

PERBANDINGAN ANTARA ALGORITME ID3 DAN NAIVE BAYESIAN CLASSIFICATION DALAM MENGELOMPOKAN STATUS GIZI

COMPARISON BETWEEN AND ID3 ALGORITHM NAIVE BAYESIAN CLASSIFICATION CLASSIFYING STATUS IN NUTRITION

Tri Astuti¹, Zanuvar Rifai², Diyo Mardiyanto³, Bagas Luka A.L⁴, Angga Tri S⁵, Widi Hastuti⁶,
Tia Dwi⁷

^{1,2,3,4,5,6,7,8} STMIK Amikom Purwokerto, Purwokerto

Jl. Letjen Pol Sumarto Purwokerto

^{1,3,4,5,6,7,8} Teknik Informatika, ² Sistem Informasi

E-mail: ¹tri_astuti@amikompurwokerto.ac.id, ²zanuar.rifai@amikompurwokerto.ac.id,

³diyo@amikompurwokerto.ac.id, ⁴bagas@amikompurwokerto.ac.id,

⁵angga@amikompurwokerto.ac.id, ⁶widi@amikompurwokerto.ac.id,

⁷tia@amikompurwokerto.ac.id

Abstrak

Perubahan berat badan balita dari waktu ke waktu merupakan petunjuk awal dari perubahan status gizi. Penting bagi orang tua untuk mengontrol status gizi anaknya. Untuk orang tua yang sudah mengetahui status gizi anaknya, dapat mengontrol berat badan sehingga dapat mencapai berat badan yang ideal menurut KMS. Untuk itu diperlukan metode yang tepat dalam menentukan status gizi pada balita. Metode yang digunakan dalam penelitian ini meliputi Pengumpulan data yang dilakukan dengan observasi dan wawancara, kemudian data akan diuji menggunakan aplikasi WEKA 3.8.1 dalam beberapa algoritma. Yang dimana nantinya algoritma yang memiliki akurasi tertinggi dapat digunakan sebagai metode dalam mengelompokkan status gizi balita. Dari hasil pengujian beberapa algoritma, didapatkan algoritma Naive Bayesian Classification yang memiliki akurasi tertinggi sebesar 91,4634% dibandingkan dengan algoritma yang lain. Sehingga algoritma Naive Bayesian Classification dapat digunakan sebagai metode dalam mengelompokkan status gizi pada balita.

Kata kunci : Naive Bayesian Classification, ID3, Data Mining, Status Gizi

Abstract

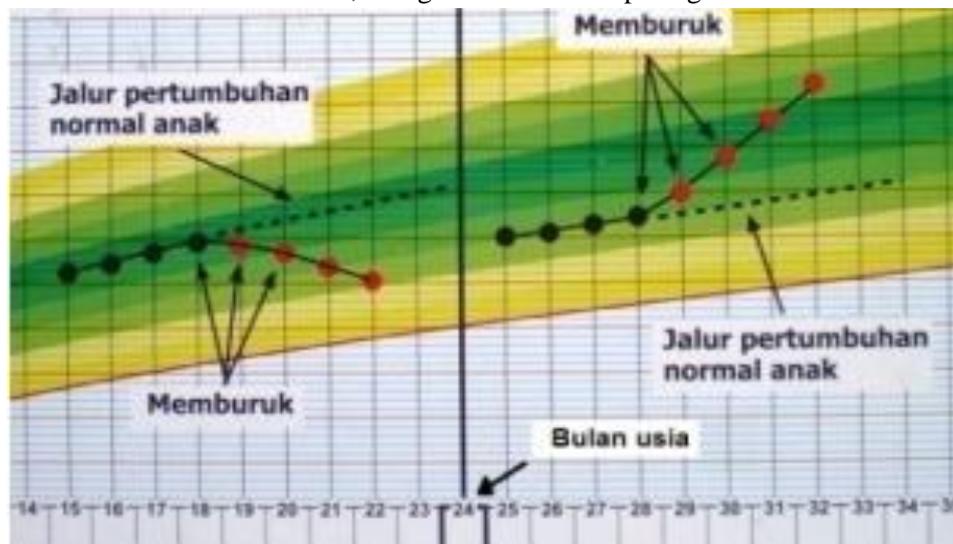
Changes in body weight of infants from time to time an early indication of changes in nutritional status. It is important for parents to control their nutritional status. For parents who already know the nutritional status of children, can control the weight so as to achieve ideal weight according to KMS. It is necessary for the proper method in determining the nutritional status of infants. The method used in this study includes the data collection is done by observation and interviews, then data will be tested using Weka 3.8.1 application in some algorithms. That where the latter algorithm has the highest accuracy can be used as a method to classify the nutritional status of children. From the results of testing several algorithms, obtained Naive Bayesian Classification algorithm that has the highest accuracy of 91.4634% compared with other algorithms. Naive Bayesian Classification algorithm so that it can be used as a method to classify the nutritional status of infants.

Keywords Naive Bayesian Classification, ID3, Data Mining, Nutritional status

1. PENDAHULUAN

Kondisi kesehatan dan status gizi balita merupakan salah satu tolak ukur cermin keadaan gizi masyarakat secara luas. Kasus gizi buruk tidak hanya menjadi beban keluarga juga menjadi beban negara. Kesehatan dan status gizi balita merupakan salah satu tolak ukur yang dapat mencerminkan keadaan gizimasyarakat luas. Salah satu cara mengetahui kesehatan dan pertumbuhan anak adalah dengan memantau hasil penimbangan berat badan bulannya yang tercatat dalam kartu menuju sehat(KMS). KMS adalah sebagai alat bantu bagi ibu atau orang tua dan petugas untuk memantau tingkat pertumbuhan dan perkembangan anak, serta menentukan tindakan – tindakan pelayanan kesehatan dan gizi^[1].

Berat badan adalah parameter antropometri yang sangat labil. Dala keadaan normal, keadaan kesehatan baik dan keseimbangan antara konsumsi dan kebutuhan zat gizi terjamin, maka berat badan akan bertambah mengikuti pertambahan umur. Sebaliknya dalam keadaan yang normal, terdapat dua kemungkinan perkembangan berat badan, yaitu dapat berkembang lebih cepat atau lebih lambat dari keadaan normal^[2], sebagai mana terlihat pada gambar 1.



Gambar 1. Jalur pertumbuhan pada KMS berdasarkan indeks B/U (Berat/Usia)

Pohon keputusan atau dikenal dengan *Decision Tree* adalah salah satu metode klasifikasi yang menggunakan representasi suatu struktur pohon yang berisi alternatif-alternatif untuk pemecahan suatu masalah. Pohon ini juga menunjukkan faktor – faktor yang mempengaruhi hasil alternatif dari keputusan tersebut disertai dengan estimasi hasil akhir bila kita mengambil keputusan tersebut. Peranan pohon keputusan ini adalah sebagian *Decision Support Tool* untuk membantu manusia dalam mengambil suatu keputusan. Manfaat dari *Decision Tree* adalah melakukan *break down* proses pengambilan keputusan yang kompleks menjadi lebih simpel sehingga orang yang mengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan. Konsep yang digunakan oleh *Decision Tree* adalah mengubah data menjadi suatu keputusan pohon dan aturan – aturan keputusan (*Rule*)^[3].

Model statistik merupakan salah satu model yang terpercaya sangat andal sebagai pendukung pengambilan keputusan. Konsep probabilitas merupakan salah satu bentuk model statistik. Salah satu metode yang merupakan konsep probabilitas adalah *Naive Bayesian Classification* (NBC). Pada metode ini, semua atribut akan memberika kontribusinya dalam pengambilan keputusan, dengan bobot atribut yang sama penting dan setiap atribut saling bebas satu sama lain. Apabila diberikan atribut K yang saling bebas (*independence*), nilai probabilitas dapat diberikan segai berikut.

$$P(X_1, \dots, X_k | C) = P(X_1 | C) \times \dots \times P(X_k | C)$$

Jika atribut ke-i bersifat diskret, maka $P(X_i | C)$ diestimasi sebagai frekuensi relatif dari sample yang memiliki nilai X_i sebagai atribut ke-i dalam kelas C. Namun, jika atribut ke-i bersifat kontinu, maka $P(X_i | C)$ diestimasi dengan fungsi densitas *Gauss*.

Status gizi balita diukur berdasarkan umur, berat badan (BB), dan Tinggi Badan (TB). Variabel BB dan TB tersebut disajikan dalam bentuk tiga indikator antropometri yaitu berat badan menurut umur (BB/U), Tinggi badan menurut /umur (TB/U), dan berat badan menurut tinggi badan (BB/TB). Angka berat badan dan tinggi badan setiap balita dikonversikan ke dalam bentuk nilai Z-Score dengan menggunakan buku antropometri WHO 2006 (Depkes RI 2009).

Penelitian sebelumnya yang membandingkan antar algoritma menghasilkan klasifikasi dari kinerja metode decision tree(C4.5) dan naive bayes. Atribut yang digunakan terdiri dari Gender, Usia, Berat badan, Waktu, Wilayah dan Tumbuh kembang. Dengan menggunakan masing-masing data training dan data testing sebanyak 304 data. Hasil dari penelitian yang dilakukan, berdasarkan dari nilai accuracy maupun recallnya naive bayes lebih tinggi dibandingkan dengan decision tree yaitu dengan nilai accuracy 75,66% untuk decision tree dan 76,97% untuk naive bayes. Untuk nilai recall-nya naive bayes lebih unggul yaitu 96,89% dibandingkan decision tree 89,78%. Meskipun dalam penelitian ini tingkat Precision-nya lebih tinggi decision tree yaitu 85,23% dibandingkan naive bayes 84,17%. Hasil akhir dari penelitian ini adalah metode naive bayes lebih baik digunakan dari pada metode decision tree dengan nilai total 250,67% untuk decision tree dan 258,03% untuk naive bayes^[4].

Selain perbandingan algoritma decision tree(C4.5) dan naive bayes, penelitian Imam Tahyudin(2013) algoritma SVM adalah algoritma yang tepat yang digunakan untuk memprediksi kelulusan siswa pada waktu. Tingkat akurasi untuk memprediksi algoritma SVM tinggi (hampir 100% dengan kategori klasifikasi sangat baik). Di sisi lain, hasil t-test dari algoritma SVM sangat dominan daripada algoritma lainnya^[5]. Pada penelitian tersebut penulis membandingkan algoritma Decision Tree (DT), Naif Bayes (NB), Artificial Neural Network (ANN), Support Vector Machine (SVM) dan Regresi Logistik (LR)

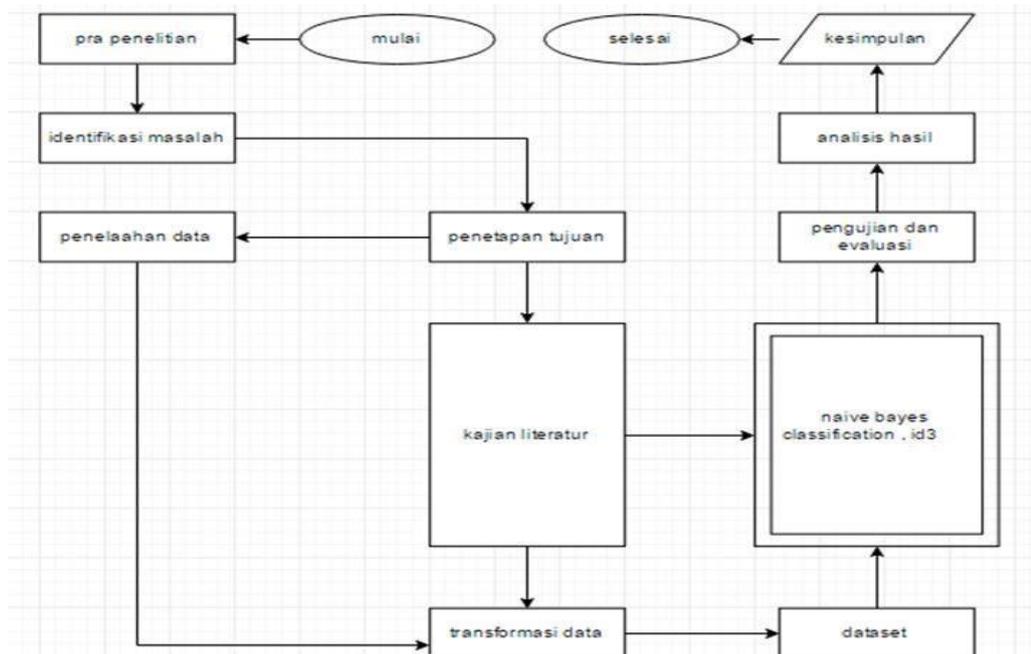
Wiwit Supriyanti, Dkk (2016) dalam judulnya Perbandingan Algoritma C4.5 dan Naive Bayes untuk memilih ketepatan pemilihan konsentrasi mahasiswa menyatakan bahwa hasil dari komparasi algoritma klasifikasi antara decision tree C4.5, dan naive bayes yang digabungkan dengan metode seleksi fitur forward selection untuk kasus ketepatan pemilihan konsentrasi mahasiswa didapatkan tingkat akurasi tertinggi dengan algoritma terpilih C4.5 dengan nilai akurasi sebesar 84,98%^[6].

P. Nancy dan R. Geetha Ramani ^[7] pada perbandingan dari beberapa metode klasifikasi data mining Rnd Pohon khususnya, ID3, K-NN, C-RT, CS-CRT, C.45 dan CS-MC4. Hasil penelitian menunjukkan bahwa algoritma Rnd Pohon memiliki tingkat kesalahan terkecil. Milan Kumari, Sunila GODARA^[8] pada perbandingan klasifikasi algoritma, yaitu RIPPER Data classifier mining, pohon keputusan, jaringan saraf tiruan (JST), mesin dukungan vektor (SVM). Hasilnya adalah algoritma SVM merupakan algoritma yang menghasilkan kesalahan terkecil dan Ketepatan terbesar. Aman kumar Sharma, dan Suruchi Sahni ^[9] klasifikasi perbandingan algoritma data mining, terutama ID3, J48, Kereta sederhana dan pohon Keputusan data email spam yang ditetapkan. Hasilnya adalah algoritma J48 memiliki nilai akurasi tertinggi. Neslihan Dogan dan Saturnus Tanrikulu ^[10] pada perbandingan algoritma data mining beberapa, terutama CHAID, MLP, Logistik, mengudara dan Naif Bayesian yang diterapkan untuk kumpulan data. Dihadirkan kesimpulan bahwa model terbaik yang digunakan untuk prediksi adalah model regresi. Penelitian lebih lanjut dilakukan oleh Silvia Marselina Suhartinah dan Ernastuti ^[11] pada penerapan data mining untuk memprediksi kelulusan menggunakan Naive Bayes dan C4.5 algoritma. Hasilnya adalah C4.5 algoritma memiliki tingkat akurasi yang lebih dari Naif Bayes.

2. METODE PENELITIAN

A. Tahapan Penelitian

Secara rinci tahapan penelitian sebagai berikut



Gambar 2. Tahapan Penelitian

B. Pengambilan Data

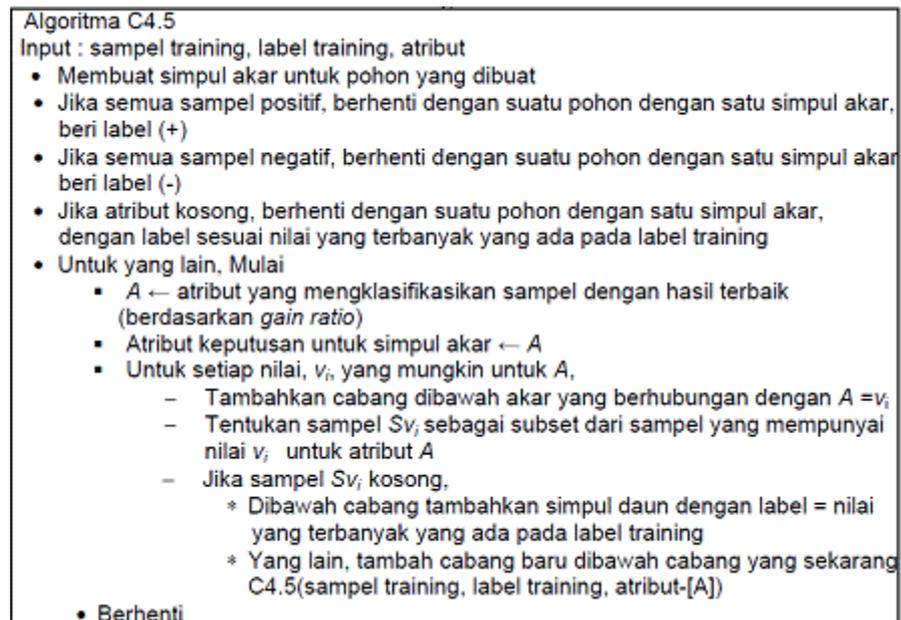
Penelitian ini memakai data Form Pencatatan Posyandu 2017, yang didapatkan dari Posyandu Desa Gumelem Wetan Kecamatan Susukan, Banjarnegara. Informasi lain kami dapatkan dari hasil wawancara dengan Kader posyandu dan bidan.

C. Metode yang diusulkan

Dalam penelitian ini akan menggunakan perhitungan Algoritma Decision Tree dan Naive Bayesian Classification, yang nantinya akan kami bandingkan antara hasil yang didapatkan oleh kedua algoritma tersebut. Dimana algoritma yang memiliki nilai akurasi yang tinggi akan digunakan sebagai metode dalam pengklasifikasian status gizi.

a. Algoritma Decision Tree

Salah satu algoritma induksi pohon keputusan yaitu ID3 (Iterative Dichotomiser 3). ID3 dikembangkan oleh J. Ross Quinlan. Dalam prosedur algoritma ID3, input berupa sampel training, label training dan atribut. Algoritma C4.5 merupakan pengembangan dari ID3. Sedangkan pada perangkat lunak *open source* WEKA mempunyai versi sendiri C4.5 yang dikenal sebagai J48.



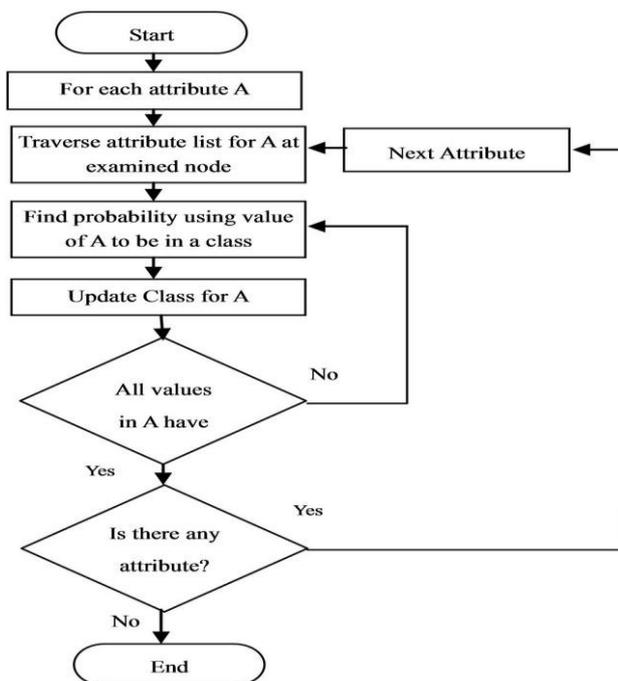
Gambar 3 Algoritma C4.5

Pohon dibangun dengan cara membagi data secara rekursif hingga tiap bagian terdiri dari data yang berasal dari kelas yang sama. Bentuk pemecahan (*split*) yang digunakan untuk membagi data tergantung dari jenis atribut yang digunakan dalam *split*. Algoritma C4.5 dapat menangani data numerik (kontinyu) dan diskret. *Split* untuk atribut numerik yaitu mengurutkan contoh berdasarkan atribut kontinyu A , kemudian membentuk minimum permulaan (*threshold*) M dari contoh-contoh yang ada dari kelas mayoritas pada setiap partisi yang bersebelahan, lalu menggabungkan partisi-partisi yang bersebelahan tersebut dengan kelas mayoritas yang sama. *Split* untuk atribut diskret A mempunyai bentuk *value* $(A) \in X$ dimana $X \subset domain(A)$.

Jika suatu set data mempunyai beberapa pengamatan dengan *missing value* yaitu *record* dengan beberapa nilai variabel tidak ada, Jika jumlah pengamatan terbatas maka atribut dengan *missing value* dapat diganti dengan nilai rata-rata dari variabel yang bersangkutan^[12]

b. Naive Bayes Classification

Ide di balik algoritma Naive Bayes adalah probabilitas posterior dari data contoh t_i dalam c_j kelas data model. The probabilitas posterior $P(t_i | c_j)$ adalah kemungkinan t_i yang dapat diberi label c_j . $P(t_i | c_j)$ dapat dihitung dengan mengalikan semua probabilitas dari semua atribut dari contoh data dalam model data: dengan p dinotasikan sebagai jumlah atribut di setiap contoh data. Probabilitas posterior dihitung untuk semua kelas, dan kelas dengan probabilitas tertinggi akan menjadi label contoh ini^[13]. Flowchart dari algoritma ini disajikan dalam Gambar 4.



Gambar 4 Flowchart of Naïve Bayes

D. Pengujian

Setelah mendapatkan data dan menemukan metode, kami melakukan penelitian terhadap data menggunakan aplikasi WEKA 3.8.1 untuk mencari akurasi tertinggi antara kedua algoritma tersebut.

3. HASIL DAN PEMBAHASAN

A. PENGUKURAN

Pada penelitian ini dilakukan pengukuran status gizi terhadap 82 sampel data balita di Posyandu Budi Lestari 3 dan 4 Desa Gumelem Wetan. Usia sampel berkisar antara 1-5 tahun. Terdapat 3 variabel pengukuran, yaitu berat badan(kg), umur(bln) dan jenis kelamin. Status gizi juga diketahui berdasarkan pengukuran tersebut. Data Dapat dilihat pada tabel berikut.

Table 1. Data Sample Balita (sumber : Posyandu Budi Lestari Ds. Gumelem Wetan)

No	NAMA ANAK	JENIS KELAMIN	UMUR	BB	STATUS GIZI
1	Ahmad Ata A	L	48	16,2	Baik
2	Akeila Rizqi yasha	P	50	12,6	Kurang
3	Reza Afrinskyah	L	45	14,3	Baik
4	Azka Aldric	L	43	16	Baik
5	Naura Yasmina	P	43	13,3	Baik
6	Faizal Hanif	L	42	14,1	Baik
7	Bagas Rianto	L	38	11,2	Kurang
8	Atalah Aditya	L	31	11,9	Baik
9	Bening Purnama	P	31	10,3	Baik
10	Raihana Aulia	P	28	9	Baik
11
12
82	Juan Abid Istara	L	3	6.0	Baik

Data tersebut nantinya akan diuji menggunakan aplikasi WEKA 3.8.1 untuk

menentukan algoritma mana yang lebih tepat dalam mengklasifikasikan status gizi pada balita.

B. PENGUJIAN

Sebelum diuji menggunakan WEKA, dataset diubah formatnya menjadi ekstensi .arff dengan menggunakan notepad++ atau dengan aplikasi lain. Seperti pada gambar 5.

1	@relation gizi	33	1,60,17.4,baik	65	1,23,11.2,baik
2	@attribute jenis_kelamin{p,l}	34	p,59,17.4,baik	66	p,23,9.3,baik
3	@attribute umur numeric	35	p,58,18.8,baik	67	1,22,10.6,baik
4	@attribute bb numeric	36	p,54,15.6,baik	68	1,21,10.5,baik
5	@attribute status{baik,kurang,buruk}	37	p,53,15,baik	69	1,21,10.7,baik
6		38	p,48,12.4,baik	70	1,20,10,baik
7	@data	39	p,47,14.2,baik	71	1,20,11.6,baik
8	1,48,16.2,baik	40	p,44,16.3,baik	72	1,20,10,baik
9	p,50,12.6,kurang	41	p,42,14.1,baik	73	p,20,9.9,baik
10	1,45,14.3,baik	42	p,41,12.3,baik	74	p,19,9.5,baik
11	1,43,16,baik	43	1,40,15,baik	75	1,18,8,kurang
12	p,43,13.3,baik	44	p,40,12.5,baik	76	1,17,10.7,baik
13	1,42,14.1,baik	45	1,39,12.3,baik	77	p,17,9,baik
14	1,38,11.2,kurang	46	1,38,16.1,baik	78	p,16,10,baik
15	1,31,11.9,baik	47	1,38,12.5,baik	79	p,16,8.2,baik
16	p,31,10.3,baik	48	1,36,14.8,baik	80	1,15,9.6,baik
17	p,28,9,baik	49	p,35,12.2,baik	81	p,14,8.4,baik
18	1,27,9.8,baik	50	1,35,12,baik	82	p,13,7.7,baik
19	p,25,9.5,baik	51	p,35,11.7,baik	83	p,9,8.2,baik
20	p,24,10.8,baik	52	p,32,13.2,baik	84	p,9,6.5,kurang
21	1,24,11.9,baik	53	p,32,13.5,baik	85	p,5,6.7,baik
22	1,23,9.4,kurang	54	1,32,12.6,baik	86	1,5,8.3,baik
23	1,23,10.2,baik	55	p,30,14,baik	87	1,5,6.5,baik
24	1,20,8.9,baik	56	p,30,10.4,kurang	88	p,4,5.8,baik
25	p,14,8.5,baik	57	1,30,14.2,baik	89	1,3,6,baik
26	1,12,7.6,baik	58	1,29,13.2,baik		
27	p,11,6.2,buruk	59	1,29,12.1,baik		
28	p,9,7.3,baik	60	p,28,10,kurang		
29	p,8,7.5,baik	61	1,28,13.1,baik		
30	p,8,7.3,baik	62	p,25,14.5,baik		
31	1,5,6.6,baik	63	1,24,10.5,baik		
32	1,1,5,baik	64	p,24,11.1,baik		

Gambar 5. Dataset dalam format .arff pada notepad++

Pengujian dengan algoritma *Decision Tree*(ID3) menggunakan aplikasi WEKA 3.8.1 dapat dilihat pada gambar 6.

Correctly Classified Instances	74	90.2439 %
Incorrectly Classified Instances	8	9.7561 %
Kappa statistic	0	
Mean absolute error	0.1188	
Root mean squared error	0.2437	
Relative absolute error	91.1769 %	
Root relative squared error	99.8297 %	
Total Number of Instances	82	

Gambar 6. Hasil Pengujian Data Menggunakan Algoritma ID3

Pengujian dengan algoritma Naive Bayesian Classification (NBC) menggunakan aplikasi WEKA 3.8.1 dapat dilihat pada gambar 7.

Correctly Classified Instances	75	91.4634 %
Incorrectly Classified Instances	7	8.5366 %
Kappa statistic	0.2126	
Mean absolute error	0.1061	
Root mean squared error	0.2243	
Relative absolute error	81.4719 %	
Root relative squared error	91.893 %	
Total Number of Instances	82	

Gambar 7. Hasil Pengujian Data Menggunakan Algoritma ID3

Pada hasil pengujian, dapat dilihat bahwa algoritma Naive Bayesian Classification memiliki nilai Correctly Classified Instances yang lebih besar yaitu dengan nilai 91,4634%. dibandingkan dengan algoritma Decision Tree(ID3) yang hanya sebesar 90,2439%.

4. KESIMPULAN

Berdasarkan hasil Pembahasan dan penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma Naive Bayesian Classification (NBC) memiliki akurasi yang lebih besar dengan nilai persentase sebesar 91,4634% dibandingkan dengan Algoritma Decision Tree (ID3) yang hanya mempunyai nilai persentase sebesar 90,2439%, sehingga algoritma Naive Bayesian Classification (NBC) dapat digunakan sebagai salah satu metode untuk klasifikasi status gizi yang baik berdasarkan data yang telah tersedia.

5. SARAN

Untuk penelitian mendatang, pemilihan algoritma klasifikasi selain ID3 dan Naive Bayes, seperti k-NN, C4.5, Support Vector Machine, Penggunaan variabel yang lain juga memungkinkan untuk kasus yang sama.

UCAPAN TERIMAKASIH

Penulis mengucapkan terimakasih kepada STMIK AMIKOM Purwokerto yang telah memberikan dukungan financial.

DAFTAR PUSTAKA

- [1] Ad Sediaoetama, Ilmu Gizi. Jakarta: Dian Rakyat, 2004
- [2] Soekirman, Ilmu Gizi dan Aplikasi untuk Keluarga dan Masyarakat. Jakarta : Departemen Pendidikan Nasional, 2000
- [3] Ariadni Ratih, Arieshanti Isye, 2015; "Implementasi Metode Pohon Keputusan Untuk Klasifikasi Data Dengan Nilai Fitur yang Tidak Pasti, Teknik Informatika, Institut Teknologi Sepuluh November, Surabaya, 60111".
- [4] Listiana Mila, Sudjalwo, dan Gunawan Dedi, 2015; "Perbandingan Algoritma Decision Tree(C4.5) dan Naive Bayes Pada Data Mining Untung Identifikasi Tumbuh Kembang Anak Balita, Jurusan Teknik Informatika, Universitas Muhammadiyah Surakarta
- [5] Tahyudin Imam, Utami Ema, dan Amborowati Armadyah, 2013; "Comparasion Clasification Algorithm Of Data mining to Predict the Graduation Students On Tim, Information Systems Iternational Conference(ISICO), 2-4 Desember 2013

- [6] Supriyanti Wiwit, Kusrini, Amborowati Armadyah, 2016; " Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes Untuk Ketepatan Pemilihan Konsentrasi Mahasiswa, Jurnal INFORMA Politeknik Indonusa Surakarta ISSN: 2442-7942 Vol. 1 Nomor 3 Tahun 2016
- [7] P. Nancy, Ramani, and R. Geetha. 2011. "A Comparison on Performance of Data Mining Algorithms in Clasification of Social Network Data," International Journal of computer Applications (0975 – 8887) Vol. 32 No. 8 October 2011.
- [8] M. Kumari and S. Godara, "Comparative Study of Data Mining Clasification Methods in Cardiovascular Diseases Prediction," IJCST Vol 2. Issue 2, June 2011. ISSN : 2229 – 4333(Print) – 0976 – 8491 (online), 2011.
- [9] Sharma, K. Aman and S. Sahni, "A Comparative Study of Clasification Algorithms for Spam Email Data Analysis," International Journal On Computer Science and Engineering (IJCSE). ISSN : 0975 – 3397 Vol. 3 No. 5 may 2011
- [10] D. Neslihan and T. Zuhail, "A comparative framework for evaluating clasification algorithms," Proceedings of The World Congress on Engineering 2010 Vol. I WCE 2010, June 30 – july 2, 2010, London, U.K. 2010.
- [11] S.M. Suhartinah and Ernastuti. 2010. "Graduation Prediction of Guna Darma University Students Using Algorithm Naive Bayes and C4.5 Algorithm," Jurnal Magister Sistem Informasi. Universitas Guna Darma. Jakarta. 2005.
- [12] Santosa, Budi. 2007. Data Mining : *Teknik Pemanfaatan Data untuk keperluan Bisnis*. Graha Ilmu. Yogyakarta.
- [13] Karim Masud, M.Rahman Rashedur. 2013. " Decision Tree and Naïve Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing" Journal Of Software Engineering and Aplications, 2013,06, 196-206