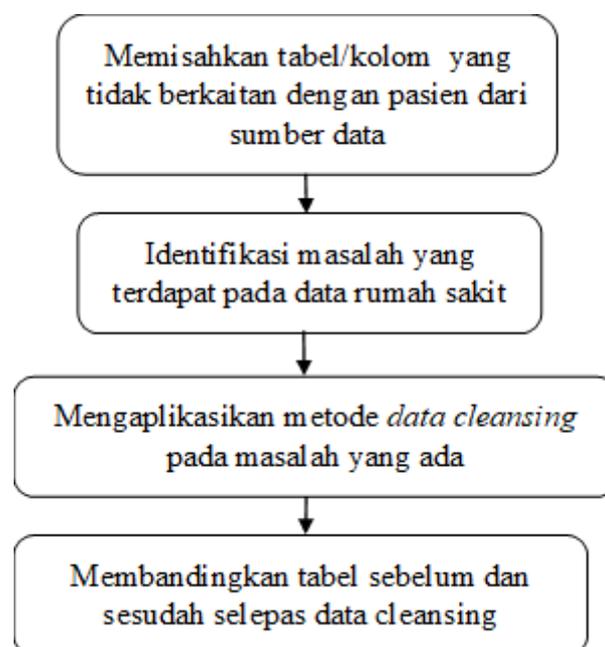


BAB III ANALISIS DATA DAN METODE

3.1 Tahapan Data Cleansing

Penelitian ini akan dilakukan beberapa tahapan atau langkah – langkah penelitian berikut ini:



Gambar 3.1 Langkah penelitian

Berikut adalah penjelasan tahapan – tahapan dari penelitian ini

3.1.1 Memisahkan tabel/kolom yang tidak berkaitan dengan pasien

Data yang akan diolah dengan *data cleansing* berasal dari *database* sistem informasi rumah sakit di suatu rumah sakit di Indonesia. *Database* merupakan *database Oracle* dengan besar data 500 Mb. Terdapat 171 tabel dalam *database* tersebut. Daftar tabel dengan jumlah kolom dan jumlah *records* dapat dilihat di LAMPIRAN.

Sesuai dengan batasan masalah dari penelitian ini yaitu meneliti tabel dari data sistem rumah sakit yang hanya berkaitan dengan pasien, maka dari 171 tabel tersebut tabel yang digunakan hanya 14 tabel. Tabel – tabel tersebut ialah:

a. ANTRIAN

Tabel ANTRIAN memiliki 214 *records* dengan 5 atribut dari tabel. Berikut atribut tersebut dengan tipe datanya:

Column Name	Data Type	Nullable	Default	Primary Key
ID	NUMBER(10,0)	No	-	1
ID_KUNJUNGAN	NUMBER(10,0)	No	-	-
ID_ANTRI	NUMBER(10,0)	No	-	-
ID_BED	NUMBER(10,0)	No	-	-
STATUS	VARCHAR2(255)	No	-	-
				1 - 5

Gambar 3.2 Tabel ANTRIAN

b. ASURANSI_KEPESERTAAN_VISIT

Tabel ASURANSI_KEPESERTAAN_VISIT memiliki 500 *records* dengan 5 atribut dari tabel. Berikut atribut tersebut dengan tipe datanya:

Column Name	Data Type	Nullable	Default	Primary Key
ID	NUMBER(10,0)	No	-	1
NO_POLIS	VARCHAR2(20)	No	-	-
ID_VISIT	NUMBER(24,0)	No	-	-
ID_ASURANSI_PRODUK	NUMBER(10,0)	No	-	-
ID_PERUSAHAAN_ASURANSI	NUMBER	Yes	-	-
				1 - 5

Gambar 3.3 Tabel ASURANSI_KEPESERTAAN_VISIT

c. BILLING

Tabel BILLING memiliki 500 *records* dengan 5 atribut dari tabel. Berikut atribut tersebut dengan tipe datanya:

Column Name	Data Type	Nullable	Default	Primary Key
ID	NUMBER(10,0)	No	-	1
ID_VISIT	NUMBER(24,0)	No	-	-
WAKTU	TIMESTAMP(6)	No	-	-
ID_PEGAWAI_PETUGAS	NUMBER(10,0)	No	-	-
STATUS PEMBAYARAN	VARCHAR2(4000)	No	-	-
TOTAL_TAGIHAN	FLOAT	Yes	-	-
JENIS	VARCHAR2(10)	Yes	-	-
1 - 7				

Gambar 3.4 Tabel BILLING

d. HASIL_PEMERIKSAAN_LAB

Tabel HASIL_PEMERIKSAAN_LAB memiliki 500 *records* dengan 8 atribut dari tabel. Berikut atribut tersebut dengan tipe datanya:

Column Name	Data Type	Nullable	Default	Primary Key
ID	NUMBER(11,0)	No	-	1
TANGGAL	TIMESTAMP(6)	No	-	-
KETERANGAN	VARCHAR2(4000)	Yes	-	-
ID_VISIT	NUMBER(24,0)	No	-	-
ID_PEGAWAI	NUMBER(10,0)	Yes	-	-
ID_DOKTER_PENGIRIM	NUMBER(10,0)	Yes	-	-
JENIS_DOKTER_PENGIRIM	VARCHAR2(50)	Yes	-	-
DOKTER_PENGIRIM_LUAR	VARCHAR2(500)	Yes	-	-
1 - 8				

Gambar 3.5 Tabel HASIL_PEMERIKSAAN_LAB

e. KUNJUNGAN

Tabel KUNJUNGAN memiliki 500 *records* dengan 17 atribut dari tabel. Berikut atribut tersebut dengan tipe datanya:

Column Name	Data Type	Nullable	Default	Primary Key
ID	NUMBER(24,0)	No	-	1
NO_ANTRIAN	NUMBER(24,0)	No	-	-
WAKTU	TIMESTAMP(6)	No	-	-
ID_PASIEN	NUMBER(10,0)	No	-	-
NO_KUNJUNGAN_PASIEN	NUMBER(10,0)	No	'1'	-
ID_LAYANAN	NUMBER(10,0)	Yes	-	-
ID_PENDUDUK_DPJP	NUMBER(10,0)	Yes	-	-
ID_PENDUDUK_PENANGGUNGJAWAB	NUMBER(10,0)	Yes	NULL	-
ID_PENDUDUK_PENGANTAR	NUMBER(10,0)	Yes	NULL	-
ID_RUJUKAN	NUMBER(10,0)	Yes	NULL	-
RENCANA_CARA_BAYAR	VARCHAR2(4000)	No	-	-
ID_BED	NUMBER(10,0)	Yes	-	-
STATUS	VARCHAR2(4000)	No	-	-
WAKTU_KELUAR	TIMESTAMP(6)	Yes	NULL	-
ID_VISIT	NUMBER(24,0)	No	-	-
ID_KUNJUNGAN_INDUK	NUMBER(24,0)	Yes	-	-
ID_PENDUDUK_KELUARGA	NUMBER(10,0)	Yes	-	-

Gambar 3.6 Tabel KUNJUNGAN

f. KUNJUNGAN_BPJS

Tabel KUNJUNGAN_BPJS memiliki 42 *records* dengan 22 atribut dari tabel. Berikut atribut tersebut dengan tipe datanya:

Column Name	Data Type	Nullable	Default	Primary Key
ID	NUMBER(24,0)	No	-	1
ID_BILLING	NUMBER(10,0)	No	-	-
ID_POLI_SEP	NUMBER(2,0)	Yes	-	-
KODE_PISAT	VARCHAR2(10)	No	-	-
KODE_JENIS_BPJS	NUMBER(2,0)	Yes	-	-
NAMA_JENIS_BPJS	VARCHAR2(100)	Yes	-	-
JENIS_PERAWATAN	NUMBER(2,0)	Yes	-	-
ID_KELAS_BPJS	NUMBER(2,0)	Yes	-	-
KELAS_BPJS	VARCHAR2(50)	Yes	-	-
ID_PROVIDER	VARCHAR2(100)	Yes	-	-
NAMA_PROVIDER	VARCHAR2(100)	Yes	-	-
NO_RUJUKAN	VARCHAR2(200)	Yes	-	-
TANGGAL_RUJUKAN	TIMESTAMP(6)	Yes	-	-
TANGGAL_SEP	TIMESTAMP(6)	No	-	-
ID_DIAGNOSA_AWAL	NUMBER(11,0)	No	-	-
CATATAN	CLOB	Yes	-	-
ID_VISIT	NUMBER(10,0)	No	-	-
NO_SEP	VARCHAR2(100)	Yes	-	-
LAKALANTAS	NUMBER(2,0)	Yes	-	-
LOKASILAKA	VARCHAR2(200)	Yes	-	-
ID_ASAL_RUJUKAN	VARCHAR2(200)	Yes	-	-
ASAL_RUJUKAN	VARCHAR2(1000)	Yes	-	-

1 - 22

Gambar 3.7 Tabel KUNJUNGAN_BPJS

g. LOG_BATAL_KUNJUNGAN

Tabel LOG_BATAL_KUNJUNGAN memiliki 15 *records* dengan 5 atribut dari tabel. Berikut atribut tersebut dengan tipe datanya:

Column Name	Data Type	Nullable	Default	Primary Key
ID	NUMBER(24,0)	No	-	1
TANGGAL_BATAL	TIMESTAMP(6)	No	-	-
ALASAN	VARCHAR2(1000)	Yes	-	-
INFO	VARCHAR2(4000)	Yes	-	-
ID_VISIT	NUMBER(10,0)	No	-	-
				1 - 5

Gambar 3.8 Tabel LOG_BATAL_KUNJUNGAN

h. PASIEN

Tabel g. PASIEN memiliki 500 *records* dengan 2 atribut dari tabel. Berikut atribut tersebut dengan tipe datanya:

Column Name	Data Type	Nullable	Default	Primary Key
ID	NUMBER(10,0)	No	-	1
ID_PENDUDUK	NUMBER(10,0)	No	-	-
				1 - 2

Gambar 3.9 Tabel PASIEN

i. RAWAT_DARURAT

Tabel RAWAT_DARURAT memiliki 57 *records* dengan 37 atribut dari tabel. Berikut atribut tersebut dengan tipe datanya:

Column Name	Data Type	Nullable	Default	Primary Key
ID	NUMBER(10,0)	No	-	1
ID_VISIT	NUMBER(24,0)	No	-	-
WAKTU	TIMESTAMP(6)	No	-	-
ID_ALASAN_DATANG	NUMBER(10,0)	No	-	-
KETERANGAN	CLOB	Yes	-	-
WAKTU Kejadian	TIMESTAMP(6)	Yes	-	-
TEMPAT Kejadian	VARCHAR2(50)	Yes	-	-
WAKTU_TIBA	TIMESTAMP(6)	Yes	-	-
TRANSPORTASI	VARCHAR2(50)	Yes	-	-
ID_PETUGAS	NUMBER(10,0)	No	-	-
WAKTU_PERIKSA	TIMESTAMP(6)	Yes	-	-
ID_JENIS_KASUS	NUMBER	Yes	-	-
WAKTU_MULAI	TIMESTAMP(6)	Yes	-	-
WAKTU_SELESAI	TIMESTAMP(6)	Yes	-	-
KATEGORI	NUMBER(10,0)	Yes	-	-
KET_KATEGORI	VARCHAR2(50)	Yes	-	-
WARNA	VARCHAR2(20)	Yes	-	-
ANAMNESE	CLOB	Yes	-	-
KEADAAN_UMUM	CLOB	Yes	-	-
KEADAAN_KHUSUS	CLOB	Yes	-	-
PENUNJANG	CLOB	Yes	-	-
OKSIGEN	VARCHAR2(50)	Yes	-	-
INFUS	VARCHAR2(50)	Yes	-	-
ORAL	VARCHAR2(50)	Yes	-	-
INJEKSI	VARCHAR2(50)	Yes	-	-
ID_TINDAK_LANJUT	NUMBER(10,0)	No	-	-
CATATAN	CLOB	Yes	-	-
ID_USER	NUMBER(10,0)	No	-	-
RUJUKAN	NUMBER(10,0)	No	-	-
ID_KEADAAN_KELUAR	NUMBER(10,0)	Yes	-	-
ID_CARA_KELUAR	NUMBER(10,0)	Yes	-	-
ID_KELUAR_PEMERIKSAAN_LANJUT	NUMBER(24,0)	Yes	-	-
ID_DOKTER_JAGA	NUMBER(10,0)	Yes	-	-
ID_PERAWAT_JAGA	NUMBER(10,0)	Yes	-	-
SHIFT	VARCHAR2(40)	Yes	-	-
ID_BED	NUMBER(10,0)	Yes	-	-
ID_RS_RUJUKAN	NUMBER(10,0)	Yes	-	-

1 - 37

Gambar 3.10 Tabel RAWAT_DARURAT

j. RAWAT_JALAN

Tabel RAWAT_ JALAN memiliki 63 *records* dengan 24 atribut dari tabel. Berikut atribut tersebut dengan tipe datanya:

Column Name	Data Type	Nullable	Default	Primary Key
ID	NUMBER(24,0)	No	-	1
ID_VISIT	NUMBER(24,0)	No	-	-
WAKTU	TIMESTAMP(6)	No	-	-
ID_BED	NUMBER	Yes	-	-
ID_DOKTER	NUMBER(10,0)	Yes	-	-
ANAMNESE	CLOB	Yes	-	-
ID_JENIS_KASUS	NUMBER(10,0)	Yes	-	-
ID_TINDAK_LANJUT	NUMBER(10,0)	Yes	-	-
CATATAN	CLOB	No	-	-
ID_USER	NUMBER(10,0)	Yes	-	-
TENSI	VARCHAR2(30)	No	-	-
NADI	VARCHAR2(30)	No	-	-
SUHU	VARCHAR2(30)	No	-	-
NAFAS	VARCHAR2(30)	No	-	-
TINGGI	VARCHAR2(30)	No	-	-
BERAT	VARCHAR2(30)	No	-	-
ID_KEADAAN_KELUAR	NUMBER(10,0)	Yes	-	-
ID_CARA_KELUAR	NUMBER(10,0)	Yes	-	-
ID_KELUAR_PEMERIKSAAN_LANJUT	NUMBER(10,0)	Yes	-	-
ID_RUJUKAN	NUMBER(10,0)	Yes	-	-
ID_UNIT_TUJUAN	NUMBER(10,0)	Yes	-	-
WAKTU_KONTROL	DATE	Yes	-	-
WAKTU_DIPERIKSA	TIMESTAMP(6)	Yes	-	-
TERAPI	CLOB	Yes	-	-
				1 - 24

Gambar 3.11 Tabel RAWAT_JALAN

k. PEMBAYARAN

Tabel PEMBAYARAN memiliki 500 *records* dengan 8 atribut dari tabel. Berikut atribut tersebut dengan tipe datanya:

Column Name	Data Type	Nullable	Default	Primary Key
ID	NUMBER(24,0)	No	-	1
ID_VISIT	NUMBER(24,0)	No	-	-
TOTAL_TAGIHAN	FLOAT	No	-	-
JUMLAH_BAYAR	FLOAT	No	-	-
SISA_TAGIHAN	FLOAT	No	-	-
TOTAL_BILLING_PENJUALAN	FLOAT	No	-	-
KLAIM_BPJS	FLOAT	No	-	-
DISKON	FLOAT	No	-	-
				1 - 8

Gambar 3.12 Tabel PEMBAYARAN

l. PENDUDUK

Tabel PENDUDUK memiliki 500 *records* dengan 10 atribut dari tabel. Berikut atribut tersebut dengan tipe datanya:

Column Name	Data Type	Nullable	Default	Primary Key
ID	NUMBER(10,0)	No	-	1
NO_IDENTITAS	VARCHAR2(30)	Yes	-	-
NAMA	VARCHAR2(50)	No	-	-
JENIS_KELAMIN	VARCHAR2(4000)	No	-	-
GOL_DARAH	CHAR(2)	Yes	-	-
TANGGAL_LAHIR	DATE	Yes	-	-
NO_KARTU_KELUARGA	VARCHAR2(20)	Yes	-	-
POSISI_DI_KELUARGA	VARCHAR2(4000)	Yes	-	-
SIP	VARCHAR2(50)	Yes	-	-
IBU_KANDUNG	VARCHAR2(300)	Yes	-	-
				1 - 10

Gambar 3.13 Tabel PENDUDUK

m. PENJUALAN_RESEP

Tabel PENJUALAN_RESEP memiliki 500 *records* dengan 4 atribut dari tabel. Berikut atribut tersebut dengan tipe datanya:

Column Name	Data Type	Nullable	Default	Primary Key
ID	NUMBER(10,0)	No	-	1
ID_PENJUALAN	NUMBER(10,0)	No	-	-
ID_VISIT	NUMBER(10,0)	Yes	-	-
ID_DOKTER	NUMBER(10,0)	Yes	-	-
				1 - 4

Gambar 3.14 Tabel PENJUALAN_RESEP

n. VISIT

Tabel VISIT memiliki 500 *records* dengan 6 atribut dari tabel. Berikut atribut tersebut dengan tipe datanya:

Column Name	Data Type	Nullable	Default	Primary Key
ID	NUMBER(24,0)	No	-	1
ID_PASIEN	NUMBER(10,0)	No	-	-
JENIS_KUNJUNGAN	NUMBER(2,0)	No	-	-
WAKTU	TIMESTAMP(6)	No	-	-
STATUS	NUMBER(2,0)	Yes	-	-
CARA_BOOKING	NUMBER(2,0)	Yes	-	-
				1 - 6

Gambar 3.15 Tabel VISIT

3.1.2 Identifikasi masalah yang terdapat pada data sistem rumah sakit

Permasalahan yang ditemukan berdasarkan observasi pada tiap tabel dan membandingkan tiap baris pada tabel data sistem informasi rumah sakit, penelitian ini mendapatkan beberapa permasalahan pada data kotor. Sehingga perlu dibersihkan menggunakan metode yang tepat sesuai permasalahannya. Untuk permasalahan yang ditemukan dibantu oleh kakas analisis, yaitu RapidMiner dikarenakan banyaknya data yang tersimpan tiap tabel dan agar lebih akurat dalam mengolah data.

Salah satu indikasi dalam data berkualitas adalah Akurasi. Akurasi didefinisikan sebagai hasil bagi dari jumlah nilai yang benar dalam data koleksi dan jumlah keseluruhan nilai. Sehingga data tidak dianggap akurat karena adanya duplikat data (Müller & Freytag, 2003).

Selain duplikasi data, permasalahan data kotor juga disebabkan oleh data tidak konsisten dan data kosong. Data tidak konsisten akan mempengaruhi Konsistensi yang merupakan indikasi dalam data berkualitas. Begitu juga dengan data kosong yang akan mempengaruhi

kualitasnya data, yaitu Kelengkapan. Sehingga melakukan data *cleansing* akan diperlukan untuk mengatasi permasalahan tersebut.

Dengan begitu, penelitian ini akan menjelaskan secara rinci mengenai permasalahan yang ditemukan dalam data sistem informasi rumah sakit ini adalah sebagai berikut:

Duplikasi Data

Permasalahan dalam mendeteksi dan mengeliminasi duplikasi data merupakan salah satu hal masalah yang utama dalam area *data cleansing* dan kualitas data pada *database*. Duplikasi data adalah satu atau lebih data yang memiliki kemiripan dan merujuk pada satu data atau objek yang unik (Tamilselvi & Gifta, 2011).

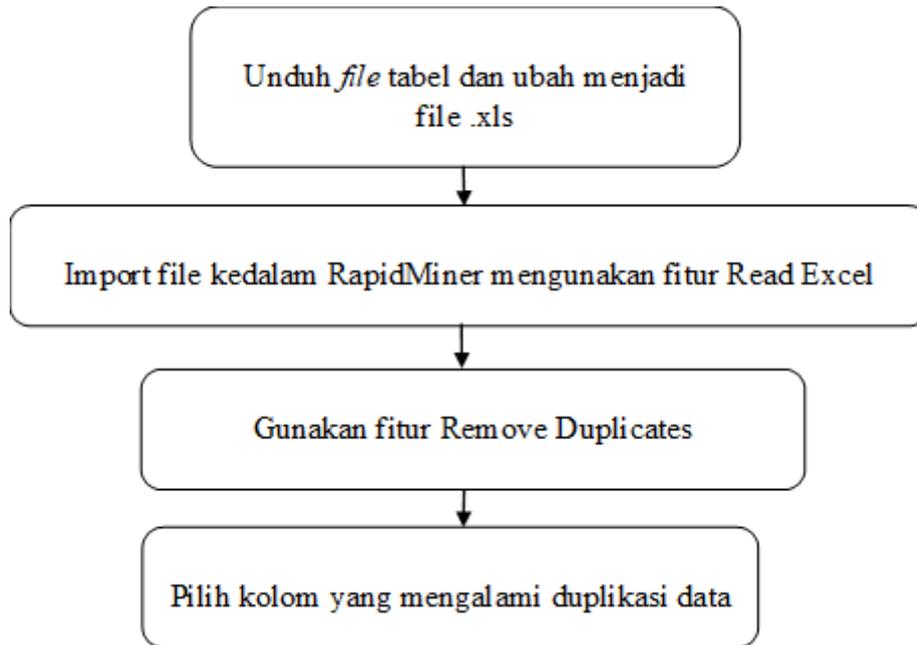
Sebagai sampel, dengan menggunakan tabel RAWAT_JALAN, hasil yang didapatkan adalah adanya duplikasi data pada tabel RAWAT_JALAN. Duplikasi data dapat dilihat sebagai berikut:

EDIT	ID	ID_VISIT	WAKTU	ID_BED	ID_DOKTER	ANAMNESE	ID_JENIS_KASUS	ID_TINDAK_LANJUT
	2263	82322	14-FEB-18 01.48.06.000000 PM	189	154270	-	-	55
	2264	82322	14-FEB-18 01.48.06.000000 PM	189	154270	-	-	55
	2265	82322	14-FEB-18 01.48.06.000000 PM	189	154270	-	-	55

Gambar 3.16 Duplikasi data pada tabel RAWAT_JALAN

Ditemukan pada ID 2263 – 2265 mengalami duplikasi data karena isi data selain pada kolom ID sama. Hal tersebut dapat dikatakan sebagai duplikasi data karena data pada kolom WAKTU sama persis, sehingga tidak relevan.

Agar lebih akurat, penelitian ini menggunakan RapidMiner sebagai alat bantu dalam menemukan duplikasi data. Langkah – langkah dalam menemukan duplikasi data sebagai berikut:



Gambar 3.17 Langkah pencarian duplikasi data

Berikut ini penjelasan dari tahap penemuan duplikasi data:

- a. Unduh *file* tabel dan ubah menjadi file *.xls*

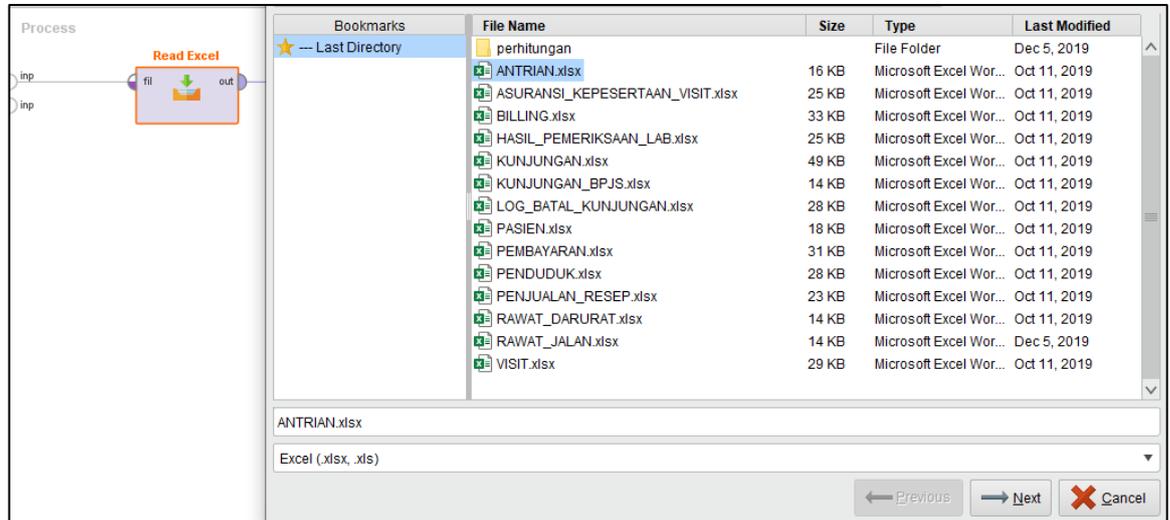
File diunduh dari database Oracle. File yang diunduh biasanya berformat *.csv* sehingga harus diubah dulu menjadi *spreadsheet* menggunakan Microsoft Excel dikarenakan RapidMiner menggunakan format *.xls*.

ROLE	2266	82404	14-FEB-18 01.49.45.000000 PM	186	52399	-	-
ROLE_PERMISSION	2267	82404	14-FEB-18 01.52.56.000000 PM	186	52399	-	-
RUJUKAN	2268	82405	14-FEB-18 01.56.30.000000 PM	189	52399	-	-
RUJUKAN_RS	2269	82406	14-FEB-18 01.57.58.000000 PM	186	52399	-	-
RUMAH_SAKIT	2270	82406	14-FEB-18 02.03.19.000000 PM	186	52399	-	-
SALDO							
SATUAN							
SATUAN_LAB							
SEDIAAN							
SEP_POLI_TUJUAN							
SETTING							
SPESIALISASI							
STOK							
STOK_OPNAME							
STOK_UNIT							
STOK_UNIT_AKHIR							
SUB_FARMAKOLOGI							

Gambar 3.18 Klik *Download*

b. Import file kedalam RapidMiner menggunakan fitur Read Excel

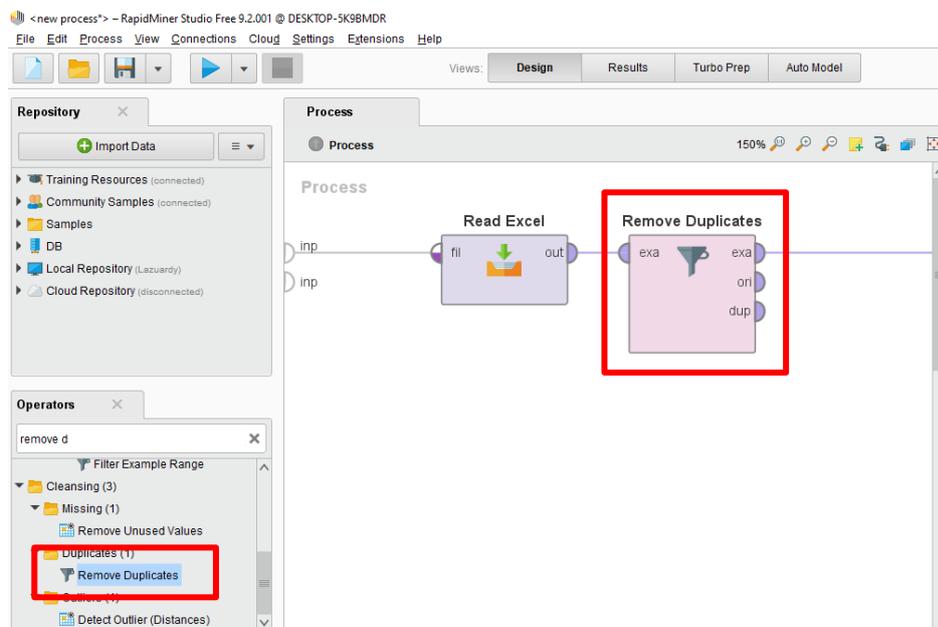
Fitur Read Excel terdapat pada menu Operators > Data Access > Files > Read > Read Excel.



Gambar 3.19 Penggunaan *Read Excel*

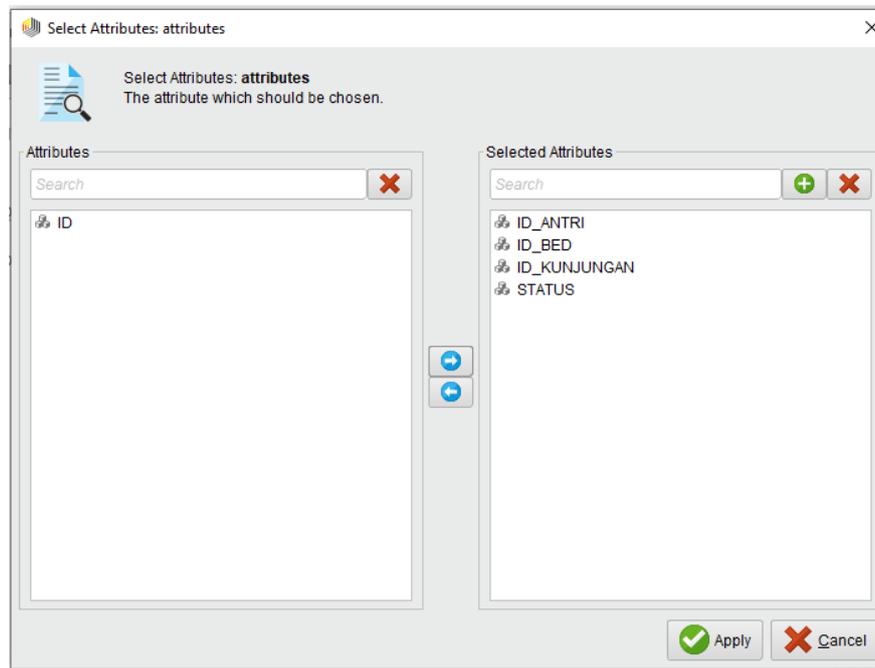
c. Lalu gunakan fitur Remove Duplicates

Fitur Remove Duplicates terdapat pada Operators > Cleansing > Duplicates > Remove Duplicates.



Gambar 3.20 Penggunaan *Remove Duplicates*

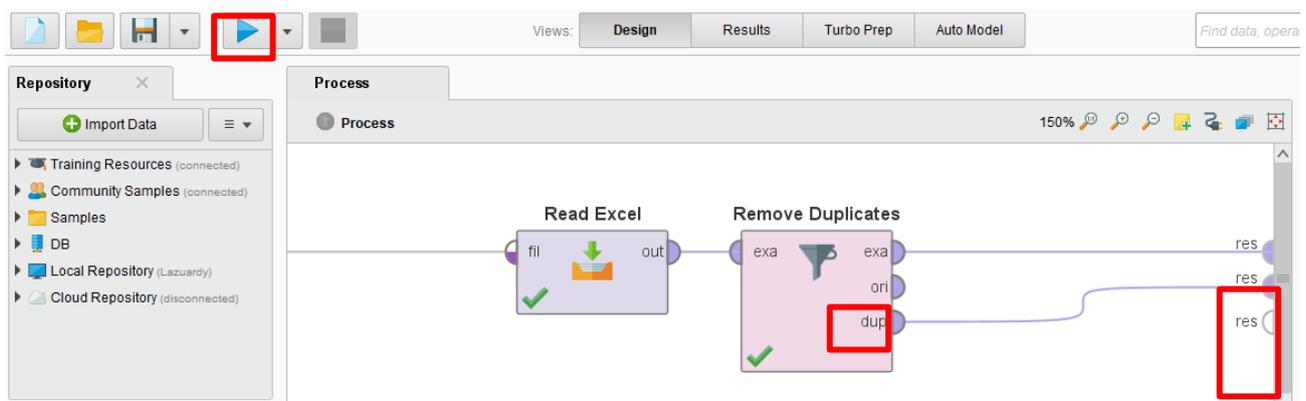
- d. Pilih kolom yang mengalami duplikasi data
Pilih berdasarkan atribut dalam tabel.



Gambar 3.21 Pilih indikator duplikasi data

- e. Sambungkan tulisan 'Dup' ke 'Res' lalu jalankan program

'Dup' pada Remove Duplicate bertujuan untuk menampilkan duplikasi data. Sedangkan 'Res' merupakan kependekan dari result yang berarti hasil dari olah data tersebut. Dijalankan dengan meng-klik tombol Start Execution yang bertanda segitiga biru.



Gambar 3.22 Hasil akhir

Data tidak konsisten

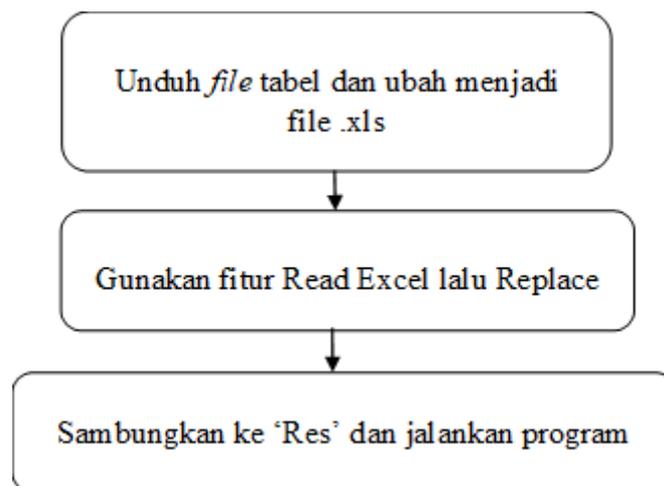
Tidak konsistennya data berkaitan dengan penggunaan nilai, unit, dan singkatan yang tidak seragam. Penggunaan nilai yang tidak konsisten akan memengaruhi keakuratan dan ketepatan suatu data (Müller & Freytag, 2003).

Sebagai sampel, tidak konsistennya data ditemukan pada tabel ASURANSI_KEPESERTAAN_VISIT. Ditemukan bahwa pada kolom NO_POLIS ada yang berisikan dengan '-' selain angka dengan menelusuri relasi tabel. Hal ini tentunya akan merusak sistem karena NO_POLIS merupakan *foreign key*.

NO_POLIS
1018624228
191418554
-
-

Gambar 3.23 Data tidak konsisten pada atribut NO_POLIS

Dengan begitu, untuk menemukan data tidak konsisten lainnya dapat menggunakan RapidMiner sebagai alat bantu. Berikut cara – cara menemukan data tidak konsisten:

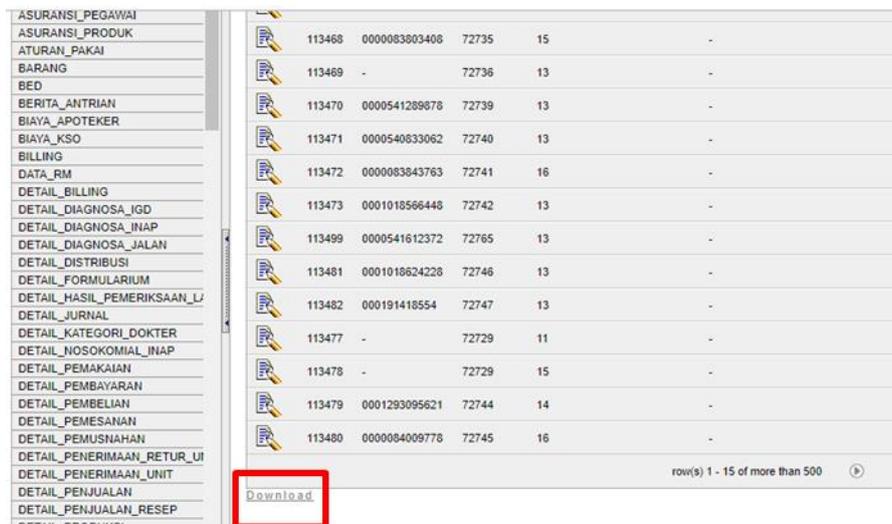


Gambar 3.24 Langkah menemukan data tidak konsisten

Berikut ini penjelasan dari tahap penemuan data tidak konsisten:

- a. Unduh *file* tabel dan ubah menjadi file .xls

Caranya sama dengan pencarian duplikat data. File tersebut diunduh dari database Oracle. File yang diunduh biasanya berformat .csv sehingga harus diubah dulu menjadi .xls menggunakan Microsoft Excel dikarenakan RapidMiner menggunakan format .xls.

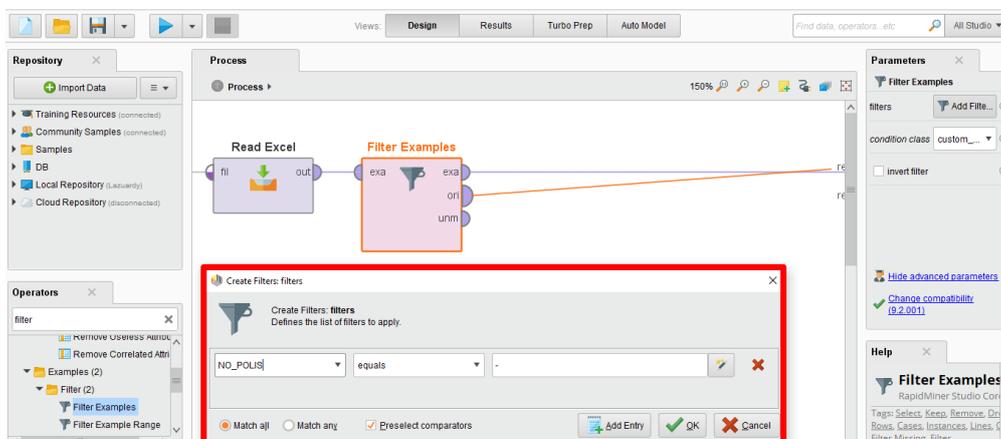


ASURANSI_PEGAWAI	ASURANSI_PRODUK	ATURAN_PAKAI	BARANG	BED	BERITA_ANTRIAN	BIAYA_APOTEKER	BIAYA_KSO	BILLING	DATA_RM	DETAIL BILLING	DETAIL DIAGNOSA_IJD	DETAIL DIAGNOSA_INAP	DETAIL DIAGNOSA_JALAN	DETAIL_DISTRIBUSI	DETAIL FORMULARIUM	DETAIL HASIL PEMERIKSAAN_LU	DETAIL JURNAL	DETAIL KATEGORI_DOKTER	DETAIL NOSOKOMIAL_INAP	DETAIL PEMAKAIAN	DETAIL PEMBAYARAN	DETAIL PEMBELIAN	DETAIL PEMESANAN	DETAIL PEMUSNAHAN	DETAIL PENERIMAAN_RETUR_UI	DETAIL PENERIMAAN_UNIT	DETAIL PENJUALAN	DETAIL PENJUALAN_RESEP	DETAIL PRODUK
113468	0000083803408	72735	15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
113469	-	72736	13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
113470	0000541289878	72739	13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
113471	0000540833062	72740	13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
113472	0000083843763	72741	16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
113473	0001018566448	72742	13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
113499	0000541612372	72765	13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
113481	0001018624228	72746	13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
113482	000191418554	72747	13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
113477	-	72729	11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
113478	-	72729	15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
113479	0001293095621	72744	14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
113480	0000084009778	72745	16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	

Gambar 3.25 Klik *Download*

- b. Lalu gunakan fitur Read Excel lalu Replace

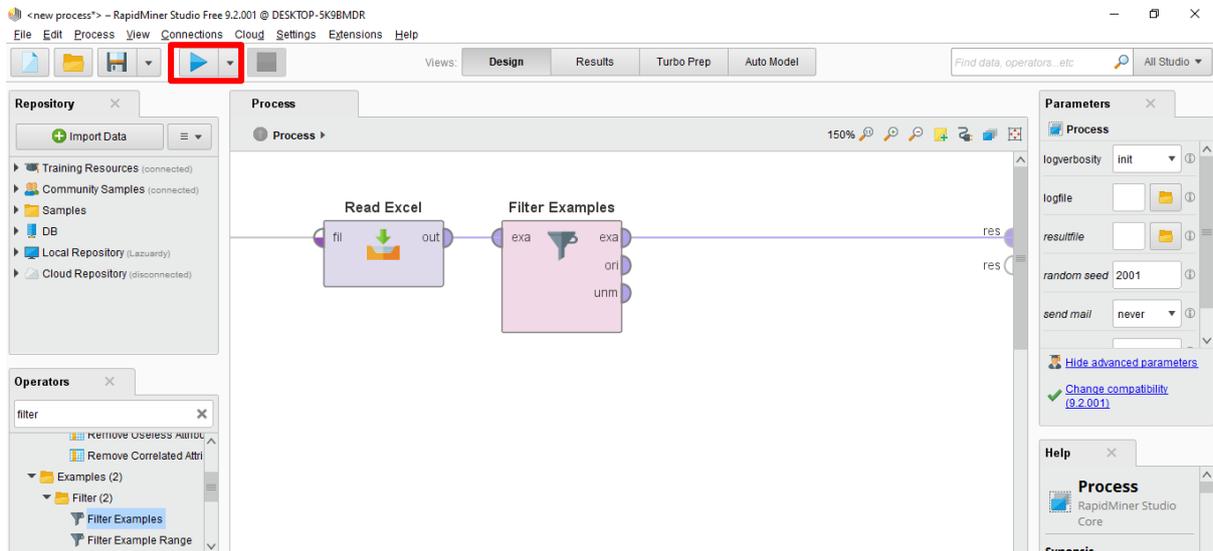
Fitur Read Excel untuk membaca data dan Fitur Replace sebagai filtrasi data yang tidak konsisten. Lalu isi indikator dalam kolom tabel yang dianggap sebagai anomali seperti tanda ‘-’.



Gambar 3.26 Penggunaan *Read Excel*

c. Selanjutnya sambungkan ke 'Res' dan jalankan program

Sebagai hasil yang didapatkan, data yang ditampilkan hanya yang mengalami duplikasi data.



Gambar 3.27 Penggunaan *Remove Duplicates*

Data kosong

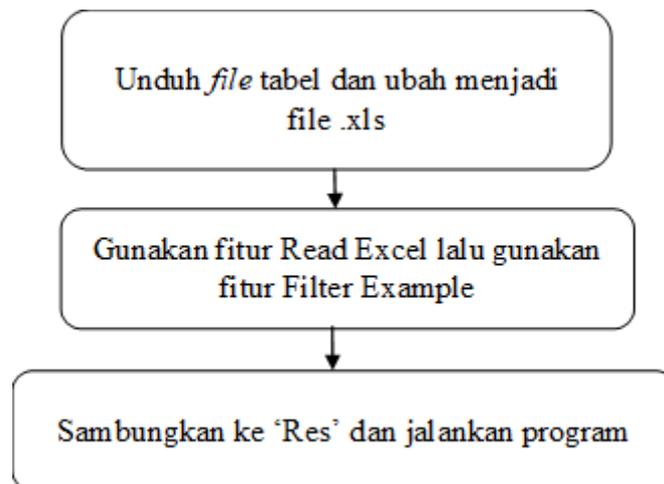
Data yang hilang atau tidak konsisten telah menjadi masalah yang meluas dalam analisis data sejak pengumpulan data asal. Deteksi sederhana data yang berisi nilai yang hilang dapat digunakan ketika pola data yang hilang tidak acak muncul dengan cara menghapus data yang terdapat data kosong (Brown & Kros, 2015).

Sebagai sampel, data kosong ditemukan pada tabel RAWAT_JALAN dengan cara memilih kolom ID_BED dan menyamakan dengan data kosong. Pada kolom ID_BED dan ID_DOKTER terdapat data kosong dengan ditandai dengan "?". Seharusnya dalam tabel RAWAT_JALAN, kolom ID_BED dan ID_DOKTER berisi kan nomor ID dan kolom – kolom ini menjadi *foreign key*.

ID	ID_VISIT	WAKTU	ID_BED	ID_DOKTER
2233	82352	31-JAN-18 03...	186	52399
2172	82302	18-JAN-18 12...	?	?
2173	81934	12-DEC-18 1...	?	?
2178	13426	19-JAN-18 12...	?	?
2179	82302	19-JAN-18 12...	?	?
2180	13426	19-JAN-18 12...	?	?
2181	13426	19-JAN-18 12...	?	?
2182	13426	19-JAN-18 12...	?	?
2183	13426	19-JAN-18 12...	?	?
2191	82107	23-JAN-18 12...	374	154270

Gambar 3.28 Data kosong pada atribut ID_BED

Dengan begitu, untuk menemukan data tidak konsisten lainnya dapat menggunakan RapidMiner sebagai alat bantu. Berikut cara – cara menemukan data tidak kosong:

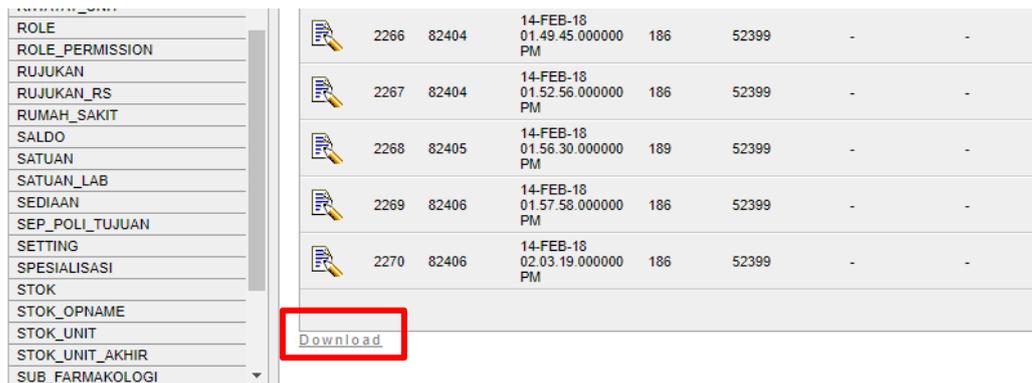


Gambar 3.29 Langkah menemukan data kosong

Berikut ini penjelasan dari tahap penemuan data kosong:

- a. Unduh *file* tabel dan ubah menjadi file .xls

Caranya sama dengan pencarian duplikat data. File tersebut diunduh dari database Oracle. File yang diunduh biasanya berformat .csv sehingga harus diubah dulu menjadi .xls menggunakan Microsoft Excel dikarenakan RapidMiner menggunakan format .xls.

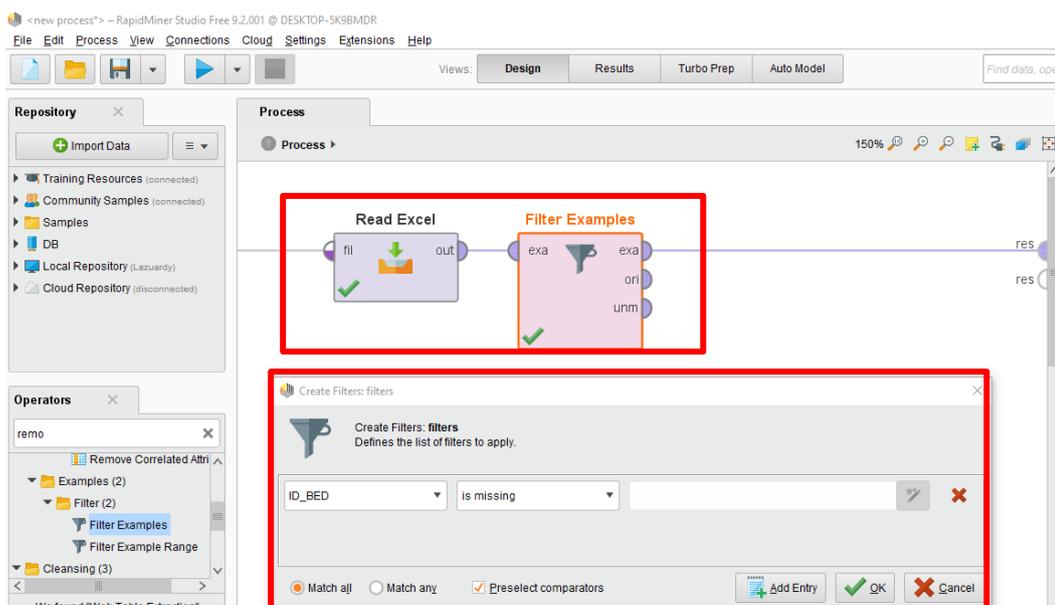


ROLE	ROLE_PERMISSION	RUJUKAN	RUMAH_SAKIT	SALDO	SATUAN	SATUAN_LAB	SEDIAAN	SEP_POLI_TUJUAN	SETTING	SPEKIALISASI	STOK	STOK_OPNAME	STOK_UNIT	STOK_UNIT_AKHIR	SUB_FARMAKOLOGI
2266	82404	14-FEB-18 01.49.45.000000 PM	186	52399	-	-									
2267	82404	14-FEB-18 01.52.56.000000 PM	186	52399	-	-									
2268	82405	14-FEB-18 01.56.30.000000 PM	189	52399	-	-									
2269	82406	14-FEB-18 01.57.58.000000 PM	186	52399	-	-									
2270	82406	14-FEB-18 02.03.19.000000 PM	186	52399	-	-									

Gambar 3.30 Klik *Download*

- b. Lalu gunakan fitur Read Excel gunakan fitur Filter Example

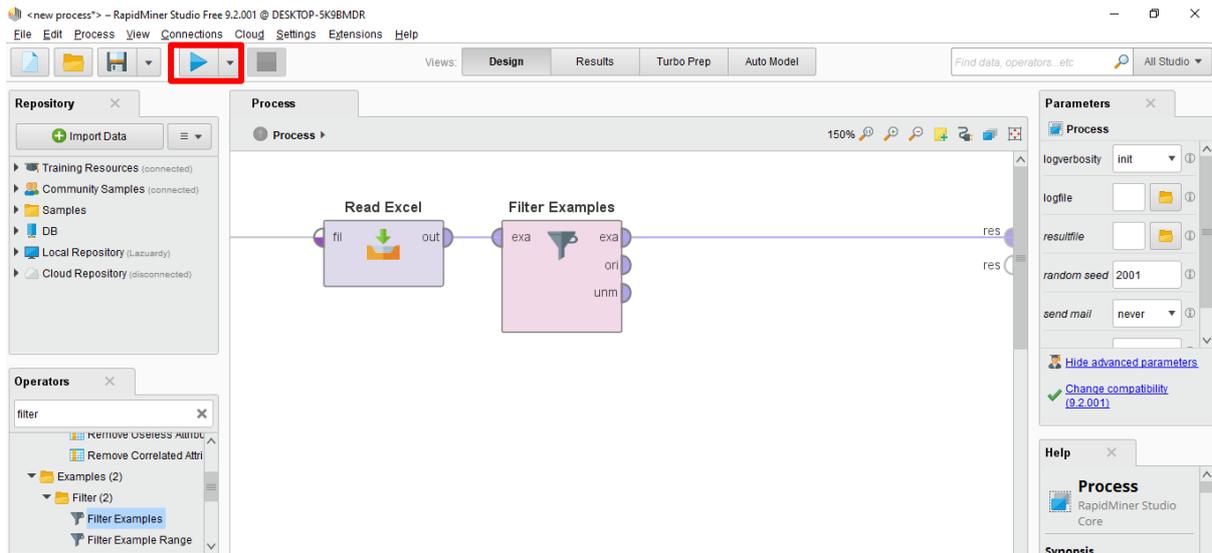
Fitur Read Excel untuk membaca data dan Fitur Filter Example sebagai filtrasi data yang kosong. Lalu isi indikator dalam kolom tabel yang dianggap sebagai anomali dengan 'is missing'.



Gambar 3.31 Isi indikator yang dianggap anomali

c. Selanjutnya sambungkan ke 'Res' dan jalankan program

Sebagai hasil yang didapatkan, data yang ditampilkan hanya yang mengalami data kosong. Data kosong akan ditandai dengan "?".



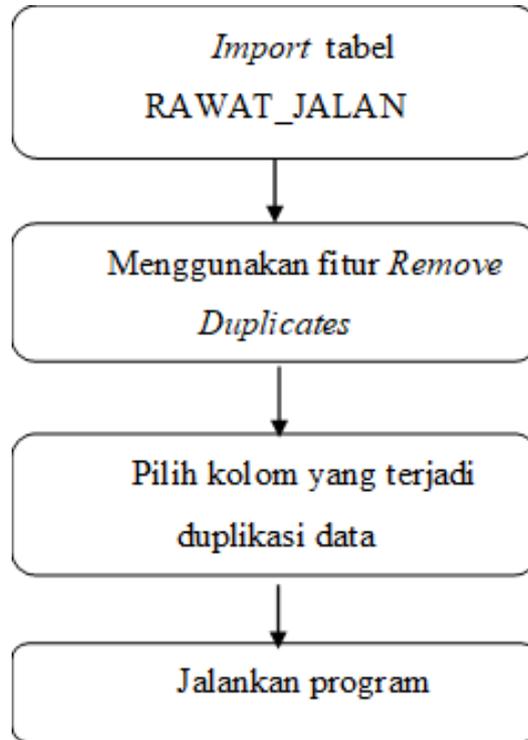
Gambar 3.32 Jalankan program

3.1.3 Mengaplikasikan metode data cleansing pada masalah yang ada

Dengan permasalahan yang ada pada data sistem rumah sakit dapat diatasi dengan menggunakan beberapa metode yang sudah disebutkan di atas. Pengimplementasian metode data cleansing menggunakan RapidMiner sebagai pembantu dalam mengolah data.

Penggunaan RapidMiner dalam mengatasi permasalahan yang terdapat pada data sistem rumah sakit:

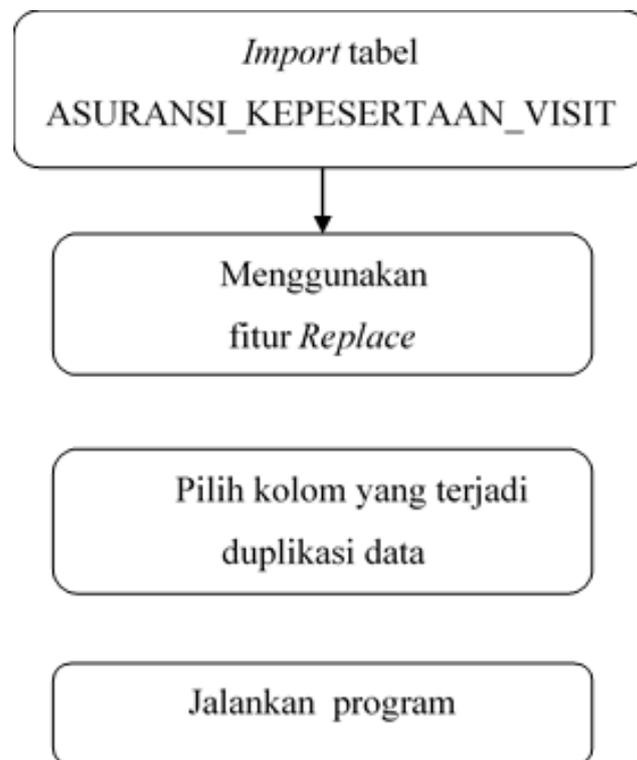
a. Menghilangkan duplikasi data dengan menggunakan metode *Duplicate Elimination*



Gambar 3.33 Tahap mengeliminasi data duplikasi

Dalam menghilangkan duplikasi data pada data sistem rumah sakit, dapat menggunakan RapidMiner sebagai kakas analisis. Duplikasi data dapat dihilangkan dengan fitur Remove Duplicates. Sebagai contoh duplikasi data terjadi pada tabel RAWAT_JALAN. Maka sebagai langkah awal yang dilakukan adalah import file .xls menggunakan fitur Read Excel pada tabel RAWAT_JALAN. Lalu menggunakan fitur Remove Duplicates. Selanjutnya memilih nama kolom yang terjadi duplikasi data pada tabel RAWAT_JALAN, yaitu ID_BED, ID_DOKTER, ID_TINDAK_LANJUT, ID_VISIT dan WAKTU. Langkah terakhir, jalankan program.

- b. Membuat data menjadi konsisten dengan metode *Inconsistency Detection*



Gambar 3.34 Tahapan menghilangkan data tidak konsisten pada RapidMiner

Permasalahan data yang tidak konsisten terjadi pada tabel ASURANSI_KEPESERTAAN_VISIT pada kolom NO_POLIS. Dengan menggunakan RapidMiner, langkah pertama adalah import file .xls menggunakan fitur Read Excel pada tabel ASURANSI_KEPESERTAAN_VISIT. Selanjutnya data tidak konsisten dapat diatasi menggunakan fitur Replace. Lalu ubah data ‘-‘ menjadi NO_POLIS yang default yaitu, “123456789”. Langkah terakhir, jalankan program.

c. data yang kosong dengan metode *Handling Missing Entries*



Gambar 3.35 Tahapan menghilangkan data tidak konsisten pada RapidMiner

Sebagai contoh data kosong terjadi pada tabel RAWAT_JALAN. Data kosong dapat dihilangkan dengan menggunakan fitur Remove Missing Values pada RapidMiner. Langkah pertama adalah import file .xls menggunakan fitur Read Excel pada tabel RAWAT_JALAN. Lalu menggunakan fitur Remove Missing Values dan masukkan atribut yang terdapat data kosong seperti ID_BED dan ID_DOKTER. Langkah terakhir, jalankan program.