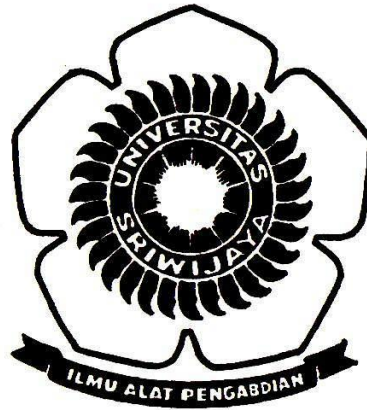


**PENERAPAN ALGORITMA RELIEF-F UNTUK *FEATURE SELECTION*
PADA PREDIKSI KESESUAIAN TINGKAT PENDIDIKAN DENGAN
BIDANG PEKERJAAN PADA ALUMNI UNIVERSITAS SRIWIJAYA**

SKRIPSI

**Sebagai Salah Satu Syarat Memperoleh Gelar
Sarjana Sains Bidang Studi Matematika**



Oleh

**KERENILA AGUSTIN
NIM 08011181520024**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SRIWIJAYA
OKTOBER 2019**

Lembar Pengesahan

**PENERAPAN ALGORITMA RELIEF-F UNTUK *FEATURE SELECTION*
PADA PREDIKSI KESESUAIAN TINGKAT PENDIDIKAN DENGAN
BIDANG PEKERJAAN PADA ALUMNI UNIVERSITAS SRIWIJAYA**

SKRIPSI

**Sebagai Salah Satu Syarat Memperoleh Gelar
Sarjana Sains Bidang Studi Matematika**

Oleh

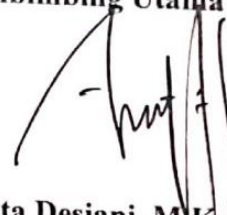
**KERENILA AGUSTIN
NIM.08011181520024**

Pembimbing Kedua



**Drs. Sugandi Yahdin, M.M
NIP. 19580727 198603 1 003**

**Indralaya, Oktober 2019
Pembimbing Utama**



**Anita Desiani, M.Kom
NIP. 19771211 200312 2 002**

**Mengetahui,
Ketua Jurusan Matematika**



**Drs. Sugandi Yahdin, M.M
NIP. 19580727 198603 1 003**

LEMBAR PERSEMBAHAN

Motto

”Khoirunnas anfa'uhum linnas”

Artinya : Sebaik-baik manusia adalah yang paling bermanfaat bagi manusia lain.

(HR Ahmad dan Thabrani)

“PRAY”

not only because you need something but because you have a lot to be thankful for

(Anonim)

Skripsi ini kupersembahkan kepada :

1. Allah SWT
2. Kedua Orang Tuaku
3. Saudaraku
4. Keluarga Besarku
5. Para Pemberi Ilmu
6. Sahabat-sahabatku
7. Almamaterku

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh.

Alhamdulillah penulis ucapkan syukur kepada Allah SWT karena berkat rahmat, karunia, kasih sayang dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul **"Penerapan *Feature Selection* Algoritma Relief-f dalam Meningkatkan Akurasi Prediksi Kesesuaian Tingkat Pendidikan dengan Bidang Pekerjaan pada Alumni Universitas Sriwijaya"**. Sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains Bidang Studi Matematika di Fakultas Matematika dan Ilmu Pengeratahuan Alam Universitas Sriwijaya.

Dengan segala hormat dan kerendahan hati penulis mengucapkan terima kasih yang sebesar-besarnya kepada Ayahanda Suseno dan Ibunda Misti Rahayu serta Saudara-saudaraku Sony Suryana, Feri Fernando, dan Trialin Safitri atas segala kasih sayang, perhatian, dukungan, serta doanya. Penulisan skripsi ini tidak lepas bantuan sebagai pihak baik secara langsung maupun tidak langsung, untuk itu penulis menyampaikan terimakasih sebesar-besarnya kepada :

1. Bapak Drs. Sugandi Yahdin, M.M, selaku Ketua Jurusan Matematika dan Ibu Des Alwine Zayanti, M.Si selaku Sekertaris Jurusan Matematika atas bimbingan yang telah diberikan selama penulis belajar di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya.
2. Ibu Anita Desiani, M.Kom selaku Pembimbing Utama dan Bapak Drs. Sugandi Yahdin, M.M selaku Pembimbing Kedua. Terima kasih telah bersedia

menyediakan waktu, pikiran, nasehat, motivasi dan saran serta kesabaran memberikan arahan dan bimbingan terbaik penulis dalam penyusunan skripsi ini.

3. Ibu Eka Susanti, M.Si, Ibu Irmeilyana, M.Si, dan Ibu Indrawati, M.Si selaku Dosen Penguji yang telah bersedia meluangkan waktu dalam memberikan tanggapan, kritik dan saran yang bermanfaat dalam perbaikan dan penyelesaian skripsi ini, serta Drs. Ali Amran, M.T selaku Dosen Pembimbing Akademik yang telah memberikan pengarahan akademik kepada penulis di Jurusan Matematika FMIPA Universitas Sriwijaya
4. Seluruh Dosen di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya yang telah memberkan ilmu yang bermanfaat, nasehat, serta bimbingan selama penulis menjalani perkuliahan.
5. Bapak Irwansyah selaku admin dan Ibu Hamidah selaku pegawai tata usaha jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam yang telah membantu penulis selama masa perkuliahan.
6. Pihak *Career Development Center* (CDC) Universitas Sriwijaya terima kasih telah memberikan kesempatan kepada penulis memberikan data kuisinoer penelitian.
7. Sahabat-sahabat penulis di bangku perkuliahan Mba Vinda, Eka, Wili, Annisa, Yusti, Novika, Shaly, Vidya, Ria, Indah, Elsa, Feren, Marnita, Budi, Febrizal dan Seluruh teman-teman angkatan 2015.

8. Adik-adik Elsa V, Apri, Ririz, Kariah, Oliv, Khairanil, Enyta, Aisyah Lian, Eling, Anita terima kasih atas semangat, do'a, motivasi dan dukungan yang telah diberikan.
9. Teman-teman Organisasi LDF KOSMIC, DPM KM FMIPA, dan DPM KM UNSRI untuk semangat, pengalaman serta bantuan yang diberikan kepada penulis.
10. Semua Pihak yang tidak bisa disebutkan satu persatu yang telah memberikan bantuan, do'a, serta semangat dalam menyelesaikan skripsi ini.

Penulis mengharapkan semoga skripsi ini dapat menambah pengetahuan dan bermanfaat bagi mahasiswa/mahasiswi Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya.

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

Indralaya, Oktober 2019

Penulis

APPLICATION OF RELIEF-F ALGORITHM FOR FEATURE SELECTION ON THE PREDICTION OF EDUCATION LEVELS WITH THE FIELD OF WORK AT ALUMNI SRIWIJAYA UNIVERSITY

By:

Kerenila Agustin
08011181520024

ABSTRACT

Sriwijaya University has a talent and career development center for alumni. The *Career Development Center* (CDC) website provides a questionnaire form for alumni to fill out. The features on the questionnaire form have relevance about the level of education of alumni with their fields of work such as lectures, research projects, internships, practicum, field work, discussions, time to get a job, fields of science, outside the field of science, English, internet, computers, and the relationship of study programs with work. Based on data from the CDC in 2013 to 2015 there were 1143 alumni of Sriwijaya University who filled out the questionnaire form completely, but there were 1019 alumni who had jobs. *Feature Selection* Relief-f algorithm is a feature selection algorithm based on the amount of data or *records*. Feature selection is done by calculating the weight difference for each data chosen at random with data selected as *near hit* (nearest neighbor selected data in the same class) and *near miss* (nearest neighbor selected data in different classes). This study uses the *k-fold cross-validation* of the naive bayes and KNN methods to see the success rate of the *feature selection function* Relief-f. The accuracy results of the data before *feature selection* process was 73.43% for the naive bayes method and 66.24% for the KNN method, after *feature selection* process increased to 74.38% for the naive bayes method and 72.22% for the KNN method. The best features were selected as many as 8 features, namely the relationship of study programs with work, science, English, research projects, outside the field of science, field work, internships, and discussions. From the accuracy obtained it can be concluded that the *feature selection* Relief-f algorithm works well in the feature selection process and improves accuracy.

Keywords: *feature selection*, Relief-f, education level

PENERAPAN ALGORITMA RELIEF-F UNTUK *FEATURE SELECTION* PADA PREDIKSI KESESUAIAN TINGKAT PENDIDIKAN DENGAN BIDANG PEKERJAAN PADA ALUMNI UNIVERSITAS SRIWIJAYA

Oleh :

**Kerenila Agustin
08011181520024**

ABSTRAK

Universitas Sriwijaya memiliki pusat pengembangan bakat dan karir bagi alumni. pada website *Career Development Centre* (CDC) menyediakan formulir kuisioner untuk diisi alumni. Fitur-fitur pada formulir kuisioner tersebut memiliki relevansi tentang tingkat pendidikan alumni dengan bidang pekerjaan mereka seperti perkuliahan, proyek riset, magang, praktikum, kerja lapangan, diskusi, waktu mendapatkan pekerjaan, bidang ilmu, diluar bidang ilmu, bahasa inggris, internet, komputer, dan hubungan program studi dengan pekerjaan. Berdasarkan data pada CDC tahun 2013 sampai 2015 terdapat 1143 alumni Universitas Sriwijaya yang mengisi formulir kuisioner dengan lengkap, tetapi alumni yang memiliki pekerjaan sebanyak 1019 alumni. *Feature Selection* algoritma Relief-f merupakan algoritma pemilihan fitur yang berbasis pada banyaknya data atau *record*. Pemilihan fitur dilakukan dengan menghitung perbedaan bobot untuk tiap data yang terpilih secara acak dengan data yang terpilih sebagai *near hit* (tetangga terdekat data terpilih pada kelas yang sama) dan *near miss* (tetangga terdekat data terpilih pada kelas yang berbeda). Penelitian ini menggunakan *k-fold cross-validation* metode naive bayes dan KNN untuk melihat tingkat keberhasilan dari *feature selection* algoritma Relief-f. Hasil akurasi pada data sebelum proses *feature selection* sebesar 73,43% untuk metode naive bayes dan 66,24% untuk metode KNN, setelah proses *feature selection* meningkat menjadi 74,38% untuk metode naive bayes dan 72,22% untuk metode KNN. Fitur terbaik yang terpilih sebanyak 8 fitur yaitu hubungan program studi dengan pekerjaan, bidang ilmu, bahasa inggris, proyek riset, luar bidang ilmu, kerja lapangan, magang, dan diskusi. Berdasarkan akurasi yang diperoleh dapat disimpulkan bahwa *feature selection* algoritma Relief-f bekerja dengan baik dalam proses pemilihan fitur dan meningkatkan akurasi.

Kata Kunci: *feature selection*, Relief-f, tingkat pendidikan

DAFTAR ISI

	Halaman
HALAMAN JUDUL	
HALAMAN PENGESAHAN	i
HALAMAN PERSEMBAHAN	ii
KATA PENGANTAR	iii
ABSTRACT	vi
ABSTRAK	vii
DAFTAR ISI	viii
DAFTAR TABEL	x
DAFTAR	
BAB I PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Pembatasan Masalah	3
1.4 Tujuan	4
1.5 Manfaat	4
BAB II TINJAUAN PUSTAKA	
2.1 <i>Feature Selection</i>	6
2.2 Algoritma Relief-f.....	6
2.3 Klasifikasi	9
2.4 Evaluasi Metode <i>Feature Selection</i>	9
BAB III METODOLOGI PENELITIAN	
3.1 Tempat.....	12
3.2 Waktu	12

3.2 Alat.....	12
3.3 Metode Penelitian.....	12
 BAB IV HASIL DAN PEMBAHASAN	
4.1 Deskripsi Data.....	15
4.2 Proses <i>Feature Selection</i> Algoritma Relief-f.....	19
4.2.1 <i>Feature Selection</i> Algoritma Relief-f dengan perhitungan manual	19
4.2.2 <i>Feature Selection</i> Algoritma Relief-f pada Data Penelitian	30
4.3 Evaluasi Metode <i>Feature Selection</i>	31
4.3.1 Evaluasi pada Data Sebelum <i>Feature Selection</i>	33
4.3.2 Evaluasi pada Data Hasil <i>Feature Selection</i>	43
4.4 Analisis Hasil dan Kesimpulan	52
 BAB V KESIMPULAN	
5.1 Kesimpulan	55
5.2 Saran.....	56
 DAFTAR PUSTAKA	 57

DAFTAR TABEL

	Halaman
Tabel 2.1 Model <i>Confussion Matrix</i>	10
Tabel 2.2 Kategori Nilai Akurasi	11
Tabel 4.1 Keterangan Fitur.....	15
Tabel 4.2 Data Nilai Masing-Masing Fitur	17
Tabel 4.3 Data Contoh untuk Perhitungan Manual.....	19
Tabel 4.4 Nilai Bobot F2 pada Setiap Iterasi	24
Tabel 4.5 Nilai Bobot F5 pada Setiap Iterasi	26
Tabel 4.6 Nilai Bobot F12 pada Setiap Iterasi	28
Tabel 4.7 Hasil Bobot untuk Setiap Fitur.....	29
Tabel 4.8 Nilai Bobot Setiap Fitur dengan Algoritma Relief-f.....	30
Tabel 4.9 Hasil Perbandingan Nilai Akurasi pada Setiap Seleksi Fitur.....	32
Tabel 4.10 <i>Confusion Matrix</i> Metode Naive Bayes pada Data sebelum <i>Feature Selection</i>	33
Tabel 4.11 Hasil Evaluasi Data dengan Metode Naive Bayes sebelum <i>Feature Selection</i>	38
Tabel 4.12 <i>Confusion Matrix</i> Metode Naive KNN pada Data sebelum <i>Feature Selection</i>	39
Tabel 4.13 Hasil Evaluasi Data dengan Metode KNN sebelum <i>Feature Selection</i>	43
Tabel 4.14 <i>Confusion Matrix</i> Metode Naive Naive Bayes pada Data Hasil <i>Feature Selection</i>	43
Tabel 4.15 Hasil Evaluasi Data dengan Metode Naive Bayes setelah <i>Feature Selection</i>	47
Tabel 4.16 <i>Confusion Matrix</i> Metode Naive KNN pada Data Hasil <i>Feature Selection</i>	47
Tabel 4.17 Hasil Evaluasi Data dengan Metode KNN setelah <i>Feature Selection</i>	52

Tabel 4.18 Perbandingan Hasil *Feature Selection* Algoritma Relief-f pada

Data 52

BAB I

PENDAHULUAN

1.1 Latar Belakang

Feature Selection adalah suatu proses memilih subset dari fitur atau atribut yang penting dengan menggunakan kriteria tertentu dan salah satu dari *pre-processing* data (Liu, Motoda, Setiono, & Zhao, 2010). *Feature selection* biasanya digunakan untuk mereduksi data dengan dimensi yang banyak dan digunakan untuk menghilangkan fitur-fitur yang tidak memiliki relevansi terhadap dataset, sehingga dengan *feature selection* dapat menghemat jumlah memori dan waktu yang digunakan agar lebih efisien. *Feature Selection* merupakan cara yang efektif untuk melakukan reduksi data dan menjadi langkah penting yang perlu dilakukan supaya aplikasi data mining berhasil dengan baik (Irawan, Baizal, & Perdana, 2011).

Algoritma dari *feature Selection* memiliki beberapa jenis diantaranya Relief, Relief-f, RRelief-f, Relieved-f, I-Relief, TuRF, ERelief-f, Reliefseq, Multi-SURF dan masih banyak perkembangan jenis Relief lainnya (Urbanowicz, Meeker, Cava, Olson, & Moore, 2018). Pada penelitian ini menggunakan Relief-f. Relief-f merupakan algoritma yang dikembangkan dari algoritma Relief. Algoritma Relief memiliki kekurangan yaitu tidak bisa menangani *incomplete data* (data tidak lengkap) dan terbatas pada 2 kelas, sedangkan Relief-f sendiri bisa menangani dataset yang *multiclass* (lebih dari 2 kelas) dan *incomplete data* (data tidak lengkap) serta bisa menangani data dalam bentuk diskrit ataupun kontinu (Kononenko, 1994).

Feature selection algoritma Relief-f banyak digunakan diberbagai bidang riset diantaranya oleh Miftahuddin dkk (2012) menggunakan *feature selection* untuk menganalisis pertumbuhan produktivitas tumbuhan buah di beberapa Kabupaten yang berada pada Provinsi Aceh. Anggraeny (2017) menggunakan *feature selection* algoritma Relief-f untuk mendiagnosa penyakit hepatitis. Kesuma (2011) menggunakan *feature Selection* algoritma Relief untuk menganalisis fitur mana yang optimal pada data indeks kesehatan masyarakat. Kinerja *feature selection* algoritma Relief-f pernah digunakan oleh Irawan dkk (2011) untuk menganalisis perbandingan Klasifikasi pada beberapa dataset *multiclass*.

Career Development Center (CDC) merupakan pusat pengembangan karir yang ada di Universitas Sriwijaya, dimana CDC dibentuk pada tahun 2013 untuk menyikapi rendahnya nilai capaian point pelacakan lulusan terhadap borang AIPT. CDC menyediakan data *tracer* alumni pada *website* (<http://cdc.unsri.ac.id>) dimana CDC menyebarkan kuisoner online yang terdiri atas 17 pertanyaan yang mengacu pada standar DIKTI. Kuisoner tersebut berisikan pertanyaan-pertanyaan tentang penilaian kondisi dan regulasi belajar yang alumni Unsri alami dalam masa belajar dengan dunia kerja yang mereka hadapi, sehingga dokumen *tracer* alumni dapat memberi manfaat baik untuk pengelola unsri maupun untuk pengguna (stakeholder) lulusan Unsri (Gofar & Susmanto, 2018). Data yang digunakan adalah data *tracer* alumni pada tahun 2016 dan 2017, data tersebut diisi oleh alumni tahun 2013, 2014 dan 2015 dengan jumlah data sebanyak 1143 data . Data yang tersedia dalam bentuk dimensi yang besar dengan jumlah yang banyak serta atribut pendukung yang

beragam, maka bisa dimanfaatkan untuk mengetahui faktor apa saja yang dapat mendukung untuk meningkatkan kualitas lulusan Universitas Sriwijaya. Algoritma Relief-f digunakan pada penelitian ini dalam menyeleksi fitur untuk memprediksi dan melihat kesesuaian tingkat pendidikan dengan bidang pekerjaan pada alumni Universitas Sriwijaya dengan menggunakan atribut atau fitur-fitur terpilih yang ada pada data CDC untuk mendukung kesesuaian tersebut.

1.2 Perumusan Masalah

Bagaimana menerapkan algoritma Relief-f untuk menyeleksi fitur pada prediksi kesesuaian tingkat pendidikan dengan bidang pekerjaan alumni Universitas Sriwijaya dengan melihat tingkat akurasi yang diperoleh.

1.3 Pembatasan Masalah

Penelitian ini dibatasi pada data alumni Universitas Sriwijaya untuk lulusan 2013, 2014 dan 2015. Fitur yang digunakan sesuai dengan fitur yang tersedia pada data *tracer* alumni pada *website*. Adapun fitur-fitur yang digunakan adalah :

1. Tingkat pembelajaran pada perkuliahan
2. Tingkat kompetensi dalam proyek riset
3. Tingkat pembelajaran pada magang
4. Tingkat pembelajaran pada praktikum
5. Tingkat pembelajaran pada kerja lapangan
6. Tingkat pembelajaran pada diskusi

7. Waktu mendapat kerja
8. Tingkat kompetensi bidang ilmu
9. Tingkat kompetensi luar bidang ilmu
10. Tingkat kompetensi bahasa inggris
11. Tingkat kompetensi kemampuan internet
12. Tingkat kompetensi komputer
13. Hubungan program studi dengan pekerjaan
14. Tingkat pendidikan yang sesuai dengan bidang pekerjaan

1.4 Tujuan

Menerapkan algoritma Relief-f untuk menyeleksi fitur-fitur yang paling berpengaruh pada prediksi kesesuaian tingkat pendidikan dengan bidang pekerjaan alumni Universitas Sriwijaya dengan melihat nilai akurasi prediksi yang diperoleh.

1.5 Manfaat

Manfaat dari hasil penelitian ini adalah :

1. Mengetahui faktor apa saja yang paling mempengaruhi alumni Universitas Sriwijaya dalam mencari pekerjaan yang sesuai dengan bidang pendidikannya.
2. Membantu pihak Universitas Sriwijaya untuk memberikan program-program untuk mempersiapkan alumni Universitas Sriwijaya yang siap bersaing dalam dunia kerja.

3. Dapat memberikan rekomendasi bagi pihak Universitas Sriwijaya untuk meningkatkan kualitas lulusan mahasiswa Unsri agar bisat bersaing di dunia kerja.

DAFTAR PUSTAKA

- Anggraeny, F. T., Purbasari, I. Y., & Suryaningsih, E. (2017). ReliefF Feature Selection and Bayesian Network Model for Hepatitis Diagnosis. In *3rd International Conferences on Information Technology and Business (ICITB)*, 7th Dec 2017 (hal. 113–118).
- Gofar, N., & Susmanto, P. (2018). *Tracer Study Universitas Sriwijaya Tahun 2018 (Lulusan Tahun 2016)*. Palembang: Noer Fikri.
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining Concepts and Techniques* (3rd ed.). Waltham: Morgan Kaufmann.
- Irawan, D. A., Baizal, Z. A., & Perdana, E. G. (2011). Analisis dan Implementasi Algoritma Relief-f untuk Feature Selection pada Klasifikasi Dataset Multiclass, 1–6.
- Kesuma, Z. M. (2011). Feature Selection Data Indeks Kesehatan Masyarakat Menggunakan Algoritma Relief, *11*(1), 61–66.
- Kira, K., & Rendell, L. A. (1992). A Practical Approach to Feature Selection. In *9th International Workshop on Machine Intelligence, Aberdeen* (hal. 249–256). Scotland: Morgan Kaufmann Publishers, Inc.
- Kononenko, I. (1994). Estimating Attributes : Analysis and Extensions of Relief. In *European Conference on Machine Learning* (hal. 171–182).
- Liu, H., Motoda, H., Setiono, R., & Zhao, Z. (2010). Feature Selection : An Ever Evolving Frontier in Data Mining, *JMLR: Workshop and Conference Proceedings* 10:4-13.
- Lorena Br Gintig, S., Zarman, W., & Darmawan, A. (2014). Teknik Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighborhood, *3*(2), 29–34.
- Robnik-sikonja, M. (2014). Theoretical and Empirical Analysis of ReliefF and RReliefF Theoretical and Empirical Analysis of ReliefF and RReliefF \checkmark . *Machine Learning Journal*, *53*(October 2003), 23–69.
- Sun, Y., Lou, X., & Bao, B. (2011). A Novel Relief Feature Selection Algorithm Based on Mean-Variance Model, *16*(December), 3921–3929.
- Urbanowicz, R. J., Meeker, M., Cava, W. La, Olson, R. S., & Moore, H. (2018).

Relief-Based Feature Selection: Introduction and Review. *Journal of Biomedical Informatics*.

Wahyuningsih, S. (2018). Perbandingan Metode K-Nearest Neighbor , Naïve Bayes dan Decision Tree untuk Prediksi Kelayakan Pemberian Kredit. In *Konferensi Nasional Sistem Informasi 2018* (hal. 619–623).