

**PROTOTYPE SISTEM INFORMASI PREDIKSI
PENEMPATAN KARIR ALUMNI MENGGUNAKAN
ALGORITMA NAIVE BAYES
Studi Kasus : BINUS Career Center**



**PROGRAM STUDI MAGISTER ILMU KOMPUTER (MKOM)
PROGRAM PASCASARJANA
UNIVERSITAS BUDI LUHUR
JAKARTA 2016**



PROGRAM STUDI MAGISTER ILMU KOMPUTER
PROGRAM PASCASARJANA
UNIVERSITAS BUDI LUHUR

LEMBAR PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : **ROHIMAH**
NIM : **1311600918**
Konsentrasi : **TEKNOLOGI SISTEM INFORMASI**
Jenjang Studi : **STRATA-2**
Menyatakan bahwa Tesis yang berjudul:

**PROTOTYPE SISTEM INFORMASI PREDIKSI PENEMPATAN KARIR
ALUMNI MENGGUNAKAN ALGORITMA NAIVE BAYES.**

Studi Kasus : BINUS Career Center

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

Pernyataan ini saya buat dengan penuh tanggung jawab dan saya bersedia menerima konsekuensi apapun sesuai aturan yang berlaku apabila di kemudian hari pernyataan ini tidak benar.

Jakarta, Januari 2016



(ROHIMAH)



PROGRAM STUDI: MAGISTER ILMU KOMPUTER
PROGRAM PASCASARJANA
UNIVERSITAS BUDI LUHUR

LEMBAR PENGESAHAN TESIS

Nama Mahasiswa : Rohimah
Nomor Induk Mahasiswa : 1311600918
Konsentrasi : Teknologi Sistem Informasi
Jenjang Studi : Strata-2
Judul : **Prototipe Sistem Informasi Prediksi
Penempatan Karir Alumni Menggunakan
Algoritma Naive Bayes. Studi Kasus : Binus
Career Center**

Jakarta, Januari 2016

Tim Penguji :

Tanda Tangan :

Ketua,

(Prof. Dr. Moedjiono, M.Sc)

Anggota,

(Dr., Ir. Nazori Az, M.T)

Pembimbing Utama,

(Arief Wibowo, M.Kom)

Ketua Program Studi

(Dr., Ir. Nazori Az, M.T)

ABSTRAK

Jumlah lulusan perguruan tinggi di Indonesia setiap tahunnya, mulai dari lulusan diploma, sarjana, magister dan dokter sangatlah banyak. Namun, selama ini tidak banyak perguruan tinggi yang memiliki pemetaan jejak lulusan alumninya. Kondisi ini mengakibatkan hampir tidak ada umpan balik dari lulusan untuk memperbaiki kualitas pendidikan perguruan tinggi itu sendiri. Dan Kualitas kompetensi pekerja saat ini tidak memenuhi kriteria pasar tenaga kerja dan rendahnya tingkat produktivitas tenaga kerja, lemahnya komunikasi antara pasar tenaga kerja pendidikan, perubahan struktur sosial ekonomi dan politik global sangatlah berpengaruh besar di pasar tenaga kerja. Pelacak hasil studi dapat digunakan oleh perguruan tinggi untuk menentukan keberhasilan proses pendidikan yang telah dilakukan terhadap anak didiknya. Oleh karena itu, universitas perlu teknologi untuk mendukung optimalisasi penggunaan pelacak studi. Untuk menjawab semua itu Teknologi Data Mining hadir sebagai solusi. Pada penelitian ini, akan dilakukan klasifikasi berdasarkan tipe perusahaan tempat alumni bekerja dengan menggunakan 100 data alumni yang dibagi kedalam 2 kelompok, yaitu 70 data training dan 30 data testing Metode yang akan digunakan adalah Naive Bayes untuk kemudian dapat disimpulkan menjadi aturan-aturan dan diimplementasikan dalam suatu prototipe sistem informasi pemetaan alumni bekerja. Dengan algoritma Naive Bayes dihasilkan tingkat akurasi sebesar 77.14%. Hasil dari proses klasifikasi dievaluasi dengan menggunakan cross validation sebesar 0.489, confusion matrix sebesar 31 jumlah angka benar dalam proses klasifikasi, ROC Curve sebesar 0.933 dan testing program untuk mengetahui tingkat akurasi yang dihasilkan yang paling akurat untuk prediksi lulusan bekerja.

Kata Kunci: Tipe Perusahaan, Klasifikasi, Data Mining, Naive Bayes, K-Fold Cross Validation.

ABSTRAC

Graduates of institutions of higher education in Indonesia holding associate degree, bachelor's degree, master's degree, and doctoral degree are growing significantly in numbers. However, many of said institutions of higher education have not yet been equipped with a graduate survey or tracer study system. Consequently, this situation has led to a condition of virtually nonexistent feedbacks from the graduates on enhancing the quality of the higher education itself. Meanwhile, in the contemporary setting, the competences of existing workers have failed to meet the labor-market criteria. Additionally, low productivity rate of workers as well as poor communications among the actors in the labor market of education sector, compounded with global socioeconomic and geopolitical changes at the structural level are shown to be highly influential to the changing status of the labor market. A tracer study for college students is useful and can be utilized by institutions of higher education in order to determine the success of the educational processes of which the graduates had undergone. Therefore, higher education institutions need to embrace the technologies to support the optimization of the utilization of a tracer study. Data Mining technology is one of the solutions. In this research, classifications that are based upon the types of company in which the alumni work will be made by using the data of 100 alumni divided into two groups: 70 training data, and 30 testing data. The method that is going to be used in this research is Naive Bayes, which is going to be summarized as rules and is going to be implemented in a working alumni tracking information system prototype. By utilizing Naive Bayes, an accuracy rate of 77.14% was reached. The results of the classification processes are evaluated by using cross validation rate of 0.489, confusion matrix with precision rate of 31 correctly classified data, ROC Curve rate of 0.933, and program testing to determine the most precise accuracy rate for the purpose of generating effective alumni predictions.

Keywords: Company Types, Classifications, Data Mining, Naive Bayes, K-Fold Cross Validation

KATA PENGANTAR

Subhanallah Walhamdulillah Walaillahaillah Allahuakbar, puji syukur peneliti panjatkan kepada Allah SWT karena atas berkah dan rahmatnyalah penelitian proposal tesis dengan judul “Prototipe Sistem Informasi Pemetaan Alumni Bekerja Menggunakan Algoritma Naive Bayes, Studi Kasus BINUS Career Center” ini dapat terselesaikan dengan baik. Proposal tesis ini disusun dengan maksud untuk memenuhi salah satu persyaratan dalam memperoleh gelar Magister Komputer.

Dalam penyusunan proposal tesis ini, peneliti menyampaikan terima kasih yang tulus kepada:

1. Kepada Almarhum ayahanda tercinta yang selalu menginspirasi saya untuk dapat melanjutkan kejenjang S2, yang merupakan cita-cita beliau semasa hidupnya. Semoga beliau bisa tersenyum bangga dengan ini semua dan semoga Alloh SWT mengampuni semua dosa-dosa beliau, menerima amal ibadah beliau, dan memberikan beliau tempat yang terbaik disisi Allah SWT.
2. Kepada ibunda tercinta, kakak-kakakku tersayang yang telah menginspirasi sekaligus pemberi semangat kepada peneliti untuk dapat menyelesaikan proposal tesis ini.
3. Bapak Arief Wibowo, S.Kom, M.Kom selaku dosen pembimbing tesis yang telah membimbing dan memotivasi peneliti dalam menyelesaikan tesis ini.
4. Teman – teman, sahabat dan pihak – pihak lain yang tidak bisa peneliti sebutkan satu-persatu. Saya ucapkan terima kasih dan penghargaan yang sebesar – besarnya.

Peneliti menyadari bahwa penelitian tesis ini jauh dari sempurna. Untuk itu peneliti berharap mendapatkan kritik dan saran yang dapat berguna untuk membangun demi kesempurnaan penelitian tesis.

Jakarta, Januari 2016

Rohimah

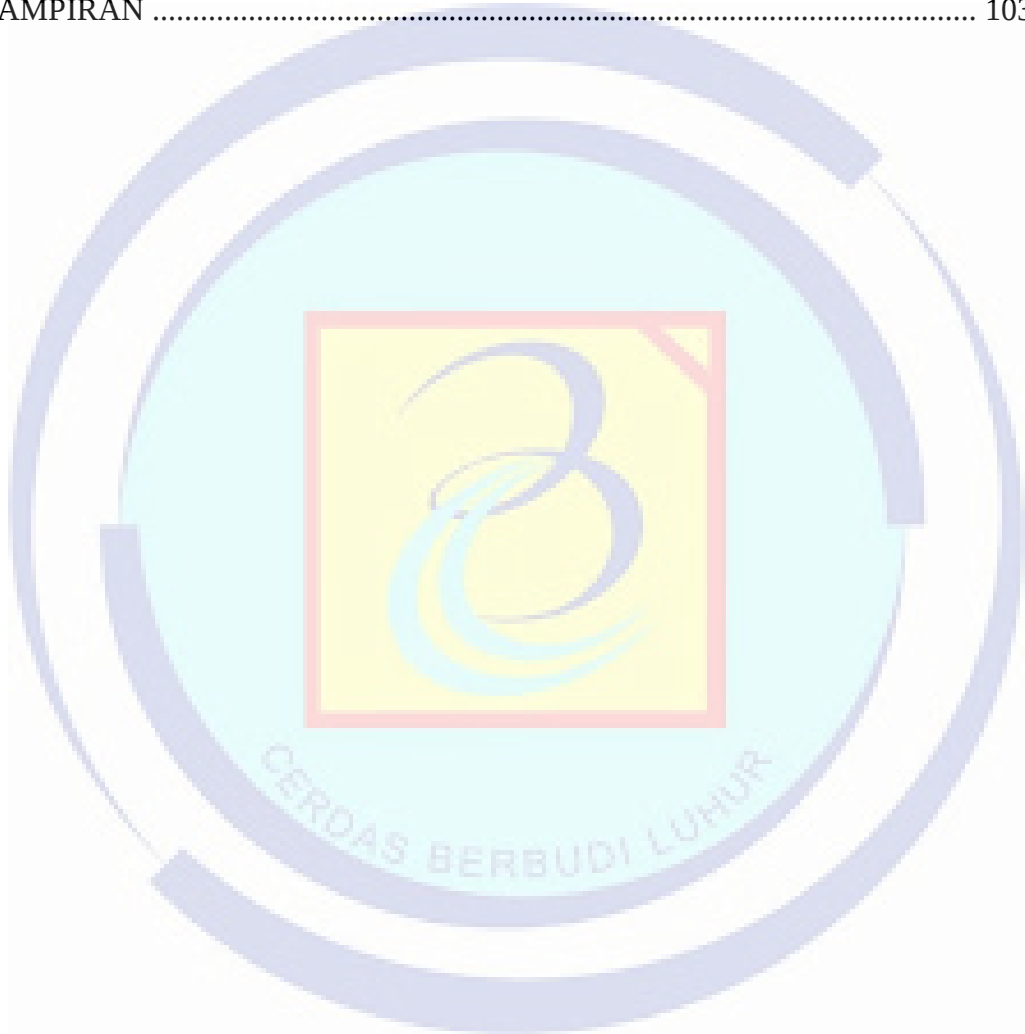
DAFTAR ISI

ABSTRAK	iii
KATA PENGANTAR	v
DAFTAR ISI.....	vi
DAFTAR GAMBAR	x
DAFTAR TABEL.....	xii
DAFTAR LAMPIRAN.....	xiii
BAB I PENDAHULUAN.....	1
1.2 Masalah Penelitian.....	2
1.2.1 Identifikasi Masalah.....	2
1.2.2 Batasan Masalah.....	2
1.2.3 Rumusan Masalah.....	2
1.3 Tujuan dan Manfaat Penelitian.....	3
1.3.1 Tujuan Penelitian	3
1.3.2 Manfaat Penelitian	3
1.4 Tata Urut Penulisan	3
1.5 Daftar Pustaka	4
BAB II LANDASAN TEORI DAN KERANGKA PEMIKIRAN.....	6
2.1 Tinjauan Pustaka	6
2.1.1 Data Mining.....	6
2.1.1.1 Definisi Data Mining	6
2.1.1.2 Kelompok Data Mining	8
2.1.1.3 Metode Klasifikasi Data Mining	9
2.1.2 Bayesian Classifier.....	13
2.1.3 K-Nearest Neighbor.....	14

2.1.4 C4.5	16
2.1.5 Evaluasi dan Validasi Metode klasifikasi Data Mining	17
2.1.5.1 K-Fold Cross Validation.....	18
2.1.6 Evaluasi Model	19
2.1.6.1 Pengujian Model.....	19
2.1.7 Slovin.....	19
2.1.5.1 Metode Slovin.....	20
2.1.6 WEKA	21
2.1.6.1 Format Input WEKA	22
2.1.6.2 Algoritma <i>Naive Bayes</i> pada WEKA	23
2.2 Tinjauan Studi	23
2.3 Tinjauan objek penelitian	34
2.3.1 Gambaran Umum BINUS University	34
2.3.2 Visi BINUS University	34
2.2.3 Misi BINUS University	34
2.3.4 Struktur Organisasi	35
2.4 Perangkat Keras	36
2.5 Kerangka Konsep	36
2.6 Hipotesis	38
BAB III DESAIN PENELITIAN.....	39
3. 1 Metode Penelitian	39
3. 2 Metode Pemilihan Sample.....	39
3. 3 Metode Pengumpulan Data	40
3. 4 Pengolahan Data Awal	41
3.5 Instrumentasi	41
3.6 Teknik Analisa, Perancangan dan Pengujian	42

3.6.1 Teknik Analisa Data	42
3.6.2 Teknik Pengujian Model	44
3.8 Langkah-Langkah Penelitian	45
3.9 Penjelasan Bagan Penelitian	46
3.10Jadwal Penelitian	47
BAB IV PEMBAHASANA DAN PENELITIAN	49
4.1 Pengelompokan dan Analisis.....	49
4.1.1 Pengelompokan Data.....	49
4.1.2 Perhitungan Algoritma C4.5	54
4.1.3 Perhitungan Algoritma K-Nearest Neighbor	57
4.1.4 Perhitungan Algoritma Naive Bayes	62
4.2 Evaluasi dan Validasi	66
4.2.1 Confussion Matrix algoritma C4.5.....	66
4.2.2 Kurva ROC/AUC algoritma C4.5.....	67
4.2.3 Confussion Matrix algoritma K-Nearest Neighbor.....	68
4.2.4 Kurva ROC/AUC algoritma K-Nearest Neighbor	69
4.2.5 Confussion Matrix algoritma Naive Bayes	69
4.2.6 Kurva ROC/AUC algoritma Naive Bayes	70
4.2.7 Cross Validation.....	71
4.2.8 Analisa Evaluasi Kompasari Model	72
4.2.9 Penerapan Hasil Algoritma Terpilih.....	72
4.2.10 Logical Design.....	75
4.3 Implikasi Penelitian	94
4.3.1 Aspek Sistem.....	94
4.3.2 Aspek Manajerial	94
4.3.3 Aspek Penelitian Lanjut	95

4.3.4 Rencana Implementasi	95
BAB V PENUTUP.....	98
5.1 Kesimpulan.....	98
5.2 Saran	98
DAFTAR PUSTAKA	99
LAMPIRAN	103



DAFTAR GAMBAR

Gambar	Halaman
II.1 Knowledge Discovering in Database	6
II.2 Ilustrasi Kasus Algoritma K-Nearest Neighbor	15
II.3 Slovin	20
II.4 Tampilan awal GUI WEKA	22
II.5 Struktur Organisasi	34
II.6 Kerangka Konsep	36
III.1 Proses Cross Validation Industries for Data Mining (CRISP-DM).....	41
III.2 Tahap-tahap Penelitian	44
IV.1 Hasil WEKA, C4.5	56
IV.2 Hasil WEKA, K-Nearest Neighbor	62
IV.3 Hasil WEKA, Naive Bayes	66
IV.4 Confussion Matrix Algoritma C4.5	66
IV.5 ROC/AUC Algoritma C4.5	67
IV.6 Confussion Matrix Algoritma K-Nearest Neighbor	68
IV.7 ROC/AUC Algoritma K-Nearest Neighbor.....	69
IV.8 Confussion Matrix Algoritma Naive Bayes	69
IV.9 ROC/AUC Algoritma Naive Bayes	70
IV.10 Cross Validation C4.5	70
IV.11 Cross Validation K-Nearest Neighbor	71
IV.12 Cross Validation Naive Bayes	71
IV.13 Hasil WEKA, Data Testing	74
IV.14 ROC/AUC Data Training	74
IV.15 Cross Validation Data Testing	75
IV.16 Use Case Admin	76
IV.17 Use Case Staff	76
IV.18 Diagram Aliran Data	77
IV.19 Class Diagram	78
IV.20 Menu Navigasi	81
IV.21 Tampilan Dashboard	81

IV.22 Tampilan Upload Training	82
IV.23 Tampilan List Data Training	82
IV.24 Tampilan Input Prediksi	83
IV.25 Tampilan Hasil Prediksi	83
IV.26 Tampilan Hasil Prediksi	84
IV.27 Tampilan Rules Prediksi Root Class	84
IV.28 Tampilan Rules Jenis Kelamin	85
IV.29 Tampilan Rules Usia	85
IV.30 Tampilan Rules Jurusan	86
IV.31 Tampilan Rules Lama Studi	86
IV.32 Tampilan Rules IPK	87
IV.33 Tampilan Rules SKS	87
IV.34 Tampilan Rules Jumlah Mata Kuliah Mengulang	88
IV.35 Tampilan Rules SMA/SMEA	88
IV.36 Tampilan Rules IPA/IPS	89
IV.34 Tampilan Rules Perusahaan	89
IV.35 Report	90
IV.36 Rencana Jaringan.....	93

DAFTAR TABEL

Tabel	Halaman
II.1 Studi Literatur Jurnal Pilihan	26
II.2 Spesifikasi Hardware	35
III.1 Data Alumni 4 tahun terakhir	39
III.2 Data Atribut Alumni	42
III.3 Rencana Jadwal Penelitian	45
IV.1 Sampel Data Training	50
IV.2 Perhitungan Menentukan Entropy dan Gain	55
IV.3 Data Training Lulusan Bekerja	57
IV.4 Data Numerik Lulusan Bekerja	58
IV.5 Data Testing K-Nearest Neighbor	59
IV.6 Komparasi	72
IV.7 Data Testing	72
IV.8 File Training	78
IV.9 File Prediski	79
IV.10 File Attribute	79
IV.11 File Classifier	79
IV.12 File Result	80
IV.13 File Admin/Staff	80
IV.14 File Config	80
IV.15 File Field	80
IV.16 Employment Rate Report	91
IV.16 Rencana Implementasi Sistem Pemetaan Lulusan Bekerja	95
IV.17 Hardware dan Software	97

DAFTAR LAMPIRAN

Lampiran	Halaman
Daftar Pertanyaan Survey Lulusan Bekerja	101
Data Training	103
Data Testing	108



DAFTAR PUSTAKA

- [Moedjiono 2010] Pedoman Penyusunan Tesis V4.2
- [Aprilla, D., Wicaksana, I. W. S., Baskoro, D. A., & Ambarwati, L. 2013] *Belajar Data Mining Dengan Weka*. (R. Sanjaya, Ed.). Jakarta.
- [Bramer, M. 2012] *Principles of Data Mining*. (I. Mackie, Ed.) *Saudi Med J* (Vol. 33). London, UK: Springer. <http://doi.org/10.1073/pnas.0703993104>
- [Brank, J., Mladenic, D., Grobelnik, M., & Milic-Frayling, N. 2008] Feature Selection for the Classification of Large Document Collections. *J. Ucs*, 14(10), 1562–1596. Retrieved from http://jucs.org/jucs_14_10/feature_selection_for_the/jucs_14_10_1562_1596_brank.pdf
- [Gorunescu, F. (Ed.). 2010] *Data Mining Concepts, Model and Techniques*. Springer.
- [Hall, M. a, & Smith, L. a. 1998] Feature subset selection: A correlation based filter approach. *Progress in Connectionist-Based Information Systems, Vols 1 and 2*, 855–858. Retrieved from <Go to ISI>://WOS:000085593600208
- [Han, Jiawei, Kamber, M. 2012] *Data Mining Concepts and Techniques* (Third Edit). Waltham, USA: Morgan Kaufmann, Elsevier.
- [Kusrini, Luthfi, E. T. 2009] *Algooritma Data Mining*. (T. A. Prabawati, Ed.). Yogyakarta.
- [Larose, D. T. 2005] *Discovering Knowledge In Data*.
- [Maimon, O., & Rokach, L. 2010] *Data Mining and Knowledge Discovery Handbook*. (O. Maimon & L. Rokach, Eds.) (Second Edi). Israel: Springer.
- [Mitchell, T. M. 2006] The Discipline of Machine Learning. *Machine Learning*, 17(July), 1–7. <http://doi.org/10.1080/026404199365326>
- [Schomburg, H. 2003] *Handbook for Graduate Tracer Studies*. Germany: Deutsche Gesellschaft Fur Technische Zusammenarbeit (GTZ) GmbH.
- [Shang, W., Huang, H., Zhu, H., Lin, Y., Qu, Y., & Wang, Z. 2007] A novel feature selection algorithm for text categorization. *Expert Systems with Applications*, 33(1), 1–5. <http://doi.org/10.1016/j.eswa.2006.04.001>
- [Smith, E. P., Lipkovich, I., & Ye, K. 2002] Weight of Evidence (WOE): Quantitative estimation of probability of impact. *Blacksburg, VA: Virginia*