
KOMPARASI KLASIFIKASI PADA PREDIKSI PENDAPATAN RUMAH TANGGA

Evy Priyanti

Program Studi Sistem Informasi Akuntansi
Universitas Bina Sarana Informatika (UBSI)
Jl. Kamal Raya No 18 Ringroad Barat, Cengkareng, Jakarta Barat
evy.evp@bsi.ac.id

ABSTRAK

Kebutuhan akan kehidupan sehari-hari tidak terlepas dari pendapatan yang dihasilkan, baik pendapatan harian, mingguan atau bulanan. Oleh karena itu prediksi pendapatan rumah tangga sangat penting dikarenakan akan membantu dalam menciptakan pendapatan yang lebih baik, dalam memprediksi suatu data dapat dilakukan dengan beberapa algoritma diantaranya dengan algoritma K-nearest neighbor dan algoritma Neural Network. Pada penelitian kali ini akan dikomparasi bagaimana algoritma K-Nearest neighbor dengan Neural network dalam memprediksi pendapatan rumah tangga pada sensus yang dilakukan di Bereu pada tahun 1996. Algoritma K-Nearest neighbor menghasilkan nilai akurasi sebesar 70,49% sedangkan algoritma Neural Network menghasilkan akurasi sebesar 83,62%, hal ini membuktikan bahwa algoritma Neural Network dapat bekerja lebih baik dalam memprediksi pendapatan rumah tangga di Bereu yang dilakukan oleh Ronny Kohavi dan Barry Becker pada 1 Mei 1996 yang terdiri dari 48842 dataset dan 14 atribut. Beberapa atribut menjadi faktor penentu dalam menciptakan pendapatan yang lebih tinggi, diantaranya status pernikahan yang utuh yang terdiri dari minimal suami dan istri dalam satu atap dan pendidikan yang tinggi mendapatkan peluang untuk mendapatkan penghasilan yang lebih tinggi dari pasangan itu sendiri yang nantinya dapat mempengaruhi pendapatan rumah tangga lebih dari \$50.000/tahun. Selain itu faktor pengalaman bekerja juga menjadi salah satu faktor penentu tingginya pendapatan rumah tangga. Faktor ketidakharmonisan dalam rumah tangga juga menjadi salah satu faktor pendapatan yang kurang dari \$50.000/tahun.

ABSTRACT

The need for daily life is inseparable from the income generated, whether daily, weekly or monthly income. Therefore prediction of household income is very important because it will help in creating a better income, in predicting a data can be done with several algorithms including the K-nearest neighbor algorithm and Neural Network algorithm. This study will compare how the K-Nearest neighbor algorithm with Neural Network predicts household income in the census conducted in Bereu in 1996. The K-Nearest neighbor algorithm produces an accuracy value of 70.49% while the Neural Network algorithm produces an accuracy of 83.62%, this proves that the Neural Network algorithm can work better in predicting household income in Bereu conducted by Ronny Kohavi and Barry Becker on May 1, 1996 consisting of 48842 datasets and 14 attributes. Some attributes are the deciding factors in creating higher income, including complete marital status consisting of a minimum of husband and wife under one roof and higher education to get opportunities to earn higher income from the couple themselves which can later affect household income more than \$ 50,000 / year. In addition, work experience is also one of the determinants of high household income. The disharmony factor in the household is also a fact that income is less than \$ 50,000 / year.

Keyword :Household, Neural Network, K-NN

1. Pendahuluan

Sensus yang dilakukan oleh Ronny Kohavi and Barry Becker pada 1 Mei 1996 ini terdiri dari 48842 data. Sensus ini dilakukan di Bereu didapatkan data pendapatan untuk beberapa Negara di dunia dengan besaran yang sudah diklasifikasikan menjadi beberapa bagian yaitu pendapatan yang lebih besar dari \$50.000/tahun, pendapatan yang sama besar \$50.000/tahun dan pendapatan yang kurang dari \$50.000/tahun. Data yang didapat terdiri dari beberapa atribut yaitu *age, workclass, state-gov, fnlwtg, education, education-num, marital-status, occupation, relationship, race, sex, capital-gain, capital-loss, hours-per-week, native-country*. (Priyanti, 2017)

Beberapa faktor yang mempengaruhi pendapatan rumah tangga lebih dari \$50.000/tahun diantaranya adalah jenis keluarga, bahwasannya keluarga yang utuh yang minimal terdiri dari suami dan istri dalam satu atap memiliki peluang mendapatkan pendapatan rumah tangga yang lebih besar dibandingkan dengan jenis keluarga yang tidak utuh. Pendidikan, seseorang yang memiliki pendidikan yang lebih tinggi memiliki peluang untuk mendapatkan penghasilan yang lebih tinggi. Lama bekerja, semakin lama seorang pekerja semakin besar pula penghasilan yang didapatkannya dan ini membuktikan bahwa seseorang dengan pengalaman yang lebih banyak dapat menghasilkan pendapatan yang lebih banyak pula. (priyanti, 2017).

Pada penelitian kali ini akan dibuktikan bahwa algoritma neural network akan meningkatkan akurasi dari data pendapatan penduduk ada di Bereu.

Menurut Heaton (2010:p27) Neural Network adalah jaringan saraf yang mensimulasikan jaringan saraf biologis manusia kedalam arsitektur komputer dan arsitektur algoritma baru terhadap komputer konvensional. Hal ini memungkinkan penggunaan operasi komputasi (penambahan, pengurangan, dan elemen logika fundamental) yang sangat sederhana untuk memecahkan masalah yang kompleks, matematis yang tidak jelas, masalah nonlinear atau masalah stokastik.

Penelitian sebelumnya Valentin Mihov dalam penelitiannya yang berjudul *Adult Income Dataset Analysis with IPython* menggunakan *Learning Model Logistic regression* pada tahun 2015 yang hanya memisahkan antara data yang berhubungan dengan data yang tidak berhubungan. Fitur yang bertanda negatif terdiri dari Never married, Own child, priv-house-serv dan divorce sedangkan fitur yang bernilai positif terdiri dari capital gain, married-civ-spouse,

age, hours per week, dan exec-managerial, selanjutnya fitur yang bernilai negatif akan dipisahkan supaya proses analisis dapat berjalan lebih baik.

2. Metode Penelitian

Atribut-atribut yang digunakan pada penelitian kali ini didapat dari UCI dataset yang terdiri dari :

age: dengan jenis continuous.

workclass: terdiri dari Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov, Without-pay, Never-worked.

fnlwtg: dengan jenis continuous.

education: terdiri dari Bachelors, Some-college, 11th, HS-grad, Prof-school, Assoc-acdm, Assoc-voc, 9th, 7th-8th, 12th, Masters, 1st-4th, 10th, Doctorate, 5th-6th, Preschool.

education-num: dengan jenis continuous.

marital-status: terdiri dari Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent, Married-AF-spouse.

occupation: terdiri dari Tech-support, Craft-repair, Other-service, Sales, Exec-managerial, Prof-specialty, Handlers-cleaners, Machine-op-inspct, Adm-clerical, Farming-fishing, Transport-moving, Priv-house-serv, Protective-serv, Armed-Forces.

relationship: terdiri dari Wife, Own-child, Husband, Not-in-family, Other-relative, Unmarried.

race: terdiri dari White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black.

sex: terdiri dari Female, Male.

capital-gain: dengan jenis continuous.

capital-loss: dengan jenis continuous.

hours-per-week: dengan jenis continuous.

native-country: terdiri dari United-States, Cambodia, England, Puerto-Rico, Canada, Germany, Outlying-US(Guam-USVI-etc), India, Japan, Greece, South, China, Cuba, Iran, Honduras, Philippines, Italy, Poland, Jamaica, Vietnam, Mexico, Portugal, Ireland, France, Dominican-Republic, Laos, Ecuador, Taiwan, Haiti, Columbia, Hungary, Guatemala, Nicaragua, Scotland, Thailand, Yugoslavia, El-Salvador, Trinidad&Tobago, Peru, Hong, Holand-Netherlands.

Atribut-atribut diatas dideskripsikan sebagai berikut :

Usia terdiri dari usia 17-90 tahun.

Kelas pekerja terdiri dari pegawai swasta, wiraswasta yang bukan dari perusahaan, pemerintah federal, pemerintah local, pemerintah pusat, pekerja tanpa bayaran, dan tidak pernah bekerja.

Sample adalah nomor sample

Pendidikan terdiri dari Sarjana, Beberapa perguruan tinggi, kelas 11, HS-grad, Prof-

school, Assoc-acdm, Assoc-voc, 9th, 7th-8th, 12th, Masters, 1st-4th, 10th, Doctorate, 5th-6th, Prasekolah

Nomor pendidikan terdiri dari 1 sampai dengan 16.

Status pernikahan terdiri dari Married-civ-spouse=menikah dengan warga sipil, Divorced=bercerai, Never-married=memiliki pasangan tetapi tidak menikah secara resmi, Separated=terpisah, Widowed=janda atau duda yang belum menikah lagi, Married-spouse-absent=pasangan yang menikah akan tetap tidak satu rumah tangga bisa jadi karena dipisahkan atau belum resmi bercerai akan tetapi sudah tidak 1 rumah, Married-AF-spouse= menikah dengan angkatan/prajurit.

Jenis keluarga juga terdiri dari dua bagian yaitu keluarga lengkap dan keluarga tidak lengkap. Yang terdiri dari keluarga tidak lengkap adalah divorce, never-married, separated, married – spouse-absen

Keluarga lengkap terdiri dari married-civ-spouse dan married-AF-spouse.

Pekerjaan: Tech-support, Craft-repair, Other-service, Sales, Exec-managerial, Prof-specialty, Handlers-cleaners, Machine-op- inspct, Adm-clerical, Farming-fishing,

Dikarenakan fungsi aktivasi pada neural network yang dipakai fungsi aktivasi biner, data harus ditransformasikan dulu karena batasan keluaran fungsi aktivasi sigmoid adalah [0,1], tapi akan lebih baik jika ditransformasikan ke interval yang lebih kecil, missal pada interval Berikut perhitungan transform dengan menggunakan fungsi aktivasi sigmoid:

Berdasarkan data pendapatan rumah tangga dengan atribut *age* diketahui

nilai minimum (a) = 17

nilai maximum (b) = 90

pendapatan rumah tangga yang akan di

transform $(x) = x^l = \frac{0.8(39-17)}{90-17} + 0.1$

3. Hasil dan Pembahasan

Algoritma K-Nearest Neighbors dikenal sebagai algoritma lazy learning karena membutuhkan lebih sedikit waktu untuk pelatihan. Komputasinya seperti algoritma pembelajaran eager-based learning yang membutuhkan waktu lebih sedikit selama klasifikasi. (singh, 2017). Pelatihan dalam k-nn untuk memprediksi label kelas kategorikal (Nikam, 2015). Pada K-Nearest neighbour membutuhkan nilai k tetangga atau nilai k disekitarnya instance data uji baru untuk diklasifikasikan keladalam kelas maksimum, disarankan kelas k bernilai ganjil supaya akurasi klasifikasi dapat memperoleh nilai yang lebih tinggi (nurwanto, 2016).

Berikut algoritma k-nearest neighbour:

Transport-moving, Priv-house-serv, Protective-serv, Armed-Forces.

hubungan: istri, anak kandung, suami, bukan bagian dari keluarga inti, relasi lainnya, tidak menikah.

ras: White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black.

jenis kelamin: wanita, pria.

keuntungan dalam bentuk uang: bernilai antara 1 sampai dengan 99.999
kehilangan modal: bernilai antara 1 sampai dengan 3.004.

jam per minggu: bernilai antara 1 sampai dengan 99.

Negara asal: United-States, Cambodia, England, Puerto-Rico, Canada, Germany, Outlying-US(Guam-USVI-etc), India, Japan, Greece, South, China, Cuba, Iran, Honduras, Philippines, Italy, Poland, Jamaica, Vietnam, Mexico, Portugal, Ireland, France, Dominican-Republic, Laos, Ecuador, Taiwan, Haiti, Columbia, Hungary, Guatemala, Nicaragua, Scotland, Thailand, Yugoslavia, El-Salvador, Trinidad&Tobago, Peru, Hong, Holand-Netherlands.

[0.1,0.9] (Siang ,2009:p25). Maka, pada data pendapatan rumah tangga yang ada dilakukan transform data dengan interval [0.1,0.9], dengan rumus sebagai berikut:

$$x^l = \frac{0.8(x - a)}{b - a} + 0.1$$

$$x^l = \frac{0.8(22)}{73} + 0.1$$

$$x^l = 0.889$$

Selanjutnya akan dilakukan pengujian dan pelatihan apakah dengan nilai transform pendapatan rumah tangga dengan nilai atribut 39 dan nilai transform 0.889 memiliki nilai akurasi yang tinggi.

Calculate optimal k with least misclassification error.

Create empty list predictions

For each t in testSet

Find neighbors:

Create empty list distances

For each x in trainingSet

dist= Euclidean distance measure between x and test instance t

distances.append(trainingSet[x], dist)

End For

Sort distances in ascending order.

Create list neighbors by taking k subset of training points from distances

Return neighbors

result = maximum voted class in the neighbors

predictions.append(result)

End For

Calculate accuracy using predicted list predictions of target values and actual target values

	true <=50K	true >50K	true >=50	class precision
pred. <=50K	3901	957	1	80.28%
pred. >50K	943	644	0	40.58%
pred. >=50	2	0	0	0.00%
class recall	80.50%	40.22%	0.00%	

Gambar 3.1. Hasil Akurasi K-NN

Rule Model yang dihasilkan pada K-NN sebagai berikut :

if age \leq 31.500 then \leq 50K (2076 / 161 / 1)

if education-num \leq 11.500 and relationship $>$ 3.500 then \leq 50K (566 / 25 / 0)

if education-num \leq 11.500 and capital-gain \leq 4243.500 then \leq 50K (1595 / 546 / 0)

if capital-gain $>$ 5095.500 then $>$ 50K (10 / 277 / 0)

if marital-status $>$ 2.500 and sex $>$ 1.500 and fnlwgt \leq 179660.500 then \leq 50K (80 / 5 / 0)

if capital-loss $>$ 1794.500 then $>$ 50K (6 / 88 / 0)

if hours-per-week \leq 42.500 and fnlwgt $>$ 206248.500 and relationship \leq 1.500 and fnlwgt \leq 305627.500 then \leq 50K (36 / 4 / 0)

if marital-status \leq 2.500 and hours-per-week $>$ 41 and marital-status $>$ 1.500 and occupation \leq 3 and hours-per-week \leq 57.500 then $>$ 50K (9 / 61 / 0)

if marital-status $>$ 2.500 and education-num \leq 12.500 then \leq 50K (27 / 1 / 0)

if workclass \leq 3.500 and occupation $>$ 4.500 and age \leq 40.500 and hours-per-week \leq 41 and education-num \leq 13.500 and fnlwgt $>$ 224890 then \leq 50K (11 / 2 / 0)

if relationship $>$ 1.500 and education $>$ 1.500 and age $>$ 33.500 and fnlwgt \leq 202449.500 and fnlwgt $>$ 121054.500 and fnlwgt \leq 143877 then $>$ 50K (0 / 17 / 0)

if education-num \leq 14.500 and hours-per-week \leq 35.500 and age $>$ 54.500 and age $>$ 61.500 then \leq 50K (17 / 3 / 0)

if education-num \leq 14.500 and marital-status $>$ 2.500 and hours-

per-week \leq 44 and workclass $>$ 2.500 and age \leq 50 then \leq 50K (22 / 1 / 0)

if education-num $>$ 14.500 and age \leq 52.500 and fnlwgt \leq 346399.500 and fnlwgt $>$ 106099.500 then $>$ 50K (5 / 41 / 0)

if workclass \leq 3.500 and native-country $>$ 4.500 and native-country \leq 9.500 then \leq 50K (14 / 2 / 0)

if workclass \leq 3.500 and hours-per-week $>$ 58 and fnlwgt $>$ 194906 and workclass \leq 2.500 then \leq 50K (9 / 2 / 0)

if relationship \leq 3.500 and relationship $>$ 2.500 and age \leq 37.500 then $>$ 50K (2 / 15 / 0)

if relationship $>$ 3.500 and marital-status \leq 1.500 then \leq 50K (35 / 2 / 0)

if fnlwgt $>$ 119015 and relationship $>$ 1.500 and fnlwgt \leq 201918 and occupation \leq 6.500 and age \leq 50.500 and age $>$ 40.500 then $>$ 50K (5 / 39 / 0)

if fnlwgt \leq 204424.500 and age \leq 47.500 and fnlwgt $>$ 150182 and age $>$ 37.500 and fnlwgt $>$ 190655.500 then \leq 50K (12 / 0 / 0)

if fnlwgt $>$ 188539.500 and fnlwgt \leq 243494 and age \leq 46.500 and age $>$ 39.500 and workclass $>$ 2.500 then $>$ 50K (2 / 16 / 0)

if capital-gain $>$ 4518 then \leq 50K (18 / 1 / 0)

if workclass $>$ 3.500 and fnlwgt \leq 160164.500 and education-num $>$ 13.500 and age \leq 47.500 then $>$ 50K (0 / 8 / 0)

if age \leq 47.500 and occupation \leq 1.500 and fnlwgt \leq 187078.500 then \leq 50K (18 / 1 / 0)

if sex \leq 1.500 and fnlwgt \leq 34398 and workclass \leq 4.500 then $>$ 50K (1 / 14 / 0)

if fnlwgt \leq 90725.500 and fnlwgt $>$ 65357 and hours-per-week \leq 41 then \leq 50K (13 / 2 / 0)

if relationship $>$ 1.500 and hours-per-week \leq 49 and occupation \leq 6.500 and fnlwgt \leq 273425 and age \leq 35.500 and fnlwgt $>$ 157517 then $>$ 50K (1 / 14 / 0)

if relationship \leq 1.500 and marital-status \leq 2 and sex $>$ 1.500 and hours-per-week \leq 47.500 then \leq 50K (19 / 1 / 0)

if relationship $>$ 1.500 and age $>$ 47.500 and occupation $>$ 5.500 and fnlwgt $>$ 173487 and education \leq

```

5.500 and age ≤ 57.500 then >50K
(3 / 18 / 0)
if fnlwgt ≤ 205814 and age > 60.500
then >50K (5 / 17 / 0)
if age > 44.500 and occupation > 3
and fnlwgt > 163993 and hours-per-
week > 44.500 and fnlwgt ≤ 344657
then ≤50K (21 / 5 / 0)
if fnlwgt > 90725.500 and hours-
per-week > 39 and age > 48.500 and
fnlwgt ≤ 272693.500 and hours-per-
week ≤ 49 and race ≤ 1.500 and age
≤ 54.500 then >50K (2 / 16 / 0)
if relationship ≤ 1.500 and
education-num ≤ 12.500 then ≤50K
(8 / 0 / 0)
if age > 44.500 and workclass ≤
3.500 and fnlwgt > 102468 and
occupation ≤ 1.500 then ≤50K (9
/ 1 / 0)
if fnlwgt > 265806 and fnlwgt ≤
313666 and hours-per-week ≤ 65 and
age ≤ 44.500 then >50K (1 / 17 /
0)
if hours-per-week ≤ 47 and fnlwgt >
140014 and fnlwgt ≤ 200941.500 and
fnlwgt ≤ 168774 and occupation >
6.500 then ≤50K (7 / 1 / 0)
if fnlwgt > 208415.500 and fnlwgt ≤
316152 and workclass ≤ 3.500 and age
≤ 44.500 and fnlwgt > 245527.500
then ≤50K (15 / 1 / 0)
if age ≤ 44.500 and age > 36.500 and
fnlwgt > 76409 and fnlwgt ≤ 120762
and hours-per-week ≤ 43.500 then
>50K (2 / 18 / 0)
if hours-per-week ≤ 47 and sex ≤
1.500 and relationship ≤ 1.500 and
age > 49 then ≤50K (6 / 0 / 0)
if hours-per-week > 47 and fnlwgt >
156146 and occupation > 6.500 then
>50K (1 / 8 / 0)
if fnlwgt ≤ 116039 and fnlwgt >
110042 and fnlwgt ≤ 112875 then
≤50K (7 / 1 / 0)
if fnlwgt ≤ 101759.500 and fnlwgt >
50606 and age > 51 then ≤50K (5
/ 0 / 0)
if hours-per-week > 47 and fnlwgt >
171588 and native-country ≤ 5.500
and fnlwgt > 349186 then >50K (1
/ 12 / 0)
if age ≤ 47.500 and education ≤
1.500 and age > 43.500 and fnlwgt >
85717 and workclass ≤ 5.500 and
marital-status > 1.500 then ≤50K
(15 / 1 / 0)
if workclass > 3.500 and fnlwgt ≤
326967 and fnlwgt > 169132.500 and
age > 36.500 then >50K (3 / 13 /
0)

```

```

if race > 1.500 and workclass >
2.500 and fnlwgt > 118569.500 and
fnlwgt ≤ 337590.500 then ≤50K (9
/ 0 / 0)
if hours-per-week > 47 and fnlwgt ≤
215753 and age ≤ 41.500 and age >
33.500 and education > 2.500 then
>50K (0 / 8 / 0)
if fnlwgt ≤ 99382 and age ≤ 37.500
and fnlwgt > 56670.500 then ≤50K
(5 / 0 / 0)
if education ≤ 1.500 and age >
43.500 and fnlwgt ≤ 363603 and
fnlwgt > 199825.500 and age > 53.500
then >50K (0 / 7 / 0)
if fnlwgt > 150593.500 and marital-
status ≤ 2.500 and fnlwgt ≤
156088.500 then ≤50K (5 / 0 / 0)
if fnlwgt ≤ 157065 and fnlwgt >
90725.500 and occupation ≤ 5 and
hours-per-week ≤ 47.500 and age ≤
59.500 then >50K (3 / 17 / 0)
if hours-per-week ≤ 47 and fnlwgt ≤
343569.500 and fnlwgt > 102468 and
fnlwgt ≤ 150330 then ≤50K (10 /
1 / 0)
if fnlwgt > 205949 and education >
5.500 and occupation > 3 then ≤50K
(9 / 1 / 0)
if hours-per-week > 46.500 and
fnlwgt > 99382 and fnlwgt ≤
181590.500 and hours-per-week > 51
and occupation > 3 then >50K (3 /
13 / 0)
if fnlwgt ≤ 343569.500 and hours-
per-week ≤ 46.500 and occupation ≤
8.500 and fnlwgt > 174586 and hours-
per-week > 37 and fnlwgt ≤ 203880
then ≤50K (10 / 2 / 0)
if hours-per-week ≤ 58 and
workclass ≤ 3.500 and fnlwgt >
170125 and fnlwgt ≤ 181590.500 and
hours-per-week > 32.500 then >50K
(0 / 7 / 0)
if hours-per-week > 57.500 and age
≤ 47.500 then ≤50K (10 / 2 / 0)
if workclass ≤ 4.500 and age >
33.500 and fnlwgt > 234132 and
education > 1.500 then >50K (2 /
9 / 0)
if fnlwgt > 165376.500 and fnlwgt ≤
197037 then ≤50K (7 / 1 / 0)
if hours-per-week > 53.500 and
relationship > 1.500 then >50K (0
/ 6 / 0)
if relationship ≤ 1.500 and
education > 2.500 and hours-per-
week > 37.500 then ≤50K (8 / 0 /
0)

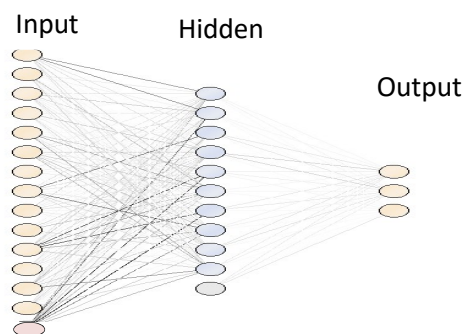
```

```

if fnlwgt ≤ 370677 and sex > 1.500
and fnlwgt ≤ 335257 then >50K (1
/ 11 / 0)
if age > 50.500 then ≤50K (6 / 1
/ 0)
if hours-per-week ≤ 42.500 and
occupation ≤ 8.500 and education-
num ≤ 13.500 and fnlwgt > 416296
then ≤50K (4 / 0 / 0)
if age > 37.500 and age ≤ 40.500 and
fnlwgt ≤ 182884 then >50K (0 / 6
/ 0)
if fnlwgt ≤ 102068.500 and age ≤
43.500 and age > 36 then ≤50K (7
/ 0 / 0)
if age > 42.500 and workclass ≤
4.500 and age > 45.500 then >50K
(1 / 8 / 0)
if age > 45.500 then ≤50K (4 / 0
/ 0)

```

Nilai uji dataset menggunakan algoritma neural network didapatkan data akurasi yang terlihat pada gambar 3.2. sebagai berikut:



Gambar 3.2. Arsitektur Jaringan yang didapat dari Hasil Eksperimen Neural Network

Pada gambar 3.2. terlihat 10 buah hidden layer dan 1 buah Threshold untuk mendapatkan nilai akurasi sebesar 83.62% pada algoritma Neural Network dengan 3 buah output yang akan dihasilkan dan 14 buah input yang didapat dari atribut age, workclass, fnlwgt, education, education-num, marital-status, occupation, relationship, race, sex, capital-gain, capital-loss, hours-per-week, native-country dan 1 buah Threshold. Nilai dari setiap Node sebagai berikut:

Output
Class ≤50K (Sigmoid)
Node 1: 2.202

```

if workclass > 3.500 and fnlwgt ≤
234129 then >50K (1 / 6 / 0)
if fnlwgt ≤ 343569.500 and fnlwgt >
126570 and fnlwgt ≤ 210524.500 then
≤50K (6 / 2 / 0)
if occupation ≤ 6.500 then >50K
(11 / 15 / 0)
else ≤50K (3 / 0 / 0)

```

correct: 5586 out of 6447 training examples.

Sedangkan pada Neural Network Parameter penunjang seperti penentuan training cycle, learning rate, dan momentum sangat berpengaruh terhadap akurasi yang akan dihasilkan pada eksperimen yang dilakukan. Setelah itu akan muncul nilai akurasi, precision, recall, AUC (Optimistic), AUC, AUC (Pessimistic) yang didapat dari hasil eksperimen yang telah dilakukan.

Node 2: 1.918
Node 3: 5.125
Node 4: 8.848
Node 5: 3.301
Node 6: 1.703
Node 7: -4.852
Node 8: 2.791
Node 9: 5.925
Node 10: 3.626
Threshold: -3.293

Class >50K (Sigmoid)

Node 1: -2.202
Node 2: -1.918
Node 3: -5.126
Node 4: -8.849
Node 5: -3.301
Node 6: -1.703
Node 7: 4.853
Node 8: -2.791
Node 9: -5.925
Node 10: -3.626
Threshold: 3.293

Class ≥50 (Sigmoid)

Node 1: -0.929
Node 2: -1.012
Node 3: -0.978
Node 4: -0.935
Node 5: -1.827
Node 6: -1.184
Node 7: -1.146
Node 8: -1.013
Node 9: -0.962
Node 10: -0.899
Threshold: -5.399

	true <=50K	true >50K	true >=50	class precision
pred <=50K	4470	679	1	86.80%
pred >50K	376	922	0	71.03%
pred >=50	0	0	0	0.00%
class recall	92.24%	57.59%	0.00%	

Gambar 3.4 Hasil Akurasi Neural Network

Nilai akurasi yang didapatkan dari hasil prediksi perhitungan menggunakan algoritma Neural Network terhadap data pendapatan rumah tangga sesuai dengan gambar 3.4.. Dengan perhitungan akurasi sebagai berikut:

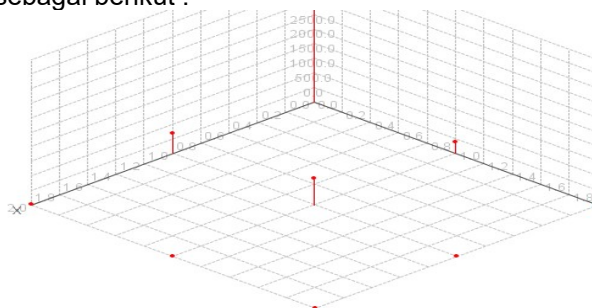
$$Accuracy = \frac{4470 + 922}{4470 + 679 + 376 + 922} = 0.83$$

$$Precision = \frac{376 + 922}{922} = 0.71$$

$$Recall = \frac{679 + 922}{922} = 0.57$$

$$F - Measure = \frac{2 * 0.57 * 0.71}{0.57 + 0.71} = 0.63$$

Selanjutnya akan didapatkan hasil dari confusion matrix yaitu pada gambar 3.5. sebagai berikut :



Gambar 3.5 Hasil Confusion matrix pada akurasi 83.62% dengan Rapid Miner

Hasil Confusion Matrix pada gambar 3.5 dari nilai akurasi Neural network sebesar 83.62% menunjukkan bahwa nilai output berupa

Referensi

F. Nurwanto, I. Ardiyanto and S. Wibirama, "Light sport exercise detection based on smartwatch and smartphone using k-Nearest Neighbor and Dynamic Time

nilai threshold untuk pendapatan kurang dari sama dengan \$50.000 sebesar -3.293, dan untuk output pendapatan yang lebih besar dari \$50.000 nilai threshold sebesar 3.293 sedangkan untuk output pendapatan yang lebih besar sama dengan \$50.000 nilai threshold sebesar -5.399.

PerformanceVector

PerformanceVector:
 accuracy: 83.62% +/- 1.49% (mikro: 83.62%)
 ConfusionMatrix:
 True: <=50K >50K >=50
 <=50K: 4470 679 1
 >50K: 376 922 0
 >=50: 0 0 0
 kappa: 0.530 +/- 0.050 (mikro: 0.532)
 ConfusionMatrix:
 True: <=50K >50K >=50
 <=50K: 4470 679 1
 >50K: 376 922 0
 >=50: 0 0 0

Gambar 3.6. hasil Performance Vector Neural Network pada data pendapatan rumah tangga dengan Rapid Miner

Gambar 3.6. menunjukkan performance vector dari data pendapatan rumah tangga menggunakan algoritma neural network yang menunjukkan nilai akurasi sebesar 83.62% serta penjabaran nilai untuk confusion matrix.

4. Kesimpulan dan Saran

Penelitian ini dilakukan komparasi antara K-nearest neighbour dengan algoritma neural network yang akan menghasilkan nilai akurasi pada data pendapatan rumah tangga di beberapa Negara di dunia sebesar 83.62% pada neural network dan akurasi sebesar 70,49% pada algoritma K-Nearest neighbor dengan demikian dapat dipastikan neural network dapat lebih tepat dalam memprediksi data pada pengolahan dataset pendapatan rumah tangga yang diambil dari UCI dataset pada Data sensus pendapatan yang dilakukan oleh Ronny Kohavi and Barry Becker pada sensu Beureu.

Warping algorithm," 2016 8th International Conference on Information Technology and Electrical Engineering (ICITEE), Yogyakarta, 2016, pp. 1-5.
 Gorunescu, Florin. (2011). *Data Mining: Concepts, Models and Techniques.*

- Verlag Berlin Heidelberg, Springer. Jerman.
- Guillet, Fabrice. Hamilton, Howard J. (2007). Quality Measures in Data Mining. Verlag Berlin Heidelberg, Springer. Jerman.
- Han, J. & Kamber, Micheline. (2007). Data Mining Concepts, Models and Techniques. Second Edition, Morgan Kaufmann Publisher. Elsevier.
- Heaton, Jeff. (2010). Programming Neural Networks With Encog 2 In Java. Heaton Research. Inc, USA.
- Larose, D. (2005). Discovering Knowledge in Data. New Jersey, John Wiley & Sons. Inc.
- Liao, Warren. T. & Triantaphyllou. Evangelos. (2007). Recent Advances in Data Mining of Enterprise Data: Algorithms and Applications. Series: Computer and Operation Research. 6. 190.
- Lim TS, Loh WY, Shih YS. (1999). A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. Kluwer Academic Publishers: Boston.
- Maimon, Oded & Rokach, Lior. (2010). Data Mining and Knowledge Discovery Handbook, Springer, New York.
- Mihov, Valentin. (2015). Adult Income Data Set Analysis with IPython. Sofia University, Bulgaria.
- Min, Hui Tsai, et al. (2010). Profiling U.S. Household Income. https://faculty.biu.ac.il/~yahavi1/Projects/CP2010T1_rep.pdf.
- Myatt, Glenn J. (2007). Making sense of data : A Practical Guide to Exploratory data analysis and Data Mining. John Wiley & Sons Inc, New Jersey.
- Priyanti, Evy. (2017). Implementasi Neural Network pada prediksi pendapatan rumah tangga. AMIK BSI. Swabumi.
- Shukla, Anupam. Tiwari, Ritu. & Kala, Rahul. (2010). Real Life Application of Soft Computing. New York: Taylor and Francis Groups, LLC.
- Siang, Jong Jek (2009). Jaringan Syaraf Tiruan dan Pemrogramannya menggunakan MATLAB. Penerbit Andi. Yogyakarta.
- Singh Asmita, Malka N. Halgamuge, Rajasekaran Lakshmanan. (2017). Impact of Different Data Types on Classifier Performance of Random Forest, Naïve Bayes, and K-Nearest Neighbors Algorithms. School of Computing and Mathematics Charles Sturt University Melbourne, Australia. (IJACSA) International Journal of Advanced Computer Science and Applications
- S. S. Nikam. A Comparative Study of Classification Techniques in Data Mining Algorithms. Orient. J. Comp. Sci. and Technol; 8(1), April 2015
- Topiwalla, Mohammed. (2013). Machine Learning on UCI Adult data set using various classifier algorithms and scaling up the accuracy using extreme gradient boosting. University of SP Jain School of Global Management.
- Vercellis, C. (2009). Business Intelligence: Data Mining and Optimization for Decision Making. Wiley.
- Witten, I. Frank, E., & Hall. (2011). Data Mining: Practical Machine Learning and tools. Morgan Kaufmann Publisher, Burlington.