

**KAJIAN TIME SERIES DATAMINING
DENGAN MEMBANDINGKAN ALGORITMA MULTILAYER
PERCEPTRON DAN SUPPORT VECTOR MACHINE GUNA
MEMPREDIKSI KONSUMSI DATA CDMA 2000 EV-DV:
STUDI KASUS PERUSAHAAN TELEKOMUNIKASI ABC**



Oleh:

Tommy Hutomo

1111600894

**PROGRAM STUDI: MAGISTER ILMU KOMPUTER (MKOM)
PROGRAM PASCASARJANA
UNIVERSITAS BUDI LUHUR**

JAKARTA

2014

**KAJIAN TIME SERIES DATAMINING
DENGAN MEMBANDINGKAN ALGORITMA MULTILAYER
PERCEPTRON DAN SUPPORT VECTOR MACHINE GUNA
MEMPREDIKSI KONSUMSI DATA CDMA 2000 EV-DV:
STUDI KASUS PERUSAHAAN TELEKOMUNIKASI ABC**

TESIS

**Diajukan sebagai salah satu persyaratan
memperoleh gelar Magister Ilmu Komputer (M.Kom)**



Oleh:

Tommy Hutomo

1111600894

PROGRAM STUDI: MAGISTER ILMU KOMPUTER (MKOM)

PROGRAM PASCASARJANA

UNIVERSITAS BUDI LUHUR

JAKARTA

2014



PROGRAM STUDI : MAGISTER ILMU KOMPUTER (MKOM)
PROGRAM PASCA SARJANA
UNIVERSITAS BUDI LUHUR

LEMBAR PERNYATAAN

Nama Mahasiswa : Tommy Hutomo
Nomor Induk Mahasiswa : 1111600894
Konsentrasi : Teknik informatika
Program : Pasca Sarjana

Menyatakan tesis yang berjudul:

Kajian Time Series Datamining Dengan Membandingkan Algoritma Multilayer Perceptron dan Support Vector Machine Guna Memprediksi Konsumsi Data CDMA 2000 EV-DV : Studi Kasus Perusahaan Telekomunikasi ABC

1. Merupakan hasil karya tulis ilmiah sendiri dan bukan merupakan karya yang pernah diajukan untuk memperoleh gelar akademik oleh pihak lain.
2. Saya ijin untuk dikelola oleh Universitas Budi Luhur sesuai dengan norma hukum dan etika yang berlaku.

Pertanyaan ini saya buat dengan penuh tanggung jawab dan saya siap menerima konsekuensi apapun sesuai aturan yang berlaku, apabila dikemudian hari pernyataan ini tidak benar.

Jakarta, 23 Februari 2014



Tommy Hutomo



PROGRAM STUDI : MAGISTER ILMU KOMPUTER (MKOM)
PROGRAM PASCA SARJANA
UNIVERSITAS BUDI LUHUR

LEMBAR PENGESAHAN

Nama Mahasiswa : Tommy Hutomo
Nomor Induk Mahasiswa : 1111600894
Konsentrasi : Teknik informatika
Judul Tesis : Kajian Time Series Datamining Dengan Membandingkan
Algoritma Multilayer Perceptron dan Support Vector Machine
Guna Memprediksi Konsumsi Data CDMA 2000 EV-DV : Studi
Kasus Perusahaan Telekomunikasi ABC

Jakarta, 22 Februari 2014

Tim Penguji:

Tanda Tangan:

Ketua,

Dr., Ir., Nazori A.Z., M. T.

Anggota,

Samidi., M.kom., MM

Pembimbing,

Dana Indra Sensuse., Ir., MLIS., PhD

ABSTRAK

Industri telekomunikasi selular berkembang pesat di Indonesia, masalah muncul karena area di Indonesia tidak memiliki kemampuan yang sama dan terkadang tidak sebanding dengan kebutuhan pelanggan yang ada. Hal ini disebabkan berbagai alasan, salah satu diantaranya perencanaan kapasitas selular yang tidak sempurna dalam memprediksi konsumsi layanan pada area tersebut. PT ABC adalah salah satu perusahaan telekomunikasi CDMA2000 EV-DV di Indonesia. Peneliti menawarkan alternatif solusi berupa prediksi konsumsi layanan data pada jaringan CDMA 2000 EV-DV dengan mendekati *time series datamining* menggunakan algoritma terbaik hasil perbandingan *multilayer perceptron* dan *support vector machine*. Metode pendekatan datamining yang digunakan adalah metodologi CRISP-DM. Pada penelitian ini ditemukan bahwa algoritma *support vector machine* dapat memberikan akurasi yang lebih tinggi dari algoritma *multilayer perceptron*. Dalam melakukan memprediksi konsumsi layanan data pada jaringan CDMA 2000 algoritma *support vector machine* dapat memberikan PTA sebesar 0.842, MSE sebesar $260,752 \times 10^6$, MAPE sebesar 2.93 dan MAD sebesar 6,044,234,170.00. Penelitian ini diakhiri dengan pembangunan *prototype* yang mengacu pada metodologi *prototyping*, yang mana *prototype* tersebut diujikan kembali menggunakan *software quality assurance quality factor* ISO 9126, dengan skor SQA yang dihasilkan 80.225.

Kata kunci : *CDMA 2000 EV-DV, Time series Datamining, Multilayer Perceptron, Support V ector Machine, CRISP-DM, Prototyping, Software Quality Assurance, ISO 9126*

ABSTRACT

Mobile telecommunications industry in Indonesia is growing rapidly , a problem arises because the area in Indonesia does not have the same ability and sometimes not worth the needs of existing customers . This is due to various reasons , one of which cellular capacity planning is not well perform in predicting consumption of services in the area . PT ABC is one of the telecommunications companies CDMA2000 EV - DV in Indonesia . Researchers offer an alternative solution in the form of predictions consumption of data services on the CDMA 2000 EV - DV with time series data mining approach best results comparison between multilayer perceptron and support vector machine . Datamining approach method used is the CRISP - DM methodology .This study found that the support vector machine algorithm can provide higher accuracy than the multilayer perceptron algorithm. In doing predict consumption of data services in CDMA 2000 networks, support vector machine algorithm can provide PTA 0.842, MSE $260,752 \times 10^6$, MAPE 2.93 and MAD 6,044,234,170.00. This study concludes with the development of a prototype that refer to the prototyping methodology, in which the prototype was tested again using a quality factor of software quality assurance ISO 9126, with the resulting score is 80 225

Keyword : *CDMA 2000 EV-DV, Time series Datamining, Multilayer Perceptron, Support V ector Machine, CRISP-DM, Prototyping, Software Quality Assurance, ISO 9126*

KATA PENGANTAR

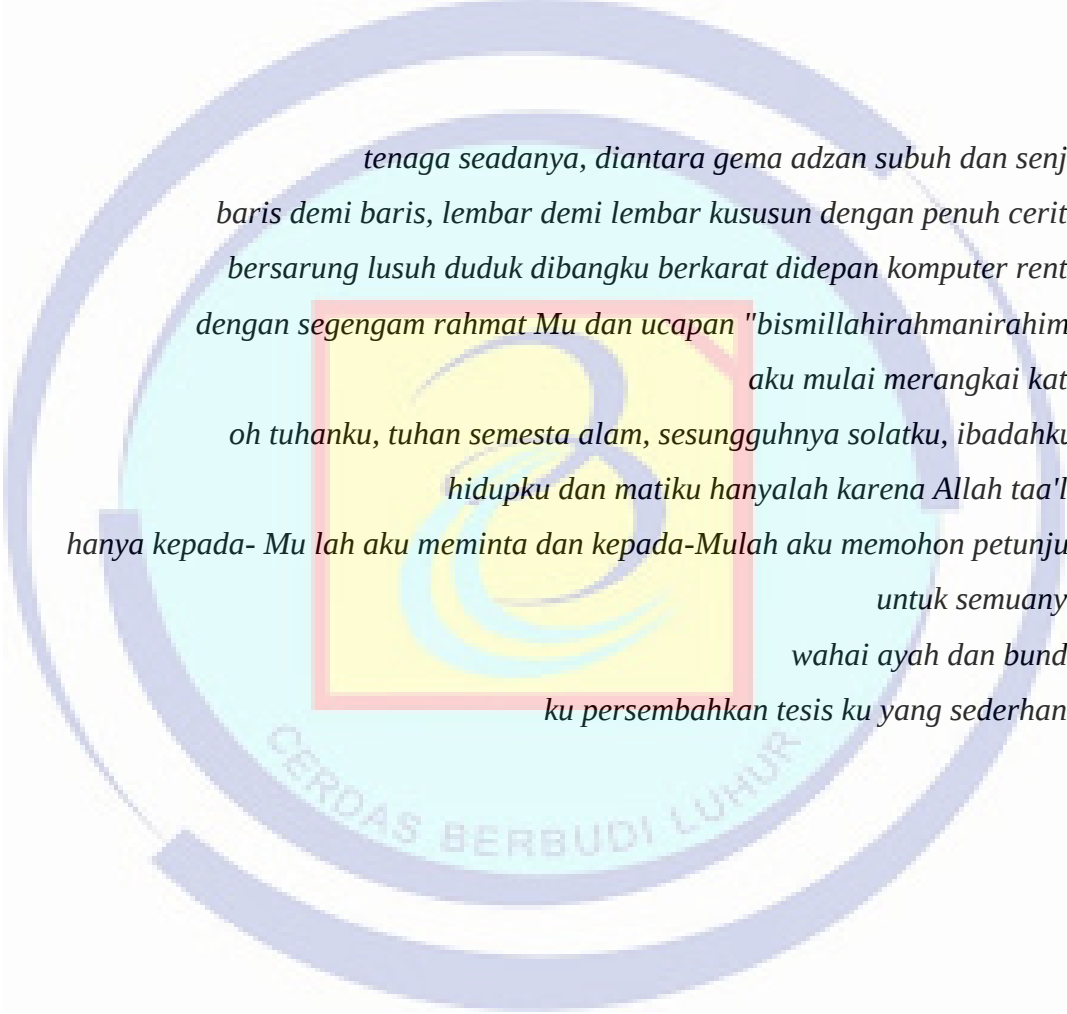
Puji syukur penulis panjatkan kepada Allah SWT yang telah memberikan rahmat dan hidayahnya sehingga penulis dapat menyelesaikan penelitian tesis yang berjudul *Kajian Time Series Datamining Dengan Membandingkan Algoritma Multilayer Perceptron dan Support Vector Machine guna Memprediksi Konsumsi Data CDMA 2000 EV-DV : Studi Kasus Perusahaan Telekomunikasi ABC*. Tak pula shalawat dan salam penulis haturkan kepada baginda Rasulullah Muhammad SAW. Adapun tujuan penulisan isi adalah sebagai salah satu syarat untuk menuntaskan pendidikan pada Program Studi Magister Ilmu Komputer, Universitas Budi Luhur.

Penulis menghaturkan terima kasih kepada semua pihak yang telah memberi dukungan dan bantuan kepada penulis dalam menulis peneltian tesis ini :

1. Bapak Ir. Dana Indra Sensuse, MLIS, PhD, selaku dosen pembimbing tesis yang telah membimbing dan meluangkan waktu untuk membimbing penulis dalam mengerjakan tesis ini.
2. Bapak Dr. Moedjiono, M.Sc., selaku Dekan program studi Magister Ilmu Komputer Program Pasca Sarjana di Universitas Budi Luhur yang telah memberikan kesempatan untuk mengali ilmu dan menyelesaikan tesis ini.
3. Bapak Dr. Ir. Nazori Az., MT., selaku Ketua Jurusan program studi Magister Ilmu Komputer Program Pasca Sarjana di Universitas Budi Luhur yang telah memberikan dukungan untuk menyelesaikan tesis ini.
4. Alm. Ir Soewantoro dan Dra Hj Munichah, selaku orang tua penulis yang telah memberikan kasih sayang dan semangatnya kepada penulis sehingga tesis ini dapat diselesaikan.

Penulis yang mahluk Allah SWT penuh kealfaan dan jauh dari kesempurnaan. Dengan penuh kesederhanaan penulis mengharapkan kritik dan saran yang membangun untuk perbaikan penulis kedepannya.

Jakarta,
Tommy Hutomo



*tenaga seadanya, diantara gema adzan subuh dan senja
baris demi baris, lembar demi lembar kusun dengan penuh cerita
bersarung lusuh duduk dibangku berkarat didepan komputer renta
dengan segengam rahmat Mu dan ucapan "bismillahirrahmanirrahim"
aku mulai merangkai kata
oh tuhanku, tuhan semesta alam, sesungguhnya solatku, ibadahku,
hidupku dan matiku hanyalah karena Allah taa'la
hanya kepada- Mu lah aku meminta dan kepada-Mulah aku memohon petunjuk
untuk semuanya
wahai ayah dan bunda
ku persembahkan tesis ku yang sederhana*

DAFTAR ISI

	Halaman
ABSTRAK	i
ABSTRACT	ii
KATA PENGANTAR.....	iii
DAFTAR ISI	v
DAFTAR GAMBAR.....	viii
DAFTAR TABEL	x
DAFTAR LAMPIRAN	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Masalah Penelitian	3
1.2.1 Identifikasi Masalah.....	3
1.2.2 Batasan Masalah	4
1.2.3 Rumusan Masalah.....	5
1.3 Tujuan dan Manfaat Penelitian	5
1.3.1 Tujuan Penelitian	5
1.3.2 Manfaat Penelitian	5
1.4 Sistematika Penulisan	6
1.5 Daftar Istilah	7
BAB II LANDASANPEMIKIRAN.....	9
2.1 Tinjauan Pustaka	9
2.1.1 Datamining.....	9
2.1.1.1 <i>Cross Industry Standard Process for Data Mining</i> .	11
2.1.1.2 Komponen <i>Datamining</i>	13
2.1.1.3 <i>Prediksi Time Series</i>	14
2.1.2 <i>Multilayer Perceptron</i>	16
2.1.2.1 Arsitektur jaringan saraf tiruan	16
2.1.2.2 Metode pembelajaran <i>Backpropagation</i>	17
2.1.3 <i>Support Vector Machine</i>	20
2.1.3.1 <i>Support Vector</i>	20
2.1.3.2 <i>Kernel Trick</i>	21
2.1.3.3 <i>Sequential Minimal Optimization</i>	23
2.1.4 CDMA 2000 EV-DV	23
2.1.4.1 Arsitektur jaringan sistem selular CDMA 2000 1x ..	23
2.1.4.2 Kanal sistem selular CDMA 2000 1x EV-DV.....	25
2.1.4.3 <i>Data Call Procedure</i> CDMA 2000 EV-DV	28
2.1.5 Pendekatan <i>Windowing</i>	29
2.1.6 <i>Predicted Trend Accuracy (PTA)</i>	30
2.1.7 <i>Bivariate corellation</i> dan standar deviasi populasi	30

2.1.8 <i>Object Oriented Analysis and Design</i>	31
2.1.9 <i>Prototyping Methodologies</i>	34
2.1.10 <i>Software Quality Assurance</i>	35
2.1.10.1 <i>ISO 9126 Quality Factor</i>	38
2.2 Tinjauan Studi	39
2.2.1 Model penelitian Pandey	39
2.2.2 Metode Penelitian Rohana	40
2.2.3 Metode Penelitian Saigal et al	41
2.2.4 Metode Penelitian Mantri	43
2.2.5 Ringkasan Penelitian Terkait	44
2.3 Tinjauan Objek Penelitian	46
2.3.1 Aspek Organisasi	46
2.3.2 Aspek Sistem	47
2.4 Kerangka Konsep	48
2.5 Hipotesis Penelitian	49
BAB III METODOLOGI DAN RANCANGAN PENELITIAN	51
3.1 Metode Penelitian	51
3.2 Metode Pemilihan Sampel	51
3.3 Metode pengumpulan data	52
3.4 Instrumentasi	52
3.5 Langkah Penelitian	53
3.5.1 <i>Bussiness Understanding</i>	54
3.5.2 <i>Data Understanding</i>	55
3.5.3 <i>Data Preparation/Preprocessing</i>	62
3.5.3.1 <i>Data aggregation dan selection</i>	62
3.5.3.2 <i>Data Transformation dan Cleansing</i>	63
3.5.4 Modelling	66
3.5.4.1 Penentuan <i>training dan testing set</i>	66
3.5.4.2 Pendefinisian interproses basis data	67
3.5.4.3 Penentuan <i>role data</i>	69
3.5.4.4 Pendefinisian <i>Windowing operator</i>	69
3.5.4.5 <i>Training dan Testing</i>	70
3.5.5 <i>Evaluation</i>	71
3.5.6 <i>Deployment</i>	72
3.5.6.1 Pemilihan Algoritma	72
3.5.6.2 <i>Prototyping</i>	73
3.5.6.3 <i>Software Quality Assurance</i>	73
3.6 Jadwal Penelitian	75
BAB IV PEMBAHASAN PENELITIAN	77
4.1 Hasil Penelitian	77
4.1.1 Hasil Prediksi Algoritma <i>Multilayer Perceptron</i>	83
4.1.2 Hasil Prediksi Algoritma <i>Support Vector Machine</i>	86

4.2	Evaluasi.....	88
4.2.1	Evaluasi Algoritma <i>Multilayer Perceptron</i>	88
4.2.1.1	Evaluasi dengan PTA.....	88
4.2.1.2	Evaluasi dengan MSE	89
4.2.1.3	Evaluasi dengan MAPE	90
4.2.1.4	Evaluasi dengan MAD	91
4.2.2	Evaluasi Algoritma <i>Support Vector Machine</i>	91
4.2.2.1	Evaluasi dengan PTA.....	91
4.2.2.2	Evaluasi dengan MSE	92
4.2.2.3	Evaluasi dengan MAPE	93
4.2.2.4	Evaluasi dengan MAD	93
4.2.3	Kesimpulan Evaluasi	94
4.3	Penerapan Algoritma Terpilih.....	95
4.4	<i>Prototyping</i>	95
4.4.1	Perancangan <i>Prototype</i>	95
4.4.2	Pembangunan <i>Prototype</i>	99
4.5	<i>Software Quality Assurance</i>	101
4.6	Implikasi Penelitian	103
4.6.1	Aspek Manajerial	103
4.6.2	Aspek Penelitian Lanjutan	104
BAB IV	PENUTUP.....	105
5.1	Kesimpulan	105
5.2	Saran	105
DAFTAR	PUSTAKA.....	106

DAFTAR GAMBAR

Gambar	Halaman
I.1. Jumlah pelanggan seluler	1
I.2. ARPU untuk layanan data dan suara.....	2
II.1 Klasifikasi datamining	11
II.2 <i>Life Cycle</i> CRISP-DM	12
II.3 Arsitektur <i>Datamining</i>	14
II.4 Komponen <i>Time Series</i>	15
II.5 Lapisan-lapisan pada jaringan saraf tiruan	17
II.6 <i>Hyperplane</i> yang misahkan <i>training set</i> 2 dimensi[.....	21
II.7 <i>Hyperplane</i> non linier	21
II.8 <i>Soft-margin hyperplane</i>	22
II.9 Arsitektur CDMA	25
II.10 Alur persedur data call pada CDMA	28
II.11 <i>Windowing Approach</i>	29
II.12 Contoh <i>Use case Diagram</i>	32
II.13 Contoh <i>Class diagram</i>	33
II.14 Contoh <i>Sequence Diagram</i>	34
II.15 <i>Prototyping-based Methodologies</i>	35
II.16 Ilustrasi <i>incremental testing</i>	37
II.17 Model penelitian Pandey	39
II.18 Model penelitian Tatang Rohana.....	41
II.19 Grafik hasil penelitian Saigal et al.	42
II.20 Model penelitian Saigal et al	42
II.21 Model penelitian Mantri	43
II.22 Arsitektur gudang data pada perusahaan ABC	47
II.23 Kerangka Konsep.....	50
III.1 Tahapan penelitian.....	54
III.2 Tampilan awal RapidMiner 5.3.008.....	67
III.3 Operator pembaca basis pada RapidMiner 5.03.008.....	68
III.4 Input parameter koneksi basis data pada RapidMiner 5.03.008.....	68
III.5 Definisi <i>role</i> data	69
III.6 Definisi <i>windowing operator</i>	70
III.7 <i>Training</i> dan <i>Testing Multilayer Perceptron</i>	71
III.8 <i>Training</i> dan <i>Testing Support Vector Machine</i>	71
IV.1 Grafik jumlah <i>subscriber</i> hasil <i>cleansing</i> dan <i>aggregasi</i>	78
IV.2 Grafik Transaksi SMS hasil <i>cleansing</i> dan <i>aggregasi</i>	78
IV.3 Grafik durasi <i>Outgoing</i> hasil <i>cleansing</i> dan <i>aggregasi</i>	79
IV.4 Grafik <i>traffic</i> RSCH hasil <i>cleansing</i> dan <i>aggregasi</i>	80

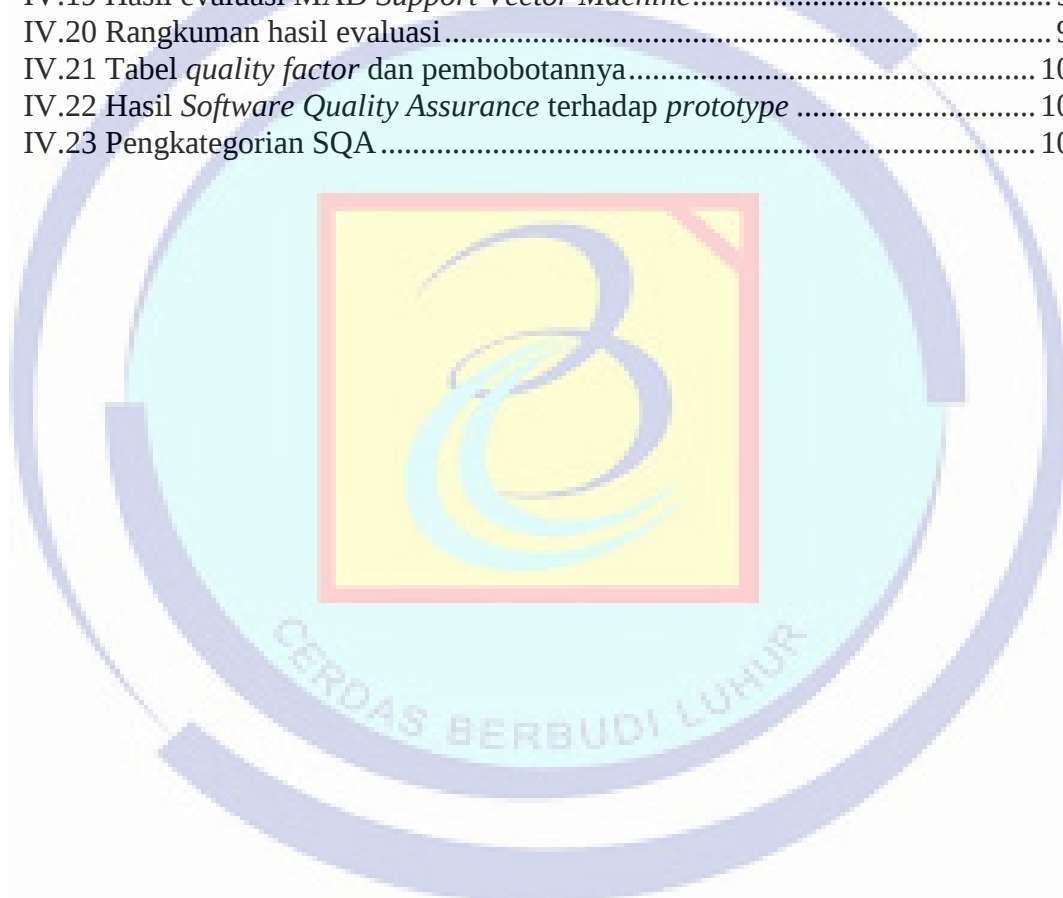
IV.5 Grafik <i>traffic</i> RACH hasil <i>cleansing</i> dan agregasi.....	80
IV.6 Grafik <i>traffic</i> FSCH hasil <i>cleansing</i> dan agregasi.....	81
IV.7 Grafik Konsumsi data hasil <i>cleansing</i> dan agregasi.....	81
IV.8 <i>Use case prototype</i> yang akan dibangun.....	95
IV.9 <i>Sequence diagram prototype</i> yang akan dibangun.....	96
IV.10 <i>Class diagram prototype</i> yang akan dibangun.....	97
IV.11 Diagram alur <i>prototype</i> yang akan dibangun.....	98
IV.12 Rancangan tampilan <i>prototype</i> yang akan dibangun.....	99
IV.13 Tampilan <i>prototype</i> yang telah dibangun.....	99
IV.14 Tampilan <i>prototype</i> saat mengeluarkan pesan error	100
IV.15 Tampilan <i>log process prototype</i> saat prediksi selesai dilakukan	101



DAFTAR TABEL

Tabel	Halaman
II.1 Fungsi-fungsi kernel yang sering digunakan	22
II.2 Koeefisien korelasi	31
II.3 Hasil penelitian Guarav pandey	40
II.4 Hasil penelitian Mantri	44
II.5 Ringkasan penelitian yang telah dilakukan sebelumnya	44
II.6 Perbedaan penelitian yang dilakukan dengan penelitian sebelumnya	45
II.7 Spesifikasi Perangkat keras	48
II.8 Perangkat Lunak yang digunakan pada penelitian	48
III.1 Kandidat data	55
III.2 Contoh data tabel <i>fact_rated_data_summary</i>	57
III.3 Contoh tabel <i>outgoing</i> (a) dan data tabel <i>incoming</i> (b)	58
III.4 Contoh data tabel <i>fact_rated_sms_summary</i>	59
III.5 Contoh data tabel <i>fact_subscriber</i>	59
III.6 Contoh data tabel <i>fact_topup_summary</i>	60
III.7 Contoh data tabel <i>fact_unrated</i>	61
III.8 Hasil analisa korelasi	62
III.9 Atribut dataset	65
III.10 Kombinasi rentang waktu <i>training</i> dan evaluasi	66
III.11 Perangkat keras dan lunak yang digunakan	67
III.12 Pembobotan SQA	74
III.12 Pengkategorian SQA	75
III.13 Jadwal Penelitian	76
IV.1 Kombinasi <i>training set</i> dan <i>testing set</i>	77
IV.2 Parameter yang digunakan untuk simulasi <i>Multilayer Perceptron</i>	82
IV.3 Parameter yang digunakan untuk simulasi <i>Support Vector Machine</i>	83
IV.4 <i>Sample</i> hasil prediksi algorima <i>Multilayer Perceptron</i> dengan 75% <i>training set</i>	83
IV.5 <i>Sample</i> hasil prediksi algorima <i>Multilayer Perceptron</i> dengan 80% <i>training set</i>	84
IV.6 <i>Sample</i> hasil prediksi algorima <i>Multilayer Perceptron</i> dengan 85% <i>training set</i>	85
IV.7 <i>Sample</i> hasil prediksi algorima <i>Multilayer Perceptron</i> dengan 90% <i>training set</i>	85
IV.8 <i>Sample</i> hasil prediksi algorima <i>Support Vector Machine</i> dengan 75% <i>training set</i>	86
IV.9 <i>Sample</i> hasil prediksi algorima <i>Support Vector Machine</i> dengan 80% <i>training set</i>	87
IV.10 <i>Sample</i> hasil prediksi algorima <i>Support Vector Machine</i> dengan 85% <i>training set</i>	87

IV.11	Sample hasil prediksi algoritma Support Vector Machine dengan 90% training set.....	88
IV.12	Tabel hasil evaluasi PTA Multilayer Perceptron.....	89
IV.13	Tabel hasil evaluasi MSE Multilayer Perceptron (skala 10^6).....	89
IV.14	Tabel hasil evaluasi MAPE Multilayer Perceptron	90
IV.15	Tabel hasil evaluasi MAD Multilayer Perceptron.....	91
IV.16	Hasil evaluasi PTA Support Vector Machine	91
IV.17	Hasil evaluasi MSE Support Vector Machine(skala 10^6).....	92
IV.18	Hasil evaluasi MAPE Support Vector Machine	93
IV.19	Hasil evaluasi MAD Support Vector Machine.....	94
IV.20	Rangkuman hasil evaluasi.....	94
IV.21	Tabel quality factor dan pembobotannya.....	101
IV.22	Hasil Software Quality Assurance terhadap prototype	102
IV.23	Pengkategorian SQA.....	103



DAFTAR LAMPIRAN

Lampiran	Halaman
1. Pengujian SQA oleh Kepala Devisi Datawarehouse.....	111
2. Pengujian SQA oleh Administrator Datawarehouse	112
3. Pengujian SQA oleh Data Analyst	113
4. Pengujian SQA oleh Bussiness Analyst	114
5. Data hasil <i>cleansing</i> dan agregasi.....	115
6. Hasil Prediksi <i>Multilayer Perceptron</i> menggunakan 85% <i>Training set</i>	129
7. Hasil Prediksi <i>Support Vector Machine</i> menggunakan 90% <i>Training set</i>	131
8. Source Code.....	132

DAFTAR PUSTAKA

- [Adamson 2010] Adamson, C., *Star Schema - The Complete Reference*, McGraw-Hill, 2010
- [Ali 2007] Ali, M., *Introduction to CDMA Wireless Communication*, ELSEVIER, 2007
- [Bowen 1985] Bowen, T. P., Wigle, Gary, B., Tsai, T., J., *Spesification of Software Quality Attributes: software Quality Evaluation Guide Book*, Boeing Aerospace Company, 1985
- [Booch 1998] Booch, G., *Object Oriented Design and Analysis with application*, Addison-Welly, 1998
- [Christianini 2000] Christianini, Nello, Taylor, J., S., *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2000
- [Crone 2005] Crone, S. , F, *Forecasting with Artificial Neural Networks*, Lancaster University, 2005
- [Dawson 2009] Dawson, C., W., *Project in Computing and Information System a Student Guide 2nd Edition*, Addison-Wesley, 2009
- [Dennis 2005] Dennis, A, Wixom, B., H., Tegarden, D., *System Analysis and Design with UML v2.0*, Wiley, 2005