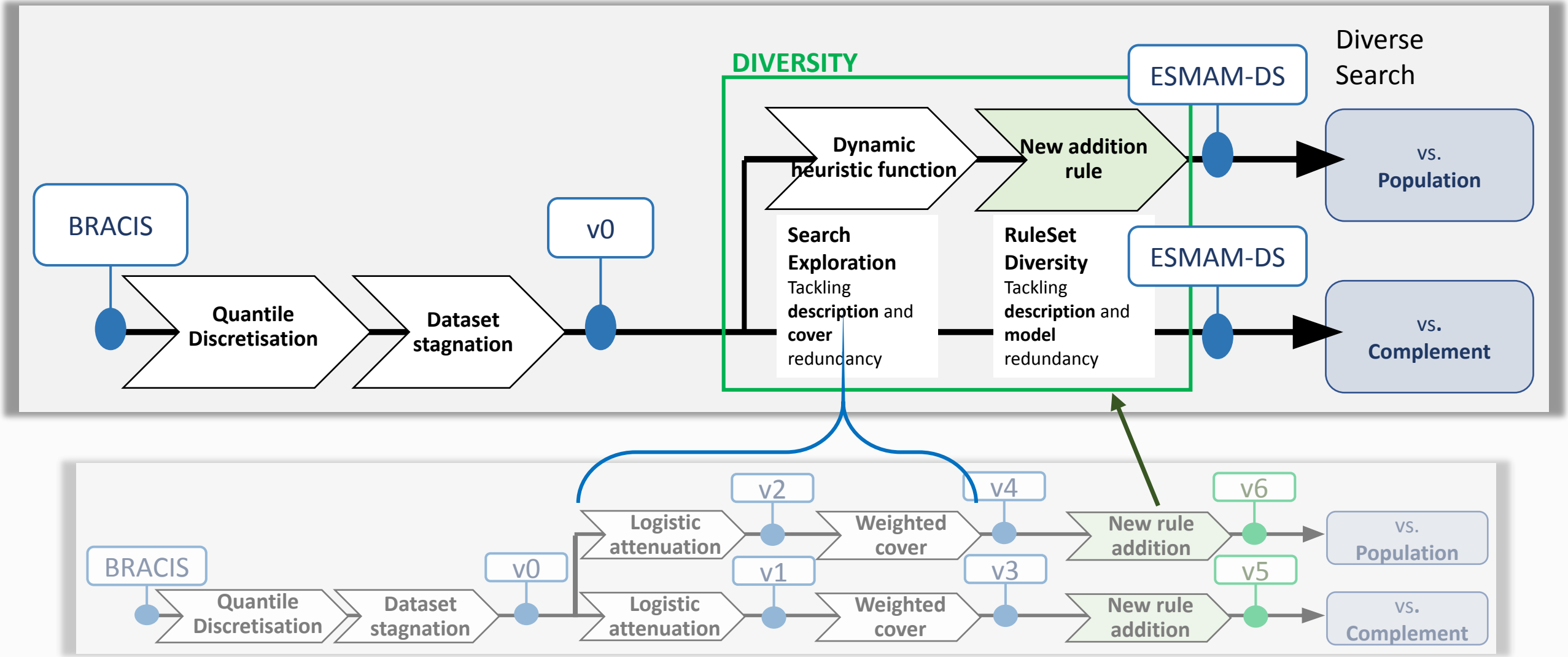
proposal: **CHANGE THE FUNCTION FOR ADDING RULES TO THE LIST**



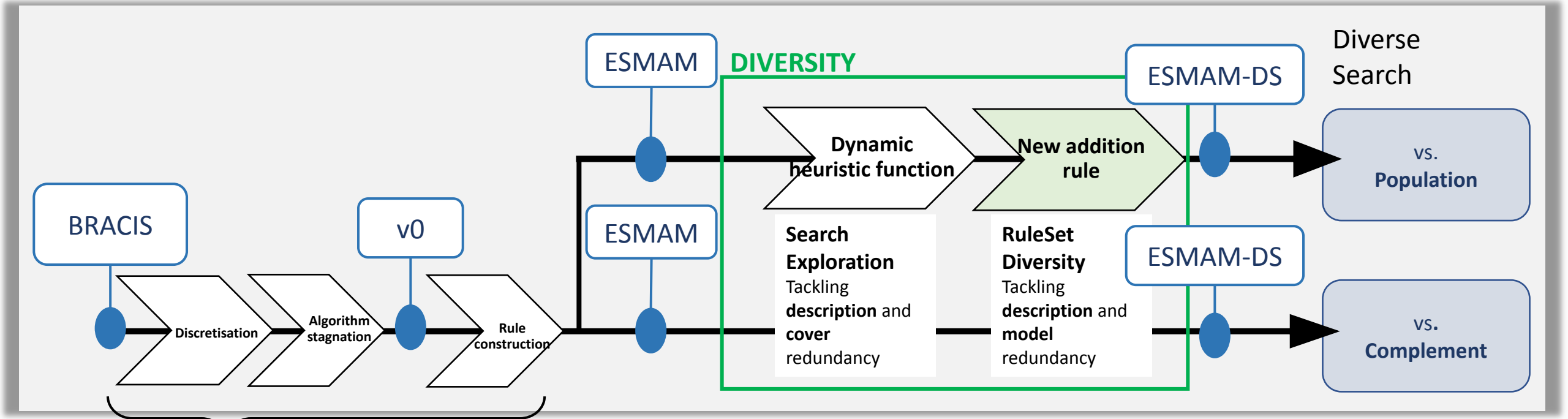
FINAL VERSION: ESMAM-DS (Diverse Search)



FINAL VERSION: ESMAM-DS (Diverse Search)



FINAL VERSION: ESMAM-DS (Diverse Search)



STRUCTURAL ADJUSTMENTS

- **Discretisation**: using quantile (equal frequency) [instead of Kmeans]
- **Algorithm stagnation:** removed max_uncovered_cases: executes until covers all, or dataset stagnation (20 its – no change in coverage) or heuristic goes to zero
- **Rule construction:** iteratively constructed by sorting from the set of items related to the partial-rule coverage [instead from the hole set of items]
    - Min_cases_per_rule: a percentual (5%) from population

TACKLING REDUNDANCY: SEARCH EXPLORATION

DIVERSITY

Dynamic heuristic function

Search Exploration
Tackling **description** and **cover** redundancy

TRANSITION RULE

$$p(c_{ij}) = \frac{\tau_{ij} * \eta_{ij}}{\sum_{\forall i,j}(\tau_{ij} * \eta_{ij})} *$$

Approached degrees of redundancy:

$$\delta_{ij} = 1 - \frac{1}{1 + e^{-(x_{ij}-5)}}$$

Logistic attenuation δ_{ij}: (DESCRIPTION)

- x_{ij} the counting of the term c_{ij} presence in the final rule model.

$$\psi_{ij} = \frac{1}{|c_{ij}|} \sum_{e \in c_{ij}} 0.9^{x(e,R)}$$

The cover-based attenuation: (COVERAGE)

- $|c_{ij}|$ is the size of the term c_{ij}, i.e. the number of examples e it covers; and
- $x(e,R)$ is how many times the example e is covered by any rule in the final rule model R.

$$\eta_{ESMAM} = \frac{\log_2 2 - H(W|a_i = v_{ij})}{\sum_{\forall i,j} \log_2 2 - H(W|a_i = v_{ij})} = \frac{\zeta_{ij}^{S}}{\sum_{\forall i,j} \zeta_{ij}^{S}}$$

$$\eta_{ESMAM-DS} = \frac{\psi_{ij} * \delta_{ij} * \zeta_{ij}^{D}}{\sum_{\forall i,j}(\psi_{ij} * \delta_{ij} * \zeta_{ij}^{D})}$$

[S] STATIC: over initial set (initialization)
[D] DINAMIC: over the uncovered set

OBS.: A term only has $\zeta_{ij}! = 0$ if it covers more than minimum cases

TACKLING REDUNDANCY: SEARCH EXPLORATION

DIVERSITY

New addition rule

RuleSet Diversity
Tackling **description** and **model** redundancy

```
R0: ('X204015_s_at', '[7.00,8.00)')
R1: ('X204015_s_at', '[7.00,8.00)') & ('X217815_at', '[10.00,10.00]')

R5: ('X202240_at', '[6.00,7.00)')
R6: ('X202240_at', '[4.00,6.00)')
```

SIMILAR **MODELS**
&
SIMILARITIES IN **DESCRIPTION**

→ KEEP **GENERALIZATION**

INITIAL DEFINITIONS:

$dataset: \Omega$

$A = (a_1, \dots, a_n)$ $-$ *descriptive attributes*

$V^i = (v_1^i, \dots, v_k^i) \forall i \in (1,n)$ $-$ *values of* a_i *domain*

$T = \{(a_i = \{v_j^i\}) | \forall a_i \in A \;\&\; v_j^i \in V^i\}$ – the set of Terms

$\boldsymbol{R^A}$ $\boldsymbol{-}$ ***the set of attributes in R description***

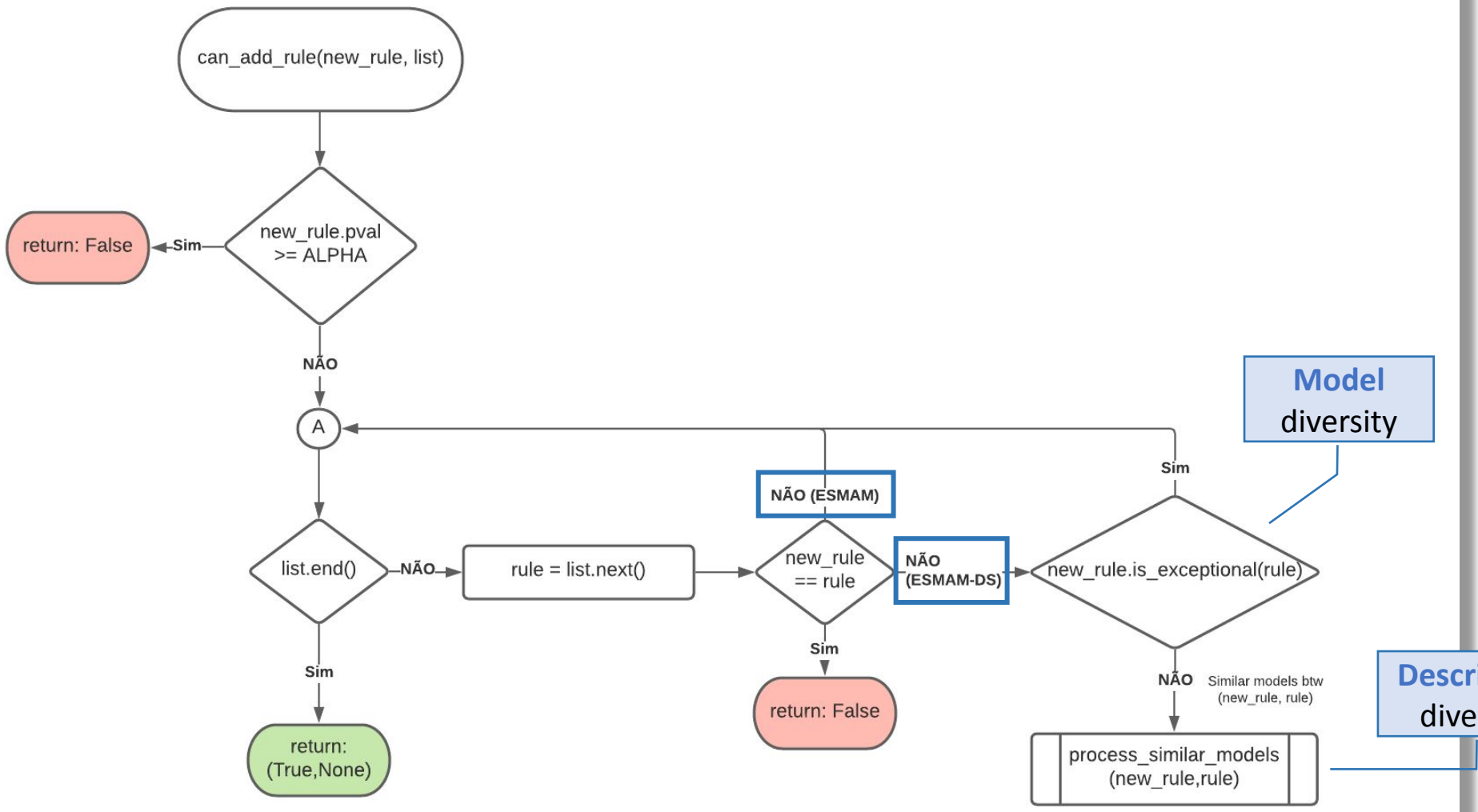
$\boldsymbol{R^T}$ $\boldsymbol{-}$ ***the set of terms in R description***

Rule R_1 **IS IN** R_2

$$R_1 \subset R_2 \leftrightarrow \begin{cases} R_1^A \cap R_2^A = R_2^A \\ R_1^T \subseteq R_2^T \end{cases}$$

DESCRIPTION GENERALITY

TACKLING REDUNDANCY: RULE SELECTION

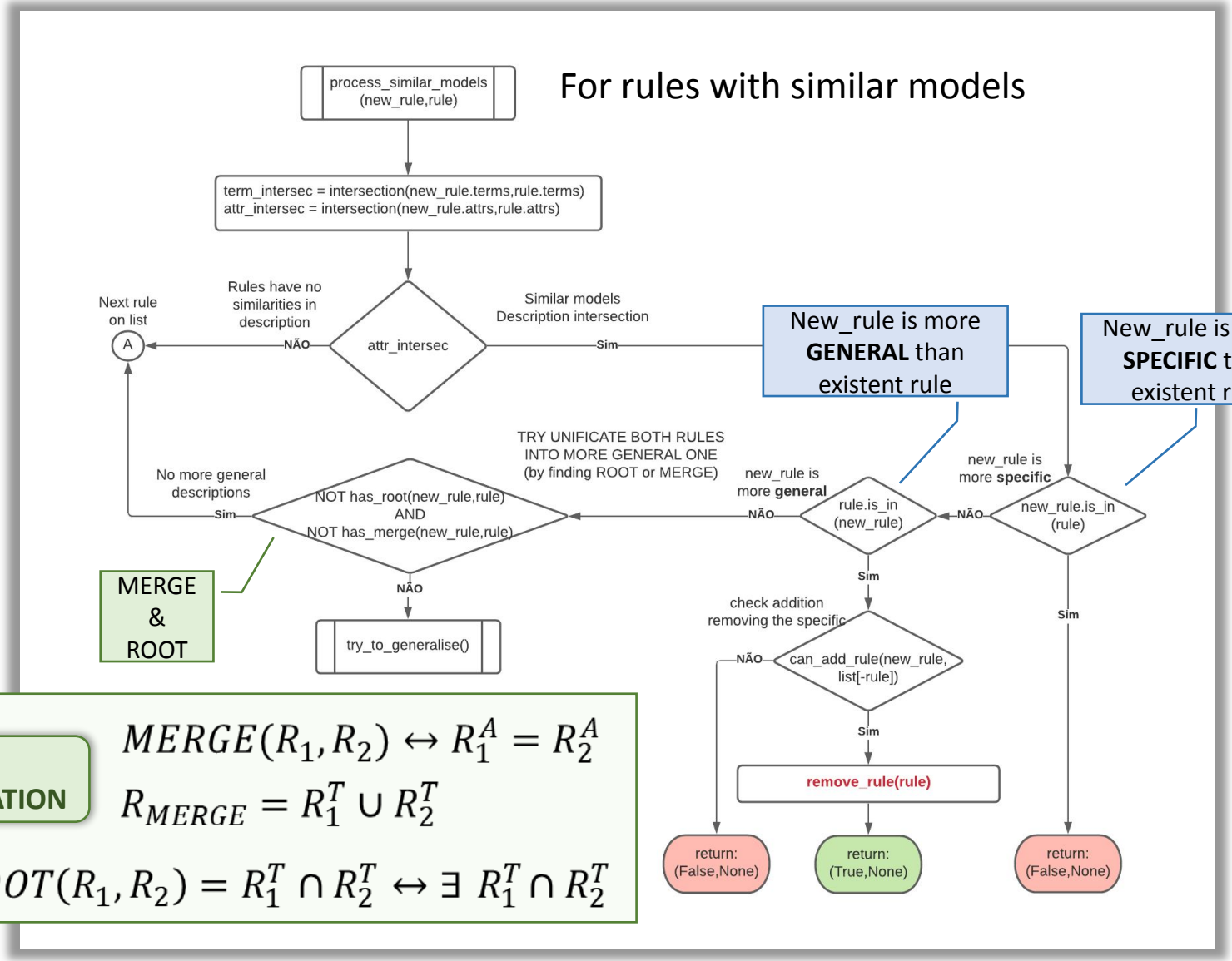


DIVERSITY

New addition rule

RuleSet Diversity
Tackling **description** and **model** redundancy

TACKLING REDUNDANCY: RULE SELECTION



DIVERSITY

New addition rule

RuleSet Diversity
Tackling **description** and **model** redundancy

GENERALIZATION

Rule R_1 **IS IN** R_2

$$R_1 \subset R_2 \leftrightarrow \begin{cases} R_1^A \cap R_2^A = R_2^A \\ R_1^T \subseteq R_2^T \end{cases}$$

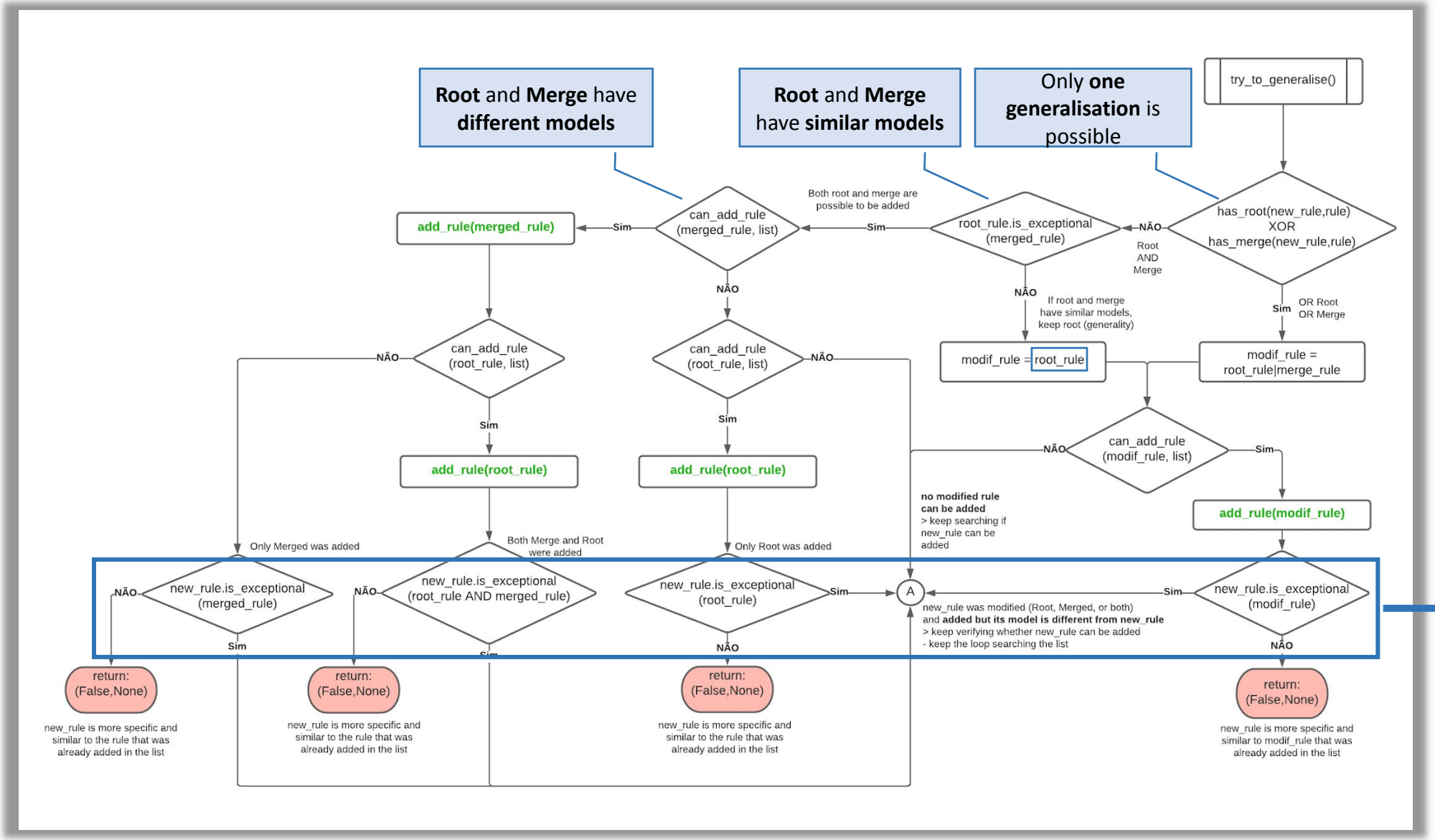
OTHER GENERALIZATION

$$MERGE(R_1, R_2) \leftrightarrow R_1^A = R_2^A$$

$$R_{MERGE} = R_1^T \cup R_2^T$$

$$ROOT(R_1, R_2) = R_1^T \cap R_2^T \leftrightarrow \exists\ R_1^T \cap R_2^T$$

TACKLING REDUNDANCY: RULE SELECTION



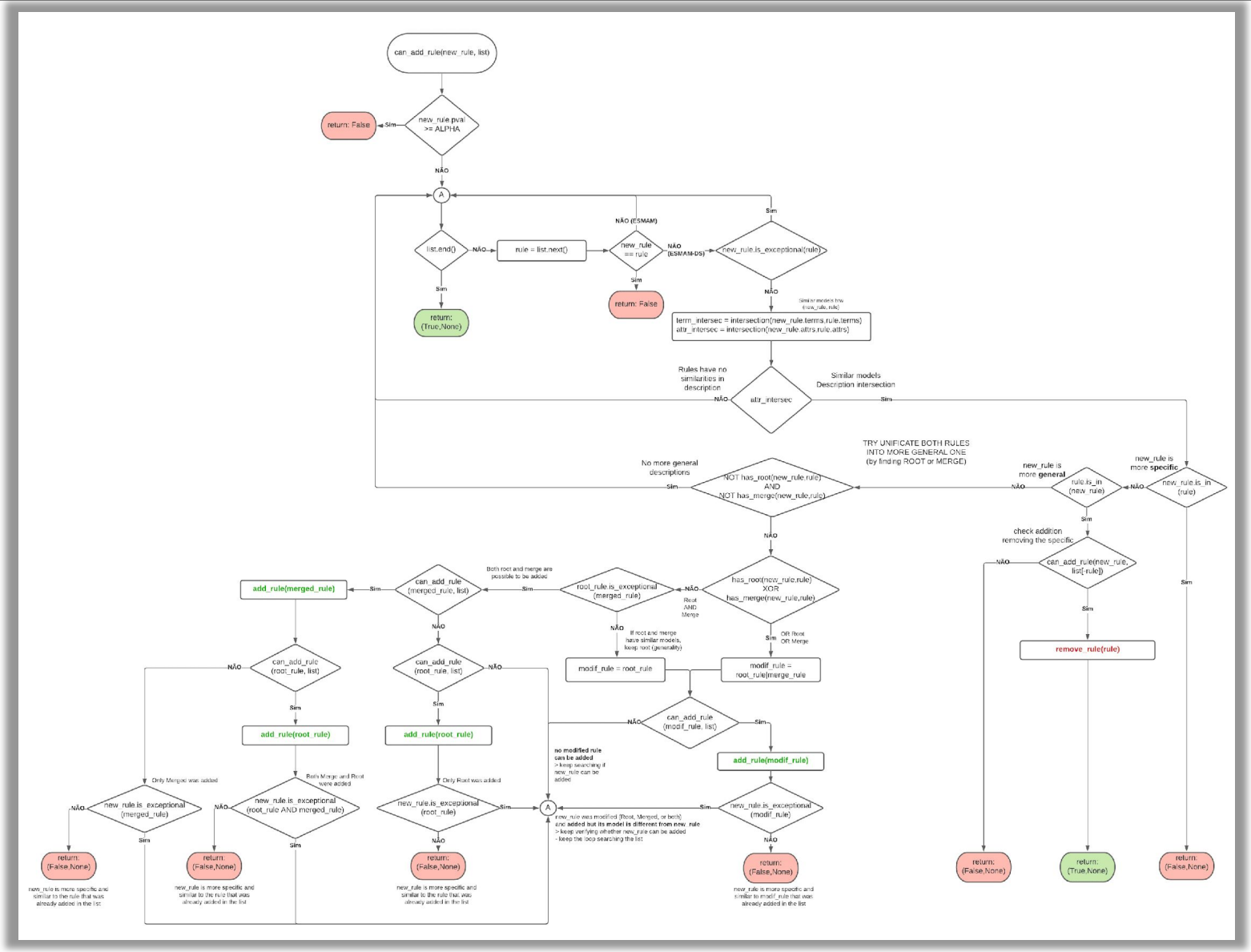
DIVERSITY

New addition rule

RuleSet Diversity
Tackling **description** and **model** redundancy

new_rule was generalised and the generalisation was added
- If new_rule model is different from the added generalisation > goes to next in list

TACKLING REDUNDANCY: RULE SELECTION

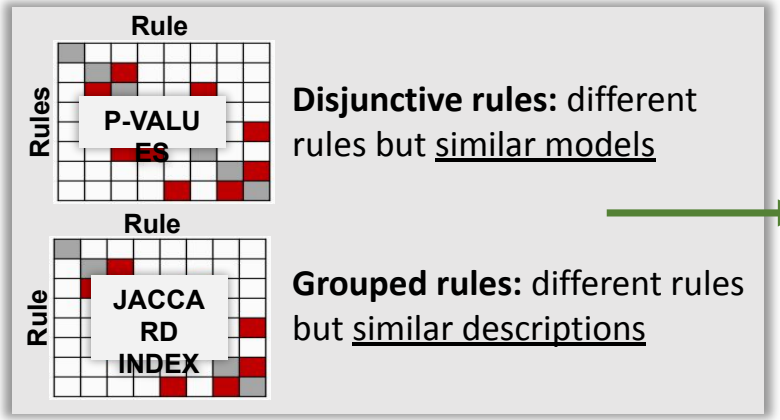


DIVERSITY

New addition rule

RuleSet Diversity
Tackling **description** and **model** redundancy

RESULTS DISPLAY



Disjunctive rules: different rules but <u>similar models</u>

Grouped rules: different rules but <u>similar descriptions</u>

RuleSet.txt

esmam-ds-cpm_breast-cancer_exp9_RuleSet.txt - Bloco de Notas

Arquivo Editar Formatar Exibir Ajuda

```
DISCOVERED SUBGROUPS

R0: (X204015_s_at=[7.00,8.00))
[SM] R4: (X201663_s_at=[9.00,9.00)) & (X210028_s_at=[9.00,9.00]) & (X205848_at=[5.00,6.00))
[SM] R6: (X221816_s_at=[10.00,10.00)) & (X201663_s_at=[9.00,9.00))
[SM] R8: (X205034_at=[8.00,8.00]) & (X210028_s_at=[9.00,9.00]) & (X200965_s_at=[10.00,10.00)) & (X211762_s_at=[10.00,11.00])
[SM] R9: (X218883_s_at=[9.00,10.00]) & (X209500_x_at=[9.00,9.00)) & (X221634_at=[7.00,7.00)) & (X204631_at=[5.00,5.00))
[SM] R10: (X205034_at=[8.00,8.00]) & (X200965_s_at=[10.00,10.00)) & (X210028_s_at=[9.00,9.00]) & (X202240_at=[7.00,8.00]) & (X219588_s
[SM] R11: (size=[2.00,2.00])

R1: (X203306_s_at=[9.00,10.00])
[SM] R3: (X202240_at=[6.00,7.00))
[SM] R5: (X217404_s_at=[3.00,5.00)) & (X217767_at=[12.00,12.00])
[SM] R7: (X221928_at=[6.00,7.00]) & (er=positive)

R2: (size=[1.00,2.00)) & (X217019_at=[7.00,7.00])

---------
[SM] similar model
[SD] similar description
```

Ln 14, Col 50 | 100% | Windows (CRLF) | UTF-8

HYPER-PARAMETERS

```
MIN_SIZE_SUBGROUP = 0.05
WEIGH_SCORE = 0.9
F_LOGISTIC_OFFSET = 5
ALPHA = 0.05

ITS_TO_STAGNATION = 20
RULES_TO_CONVERGENCE = 10
NUM_OF_ANTS = 3000

JACCARD_THRESHOLD = 0.5
```

Salvamento Automático | esmam-ds-cpm_breast-cancer_exp9_RuleModel.csv | Juliana Mattos

Arquivo | Página Inicial | Inserir | Layout da Página | Fórmulas | Dados | Revisão | Exibir | Desenvolvedor | Ajuda | Compartilhar | Comentários

M6

	sg	baseline	fitness	pvalue	mean_sg	mean_pop	mean_cpm	size_sg	size_pop	size_cpm
2	(X204015_s_at	complement	0.999423	0.000577	2907.148148	3935.316327	4099.579882	27	196	169
3	(X203306_s_at	complement	0.994025	0.005975	4219.22449	3935.316327	3651.408163	98	196	98
4	(size=[1.00,2.0	complement	0.999827	0.000173	4704.452381	3935.316327	3725.551948	42	196	154
5	(X202240_at=[(	complement	0.993422	0.006578	4778.021739	3935.316327	3676.886667	46	196	150
6	(X201663_s_at	complement	0.999753	0.000247	2618.5625	3935.316327	4052.361111	16	196	180
7	(X217404_s_at	complement	0.972896	0.027104	4510.657143	3935.316327	3810.242236	35	196	161
8	(X221816_s_at	complement	0.999932	0.000068	2914.21875	3935.316327	4134.554878	32	196	164
9	(X221928_at=[(	complement	0.995588	0.004412	4381.095238	3935.316327	3600.982143	84	196	112
10	(X205034_at=[(	complement	0.999912	0.000088	2612.5	3935.316327	4137.629412	26	196	170
11	(X218883_s_at	complement	0.999984	0.000016	2322.583333	3935.316327	4040.494565	12	196	184
12	(X205034_at=[(	complement	1	0	1913.133333	3935.316327	4102.900552	15	196	181
13	(size=[2.00,2.0	complement	0.999574	0.000426	3661.294574	3935.316327	4462.910448	129	196	67

esmam-ds-cpm_breast-cancer_exp9

Pronto

RuleModel.csv

Salvamento Automático | esmam-ds-cpm_breast-cancer_exp9_SurvivalModels.csv | Juliana Mattos

Arquivo | Página Inicial | Inserir | Layout da Página | Fórmulas | Dados | Revisão | Exibir | Desenvolvedor | Ajuda | Compartilhar | Comentários

S1

times	population	R0	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11
0	1	1	1	1	1	1	1	1	1	1	1	1	1
125	0.994898	1	0.989796	0.97619	1	0.9375	0.971429	0.96875	1	1	1	1	1
269	0.989796	0.962963	0.989796	0.97619	1	0.9375	0.971429	0.96875	1	1	1	1	0.992248
289	0.984694	0.962963	0.989796	0.97619	1	0.9375	0.971429	0.96875	1	1	1	1	0.984496
394	0.979592	0.962963	0.989796	0.97619	1	0.9375	0.971429	0.9375	1	0.961538	1	0.933333	0.976744
404	0.97449	0.925926	0.989796	0.97619	1	0.9375	0.971429	0.9375	1	0.923077	0.916667	0.866667	0.968992
421	0.969388	0.888889	0.979592	0.97619	1	0.875	0.971429	0.9375	1	0.884615	0.916667	0.8	0.96124
434	0.964286	0.888889	0.969388	0.97619	1	0.8125	0.971429	0.90625	1	0.884615	0.833333	0.8	0.953488
524	0.959184	0.888889	0.969388	0.97619	1	0.8125	0.971429	0.875	1	0.846154	0.833333	0.733333	0.945736
528	0.954082	0.851852	0.969388	0.97619	1	0.8125	0.971429	0.875	1	0.807692	0.75	0.666667	0.937984
530	0.94898	0.851852	0.969388	0.97619	1	0.8125	0.971429	0.875	1	0.807692	0.75	0.666667	0.930233
649	0.943878	0.851852	0.969388	0.97619	1	0.75	0.971429	0.84375	1	0.769231	0.75	0.666667	0.922481
690	0.938776	0.851852	0.959184	0.97619	1	0.75	0.971429	0.84375	1	0.730769	0.75	0.666667	0.914729
723	0.933673	0.851852	0.959184	0.97619	1	0.75	0.971429	0.8125	1	0.730769	0.75	0.666667	0.906977
730	0.928571	0.851852	0.94898	0.97619	1	0.75	0.971429	0.8125	0.988095	0.730769	0.75	0.666667	0.899225
794	0.923469	0.851852	0.94898	0.97619	1	0.6875	0.971429	0.78125	0.988095	0.730769	0.666667	0.666667	0.891473
796	0.918367	0.851852	0.938776	0.97619	1	0.6875	0.942857	0.75	0.988095	0.692308	0.666667	0.666667	0.891473
803	0.913265	0.814815	0.938776	0.97619	1	0.6875	0.942857	0.71875	0.988095	0.653846	0.583333	0.6	0.883721
805	0.908163	0.777778	0.938776	0.97619	1	0.6875	0.942857	0.71875	0.988095	0.653846	0.583333	0.6	0.875969
880	0.903061	0.740741	0.938776	0.97619	1	0.6875	0.942857	0.71875	0.988095	0.653846	0.583333	0.6	0.875969
910	0.897959	0.740741	0.938776	0.97619	1	0.6875	0.942857	0.71875	0.988095	0.653846	0.583333	0.6	0.868217
958	0.897959	0.740741	0.938776	0.97619	1	0.6875	0.942857	0.71875	0.988095	0.653846	0.583333	0.6	0.868217
994	0.892828	0.703704	0.938776	0.97619	1	0.6875	0.942857	0.71875	0.988095	0.653846	0.583333	0.6	0.860395
1023	0.887697	0.703704	0.938776	0.97619	1	0.6875	0.942857	0.71875	0.988095	0.653846	0.583333	0.533333	0.852574

esmam-ds-cpm_breast-cancer_exp9

SurvivalModels.csv

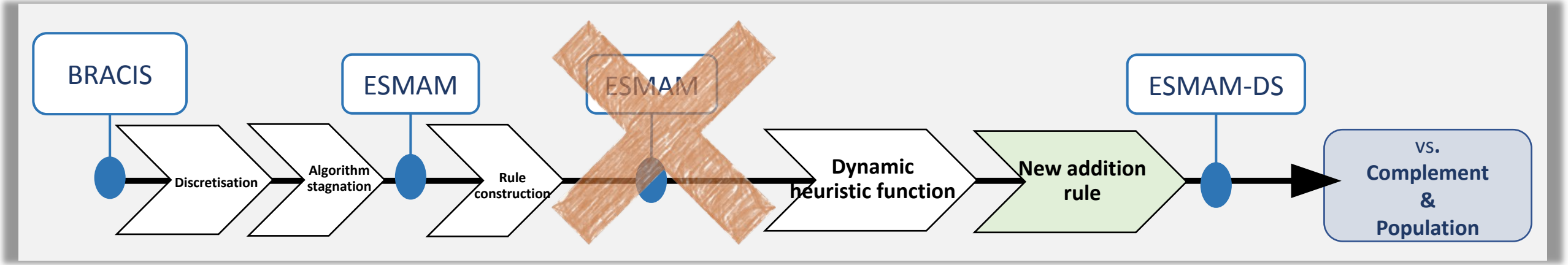
PARAMETERS

```
SG_BASELINE = ['complement',
               'population']
DIVERSE_SEARCH = [True,
                  False]
```

STATISTICAL TESTS

14 datasets
30 experiments

→ Keeping the same seed for equivalent executions (same dataset & experiment)



NEXT: adjust codes for computing results

⚠ **DISCUSS:** which metrics and results are to compute