

PREDIKSI CUSTOMER CHURN MENGGUNAKAN ARTIFICIAL NEURAL NETWORK (ANN)

Masayu Franstika¹⁾, Mochammad Aditya Putra Suhendar²⁾, Adisya Ridia Nurahma³⁾, Junpito Salim⁴⁾

Program Studi Sains Data, Jurusan Sains, Institut Teknologi Sumatera

¹masayu.120450016@student.itera.ac.id

²mochammad.120450058@student.itera.ac.id

³adisya.120450056@student.itera.ac.id

⁴junpito.120450086@student.itera.ac.id

Abstrak— Pertumbuhan pasar yang cepat di setiap sektor memimpin ke basis pelanggan yang lebih besar untuk penyedia layanan yang lebih banyak pesaing dengan model bisnis baru dan inovatif, serta layanan yang lebih baik meningkatkan biaya akuisisi pelanggan. Customer churn adalah sebuah kecenderungan pelanggan untuk pergi meninggalkan sebuah layanan atau perusahaan. Disini kami menggunakan salah satu metode yang dapat digunakan yaitu Artificial Neural Network (ANN). ANN merupakan teknik jaringan saraf tiruan yang mengelola informasi yang terinspirasi dari cara kerja sistem saraf biologis, khususnya pemrosesan informasi pada sel otak manusia. Hasil yang diperoleh dengan akurasi klasifikasi mencapai 90%. Penelitian ini berhasil dan cocok untuk diterapkan dalam metode ini karena dilihat dari tingkat keakurasiannya yang cukup tinggi.

Kata kunci— ANN, Customer Churn, Fungsi Aktivasi,

I. PENDAHULUAN

A. LATAR BELAKANG

Churn merupakan suatu tingkatan atau persentase pelanggan yang menghentikan kegiatan berlangganan yang mereka lakukan dengan suatu bisnis dalam waktu atau periode tertentu yang sangat erat kaitannya dengan keuntungan dan pertumbuhan pada sebuah perusahaan. Dalam industri telekomunikasi kehilangan pelanggan merupakan isu penting yang menjadi tantangan utama oleh perusahaan. Jika perusahaan kehilangan pelanggan maka perusahaan tersebut akan mendapat kerugian yang besar karena pada kenyataannya untuk mendapatkan pelanggan baru membutuhkan biaya 5 kali lebih mahal daripada mempertahankan

pelanggan lama dan membuat pelanggan baru juga 16 kali lebih mahal. Jadi, demi mengurangi kerugian besar yang disebabkan pencarian pelanggan baru terus menerus maka customer churn ini harus dihentikan agar dana yang ada bisa digunakan lebih baik untuk mempertahankan pelanggan yang sudah ada karena sumber revenue utama dari bisnis telekomunikasi ini ada pelanggan tersebut.

Artificial Neural Network (ANN) adalah suatu jaringan yang memodelkan sistem saraf otak manusia yang disebut neuron. Neural Network dan algoritma komputer konvensional tidaklah saling bersaing tetapi saling melengkapi. Beberapa tugas atau masalah lebih cocok diselesaikan dengan pendekatan algoritmik seperti halnya operasi aritmatika, di sisi lain ada tugas-tugas yang lebih cocok untuk jaringan saraf.

Oleh karena itu sangatlah penting bagi industri telekomunikasi untuk memanfaatkan metode analisis canggih guna memahami perilaku pelanggan sehingga bisa diprediksi apakah mereka akan pergi atau tidak. Dalam hal ini kita akan menggunakan metode Artificial Neural Network (ANN) yakni suatu metode dalam teknik machine learning yang termasuk ke dalam materi Artificial Intelligence. Adapun Hasil dari teknik prediksi dapat memprediksi churn telekomunikasi dengan akurasi 0.9 atau 90% yang menunjukkan faktor churn.

B. RUMUSAN MASALAH

1. Variabel apa yang mempengaruhi churn seorang customer?
2. Berapakah hasil akurasi yang diperoleh dan apakah metode ANN cocok diterapkan dalam mengatasi churn dalam sebuah industri?

C. DATA

Sumber dataset pelanggan perusahaan telekomunikasi diperoleh dari kaggle.com sebuah website yang menyediakan dataset-dataset open source. Data ini menyimpan berbagai atribut terkait layanan yang digunakan dan dicatat pada setiap pelanggan. Data yang digunakan sudah dalam keadaan bersih dimana variabel unique text sudah diubah kedalam format angka dengan keterangan kolom sebagai berikut:

1. Churn = Keterangan customer churn.
2. AccountWeeks = Jangka waktu customer berlangganan.
3. ContractRenewal = Customer melakukan pembaruan kontrak.
4. DataPlan = Customer memilih suatu rencana data.
5. DataUsage = Besaran customer menghabiskan data.
6. CustServCalls = Customer melakukan panggilan bantuan.
7. MonthlyCharge = Besaran tagihan bulanan.
8. OverageFee = Jumlah uang ekstra yang harus dibayarkan untuk menggunakan sesuatu lebih dari yang diharapkan atau disepakati.
9. RoamMins = Jumlah menit skenario dimana customer masih dapat melakukan dan menerima panggilan suara, mengirim dan menerima data, atau mengakses layanan lain saat berada di luar jangkauan jaringan utama.

	Churn	AccountWeeks	ContractRenewal	DataPlan	DataUsage	CustServCalls	DayMins	DayCalls	MonthlyCharge	OverageFee	RoamMins
0	0	128	1	1	2.7	1	265.1	110	89.0	9.87	10.0
1	0	107	1	1	3.7	1	161.6	123	82.0	9.78	13.7
2	0	137	1	0	0.0	0	243.4	114	52.0	6.06	12.2
3	0	84	0	0	0.0	2	299.4	71	57.0	3.10	6.6
4	0	75	0	0	0.0	3	166.7	113	41.0	7.42	10.1

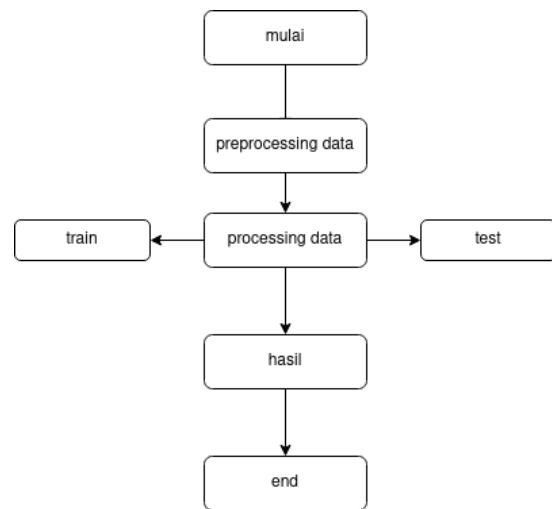
Gambar 1. Tabel Data

II. METODE

A GAMBARAN UMUM

Dalam penelitian ini data pelanggan perusahaan telekomunikasi dikumpulkan kemudian dilakukan processing pada dataset tersebut menggunakan algoritma Artificial Neural Network. kemudian hasil operasi menggunakan algoritma

Artificial Neural Network tersebut akan dievaluasi menggunakan evaluasi matriks.



Gambar 2. Flowchart

Metode yang digunakan dalam penelitian ini adalah Artificial Neural Network (ANN). Artificial Neural Network Artificial (ANN) merupakan teknik jaringan saraf tiruan yang mengelola informasi yang terinspirasi dari cara kerja sistem saraf biologis, khususnya pemrosesan informasi pada sel otak manusia. Kumpulan data yang kami gunakan berasal dari website Kaggle tentang Customer Churn yang terdapat 3333 baris dan 11 kolom didalamnya.

Pemodelan jaringan pada ANN ada 3 macam, yaitu : single layer, multi layer, dan recurrent network. Sebelum dilakukan pemodelan, data memasuki tahapan preprocessing terlebih dahulu, adapun sebagai berikut hasil dari proses preprocessing dan informasi yang didapatkan dari data sebelum melakukan ke tahap pemodelan berikut adalah preprocessing data yang kita lakukan :

A. Informasi Tipe Data

Data yang kami gunakan memberikan informasi berupa tipe data apa saja yang ada dalam data tersebut, dengan menggunakan fungsi dtypes dari pandas kita bisa mendapat informasi jenis data apa saja dari setiap variabel yang ada. Pada data tersebut sebagian merupakan data integer dan sebagian berupa float.

B. Pengecekan Missing Value

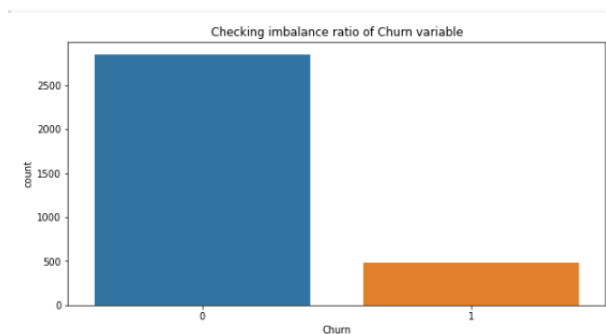
Pengecekan ini untuk melihat data terdapat missing value. Informasi yang didapat melalui data, menunjukkan bahwa terdapat beberapa data berisikan missing value. Pengecekan missing value ini menggunakan fungsi `isnull` pada `pandas`.

```
Churn          0
AccountWeeks   0
ContractRenewal 0
DataPlan       0
DataUsage      0
CustServCalls  0
DayMins        0
DayCalls       0
MonthlyCharge  0
OverageFee     0
RoamMins       0
dtype: int64
```

Gambar 3. Pengecekan Missing Value

C. Scaling nilai dari data numerical

Scaling data ini dilakukan untuk menormalisasikan range pada fitur—fitur data.



Gambar 4. Balance test data target

D. Fungsi Aktivasi

Fungsi aktivasi merupakan fungsi yang digunakan pada jaringan saraf untuk mengaktifkan atau tidak mengaktifkan neuron. Karakteristik yang harus dimiliki oleh fungsi aktivasi jaringan perambatan balik antara lain harus kontinu, terdiferensiasi, dan tidak menurun secara monotonis. Fungsi aktivasi yang di analisis adalah sigmoid biner. Fungsi aktivasi sigmoid biner memiliki nilai pada range 0 sampai 1. Ditunjukkan pada gambar 2, didefinisikan sebagai berikut :

$$a = \frac{1}{(1+e^{-n})}$$

E. Metrics Evaluation

Confusion Matrix adalah pengukuran performa untuk masalah klasifikasi *machine learning* dimana keluaran dapat berupa dua kelas atau lebih. *Confusion Matrix* adalah tabel dengan 4 kombinasi berbeda dari nilai prediksi dan nilai aktual. Ada empat istilah yang merupakan representasi hasil proses klasifikasi pada *confusion matrix* yaitu True Positif, True Negatif, False Positif, dan False Negatif.

1. True Positive (TP) memprediksi positif dan itu benar
2. True Negative (TN) memprediksi negatif dan itu benar
3. False Positive (FP) (Kesalahan Tipe 1) memprediksi positif dan itu salah.
4. False Negative (FN): (Kesalahan Tipe 2) memprediksi negatif dan itu salah.

maka dapat dihitung nilai *accuracy*, *precision*, *recall* dan F-1 score :

$$Accuracy = (TP+TN) / (TP+FP+FN+TN)$$

$$Precision = (TP) / (TP + FP)$$

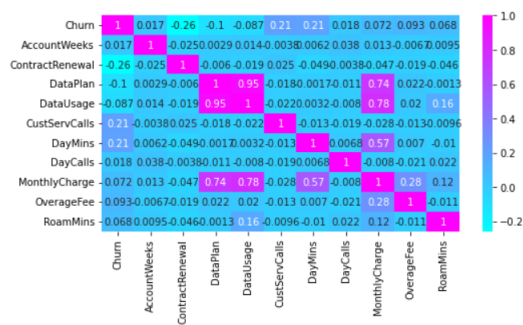
$$Recall = TP / (TP + FN)$$

$$F-1 \text{ Score} = (2 * Recall * Precision) / (Recall + Precision)$$

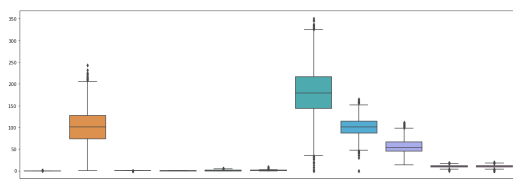
III. HASIL DAN PEMBAHASAN

A. Hasil data cleaning dan EDA

Hasil penelitian yang telah diujikan pada tahap ini menggunakan metode klasifikasi ANN. Sebelum masuk ke tahap pengujian dilakukan tahap preprocessing. Dimulai dengan pemeriksaan struktur data, dimana semua kolom sudah memiliki tipe yang sesuai. Selanjutnya memeriksa data yang hilang, lalu setelah melakukan data processing dilakukan juga Exploratory Data Analysis (EDA). Pertama melihat korelasi yang ada pada setiap variabel.



Gambar 5. Korelasi antar variabel pada dataset



Gambar 6. Boxplot antar variabel pada dataset

setelah melakukan preprocessing dan Exploratory Data Analysis (EDA) maka dataset yang akan kita jadikan model akan menjadi seperti berikut.

	AccountWeeks	ContractRenewal	DataPlan	DataUsage	CustServCalls	DayMins	DayCalls	MonthlyCharge	OverageFee	RoamMins
1286	0.438017	1	1.0	0.375926	0.111111	0.325827	0.800000	0.445015	0.549203	0.375
1765	0.454545	1	0.0	0.000000	0.444444	0.377993	0.757576	0.287770	0.608026	0.635
2961	0.400826	1	1.0	0.400000	0.666667	0.478905	0.490909	0.520041	0.448598	0.400
1271	0.330579	0	1.0	0.394444	0.111111	0.675884	0.460606	0.722508	0.726223	0.395
869	0.520661	0	0.0	0.048148	0.222222	0.515678	0.690909	0.386434	0.576141	0.370
437	0.409091	1	0.0	0.059259	0.000000	0.792474	0.460606	0.536485	0.485981	0.415
2197	0.545455	1	0.0	0.000000	0.222222	0.616306	0.406061	0.431655	0.610775	0.155
519	0.165289	1	0.0	0.038889	0.333333	0.578392	0.587879	0.371017	0.422760	0.450
2669	0.595041	1	0.0	0.000000	0.222222	0.368871	0.587879	0.246660	0.509621	0.055
2078	0.409091	1	0.0	0.000000	0.111111	0.373717	0.654545	0.246660	0.484332	0.215

Gambar 7. Tabel Data setelah preprocessing dan EDA

B. Hasil Modeling

Setelah dilakukan data preprocessing maka kita melakukan modeling data pada Artificial Neural Network yang mana pemodelan yang kita lakukan memiliki fitur :

Model: "sequential"		
Layer (type)	Output Shape	Param
dense (Dense)	(None, 10)	110
dense_1 (Dense)	(None, 5)	55
dense_2 (Dense)	(None, 1)	6
Total params: 171		
Trainable params: 171		
Non-trainable params: 0		

Gambar 8. Model ANN

a. 10 layer input

dengan sample weight pada tiap kali algoritma melakukan forward propagation sebagai berikut

```
1 w,b = model.layers[0].get_weights()
2 print(w)
```

```
[ [ 0.43588638 0.08314252 -0.09074402 0.33463043 0.00267327 0.24398774
-0.50781566 -0.2626445 -0.49180514 0.18333 ]
[ -0.49858862 0.08231109 0.35839796 0.00115585 -0.38283634 -0.3895032
-0.5319647 0.5233276 0.48014534 -0.31508455 ]
[ -0.3854712 0.09018326 -0.07743797 -0.2216644 -0.08078322 -0.13368437
0.03294235 -0.44784933 -0.02267623 0.24893767 ]
[ -0.29321784 0.22749758 0.42060333 0.53990977 -0.34904152 -0.54759955
0.28155404 -0.08885898 0.09207886 0.27442986 ]
[ 0.3857513 0.0858115 -0.34896946 -0.32525533 0.50989723 0.49950767
-0.4950437 -0.03048897 -0.07584023 0.45512176 ]
[ -0.42415753 -0.01109284 0.1754505 0.24484873 0.11710894 0.27860332
0.1221354 -0.29739806 0.3930549 0.27432787 ]
[ 0.13428897 -0.2732862 -0.15000191 0.19900686 0.15587169 0.4065678
0.18134195 0.38143396 0.02353483 -0.33289784 ]
[ 0.37204605 -0.1899454 -0.46248502 -0.16776946 -0.43112803 0.09596086
0.2209301 0.52622914 0.449139 0.17654938 ]
[ 0.3971141 -0.31953222 -0.33555815 0.09088951 -0.27612644 -0.00748748
-0.2853826 -0.09065518 0.07135209 -0.10225385 ]
[ 0.297809 0.22960687 0.17947859 0.46406543 0.4513082 0.02385569
-0.05304453 -0.00876892 -0.01106846 0.14890212 ] ]
```

Gambar 9. 10 layer input

b. 5 hidden layer

```
1 w,b = model.layers[1].get_weights()
2 print(w)
```

```
[ [ -0.4959336 -0.39264423 -0.42337114 -0.38691017 0.10372692 ]
[ 0.3251524 0.29910338 -0.5676598 -0.62822366 -0.5792678 ]
[ -0.3641184 0.14030671 -0.4702376 -0.38589233 -0.2762614 ]
[ 0.430924 -0.22629035 0.03233808 0.5420887 -0.33772412 ]
[ 0.54851025 -0.05171889 0.21100408 -0.49148214 0.08014292 ]
[ 0.06684017 -0.08937436 -0.44885993 0.47128314 0.52681774 ]
[ 0.09544921 0.00983912 -0.14390531 -0.33014426 -0.59067154 ]
[ -0.23180634 -0.2714773 -0.0340898 0.6171307 0.34465408 ]
[ 0.4277584 0.31844407 -0.45083392 -0.3829108 0.54477876 ]
[ 0.36863 0.24998039 -0.6205236 -0.23261082 0.4128533 ] ]
```

Gambar 10. 5 hidden layer

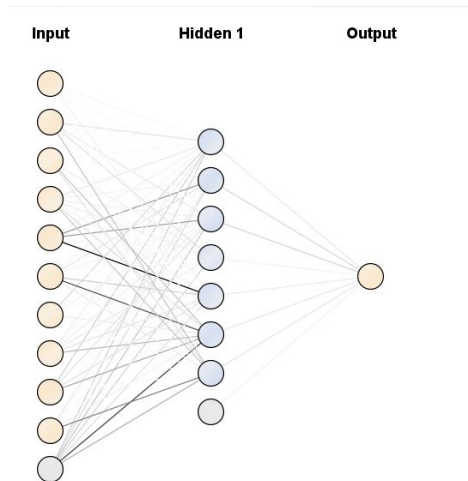
c. 1 output layer

```
[ ] w,b = model.layers[2].get_weights()
print(w)
```

```
[ [-0.51382923 ]
[ -0.1829629 ]
[ -0.42424798 ]
[ -0.4597423 ]
[ 0.8970406 ] ]
```

Gambar 11. 1 output layer

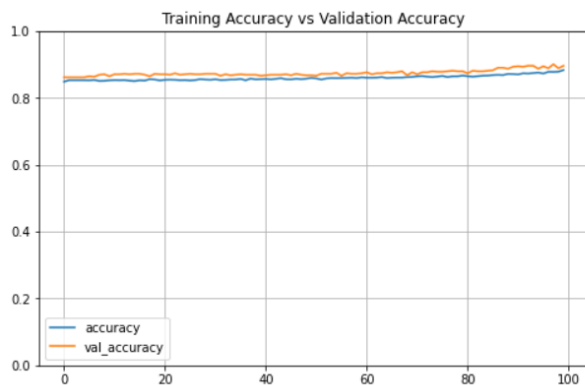
Pada layer ini akan dihasilkan 1 output dari node Hidden 1 yang dimana node Hidden 1 berasal dari beberapa input.



Gambar 12. Model ANN single layer

C. Hasil Training

Dengan menggunakan sebanyak 100 epoch yakni kemampuan dari neural network untuk melakukan backward dan forward propagation maka di dapat hasil learning rate sebagai berikut



Gambar 13. curva training accuracy dan validation accuracy



Gambar 14. curva training loss dan validation loss

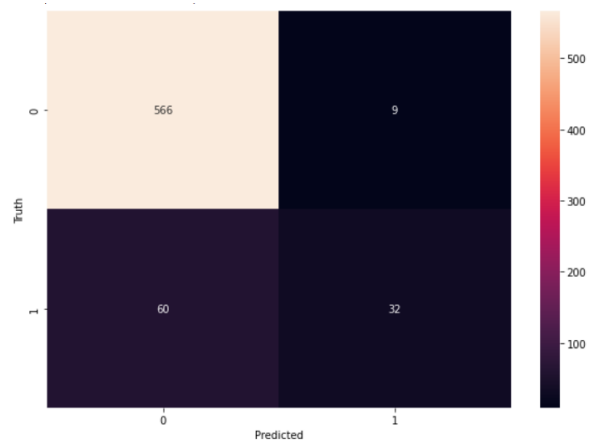
Dimana dalam chart di tunjukkan bahwa curva accuracy dan val_accuracy tidak saling jauh dan begitu juga pada curva training loss dan val_loss berarti modeling yang kita lakukan well fitting.

D. Hasil evaluasi model

	precision	recall	f1-score	support
0	0.90	0.98	0.94	575
1	0.78	0.35	0.48	92
accuracy			0.90	667
macro avg	0.84	0.67	0.71	667
weighted avg	0.89	0.90	0.88	667

Gambar 13. Tabel hasil prediksi

dengan sample weight pada tiap kali algoritma melakukan forward propagation sebagai berikut



Gambar 15. Confusion Matrix

IV. KESIMPULAN

Berdasarkan penelitian yang dilakukan menggunakan metode ANN dapat ditarik kesimpulan dari sebagai berikut :

1. Penerapan metode ANN untuk Model dibuat dengan memperhitungkan semua variabel yang ada seperti single layer, multi layer, dan recurrent network. Model ANN dapat membantu perencanaan membuat keputusan berdasarkan keputusan awal pada proses perancangan.
2. Hasil yang diperoleh dengan akurasi klasifikasi mencapai 90%. Penelitian ini berhasil dan cocok untuk diterapkan dalam metode ini karena dilihat dari tingkat keakurasiannya yang cukup tinggi.
3. Jenis variabel yang memiliki korelasi paling tinggi

REFERENSI

[1]

https://www.researchgate.net/profile/Mohammad-Alshammari-2/publication/336266496_Design_and_Learning_Effectiveness_Evaluation_of_Gamification_in_e-Learning_Systems/links/5daee953a6fdccc99d92b461/Design-and-Learning-Effectiveness-Evaluation-of-Gamification-in-e-Learning-Systems.pdf#page=142

[2]

[ANALISIS FUNGSI AKTIVASI SIGMOID BIPOLAR DAN SIGMOID BIPOLAR DALAM ALGORITMA BACKPROPAGATION PADA PREDIKSI KEMAMPUAN SISWA - CORE](#)

LAMPIRAN

BERIKUT KAMI LAMPIRKAN KODE YANG TELAH DIBUAT:

https://colab.research.google.com/drive/1XuJCMCX-BBD_iZOE9mKE_cACJq8k7LH8?usp=sharing

BERIKUT KAMI LAMPIRKAN DATASET YANG DIGUNAKAN:

<https://www.kaggle.com/datasets/barun2104/telecom-churn?resource=download>

BERIKUT KAMI LAMPIRKAN HASIL PERHITUNGAN MANUAL:

AccountWeeks	ContractRenewal	DataPlan	DataUsage	CustServCalls	DayMins	DayCalls	MonthlyChrg	OverageFee	RoomMins	Churn
0,52479339	1	1	0,5	0,111	0,755	0,667	0,77	0,54	0,2	0

Input	10									
Hidden	5									
Output	1									

Bobot random										
Bobot	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10
V1	0,5	-0,4	-1	-2,5	0,03	-1,2	-4,3	-1,3	-4,2	-0,9
V2	-0,2	-0,8	-4,2	-5,5	-0,7	-0,7	-3,6	-1	0,6	-1,9
V3	-3,5	-1,4	-1,6	-0,6	-4,3	0,4	0,7	-0,4	-3,5	0,02
V4	0,9	0,08	0,4	-0,5	-0,5	-0,2	0,3	-2,2	0,1	-0,5
V5	-2,4	0,5	-2,2	-2,5	0,2	0,2	0,2	-0,7	-0,9	-0,7

Bobot	w1	w2	w3	w4	w5
Y	-1,5	-1,3	-0,5	-2,6	-0,7

Bias	1	2	3	4	5	6	7	8	9	10
Vj	0,73	0,84	0,45	0,93	0,21	-1,25	-0,78	-2,6	-0,78	0,93

Bias	1
wj	0,83

Pelatihan epoch ke-1										
data ke-1										
learning rate	0,1									

1. Operasi pada Hidden Layer						2. Operasi pada Output Layer				
PENJUMLAHAN BOBOT						y_in	-0,3962255			
z_in 1	-8,877373305									
z_in 2	-11,39835868					Yk	0,40221954			
z_in 3	-6,064383475									
z_in 4	-0,124085949									
z_in 5	-4,857904136									
PENGAKTIFAN BOBOT										
Z1	0,000139491									
Z2	1,121375-05									
Z3	0,002318801									
Z4	0,469018256									
Z5	0,007706887									
B. Tahap Backpropagation										
ERROR = TARGET - Y	-0,402219541									
KUADRAT ERROR	0,161780559									
Ek	-0,096709257									
Koreksi bobot pada unit k						Perbaruan bobot dan Prinsipap				
Δw0	-0,009670926					δ_in1	0,14506389	δ1	2,0232E-05	
Δw1	-1,349E-06					δ_in2	0,12572203	δ2	1,4098E-06	
Δw2	-1,08447E-07					δ_in3	0,04835463	δ3	0,00011186	
Δw3	-2,2425E-05					δ_in4	0,25144407	δ4	0,06261966	
Δw4	-0,004535841					δ_in5	0,06769648	δ5	0,00051771	
Δw5	-7,45327E-05									
Perbaruan Bobot Tiap Unit Keluaran										

TABEL PERUBAHAN PADA SETIAP UNIT KELUARAN (OUTPUT)										
	ΔV1	ΔV2	ΔV3	ΔV4	ΔV5					
1	1,06177E-06	7,3985E-08	5,8708E-06	0,00286239	2,7169E-05					
2	2,02322E-05	1,4098E-07	1,1186E-05	0,006261966	5,1771E-05					
3	2,02322E-06	1,4098E-07	1,1186E-05	0,006261966	5,1771E-05					
4	1,01161E-06	7,049E-08	5,8932E-06	0,003130983	2,5885E-05					
5	2,24578E-07	1,5649E-08	1,2417E-06	0,000895078	5,7466E-06					
6	1,51735E-06	1,0644E-07	8,4458E-06	0,004727785	3,5007E-05					
7	1,34949E-06	9,4034E-08	7,4614E-06	0,004176732	3,4531E-05					
8	1,55788E-06	1,0855E-07	8,6136E-06	0,004821714	3,8864E-05					
9	1,09254E-06	7,6129E-08	6,407E-06	0,00381462	2,7956E-05					
10	4,04849E-07	2,8196E-08	2,2378E-06	0,001252393	1,0354E-05					
TABEL PERUBAHAN BOBOT INPUT MENUJU HIDDEN										
	ΔV1									
1	2,02322E-06									
2	1,4098E-07									
3	1,1186E-05									
4	0,006261966									
5	5,17708E-05									
TABEL PERUBAHAN BOBOT INPUT KE HIDDEN										
	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10
V1	0,500001062	-0,399998	-0,999998	-2,49999899	0,03000022	-1,1999985	-4,3	-1,2999984	-4,2	-0,9
V2	-0,199999926	-0,7999999	-4,1999999	-6,59999993	-0,7	-0,6999999	-3,6	-0,9999999	0,6	-1,9
V3	-2,499994129	-1,3999988	-1,5999988	-0,59999441	-4,2999988	0,4000084	0,70001	-0,3999914	-3,5	0,02
V4	0,903286239	0,08626197	0,40626197	-0,49688902	-0,4993049	-0,1952722	0,30418	-2,1951783	0,10338	-0,4987
V5	-2,399972831	0,50005177	-2,1999942	-2,49997411	0,30000675	0,3000391	0,20005	-0,6999601	-0,9	-0,7

TABEL PERUBAHAN BOBOT HIDDEN KE OUTPUT					
	W1	W2	W3	W4	W5
Y	-1,500001349	-1,30000001	-0,5000224	-2,60453584	-0,7000745
	1	2	3	4	5
Vj0	0,730002023	0,84000014	0,45001119	0,926261966	0,21005177
bias					
wk	0,820329074				