

Analisis Algoritma Klasifikasi Neural Network Untuk Diagnosis Penyakit Diabetes

Jajang Jaya Purnama¹, Sri Rahayu², Siti Nurdiani³, Tuti Haryanti⁴, Nissa Almira Mayangky⁵

¹Teknologi Komputer, Univeristas Bina Sarana Informatika
Jakarta, Indonesia
Jajangja2412@bsi.ac.id

^{2,3,4,5}Sistem Informasi, STMIK Nusa Mandiri
Jakarta, Indonesia
sriahayu.rry@nusamandiri.ac.id², siti.sxd@nusamandiri.ac.id³, tuti@nusamandiri.ac.id⁴,
nissa.nky@nusamandiri.ac.id⁵

ABSTRAK

Diabetes merupakan penyakit yang sangat mematikan terbukti dari tahun ke tahun selalu ada yang meninggal dikarenakan pasien tersebut mengidap penyakit diabetes, banyak cara penanggulangan sejak dini penyakit diabetes. Salah satunya dengan data mining klasifikasi algoritma neural network yang dapat digunakan untuk prediksi pasien mana yang terkena penyakit diabetes dan pasien mana yang tidak terkena diabetes dengan menggunakan parameter dan indikator yang ada, dan tools yang digunakan adalah tools rapid miner 9.0 yang menghasilkan accuracy sebesar = 80.00% precision sebesar = 100.00 % dan recall sebesar = 2.50 % dengan AUC sebesar = 0.605 % yang artinya klasifikasi dinyatakan cukup, dari hasil tersebut bisa diambil kesimpulan bahwa penelitian ini bisa mencegah dan bisa diketahui sejak dini mana yang termasuk penyakit diabetes mana yang tidak mengidap penyakit diabetes, dan dari penelitian ini sangat diharapkan angka kematian bisa berkurang.

Katakunci: diabetes, klasifikasi, data mining, neural network.

ABSTRACTS

Diabetes is a very proven disease from year to year there are always people who die, many ways to postpone early diabetes. One of them is data mining neural network algorithm classification which can be used to predict which patients are affected by diabetes and which patients are not affected by diabetes by using existing parameters and indicators, and the tools used are rapid miner 9.0 tools that produce accuracy = 80.00% precision = 100.00% and recall of = 2.50% with AUC of = 0.605% which means the classification is declared sufficient, From these results it can be concluded that this study can prevent and can be known from the outset which of the diabetics do not have diabetes, and from this study it is expected that the mortality rate can be reduced.

Keywords: classification, data mining, diabetes, neural network.



1. PENDAHULUAN

Diabetes mellitus adalah gangguan metabolisme yang secara genetik dan klinis termasuk heterogen dengan manifestasi berupa hilangnya toleransi karbohidrat, jika telah berkembang penuh secara klinis maka diabetes mellitus ditandai dengan hiperglikemia puasa postprandial, aterosklerosis dan penyakit vaskular mikroangiopati. (Fatimah, 2015).

Diabetes mellitus (DM) merupakan salah satu masalah kesehatan yang besar. Data dari studi global menunjukkan bahwa jumlah penderita Diabetes Mellitus pada tahun 2011 telah mencapai 366 juta orang. Jika tidak ada tindakan yang dilakukam, jumlah ini diperkirakan akan meningkat menjadi 552 juta pada tahun 2030. Diabetes mellitus telah menjadi penyebab dari 4,6 juta kematian. Selain itu pengeluaran biaya kesehatan untuk Diabetes Mellitus telah mencapai 465 miliar USD. *International Diabetes Federation* (IDF) memperkirakan bahwa sebanyak 183 juta orang tidak menyadari bahwa mereka mengidap DM. Sebesar 80% orang dengan DM tinggal di negara berpenghasilan rendah dan menengah. Pada tahun 2006, terdapat lebih dari 50 juta orang yang menderita DM di Asia Tenggara. Jumlah penderita DM terbesar berusia antara 40-59 tahun (Trisnawati & Setyorogo, 2013).

Ada beberapa jenis Diabetes Mellitus yaitu Diabetes Mellitus Tipe I, Diabetes Mellitus Tipe II, Diabetes Mellitus Tipe Gestasional, dan Diabetes Mellitus Tipe Lainnya. Jenis Diabetes Mellitus yang paling banyak diderita adalah Diabetes Mellitus Tipe 2. Diabetes Mellitus Tipe 2 (DM Tipe 2) adalah penyakit gangguan metabolik yang di tandai oleh kenaikan gulah darah akibat penurunan sekresi insulin oleh sel beta pankreas dan atau gangguan fungsi insulin (resistensi insulin) (Trisnawati & Setyorogo, 2013).

Diabetes melitus merupakan salah satu penyakit kronis yang dapat meningkatkan dengan cepat prevalensi komplikasi kronis pada lansia. Hal ini disebabkan kondisi hiperglikemia akibat ketiadaan absolut insulin atau penurunan relatif sensitivitas sel terhadap insulin, akan memicu munculnya penyakit tidak menular kronis lainnya, bahkan kematian penyandang diabetes melitus tidak jarang disebabkan oleh komplikasi.8,9 Klub Persadia Rumah Sakit Islam Jakarta Pondok Kopi Tahun 1998 _ 2005 menjelaskan bahwa komplikasi

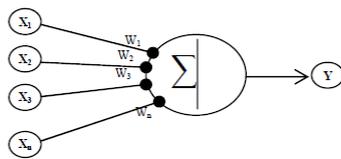
diabetes terbanyak adalah hipertensi dengan proporsi sekitar 54,2%. (Rosyada & Trihandini, 2010).

Prevalensi diabetes melitus yang terdiagnosis pada penduduk usia di atas 65 tahun menurut data Riskesdas tahun 2007 adalah 4,6%. Diabetes melitus pada lansia seringkali tidak disadari karena gejala-gejala diabetes seperti sering haus, sering berkemih, dan penurunan berat badan tersamarkan akibat perubahan fisik alamiah lansia yang mengalami penurunan, sehingga diabetes yang tidak terdiagnosis ini akan terus berkembang menjadi komplikasi yang dapat berakibat fatal.11,12 Data di Rumah Sakit Umum Daerah (RSUD) dr. H. Abdul Moeloek Provinsi Lampung menunjukkan angka prevalensi diabetes melitus mencapai 35,5% dengan 76,2% mengalami komplikasi.13 Poli lansia puskesmas kecamatan Jatinegara juga menunjukkan angka prevalensi diabetes melitus yang tak jauh berbeda yaitu mencapai 29,3% pada lansia usia 60 _ 69 tahun dan 20,8% pada lansia usia \geq 70 tahun. (Rosyada & Trihandini, 2010).

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. Istilah *data mining* memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki (Ridwan, Suyono, & Sarosa, 2013).

NN Merupakan sebuah sistem pembelajaran terhadap penerimaan informasi yang memiliki kinerja layaknya sebah jaringan syaraf pada manusia. *NN* diimplementasikan dengan menggunakan program komputer sehingga ma mpu menyelesaikan sejumlah proses perhitungan. Salah satu pengguna *NN* adalah untuk pengenalan pola, sistem pengenalan pola merupakan komponen penting dalam proses peniruan cara kerja manusia (Nurmila & Sugiharto, 2005).

Neural Network adalah prosesor yang terdistribusi paralel, terbuat dari unit-unit yang sederhana, dan memiliki kemampn untuk menyimpan pengetahuan yang diperoleh secara eksperimental dan siap pakai untuk berbagai tujuan



Gambar 1. Proses Komunikasi Antar Neuron

Pada gambar 1 diperlihatkan bahwa *neural network* terdiri atas satuan-satuan pemroses berupa neuron. Y sebagai *output* menerima *input* dari neuron $X_1, X_2, X_3, \dots, X_n$ dengan bobot $W_1, W_2, W_3, \dots, W_n$. Hasil penjumlahan seluruh *impuls neuron* dibandingkan dengan nilai ambang tertentu melalui fungsi aktivasi f setiap *neuron*. Fungsi aktivasi digunakan sebagai penentu keluaran suatu *neuron*.

Salah satu metode pelatihan dalam *NN* adalah pelatihan terbimbing (*supervised learning*). Pada pelatihan terbimbing diperlukan sejumlah masukan dan target yang berfungsi untuk melatih jaringan hingga diperoleh bobot yang diinginkan.

Maka dalam penelitian ini penulis berfokus pada algoritma klasifikasi *machine learning* yang akan digunakan adalah *neural network* yang bertujuan untuk memprediksi diagnosa penyakit diabetes yang ditujukan dengan besarnya akurasi setelah penulis melakukan penelitian menggunakan *tools* Rapid Miner.

2. METODE PENELITIAN

Penelitian ini dilakukan melalui beberapa tahapan, yaitu:

1) Menentukan Dataset

Penulis menggunakan data untuk penelitian ini adalah data diabetes yang didapat dari *Machine Learning Repository* UCI (Universitas California Invene) dengan alamat web : <http://archive.ics.uci.edu/ml>. Data yang dipakai sebanyak 220 *sample*.

2) Studi Literatur

Setelah penulis memperoleh dataset, kemudian mencari literatur guna menjadikan tolak ukur penelitian yang dilakukan.

3) Menentukan metode klasifikasi

Metode yang dipilih untuk permasalahan yang ada, penulis melihat dari bentuk dataset yang memiliki label (*supervised*) serta studi literatur yang dipelajari memiliki kecocokan sesuai yang sama, dengan demikian penulis memilih algoritma *neural network*

karena merupakan salah satu algoritma klasifikasi prediksi yang akurasi cukup tinggi.

4) Pengolahan Data Awal (*Pre Processing*)

Sebanyak 220 Data set dari data diabetes yang didapat dari *UCI Repository* akan ditransformasikan sesuai algoritma *neural network* dengan kriteria input data untuk rapid miner 9.0

5) Menerapkan Metode *Learning* dan *Testing* menggunakan *Neural Network*

Pembelajaran atau proses *learning* proses pertama dengan cara menghitung nilai prediksi data pada waktu terdahulu. Nilai diabetes diproses *learning* terlebih dahulu untuk mengetahui apakah dalam data ada nilai nominal atau tidak, apabila ada maka ubah nilai nominal menjadi numerik karena ada *boost* tidak dapat membaca nilai nominal. Pada tahap ini data hasil pengolahan antara *training* menggunakan *Neural Network* dan *learning* menggunakan *Neural Network* (*backpropagation*) setelah itu diuji untuk mengukur berapa akurasi *trend* dari prediksi.

6) Menerapkan Metode Evaluasi dan Validasi Hasil

Proses evaluasi dan juga pengamatan hasil kerja *Neural Network* pada Rapid Miner. Validasi dilakukan untuk melakukan pengukuran hasil prediksi.

a. *Cross Validation*

Cross validation adalah pengujian standar yang dilakukan untuk memprediksi *error rate*. Data *training* dibagi secara random ke dalam beberapa bagian dengan perbandingan yang sama kemudian *error rate* dihitung bagian demi bagian, selanjutnya hitung rata-rata seluruh *error rate* untuk mendapatkan *error rate* secara keseluruhan (Kepemilikan & Bemotor, 2013)

b. Pengukuran menggunakan *ROC Curve*

Hasil perhitungan divisualisasikan dengan kurva ROC. Kurva ROC pada gambar mengekspresikan *confusion matrix* dari Gambar 2. Garis horizontal adalah *false positives* dan garis vertikal *true positives*. Terlihat pada gambar nilai AUC sebesar 0.984.

Untuk klasifikasi *data mining*, nilai AUC dapat dibagi menjadi beberapa kelompok.

a. 0.90-1.00 = klasifikasi sangat baik

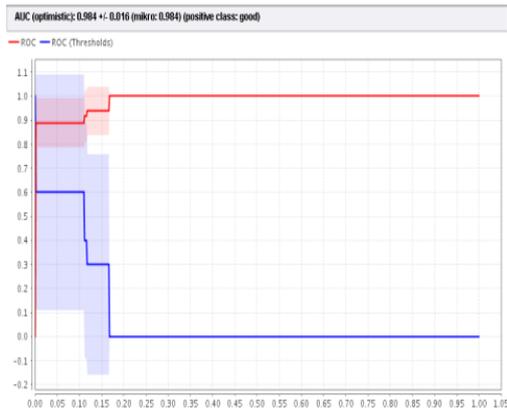
b. 0.80-0.90 = klasifikasi baik

c. 0.70-0.80 = klasifikasi cukup

d. 0.60-0.70 = klasifikasi buruk

e. 0.50-0.60 = klasifikasi salah

(Kepemilikan & Bemotor, 2013)



Sumber (Kepemilikan & Bemotor, 2013)

Gambar 2. Kurva ROC Metode KNN

c. Pengukuran Confusion Matrix

Confusion matrix merupakan sebuah tabel yang terdiri dari atas banyaknya baris data uji yang diprediksi benar dan tidak benar oleh model klasifikasi, tabel ini diperlukan untuk menentukan kinerja suatu model klasifikasi (Wijayanto, n.d.)

Tabel 1. Tabel Confusion Matrix

		Predicted Class	
		Class=1	Class=0
Actual Class	Class=1	F 11	F 10
	Class=0	F 01	F 00

Perhitungan akurasi dengan menggunakan tabel *confusion matrix* adalah sebagai berikut :

$$\text{Akurasi} = \frac{F 11 + F 00}{F 11 + F 10 + F 01 + F 00}$$

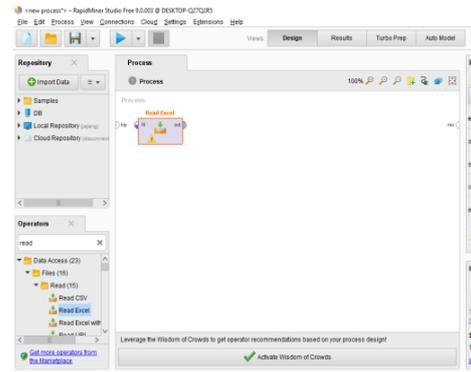
7) Menarik Kesimpulan

Setelah penulis selesai mengelola dataset dan didapat hasilnya secara kuantitatif, maka penulis dapat menarik kesimpulan apakah dataset tersebut dengan klasifikasi 1 berhasil dan 0 untuk gagal mendapatkan akurasi yang tinggi atau tidak, sehingga keberhasilan metode pengobatan.

3. PEMBAHASAN

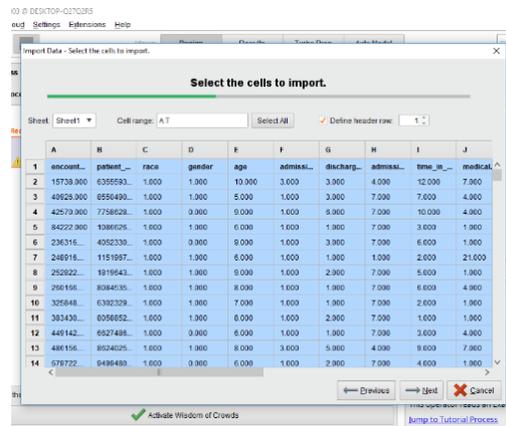
Proses *training, learning,* dan *testing* menggunakan model *neural network* dengan dataset sebanyak 220 data sebagai berikut:

- 1) Membuka aplikasi tools *rapid miner*, cari *read excel* di operator untuk memasukkan dataset



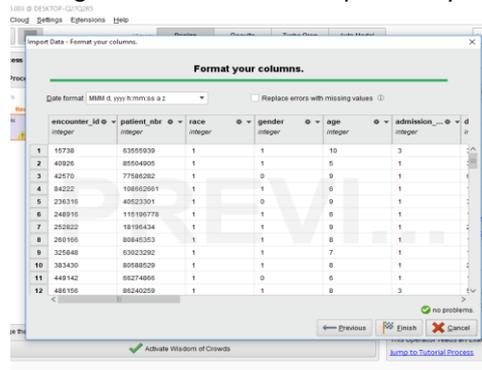
Gambar 3. Input Operator Read Excel

- 2) Memasukan data primer ke dalam *read excel* dengan atributnya adalah *encounter_id, patient_nbr, race, gender, age, admission_type_id, discharge_disposition_id, admission_source_id, time_in_hospital, medical_speciality, num_lab_procedures, num_procedures, num_medications, number_outpatient, number_emergency, number_inpatient, diag_1, diag_2, diag_3, diabetesMed,*



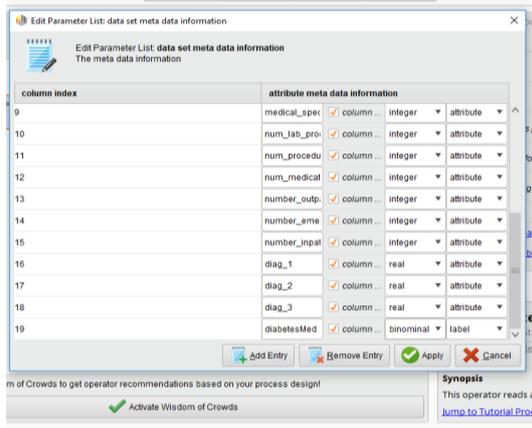
Gambar 4. Input Dataset

- 3) Setelah memasukkan dataset, tool akan mengetahui format dari setiap kolomnya.



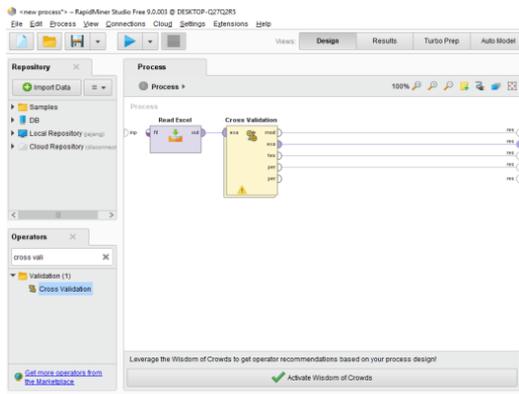
Gambar 5. Format dataset

- 4) Di kolom edit parameter data set meta data information kita bisa mengedit atribut yang kita udah masukan.



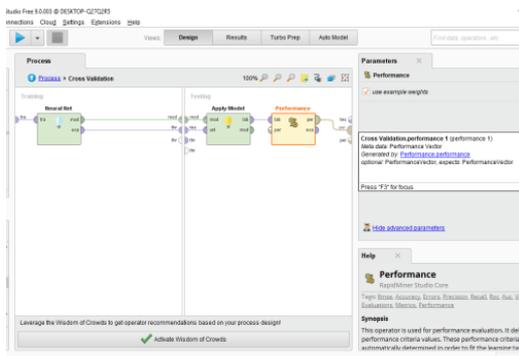
Gambar 6. Mengatur Atribut

- 5) Proses *insert* dan atur *validation* dengan number of folds 10.



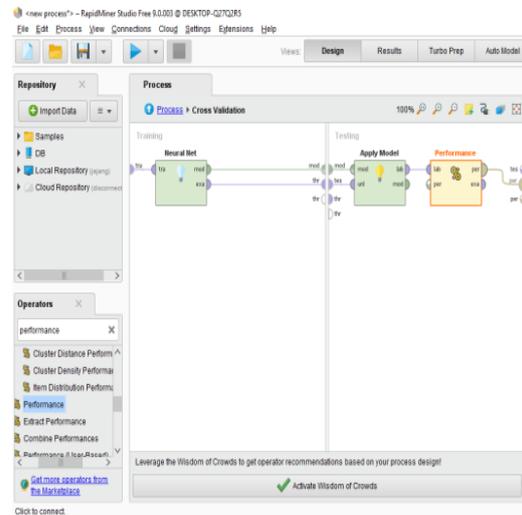
Gambar 7. Proses Mengatur Atribut

- 6) Proses berikutnya adalah proses *insert* model yang akan di *training*, pada proses *training* kali ini dengan memasukkan model algoritma neural network dengan menggunakan data set sebanyak 220.



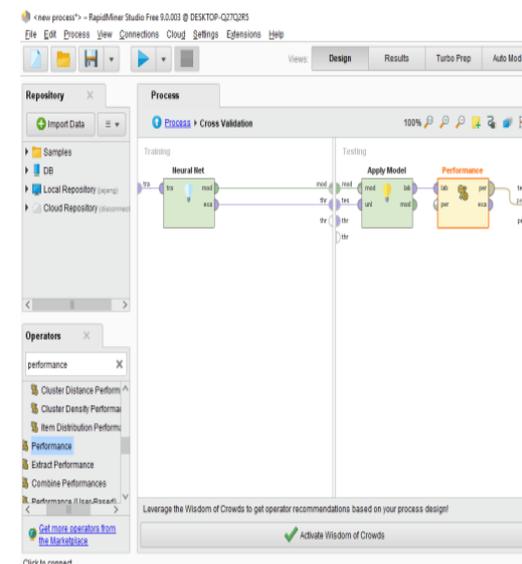
Gambar 8. Insert dan Atur Validation

- 7) Proses terakhir adalah proses testing dengan cara *insert apply* dan *performance* dengan main criterion *accuracy* dan *AUC*.



Gambar 9. Proses Insert Model NN Yang Akan di Training

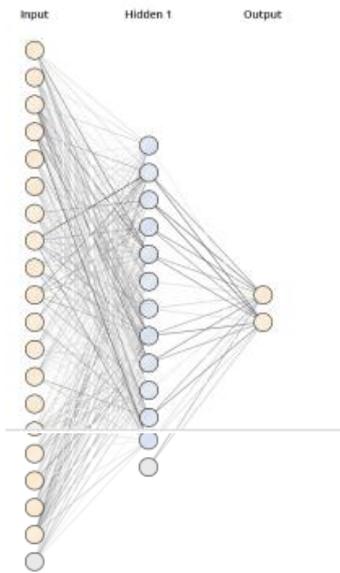
- 8) Hasil testing bisa diperoleh dengan cara insert apply model dan *performance* dengan main criterion *accuracy* dan *AUC*.



Gambar 10. Proses Testing Dengan Cara Insert Apply Model dan Performance

- 9) Hasil *testing* menunjukkan *performance accuracy* yang dihasilkan menggunakan algoritma Neural Network sebesar 80.00% dan nilai *AUC* sebesar 0.605.

- 10) Setelah proses *testing* dengan *apply* model hasil pengujian Neural Network

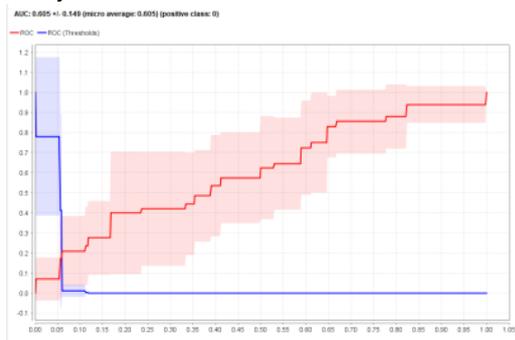


Gambar 11. Model Hasil Pengujian Neural Network

Evaluasi dan Validasi

Penelitian ini bertujuan untuk menguji akurasi dari analisa penyakit diabetes menggunakan algoritma *neural network* data yang dianalisa adalah data diabetes yang diperoleh data dari *UCI Repository*. Dengan bantuan *rapid miner* ternyata dapat diketahui algoritma *neural network* mempunyai akurasi yang tinggi.

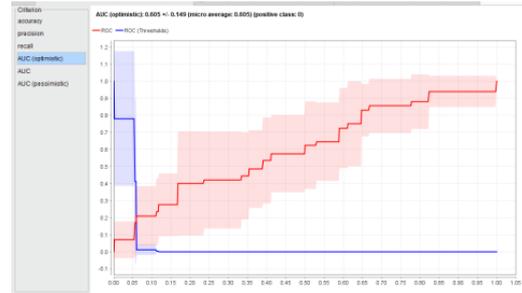
a. Uji Roc Curva



Gambar 12. AUC Neural Network

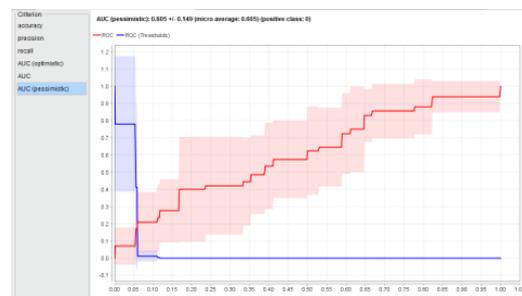
Menggambarkan grafik *under curve* (AUC) hasil validasi *Neural Network* dengan hasil AUC sebesar 0,605. Ini menunjukkan nilai dibawah *curve* sebesar 0,605.

b. Menggambarkan grafik area *under curve* (AUC) *optimistic* hasil validasi *Neural Network* dengan nilai yang tertera dalam gambar sebesar = 0,605.



Gambar 13. AUC Optimistic

c. Menggambarkan grafik area *under curve* (AUC) *pesimistic* hasil validasi *Neural Network* sebesar = 0,605.



Gambar 14. AUC Pesimistic

d. Uji Confusion Matrix

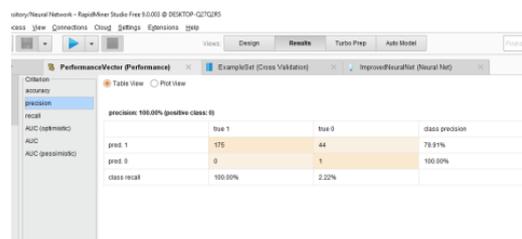
accuracy: 80.00% +/- 3.02% (micro average: 80.00%)

	true 1	true 0	class precision
pred 1	175	44	79.91%
pred 0	0	1	100.00%
class recall	100.00%	2.22%	

Gambar 15. Uji Confusion Matrix

Berdasarkan gambar diatas dapat ketahui bahwa hasil *accuracy* dari metode klasifikasi *Neural Network* sebesar 80.00 % ini menunjukkan bahwa hasil akurasi yang diperoleh termasuk kedalam kategori baik.

e. Menggambarkan precision dari metode klasifikasi *Neural Network* sebesar 100.00 % (positive class: 0) ini menunjukkan bahwa hasil akurasi yang diperoleh masuk dalam kategori sangat baik.



Gambar 16. Precision Klasifikasi NN

f. Menggambarkan hasil *recall* dari metode klasifikasi *Neural Network* sebesar = 2.50 %

	True 1	True 0	Class Precision
Recall	recall 2.50% +/- 7.50% (micro average: 2.22%) (positive class: 0)		
AUC (optimistic)	175	44	79.91%
AUC	0	1	100.00%
AUC (pessimistic)	100.00%	2.22%	

Gambar 17. Recall Klasifikasi Neural Network

g. Memberikan informasi secara keseluruhan tentang hasil dan validasi dengan metode klasifikasi *Neural Network* yang terdiri dari hasil *accuracy*, *precision*, *recall* dan juga nilai *AUC* dan hasil lainnya seperti performance vektor (confusion matrix).

```

PerformanceVector
PerformanceVector:
accuracy: 80.00% +/- 3.02% (micro average: 80.00%)
ConfusionMatrix:
True: 1 0
I: 175 44
O: 0 1
precision: 100.00% (positive class: 0)
ConfusionMatrix:
True: 1 0
I: 175 44
O: 0 1
recall: 2.50% +/- 7.50% (micro average: 2.22%) (positive class: 0)
ConfusionMatrix:
True: 1 0
I: 175 44
O: 0 1
AUC (optimistic): 0.605 +/- 0.149 (micro average: 0.605) (positive class: 0)
AUC: 0.605 +/- 0.149 (micro average: 0.605) (positive class: 0)
AUC (pessimistic): 0.605 +/- 0.149 (micro average: 0.605) (positive class: 0)
    
```

Gambar 18. Performance Vektor Neural Network

4. KESIMPULAN

Berdasarkan penelitian yang dilakukan oleh penulis menggunakan tools rapid 9.0 terhadap dataset diabetes yang didapat dari *UCI repository* yang diuji dengan metode *Neural Network* menghasilkan nilai *accuracy* = 80.00 % , *precision* = 100.00 % dan *recall* = 2.50% dengan *AUC* sebesar = 0.605 yang menunjukkan bahwa hasil klasifikasinya cukup, sehingga dari dataset tersebut dapat diprediksi mana saja yang mengidap penyakit diabetes dan mana yang bukan, sehingga dapat diprediksi dan menjadi tolak ukur diagnosis sehingga dapat dideteksi lebih dini.

5. REFERENSI

Fatimah, R. N. (2015). DIABETES MELITUS TIPE 2. *J MAJORITY*, 4, 93–101.

Kepemilikan, K., & Bemotor, K. (2013). Penerapan algoritma k-nearest neighbor untuk penentuan resiko kredit kepemilikan kendaraan bemotor. *Jurnal Penelitian Ilmu Komputer, System Embedded & Logic*, 1(1), 65–76.

Nurmila, N., & Sugiharto, A. (2005). Algoritma back propagation neural network untuk pengenalan pola karakter huruf jawa. *Masyarakat Informatika*, 1, 1–10.

Ridwan, M., Suyono, H., & Sarosa, M. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *EECCIS*, 7(1), 59–64.

Rosyada, A., & Trihandini, I. (2010). Determinan Komplikasi Kronik Diabetes Melitus pada Lanjut Usia Determinan of Diabetes Mellitus Chronic Complications on Elderly. *Jurnal Kesehatan Masyarakat Nasional*, 395–401.

Trisnawati, S. K., & Setyorogo, S. (2013). Faktor Risiko Kejadian Diabetes Melitus Tipe II Di Puskesmas Kecamatan Cengkareng Jakarta Barat Tahun 2012. *Jurnal Ilmiah Kesehatan*, 5(1), 6–11.

Wijayanto, H. (n.d.). KLASIFIKASI BATIK MENGGUNAKAN METODE K-NEAREST NEIGHBOUR BERDASARKAN GRAY LEVEL CO-OCCURRENCE MATRICES (GLCM), (5).