**2. Data exploration**

* Total number of rows: 35945
* Total number of attributes: 4
* Datatypes & total number of unique, non-NULL values:
  + open: string/text, 2
  + single\_blind: string/text, 2
  + double\_blind: string/text, 2
  + masking: string/text, 4
* Total number of rows in which single\_blind takes value ‘Yes’ and masking takes value ‘1’: 287
* Total number of rows in which masking takes a value containing number ‘2’: 19767
* Total number of rows containing at least one NULL-value: 2083
* Total number of rows containing a NULL-value per attribute:
  + open: 1547
  + single\_blind: 1779
  + double\_blind: 1622
  + masking: 254

**3. Edit rules**

**3.1 Definition**

In the following table, we have listed, for each edit rule given in the assignment of the lab session, which attributes are involved in the corresponding rule.

|  |  |
| --- | --- |
| **edit rule** | **involved attributes** |
|  | open, double\_blind |
|  | single\_blind, double\_blind |
|  | open, single\_blind |
|  | single\_blind, masking |
|  | double\_blind, masking |
|  | double\_blind, masking |
|  | single\_blind, masking |

In the following table, we have listed, for each row given in the assignment of the lab session, which of the original edit rules are failed by the corresponding row.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **open** | **single\_blind** | **double\_blind** | **masking** | **fails/potentially fails edit rules...** |
| Yes | No | No | 2 |  |
| No | No | Yes | >2 | / |
| Yes | Yes | No | 0 | , |
| Yes | No | Yes | 2 |  |
| Yes | No | NULL | 0 | , |

Because the value of double\_blind is missing in the last row, we can not say for sure that it fails and , but it can fail these rules, if its value equals ‘Yes’.

Error detection on the clinical trials dataset:

* Total number of rows that fail at least one edit rule: 1222
* Total number of rows that fail : 301
* Total number of rows that fail and : 28
* Rows that fail the most edit rules:
  + open: Yes, single\_blind: Yes, double\_blind: Yes, masking: ‘0’
* Total number of rows that fail at least two edit rules: 72

**3.2 Error localization**

In the following table, we have listed, for each row given in the assignment of the lab session, what the minimal set covers of failing edit rules are for the corresponding row if all costs are equal to 1.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **open** | **single\_blind** | **double\_blind** | **masking** | **minimal set covers** |
| Yes | No | No | 2 | {double\_blind}, {masking} |
| No | No | Yes | >2 | {} |
| Yes | Yes | No | 0 | {single\_blind} |
| Yes | No | Yes | 2 | {open}, {double\_blind} |

In the following table, we have listed the minimal set covers of failing edit rules for each row if the costs for open and single\_blind change from 1 to 2.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **open** | **single\_blind** | **double\_blind** | **masking** | **minimal set covers** |
| Yes | No | No | 2 | {double\_blind}, {masking} |
| No | No | Yes | >2 | {} |
| Yes | Yes | No | 0 | {single\_blind} |
| Yes | No | Yes | 2 | {double\_blind} |

**3.3 Implication of edit rules**

In the following table, we have listed for each minimal set cover of a failing row, whether the minimal set cover is a correct solution or not. We also have provided an additional explanation, related to the first row, below the table.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **open** | **single\_blind** | **double\_blind** | **masking** | **minimal set covers** | **correct solution?** |
| Yes | No | No | 2 | {double\_blind} | No |
| Yes | No | No | 2 | {masking} | Yes |
| Yes | Yes | No | 0 | {single\_blind} | Yes |
| Yes | No | Yes | 2 | {open} | Yes |
| Yes | No | Yes | 2 | {double\_blind} | No |

The minimal set cover {double\_blind} of row 1 is not a correct solution, because changing the value of double\_blind from ‘No’ to ‘Yes’ implies that this row fails rule .

The minimal set cover {masking} of row 1 is a correct solution, because there exists at least one value of masking (‘0’), that makes this row satisfy all rules.

In the following table, we have listed the implied edit rule starting from the set and generator .

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | **essentially new?** |
| , | double\_blind | {Yes} {Yes} | No |
| , , | masking | *tautology* | No |
| , | double\_blind | *tautology* | No |
| , | single\_blind | {Yes} {1} | Yes |
| *,* | masking | {No} {No} {1,2} | No |

**3.4 Complete set generation**

In the following table, we have listed for each rule given in the assignment of the lab session, the rules of Table 1 that are dominating this rule.

|  |  |
| --- | --- |
|  | **dominating rules** |
| {Yes} {No} {Yes} {1} | , , |
| {Yes} {1} | / |
| {Yes} {Yes} {1,2} |  |

**FCF execution**

*Node (1)*

* Selected rules: &
* Essentially new rules: none
* Children of this node should not be visited anymore

*Node (2)*

* Selected rules: , , &
* Essentially new rules:
  + & : {Yes} {1} -> redundant to
  + & : {Yes} {1} ->

*Node (23)*

* Selected rules: , ,
* Essentially new rules:
  + & : {Yes} {2} ->

*Node (234)*

* Selected rules: ,
* Essentially new rules: none
* Node has no children

*Node (24)*

* Selected rules: , ,
* Essentially new rules: none
* Node has no children

*Node (3)*

* Selected rules: , , ,
* Essentially new rules:
  + & : {Yes} {2} -> redundant to (equal to)
  + & : {Yes} {2} ->

*Node (34)*

* Selected rules: , , ,
* Essentially new rules: none
* Node has no children

*Node (4)*

* Selected rules: , , ,
* Essentially new rules: none
* Node has no children

**Result: 3 essentially new, non-redundant rules**

* : {Yes} {1}
* : {Yes} {2}
* : {Yes} {2}

In the following table, we have listed, for each row given in the assignment of the lab session, which of the edit rules in the sufficient set are failed by the corresponding row, as well as the minimal set covers (and therefore, solutions) of the set of rules failed by the corresponding row.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **open** | **single\_blind** | **double\_blind** | **masking** | **fails/potentially fails edit rules...** | **minimal set covers** |
| Yes | No | No | 2 | , | {masking} |
| No | No | Yes | >2 | / | {} |
| Yes | Yes | No | 0 | , | {single\_blind} |
| Yes | No | Yes | 2 | , | {open} |
| Yes | No | NULL | 0 | , | {double\_blind} |