

## **BAB IV**

### **HASIL DAN PEMBAHASAN**

#### **4.1 Hasil Penelitian**

Pada penelitian ini menggunakan Algoritma *K-Nearest Neighbor* dengan menerapkan perhitungan *Euclidean Distance* untuk memprediksi kelayakan penerima bantuan rehabilitasi Rumah Tidak Layak Huni (RTLH) memiliki lima tahapan penelitian. Kelima tahapan tersebut antara lain tahap pengumpulan data, tahap pengolahan data awal, tahap pengujian model *k-fold cross validation*, tahap Algoritma *K-Nearest Neighbor* serta Evaluasi dan validasi. Adapun penjelasan yang lebih lengkap terkait tahapan penelitian tersebut akan diuraikan sebagai berikut.

##### **4.1.1 Pengumpulan Data**

Tahapan pengumpulan data pada penelitian ini dilakukan di Dinas Perumahan Rakyat dan Kawasan Permukiman (DISPERKIM) Jepara dibagian Administrasi Bidang Perumahan. Tahapan ini dilakukan dengan cara observasi dan wawancara perihal Penerimaan Bantuan Rehabilitasi Rumah Tidak Layak Huni (RTLH). Data diperoleh sebanyak 2.962 dalam bentuk Microsoft Excel sehingga dapat mempermudah dalam pengolahan data. Data tersebut diambil dari pendaftar dan penerima bantuan rehabilitasi Rumah Tidak Layak Huni (RTLH) tahun 2019, terdiri dari 2.932 data yang sudah diketahui kategori hasilnya dan 30 data yang belum diketahui hasilnya.

##### **4.1.2 Pengolahan Data Awal**

Pada tahapan pengumpulan data diperoleh data *training* sebanyak 2932 data dan data *testing* sebanyak 30 data dengan total data 2.962 data yang memiliki atribut nama, NIK, kecamatan, desa, RT, RW, fasilitator, hasil verifikasi, kepemilikan rumah dan tanah, ukuran rumah, jumlah anggota keluarga, pondasi, balok, sanitasi, jendela, ventilasi, kondisi atap, material atap, kondisi lantai,

material lantai, kondisi dinding dan material dinding. Namun tidak semua atribut digunakan karena harus melalui tahapan pengolahan data awal. Adapun proses pengolahan data awal menggunakan *software* Microsoft Excel terlebih dahulu baru setelah didapatkan hasil data yang sesuai untuk dilakukan penelitian maka dilanjutkan dengan pengujian menggunakan *software* RapidMiner. Data tersebut kemudian dilakukan *Pre-processing* data sebagai berikut:

#### 4.1.2.1 Seleksi Atribut

Pada tahap ini dilakukan pengambilan atribut tertentu dari semua atribut pada data awal, dimana atribut yang diambil berpengaruh pada proses perhitungan. Penyeleksian data dilakukan melalui *software* Microsoft Excel terlebih dahulu. Untuk data awal RTLH dan Non RTLH sebelum diseleksi terdapat 22 atribut sebagai berikut.

LAPORAN HASIL SURVEY APBD ANGGARAN TAHUN 2019																						
NO	NAMA	NIK	KECAMATAN	DESA	RT	RW	FASILITATOR	HASIL VERIFIKASI	KEPEMILIKAN TANAH & TANAH	UKURAN RUMAH	JUMLAH ANGGOTA KELUARGA	KESELAMATAN DAN KESEHATAN					KONDISI RUMAH					
												PONDASI	BALOK	SANITASI	JENDELA	VENTILASI	KONDISI ATAP	MATERIAL	KONDISI LANTAI	MATERIAL	KONDISI DINDING	
1	TEMU		BANGSRI	BANGSRI	RT 04	RW 12	ARIS AHMAD MUSTAFAK	NON RTLH	MLIK SENDIRI	35	2	LAYAK	LAYAK	TKD	ADA	ADA	BOCOR	GENTENG	PUSAK BERAT	TANAH	TIDAK RUSAK	BATA
2	ABDUL ALM		BANGSRI	BANGSRI	RT 04	RW 13	ARIS AHMAD MUSTAFAK	RTLH	MLIK SENDIRI	35	2	TKD	TKD	TKD	ADA	ADA	BOCOR BERAT	GENTENG	PUSAK BERAT	TANAH	RUSAK SEDANG	KAYU
3	RTI KHASA DAM	320084902500	BANGSRI	BANGSRI	RT 01	RW 12	ARIS AHMAD MUSTAFAK	RTLH	MLIK SENDIRI	35	1	LAYAK	LAYAK	TKD	ADA	ADA	BOCOR	GENTENG	PUSAK BERAT	PLESTER	RUSAK BERAT	BATA/BAMBU
4	TUMILAN	32008060250001	BANGSRI	BANGSRI	RT 03	RW 12	ARIS AHMAD MUSTAFAK	RTLH	MLIK SENDIRI	35	2	LAYAK	LAYAK	TKD	ADA	ADA	BOCOR SEDANG	GENTENG	PUSAK SDG	TANAH	RUSAK BRT	TRIPLEK
5	SITI AMINAH		BANGSRI	BANGSRI	RT 03	RW 15	ARIS AHMAD MUSTAFAK	RTLH	MLIK SENDIRI	35	2	TKD	TKD	TKD	ADA	ADA	BOCOR	GENTENG	PUSAK BERAT	TANAH	RUSAK BRT	KAYU
6	PANCAH		BANGSRI	BANGSRI	RT 05	RW 15	ARIS AHMAD MUSTAFAK	RTLH	MLIK SENDIRI	48	2	TKD	TKD	TKD	ADA	ADA	BOCOR BRT	GENTENG	PUSAK BERAT	TANAH	RUSAK BRT	KAYU
7	HENDRI K. SANTO SO	320080204630002	BANGSRI	BONDO	RT 02	RW 05	ACHMAD BAGAS VIRATAMA	NON RTLH	MLIK SENDIRI	80	4	LAYAK	LAYAK	LAYAK	ADA	ADA	TIDAK BOCOR	GENTENG	TIDAK RUSAK	PLESTER	TIDAK RUSAK	BATA
8	SARFA VI	320081907420001	BANGSRI	BANJAR AGUNG	RT 03	RW 02	ZULISTIANA MUFADAH	RTLH	MLIK SENDIRI	48	1	TIDAK LAYAK	TIDAK LAYAK	ADAM CEMPLUNG	TIDAK ADA	TIDAK ADA	BERAT	GENTENG	BERAT	TANAH	BERAT	KAYU
9	WAWINI	320080509700002	BANGSRI	BANJAR AGUNG	RT 03	RW 02	ZULISTIANA MUFADAH	RTLH	MLIK SENDIRI	24	2	LAYAK	LAYAK	ADAM LEHER ANGSA	ADA	ADA	TIDAK BOCOR	GENTENG	BERAT	TANAH	TIDAK RUSAK	BATA
10	JUMARI YAH	320080508080001	BANGSRI	BANJAR AGUNG	RT 03	RW 02	ZULISTIANA MUFADAH	NON RTLH	MLIK SENDIRI	35	2	LAYAK	LAYAK	ADAM LEHER ANGSA	ADA	ADA	TIDAK BOCOR	GENTENG	SEDANG	PLESTER	TIDAK RUSAK	PLESTER
11	KEMILAH	320080509700004	BANGSRI	BANJAR AGUNG	RT 01	RW 04	ZULISTIANA MUFADAH	RTLH	MLIK SENDIRI	35	1	TIDAK LAYAK	TIDAK LAYAK	ADAM LEHER ANGSA	TIDAK ADA	TIDAK ADA	BERAT	GENTENG	BERAT	TANAH	BERAT	KAYU
12	BAROKAH	320080509700004	BANGSRI	BANJAR AGUNG	RT 02	RW 04	ZULISTIANA MUFADAH	RTLH	MLIK SENDIRI	24	3	LAYAK	LAYAK	TKD	ADA	ADA	BERAT	TIDAK ADA ATAP	BERAT	TANAH	TIDAK RUSAK	BATA
13	MASRUROTUN	32008050704800002	BANGSRI	BANJAR AGUNG	RT 02	RW 04	ZULISTIANA MUFADAH	RTLH	MLIK SENDIRI	24	3	LAYAK	LAYAK	ADAM LEHER ANGSA	TIDAK ADA	TIDAK ADA	SEDANG	GENTENG	BERAT	PLESTER	BERAT	KAYU

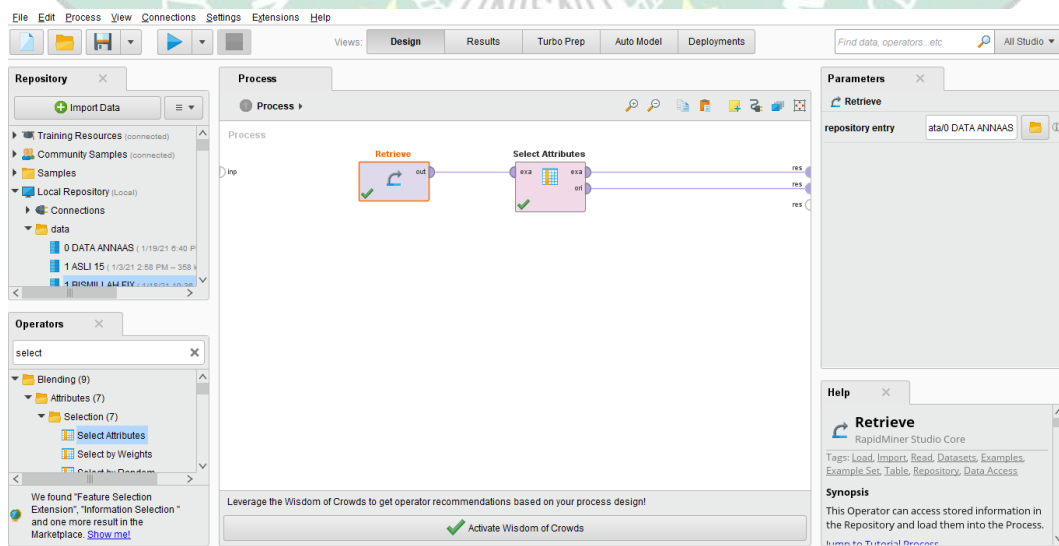
Gambar 4.1 Data RTLH dan Non RTLH Sebelum Seleksi Atribut

Selanjutnya memilih atribut yang digunakan untuk penelitian yaitu sebanyak 13 atribut terdiri dari nama, kepemilikan rumah dan tanah, ukuran rumah, jumlah anggota keluarga, pondasi, balok, sanitasi, jendela, ventilasi, material atap, material lantai, material dinding dan hasil verifikasi dengan *output* RTLH atau NON RTLH sebagai berikut.

NO	NAMA	KEPEMILIKAN RUMAH DAN TANAH	JUMLAH ANGGOTA KELUARGA	UKURAN RUMAH	PONDASI	BALOK	SANTIASI	JENDELA	VENTILASI	MATERIA L ATAP	MATERIA LANTAI	MATERIA LDINDING	HASIL
1	TEMU	MILIK SENDIRI	2	35	LAYAK	LAYAK	LAYAK	ADA	ADA	GENTENG	TANAH	BATA EXPOSE	NON RTLH
2	ABDUL ALIM	MILIK SENDIRI	2	35	TIDAK LAYAK	TIDAK LAYAK	ADA	ADA	ADA	GENTENG	TANAH	KAYU	RTLH
3	SITI KHASANAH	MILIK SENDIRI	1	35	LAYAK	LAYAK	TIDAK ADA	ADA	ADA	GENTENG	PLESTER	BATA EXPOSE	RTLH
4	TUMIJAN	MILIK SENDIRI	2	35	TIDAK LAYAK	LAYAK	TIDAK ADA	ADA	ADA	GENTENG	PLESTER	TRIPLEK	RTLH
5	SITI AMINAH	MILIK SENDIRI	2	35	TIDAK LAYAK	TIDAK LAYAK	TIDAK ADA	ADA	ADA	GENTENG	TANAH	KAYU	RTLH
6	PAIMAH	MILIK SENDIRI	2	48	TIDAK LAYAK	TIDAK LAYAK	TIDAK ADA	ADA	ADA	GENTENG	TANAH	KAYU	RTLH
7	HENDRIK SANTOSO	MILIK SENDIRI	4	80	LAYAK	LAYAK	LAYAK	ADA	ADA	GENTENG	PLESTER	BATA EXPOSE	NON RTLH
8	SARKAWI	MILIK SENDIRI	1	48	TIDAK LAYAK	TIDAK LAYAK	ADA	TIDAK ADA	TIDAK ADA	GENTENG	TANAH	KAYU	RTLH
9	WAKINI	MILIK SENDIRI	2	24	LAYAK	LAYAK	ADA	ADA	ADA	GENTENG	TANAH	BATA EXPOSE	RTLH
10	JUMARIYAH	MILIK SENDIRI	2	35	LAYAK	LAYAK	ADA	ADA	ADA	GENTENG	PLESTER	BATA PLESTER	NON RTLH
11	KEMUAH	MILIK SENDIRI	1	35	TIDAK LAYAK	TIDAK LAYAK	ADA	TIDAK ADA	TIDAK ADA	GENTENG	TANAH	KAYU	RTLH
12	BAROKAH	MILIK SENDIRI	3	24	LAYAK	LAYAK	TIDAK ADA	ADA	ADA	RUMBIA	TANAH	BATA EXPOSE	RTLH

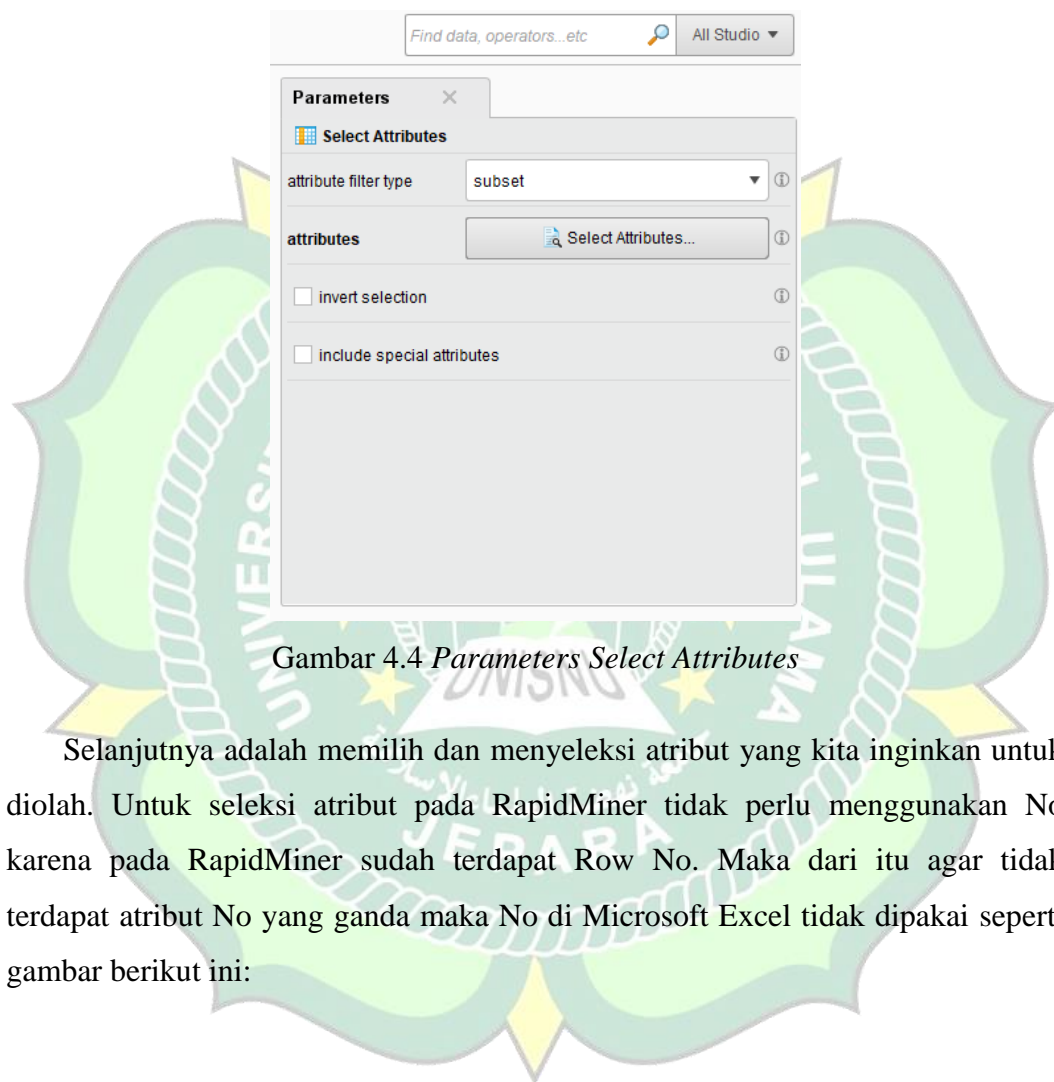
Gambar 4.2 Data RTLH dan Non RTLH Setelah Seleksi Atribut

Data yang telah melewati tahap seleksi atribut dan pengkategorian data melalui *Software* Microsoft Excel serta penghapusan data *outlier* menggunakan IBM SPSS Statistics selanjutnya baru bisa diaplikasikan dengan *software* RapidMiner. Berikut adalah cara penyeleksian atribut yang dilakukan menggunakan *software* RapidMiner versi 9.8:



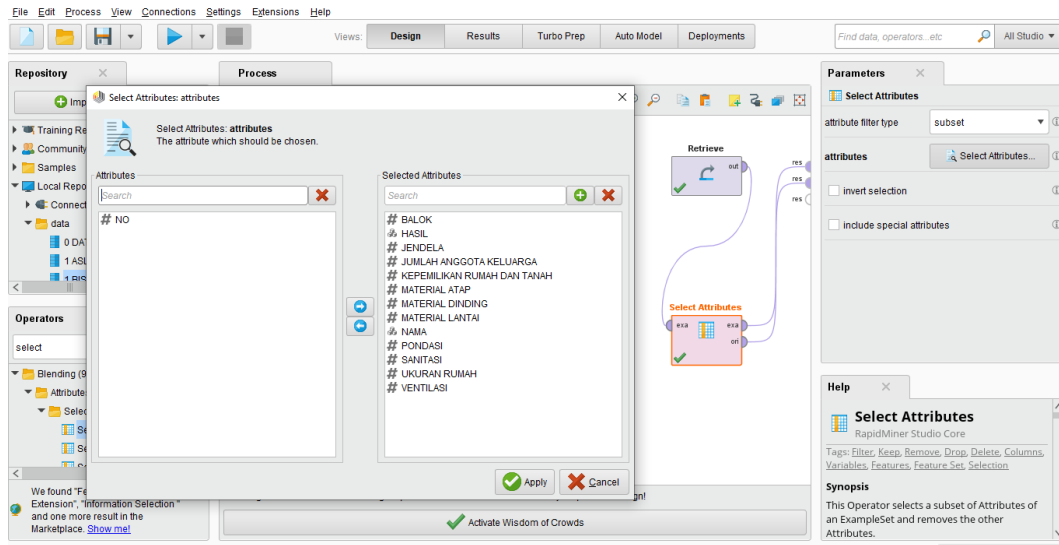
Gambar 4.3 Permodelan Seleksi Atribut

Setelah membuat permodelan seleksi atribut pada *software* RapidMiner versi 9.8, langkah selanjutnya adalah dilakukan pengaturan pada *parameters*. Pada penelitian ini *attribute type filter* yang digunakan adalah *subset*, hal ini dikarenakan dapat memilih dan menyeleksi atribut yang kita inginkan untuk diolah seperti gambar berikut ini:



Gambar 4.4 *Parameters Select Attributes*

Selanjutnya adalah memilih dan menyeleksi atribut yang kita inginkan untuk diolah. Untuk seleksi atribut pada RapidMiner tidak perlu menggunakan No karena pada RapidMiner sudah terdapat Row No. Maka dari itu agar tidak terdapat atribut No yang ganda maka No di Microsoft Excel tidak dipakai seperti gambar berikut ini:



Gambar 4.5 Seleksi Atribut

Pada tahapan seleksi atribut setelah menghilangkan atribut No maka diperoleh hasil seperti gambar berikut ini:

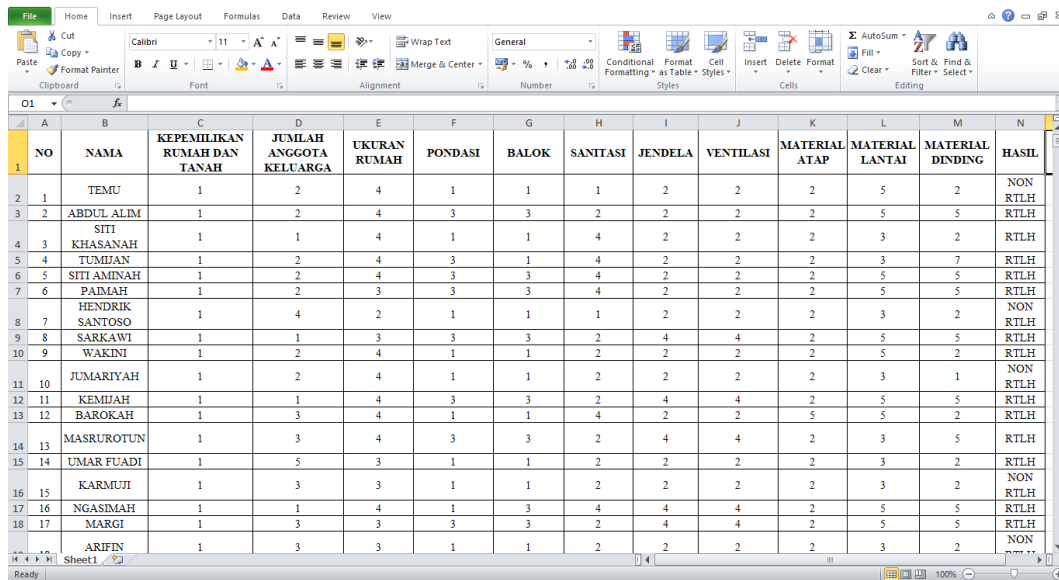
Ro...	NAMA	KEPEMILIKA...	JUMLAH AN...	UKURAN RU...	PONDASI	BALOK	SANITASI	JENDELA	VENTILASI	MATERIAL A...	MATERIAI
1	ABDUL ALIM	1	2	4	3	3	2	2	2	2	5
2	SITI KHASAN...	1	1	4	1	1	4	2	2	2	3
3	TUMUJAN	1	2	4	3	1	4	2	2	2	3
4	HENDRIK SA...	1	4	2	1	1	1	2	2	2	3
5	WAKINI	1	2	4	1	1	2	2	2	2	5
6	JUMARIYAH	1	2	4	1	1	2	2	2	2	3
7	KEMUJAH	1	1	4	3	3	2	2	4	2	5
8	MASRURROT...	1	3	4	3	3	2	2	4	2	3
9	UMAR FUADI	1	5	3	1	1	2	2	2	2	3
10	KARIMLUJI	1	3	3	1	1	2	2	2	2	3
11	NGASIMAH	1	1	4	1	3	4	2	4	2	5
12	MARGI	1	3	3	3	3	2	2	4	2	5
13	NOR LATIF	1	6	3	1	1	2	2	2	2	1
14	WAGINAH	1	1	2	1	3	2	2	2	2	5
15	NGATIMIN	1	3	3	1	1	2	2	2	2	3
16	RUMEK	1	1	4	3	3	2	2	2	2	5

Gambar 4.6 Hasil Seleksi Atribut

#### 4.1.2.2 Menyusun dan Mengkategorikan Data

Pada tahapan ini atribut yang telah diseleksi maka dikonversi berdasarkan kriteria variabel yang sudah ditentukan menggunakan *software* Microsoft Excel. Kriteria variabel dapat dilihat pada table 3.1. Adapun atribut yang dikonversi yaitu kepemilikan rumah dan tanah, ukuran rumah, pondasi, balok, sanitasi, jendela,

ventilasi, material atap, material lantai, dan material dinding. Sedangkan tipe data masing-masing atribut yang nantinya digunakan pada *software* RapidMiner dapat dilihat pada table 4.1, sehingga dari tahapan ini diperoleh hasil sebagai berikut:



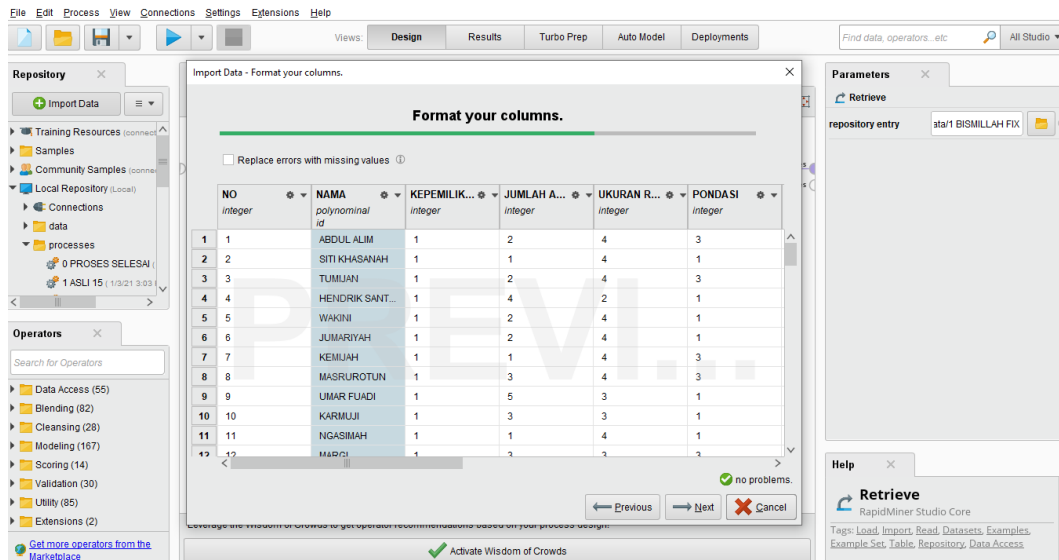
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	NO	NAMA	KEPEMILIKAN RUMAH DAN TANAH	JUMLAH ANGGOTA KELUARGA	UKURAN RUMAH	PONDASI	BALOK	SANITASI	JENDELA	VENTILASI	MATERIAL ATAP	MATERIAL LANTAI	MATERIAL DINDING	HASIL
2	1	TEMU	1	2	4	1	1	1	2	2	2	5	2	NON RTLH
3	2	ABDUL ALIM SITI	1	2	4	3	3	2	2	2	2	5	5	RTLH
4	3	KHASANAH	1	1	4	1	1	4	2	2	2	3	2	RTLH
5	4	TUMIJAN	1	2	4	3	1	4	2	2	2	3	7	RTLH
6	5	SITI AMINAH	1	2	4	3	3	4	2	2	2	5	5	RTLH
7	6	PALMAH	1	2	3	3	3	4	2	2	2	5	5	RTLH
8	7	HENDRIK SANTOSO	1	4	2	1	1	1	2	2	2	3	2	NON RTLH
9	8	SARKAWI	1	1	3	3	3	2	4	4	2	5	5	RTLH
10	9	WAKINI	1	2	4	1	1	2	2	2	2	5	2	RTLH
11	10	JUMARIYAH	1	2	4	1	1	2	2	2	2	3	1	NON RTLH
12	11	KEMIJAH	1	1	4	3	3	2	4	4	2	5	5	RTLH
13	12	BAROKAH	1	3	4	1	1	4	2	2	5	5	2	RTLH
14	13	MASRUROTUN	1	3	4	3	3	2	4	4	2	3	5	RTLH
15	14	UMAR FUADI	1	5	3	1	1	2	2	2	2	3	2	RTLH
16	15	KARMUJI	1	3	3	1	1	2	2	2	2	3	2	NON RTLH
17	16	NGASIMAH	1	1	4	1	3	4	4	4	2	5	5	RTLH
18	17	MARGI	1	3	3	3	3	2	4	4	2	5	5	RTLH
19	18	ARIFIN	1	3	3	1	1	2	2	2	2	3	2	NON RTLH

Gambar 4.7 Hasil Penyusunan dan Pengkategorian Data

Tabel 4.1 Tipe Data

No	Atribut	Tipe Data	Role
1	Nama	Polynomial	ID
2	Kepemilikan Rumah dan Tanah	Integer	
3	Jumlah Anggota Keluarga	Integer	
4	Ukuran Rumah	Integer	
5	Pondasi	Integer	
6	Balok	Integer	
7	Sanitasi	Integer	
8	Jendela	Integer	
9	Ventilasi	Integer	
10	Material Atap	Integer	
11	Material Lantai	Integer	
12	Material Dinding	Integer	
13	Hasil	Polynomial	Label

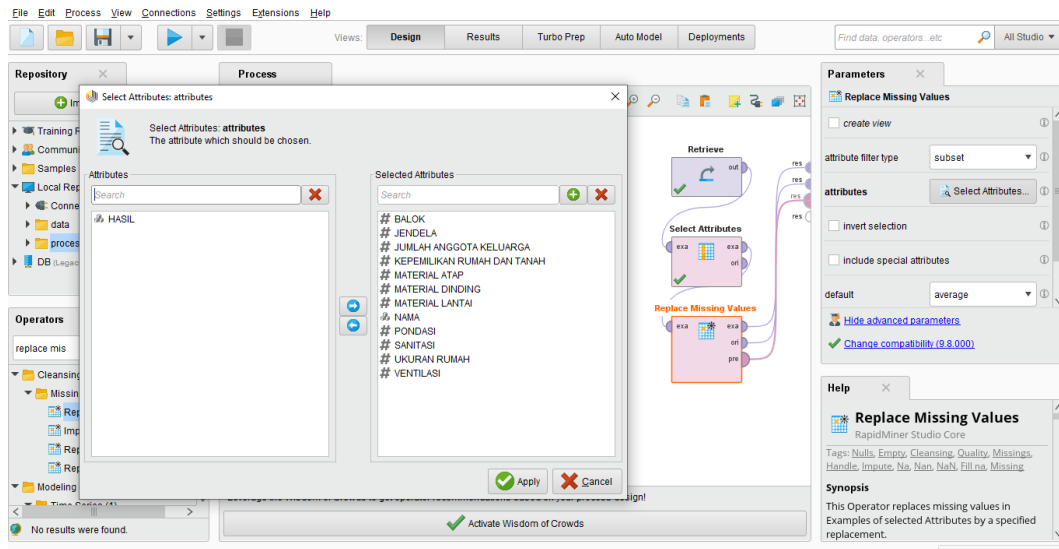
Gambar dibawah ini merupakan pengaplikasian tipe data kedalam format kolom yang ada di RapidMiner.



Gambar 4.8 Mengubah Tipe Data Pada Format Kolom

#### 4.1.2.3 Pembersihan Data

Tahapan ini dilakukan untuk mengidentifikasi dan menghilangkan data yang kosong (*missing value*). Sebelumnya pada tahapan seleksi atribut dan pengkategorian data menggunakan *software* Microsoft Excel serta penghapusan data *outlier* menggunakan *software* IBM SPSS Statistics sudah dilakukan pembersihan data yang kosong. Tetapi untuk memaksimalkan hasil akurasi maka dilakukan pengecekan data lagi untuk mengantisipasi adanya data yang kosong menggunakan operator *replace missing values* Pada *software* RapidMiner. Atribut yang dilakukan pembersihan antara lain nama, kepemilikan rumah dan tanah, ukuran rumah, pondasi, balok, sanitasi, jendela, ventilasi, material atap, material lantai, dan material dinding. Sedangkan atribut hasil tidak dilakukan pembersihan dikarenakan terdapat *missing labels* sebagai data *testing* untuk pengujian. Tentunya data yang akan diaplikasikan pada RapidMiner merupakan data yang telah dilakukan *preprocessing* data pada *software* Microsoft Excel terlebih dahulu. Tahapan pembersihan data pada RapidMiner dapat dilihat pada gambar berikut ini:



Gambar 4.9 Permodelan *Replace Missing Values*

Seperti halnya pada permodelan seleksi atribut, pada tahapan ini *attribute type filter* yang digunakan adalah subset, sehingga diperoleh hasil sebagai berikut:

Row No.	NAMA	HASIL	KEPEMILIK...	JUMLAH AN...	UKURAN RU...	PONDASI	BALOK	SANITASI	JENI
1	ABDUL ALIM	RTLH	1	2	4	3	3	2	2
2	SITI KHASAN...	RTLH	1	1	4	1	1	4	2
3	TUMIJAN	RTLH	1	2	4	3	1	4	2
4	HENDRIK SA...	NON RTLH	1	4	2	1	1	1	2
5	WAKINI	RTLH	1	2	4	1	1	2	2
6	JUMARIYAH	NON RTLH	1	2	4	1	1	2	2
7	KEMIJAH	RTLH	1	1	4	3	3	2	2
8	MASRUROT...	RTLH	1	3	4	3	3	2	2
9	UMAR FUADI	RTLH	1	5	3	1	1	2	2
10	KARMUJI	NON RTLH	1	3	3	1	1	2	2
11	NGASIMAH	RTLH	1	1	4	1	3	4	2
12	MARGI	RTLH	1	3	3	3	3	2	2
13	NOR LATIF	NON RTLH	1	6	3	1	1	2	2
14	WAGINAH	RTLH	1	1	2	1	3	2	2

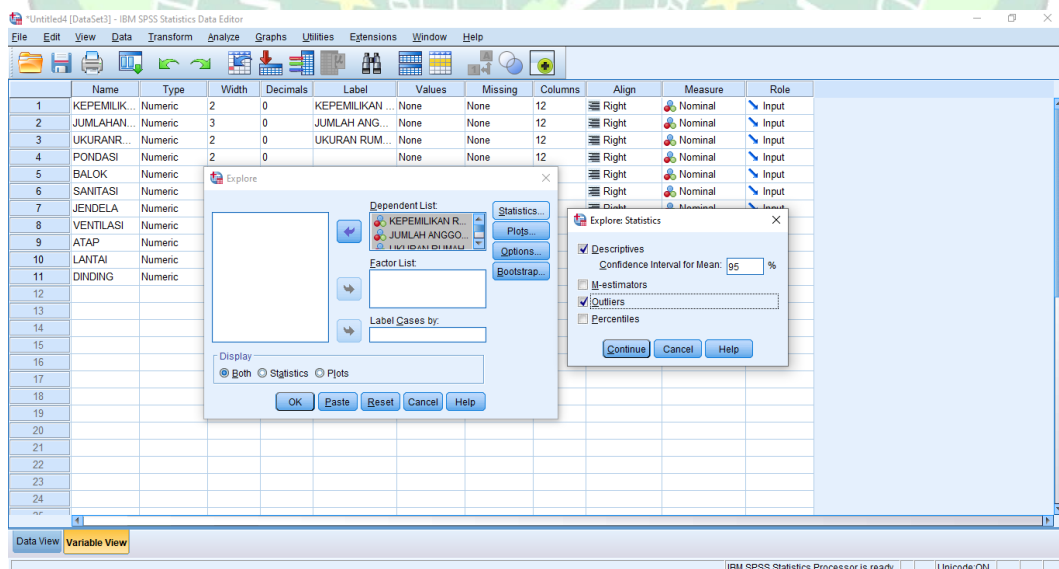
Gambar 4.10 Hasil Pembersihan Data

Pada tahap ini juga dilakukan pembersihan data *outlier* yang menggunakan *software* IBM SPSS Statistics versi 25 dimana untuk mengidentifikasi data yang memiliki selisih jarak sangat berbeda jauh dari sekelompok data lainnya. Sebelumnya harus mengimport data terlebih dahulu, untuk data yang digunakan



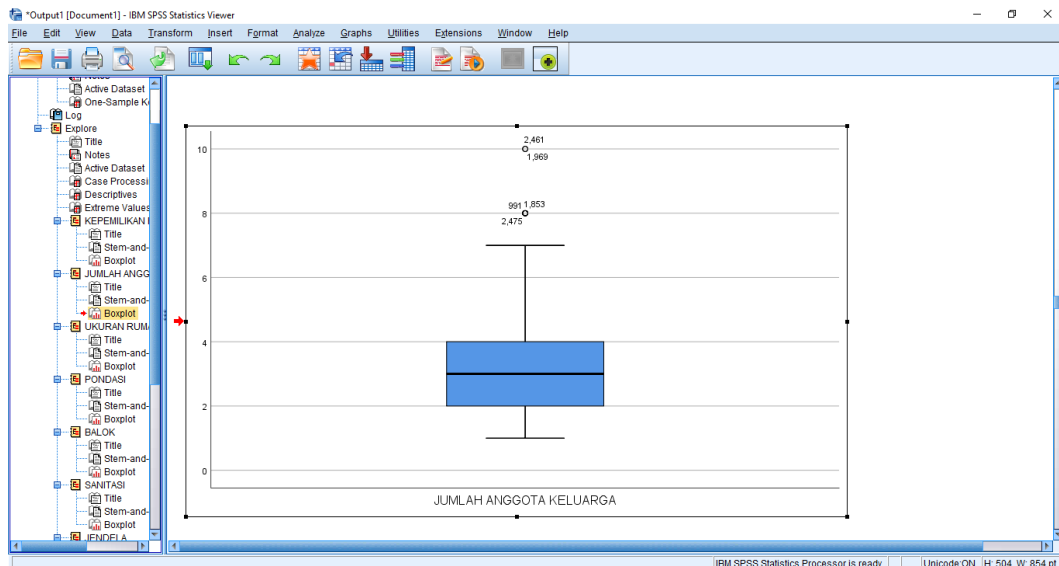
dalam bentuk file Microsoft Excel. Setelah itu untuk uji *outlier* peneliti menggunakan operator *Analyze* kemudian pilih *Descriptive Statistics* dan *Explore*. Pada *Explore* terdapat beberapa atribut lalu semua atribut dipindahkan ke *Dependent List*. Kemudian pilih operator *Statistics* dan pilih *outlier*. Untuk melihat data *outlier* pilih *Boxplot* dari masing-masing atribut. Pada SPSS ditampilkan 2 jenis data *outlier*, pertama data yang diberikan symbol bintang (\*) yang memiliki maksud data tersebut memiliki selisih yang jauh dan harus dihilangkan, kedua data yang diberikan symbol bulat (o) yang memiliki maksud data tersebut masih ada toleransinya tetapi baiknya tetap dihapus. Setelah diketahui data keberapa yang termasuk *outlier* maka tinggal dihapus pada data di Microsoft Excel. Karena dengan menghapus data *outlier* akan mempengaruhi nilai akurasi pada *software* RapidMiner.

Setelah dilakukan pembersihan data *outlier* maka jumlah data *training* menjadi 1.287 data dari semula 2.932 data. Untuk data testing tetap sebanyak 30 data. Untuk tahapan mendeteksi adanya data *outlier* pada data training RTLH dan Non RTLH 2019 sebagai berikut.



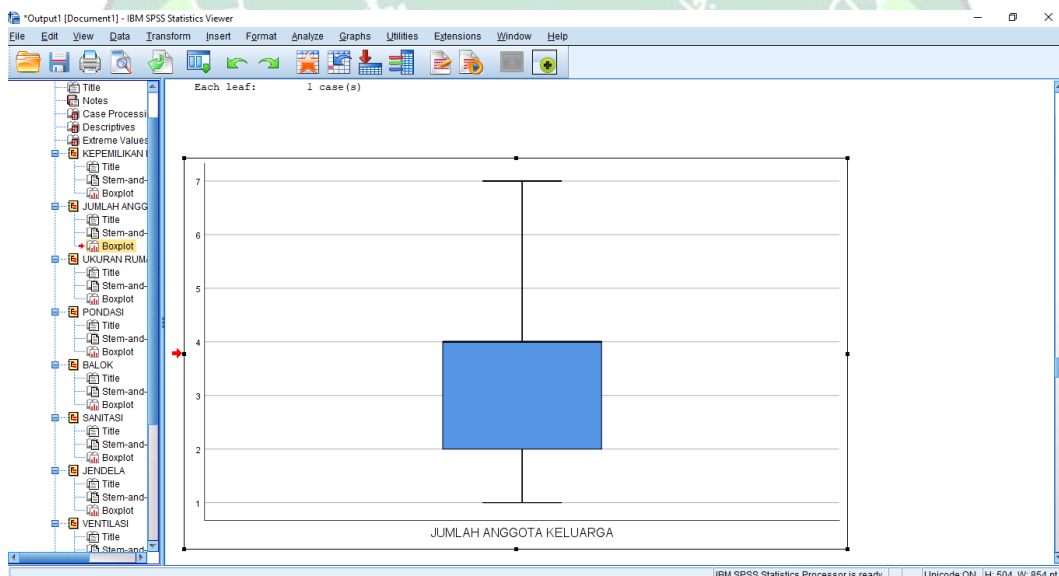
Gambar 4.11 Mengidentifikasi Data *Outlier*

Setelah dilakukan pendeteksian *outlier* pada data *training* RTLH dan Non RTLH 2019 maka dihasilkan data *outlier* yang diberikan simbol bulat (o) sebagai berikut.



Gambar 4.12 Permodelan *Boxplot*

Gambar dibawah ini merupakan hasil pembersihan data *outlier*. Jika sudah tidak ada data yang diberi tanda bintang (\*) dan bulat (o) maka atribut tersebut sudah bersih dari data *outlier*.



Gambar 4.13 Hasil Pembersihan Data *Outlier*

#### 4.1.2.4 Normalisasi Data

Normalisasi data dilakukan pada atribut kepemilikan rumah dan tanah, jumlah anggota keluarga, ukuran rumah, pondasi, balok, sanitasi, jendela, ventilasi, material atap, material lantai dan material dinding. Pada tahapan normalisasi data ini juga dilakukan pada Microsoft Excel guna menyocokkan hasil normalisasi ketika di RapidMiner. Normalisasi data menggunakan *min-max normalization* yang ditunjukkan pada rumus berikut ini:

$$V' = \frac{V - \text{min}A}{\text{max}A - \text{min}A} (\text{new max}A - \text{new min}A) + \text{new min}A \dots \dots \dots (6)$$

Dimana:

$V'$  = Nilai dari data baru hasil dari normalisasi

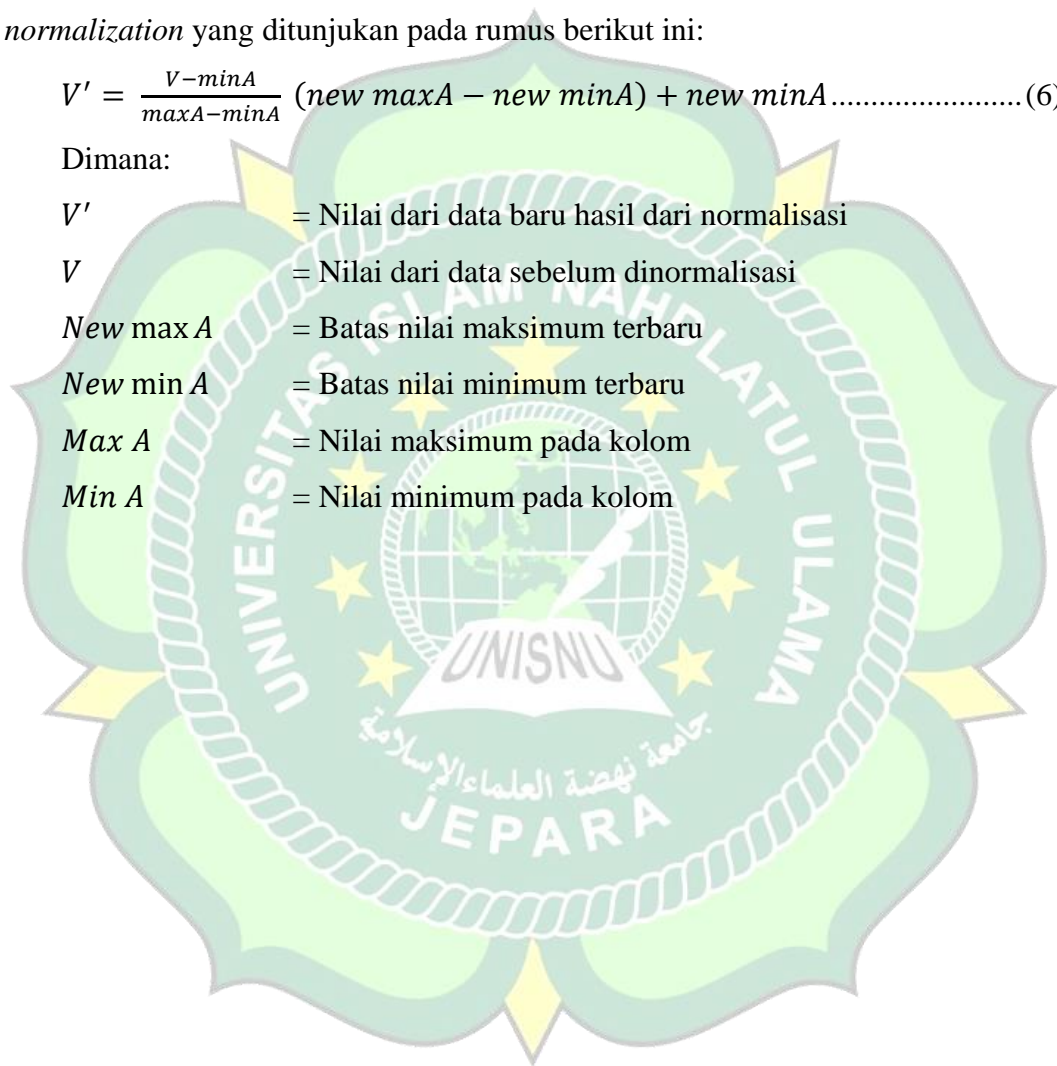
$V$  = Nilai dari data sebelum dinormalisasi

$\text{New max} A$  = Batas nilai maksimum terbaru

$\text{New min} A$  = Batas nilai minimum terbaru

$\text{Max} A$  = Nilai maksimum pada kolom

$\text{Min} A$  = Nilai minimum pada kolom



Tabel 4.2 Data *Training* dan Data *Testing* RTLH dan Non RTLH 2019

NO	NAMA	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	HASIL
1	ABDUL ALIM	1	2	4	3	3	2	2	2	2	5	5	RTLH
2	SITI KHASANAH	1	1	4	1	1	4	2	2	2	3	2	RTLH
3	TUMIJAN	1	2	4	3	1	4	2	2	2	3	7	RTLH
4	HENDRIK SANTOSO	1	4	2	1	1	1	2	2	2	3	2	NON RTLH
5	WAKINI	1	2	4	1	1	2	2	2	2	5	2	RTLH
6	JUMARIYAH	1	2	4	1	1	2	2	2	2	3	1	NON RTLH
7	KEMIYAH	1	1	4	3	3	2	2	4	2	5	5	RTLH
8	MASRUROT UN	1	3	4	3	3	2	2	4	2	3	5	RTLH
9	UMAR FUADI	1	5	3	1	1	2	2	2	2	3	2	RTLH
10	KARMUJI	1	3	3	1	1	2	2	2	2	3	2	NON RTLH
11	NGASIMAH	1	1	4	1	3	4	2	4	2	5	5	RTLH
12	MARGI	1	3	3	3	3	2	2	4	2	5	5	RTLH
13	NOR LATIF	1	6	3	1	1	2	2	2	2	1	1	NON RTLH
14	WAGINAH	1	1	2	1	3	2	2	2	2	5	5	RTLH
15	NGATIMIN	1	3	3	1	1	2	2	2	2	3	2	NON RTLH
16	RUMEK	1	1	4	3	3	2	2	2	2	5	5	RTLH
17	TASIYAH	1	1	4	3	3	2	2	4	2	5	8	RTLH
18	SANI	1	1	4	3	3	4	2	4	2	5	5	RTLH
19	TARMI	1	3	3	1	1	2	2	2	2	3	2	NON RTLH
20	NURUL WIDAYATI	1	2	4	1	1	2	2	2	2	3	2	NON RTLH
:	:	:	:	:	:	:	:	:	:	:	:	:	:
1313	PURWATI	3	4	2	1	3	3	2	2	2	5	5	?
1314	BOWO	5	7	3	2	2	1	1	3	3	3	4	?
1315	JALAL	3	6	4	3	1	2	1	3	4	2	3	?
1316	PRAKAS	2	7	5	3	1	1	2	1	1	3	8	?
1317	ALI MAKSUM	3	5	4	1	1	3	3	2	4	2	1	?

Keterangan:

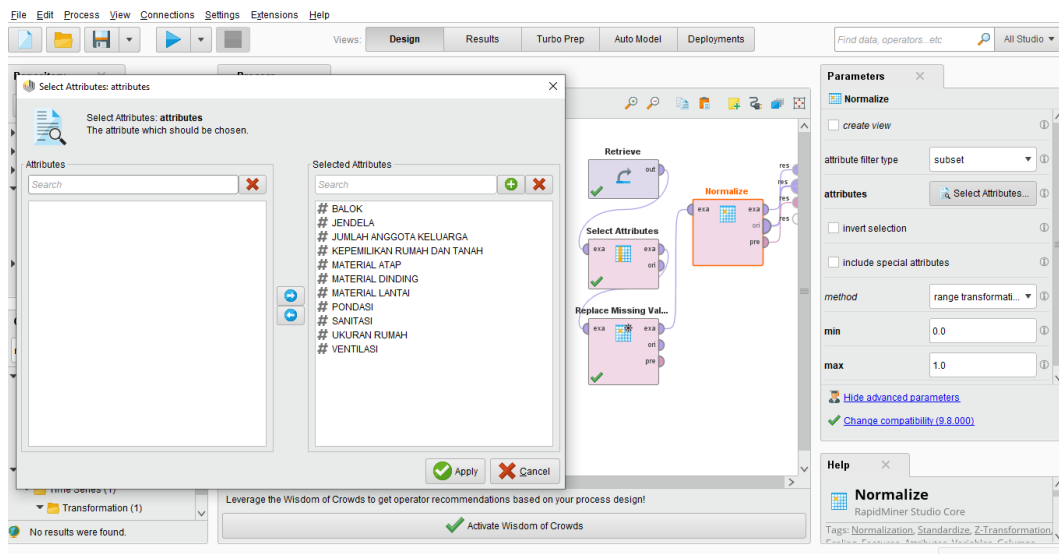
- A1 : Kepemilikan rumah dan tanah
- A2 : Jumlah anggota keluarga
- A3 : Ukuran rumah
- A4 : Pondasi
- A5 : Balok
- A6 : Sanitasi
- A7 : Jendela
- A8 : Ventilasi
- A9 : Material Atap
- A10 : Material Lantai
- A11 : Material Dinding

Proses normalisasi data yang dilakukan dengan cara membuat data yang sudah ada menjadi nilai yang lebih kecil sehingga mengoptimalkan dalam proses perhitungan. Data hasil penelitian yang sudah diolah sebelumnya dinormalisasi dengan cara menjadikan data range 0 sampai 1. Dimana nilai minimal dari data menjadi 0 dan nilai maksimal dari data menjadi 1 serta data diantara nilai minimal dan nilai maksimal menyesuaikan antara range yang digunakan. Berikut normalisasi data menggunakan *software* Microsoft Excel.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	NO	NAMA	KEPEMILIKAN RUMAH & TANAH	JUMLAH ANGGOTA KELUARGA	UKURAN RUMAH	PONDASI	BALOK	SANTIASI	JENDELA	VENTILASI	MATERIA L ATAP	MATERIAL LANTAI	MATERIAL DINDING	HASIL
1326	1	ABDUL ALIM	0	0,125	0,75	0,667	0,667	0,333	0,333	0,333	0,25	1	0,571	RTLH
1327	2	SITI	0	0	0,75	0	0	1	0,333	0,333	0,25	0,5	0,143	RTLH
1328	3	KHASANAH	0	0,125	0,75	0,667	0	1	0,333	0,333	0,25	0,5	0,857	RTLH
1329	4	TUMIJAN	0	0,125	0,75	0	0	0	0,333	0,333	0,25	1	0,143	NON RTLH
1330	5	HENDRIK SANTOSO	0	0,375	0,25	0	0	0	0,333	0,333	0,25	0,5	0,143	RTLH
1331	6	WAKINI	0	0,125	0,75	0	0	0,333	0,333	0,333	0,25	1	0,143	NON RTLH
1332	7	JUMARIYAH	0	0,125	0,75	0	0	0,333	0,333	0,333	0,25	0,5	0	RTLH
1333	8	KEMIJAH	0	0	0,75	0,667	0,667	0,333	0,333	1	0,25	1	0,571	RTLH
1334	9	MASRURUTUN	0	0,25	0,75	0,667	0,667	0,333	0,333	1	0,25	0,5	0,571	RTLH
1335	10	UMAR FUADI	0	0,5	0,5	0	0	0,333	0,333	0,333	0,25	0,5	0,143	RTLH
1336	11	KARMLJI	0	0,25	0,5	0	0	0,333	0,333	0,333	0,25	0,5	0,143	NON RTLH
1337	12	NGASIMAH	0	0	0,75	0	0,667	1	0,333	1	0,25	1	0,571	RTLH
1338	13	MARGI	0	0,25	0,5	0,667	0,667	0,333	0,333	1	0,25	1	0,571	RTLH
1339	14	NOR LATIF	0	0,625	0,5	0	0	0,333	0,333	0,333	0,25	0	0	NON RTLH
1340	15	WAGINAH	0	0	0,25	0	0,667	0,333	0,333	0,333	0,25	1	0,571	RTLH
1341	16	NGATIMIN	0	0,25	0,5	0	0	0,333	0,333	0,333	0,25	0,5	0,143	NON RTLH
1342	17	RUMEK	0	0	0,75	0,667	0,667	0,333	0,333	0,333	0,25	1	0,571	RTLH
1343	18	TASIYAH	0	0	0,75	0,667	0,667	0,333	0,333	0,333	0,25	1	1	RTLH

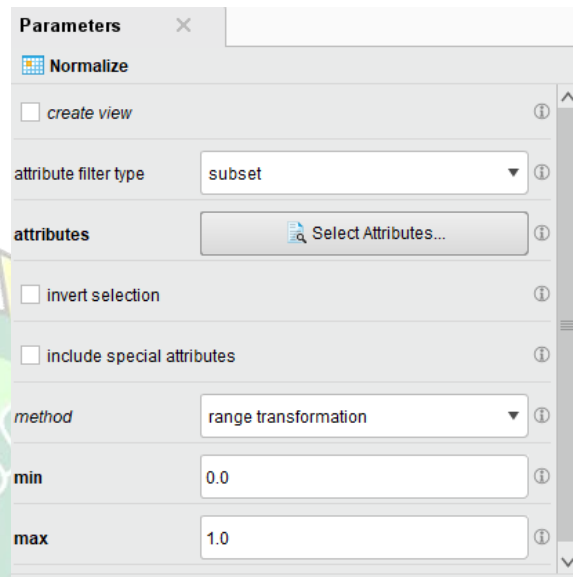
Gambar 4.14 Hasil Normalisasi Data Dengan Ms. Excel

Normalisasi data juga dilakukan pada *software* RapidMiner versi 9.8 menggunakan operator *normalize*. Sedangkan pada parameters *attribute filter type* pilih *subset*. Jika diaplikasikan pada RapidMiner versi 9.8, maka *method* yang digunakan adalah *range transformation*. Permodelan normalisasi data dapat dilihat pada gambar berikut ini:



Gambar 4.15 Permodelan Normalisasi Data

Dari permodelan normalisasi data diatas dapat dilihat batas nilai minimum dan maksimum. Pada tahapan ini peneliti menentukan batas nilai minimum 0, sedangkan batas nilai maksimumnya adalah 1, yang dapat dilihat pada gambar berikut ini:



Gambar 4.16 Batas Nilai Minimum dan Nilai Maksimum

Berikut adalah perhitungan normalisasi data dengan ketentuan sebagai berikut:

Batas nilai minimum : 0

Batas nilai maksimum: 1

Sedangkan nilai minimum dan nilai maksimum pada masing-masing atribut dapat dilihat pada tabel berikut ini:

Tabel 4.3 Nilai Minimum dan Nilai Maksimum

No	Atribut	Nilai Minimum	Nilai Maksimum
1	Kepemilikan Rumah dan Tanah	1	5
2	Jumlah Anggota Keluarga	1	9
3	Ukuran Rumah	1	5
4	Pondasi	1	4
5	Balok	1	4
6	Sanitasi	1	4
7	Jendela	1	4
8	Ventilasi	1	4
9	Atap	1	5
10	Lantai	1	5
11	Dinding	1	8

Perhitungan baris pertama adalah sebagai berikut:

**Baris 1 kolom 1**

$$\begin{aligned}
 V' &= \frac{V - \text{min kolom 1}}{\text{max kolom 1} - \text{min kolom 1}} (\text{new max}A - \text{new min}A) + \text{new min}A \\
 &= \frac{1 - 1}{5 - 1} (1 - 0) + 0 \\
 &= 0
 \end{aligned}$$

**Baris 1 kolom 2**

$$\begin{aligned}
 V' &= \frac{V - \text{min kolom 2}}{\text{max kolom 2} - \text{min kolom 2}} (\text{new max}A - \text{new min}A) + \text{new min}A \\
 &= \frac{2 - 1}{9 - 1} (1 - 0) + 0 \\
 &= 0,125
 \end{aligned}$$

**Baris 1 kolom 3**

$$\begin{aligned}
 V' &= \frac{V - \text{min kolom 3}}{\text{max kolom 3} - \text{min kolom 3}} (\text{new max}A - \text{new min}A) + \text{new min}A \\
 &= \frac{4 - 1}{5 - 1} (1 - 0) + 0 \\
 &= 0,75
 \end{aligned}$$



**Baris 1 kolom 4**

$$\begin{aligned}
 V' &= \frac{V - \text{min kolom 4}}{\text{max kolom 4} - \text{min kolom 4}} (\text{new maxA} - \text{new minA}) + \text{new minA} \\
 &= \frac{3 - 1}{4 - 1} (1 - 0) + 0 \\
 &= 0,667
 \end{aligned}$$

**Baris 1 kolom 5**

$$\begin{aligned}
 V' &= \frac{V - \text{min kolom 5}}{\text{max kolom 5} - \text{min kolom 5}} (\text{new maxA} - \text{new minA}) + \text{new minA} \\
 &= \frac{3 - 1}{4 - 1} (1 - 0) + 0 \\
 &= 0,667
 \end{aligned}$$

**Baris 1 kolom 6**

$$\begin{aligned}
 V' &= \frac{V - \text{min kolom 6}}{\text{max kolom 6} - \text{min kolom 6}} (\text{new maxA} - \text{new minA}) + \text{new minA} \\
 &= \frac{2 - 1}{4 - 1} (1 - 0) + 0 \\
 &= 0,333
 \end{aligned}$$

**Baris 1 kolom 7**

$$\begin{aligned}
 V' &= \frac{V - \text{min kolom 7}}{\text{max kolom 7} - \text{min kolom 7}} (\text{new maxA} - \text{new minA}) + \text{new minA} \\
 &= \frac{2 - 1}{4 - 1} (1 - 0) + 0 \\
 &= 0,333
 \end{aligned}$$

**Baris 1 kolom 8**

$$\begin{aligned}
 V' &= \frac{V - \text{min kolom 8}}{\text{max kolom 8} - \text{min kolom 8}} (\text{new maxA} - \text{new minA}) + \text{new minA} \\
 &= \frac{2 - 1}{4 - 1} (1 - 0) + 0 \\
 &= 0,333
 \end{aligned}$$

**Baris 1 kolom 9**

$$\begin{aligned}
 V' &= \frac{V - \min \text{kolom } 9}{\max \text{kolom } 9 - \min \text{kolom } 9} (\text{new max}A - \text{new min}A) + \text{new min}A \\
 &= \frac{2 - 1}{5 - 1} (1 - 0) + 0 \\
 &= 0,25
 \end{aligned}$$

**Baris 1 kolom 10**

$$\begin{aligned}
 V' &= \frac{V - \min \text{kolom } 10}{\max \text{kolom } 10 - \min \text{kolom } 10} (\text{new max}A - \text{new min}A) + \text{new min}A \\
 &= \frac{5 - 1}{5 - 1} (1 - 0) + 0 \\
 &= 1
 \end{aligned}$$

**Baris 1 kolom 11**

$$\begin{aligned}
 V' &= \frac{V - \min \text{kolom } 11}{\max \text{kolom } 11 - \min \text{kolom } 11} (\text{new max}A - \text{new min}A) + \text{new min}A \\
 &= \frac{5 - 1}{8 - 1} (1 - 0) + 0 \\
 &= 0,571
 \end{aligned}$$

Perhitungan normalisasi diatas juga berlaku untuk baris kedua dan seterusnya, sehingga diperoleh hasil sebagai berikut:

R...	NAMA	KEPE...	JUMLAH...	UKURAN ...	PONDASI	BALOK	SANITASI	JENDELA	VENTILASI	MATERIAL A...	MATERIAL L...	MATERIAL E
1	ABDUL ALIM	0	0.125	0.750	0.667	0.667	0.333	0.333	0.333	0.250	1	0.571
2	SITI KHAS...	0	0	0.750	0	0	1	0.333	0.333	0.250	0.500	0.143
3	TUMIJAN	0	0.125	0.750	0.667	0	1	0.333	0.333	0.250	0.500	0.857
4	HENDRIK ...	0	0.375	0.250	0	0	0	0.333	0.333	0.250	0.500	0.143
5	WAKINI	0	0.125	0.750	0	0	0.333	0.333	0.333	0.250	1	0.143
6	JUMARIYAH	0	0.125	0.750	0	0	0.333	0.333	0.333	0.250	0.500	0
7	KEMIJAH	0	0	0.750	0.667	0.667	0.333	0.333	1	0.250	1	0.571
8	MASRURO...	0	0.250	0.750	0.667	0.667	0.333	0.333	1	0.250	0.500	0.571
9	UMAR FUA...	0	0.500	0.500	0	0	0.333	0.333	0.333	0.250	0.500	0.143
10	KARIMJUI	0	0.250	0.500	0	0	0.333	0.333	0.333	0.250	0.500	0.143
11	NGASIMAH	0	0	0.750	0	0.667	1	0.333	1	0.250	1	0.571
12	MARGI	0	0.250	0.500	0.667	0.667	0.333	0.333	1	0.250	1	0.571
13	NOR LATIF	0	0.625	0.500	0	0	0.333	0.333	0.333	0.250	0	0
14	WAGINAH	0	0	0.250	0	0.667	0.333	0.333	0.333	0.250	1	0.571
15	NGATIMIN	0	0.250	0.500	0	0	0.333	0.333	0.333	0.250	0.500	0.143
16	RUMEK	0	0	0.750	0.667	0.667	0.333	0.333	0.333	0.250	1	0.571

Gambar 4.17 Hasil Normalisasi Data Dengan RapidMiner

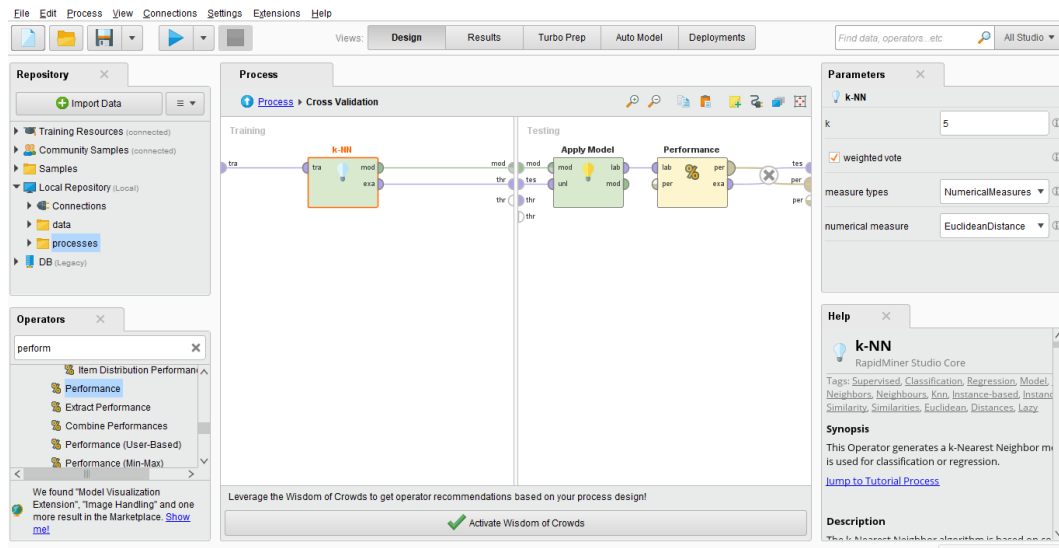
### 4.1.3 Pengujian Model *K-fold cross validation*

Pada tahapan ini dataset akan terbagi secara otomatis pada RapidMiner menjadi 10 bagian dan dilakukan percobaan sebanyak 10 kali atau biasa disebut *10-fold cross validation*. Didalam tahap ini data yang digunakan adalah data *training* dan data *testing* yang tentunya sudah dilakukan *preprocessing* data sebelumnya menggunakan *software* Microsoft Excel. Permodelan *10-fold cross validation* dapat dilihat pada gambar berikut ini:



Gambar 4.18 Permodelan *K-Fold Cross Validation*

Pada operator *Cross Validation* kemudian didouble klik untuk dilakukan permodelan data *training* dan data *testing* menggunakan algoritma *K-Nearest Neighbor* dengan perhitungan jarak menggunakan *Euclidean Distance*.



Gambar 4.19 Permodelan Data Training dan Data Testing

Dalam penelitian ini akan dilakukan pengujian model menggunakan *10-fold cross validation* sebanyak 10 kali dengan nilai  $k$  pada metode *K-Nearest Neighbor* yang berbeda-beda untuk melihat perbandingan tingkat akurasi. Berikutnya adalah hasil pengujian model yang dilakukan:

1. Pengujian pertama dilakukan menggunakan nilai  $k = 1$ , maka memperoleh nilai akurasi 94,87% seperti gambar dibawah ini:

The screenshot shows the 'Performance' panel of RapidMiner Studio. The 'Table View' is selected, displaying the following performance metrics:

accuracy: 94.87% +/- 1.94% (micro average: 94.87%)			
	true RTLH	true NON RTLH	class precision
pred. RTLH	546	44	92.54%
pred. NON RTLH	22	675	96.84%
class recall	96.13%	93.88%	

Gambar 4.20 Hasil Pengujian Model Pertama

2. Pengujian kedua dilakukan menggunakan nilai  $k = 3$ , maka memperoleh nilai akurasi 97,67% seperti gambar dibawah ini:

PerformanceVector (Performance)

accuracy: 97.67% +/- 1.04% (micro average: 97.67%)

	true RTLH	true NON RTLH	class precision
pred. RTLH	548	10	98.21%
pred. NON RTLH	20	709	97.26%
class recall	96.48%	98.61%	

Gambar 4.21 Hasil Pengujian Model Kedua

3. Pengujian ketiga dilakukan menggunakan nilai  $k = 5$ , maka memperoleh nilai akurasi 97,67% seperti gambar dibawah ini:

PerformanceVector (Performance)

accuracy: 97.67% +/- 1.16% (micro average: 97.67%)

	true RTLH	true NON RTLH	class precision
pred. RTLH	546	8	98.56%
pred. NON RTLH	22	711	97.00%
class recall	96.13%	98.89%	

Gambar 4.22 Hasil Pengujian Model Ketiga

4. Pengujian keempat dilakukan menggunakan nilai  $k = 7$ , maka memperoleh nilai akurasi 97,98% seperti gambar dibawah ini:

accuracy: 97.98% +/- 0.98% (micro average: 97.98%)			
	true RTLH	true NON RTLH	class precision
pred. RTLH	548	6	98.92%
pred. NON RTLH	20	713	97.27%
class recall	96.48%	99.17%	

Gambar 4.23 Hasil Pengujian Model Keempat

5. Pengujian kelima dilakukan menggunakan nilai  $k = 9$ , maka memperoleh nilai akurasi 97,75% seperti gambar dibawah ini:

accuracy: 97.75% +/- 1.30% (micro average: 97.75%)			
	true RTLH	true NON RTLH	class precision
pred. RTLH	545	6	98.91%
pred. NON RTLH	23	713	96.88%
class recall	95.95%	99.17%	

Gambar 4.24 Hasil Pengujian Model Kelima

6. Pengujian keenam dilakukan menggunakan nilai  $k = 11$ , maka memperoleh nilai akurasi 97,28% seperti gambar dibawah ini:

PerformanceVector (Performance)

accuracy: 97.28% +/- 1.57% (micro average: 97.28%)

	true RTLH	true NON RTLH	class precision
pred. RTLH	539	6	98.90%
pred. NON RTLH	29	713	96.09%
class recall	94.89%	99.17%	

Gambar 4.25 Hasil Pengujian Model Keenam

7. Pengujian ketujuh dilakukan menggunakan nilai  $k = 13$ , maka memperoleh nilai akurasi 97,28% seperti gambar dibawah ini:

PerformanceVector (Performance)

accuracy: 97.28% +/- 1.57% (micro average: 97.28%)

	true RTLH	true NON RTLH	class precision
pred. RTLH	539	6	98.90%
pred. NON RTLH	29	713	96.09%
class recall	94.89%	99.17%	

Gambar 4.26 Hasil Pengujian Model Ketujuh

8. Pengujian kedelapan dilakukan menggunakan nilai  $k = 15$ , maka memperoleh nilai akurasi 97,13% seperti gambar dibawah ini:

PerformanceVector (Performance)

accuracy: 97.13% +/- 1.10% (micro average: 97.13%)

	true RTLH	true NON RTLH	class precision
pred. RTLH	537	6	98.90%
pred. NON RTLH	31	713	95.83%
class recall	94.54%	99.17%	

Gambar 4.27 Hasil Pengujian Model Kedelapan

9. Pengujian kesembilan dilakukan menggunakan nilai  $k = 17$ , maka memperoleh nilai akurasi 96,66% seperti gambar dibawah ini:

PerformanceVector (Performance)

accuracy: 96.66% +/- 1.47% (micro average: 96.66%)

	true RTLH	true NON RTLH	class precision
pred. RTLH	531	6	98.88%
pred. NON RTLH	37	713	95.07%
class recall	93.49%	99.17%	

Gambar 4.28 Hasil Pengujian Model Kesembilan





#### 4.1.4 Algoritma *K-Nearest Neighbor* (K-NN)

Untuk penerapan algoritma *K-Nearest Neighbor* juga dilakukan pada *software* Microsoft Excel guna untuk memprediksi kelayakan penerima bantuan Rehabilitasi Rumah Tidak Layak Huni (RTLH). Kemudian dilakukan pengujian kembali menggunakan RapidMiner untuk mencocokkan hasil dari prediksi yang telah dihasilkan antara di Microsoft Excel dan di RapidMiner. Data yang digunakan untuk pengujian RapidMiner merupakan data *preprocessing* terakhir yang dilakukan pada Microsoft Excel. Berikut ini adalah dataset penelitian yang akan diimplementasikan terhadap algoritma *K-Nearest Neighbor* (K-NN):

Tabel 4.5 Dataset Penelitian RTLH dan Non RTLH 2019

NO	NAMA	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	HASIL
1	Abdul Alim	0	0,12 5	0,75	0,66 7	0,66 7	0,33 3	0,33 3	0,33 3	0,25	1	0,57 1	RTLH
2	Siti Khasanah	0	0	0,75	0	0	1	0,33 3	0,33 3	0,25	0,5	0,14 3	RTLH
3	Tumijan	0	0,12 5	0,75	0,66 7	0	1	0,33 3	0,33 3	0,25	0,5	0,85 7	RTLH
4	Hendrik Santoso	0	0,37 5	0,25	0	0	0	0,33 3	0,33 3	0,25	0,5	0,14 3	NON RTLH
5	Wakini	0	0,12 5	0,75	0	0	0,33 3	0,33 3	0,33 3	0,25	1	0,14 3	RTLH
:	:	:	:	:	:	:	:	:	:	:	:	:	:
1288	Zumali	0	1	0,75	0,66 7	1	1	0,33 3	1	0,25	0,25	0,57 1	?
1289	Eko Ardianto	1	0,37 5	0,5	0,66 7	0	0,33 3	0,33 3	0,33 3	0	0,5	0	?
1290	Sutawi	0	0,5	0,75	0	0	1	0,33 3	0,33 3	0,5	0	0,14 3	?
1291	Jutono	0,25	0,62 5	0,5	0,33 3	0,33 3	0,33 3	0,33 3	0,33 3	0,75	1	0,28 6	?
1292	Soderi	0	0,75	0,25	0	0,66 7	1	1	0,33 3	1	0	0,42 9	?
1293	Anton Suhartono	0,5	0	0,75	1	0	1	0,33 3	0,33 3	0,75	0,5	0,57 1	?
1294	Sumiati	0	0,25	0,5	0	0	1	0,33 3	0,33 3	0,25	0,5	0,14 3	?
1295	Kiswati	0	0,87 5	0,5	0	0,66 7	0,33 3	1	0,33 3	0,25	0,5	0,14 3	?
1296	Khairul Arif	0	1	0,5	1	0	0,33 3	0,33 3	0,33 3	1	0,5	0,85 7	?

NO	NAMA	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	HASIL
1297	Kasminah	0	0,75	0,25	0	0	0,33 3	1	0,33 3	0,25	1	1	?
1298	Suwarno	0	0,25	0,75	0	1	1	0,33 3	0,33 3	0,75	1	0,14 3	?
1299	Sunardi	0	0	0,75	0,66 7	0	1	0,33 3	0,33 3	0,75	0,5	1	?
1300	Tasmo	0,75	0,12 5	0,5	0	0	1	1	1	0	1	0,42 9	?
1301	Sulawi	0	0,5	0,75	0	0,66 7	0,33 3	1	0,33 3	0,25	0,5	0,14 3	?
1302	Suharti	1	0,75	0,5	0	0	0,33 3	0,33 3	0,33 3	0,25	0,5	0,71 4	?
1303	Munawaroh	0	0,37 5	0,25	0	0	0,33 3	0,33 3	0,33 3	0,25	0,25	0,42 9	?
1304	Masud	0,25	0,12 5	0,5	1	0,33 3	0,66 7	0	0,66 7	1	0,5	0,71 4	?
1305	Mahmudi	0	0,5	1	0,33 3	0	1	0,33 3	0	0,75	0	0,57 1	?
1306	Muyas	0,5	0,62 5	0,75	0,66 7	0	0	0,66 7	0	0,5	0,75	0,85 7	?
1307	Mashar	0,25	0,25	0	0	0,66 7	0,33 3	0	0,33 3	0	1	0,42 9	?
1308	Zumaroh	0,5	0	0,25	0	0,33 3	0,66 7	0,33 3	0,66 7	0,5	0,25	0,28 6	?
1309	Ismah	0,75	0,87 5	0,5	0,33 3	0,66 7	0,66 7	0,66 7	1	0	0,5	0,14 3	?
1310	Lilik	0,5	0,62 5	0,25	0,66 7	0	0,33 3	1	0,66 7	0,75	0,25	0,71 4	?
1311	Khotimah	0	0,25	0,75	1	0,33 3	0,33 3	0,33 3	0,33 3	0	0,75	0,57 1	?
1312	Indah	0	0,12 5	0	0,66 7	1	1	0,66 7	0	0,25	0,5	1	?
1313	Purwati	0,5	0,37 5	0,25	0	0,66 7	0,66 7	0,33 3	0,33 3	0,25	1	0,57 1	?
1314	Bowo	1	0,75	0,5	0,33 3	0,33 3	0	0	0,66 7	0,5	0,5	0,42 9	?
1315	Jalal	0,5	0,62 5	0,75	0,66 7	0	0,33 3	0	0,66 7	0,75	0,25	0,28 6	?
1316	Prakas	0,25	0,75	1	0,66 7	0	0	0,33 3	0	0	0,5	1	?
1317	Ali Maksum	0,5	0,5	0,75	0	0	0,66 7	0,66 7	0,33 3	0,75	0,25	0	?

Keterangan:

- A1 : Kepemilikan rumah dan tanah
- A2 : Jumlah anggota keluarga
- A3 : Ukuran rumah
- A4 : Pondasi
- A5 : Balok
- A6 : Sanitasi
- A7 : Jendela
- A8 : Ventilasi
- A9 : Atap
- A10 : Lantai
- A11 : Dinding

#### 4.1.4.1 Menetapkan Parameter K

Berdasarkan hasil pada tahapan pengujian model *10-fold cross validation*, nilai akurasi tertinggi diperoleh dengan nilai  $k = 7$ . Hal tersebut kemudian diimplementasikan kedalam langkah-langkah algoritma *K-Nearest Neighbor* untuk menentukan hasil penerima bantuan rehabilitasi Rumah Tidak Layak Huni (RTLH) terhadap data uji yang belum diketahui kategori penerimaannya. Setelah dilakukan pengujian menggunakan *software* RapidMiner dan didapatkan hasil akurasi tertinggi pada  $k = 7$  maka pada penentuan  $k$  di Microsoft Excel langsung dibuat menyesuaikan menjadi  $k = 7$ .

#### 4.1.4.2 Menghitung Jarak Euclidean Distance

Setelah menetapkan parameter  $k$  yaitu 7, selanjutnya adalah menghitung jarak masing-masing objek pada data *training* terhadap data *testing* menggunakan rumus *Euclidean Distance*. Perhitungan jarak *Euclidean Distance* dilakukan juga pada *software* Microsoft Excel yang nantinya digunakan untuk menentukan hasil prediksi penerima bantuan rehabilitasi rumah tidak layak huni (RTLH). Berikut adalah perhitungan jarak masing-masing objek yang dicontohkan pada data *testing* pertama:

1. Perhitungan data *testing* pertama terhadap data *training* pertama

$$\begin{aligned}
 d(1,1288) &= \sqrt{\sum_{i=1}^n (P_1 - Q_{1288})^2} \\
 d(1,1288) &= \sqrt{(A1 \text{ baris } 1 - A1 \text{ baris } 1288)^2 + (A2 \text{ baris } 1 - A2 \text{ baris } 1288)^2 + \\
 &\quad (A3 \text{ baris } 1 - A3 \text{ baris } 1288)^2 + (A4 \text{ baris } 1 - A4 \text{ baris } 1288)^2 + \\
 &\quad (A5 \text{ baris } 1 - A5 \text{ baris } 1288)^2 + (A6 \text{ baris } 1 - A6 \text{ baris } 1288)^2 + \\
 &\quad (A7 \text{ baris } 1 - A7 \text{ baris } 1288)^2 + (A8 \text{ baris } 1 - A8 \text{ baris } 1288)^2 + \\
 &\quad (A9 \text{ baris } 1 - A9 \text{ baris } 1288)^2 + (A10 \text{ baris } 1 - A10 \text{ baris } 1288)^2 + \\
 &\quad (A11 \text{ baris } 1 - A11 \text{ baris } 1288)^2} \\
 &= \sqrt{(0 - 0)^2 + (0,125 - 1)^2 + (0,75 - 0,75)^2 + \\
 &\quad (0,667 - 0,667)^2 + (0,667 - 1)^2 + (0,333 - 1)^2 + \\
 &\quad (0,333 - 0,333)^2 + (0,333 - 1)^2 + (0,25 - 0,25)^2 + \\
 &\quad (1 - 0,25)^2 + (0,571 - 0,571)^2} \\
 &= 1,526
 \end{aligned}$$

2. Perhitungan data *testing* pertama terhadap data *training* kedua

$$\begin{aligned}
 d(2,1288) &= \sqrt{\sum_{i=1}^n (P_2 - Q_{1288})^2} \\
 d(2,1288) &= \sqrt{(A1 \text{ baris } 2 - A1 \text{ baris } 1288)^2 + (A2 \text{ baris } 2 - A2 \text{ baris } 1288)^2 + \\
 &\quad (A3 \text{ baris } 2 - A3 \text{ baris } 1288)^2 + (A4 \text{ baris } 2 - A4 \text{ baris } 1288)^2 + \\
 &\quad (A5 \text{ baris } 2 - A5 \text{ baris } 1288)^2 + (A6 \text{ baris } 2 - A6 \text{ baris } 1288)^2 + \\
 &\quad (A7 \text{ baris } 2 - A7 \text{ baris } 1288)^2 + (A8 \text{ baris } 2 - A8 \text{ baris } 1288)^2 + \\
 &\quad (A9 \text{ baris } 2 - A9 \text{ baris } 1288)^2 + (A10 \text{ baris } 2 - A10 \text{ baris } 1288)^2 + \\
 &\quad (A11 \text{ baris } 2 - A11 \text{ baris } 1288)^2} \\
 &= \sqrt{(0 - 0)^2 + (0 - 1)^2 + (0,75 - 0,75)^2 + \\
 &\quad (0 - 0,667)^2 + (0 - 1)^2 + (1 - 1)^2 + \\
 &\quad (0,333 - 0,333)^2 + (0,333 - 1)^2 + (0,25 - 0,25)^2 + \\
 &\quad (0,5 - 0,25)^2 + (0,143 - 0,571)^2} \\
 &= 1,771
 \end{aligned}$$

3. Perhitungan data *testing* pertama terhadap data *training* ketiga

$$\begin{aligned}
 d(3,1288) &= \sqrt{\sum_{i=1}^n (P_3 - Q_{1288})^2} \\
 d(3,1288) &= \sqrt{(A1 \text{ baris } 3 - A1 \text{ baris } 1288)^2 + (A2 \text{ baris } 3 - A2 \text{ baris } 1288)^2 + \\
 &\quad (A3 \text{ baris } 3 - A3 \text{ baris } 1288)^2 + (A4 \text{ baris } 3 - A4 \text{ baris } 1288)^2 + \\
 &\quad (A5 \text{ baris } 3 - A5 \text{ baris } 1288)^2 + (A6 \text{ baris } 3 - A6 \text{ baris } 1288)^2 + \\
 &\quad (A7 \text{ baris } 3 - A7 \text{ baris } 1288)^2 + (A8 \text{ baris } 3 - A8 \text{ baris } 1288)^2 + \\
 &\quad (A9 \text{ baris } 3 - A9 \text{ baris } 1288)^2 + (A10 \text{ baris } 3 - A10 \text{ baris } 1288)^2 + \\
 &\quad (A11 \text{ baris } 3 - A11 \text{ baris } 1288)^2}
 \end{aligned}$$

$$\begin{aligned}
&= \sqrt{(0-0)^2 + (0,125-1)^2 + (0,75-0,75)^2 +} \\
&\quad (0,667-0,667)^2 + (0-1)^2 + (1-1)^2 +} \\
&\quad (0,333-0,333)^2 + (0,333-1)^2 + (0,25-0,25)^2 +} \\
&\quad (0,5-0,25)^2 + (0,857-0,571)^2} \\
&= 1,534
\end{aligned}$$

4. Perhitungan data *testing* pertama terhadap data *training* keempat

$$\begin{aligned}
d(4,1288) &= \sqrt{\sum_{i=1}^n (P_4 - Q_{1288})^2} \\
d(4,1288) &= \sqrt{(A1 \text{ baris } 4 - A1 \text{ baris } 1288)^2 + (A2 \text{ baris } 4 - A2 \text{ baris } 1288)^2 +} \\
&\quad (A3 \text{ baris } 4 - A3 \text{ baris } 1288)^2 + (A4 \text{ baris } 4 - A4 \text{ baris } 1288)^2 +} \\
&\quad (A5 \text{ baris } 4 - A5 \text{ baris } 1288)^2 + (A6 \text{ baris } 4 - A6 \text{ baris } 1288)^2 +} \\
&\quad (A7 \text{ baris } 4 - A7 \text{ baris } 1288)^2 + (A8 \text{ baris } 4 - A8 \text{ baris } 1288)^2 +} \\
&\quad (A9 \text{ baris } 4 - A9 \text{ baris } 1288)^2 + (A10 \text{ baris } 4 - A10 \text{ baris } 1288)^2 +} \\
&\quad (A11 \text{ baris } 4 - A11 \text{ baris } 1288)^2} \\
&= \sqrt{(0-0)^2 + (0,375-1)^2 + (0,25-0,75)^2 +} \\
&\quad (0-0,667)^2 + (0-1)^2 + (0-1)^2 +} \\
&\quad (0,333-0,333)^2 + (0,333-1)^2 + (0,25-0,25)^2 +} \\
&\quad (0,5-0,25)^2 + (0,143-0,571)^2} \\
&= 1,943
\end{aligned}$$

5. Perhitungan data *testing* pertama terhadap data *training* kelima

$$\begin{aligned}
d(5,1288) &= \sqrt{\sum_{i=1}^n (P_5 - Q_{1288})^2} \\
d(5,1288) &= \sqrt{(A1 \text{ baris } 5 - A1 \text{ baris } 1288)^2 + (A2 \text{ baris } 5 - A2 \text{ baris } 1288)^2 +} \\
&\quad (A3 \text{ baris } 5 - A3 \text{ baris } 1288)^2 + (A4 \text{ baris } 5 - A4 \text{ baris } 1288)^2 +} \\
&\quad (A5 \text{ baris } 5 - A5 \text{ baris } 1288)^2 + (A6 \text{ baris } 5 - A6 \text{ baris } 1288)^2 +} \\
&\quad (A7 \text{ baris } 5 - A7 \text{ baris } 1288)^2 + (A8 \text{ baris } 5 - A8 \text{ baris } 1288)^2 +} \\
&\quad (A9 \text{ baris } 5 - A9 \text{ baris } 1288)^2 + (A10 \text{ baris } 5 - A10 \text{ baris } 1288)^2 +} \\
&\quad (A11 \text{ baris } 5 - A11 \text{ baris } 1288)^2} \\
&= \sqrt{(0-0)^2 + (0,125-1)^2 + (0,75-0,75)^2 +} \\
&\quad (0-0,667)^2 + (0-1)^2 + (0,333-1)^2 +} \\
&\quad (0,333-0,333)^2 + (0,333-1)^2 + (0,25-0,25)^2 +} \\
&\quad (1-0,25)^2 + (0,143-0,571)^2} \\
&= 1,961
\end{aligned}$$

6. Perhitungan data *testing* pertama terhadap data *training* keenam

$$d(6,1288) = \sqrt{\sum_{i=1}^n (P_6 - Q_{1288})^2}$$

$$d(6,1288) = \sqrt{\begin{aligned} &(A1 \text{ baris } 6 - A1 \text{ baris } 1288)^2 + (A2 \text{ baris } 6 - A2 \text{ baris } 1288)^2 + \\ &(A3 \text{ baris } 6 - A3 \text{ baris } 1288)^2 + (A4 \text{ baris } 6 - A4 \text{ baris } 1288)^2 + \\ &(A5 \text{ baris } 6 - A5 \text{ baris } 1288)^2 + (A6 \text{ baris } 6 - A6 \text{ baris } 1288)^2 + \\ &(A7 \text{ baris } 6 - A7 \text{ baris } 1288)^2 + (A8 \text{ baris } 6 - A8 \text{ baris } 1288)^2 + \\ &(A9 \text{ baris } 6 - A9 \text{ baris } 1288)^2 + (A10 \text{ baris } 6 - A10 \text{ baris } 1288)^2 + \\ &(A11 \text{ baris } 6 - A11 \text{ baris } 1288)^2 \end{aligned}}$$

$$= \sqrt{\begin{aligned} &(0 - 0)^2 + (0,125 - 1)^2 + (0,75 - 0,75)^2 + \\ &(0 - 0,667)^2 + (0 - 1)^2 + (0,333 - 1)^2 + \\ &(0,333 - 0,333)^2 + (0,333 - 1)^2 + (0,25 - 0,25)^2 + \\ &(0,5 - 0,25)^2 + (0 - 0,571)^2 \end{aligned}}$$

$$= 1,868$$

7. Perhitungan data *testing* pertama terhadap data *training* ketujuh

$$d(7,1288) = \sqrt{\sum_{i=1}^n (P_7 - Q_{1288})^2}$$

$$d(7,1288) = \sqrt{\begin{aligned} &(A1 \text{ baris } 7 - A1 \text{ baris } 1288)^2 + (A2 \text{ baris } 7 - A2 \text{ baris } 1288)^2 + \\ &(A3 \text{ baris } 7 - A3 \text{ baris } 1288)^2 + (A4 \text{ baris } 7 - A4 \text{ baris } 1288)^2 + \\ &(A5 \text{ baris } 7 - A5 \text{ baris } 1288)^2 + (A6 \text{ baris } 7 - A6 \text{ baris } 1288)^2 + \\ &(A7 \text{ baris } 7 - A7 \text{ baris } 1288)^2 + (A8 \text{ baris } 7 - A8 \text{ baris } 1288)^2 + \\ &(A9 \text{ baris } 7 - A9 \text{ baris } 1288)^2 + (A10 \text{ baris } 7 - A10 \text{ baris } 1288)^2 + \\ &(A11 \text{ baris } 7 - A11 \text{ baris } 1288)^2 \end{aligned}}$$

$$= \sqrt{\begin{aligned} &(0 - 0)^2 + (0 - 1)^2 + (0,75 - 0,75)^2 + \\ &(0,667 - 0,667)^2 + (0,667 - 1)^2 + (0,333 - 1)^2 + \\ &(0,333 - 0,333)^2 + (1 - 1)^2 + (0,25 - 0,25)^2 + \\ &(1 - 0,25)^2 + (0,571 - 0,571)^2 \end{aligned}}$$

$$= 1,455$$

8. Perhitungan data *testing* pertama terhadap data *training* kedelapan

$$d(8,1288) = \sqrt{\sum_{i=1}^n (P_8 - Q_{1288})^2}$$

$$d(8,1288) = \sqrt{\begin{aligned} &(A1 \text{ baris } 8 - A1 \text{ baris } 1288)^2 + (A2 \text{ baris } 8 - A2 \text{ baris } 1288)^2 + \\ &(A3 \text{ baris } 8 - A3 \text{ baris } 1288)^2 + (A4 \text{ baris } 8 - A4 \text{ baris } 1288)^2 + \\ &(A5 \text{ baris } 8 - A5 \text{ baris } 1288)^2 + (A6 \text{ baris } 8 - A6 \text{ baris } 1288)^2 + \\ &(A7 \text{ baris } 8 - A7 \text{ baris } 1288)^2 + (A8 \text{ baris } 8 - A8 \text{ baris } 1288)^2 + \\ &(A9 \text{ baris } 8 - A9 \text{ baris } 1288)^2 + (A10 \text{ baris } 8 - A10 \text{ baris } 1288)^2 + \\ &(A11 \text{ baris } 8 - A11 \text{ baris } 1288)^2 \end{aligned}}$$

$$\begin{aligned}
&= \sqrt{(0 - 0)^2 + (0,25 - 1)^2 + (0,75 - 0,75)^2 +} \\
&\quad (0,667 - 0,667)^2 + (0,667 - 1)^2 + (0,333 - 1)^2 + \\
&\quad (0,333 - 0,333)^2 + (1 - 1)^2 + (0,25 - 0,25)^2 + \\
&\quad (0,5 - 0,25)^2 + (0,571 - 0,571)^2} \\
&= 1,087
\end{aligned}$$

9. Perhitungan data *testing* pertama terhadap data *training* kesembilan

$$d(9,1288) = \sqrt{\sum_{i=1}^n (P_9 - Q_{1288})^2}$$

$$\begin{aligned}
d(9,1288) &= \sqrt{(A1 \text{ baris } 9 - A1 \text{ baris } 1288)^2 + (A2 \text{ baris } 9 - A2 \text{ baris } 1288)^2 +} \\
&\quad (A3 \text{ baris } 9 - A3 \text{ baris } 1288)^2 + (A4 \text{ baris } 9 - A4 \text{ baris } 1288)^2 + \\
&\quad (A5 \text{ baris } 9 - A5 \text{ baris } 1288)^2 + (A6 \text{ baris } 9 - A6 \text{ baris } 1288)^2 + \\
&\quad (A7 \text{ baris } 9 - A7 \text{ baris } 1288)^2 + (A8 \text{ baris } 9 - A8 \text{ baris } 1288)^2 + \\
&\quad (A9 \text{ baris } 9 - A9 \text{ baris } 1288)^2 + (A10 \text{ baris } 9 - A10 \text{ baris } 1288)^2 + \\
&\quad (A11 \text{ baris } 9 - A11 \text{ baris } 1288)^2} \\
&= \sqrt{(0 - 0)^2 + (0,5 - 1)^2 + (0,5 - 0,75)^2 +} \\
&\quad (0 - 0,667)^2 + (0 - 1)^2 + (0,333 - 1)^2 + \\
&\quad (0,333 - 0,333)^2 + (0,333 - 1)^2 + (0,25 - 0,25)^2 + \\
&\quad (0,5 - 0,25)^2 + (0,143 - 0,571)^2} \\
&= 1,701
\end{aligned}$$

10. Perhitungan data *testing* pertama terhadap data *training* kesepuluh

$$d(10,1288) = \sqrt{\sum_{i=1}^n (P_{10} - Q_{1288})^2}$$

$$\begin{aligned}
d(10,1288) &= \sqrt{(A1 \text{ baris } 10 - A1 \text{ baris } 1288)^2 + (A2 \text{ baris } 10 - A2 \text{ baris } 1288)^2 +} \\
&\quad (A3 \text{ baris } 10 - A3 \text{ baris } 1288)^2 + (A4 \text{ baris } 10 - A4 \text{ baris } 1288)^2 + \\
&\quad (A5 \text{ baris } 10 - A5 \text{ baris } 1288)^2 + (A6 \text{ baris } 10 - A6 \text{ baris } 1288)^2 + \\
&\quad (A7 \text{ baris } 10 - A7 \text{ baris } 1288)^2 + (A8 \text{ baris } 10 - A8 \text{ baris } 1288)^2 + \\
&\quad (A9 \text{ baris } 10 - A9 \text{ baris } 1288)^2 + (A10 \text{ baris } 10 - A10 \text{ baris } 1288)^2 + \\
&\quad (A11 \text{ baris } 10 - A11 \text{ baris } 1288)^2} \\
&= \sqrt{(0 - 0)^2 + (0,25 - 1)^2 + (0,5 - 0,75)^2 +} \\
&\quad (0 - 0,667)^2 + (0 - 1)^2 + (0,333 - 1)^2 + \\
&\quad (0,333 - 0,333)^2 + (0,333 - 1)^2 + (0,25 - 0,25)^2 + \\
&\quad (0,5 - 0,25)^2 + (0,143 - 0,571)^2} \\
&= 1,790
\end{aligned}$$



Perhitungan diatas juga berlaku terhadap data training selanjutnya sampai dengan ke 1287 dan data testing sampai dengan ke 1317, sehingga diperoleh hasil sebagai berikut:

Tabel 4.6 Hasil Perhitungan *Euclidean Distance*

<b>NO</b>	<b>NAMA</b>	<b>EUCLIDEAN DISTANCE</b>	<b>HASIL</b>
1	ABDUL ALIM	1,526	RTLH
2	SITI KHASANAH	1,771	RTLH
3	TUMIJAN	1,534	RTLH
4	HENDRIK SANTOSO	1,943	NON RTLH
5	WAKINI	1,961	RTLH
6	JUMARIYAH	1,868	NON RTLH
7	KEMIJAH	1,455	RTLH
8	MASRUROTUN	1,087	RTLH
9	UMAR FUADI	1,701	RTLH
10	KARMUJI	1,790	NON RTLH
11	NGASIMAH	1,455	RTLH
12	MARGI	1,320	RTLH
13	NOR LATIF	1,710	NON RTLH
14	WAGINAH	1,805	RTLH
15	NGATIMIN	1,790	NON RTLH
16	RUMEK	1,601	RTLH
17	TASIYAH	1,517	RTLH
18	SANI	1,294	RTLH
19	TARMI	1,790	NON RTLH
20	NURUL WIDAYATI	1,829	NON RTLH
:	:	:	:
1285	NUR HADI	1,402	RTLH
1286	SUROTO	1,790	RTLH
1287	RINA FITRIYAH	0,961	RTLH
1288	ZUMALI	-	?

#### 4.1.4.3 Menentukan Kelas Mayoritas

Berdasarkan hasil pada tahapan 4.4.2 dalam menghitung jarak *Euclidean Distance* diatas, maka dapat ditentukan kelas mayoritas terhadap data testing pertama yang diranking dari urutan terkecil hingga terbesar. Proses penentuan kelas mayoritas pada *software* Micosoft Excel menggunakan rumus *function* tersendiri yang nantinya bisa menentukan berdasarkan  $k = 7$  maka kelas mayoritasnya mengikuti dan didapatkan hasil prediksi secara otomatis. Penentuan kelas mayoritas ditentukan satu per satu berdasarkan sejumlah 30 data *testing* yang telah ditentukan. Adapun hasil penentuan kelas mayoritas terhadap data *testing* pertama sebagai berikut:

Tabel 4.7 Penentuan Kelas Mayoritas

No	Nama	Euclidean Distance	Ranking	Hasil
941	MUNAWAR	0,614	1	RTLH
372	SURIPTO	0,652	2	RTLH
1198	JUPRI	0,673	3	RTLH
979	MAS'ADI	0,697	4	RTLH
316	SUTIK	0,745	5	RTLH
253	SUDARTO	0,792	6	RTLH
939	PARWI	0,792	7	RTLH
1288	ZUMALI	-	-	RTLH

Dari hasil penentuan kelas mayoritas pada data *testing* pertama yang bernama Zumali masuk kedalam kategori RTLH, dimana dicocokkan dengan hasil prediksi yang dilakukan pada Microsoft Excel dengan RapidMiner menunjukkan hasil yang sama sebagai berikut.

R...	NAMA	HASIL	predictio...	KEPE...	JUMLA...	UKUR...	PON...	BALOK	SANIT...	JEND...	VENTL...	MATERIA...	MATERI
124	MARYUN	NON RTLH	NON RTLH	0	0,250	0,500	0	0	0,333	0,333	0,333	0,250	0,500
125	KUSMADI	NON RTLH	NON RTLH	0	0,250	0,500	0	0	0,333	0,333	0,333	0,250	0,500
126	MOH ASNAWI...	NON RTLH	NON RTLH	0	0,125	0,750	0	0	0,333	0,333	0,333	0,250	0,500
127	M DAKHORI...	NON RTLH	NON RTLH	0	0,250	0,250	0,333	0,333	0,333	0,333	0,333	0,250	0,500
128	NUR KHOLIS	RTLH	RTLH	0	0,125	0,500	1	1	0,333	0,333	1	0,250	1
129	HARMANTO	RTLH	RTLH	0	0,250	0,750	0,667	0,667	0,333	0,333	1	0,250	0,500
130	ZUMALI	?	RTLH	0	1	0,750	0,667	1	1	0,333	1	0,250	0,250
131	SUTAWI	?	NON RTLH	0	0,500	0,750	0	0	1	0,333	0,333	0,500	0
132	SUHARTI	?	RTLH	1	0,750	0,500	0	0	0,333	0,333	0,333	0,250	0,500
133	ABDUL ALIM	RTLH	RTLH	0	0,125	0,750	0,667	0,667	0,333	0,333	0,333	0,250	1
134	SUYATI	NON RTLH	NON RTLH	0	0,125	0,750	0	0	0,333	0,333	0,333	0,250	0,500
135	DUL SALAM	NON RTLH	NON RTLH	0	0,125	0,750	0	0	0,333	0,333	0,333	0,250	0,500
136	PRAYIT	RTLH	RTLH	0	0,250	0,250	0,667	0,667	0,333	0,333	1	0,250	1
137	KUMARI	RTLH	RTLH	0	0,500	0,750	0,667	0,667	0,333	0,333	0,333	0,250	1

Gambar 4.30 Hasil Kelas Mayoritas RapidMiner

Dari tabel 4.7 maka dapat disimpulkan bahwa data *testing* pertama, pendaftar bantuan yang bernama Zumali termasuk kedalam kategori pemilik Rumah Tidak Layak Huni (RTLH) berdasarkan hasil klasifikasi dari penerapan algoritma *K-Nearest Neighbor* (*K-NN*). Perhitungan ini juga berlaku pada data testing lainnya, sehingga didapat dari hasil perhitungan prediksi secara otomatis yang dihasilkan menggunakan *software* Microsoft Excel yang telah dilakukan sebagai berikut:

Tabel 4.8 Hasil Klasifikasi Data *Testing*

No	Nama	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	Hasil
1	Zumali	0	1	0,75	0,667	1	1	0,333	1	0,25	0,25	0,571	RTLH
2	Eko Ardianto	1	0,375	0,5	0,667	0	0,333	0,333	0,333	0	0,5	0	RTLH
3	Sutawi	0	0,5	0,75	0	0	1	0,333	0,333	0,5	0	0,143	NON RTLH
4	Jutono	0,25	0,625	0,5	0,333	0,333	0,333	0,333	0,333	0,75	1	0,286	RTLH
5	Soderi	0	0,75	0,25	0	0,667	1	1	0,333	1	0	0,429	RTLH
6	Anton Suhartono	0,5	0	0,75	1	0	1	0,333	0,333	0,75	0,5	0,571	RTLH
7	Sumiati	0	0,25	0,5	0	0	1	0,333	0,333	0,25	0,5	0,143	NON RTLH
8	Kiswati	0	0,875	0,5	0	0,667	0,333	1	0,333	0,25	0,5	0,143	RTLH

No	Nama	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	Hasil
9	Khairul Arif	0	1	0,5	1	0	0,33 3	0,33 3	0,333	1	0,5	0,857	RTLH
10	Kasminah	0	0,75	0,25	0	0	0,33 3	1	0,333	0,25	1	1	RTLH
11	Suwarno	0	0,25	0,75	0	1	1	0,33 3	0,333	0,75	1	0,143	RTLH
12	Sunardi	0	0	0,75	0,66 7	0	1	0,33 3	0,333	0,75	0,5	1	RTLH
13	Tasmo	0,75	0,125	0,5	0	0	1	1	1	0	1	0,429	RTLH
14	Sulawi	0	0,5	0,75	0	0,667	0,33 3	1	0,333	0,25	0,5	0,143	RTLH
15	Suharti	1	0,75	0,5	0	0	0,33 3	0,33 3	0,333	0,25	0,5	0,714	RTLH
16	Munawaroh	0	0,375	0,25	0	0	0,33 3	0,33 3	0,333	0,25	0,25	0,429	NON RTLH
17	Masud	0,25	0,125	0,5	1	0,333	0,66 7	0	0,667	1	0,5	0,714	RTLH
18	Mahmudi	0	0,5	1	0,33 3	0	1	0,33 3	0	0,75	0	0,571	NON RTLH
19	Muyas	0,5	0,625	0,75	0,66 7	0	0	0,66 7	0	0,5	0,75	0,857	RTLH
20	Mashar	0,25	0,25	0	0	0,667	0,33 3	0	0,333	0	1	0,429	RTLH
21	Zumaroh	0,5	0	0,25	0	0,333	0,66 7	0,33 3	0,667	0,5	0,25	0,286	RTLH
22	Ismah	0,75	0,875	0,5	0,33 3	0,667	0,66 7	0,66 7	1	0	0,5	0,143	RTLH
23	Lilik	0,5	0,625	0,25	0,66 7	0	0,33 3	1	0,667	0,75	0,25	0,714	RTLH
24	Khotimah	0	0,25	0,75	1	0,333	0,33 3	0,33 3	0,333	0	0,75	0,571	RTLH
25	Indah	0	0,125	0	0,66 7	1	1	0,66 7	0	0,25	0,5	1	RTLH
26	Purwati	0,5	0,375	0,25	0	0,667	0,66 7	0,33 3	0,333	0,25	1	0,571	RTLH
27	Bowo	1	0,75	0,5	0,33 3	0,333	0	0	0,667	0,5	0,5	0,429	RTLH
28	Jalal	0,5	0,625	0,75	0,66 7	0	0,33 3	0	0,667	0,75	0,25	0,286	RTLH
29	Prakas	0,25	0,75	1	0,66 7	0	0	0,33 3	0	0	0,5	1	RTLH
30	Ali Maksum	0,5	0,5	0,75	0	0	0,66 7	0,66 7	0,333	0,75	0,25	0	NON RTLH

#### 4.1.5 Evaluasi dan Validasi

Pengujian validitas dilakukan menggunakan *confusion matrix* sebagai *performance* yang menghasilkan nilai *accuracy*, *precision*, *recall* dan AUC (*Area Under the Curve*), perhitungan dilakukan menggunakan rumus sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (7)$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (8)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (9)$$

Tabel 4.9 *Performance Vector*

	True RTLH	True Non RTLH	Class <i>Precision</i>
Pred. RTLH	548	6	98,92%
Pred. Non RTLH	20	713	97,27%
Class <i>Recall</i>	96,48%	99,17%	

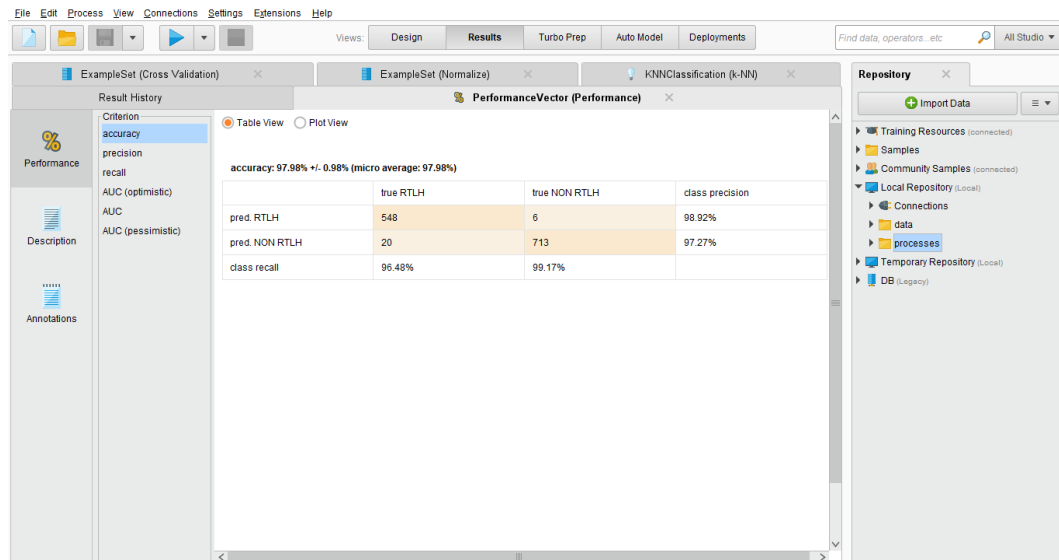
Berdasarkan hasil *performance vector* pada tabel 4.9. Maka dapat disimpulkan data training yang diprediksi dengan benar melalui algoritma *K-Nearest Neighbor (K-NN)* sejumlah 1287, sebanyak 548 orang yang benar RTLH dan 713 orang yang benar Non RTLH. Kemudian sebanyak 20 orang RTLH tetapi diklasifikasikan sebagai Non RTLH dan sebanyak 6 orang Non RTLH tetapi diklasifikasikan RTLH. Sehingga akurasi yang dihasilkan adalah sebagai berikut:

$$Accuracy = \frac{548+713}{548+713+20+6} \times 100\% = 97,98\%$$

$$Precision = \frac{548}{548+20} \times 100\% = 96,17\%$$

$$Recall = \frac{548}{548+6} \times 100\% = 98,91\%$$

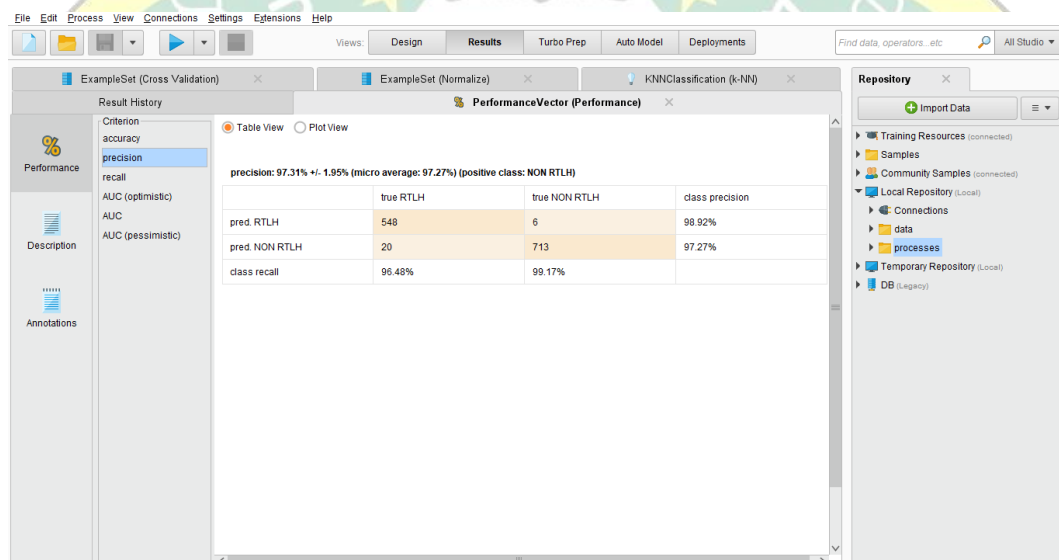
Berdasarkan uji validitas terhadap algoritma *K-Nearest Neighbor* (*K-NN*) diatas didapat nilai prosentase tingkat *accuracy* sebesar 97,98%. Berikut ini adalah hasil *performance* pada RapidMiner versi 9.8.



	true RTLH	true NON RTLH	class precision
pred. RTLH	548	6	98.92%
pred. NON RTLH	20	713	97.27%
class recall	96.48%	99.17%	

Gambar 4.31 Evaluasi dan Validasi Hasil *Accuracy*

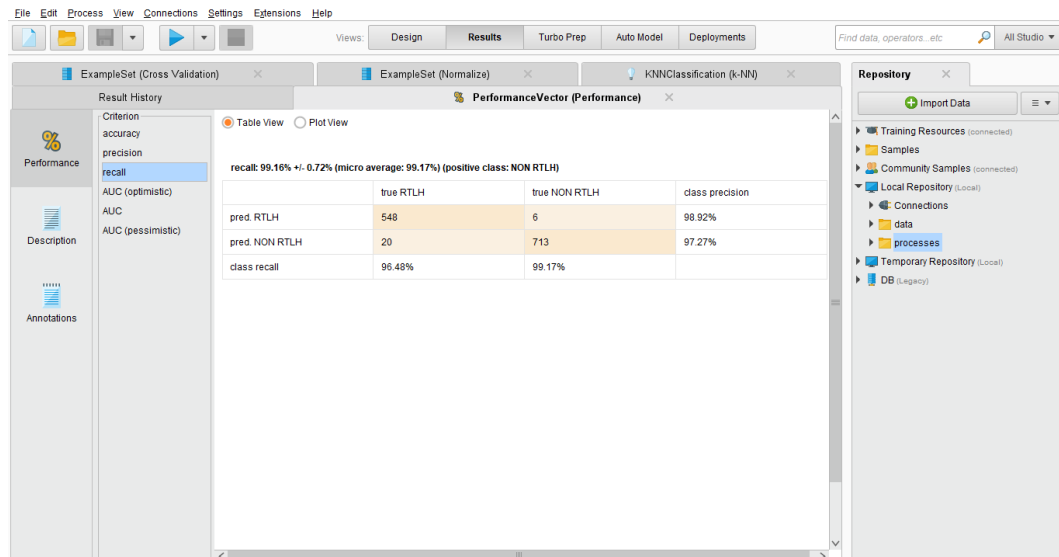
Pada gambar dibawah ini merupakan hasil uji validitas nilai prosentase *precision* sebesar 97,31%.



	true RTLH	true NON RTLH	class precision
pred. RTLH	548	6	98.92%
pred. NON RTLH	20	713	97.27%
class recall	96.48%	99.17%	

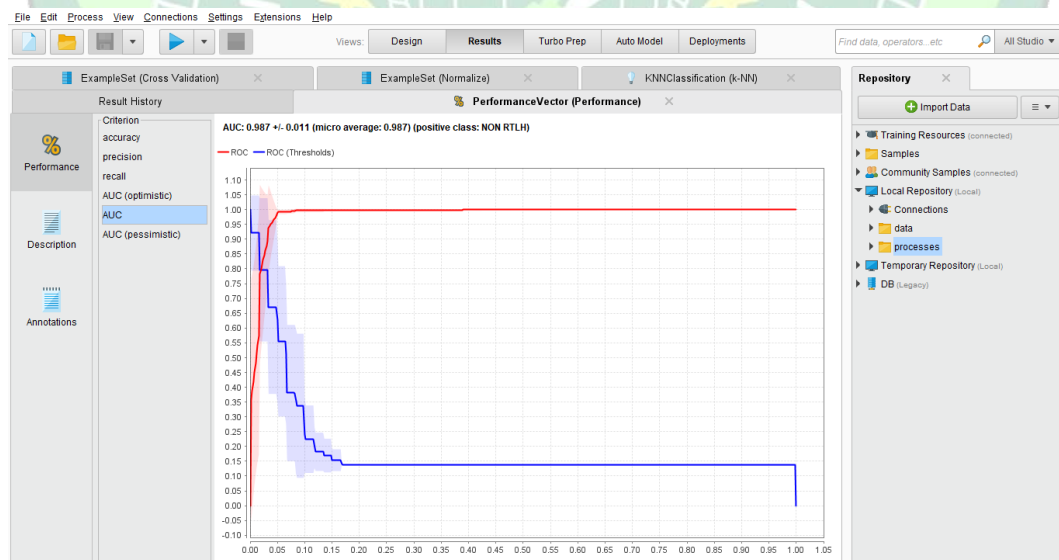
Gambar 4.32 Evaluasi dan Validasi Hasil *Precision*

Pada gambar dibawah ini merupakan hasil uji validitas nilai prosentase *recall* sebesar 99,16%.



Gambar 4.33 Evaluasi dan Validasi Hasil *Recall*

Pada gambar dibawah ini merupakan hasil uji validitas nilai *Area Under The Curve* (AUC) sebesar 0,987.



Gambar 4.34 Evaluasi dan Validasi Hasil *AUC*

## 4.2 Pembahasan

Hasil penelitian saya menerapkan Algoritma *K-Nearest Neighbor* dengan perhitungan jarak menggunakan *Euclidean Distance* serta pembersihan data

*outlier* menggunakan software IBM SPSS Statistic untuk mengoptimalkan kinerja Algoritma *K-Nearest Neighbor* sehingga didapatkan hasil akurasi yang tinggi. Dataset yang digunakan pada penelitian ini adalah data pendaftar dan penerima bantuan rehabilitasi Rumah Tidak Layak Huni (RTLH) tahun 2019 sebanyak 2962 data yang dibagi menjadi 2932 data *training* dan 30 data *testing*, dimana data diperoleh dari Dinas Perumahan Rakyat dan Kawasan Permukiman (DISPERKIM) Jepara. Penelitian ini menggunakan 13 atribut diantaranya nama, kepemilikan rumah dan tanah, ukuran rumah, jumlah anggota keluarga, pondasi, balok, sanitasi, jendela, ventilasi, material atap, material lantai, material dinding dan hasil dengan *output* RTLH atau NON RTLH.

Sebelumnya data diolah terlebih dahulu menggunakan software Microsoft Excel meliputi seleksi atribut, pengkategorian data yang sebelumnya datanya *string* diubah menjadi *numeric* kemudian dilakukan pembersihan data *outlier* dan setelah itu juga menghitung normalisasi data guna untuk menyocokkan hasil yang nantinya dilakukan normalisasi pada RapidMiner. Lalu ketika data sudah melewati beberapa tahapan *preprocessing* data dengan Microsoft Excel maka data baru bisa dilakukan pengujian melalui software RapidMiner. Pada RapidMiner juga dilakukan seleksi atribut karena pada RapidMiner sudah terdapat Row No jadi untuk No yang di Microsoft Excel tidak perlu digunakan, kemudian dilanjutkan *Replace Missing Value* untuk mengecek kembali data apabila ketika dilakukan *preprocessing* data melalui Microsoft Excel masih terdapat data yang kosong, lalu dilanjutkan *Normalize* data untuk membuat data yang sudah ada menjadi nilai yang lebih kecil sehingga mengoptimalkan dalam proses perhitungan. Data hasil penelitian yang sudah diolah sebelumnya dinormalisasi dengan cara menjadikan data range 0 sampai 1. Pada RapidMiner normalisasi data menggunakan *range transformation* dengan nilai nominal 0 dan nilai maksimal 1. Lalu dilakukan pengujian model menggunakan *k-fold cross validation* dan evaluasi dan validasi hasil menggunakan *confusion matrix*. Pengujian dilakukan sebanyak 10 kali atau biasa disebut *10-fold cross validation* pada setiap nilai *k* ganjil. Didapatkan hasil tertinggi ketika  $k = 7$ . Data penelitian yang digunakan berjumlah 2932 data *training* kemudian dilakukan pembersihan data *outlier*



menggunakan IBM SPSS Statistics menjadi 1287 data yang terdiri dari data RTLH dan Non RTLH. Setelah dilakukan *preprocessing* data menggunakan RapidMiner mendapatkan hasil akurasi yang tinggi sebesar 97,98%, nilai *precision* sebesar 97,31%, nilai *recall* sebesar 99,16% dan AUC sebesar 0,987. Sehingga dapat disimpulkan bahwa algoritma *K-Nearest Neighbor* masuk kedalam klasifikasi sangat baik. Dari hasil penelitian yang sudah dilakukan maka didapatkan sistem prediksi kelayakan penerima bantuan program rehabilitasi rumah tidak layak huni (RTLH) secara otomatis dan efisien dengan menggunakan *software* Microsoft Excel.

Dari penelitian sebelumnya yang dilakukan oleh Simatupang Fitri Juniaty dkk dalam penelitiannya yang berjudul *Klasifikasi Rumah Tidak Layak Huni di Kabupaten Brebes dengan Menggunakan Metode Learning Vector Quantization dan Naïve Bayes*. Penelitian dengan menggunakan jenis data yang sama yaitu *numeric* tetapi dengan jumlah atribut yang berbeda dihasilkan akurasi ketika menggunakan *Learning Vector Quantization* sebesar 71,43% dan menggunakan Naïve Bayes mendapatkan akurasi sebesar 95,24% [8]. Jadi dapat disimpulkan bahwa dengan jenis data yang sama didapatkan algoritma yang akurat yaitu algoritma *K-Nearest Neighbor (K-NN)*. Oleh sebab itu kelebihan algoritma *K-Nearest Neighbor* yang telah dijelaskan pada penelitian yang dilakukan Edy Nasri dan A. Selamat AW, diantaranya Algoritma *K-Nearest Neighbor* sangat Non-linear dan mudah dipahami serta diimplementasikan [19]. Setelah dilakukan pengujian langsung terhadap data *training* dan data *testing* berdasarkan kelebihan yang telah disebutkan sebelumnya bahwa benar adanya bahwa algoritma *K-Nearest Neighbor* bisa diimplementasikan dengan mudah, untuk tahapan menghitung jarak hingga didapatkan hasil berdasarkan kelas mayoritas mudah dipahami tanpa membutuhkan waktu pemahaman yang lama. Adapun penambahan kelebihan *K-Nearest Neighbor* setelah dilakukan pengujian terhadap data penelitian yaitu Algoritma *K-Nearest Neighbor* dapat mengenali data *numeric*, data *string* dan data campuran, efektif terhadap jumlah data yang besar, dan tahan terhadap data *testing* yang masih kosong.