

Penerapan *Naïve Bayes Classification* untuk Klasifikasi Tingkat Kemungkinan Obesitas Mahasiswa Sistem Informasi UIN Suska Riau

Wiwik Muslehatin¹, Muhammad Ibnu², Mustakim³

^{1,2,3}Laboratorium Data Mining Program Studi Sistem Informasi, Fakultas Sains dan Teknologi
Universitas Islam Negeri Sultan Syarif Kasim Riau

Jl.HR. Soebrantas No. 155 Simpang Baru, Tampan, Pekanbaru, Riau – Indonesia 28293
e-mail: wiwik.muslehatin@students.uin-suska.ac.id, muhammad.ibnu@students.uin-suska.ac.id,
mustakim@uin-suska.ac.id

Abstrak

Tingkat penderita obesitas pada remaja Indonesia terus meningkat dari tahun ke tahun. Hal ini terjadi karena pola makan dan gaya hidup yang tidak benar. Oleh karena itu, penelitian ini bertujuan menganalisa tingkat obesitas menggunakan algoritma klasifikasi pada teknik data mining. Penelitian ini dilakukan pada mahasiswa jurusan Sistem Informasi UIN Sultan Syarif Kasim Riau dengan mengambil sampel secara random sebanyak 88 orang. Kriteria atau atribut yang digunakan untuk mengklasifikasi kemungkinan mahasiswa menderita obesitas antara lain, lingkar perut, berat badan dan tinggi badan. Hasil pengujian menunjukkan akurasi sebesar 66,67% yaitu 16 responden beresiko untuk mendapatkan kemungkinan obesitas tingkat sedang. 69 responden tingkat gizi normal, 3 responden kekurangan gizi. Obesitas ini harus diperhatikan karena mahasiswa yang mengalami obesitas mempunyai 80% berpeluang untuk mengalami obesitas pula pada saat dewasa atau usia lanjut. Penderita obesitas juga berisiko lebih tinggi untuk masalah kesehatan, seperti penyakit jantung, stroke, diabetes, asma, dan beberapa jenis kanker.

Kata Kunci: Data mining, *Naïve bayes classification*, Obesitas

Abstract

The level of obesity for Indonesian youth increased continuously from year to year. This happens due to diet and lifestyle in a wrong way. Therefore, this study aimed to analyze the level of obesity in the classification algorithm using data mining techniques. Research is conducted on students majoring in Information Systems of State Islamic University of Sultan Syarif Kasim Riau by taking a random sample of 88 people. The criteria or attributes used to classify the likelihood of students suffer from obesity, among others, waist circumference, body weight and height. The test results showed an accuracy of 66.67% is 16 respondents at risk for getting the likelihood of obesity moderate. 69 respondents normal nutrient levels, 3 respondents malnutrition. Obesity should be considered for students who are obese have a 80% chance to be obese as well as adults or the elderly. Obese people are also at higher risk for health problems such as heart disease, stroke, diabetes, asthma, and some cancers.

Key Words : Data mining, *Naïve Bayes classification*, Obesity

1. Pendahuluan

Teknologi dimanfaatkan dalam berbagai kalangan, segala aktivitas sudah memanfaatkan teknologi sebagai penunjang dari berbagai aktivitas dimulai dari aktivitas yang ringan hingga aktivitas berat. Salah satunya teknologi digunakan dalam bidang kesehatan.

Kesehatan tubuh seseorang dikatakan baik apabila berat badan dan tinggi badan seimbang [13]. Obesitas sudah menjadi masalah di banyak Negara dikalangan orang dewasa di dunia seperti negara Amerika Serikat 30%, Inggris 22% pada tahun 2001 [2]. Ada 4 periode kritis terjadinya obesitas, yaitu pada masa prenatal, masa bayi, masa *adiposity rebound*, dan masa remaja. Obesitas yang terjadi pada masa remaja, 30% akan berlanjut sampai dewasa menjadi obesitas persisten [15]. Kelebihan berat badan hingga beberapa kilogram bisa menimbulkan risiko kesehatan yang tak bisa disepelekan [6]. Untuk itu perlu adanya pengetahuan dan

pengecambahan dalam menangani kasus obesitas pada remaja hingga dewasa. Dalam menentukan tingkat obesitas, dapat diukur melalui lingkaran pinggang dan pinggul, rasio lingkaran pinggang dan pinggul merupakan faktor yang sangat signifikan terkait dengan presentasi akut penyakit jantung koroner (PJK) [18]. Dalam mengolah data penelitian ini menggunakan teknik data mining.

Data mining merupakan salah satu teknik dalam pengolahan data yang menemukan hubungan dari data yang tidak diketahui oleh pengguna serta menyajikannya kedalam bentuk yang mudah dipahami sehingga dari hubungan data tersebut dapat dijadikan sebagai dasar dalam pengambilan keputusan [14]. Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan yaitu : Deskripsi, Estimasi, Prediksi, Klasifikasi, *Clustering*, dan Asosiasi [17]

Klasifikasi adalah suatu metode yang digunakan dalam menentukan suatu record data baru ke salah satu dari beberapa kategori yang telah di definisikan sebelumnya. [8]. Algoritma klasifikasi ini digunakan untuk mengklasifikasi tingkat kemungkinan obesitas dari mahasiswa Sistem Informasi Uin Suska Riau. Hal ini ditandai dengan adanya kebiasaan atau pola hidup yang tidak baik berat badan, dan tinggi badan. Berdasarkan hasil kuisioner dari penelitian ini diperoleh data yang akan dijadikan sebagai bahan penelitian dengan melakukan klasifikasi pada data yang telah ditentukan yang terdiri dari 88 sample mahasiswa sistem informasi uin suska riau.

Dalam menentukan kesehatan seseorang diperlukan kriteria kriteria yang ditentukan sebagai objek penelitian. Kesehatan dalam penelitian ini melibatkan pola hidup dari kegiatan yang dilakukan oleh mahasiswa. Ada perubahan pola hidup mahasiswa ketika melakukan aktivitas yang membutuhkan waktu lebih sehingga mempengaruhi kesehatan dari mahasiswa tersebut.

Berdasarkan hasil dari kuisioner yang telah disebarkan pada mahasiswa sistem informasi uin suska riau, akan dilakukan klasifikasi dari tingkat kemungkinan obesitas mahasiswa sistem informasi yang disebabkan oleh gaya hidup yang tidak sehat. Menurut penelitian dari [9] faktor penyebab dari obesitas pada remaja sangat kompleks diantaranya adalah kurang tidur, banyak menghabiskan waktu pada kegiatan yang sifatnya statis, pola makan yang tidak sehat, hasil penelitian nya menyatakan bahwa rata-rata remaja mengalami obesitas sebanyak 10,5% dari siswa yang dijadikan objek penelitian.

Berdasarkan penelitian yang telah dilakukan sebelumnya menggunakan metode algoritma *Naïve Bayes Classification* dinyatakan bahwa metode ini cocok digunakan dalam kasus ini. Oleh karena itu dalam mengelaskan status kemungkinan obesitas pada mahasiswa sistem informasi uin suska riau dapat dilakukan menggunakan metode *naïve bayes classification*. Selain itu penelitian yang dilakukan oleh [11] tentang klasifikasi status gizi dengan hasil penelitian nya sebesar 93,2%. Menurut penelitian nya *metode Naïve bayes Classification* memiliki kelebihan dengan kinerja yang baik dengan hasil pengujian sebesar 93,2%. Adapun menurut [5] kekurangan dari metode ini adalah pada saat melakukan proses training dan testing klasifikasi tidak optimal jika tidak disimpan pada memori (*database*), untuk itu perlu disimpan hasil klasifikasi pada memori (*database*) untuk mengoptimalkan klasifikasi.

2. Bahan dan metode

2.1 Data Mining

Data Mining sering juga disebut *Knowledge discovery in database*(KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan, pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan. sehingga istilah *pattern recognition* jarang digunakan karena termasuk bagian dari data mining [17]

Metode Pelatihan

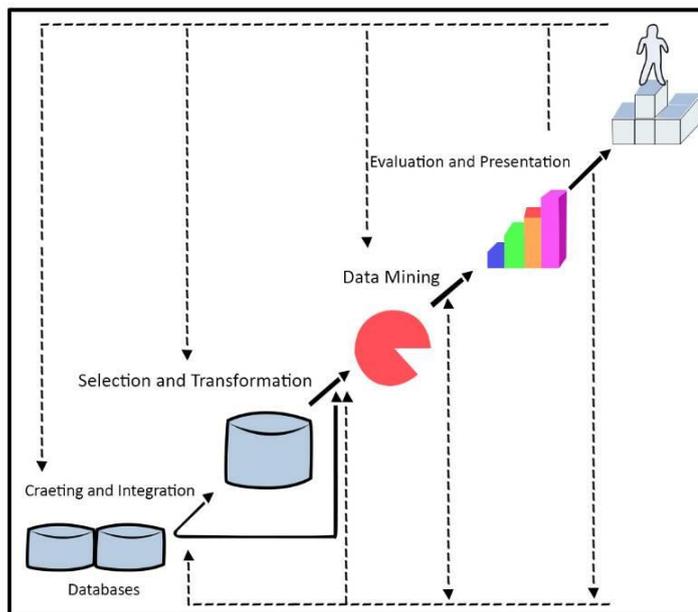
Secara garis besar metode pelatihan yang digunakan dalam teknik-teknik *data mining* dibedakan ke dalam dua pendekatan, yaitu [17]

1. *Unsupervised learning*, metode ini diterapkan tanpa adanya latihan (*training*) dan tanpa ada guru (*teacher*). Guru di sini adalah label dari data.
2. *Supervised learning*, yaitu metode belajar dengan adanya latihan dan pelatih. Dalam pendekatan ini, untuk menemukan fungsi keputusan, fungsi pemisah atau fungsi regresi, digunakan beberapa contoh data yang mempunyai output atau label selama proses

Banyak Orang-orang memperlakukan data mining sebagai sinonim untuk *popularly* digunakan penemuan pengetahuan lain dari data atau KDD, sementara yang lain melihat data mining hanya sebagai langkah penting dalam proses penemuan pengetahuan.

Proses penemuan pengetahuan ditunjukkan pada : [12]

1. Data pembersihan (untuk menghilangkan kebisingan dan data yang tidak konsisten)
2. Integrasi data (di mana beberapa *sources* data dapat digabungkan)



Gambar 1. Proses data mining (Jananto, 2013)

3. Data (di mana data yang relevan dengan tugas analisis yang diambil dari database)
4. Data transformasi (dimana data diubah dan dikonsolidasikan ke dalam bentuk yang sesuai untuk pertambangan dengan melakukan operasi Ringkasan org agregasi)
5. Data mining (proses esensial di mana metode cerdas yang diterapkan untuk mengekstrak pola data)
6. Pola evaluasi (untuk mengidentifikasi pola yang benar-benar menarik representing pengetahuan berdasarkan langkah-langkah interestingness)
7. Presentasi ilmu pengetahuan (dimana visualisasi dan teknik pengetahuan yang digunakan untuk menampilkan hasil mining kepada pengguna)

Pengetahuan presentasi (di mana visualisasi dan representasi pengetahuan teknik yang digunakan untuk menyajikan pengetahuan ditambang untuk pengguna) Langkah 1 sampai 4 berbagai bentuk preprocessing data, dimana data dipersiapkan untuk pertambangan. Langkah data mining dapat berinteraksi dengan pengguna atau base [10]. Pengetahuan pola yang menarik disajikan kepada pengguna dan dapat disimpan sebagai pengetahuan baru dalam basis pengetahuan..[7]

2.2 Algoritma Naive Bayes Classification

Teori keputusan bayes adalah pendekatan statistik yang fundamental dalam pengenalan pola (*pattern recognition*), pendekatan ini didasarkan pada kuantifikasi *trade-off* antara berbagai keputusan klasifikasi dengan menggunakan probabilitas dan ongkos yang di timbulkan dalam keputusan tersebut [17] Selain itu *Bayesian clasification* juga dapat memprediksi probabilitas keanggotaan suatu *class*. pada teorema bayes yang memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network*. *Bayesian clasification* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam *database* dengan data yang besar . [1]

Teorema Bayes memiliki bentuk umum sebagai berikut :

$$P(H | X) = \frac{P(X | H)P(H)}{P(X)} \dots\dots\dots(1)$$

Keterangan :

- X = Data dengan class yang belum diketahui
- H = Hipotesis data X merupakan suatu class spesifik
- P(H|X) = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)
- P(H) = Probabilitas hipotesis H (prior prob.)
- P(X|H) = Probabilitas X berdasarkan kondisi tersebut
- P(X) = Probabilitas dari X (Jananto,2013)

2.3 RapidMiner

RapidMiner adalah sebuah tools yang digunakan dalam teknik yang berada di lingkungan *machine learning*, data mining, text mining dan *predictive analytics* [19].

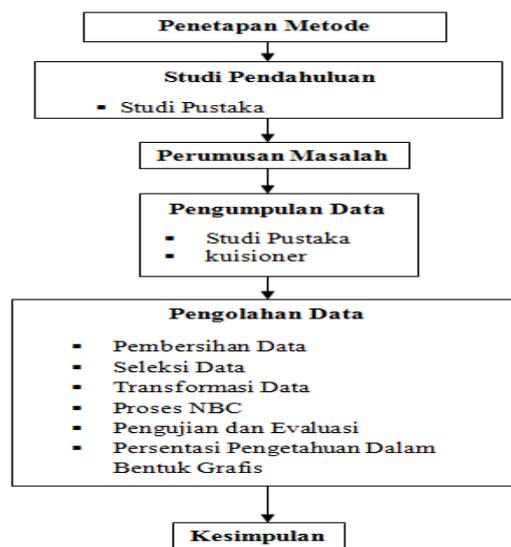
2.4. Obesitas

Obesitas atau kegemukan terjadi pada saat badan menjadi gemuk (*obese*) yang disebabkan penumpukan *adipose* (*adipocytes*: jaringan lemak khusus yang disimpan tubuh) secara berlebihan. Jadi obesitas adalah keadaan dimana seseorang memiliki berat badan yang lebih berat dibandingkan berat idealnya yang disebabkan terjadinya penumpukan lemak di tubuhnya [16]. Obesitas dapat terjadi pada siapa saja, baik balita maupun orang dewasa. Masa remaja adalah masa yang menyenangkan, namun juga masa yang kritis dan sulit, karena merupakan masa transisi atau peralihan dari masa kanak-kanak menuju masa dewasa, yang ditandai dengan perubahan aspek fisik, psikis, dan psikososial [4].

Dampak lain yang sering diabaikan adalah obesitas dapat mengganggu kejiwaan pada anak, yakni sering merasa kurang percaya diri. Apalagi jika anak sedang dalam masa remaja dan mengalami obesitas, biasanya akan menjadi pasif dan depresi karena sering tidak dilibatkan pada kegiatan yang dilakukan oleh teman sebayanya [4]. Obesitas juga mengurangi kualitas hidup, seperti *stroke*, artritis (radang sendi), batu empedu, kesulitan bernafas, masalah kulit, infertilitas, masalah psikologis, mangkir kerja dan pemanfaatan sarana kesehatan [16].

2.5. Metodologi

Metodologi yang diterapkan dalam melakukan penelitian ditunjukkan pada diagram berikut :



Gambar.2 Metodologi Penelitian

Tahap pertama yang dilakukan adalah menetapkan metode yang cocok untuk diterapkan, kemudian setelah selanjutnya melakukan studi literatur mengenai topik bahasan yang akan diangkat dalam penelitian, kemudian mengumpulkan data yang dibutuhkan, pada penelitian ini data diperoleh dari survei melalui kuisioner dan mengambil sampel secara acak, setelah itu data diseleksi mana yang layak digunakan dan mana yang tidak layak, kemudian melakukan transformasi data kedalam tools yang digunakan yaitu *Rapid Miner*, langkah selanjutnya adalah melakukan proses perhitungan menggunakan metode *Naive Bayes Classification*, setelah melakukan perhitungan selanjutnya adalah dilakukan pengujian dan evaluasi menggunakan tools yang sama, dan langkah terakhir adalah menampilkan hasil akhir penelitian dalam bentuk grafis.

3. Hasil dan Pembahasan

3.1. Pembahasan

Obesitas merupakan status gizi yang sangat tidak diharapkan oleh setiap insan. Beberapa penyakit akan tercetus dari kondisi obesitas tersebut. Pada ilmu kesehatan, penentuan status gizi seseorang dapat dinilai dengan formula tertentu dengan melibatkan variabel berat badan dan tinggi badan. Indeks Massa Tubuh (IMT) sering digunakan untuk mengukur status gizi seseorang. Persamaan berikut menunjukkan IMT [11].

$$IMT = \frac{\text{berat badan (kg)}}{\text{tinggi badan (m}^2\text{)}}$$

Rumus IMT diatas tidak melibatkan kriteria kerangka tubuh dan hanya melibatkan tinggi dan berat badan, sementara kerangka tubuh sangat mempengaruhi berat badannya. Kerangka tubuh disini seperti lingkaran perut. Karena bisa jadi status gizi seseorang berbeda walaupun mempunyai tinggi badan dan berat badan yang sama.

Oleh karena itu digunakanlah metode *Naive Bayes Classification* karena didalam metode ini semua atribut (kriteria) akan memberikan kontribusinya dalam pengambilan keputusan, dengan bobot atribut yang sama penting dan setiap atribut saling bebas satu sama lain.

Contoh penelitian sebelumnya dengan metode *naive bayes clasification* adalah penelitian yang dilakukan oleh Sri Kusuma Dewi tentang "Klasifikasi Status Gizi" dimana hasil penelitian menunjukkan NBC dapat memecahkan masalah dengan cukup baik dengan kinerja sistem sebesar 93,2%.

Dalam pembahasan pada penelitian ini diperlukan beberapa data-data kesehatan mahasiswa jurusan Sistem Informasi yang meliputi data berat badan, tinggi badan dan lingkaran perut. Status gizi juga diketahui berdasarkan pengukuran tersebut. Hasil pengukuran seperti terlihat pada Tabel dibawah.

Tabel 1. Data mahasiswa sistem informasi UIN SUSKA RIAU

No	Nama	Berat Bdn	Tinggi Bdn	lingkar perut
1	Wiwik Muslehatin	56	162	83
2	Vellyca Werdining Putri	45	157	77
3	Riat Rayendra	68	173	80
4	Risma Rustiyan R	50	151	73
5	Hadnan Hardiansyah	61	170	70
6	Hasbi	53	168	70
7	Zil Abrar	65	168	77
8	Marni	50	150	77
9	Seli Purnianda	43	153	72
10	Nathania	47	150	70
11	Adeafsa	64	164	72
12	Eva Normala	50	146	74
13	Masyfuati Masrapah	54	149	74

14	Chayo Rona F	37	148	58
15	Fitri Haryati	45	160	77
....
....
88	Andre Oktora	55	171	70

Data data tersebut akan diclasifikasi dengan metode naive bayes kedalam beberapa kelas, dimana kelas ini ditentukan berdasarkan nilai standar IMT untuk orang Asia. [11]

Adapun kelasnya sebagai berikut:

1. Kurang
2. Normal
3. Obesitas sedang
4. Obesitas berat

3.2. Hasil

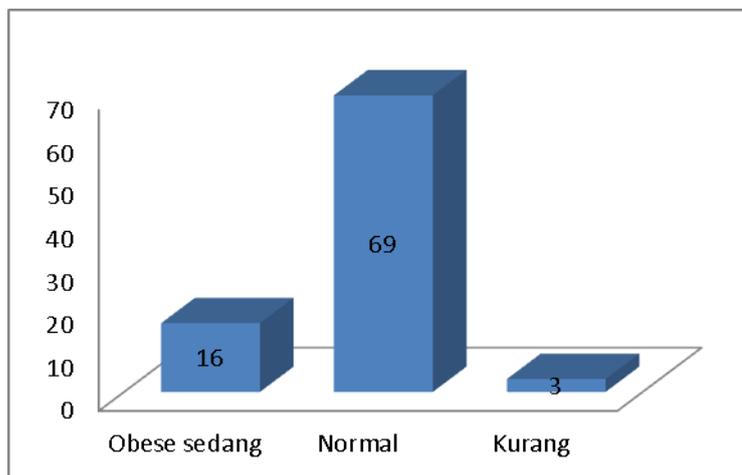
Hasil akurasi yang didapat dari perhitungan menggunakan metode Naïve Bayes Classification adalah :

Tabel 2. Klasifikasi status gizi

No	Status gizi	Tinggi Badan	Berat Badan	Lingkar perut
1	Obese sedang	162 cm	56 kg	83 cm
2	Normal	170 cm	55 kg	74 cm
3	Kurang	168 cm	49 kg	74 cm

Tabel diatas merupakan hasil dari perhitungan klasifikasi dari data mahasiswa sistem informasi UIN SUSKA RIAU menggunakan metode *Naïve Bayes Classification*. Dari data tersebut, untuk status gizi obese berat tidak tertera pada data.

Berdasarkan data diatas, Berikut adalah grafik hasil dari Classifikasi dengan Algoritma *Naive Bayes*. Data menunjukkan bahwa sebanyak 16 orang dari mahasiswa sistem informasi UIN SUSKA RIAU mengalami tingkat obesitas sedang, kemudian sebanyak 69 orang tingkat gizi normal, dan 3 orang dengan tingkat gizi kurang. Diaram tersebut dapat dilihat seperti gambar berikut ini.



Gambar 3. Diagram hasil klasifikasi

Dari hasil klasifikasi diatas dapat diketahui bahwa :

1. 16 responden beresiko untuk mendapatkan kemungkinan obesitas tingkat sedang.
2. 69 responden mempunyai tingkat gizi normal
3. 3 responden mempunyai kerurangan gizi.

4. Kesimpulan

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, dapat disimpulkan bahwa algoritma *Naive Bayesian Classification* (NBC) dapat digunakan sebagai salah satu metode untuk klasifikasi tingkat kemungkinan obesitas pada mahasiswa sistem informasi UIN SUSKA RIAU. Berdasarkan hasil pengukuran, hasil pengujian menunjukkan dengan akurasi sebesar 66,67% yaitu 16 responden beresiko untuk mendapatkan kemungkinan obesitas tingkat sedang, 69 responden mempunyai tingkat gizi normal, 3 responden mempunyai kekurangan gizi.

Ucapan Terimakasih

Terimakasih kepada Fakultas Sains dan Teknologi dan Program Studi Sistem Informasi atas segala kesempatan yang diberikan serta dukungan fasilitas dalam menunjang pembuatan *paper* ini. Kemudian juga kepada Komunitas *Puzzle Research Data Technology* (PREDATECH) atas segala inspirasi, motivasi dan bantuannya sehingga *paper* ini dapat terselesaikan dengan baik.

Daftar Pustaka

- [1] Jananto Arief, " Algoritma Naïve Bayes Untuk Mencari Perkiraan Waktu Studi Mahasiswa". *Jurnal Teknologi Informasi DINAMIK*, Vol 18, No.1 2013.
- [2] Brunner Erick J, *et al*, " Prospective Effect of Job Strain on General and Central Obesity in the Whitehall II Study". *American Journal of Epidemiology*. Volume 165 Issue 7. 2007
- [3] CMPUT 690. "Principal of Knowledge Discovery in Database <https://webdocs.cs.ualberta.ca/~zaiane/courses/cmput690/notes/Chapter1/>. University of Alberta. (Diakses tanggal 25 maret 2017)
- [4] Dewi Nur Wijayanti 2013. *Analisis Faktor Penyebab Obesitas Dan Cara Mengatasi Obesitas Pada Remaja Putri*. Skripsi, Program Studi Ilmu Keolahragaan ,Fakultas Ilmu Keolahragaan,Universitas Negeri Semarang.
- [5] Rodiyansyah Sandi Fajar, Winarko Edi,. "Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayesian Classification". *IJCCS*, Vol.6, No.1, January 2012, pp. 91-100 ISSN: 1978-1520
- [6] Hiola, *dkk*, "Pengaruh Obesitas Terhadap Terjadinya Disfungsi Seksual Pria". *Jurnal e-Biomedik (eBM)*, Volume 1, Nomor 1, 686-690 . 2013
- [7] Han. J. , M. Kamber dan J.Pei. 2012. *Data Mining: Concepts and Techniques, Third Edition*. Morgan Kaufmann Publishers. San
- [8] Hermawati.A.Fajar. 2013.*Data Mining*.Andi.Yogyakarta : Francisco.
- [9] Suryaputra Kartika, dan Nadhiro Siti Rahayu, "Perbedaan Pola Makan dan Aktivitas Fisik Antara Remaja Obesitas Dengan Non Obesitas". *Makara Kesehatan*, Vol.16, No. 1, 45-50 2012
- [10] Kusri, Luthfi, E.T. (2009). *Algoritma Data Mining* , Andi Offset. Surabaya.
- [11] Kusumadewi .S, "Klasifikasi Status Gizi Menggunakan naive Bayesian Classification". *CommIT*, Vol. 3 No. 1 96 – 11 2009
- [12] Larose, Daniel T, *Data Mining Methods and Models*. Hoboken New Jersey : Jhon Wiley& Sons, Inc, 2006.
- [13] Listiyana Aulia Dewi,. *dkk*, "Obesitas Sentral Dan Kadar Kolesterol Darah Total" *Jurnal Kesehatan Masyarakat KEMAS* Vol 9 No. 1 37-43 (2013)
- [14] Ridwan M, Suyono H, dan Sarosa M, "Penerapan Data Mining untuk evaluasi kinerja Akademik Mahasiswa menggunakan Algoritma Naive Bayes Classifier". *Jurnal EECCIS* Vol.7, No. 1, 2013
- [15] Ruslie Riska Habriel, dan Darmadi, "Analisis Regresi Logistik Untuk Faktor-Faktor Yang Mempengaruhi Status Gizi Remaja". *Majalah Kedokteran Andalas* No.1. Vol.36. 2012
- [16] Salam, A. 2010. "Faktor Resiko Kejadian Obesitas pada Remaja". *Jurnal MKMI* Vol 6 No.3, 185-190. 2010
- [17] Santosa, B. 2007. *Data Mining: Teknik Pemanfaatan Data untukKeperluan Bisnis*. Graha Ilmu. Yogyakarta.
- [18] Welborn T A dan Dhaliwa S S, "Preferred clinical measures of central obesity for predicting mortality" *European Journal of Clinical Nutrition* (2007) 61, 1373-1379; doi:10.1038/sj.ejcn.1602656; published online 14 February 2007
- [19] Y.S.Nugroho. 2009. *Data Mining Menggunakan Algoritma Naive Bayes untuk klasifikasi kelulusan Mahasiswa Universitas Diannuswantoro*. Program Studi Sistem Informasi, Fakultas Ilmu Komputer , Universitas Dianuswantoro.