

PREDIKSI DATA TIME SERIES JUMLAH PELANGGAN PSTN TELKOM DI INDONESIA BERBASISKAN GRAMMATICAL EVOLUTION

Dwi Tuti Supantari¹, Suyanto², Zakariah³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

PT. Telekomunikasi Indonesia, Tbk adalah perusahaan informasi dan komunikasi Indonesia yang menyediakan jasa dan jaringan telekomunikasi terbesar di Indonesia. Dimana core business dan penghasilan terbesarnya berasal dari penyedia layanan jaringan telepon PSTN (Public Switched Telephone Network) atau telepon kabel (fixed wire line). Namun, munculnya penerapan dan pengembangan teknologi telekomunikasi baru, yaitu GSM (Global Sistem for Mobile) dan CDMA (Code Division Multiple Access), mengakibatkan banyak konsumen telepon PSTN berpindah ke layanan telepon GSM atau CDMA dan akhirnya mengakibatkan jumlah pelanggan telepon PSTN bergejolak dengan tingkat ketidakpastian yang tinggi. Oleh karena itu, dibutuhkanlah suatu sistem yang dapat memprediksi jumlah pelanggan PSTN TELKOM. Pendekatan prediksi yang banyak dipergunakan adalah metode kuantitatif dengan subkategori metode time series. Proses prediksi akan dilakukan dengan menggunakan teknik komputasi yang berbasiskan evolusi dan genetika atau yang lebih dikenal dengan nama Evolutionary Computation (EC). Salah satu algoritma EC yang dapat digunakan untuk mencari solusi pada masalah prediksi adalah Grammatical Evolution(GE). GE mampu menghasilkan solusi untuk memprediksi pola data yang bersifat nonlinear dengan lebih tepat karena menggunakan definisi grammar Backus Naur Form (BNF) dalam pemetaan genotype ke phenotype. Solusi yang dicari pada masalah prediksi data time series ini adalah fungsi yang memiliki pola yang paling mendekati karakteristik dari data historis jumlah pelanggan PSTN TELKOM. Kemudian fungsi tersebut digunakan untuk melakukan prediksi dalam n periode berikutnya. Akurasi prediksi dihitung dengan menggunakan rumus Mean Absolute Percentage Error(MAPE). Sistem GE akan mencari fungsi prediksi yang paling akurat melalui proses evolusi yang menggunakan parameterparameter evolusi. Untuk mendapatkan sistem GE yang mampu menghasilkan fungsi prediksi yang optimal tersebut maka dilakukan pengujian terhadap definisi BNF, kombinasi parameter pada GE dan perhitungan akurasi prediksi yang dihasilkan. Berdasarkan observasi yang telah dilakukan dapat diketahui bahwa prediksi dengan menggunakan sistem yang berbasiskan Grammatical Evolution mampu menghasilkan solusi dalam bentuk fungsi prediksi dengan tingkat adaptasi yang lebih tinggi terhadap data prediksi sebenarnya dibandingkan dengan metode prediksi konvensional seperti Linear Regression with time(LR) dan Moving Average(MA), karena adanya penggunaan definisi BNF yang menyebabkan sifat non linearitas. Selain itu, pendefinisian MAPE merupakan komponen yang sangat penting dalam sistem GE. Hal ini disebabkan karena grammar yang digunakan pada BNF sangat menentukan kemungkinan pola solusi yang dibangun.

Kata Kunci : prediksi, data time series, Evolutionary Computation (EC), Grammatical Evolution (GE)

Abstract

PT Telekomunikasi Indonesia, Tbk. Is Information and Communication Technology (ICT) company that offers the biggest service and telecommunication network in Indonesia. Its biggest core business and revenue come from offering PSTN (Public Switched Telephone Network) service or cable telephone (fixed wireline). But, new telecommunication technology implementation and development, GSM (Global System for Mobile) and CDMA (Code Division Multiple Access), make many PSTN subscribers switch to GSM and CDMA services and finally make the number of PSTN subscribers fluctuates in high uncertainly. So, it is needed a system that can forecast the number of PSTN subscribers. Forecast approach that many people use is quantitative method with time series method subcategory.

Forecasting process will be done by using computation technique based on evolution and genetics or called Evolutionary Computation (EC). One of EC algorithms that can used for finding solution in forecasting is Grammatical Evolution (GE). GE can give solution to forecast nonlinear data pattern more accurately because of using Backus Naur Form (BNF) grammar definition in mapping genotype to phenotype. Solution that will be finds in this time series data forecasting is the best function which has closest pattern with historic data characteristic of PSTN subscriber. Then, this selected function will be used to forecast the number of PSTN subscribers in next n-periods. Forecast accurate will be count by using Mean Absolute Percentage Error (MAPE) formula. GE system will find the most accurate function by evolution process that uses evolution parameters. For getting GE system that can give the most optimal forecasting function, testing of BNF, GE parameters combination, and forecasting result accurate calculation, definition will be done.

Based on observation that has be done, it is known that forecasting with system based Grammatical Evolution can generate function which has the most adaptive of others konvensional forecasting method such as Linear Regression(LR) and Moving Average(MA), because it's uses BNF definition with non linearitas behaviour. Beside that, MAPE defining is the most important component in GE system because grammar that is used in BNF determines the possibility of development solution pattern.

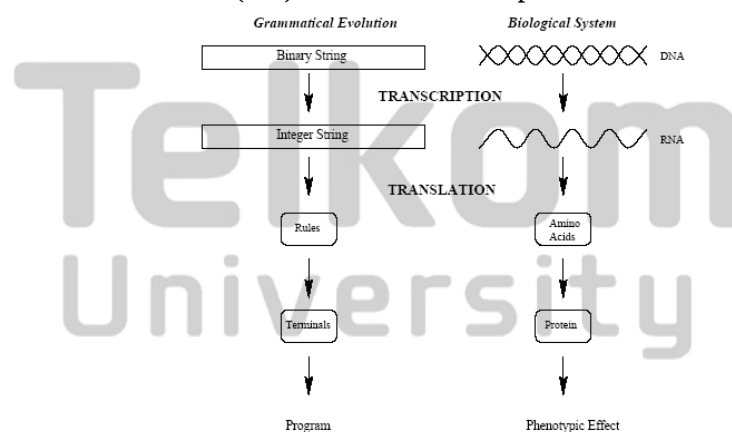
Keywords : forecasting, time series data, Evolutionary Computation (EC), Grammatical Evolution (GE)

1. Pendahuluan

1.1 Latar Belakang

PT. Telekomunikasi Indonesia, Tbk adalah perusahaan informasi dan komunikasi Indonesia yang menyediakan jasa dan jaringan telekomunikasi terbesar di Indonesia. Dimana *core business* dan penghasilan terbesarnya berasal dari penyedia layanan jaringan telepon PSTN (*Public Switched Telephone Network*) atau telepon kabel (*fixed wire line*) yang sukses sejak dimulai tahun 1974. Namun, seiring dengan perkembangan teknologi yang semakin meningkat pesat dan gaya hidup masyarakat atau pelanggan yang cenderung serba *mobile* serta kebutuhan komunikasi jarak jauh yang semakin besar, mendorong munculnya penerapan dan pengembangan teknologi telekomunikasi baru, yaitu GSM (*Global Sistem for Mobile*) dan CDMA (*Code Division Multiple Access*) yang tidak hanya dilakukan oleh TELKOM saja melainkan juga provider telekomunikasi lainnya. Dengan kelebihan dan kemudahan yang diberikan oleh teknologi telepon GSM dan CDMA, menyebabkan banyak konsumen telepon PSTN berpindah ke layanan telepon GSM atau CDMA dan pada akhirnya mengakibatkan jumlah pelanggan telepon PSTN bergejolak dengan tingkat ketidakpastian yang tinggi. Untuk menanggulangi tantangan tersebut, dibutuhkanlah suatu perencanaan strategis. Dan aspek utama dari pengambilan keputusan dalam menetapkan perencanaan adalah kemampuan untuk meramalkan atau memprediksi situasi yang akan datang[3]. Adapun pendekatan prediksi yang banyak dipergunakan adalah metode kuantitatif dengan subkategori metode *time series*[3].

Pada era komputerisasi sekarang ini, proses prediksi sudah dapat dilakukan dengan menggunakan teknik komputasi yang berbasiskan evolusi dan genetika atau yang lebih dikenal dengan nama *Evolutionary Computation* (EC). Ada beberapa algoritma pada *Evolutionary Computation* yang dapat digunakan dalam permasalahan prediksi. Namun, pada tugas akhir ini digunakan algoritma *Grammatical Evolution* (GE) untuk melakukan prediksi.



Gambar 1-1 Representasi Individu pada Grammatical Evolution

Salah satu kemampuan yang luar biasa dari *Grammatical Evolution* adalah kemampuan untuk menghasilkan suatu program/solusi otomatis. Seperti yang diilustrasikan pada Gambar1-1, sistem ini menggunakan proses pemetaan *genotype* ke *phenotype* untuk menghasilkan program/solusi [4]. Dan dalam proses pemetaan digunakan definisi *grammar Backus Naur Form* (BNF) yang bisa mewakili semua bahasa dan dikodekan dengan lebih mudah [9] sehingga solusi yang dibangun dapat lebih fleksibel. Atau dengan kata lain, GE mampu menghasilkan solusi untuk memprediksi pola data yang bersifat *non-linear* dengan lebih tepat. Selain itu, pada *Grammatical Evolution* juga terjadi proses evolusi yang akan menyeleksi individu-individu/solusi yang ada pada suatu generasi sehingga pada akhirnya tersisa/dihasilkan individu terpilih yang dapat bertahan hidup untuk menghasilkan dan mewariskan sifat-sifatnya pada generasi berikutnya. Dapat dikatakan pula bahwa individu terpilih tersebut merupakan individu/solusi yang paling tangguh/optimal pada generasinya.

1.2 Perumusan Masalah

Permasalahan yang dijadikan objek penelitian pada tugas akhir ini antara lain:

1. Bagaimana mendapatkan solusi dalam bentuk fungsi yang paling optimal untuk memprediksi jumlah pelanggan PSTN TELKOM di Indonesia
2. Bagaimana menerapkan *Grammatical Evolution* pada masalah prediksi jumlah pelanggan PSTN TELKOM di Indonesia sehingga menghasilkan prediksi yang akurat

Batasan masalah dalam tugas akhir ini, yaitu :

1. Penelitian menggunakan data historis, dimana prediksi dilakukan berdasarkan ekstrapolasi dari pola dan keterkaitan di masa lalu.
2. Data historis yang dipergunakan adalah data jumlah pelanggan PSTN Telkom di Indonesia dalam tujuh tahun terakhir (2002 - 2008) yang disajikan per bulan.
3. Proses penelitian hanya mencakup prediksi saja tidak sampai kepada pemberian solusi atau rekomendasi perencanaan.
4. Simulasi yang dibuat menggunakan program aplikasi Matlab 7.01.

1.3 Tujuan

Tujuan yang akan dicapai adalah:

1. Mengimplementasikan *Grammatical Evolution* pada kasus prediksi data *time series* jumlah pelanggan PSTN TELKOM di Indonesia untuk mendapatkan solusi dalam bentuk fungsi prediksi yang paling optimal.
2. Menganalisis pendefinisian BNF yang paling sesuai dan parameter – parameter evolusi terbaik, yaitu ukuran populasi, jumlah generasi, dan probabilitas *crossover* untuk dapat menghasilkan prediksi jumlah pelanggan PSTN TELKOM di Indonesia yang paling akurat serta menentukan maksimal jumlah periode yang dapat diprediksi dengan kesalahan prediksi yang masih bisa ditoleransi.

1.4 Metodologi Penyelesaian Masalah

Metode yang digunakan untuk menyelesaikan tugas akhir ini adalah :

1. Studi literatur
Mempelajari teori-teori dasar mengenai konsep prediksi dan *Grammatical Evolution* dari berbagai sumber.
2. Pengumpulan data-data penunjang tugas akhir
Mengumpulkan data penunjang yang dapat membantu proses prediksi, yaitu data historis jumlah pelanggan PSTN TELKOM di Indonesia dalam tujuh tahun terakhir.
3. Analisis dan perancangan sistem
Menganalisis dan merancang penerapan *Grammatical Evolution* ke dalam sistem prediksi data *time series* jumlah pelanggan PSTN TELKOM yang akan dibangun
4. Implementasi sistem
Merealisasikan perancangan sistem yang dikembangkan. Sistem direalisasikan menggunakan program aplikasi MatLab 7.4.0.287 (R2007a). Realisasi sistem dilakukan secara bertahap pada masing-masing modul dan kemudian digabungkan
5. Pengujian Sistem
Menguji terhadap hasil implementasi.
6. Analisis Hasil Pengujian
Data yang didapat dari hasil pengujian dianalisis untuk memperoleh kesimpulan.
7. Kesimpulan
Mengambil kesimpulan dari hasil analisis.

Telkom
University

5. Kesimpulan dan Saran

5.1 Kesimpulan

Dari hasil pengujian dan analisa yang telah dilakukan dapat diambil beberapa kesimpulan sebagai berikut :

1. Prediksi dengan menggunakan sistem yang berbasis *Grammatical Evolution* mampu menghasilkan solusi dalam bentuk fungsi prediksi dengan tingkat adaptasi yang lebih tinggi terhadap data prediksi sebenarnya dibandingkan dengan metode prediksi konvensional seperti *Linear Regression with time*(LR) dan *Moving Average*(MA), karena adanya penggunaan definisi BNF yang menyebabkan sifat non linearitas.
2. Kemampuan algoritma *Grammatical Evolution* untuk dapat melakukan prediksi dalam jangka waktu periode prediksi yang lebih panjang dengan tingkat kesalahan prediksi yang masih dapat ditoleransi(layak), tidak dipengaruhi oleh banyaknya data historis yang digunakan untuk membangun fungsi prediksi melainkan, kepada kecenderungan kesamaan pola data historis dengan data prediksi sebenarnya. Semakin besar kecenderungan kesamaan pola data historis dengan pola data prediksi sebenarnya maka maksimal jumlah periode prediksi yang layak semakin banyak.
3. Pendefinisian *grammar* BNF adalah komponen yang penting pada sistem *Grammatical Evolution*. Hal ini disebabkan karena *grammar* yang digunakan pada BNF sangat menentukan pola solusi yang mungkin dibangun.
4. Pendefinisian *grammar* BNF harus disesuaikan dengan karakteristik masalah yang akan dicari solusinya.
5. Dari hasil pengujian terhadap parameter *Grammatical Evolution*, diperoleh nilai-nilai parameter yang menghasilkan solusi yang optimal pada masing-masing skenario pengujian dalam pemecahan masalah prediksi data *time series* jumlah pelanggan PSTN TELKOM di Indonesia, yaitu :

Tabel 5-1 Parameter *Grammatical Evolution* Terbaik untuk Setiap Skenario Pengujian

Skenario	BNF	Ukuran Populasi	Generasi	Probabilitas crossover
1	I	100	1000	0.9
	II	100	1000	0.7
2	I	100	1000	0.9
	II	100	1000	0.9
3	I	100	1000	0.7
	II	100	1000	0.7

6. Fungsi prediksi yang optimal, yang dapat dipergunakan untuk melakukan prediksi untuk periode berikutnya serta batas maksimal jumlah periode prediksi agar tetap menghasilkan data prediksi dengan kesalahan prediksi yang masih bisa ditoleransi untuk masing-masing skenario adalah sebagai berikut :

Tabel 5-2. Fungsi Prediksi Optimal dan Maksimal Jumlah Periode Prediksi

Skenario	Fungsi Prediksi Optimal	Maksimal Jumlah Periode Prediksi (bulan)
1	$(0.6 \cdot x_3) - (0.4 \cdot x_5) - (0.4 \cdot x_4) + (0.3 \cdot x_1) + (0.9 \cdot x_2) + (0.2 \cdot e_1)$	8
2	$\cos(0.6) - \sin(\cos(0.6)) + \sin(x_1)$	5
3	$(0.2 \cdot x_3) + (0.9 \cdot x_1) + (0.5 \cdot x_2) - (0.6 \cdot x_4)$	7

6.2 Saran

Saran yang dapat digunakan untuk perkembangan selanjutnya :

1. Dapat digunakan data historis yang mempunyai lebih dari satu faktor variabel penentu hasil prediksi
2. Untuk proses evolusi pada *Grammatical Evolution* dalam menemukan individu/solusi yang optimal, dapat dilakukan pengembangan dan percobaan dengan perubahan pada jenis pindah silang yang digunakan, mekanisme penggantian populasi (*steady state replacement*), dan skema pengkodean yang lainnya.

Lampiran A: Data Pengujian

Tabel A-1. Hasil Prediksi Data Testing Skenario 1 dengan Parameter Evolusi (BNF I, UkPop 100, Gen 1000, P_c 0.9)

No	Data Testing	Data Prediksi	Kesalahan	APE
1	8526890	8526890	0	0
2	8544890	8544890	0	0
3	8562890	8562890	0	0
4	8580890	8580890	0	0
5	8595890	8595890	0	0
6	8610890	8610591	299	0.003472347
7	8625890	8625048	842	0.009761312
8	8606591	8637757	-31166	0.362117823
9	8664675	8634867	29808	0.344017519
10	8655742	8644117	11625	0.134303911
11	8658242	8666498	-8256	0.09535423
12	8663242	8691801	-28559	0.329657188
13	8679248	8670616	8632	0.099455621
14	8675258	8669196	6062	0.069876884
15	8684235	8687466	-3231	0.03720535
16	8681243	8691314	-10071	0.116008733
17	8675255	8686329	-11074	0.127650426
18	8684255	8682216	2039	0.023479274
19	8684235	8678361	5874	0.067639809
20	8693213	8681237	11976	0.137762643
21	8698200	8694124	4076	0.04686027
22	8697203	8700904	-3701	0.042553911
23	8682240	8705334	-23094	0.265991265
24	8668275	8695479	-27204	0.313834067

Tabel A-2. Hasil Prediksi Data Testing Skenario 2 dengan Parameter Evolusi (BNF II, UkPop 100, Gen 1000, P_c 0.9)

No	Data Aktual	Data Prediksi	Kesalahan	APE
1	8316603	8316603	0	0
2	8299646	8366653	-67007	0.807348
3	8256753	8350063	-93310	1.130105
4	8246778	8307703	-60925	0.738773
5	8266728	8297785	-31057	0.375687
6	8296653	8317595	-20942	0.252415
7	8310956	8347129	-36173	0.435245
8	8332979	8361140	-28161	0.337946
9	8354487	8382566	-28079	0.336095

10	8375480	8403280	-27800	0.331921
11	8395948	8423278	-27330	0.325514
12	8415890	8442532	-26642	0.316568
13	8425890	8461034	-35144	0.417095
14	8435890	8470209	-34319	0.406821
15	8444890	8479317	-34427	0.407667
16	8453890	8487448	-33558	0.396953
17	8461890	8495512	-33622	0.397334
18	8469890	8502633	-32743	0.386581
19	8476890	8509697	-32807	0.387017
20	8483890	8515837	-31947	0.376561
21	8490390	8521931	-31541	0.371491
22	8496890	8527551	-30661	0.36085
23	8502890	8533132	-30242	0.355667
24	8508890	8538256	-29366	0.345121
25	8526890	8543340	-16450	0.192919
26	8544890	8558389	-13499	0.157977
27	8562890	8573122	-10232	0.119492
28	8580890	8587511	-6621	0.07716
29	8595890	8601551	-5661	0.065857
30	8610890	8612972	-2082	0.024179
31	8625890	8624127	1763	0.020438
32	8606591	8635012	-28421	0.330224
33	8664675	8620956	43719	0.504566
34	8655742	8661861	-6119	0.070693
35	8658242	8655845	2397	0.027685
36	8663242	8657538	5704	0.065841
37	8679248	8660901	18347	0.211389
38	8675258	8671436	3822	0.044056
39	8684235	8668841	15394	0.177264
40	8681243	8674647	6596	0.07598
41	8675255	8672729	2526	0.029117
42	8684255	8668841	15414	0.177494
43	8684235	8674664	9571	0.110211
44	8693213	8674647	18566	0.213569
45	8698200	8680346	17854	0.205261
46	8697203	8683467	13736	0.157936
47	8682240	8682846	-606	0.00698
48	8668275	8673366	-5091	0.058731

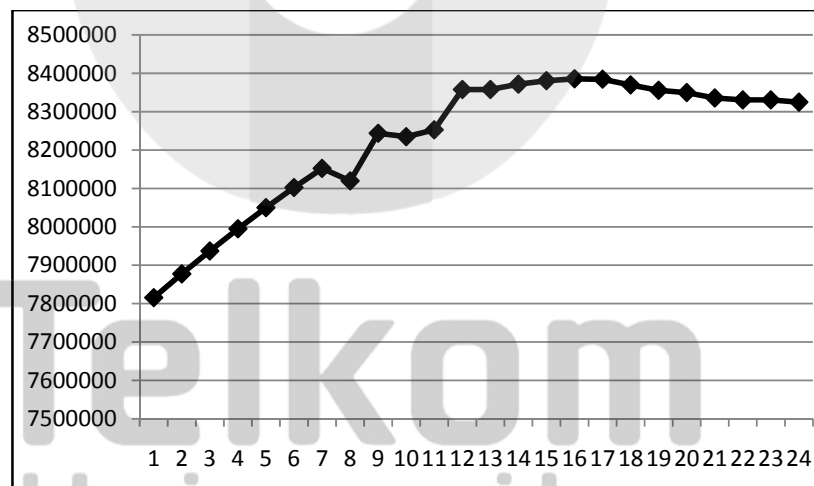
Tabel A- 3. Hasil Prediksi Data Testing Skenario 3 dengan Parameter Evolusi (BNF I, Uk.Pop 100, Gen 1000, P_c 0.7)

No	Data Aktual	Data Prediksi	Kesalahan	APE
1	8679248	8679248	0	0
2	8675258	8675258	0	0
3	8684235	8684235	0	0
4	8681243	8681243	0	0
5	8675255	8682739	-7484	0.086268358
6	8684255	8680043	4212	0.04850157
7	8684235	8679165	5070	0.058381654
8	8693213	8684244	8969	0.10317244
9	8698200	8697707	493	0.005667839
10	8697203	8701280	-4077	0.04687714
11	8682240	8704684	-22444	0.258504718
12	8668275	8686330	-18055	0.20828827

Gambar Grafik Pembagian Data Uji pada Masing-masing Skenario Pengujian

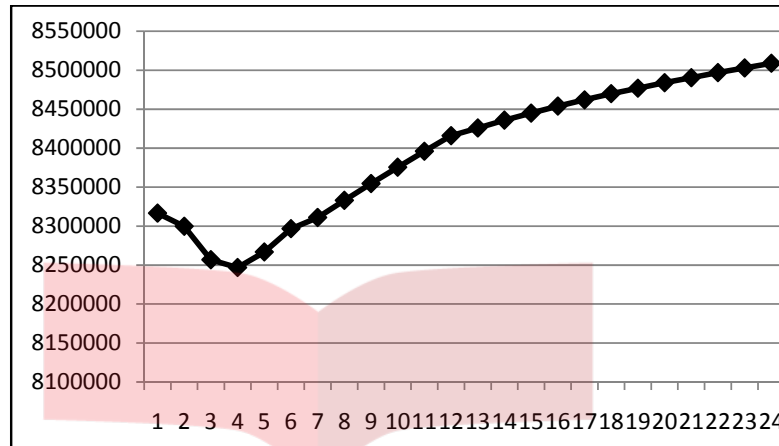
SKENARIO 1 :

1. Data Training : 24 bulan



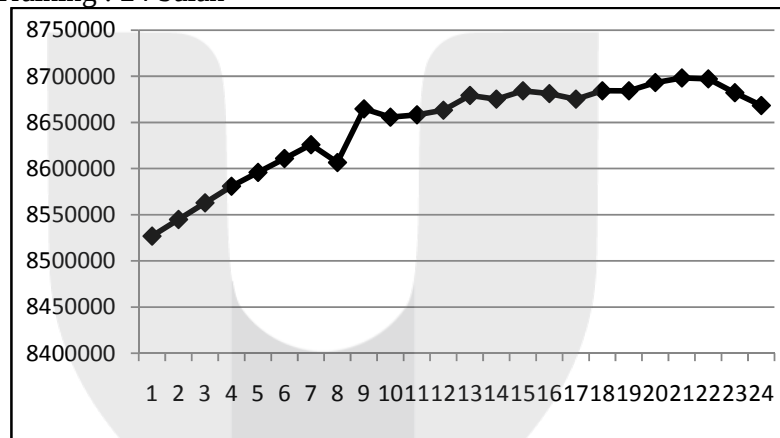
Gambar A-1. Grafik Data Training Skenario 1

2. Data Validasi : 24 bulan



Gambar A-2. Grafik Data Validasi Skenario 1

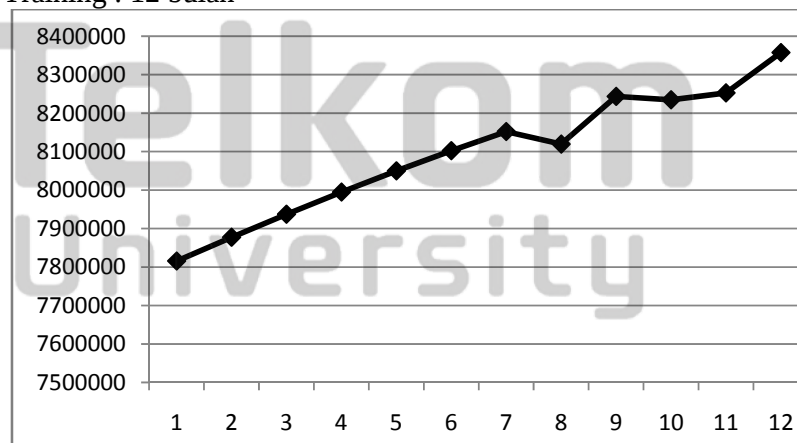
3. Data Training : 24 bulan



Gambar A-3. Grafik Data Testing Skenario 1

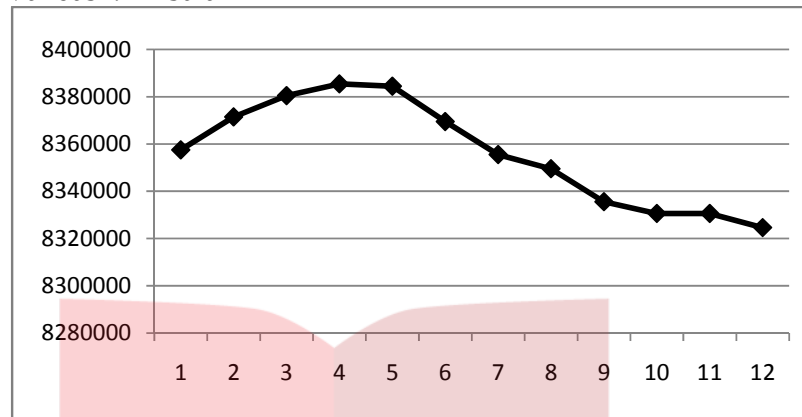
SKENARIO 2:

1. Data Training : 12 bulan



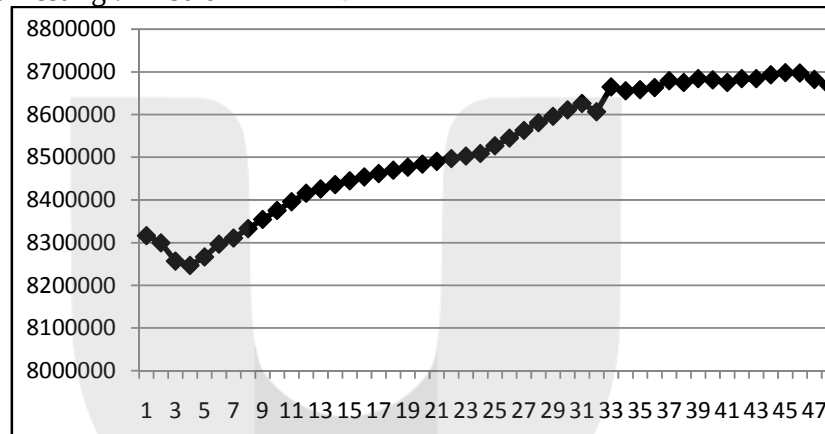
Gambar A-4 Grafik Data Training Skenario 2

2. Data Validasi : 12 bulan



Gambar A-5 Grafik Data validasi Skenario 2

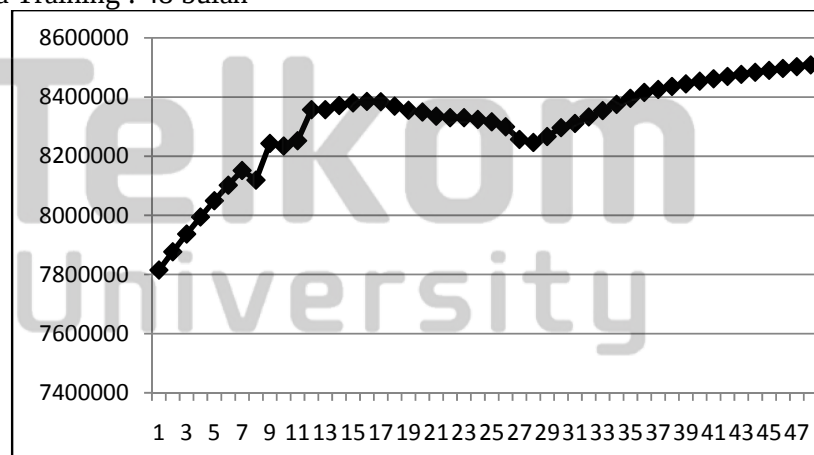
3. Data Testing : 12 bulan



Gambar A-6 Grafik Data Testing Skenario 2

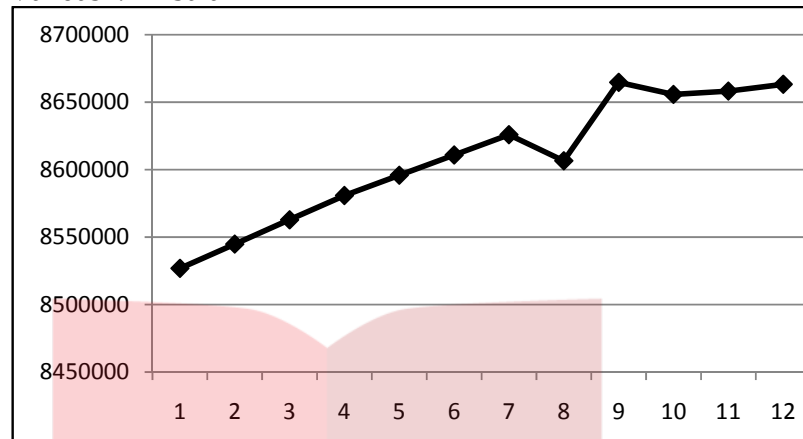
SKENARIO 3 :

1. Data Training : 48 bulan



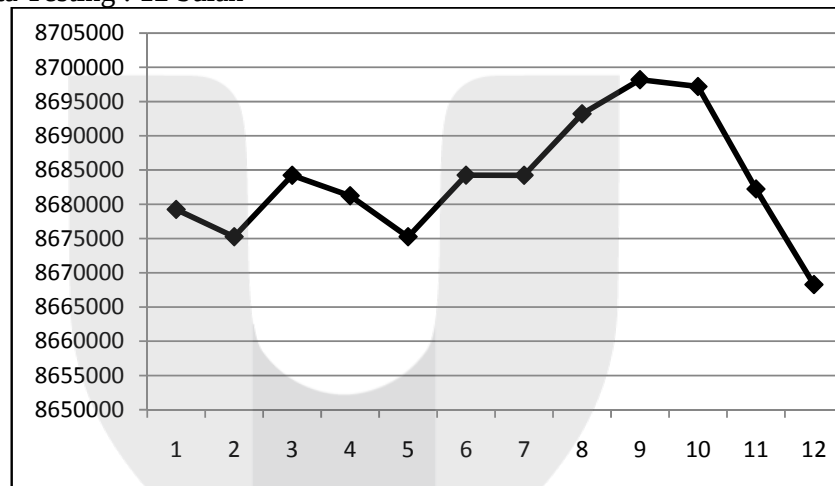
Gambar A-7 Grafik Data Training Skenario 3

2. Data Validasi : 12 bulan



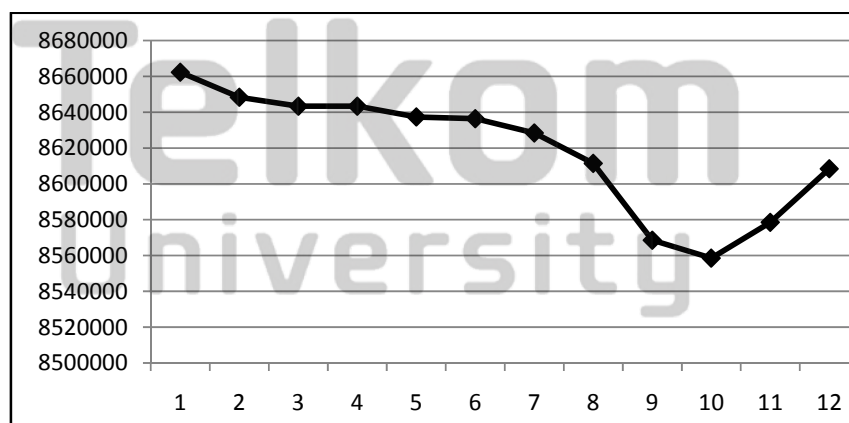
Gambar A-8 Grafik Data Validasi Skenario 3

3. Data Testing : 12 bulan



Gambar A-9 Grafik Data Testing Skenario 3

DATA PELANGGAN TAHUN 2008



Gambar A-10 Grafik Data Pelanggan Tahun 2008