



Artikel

Prediksi Kemampuan Lulusan SMK untuk Dapat Bersaing Di Dunia Kerja dengan Menggunakan Naïve Bayes: Studi Kasus SMK Buddhi Tangerang

Lianny Wydiastuty Kusuma¹

¹ Universitas Buddhi Dharma, Teknik Informatika, Banten, Indonesia

JEJAK PENGIRIMAN

Received: 28 Agustus 2019

Final Revision: 25 September 2019

Available Online: 30 September 2019

KATA KUNCI

Naïve Bayes, prakerin, siswa SMK, data mining, kejuruan

KORESPONDENSI

E-mail: liannyw_k@yahoo.com

A B S T R A K

Siswa sekolah menengah kejuruan dituntut agar bisa siap terjun langsung setelah mereka lulus. Selama masa studi mereka diberi pembekalan khusus berupa belajar bekerja pada masa praktek kerja industri, kunjungan dunia usaha/dunia industri, dan pelajaran produktif sesuai jurusan yang mereka ambil. Mampu atau tidaknya siswa yang lulus mendapat pekerjaan dipengaruhi oleh beberapa faktor dalam mengikuti syarat perekrutan di sebuah perusahaan. Di antaranya yaitu nilai rata-rata rapor, nilai rata-rata Ujian Nasional, nilai Ujian Kompetensi Kejuruan, prakerin, disiplin, tanggung jawab, skill, sikap dan komunikasi dalam penelitian ini. Dalam penelitian ini akan menggunakan algoritma Naïve Bayes dalam mengetahui pola dari variabel-variabel tersebut. Pengujian dilakukan dengan Confusion Matrix dan kurva ROC. Hasil akurasi yang diperoleh membuktikan bahwa Naïve Bayes menghasilkan persentase akurasi 98% dan nilai AUC sebesar 0.980. Dengan demikian algoritma ini dapat memprediksi berapa banyak lulusan siswa SMK Buddhi Tangerang yang layak bekerja dengan baik.

PENDAHULUAN

Kualitas pendidikan pada siswa adalah salah satu tujuan dari sekolah. Sekolah akan meningkatkan kualitasnya dari tahun ke tahun sebagai acuan dalam menentukan keberhasilan sistem pendidikannya. Salah satu aspek sebagai indikator kualitas di sekolah adalah banyaknya lulusan siswa sekolah menengah kejuruan yang diterima di perusahaan ataupun industri, yang juga menunjukkan kualitas sekolah dalam memotivasi siswa untuk memberikan jaminan kesempatan bekerja yang lebih baik di masa mendatang. Berdasarkan latar belakang masalah yang ada, masalah penelitian ini dapat diidentifikasi dengan belum adanya aplikasi untuk metode memprediksi kemampuan lulusan Sekolah Menengah Kejuruan Buddhi di dunia kerja.

Penelitian ini hanya meneliti tentang tingkat kemampuan kerja siswa Sekolah Menengah Kejuruan Buddhi Tangerang dapat bersaing di dunia kerja berdasarkan nilai rapor, nilai ujian nasional, nilai UKK, prakerin, disiplin, tanggung jawab, skill, sikap dan komunikasi. Data yang digunakan adalah data siswa tahun 2010, 2011, 2012, dan 2013 yang diperoleh dari Tata Usaha dan Kurikulum Sekolah Menengah Kejuruan Buddhi Tangerang. Rumusan masalah yang dibentuk "Bagaimana aplikasi dengan algoritma Naïve Bayes dapat memprediksi kemampuan siswa Sekolah Menengah Kejuruan Buddhi Tangerang untuk bersaing di dunia kerja?".

I. DATA MINING

Data mining merupakan kegiatan melakukan ekstraksi untuk mendapatkan informasi penting yang sifatnya implisit dan sebelumnya tidak diketahui, dari suatu data. Data mining didefinisikan sebagai proses menemukan pola dalam data. Proses ini otomatis atau (biasanya) semi-otomatis [1].

II. NAÏVE BAYES

Pendekatan Bayesian digunakan untuk menentukan kemungkinan terhadap asumsi disekitarnya. Dalam statistik Bayesian, parameter dipertimbangkan terhadap variabel yang acak dan data dipertimbangkan terhadap hasil kemungkinan. Pendekatan Bayesian pertama kali dilakukan oleh Reverend Thomas Bayes (1702-1761) pada “Essay Towards Solving a Problem in the Doctrine of Chances” yang dipublikasikan tahun 1763 [2].

Tahapan dalam metode Naïve Bayes, yakni:

- Menghitung jumlah data
- Mencari nilai probabilitas (P)

$$P(x) = \frac{E}{n}$$

- Mencari nilai mean (μ)

$$\mu = \frac{\sum_{i=1}^n x_i}{n}$$

- Mencari nilai standar deviasi

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1}}$$

- Mengklasifikasi data kontinu dengan rumus Densitas Gauss

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}}$$

- Mencari nilai posterior

$$\text{Posterior}(X) = \frac{P(X) P(\text{Atribut1} | X) P(\text{Atribut2} | X) P(\text{atribut..n} | X)}{p(X) P(\text{Atribut1} | X) P(\text{Atribut2} | X) P(\text{atribut..n} | X) + p(Y) P(\text{Atribut1} | Y) P(\text{Atribut2} | Y) P(\text{atribut..n} | Y)}$$

$$\text{Posterior}(Y) = \frac{P(Y) P(\text{Atribut1} | Y) P(\text{Atribut2} | Y) P(\text{atribut..n} | Y)}{p(X) P(\text{Atribut1} | X) P(\text{Atribut2} | X) P(\text{atribut..n} | X) + p(Y) P(\text{Atribut1} | Y) P(\text{Atribut2} | Y) P(\text{atribut..n} | Y)}$$

III. TINJAUAN STUDI

No	Penelitian	Algoritma	Hasil
----	------------	-----------	-------

1.	SMS Filtering Menggunakan Naive Bayes Classifier dan FP-Growth Algorithm Frequent Itemset Dea Delvia Arifin dan M.Arif Bijaksana, 2015 [3]	Naïve Bayes dan FP-Growth	Kolaborasi dua algoritma menghasilkan akurasi rata-rata 98,506%.
2.	Perbandingan Kinerja Algoritma Klasifikasi Naïve Bayesian, Lazy-Ibk, Zero-R, dan Decision Tree-J48 Sulidar Fitri, 2014 [4]	Naïve Bayesian, Lazy-Ibk, Zero-R, dan Decision Tree-J48	Naïve Bayesian memiliki akurasi terbaik sebesar 85,12% (mode tes cross-validation).
3.	Implementasi Metode Klasifikasi Naïve Bayes dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga Alfa Saleh, 2015 [5]	Naïve Bayes	Akurat prediksi sebesar 78,33%
4.	Perbandingan Metode Probabilistik Naive Bayesian Classifier dan Jaringan Syaraf Tiruan Learning Vector Quantization dalam Kasus Klasifikasi Penyakit Kandungan Aditya Nugraha, Ristu Saptono dan Meiyanto Eko Sulisty, 2013 [6]	Naïve Bayes dan Jaringan Syaraf Tiruan	Akurasi pengujian Naïve bayes dengan Laplacian Smoothing dengan akurasi 88%, 92.4%, 92.8% dan 92.4%
5.	Algoritma Naive Bayes untuk Mencari	Naïve bayes	Tingkat kesalahan

	Perkiraan Waktu Studi Mahasiswa Arief Jananto, 2013 [7]		prediksi 20% - 34%.
6.	Perbandingan Jaringan Syaraf Tiruan dan Naive Bayes Dalam Deteksi Seseorang Terkena Penyakit Stroke I. Rohmana, 2014 [8]	Naive bayes dan Jaringan Syaraf Tiruan	Presentase Naive Bayes 80,55% dan Jaringan Syaraf Tiruan sebesar 71,11%
7.	Perbandingan Algoritma K-Nearest Neighbors dan Naive Bayes Untuk Studi Data "Wisconsin Diagnosis Breast Cancer" Paulus Dian Wicaksana, 2015 [9]	K-Nearest Neighbors dan Naive Bayes	Akurasi naive bayes 87,5% dan k-nearest neighbor 95,8%
8.	Algoritma Klasifikasi Naive Bayes untuk Menilai Kelayakan Kredit Claudia Clarentia Ciptohartono, 2013 [10]	Naive bayes	Akurasi sebesar 92,53%
9.	Perbandingan Kinerja Algoritma Decision Tree dan Naive Bayes dalam Prediksi Kebangkrutan Dian Oktafia dan D. L. Crispina Pardede, 2010 [11]	Naive bayes dan Decision Tree	Tingkat akurasi Naive Bayes 100% dan decision tree sebesar 96.97%.
10.	Perbandingan Klasifikasi Dokumen Teks Menggunakan Metode Naive Bayes dengan K-Nearest Neighbor	Naive bayes dan Decision Tree	Tingkat akurasi Naive Bayes 100% dan decision tree sebesar 96.97%.

	Cahyo Darujati, 2010 [12]		
--	---------------------------	--	--

IV. SMK BUDDHI

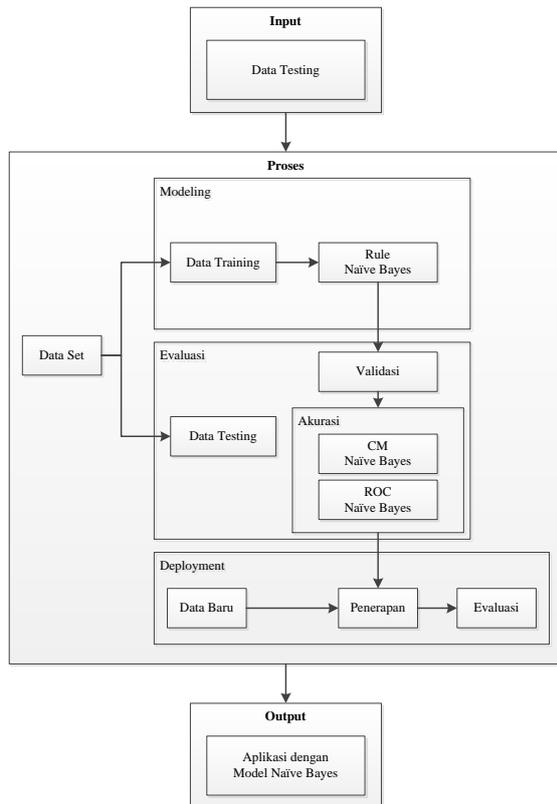
SMK Buddhi telah berdiri pada tahun 2010 dengan membuka 2 (dua) jurusan yaitu Multimedia dan Akuntansi. Sampai pada tahun 2016 ini tercatat sebanyak 114 Sekolah Menengah Kejuruan yang berdiri, baik negeri maupun swasta. Data ini diambil dalam situs resmi Dinas Pendidikan Kota Tangerang (<http://www.disdiktangerangkota.net/dataSekolah/dataMasterJenisSekolah/4>)

V. KEPUTUSAN PENERIMAAN BEKERJA

Pada penelitian ini, digunakan beberapa faktor pendukung yang dapat memprediksi tingkat diterimanya calon karyawan dalam hal ini lulusan siswa di dunia kerja. Berikut adalah faktor-faktor yang digunakan dalam penelitian ini, yaitu

- a. Rata-rata nilai rapor
- b. Nilai Ujian Nasional
- c. Nilai Ujian Kompetensi Kejuruan
- d. Praktek Kerja Industri
- e. Disiplin
- f. Tanggung Jawab
- g. Perilaku
- h. Komunikasi
- i. Keterampilan

VI. POLA PIKIR



VII. HIPOTESIS

Hipotesis dari penelitian ini adalah diduga aplikasi dengan algoritma Naïve Bayes ini dapat membantu SMK Buddhi memprediksi kemampuan lulusan siswa-siswa siap bekerja sehingga banyak perusahaan yang merekrut.

VIII. METODOLOGI PENELITIAN

Pada penelitian ini akan menggunakan dataset yang dijadikan data training maupun data testing sebanyak 100. Data training akan digunakan untuk pembentukan pola algoritma Naïve Bayes. Lalu data testing digunakan untuk menguji pola algoritma yang telah dibentuk. Proses penentuan mampu atau tidaknya seorang siswa tersebut dengan cara membagi rata semua dataset yang ada, lalu jika hasilnya lebih besar sama dengan 80, dianggap mampu, sedangkan jika hasil lebih kecil dari 80, maka dianggap belum mampu bersaing di dunia kerja. Dalam pengembangannya akan dibuat sebuah aplikasi dengan VB.net. Hasil penelitian ini

adalah implementasi algoritma Naïve Bayes dalam memprediksi kemampuan lulusan siswa SMK Buddhi dapat bersaing di dunia kerja atau industri.

IX. METODE PEMILIHAN SAMPLE

Data yang digunakan dalam penelitian ini merupakan data primer, data yang digunakan diambil berdasarkan data siswa Sekolah Menengah Kejuruan Buddhi tahun angkatan 2010, 2011, 2012 dan 2013.

X. METODE PENGUMPULAN DATA

Metode pengumpulan data yang digunakan dalam penelitian ini adalah:

- Metode Observasi
- Metode Studi Pustaka

XI. INSTRUMENTASI PENGUJIAN DAN IMPLEMENTASI

- Perangkat Keras
 - Processor Intel(R) Core(TM) i7-4700HQ CPU @ 2.40GHz
 - RAM sebesar 8 GB
 - Storage internal sebesar 1 TB
 - Sistem Operasi Windows 8.1 Pro 64-bit
- Perangkat Lunak

Perangkat lunak yang digunakan dalam penelitian ini adalah RapidMiner 7.

XII. PENGUJIAN DATASET

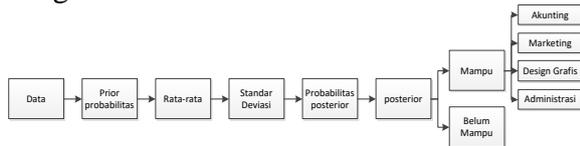
Semua dataset kemudian diuji dengan metode yang diusulkan pada aplikasi RapidMiner 7.

XIII. PENGUJIAN DENGAN SQA (SOFTWARE QUALITY ASSURANCE)

No	Metrik
1	Auditability
2	Accuracy
3	Completeness
4	Error Tolerance
5	Execution Efficiency
6	Operability
7	Simplicity
8	Training

XIV. GAMBARAN UMUM

Langkah pertama dalam penelitian ini adalah pemakaian algoritma Naïve Bayes untuk membentuk sebuah model. Model yang dihasilkan akan digunakan untuk mengetahui pola kemampuan lulusan siswa yang memiliki kemungkinan mampu dan belum mampu. Siswa yang mampu akan dikategorikan ke dalam empat bagian pekerjaan yang sesuai dengan keahlian mereka.



XV. PENGUKURAN DENGAN CONFUSION MATRIX

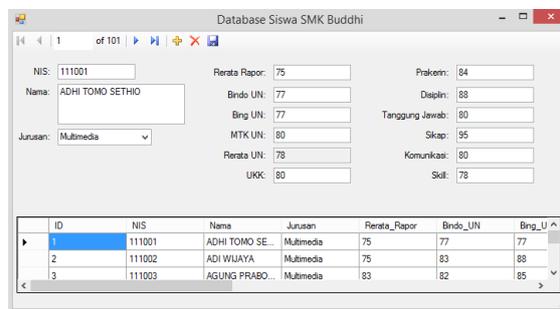
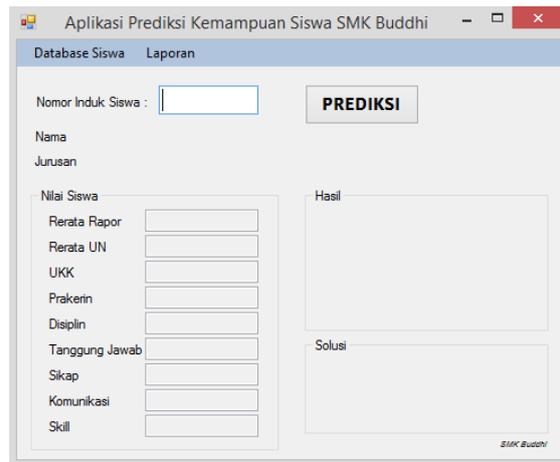
	True Belum mampu	True Mampu	Precision
Pred. Belum Mampu	48	0	100.00%
Pred. Mampu	2	50	96.15%
Class recall	96.00%	100.00%	

$$\begin{aligned}
 \text{akurasi} &= \frac{(\text{TBelum Mampu} + \text{TMampu})}{(\text{TBelum Mampu} + \text{FBelum Mampu} + \text{TMampu} + \text{FMampu})} \\
 &= \frac{(48 + 50)}{(48 + 2 + 0 + 50)} \\
 &= \frac{98}{100} = 0.98 = 98\%
 \end{aligned}$$

XVI. PENGUKURAN AUC (AREA UNDER CURVE) DENGAN KURVA ROC (RECEIVER OPERATING CHARACTERISTIC)



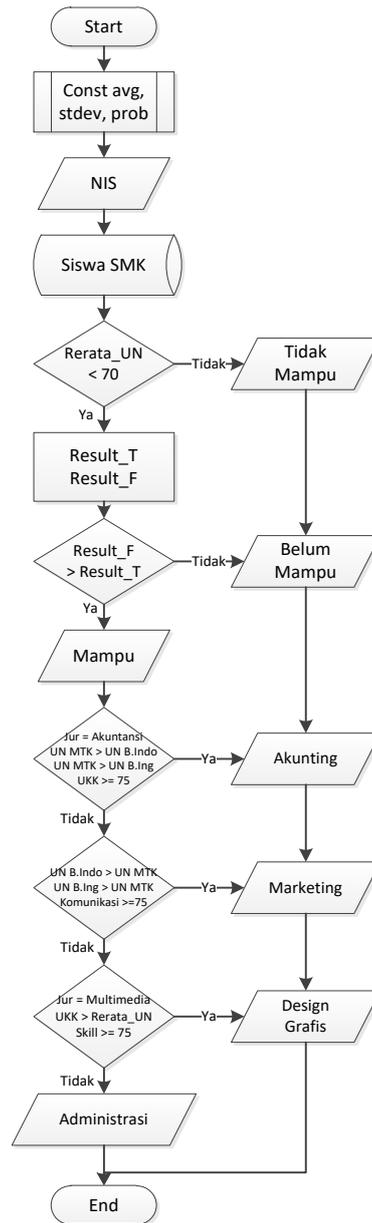
XVII. PENERAPAN ALGORITMA



Dari laporan jurusan akuntansi dengan siswa sebanyak 47, diperoleh 22 siswa diprediksi belum mampu, 25 siswa lainnya diprediksi mampu dengan kategori pekerjaan bagian akunting sebanyak 3 siswa, kategori marketing sebanyak 4 siswa, dan kategori administrasi sebanyak 18 siswa. Untuk laporan jurusan multimedia dengan siswa sebanyak 54, diperoleh 27 siswa diprediksi belum mampu, 27 siswa lainnya diprediksi mampu dengan kategori pekerjaan bagian design grafis sebanyak 9 siswa, kategori marketing sebanyak 5 siswa, dan kategori administrasi sebanyak 13 siswa. Total keseluruhan dari kedua jurusan sebanyak 101

siswa diprediksi 52 siswa mampu dan 49 siswa belum mampu.

No	Nama Field	Type	Size
1	ID	AutoNumber	Long int
2	NIS	Short Type	10
3	Nama	Short Type	40
4	Jurusan	Short Type	10
5	Rerata_Rapor	Number	Byte
6	Bindo_UN	Number	Byte
7	Bing_UN	Number	Byte
8	MTK_UN	Number	Byte
9	Rerata_UN	Number	Byte
10	UKK	Number	Byte
11	Prakerin	Number	Byte
12	Disiplin	Number	Byte
13	Tanggung_Jawab	Number	Byte
14	Sikap	Number	Byte
15	Komunikasi	Number	Byte
16	Skill	Number	Byte



User	Skor Metrik								Skor
	1	2	3	4	5	6	7	8	
1	100	80	80	100	100	80	100	60	88
2	80	100	80	80	80	80	60	80	80
3	100	80	80	60	100	80	100	80	85
4	80	80	80	80	80	80	80	80	80
5	100	80	100	80	100	60	100	80	88
6	80	80	100	80	100	80	100	80	88
7	80	60	80	80	80	80	80	80	78
Rata-Rata									84

XVIII. RENCANA IMPLEMENTASI

No	Jenis Kegiatan	Minggu Ke					
		1	2	3	4	5	6
1.	Analisa Data	X					
2.	Perancangan dan Pembuatan Aplikasi		X	X			
3.	Pelatihan				X		
4.	Implementasi				X	X	
5.	Evaluasi					X	X

XIX. KESIMPULAN

Kesimpulan yang dapat diambil dari penelitian ini adalah berdasarkan hasil evaluasi dan validasi, diketahui bahwa algoritma Naïve Bayes memiliki akurasi dan performa baik dengan nilai accuracy sebesar 98% dan nilai AUC sebesar 0.980. Skor kesimpulan aplikasi berdasarkan SQA sebesar 84 yang dinilai cukup baik.

XX. SARAN

- a. Menggunakan algoritma klasifikasi yang lain yang terdapat dalam data mining, seperti algoritma K-Nearest Neighbor, ID3, CART, Random Forest, Linear Discriminant Analysis dan Neural Network.
- b. Menambah algoritma optimasi seperti PSO (Particle Swarm Optimization), Backward Elimination, Forward Selection, dan GA (Genetic Algorithm).

REFERENCES

- [1] Witten et al., 2011. *Data Mining Practical Machine Learning Tools and Techniques* 3rd ed., Burlington: Elsevier Inc.
- [2] Larose, D.T., 2005. *Discovering Knowledge In Data : An Introduction to Data Mining*, New Jersey: John Wiley & Sons, Inc.
- [3] Arifin, D.D. & Bijaksana, M.A., 2015. SMS Filtering Menggunakan Naive Bayes Classifier dan FP-Growth Algorithm Frequent Itemset. , pp.1–8.
- [4] Fitri, S., 2014. Perbandingan Kinerja Algoritma Klasifikasi Naïve Bayesian , Lazy-Ibk , Zero-R , dan Decision Tree- J48. , 15(1), pp.33–37.
- [5] Saleh, A., 2015. Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. , 2(3), pp.207–217.
- [6] Nugraha, P.A., Saptono, R. & Sulisty, M.E., 2013. Perbandingan Metode Probabilistik Naive Bayesian Classifier dan Jaringan Syaraf Tiruan Learning Vector Quantization dalam Kasus Klasifikasi Penyakit Kandungan. , 2(2), pp.20–33.
- [7] Jananto, A., 2013. Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa. , 18(1), pp.9–16.
- [8] Rohmana, I., 2014. Perbandingan Jaringan Syaraf Tiruan dan Naive Bayes dalam Deteksi Seseorang Terkena Penyakit Stroke. , 37(2), pp.178–191.
- [9] Wicaksana, P.D., 2015. Perbandingan Algoritma K-Nearest Neighbors dan Naive Bayes Untuk Studi Data “ Wisconsin Diagnosis Breast Cancer ” A Comparison Algorithm Between K-Nearest “ Wisconsin Diagnosis Breast Cancer .”
- [10] Ciptohartono, C.C., 2013. Bayes Untuk Menilai Kelayakan Kredit. , pp.1–6.
- [11] Oktafia, D. & Pardede, D.L.C., 2010. Perbandingan Kinerja Algoritma Decision Tree dan Naive Bayes dalam Prediksi Kebangkrutan.
- [12] Darujati, C., 2010. Perbandingan Klasifikasi Dokumen Teks Menggunakan Metode Naïve Bayes dengan K-Nearest Neighbor Abstrak. , 13(1), pp.1–9.

BIOGRAPHY

Lianny Wydiastuty Kusuma pada tahun 2011 lulus S1 STMIK Buddhi Program Studi Teknik Informatika dan tahun 2016 lulus S2 di Universitas Budi Luhur. Pada tahun 2016 bekerja di Universitas Buddhi Dharma sebagai dosen Program Studi Teknik Informatika.