

ABSTRAK

Jumlah akses web sangatlah penting bagi pemilik website karena jumlah akses tersebut menentukan popularitas dan cakupan user. Namun, turunnya jumlah akses tidak dapat dicegah secara cepat karena pemilik *website* tidak mengetahui kapan seorang pengguna berhenti mengakses *website* tersebut. Maka dari itu pada tugas akhir ini akan dibangun sebuah sistem prediksi konsistensi akses *website*. Dengan adanya sistem tersebut diharapkan mampu membantu pemilik *website* mengetahui lebih dulu apakah pengguna akan pergi meninggalkan *website* atau tidak. Sistem prediksi ini dibangun berdasarkan 3 grup atribut yaitu kesamaan berdasarkan *domain*, *cosine* dan *cluster* dimana setiap grup memiliki 8 fitur dan menggunakan *classifier multinomial naïve-bayes*. Berdasarkan pengujian, sistem prediksi memiliki akurasi sebesar 48-53% dimana fitur paling berpengaruh adalah *average delta value with the same website*.

Kata kunci: konsistensi akses *website*, *machine learning*, *multinomial naïve-bayes*, *website* dan *software user behavior*



ABSTRACT

The number of website access is important for web stakeholders because that number determines the website popularity and user coverage. However, it is hard to prevent user for leaving the website because we do not know whenever the user is leaving. So, from this final project we propose a system to predict web access consistency at the time the website is being accessed. In such manner, we can prevent user leaving earlier. This system is built based on 3 similarity groups which are domain based, cosine based and cluster based where every group have 8 features and using classifier multinomial naïve-bayes. According to the result, this system have an accuracy around 48-53% and the most significant feature is average delta value with the same website.

Keywords: machine learning, multinomial naïve-bayes, web access consistency, web and software log behavior



DAFTAR ISI

LEMBAR PENGESAHAN.....	i
PERNYATAAN ORISINALITAS LAPORAN PENELITIAN	ii
PERNYATAAN PUBLIKASI LAPORAN PENELITIAN.....	iii
PRAKATA	v
ABSTRAK.....	vii
ABSTRACT.....	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xiv
DAFTAR TABEL	xx
DAFTAR SINGKATAN	xxi
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	2
1.3 Tujuan Pembahasan	2
1.4 Ruang Lingkup	2
1.5 Sumber Data	3
1.6 Sistematika Penyajian	3
BAB 2 KAJIAN TEORI	5
2.1 <i>Machine Learning</i>	5
2.2 <i>Classification</i>	7
2.3 <i>Multinomial Naïve Bayes</i>	8
2.4 <i>Weka</i>	8
2.5 Riset terkait	9
2.6 <i>Unified Modeling Language (UML)</i>	10

2.6.1 Use Case Diagram	11
2.6.2 Activity Diagram	12
BAB 3 ANALISIS DAN RANCANGAN SISTEM	14
3.1 Use Case	14
3.2 Activity Diagram	14
3.2.1 Activity Diagram Transfer Data	15
3.2.2 Activity Diagram Training Data	16
3.2.3 Activity Diagram Prediksi Konsistensi Akses Web	17
3.3 Pseudocode	18
3.3.1 Method TakeDeltaValueWithTheSameWebsite	18
3.3.2 Method TakeDeltaValueWithoutTheSameWebsite	19
3.3.3 Method AverageDeltaValueWithTheSameWebsite	21
3.3.4 Method AverageDeltaValueWithoutTheSameWebsite	22
3.3.5 Method TakeProportionWithTheSameWebsiteWithin24Hours	23
3.3.6 Method TakeProportionWithoutTheSameWebsiteWithin24Hours	24
3.3.7 Method TakeProportionWithTheSameWebsiteWithin7Days	26
3.3.8 Method TakeProportionWithoutTheSameWebsiteWithin7Days	27
3.3.9 Method TakeDeltaValueWithTheSimilarCosine	29
3.3.10 Method TakeDeltaValueWithoutTheSimilarCosine	30
3.3.11 Method AverageDeltaValueWithTheSimilarCosine	31
3.3.12 Method AverageDeltaValueWithoutTheSimilarCosine	32
3.3.13 Method TakeProportionFor24HoursWithTheSimilarCosine	34
3.3.14 Method TakeProportionFor24HoursWithoutTheSimilarCosine	35
3.3.15 Method TakeProportionFor7DaysWithTheSimilarCosine	36
3.3.16 Method TakeProportionFor7DaysWithoutTheSimilarCosine	37
3.3.17 Method TakeDeltaValueWithTheSimilarCluster	39

3.3.18 Method TakeDeltaValueWithoutTheSimilarCluster	40
3.3.19 Method AverageDeltaValueWithTheSimilarCluster.....	41
3.3.20 Method AverageDeltaValueWithoutTheSimilarCluster.....	42
3.3.21 Method TakeProportionFor24HoursWithTheSimilarCluster	44
3.3.22 Method TakeProportionFor24HoursWithoutTheSimilarCluster.....	45
3.3.23 Method TakeProportionFor7DaysWithTheSimilarCluster	46
3.3.24 Method TakeProportionFor7DaysWithoutTheSimilarCluster	47
3.4 Arsitektur Aplikasi.....	49
BAB 4 IMPLEMENTASI.....	50
4.1 Data Preprocessing.....	50
4.1.1 Proses dari Data Mentah Menjadi <i>listUser</i>	53
4.1.2 Proses dari <i>listUser</i> menjadi <i>listUserGlobal</i>	54
4.1.3 Proses dari <i>listUserGlobal</i> menjadi <i>listUserLocal</i>	56
4.2 Implementasi Kode	57
4.2.1 Method <i>takeDeltaValueWithTheSameWebsiteSingleReturn</i>	57
4.2.2 Method <i>takeDeltaValueWithoutTheSameWebsiteSingleReturn</i>	59
4.2.3 Method <i>averageDeltaValueWithTheSameWebsite</i>	61
4.2.4 Method <i>averageDeltaValueWithoutTheSameWebsite</i>	63
4.2.5 Method <i>takeProportionWithTheSameWebsiteWithin24Hours</i>	65
4.2.6 Method <i>takeProportionWithoutTheSameWebsiteWithin24Hours</i>	68
4.2.7 Method <i>takeProportionWithTheSameWebsiteWithin7Days</i>	71
4.2.8 Method <i>takeProportionWithoutTheSameWebsiteWithin7Days</i>	74
4.2.9 Method <i>takeDeltaValueWithTheSimilarCosine</i>	77
4.2.10 Method <i>takeDeltaValueWithoutTheSimilar Cosine</i>	79
4.2.11 Method <i>averageDeltaValueWithTheSimilarCosine</i>	81
4.2.12 Method <i>averageDeltaValueWithoutTheSimilarCosine</i>	84

4.2.13 Method <i>takeProportionWithTheSimilarCosineWithin24Hours</i>	86
4.2.14 Method <i>takeProportionWithTheSimilarCosineWithin7Days</i>	89
4.2.15 Method <i>takeProportionWithoutTheSimilarCosineWithin24Hours</i>	92
4.2.16 Method <i>takeProportionWithoutTheSimilarCosineWithin7Days</i>	95
4.2.17 Method <i>takeDeltaValueWithTheSimilarClusters</i>	98
4.2.18 Method <i>takeDeltaValueWithoutTheSimilar Clusters</i>	100
4.2.19 Method <i>averageDeltaValueWithTheSimilarCluster</i>	102
4.2.20 Method <i>averageDeltaValueWithoutTheSimilarCluster</i>	105
4.2.21 Method <i>takeProportionWithTheSimilarClusterWithin24Hours</i>	107
4.2.22 Method <i>takeProportionWithTheSimilarClusterWithin7Days</i>	110
4.2.23 Method <i>takeProportionWithoutTheSimilarClusterWithin24Hours</i> .	113
4.2.24 Method <i>takeProportionWithoutTheSimilarClusterWithin7Days</i>	116
4.3 Data Statistik	119
4.3.1 Data Statistik Umur	119
4.3.2 Data Statistik Pendidikan	120
4.3.3 Data Statistik Pekerjaan	121
4.3.4 Data Statistik Penghasilan	122
4.3.5 Data Statistik Provinsi	123
BAB 5 PENGUJIAN	126
5.1 Hasil Fitur Berdasarkan <i>Domain</i>	126
5.2 Hasil Fitur Berdasarkan <i>Cosine</i>	128
5.3 Hasil Fitur Berdasarkan <i>Clusters Inggris</i>	130
5.4 Hasil Fitur Berdasarkan <i>Clusters Mandarin</i>	131
5.5 Hasil 24 Fitur	133
BAB 6 SIMPULAN DAN SARAN	136
6.1 Simpulan	136

6.2 Saran	136
DAFTAR PUSTAKA.....	138



DAFTAR GAMBAR

Gambar 2.1 Grafik <i>Traditional Learning Algorithm</i>	6
Gambar 2.2 Grafik Perbandingan dengan <i>Small Neural Network (NN)</i>	6
Gambar 2.3 Grafik Perbandingan <i>Neutral Network (NN)</i> dengan <i>Algoritma Tradisional</i>	6
Gambar 2.4 <i>User Interface WEKA</i>	9
Gambar 2.5 Notasi Aktor	11
Gambar 2.6 Notasi <i>Use Case</i>	11
Gambar 2.7 Notasi Sistem.....	11
Gambar 2.8 Contoh <i>Use Case Diagram</i>	12
Gambar 2.9 Notasi <i>Action / Activity</i>	12
Gambar 2.10 Notasi <i>Initial Node</i>	12
Gambar 2.11 Notasi <i>Final Node</i>	13
Gambar 2.12 Notasi <i>Decision Node</i>	13
Gambar 3.1 <i>Use Case</i> Sistem Prediksi.....	14
Gambar 3.2 <i>Activity Diagram Transfer Data</i>	15
Gambar 3.3 <i>Activity Diagram Training Data</i>	16
Gambar 3.4 <i>Activity Diagram</i> Prediksi Konsistensi Akses <i>Web</i>	18
Gambar 3.5 <i>Method Take Delta Value</i> Dengan <i>Website Yang Sama</i>	19
Gambar 3.6 <i>Method Take Delta Value</i> Dengan <i>Website Yang Berbeda</i>	20
Gambar 3.7 <i>Method Average Delta Value</i> Dengan <i>Web Yang Sama</i>	21
Gambar 3.8 <i>Method Average Delta Value</i> Dengan <i>Website Berbeda</i>	22
Gambar 3.9 <i>Method Take Proportion 24 Jam</i> Dengan <i>Website Yang Sama</i>	23
Gambar 3.10 <i>Method Take Proportion 24 Jam</i> Dengan <i>Website Berbeda</i>	25
Gambar 3.11 <i>Method Take Proportion 7 Days</i> Dengan <i>Website Sama</i>	26
Gambar 3.12 <i>Method Take Proportion 7 Days</i> Dengan <i>Website Berbeda</i>	28
Gambar 3.13 <i>Method Take Delta Value</i> Dengan <i>Cosine Serupa</i>	29
Gambar 3.14 <i>Method Take Delta Value</i> Dengan <i>Cosine Berbeda</i>	30
Gambar 3.15 <i>Method Average Delta Value</i> Dengan <i>Cosine Serupa</i>	31
Gambar 3.16 <i>Method Average Delta Value</i> Dengan <i>Cosine Berbeda</i>	33

Gambar 3.17 <i>Method Proporsion</i> 24 Jam Dengan <i>Cosine</i> Serupa	34
Gambar 3.18 <i>Method Proporsion</i> 24 Jam Dengan <i>Cosine</i> Berbeda	35
Gambar 3.19 <i>Method Proporsion</i> 7 Hari Dengan <i>Cosine</i> Serupa	36
Gambar 3.20 <i>Method Proporsion</i> 7 Hari Dengan <i>Cosine</i> Berbeda	38
Gambar 3.21 <i>Method Take Delta Value</i> Dengan <i>Cluster</i> Serupa	39
Gambar 3.22 <i>Method Take Delta Value</i> Dengan <i>Cluster</i> Berbeda	40
Gambar 3.23 <i>Method Average Delta Value</i> Dengan <i>Cluster</i> Serupa.....	41
Gambar 3.24 <i>Method Average Delta Value</i> Dengan <i>Cluster</i> Berbeda.....	43
Gambar 3.25 <i>Method Proporsion</i> 24 Jam Dengan <i>Cluster</i> Serupa	44
Gambar 3.26 <i>Method Proporsion</i> 24 Jam Dengan <i>Cluster</i> Berbeda	45
Gambar 3.27 <i>Method Proporsion</i> 7 Hari Dengan <i>Cluster</i> Serupa.....	46
Gambar 3.28 <i>Method Proporsion</i> 7 Hari Dengan <i>Cluster</i> Berbeda.....	48
Gambar 3.29 Arsitektur Aplikasi.....	49
Gambar 4.1 Contoh Data Set.....	52
Gambar 4.2 Proses Pembacaan <i>File CSV</i>	53
Gambar 4.3 Proses Pengolahan Data Mentah menjadi <i>listUser</i>	54
Gambar 4.4 Proses penyaringan dari <i>listUser</i> menjadi <i>listUserGlobal</i>	55
Gambar 4.5 Proses Pembacaan Data Per <i>User</i> Pada <i>listUserGlobal</i>	56
Gambar 4.6 Proses Penyaringan Data Menjadi <i>listUserLocal</i>	56
Gambar 4.7 <i>Method Take Delta Value Web</i> Yang Sama	58
Gambar 4.8 <i>Method Take Delta Value</i> Beda <i>Web</i> Proses Tampung <i>User</i>	59
Gambar 4.9 <i>Method Take Delta Value</i> Beda <i>Web</i> Proses Pencarian dan Penyimpanan <i>Index Global</i>	60
Gambar 4.10 <i>Method Take Delta Value</i> Beda <i>Web</i> Proses Perhitungan Akhir	60
Gambar 4.11 <i>Method Average Delta Value Web</i> Yang Sama Proses Pencarian ..	61
Gambar 4.12 <i>Method Average Delta Value Web</i> Yang Sama Proses Perhitungan	62
Gambar 4.13 <i>Method Average Delta Value</i> Beda <i>Web</i> Proses Tampung <i>User</i>	63
Gambar 4.14 <i>Method Average Delta Value</i> Beda <i>Web</i> Proses Penampungan <i>IndexGlobal</i>	64
Gambar 4.15 <i>Method Average Delta Value</i> Beda <i>Web</i> Proses Perhitungan.....	64
Gambar 4.16 <i>Method Take Proportion Web</i> Sama 24 Jam Proses Penampungan <i>User</i>	65

Gambar 4.17 <i>Method Take Proportion Web Sama 24 Jam Proses Pencarian Index Global</i>	66
Gambar 4.18 <i>Method Take Proportion Web Sama 24 Jam Proses Perhitungan Proporsi</i>	67
Gambar 4.19 <i>Method Take Proportion Beda Web 24 Jam Proses Penampungan User</i>	68
Gambar 4.20 <i>Method Take Proportion Beda Web Proses Pencarian Index Global</i>	69
Gambar 4.21 <i>Method Take Proportion Beda Web Proses Perhitungan Proporsi</i> ..	70
Gambar 4.22 <i>Method Take Proportion Web Sama 7 Hari Proses Penampungan User</i>	71
Gambar 4.23 <i>Method Take Proportion 7 Hari Proses Pencarian Index Global</i>	72
Gambar 4.24 <i>Method Take Proportion Web Sama 7 Hari Proses Perhitungan Proporsi</i>	73
Gambar 4.25 <i>Method Take Proportion Beda Web 7 Hari Proses Penampungan User</i>	74
Gambar 4.26 <i>Method Take Proportion Beda Web 7 Hari Proses Pencarian Index Global</i>	75
Gambar 4.27 <i>Method Take Proportion Beda Web 7 Hari Proses Perhitungan Proporsi</i>	76
Gambar 4.28 <i>Method Take Delta Value Cosine Serupa Pencarian User</i>	77
Gambar 4.29 <i>Method Take Delta Value Cosine Serupa Proses Pencarian Index Global</i>	78
Gambar 4.30 <i>Method Take Delta Value Cosine Serupa Proses Perhitungan</i>	78
Gambar 4.31 <i>Method Take Delta Value Cosine Berbeda Proses Penampungan User</i>	79
Gambar 4.32 <i>Method Take Delta Value Cosine Berbeda Proses Pencarian Index Global</i>	80
Gambar 4.33 <i>Method Take Delta Value Cosine Berbeda Proses Perhitungan</i>	81
Gambar 4.34 <i>Method Average Delta Value Cosine Serupa Proses Pencarian User</i>	82

Gambar 4.35 <i>Method Average Delta Value Cosine</i> Serupa Proses Pencarian <i>Index Global</i>	83
Gambar 4.36 <i>Method Average Delta Value Cosine</i> Serupa Proses Perhitungan ..	83
Gambar 4.37 <i>Method Average Delta Value Cosine</i> Berbeda Proses Pencarian <i>User</i>	84
Gambar 4.38 <i>Method Average Delta Value Cosine</i> Berbeda Proses Pencarian <i>Index Global</i>	85
Gambar 4.39 <i>Method Average Delta Value Cosine</i> Berbeda Proses Perhitungan	86
Gambar 4.40 <i>Method Take Proportion 24 Jam Cosine</i> Serupa Proses Penampungan <i>User</i>	87
Gambar 4.41 <i>Method Take Proportion 24 Jam Cosine</i> Serupa Proses Pencarian <i>Index Global</i>	88
Gambar 4.42 <i>Method Take Proportion 24 Jam Cosine</i> Serupa Proses Perhitungan	88
Gambar 4.43 <i>Method Take Proportion 7 Hari Cosine</i> Serupa Proses Penampungan <i>User</i>	90
Gambar 4.44 <i>Method Take Proportion 7 Hari Cosine</i> Serupa Proses Pencarian <i>Index Global</i>	90
Gambar 4.45 <i>Method Take Proportion 7 Hari Cosine</i> Serupa Proses Perhitungan	91
Gambar 4.46 <i>Method Take Proportion 24 Jam Cosine</i> Berbeda Proses Penampungan <i>User</i>	92
Gambar 4.47 <i>Method Take Proportion 24 Jam Cosine</i> Berbeda Proses Pencarian <i>Index Global</i>	93
Gambar 4.48 <i>Method Take Proportion 24 Jam Cosine</i> Berbeda Proses Perhitungan	94
Gambar 4.49 <i>Method Take Proportion 7 Hari Cosine</i> Berbeda Proses Penampungan <i>User</i>	95
Gambar 4.50 <i>Method Take Proportion 7 Hari Cosine</i> Berbeda Proses Pencarian <i>Index Global</i>	96
Gambar 4.51 <i>Method Take Proportion 7 Hari Cosine</i> Berbeda Proses Perhitungan	97

Gambar 4.52 <i>Method Take Delta Value Cluster</i> Serupa Pencarian <i>User</i>	98
Gambar 4.53 <i>Method Take Delta Value Cluster</i> Serupa Proses Pencarian <i>Index Global</i>	99
Gambar 4.54 <i>Method Take Delta Value Cluster</i> Serupa Proses Perhitungan.....	99
Gambar 4.55 <i>Method Take Delta Value Cluster</i> Berbeda Proses Penampungan <i>User</i>	100
Gambar 4.56 <i>Method Take Delta Value Cluster</i> Berbeda Proses Pencarian <i>Index Global</i>	101
Gambar 4.57 <i>Method Take Delta Value Cluster</i> Berbeda Proses Perhitungan...	102
Gambar 4.58 <i>Method Average Delta Value Cluster</i> Serupa Proses Pencarian <i>User</i>	103
Gambar 4.59 <i>Method Average Delta Value Cluster</i> Serupa Proses Pencarian <i>Index Global</i>	104
Gambar 4.60 <i>Method Average Delta Value Cluster</i> Serupa Proses Perhitungan	104
Gambar 4.61 <i>Method Average Delta Value Cluster</i> Berbeda Proses Pencarian <i>User</i>	105
Gambar 4.62 <i>Method Average Delta Value Cluster</i> Berbeda Proses Pencarian <i>Index Global</i>	106
Gambar 4.63 <i>Method Average Delta Value Cluster</i> Berbeda Proses Perhitungan	107
Gambar 4.64 <i>Method Take Proportion 24 Jam Cluster</i> Serupa Proses Penampungan <i>User</i>	108
Gambar 4.65 <i>Method Take Proportion 24 Jam Cluster</i> Serupa Proses Pencarian <i>Index Global</i>	109
Gambar 4.66 <i>Method Take Proportion 24 Jam Cluster</i> Serupa Proses Perhitungan	109
Gambar 4.67 <i>Method Take Proportion 7 Hari Cluster</i> Serupa Proses Penampungan <i>User</i>	111
Gambar 4.68 <i>Method Take Proportion 7 Hari Cluster</i> Serupa Proses Pencarian <i>Index Global</i>	112
Gambar 4.69 <i>Method Take Proportion 7 Hari Cluster</i> Serupa Proses Perhitungan	112

Gambar 4.70 <i>Method Take Proportion 24 Jam Cluster Berbeda Proses Penampungan User</i>	114
Gambar 4.71 <i>Method Take Proportion 24 Jam Cluster Berbeda Proses Pencarian Index Global</i>	115
Gambar 4.72 <i>Method Take Proportion 24 Jam Cluster Berbeda Proses Perhitungan</i>	115
Gambar 4.73 <i>Method Take Proportion 7 Hari Cluster Berbeda Proses Penampungan User</i>	117
Gambar 4.74 <i>Method Take Proportion 7 Hari Cluster Berbeda Proses Pencarian Index Global</i>	118
Gambar 4.75 <i>Method Take Proportion 7 Hari Cluster Berbeda Proses Perhitungan</i>	118
Gambar 4.76 Statistik Umur <i>User</i>	120
Gambar 4.77 Statistik Edukasi <i>User</i>	121
Gambar 4.78 Statistik Pekerjaan <i>User</i>	122
Gambar 4.79 Statistik Penghasilan <i>User</i>	123
Gambar 4.80 Statistik Provinsi <i>User</i>	124

DAFTAR TABEL

Tabel 4.1 Tabel Data Umur <i>User</i>	120
Tabel 4.2 Tabel Data Edukasi <i>User</i>	121
Tabel 4.3 Tabel Data Pekerjaan <i>User</i>	122
Tabel 4.4 Tabel Data Penghasilan <i>User</i>	123
Tabel 4.5 Tabel Data Provinsi <i>User</i>	125
Tabel 5.1 Hasil Fitur Berdasarkan Domain	126
Tabel 5.2 Tabel Ranking <i>Domain</i> Dengan Feature Selection	127
Tabel 5.3 Hasil Fitur Berdasarkan <i>Cosine</i>	128
Tabel 5.4 Tabel Ranking <i>Cosine</i> Dengan Feature Selection.....	129
Tabel 5.5 Hasil Fitur Berdasarkan <i>Clusters</i>	130
Tabel 5.6 Tabel Ranking Kluster Dengan <i>Feature Selection</i>	131
Tabel 5.7 Hasil Fitur Berdasarkan <i>Clusters</i> Mandarin	131
Tabel 5.8 Tabel Ranking Kluster Dengan <i>Feature Selection</i>	133
Tabel 5.9 Hasil 24 Fitur	134
Tabel 5.10 <i>Feature Selection</i> 24 Fitur.....	135

DAFTAR SINGKATAN

URL	<i>Uniform Resource Locator</i>
CSV	<i>Comma Separated Values</i>

