



ANALISIS DATA MINAT CUSTOMER TERHADAP PRODUK DENGAN MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER (STUDI KASUS: PT JELLYFISH EDUCATION INDONESIA)

Agung Budi Susanto¹, Sudarno Wiharjo², Tika Liyana³

^{1,2,3} Pasca Sarjana Teknik Informatika, Universitas Pamulang

dosen02680@unpam.ac.id¹, sudarnogm@gmail.com², tikaliyana01@gmail.com³

Kata kunci:	Abstrak
Minat, Customer, Produk, Algoritma Naive Bayes Classifier	PT Jellyfish Education Indonesia merupakan sebuah perusahaan yang bergerak di bidang pendidikan kursus bahasa Jepang dengan berbagai produk yang ada, mulai dari produk private, regular/grup, semi private, bisnis, percapakan, kids, intensive. PT Jellyfish Education Indonesia mengalami kesulitan dalam mengelola data-data customer yang masuk dalam menentukan minat customer dalam memilih produk yang ada, sehingga produk mana yang paling diminati oleh customer tidak bisa diketahui oleh perusahaan dan menyebabkan kesulitan dalam menentukan produk mana yang akan dipasarkan atau dijual. Untuk mendukung dalam menentukan produk mana yang akan dipasarkan dan diprioritaskan perlu dilakukan sebuah penelitian dalam memprediksi minat customer yang masuk. Metode Naive Bayes Classifier membantu memberikan solusi dalam memprediksi minat customer terhadap produk yang dijual berdasarkan data pada bulan sebelumnya. Penelitian secara observasi terhadap data perusahaan tersebut mendapatkan hasil dengan menggunakan 222 data customer dan dianalisis menggunakan metode Naive Bayes Classifier yang selanjutnya diproses dengan aplikasi rapidminer yang memperoleh akurasi (rasio prediski benar dengan keseluruhan data) 72.73%.

Pendahuluan

Dalam persaingan di dunia Pendidikan sangatlah cepat tak terkecuali sebuah Lembaga kursus bahasa sebagai salah satu pendukung di dunia Pendidikan itu sendiri. Jellyfish Education Indonesia merupakan salah satu usaha yang bergerak dibidang Pendidikan sebagai Lembaga kursus Bahasa Jepang, dimana memiliki aktivitas penjualan produk yang banyak yaitu dari tahun 2016 sampai dengan sekarang. Dengan adanya data customer yang masuk setiap hari, data semakin lama akan semakin bertambah banyak. Data tersebut tidak hanya berfungsi sebagai arsip, data tersebut dapat dimanfaatkan dan diolah menjadi informasi yang berguna untuk meningkatkan penjualan produk. Permasalahan yang sering timbul yaitu sering sekali customer kesulitan dalam memilih produk yang ada, karena pemahaman mengenai produk serta target yang akan diambil dari customer untuk penyesuaian dari produk yang dipilih. Pada dasarnya kumpulan data customer memiliki informasi-informasi yang bermanfaat, yang bisa digunakan untuk mengambil sebuah keputusan dan untuk memperoleh pengetahuan baru tentang pola ketertarikan customer terhadap produk yang dijual. Pengolahan data tersebut dapat bisa dilakukan dengan menggunakan teknik-teknik

tertentu. Dengan permasalahan diatas peneliti ingin melakukan penelitian dengan menggunakan salah satu teknik yang digunakan dalam pengolahan data yaitu menggunakan metode Algoritma Naïve Bayes Classifier (NBC) dikarenakan dapat digunakan dalam penelitian kualitatif maupun kuantitatif, mampu digunakan dalam data dengan jumlah besar maupun sedikit, dan independen atau tidak tergantung dengan atribut yang ada. Dibandingkan dengan metode lain, contohnya seperti K-means harus menentukan nilai K dari awal dan sangat tergantung dengan inisialisasi penilaian awal. Hanya saja kekurangan dari Algoritma Naïve Bayes Classifier (NBC) sendiri yaitu apabila terjadi nilai 0 (kosong) maka tidak bisa dihitung walaupun bisa diatasi dengan dengan teknik penghalusan atau dilakukan pengisian nilai tersebut agar tidak 0 (kosong). Seperti pada penelitian yang dilakukan oleh Herry Derajad Wijaya dan Saruni Dwiasnati dengan judul implementasi data mining dengan algoritma Naive Bayes pada penjualan obat yang bertujuan untuk menganalisis masalah-masalah pada penentuan sebuah produk mana yang di kategorikan laku dan tidak laku, dengan sebuah kesimpulan mengetahui pola penjualan obat yang laku dan tidak laku berdasarkan jenis variasi obat yang baru (Wijaya & Dwiasnati, 2020). Algoritma Naive Bayes Classifier (NBC) adalah Jadhav et all (2016) menyatakan bahwa Naive Bayes Classifier adalah suatu model independen yang membahas mengenai klasifikasi sederhana berdasarkan teorema Bayes. Naive Bayes merupakan suatu algoritma yang dapat mengklasifikasikan suatu variable tertentu dengan menggunakan metode probabilitas dan statistik (Kurniawan, 2018). Dengan latar belakang diatas maka penulis terdorong untuk melakukan penelitian lebih lanjut terhadap permasalahan yang dihadapi oleh PT Jellyfish Education Indonesia. Penelitian ini sebagai Tesis sebagai salah satu persyaratan kelulusan pada Program Studi Magister Teknik Informatika, Universitas Pamulang. Penulis mengambil judul “Analisis data minat customer terhadap produk dengan menggunakan Algoritma Naive Bayes Classifier (Studi kasus: PT Jellyfish Education Indonesia)”.

Metode

Pada tahap ini akan dilakukan analisis kebutuhan dan spesifikasi atas masalah yang akan diselesaikan.

Jenis Penelitian

Jenis penelitian yang dilakukan dalam penelitian ini bersifat asosiatif yang bertujuan untuk mengetahui hubungan antara dua variable atau lebih yang berfungsi untuk menjelaskan, meramalkan, dan mengontrol fenomena (Sugiyono, 2013).

Penelitian Kualitatif

Penelitian kualitatif adalah penelitian yang bersifat deskriptif dan cenderung menggunakan analisis. Proses dan makna lebih ditonjolkan dalam penelitian kualitatif.

Penelitian Kuantitatif

Penelitian kuantitatif adalah penelitian ilmiah yang sistematis terhadap bagian-bagian dan fenomena serta kausalitas hubungan-hubungannya. Tujuan penelitian kuantitatif adalah mengembangkan dan menggunakan model-model matematis

Penelitian Observasi

Pengumpulan data dengan melakukan pengamatan secara langsung terhadap objek penelitian, terhadap kegiatan yang sedang berlangsung terutama objek yang sedang diteliti.

Study Kepustakaan

Dalam studi kepustakaan ini penulis mengumpulkan dan mempelajari berbagai teori dan konsep dasar yang berhubungan dengan masalah yang diteliti. Teori dan konsep dasar tersebut penulis peroleh dengan cara menelaah berbagai macam bacaan seperti buku, jurnal, dan bahan bacaan relevan lainnya.

Metode Pemilihan Sampel

Metoda pemilihan sampel merupakan cara yang dilakukan dalam mengambil sampel yang akan dijadikan bahan penelitian. Roscoe (1975) yang dikutip Uma Sekaran (2006) memberikan acuan umum untuk menentukan ukuran sampel:

1. Ukuran sampel lebih dari 30 dan kurang dari 500 adalah tepat untuk kebanyakan penelitian
2. Jika sampel dipecah ke dalam subsampel (pria/wanita, junior/senior, dan sebagainya), ukuran sampel minimum 30 untuk tiap kategori adalah tepat
3. Dalam penelitian multivariate (termasuk analisis regresi berganda), ukuran sampel sebaiknya 10x lebih besar dari jumlah variabel dalam penelitian
4. Untuk penelitian eksperimental sederhana dengan kontrol eksperimen yang ketat, penelitian yang sukses adalah mungkin dengan ukuran sampel kecil antara 10 sampai dengan 20.

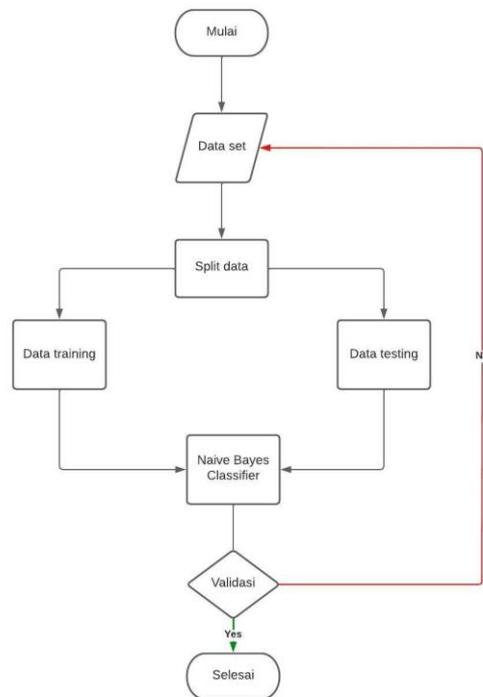
Rumus Slovin (dalam Riduwan, 2005:65)

$$n = \frac{N}{1 + Ne^2}$$

n = sampel; N = populasi; e = error margin

Gambar.1 Solvin Step

Flowchart



Gambar.2 Flowchart

1. Dimulai dengan menginput data set customer
2. Melakukan proses klasifikasi data set dan split data (data training & data testing)
3. Menghitung data set menggunakan Algoritma Naïve Bayes Classifier
4. Melakukan Validasi, apakah Akurasi yang diinginkan sudah sesuai atau tidak.
5. Jika Tidak, maka proses kembali ke data set untuk melakukan perbaikan data. Jika Ya maka Proses selesai

Analisis Data

Setelah data tersebut dikumpulkan, kemudian datatersebut dianalisis dengan menggunakan teknik pengelolaan data. Analisis data yang digunakan oleh penulis dalam

penelitian ini bertujuan untuk menjawab pertanyaan yang tercantum dalam identifikasi masalah.

Analisis data merupakan salah satu kegiatan penelitian berupa proses penyusunan dan pengelolaan data guna menafsirkan data yang telah diperoleh. Menurut Sugiyono (2016:147).

1. Persiapan Data (Data Preparation)

Adalah mempersiapkan data dan atribut yang akan digunakan selama penelitian. Data yang digunakan memiliki 1 variabel sebagai kelas yaitu status minat customer PT Jellyfish Education Indonesia “minat” dan “tidak minat”.

2. Pemodelan (Modelling)

Dalam penelitian ini akan dilakukan pengujian menggunakan klasifikasi dengan algoritma Naive Bayes untuk mengetahui hasil perhitungan yang dianalisa dan untuk mengetahui apakah fungsi bekerja dengan baik atau tidak.

3. Evaluasi (Evaluation)

Adalah melakukan evaluasi terhadap data yang sudah ada dan dilakukan pengecekan kembali.

Metode Naïve Bayes Classifier

Bayes merupakan teknik prediksi berbasis probabilitas sederhana yang berdasar pada penerapan teorema Bayes (atau aturan Bayes) dengan asumsi independensi (ketidaktergantungan) yang kuat (naif). Dengan kata lain, dalam Naïve Bayes, model yang digunakan adalah “model fitur independen” Dalam Bayes (terutama Naïve Bayes), maksud independensi yang kuat pada fitur adalah bahwa sebuah data tidak berkaitan dengan ada atau tidaknya fitur lain dalam data yang sama (Siska, 2019)

Teorema Bayes memiliki bentuk umum sebagai berikut:

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

Dalam hal ini :

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 probability)

$P(H)$ = Probabilitas hipotesis H (Prior probability)

$P(X|H)$ = Probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$ = Probabilitas dari X.

Gambar.3 Teorema Bayes

RapidMiner

Menurut (Fatmawati, 2016) Rapid Miner merupakan perangkat lunak yang dibuat oleh Dr. Markus Hofmann dari Institute of Technology Blanchardstown dan Raif Klinkenberg dari rapid-i.com dengan tampilan GUI (Graphical User Interface) sehingga memudahkan pengguna dalam menggunakan perangkat lunak ini. Perangkat lunak ini bersifat open source dan dibuat dengan menggunakan bahasa Java dibawah lisensi GNU Public License dan Rapid Miner dapat dijalankan disistem operasi manapun. Dengan menggunakan Rapid Miner, tidak dibutuhkan kemampuan koding khusus, karena semua fasilitas sudah disediakan.

Hasil dan Pembahasan

a. Atribut Penelitian

Tabel.1 Atribut Penilaian

Nama Atribut	Nilai Atribut
Jenis Kelas	private, regular, semi private, bisnis, percakapan.
Informasi	ig, brosur, baliho, teman, seminar.
Harga	Sangat tinggi, tinggi, sedang, sangat sedang, rendah.
Target	kuliah, bekerja, skill, hobi, liburan.
Level bahasa	N5, N4, N3, N2, N1

b. Perhitungan Probabilitas Prior (P(Ci))

Dataset akan diproses dengan metode Naive Bayes dengan beberapa tahapan sehingga perlu menentukan besarnya nilai dari data, yaitu dari 222 data latih yang digunakan, diketahui kelas C₀ (minat melanjutkan kursus) sebanyak 142 data, dan kelas C₁ (tidak minat melanjutkan kursus) sebanyak 80 data. Perhitungan probabilitas prior yaitu (jumlah minat/tidak minat dibagi dengan keseluruhan data).

Tabel.2 Probabilitas Prior

Perhitungan prior		Hasil
P (C ₀)	142/222	0,64
P (C ₁)	80/222	0,36

c. Perhitungan Probabilitas Posterior X bersyarat C ($P(X|C_i)$)

Tabel.3 Probabilitas Posterior

Probabilitas jenis kelas

<u>X2</u> <u>jenis kelas</u>	<u>Jumlah kejadian "Dipilih"</u>		<u>Probabilitas</u>	
	<u>minat</u>	<u>tidak minat</u>	<u>minat (C0)</u>	<u>tidak minat (C1)</u>
Private	40	11	0.28169	0.1375
Regular	64	36	0.450704	0.45
Semi Private	20	12	0.140845	0.15
Bisnis	6	11	0.042254	0.1375
Percakapan	12	10	0.084507	0.125
<u>Jumlah</u>	142	80	1	1

Probabilitas informasi

<u>X3</u> <u>Informasi</u>	<u>Jumlah kejadian "Dipilih"</u>		<u>Probabilitas</u>	
	<u>minat</u>	<u>tidak minat</u>	<u>minat (C0)</u>	<u>tidak minat (C1)</u>
Ig	21	20	0.147887	0.25
Web	66	10	0.464789	0.125
Brosur	17	18	0.119718	0.225
Baliho	22	11	0.15493	0.1375
Teman	16	21	0.112676	0.2625
<u>Jumlah</u>	142	80	1	1

Probabilitas target

<u>X5</u> <u>target</u>	<u>Jumlah kejadian "Dipilih"</u>		<u>Probabilitas</u>	
	<u>minat</u>	<u>tidak minat</u>	<u>minat (C0)</u>	<u>tidak minat (C1)</u>
Hobi	19	18	0.133803	0.0225
Liburan	4	4	0.028169	0.05
Skill	102	48	0.71831	0.6
Kuliah	6	2	0.042254	0.025
Bekerja	11	8	0.077465	0.1
<u>Jumlah</u>	142	80	1	1

Probabilitas level bahasa

<u>X6</u> <u>Level bahasa</u>	<u>Jumlah kejadian "Dipilih"</u>		<u>Probabilitas</u>	
	<u>minat</u>	<u>tidak minat</u>	<u>minat (C0)</u>	<u>tidak minat (C1)</u>
N5	49	29	0.34507	0.3625
N4	29	33	0.204225	0.4125
N3	48	8	0.330828	0.1
N2	15	4	0.105634	0.05
N1	1	6	0.007042	0.075
<u>Jumlah</u>	142	80	1	1

a. Hasil Ujian

Persamaan (1)

$$P(C0|X) = P(X|C0)*P(C0)$$

Hasil dari persamaan (1) akan dikalikan untuk mengetahui jumlah data minat:

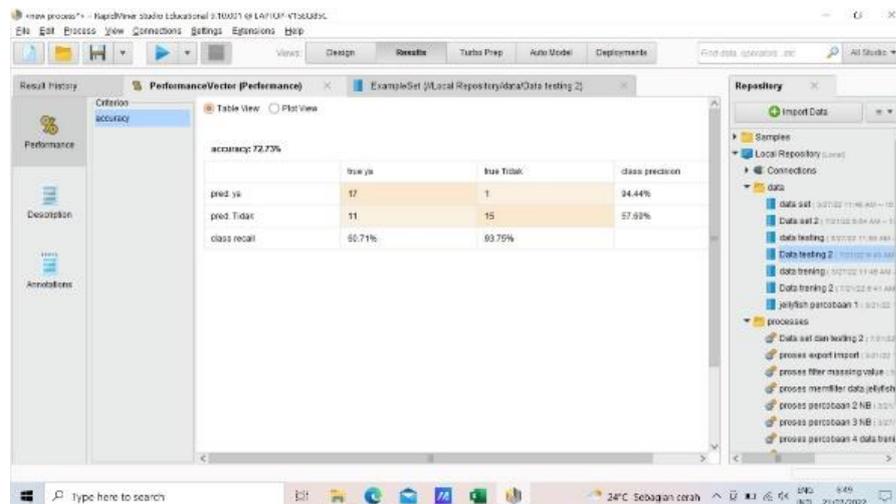
$$P(C0|X) = 0.478*0.450*0.147*0.478*0.718*0.330*0.6396 \\ = 0,0023$$

Persamaan (2)

$$P(C1|X) = P(X|C1)*P(C1)$$

Hasil dari persamaan (2) akan dikalikan untuk mengetahui jumlah data tidak minat:

$$P(C1|X) = 0,5*0,45*0,25*0,45*0,6*0,1*0,36036 \\ = 0,0005$$



	minat	tidak minat	class precision
accuracy: 72.73%			
pred. ya	17	1	34.44%
pred. tidak	11	15	57.89%
class recall	60.71%	93.75%	

Kesimpulan

Berdasarkan data yang sudah di uji menggunakan aplikasi Rapidminer menunjukkan tingkat akurasi 72.73% yang berarti “minat” terhadap produk di PT Jellyfish Education Indonesia. Dengan jumlah 17 data (True Positive) minat melanjutkan kursus bahasa Jepang, dan jumlah prediksi tidak minat melanjutkan kursus bahasa Jepang yang diklasifikasikan oleh classifier yaitu 1 data (False Positive). Dengan class precision (rasio prediksi benar positif dengan keseluruhan hasil yang diprediksi positif) 94.44%.

Berdasarkan interpretasi hasil penelitian yang didapat melalui data yang tersedia dari PT Jellyfish Education Indonesia, ada beberapa masukan yang belum dapat diimplementasikan dalam penelitian ini seperti latar belakang customer yang diukur melalui tujuan belajar, usia yang minat dalam kursus Bahasa, prediksi kelulusan pengambilan sertifikat Bahasa. Sarannya agar masukan tersebut dapat diimplementasikan pada penelitian pengembangan analisis minat customer terhadap produk di PT Jellyfish Education Indonesia dimasa mendatang.

Daftar Pustaka

- Abdurahman, M. (2018). Sistem Informasi Data Pegawai Berbasis Web Pada Kementerian Kelautan Dan Perikanan Kota Ternate. *Jurnal Ilmiah ILKOMINFO - Ilmu Komputer & Informatika*, 1(2), 70–78. <https://doi.org/10.47324/ilkominfo.v1i2.10>
- Afiatin, N. (2015). Pengaruh Perhatian Orang Tua Dan Minat Belajar Siswa Terhadap Prestasi Belajar Ilmu Pengetahuan Sosial. *Faktor Jurnal Ilmiah Kependidikan*, II(1), 1–9.
- Andoyo, A., & Rianto, R. (2015). Program Aplikasi Nilai Siswa Pada SMK Muhammadiyah Pringsewu Sebagai Penunjang Pengambilan Keputusan Siswa Berprestasi Menggunakan Visual Basic 6 . o. *Jurnal TAM (Technology Acceptance Model)*, 5(2), 58–65. <http://www.ojs.stmikpringsewu.ac.id/index.php/JurnalTam/article/view/49/49>
- Artanti, D. P., Syukur, A., Prihandono, A., & Setiadi, D. R. I. M. (2018). Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes. 8–9.
- Badii, Y. A., & Kuncoro, D. (2017). Perancangan Sistem Informasi Penjualan Dan Pembelian Serta Pengendalian Stok Barang Pada Toko Batik Kusumawardani Solo. *Jurnal Ilmiah Go Infotech*, 23(1), 19–26.
- Bagus, A., & Dinata, J. (2018). Media Pembelajaran Interaktif Untuk Materi Bahasa Jepang Level Dasar. 2(1), 438–445. <https://ejournal.itn.ac.id/index.php/jati/article/view/1702>
- Faisal, M. R., Afidh, R. F. F., & Arianto, A. (2017). Sistem Pengolahan Data Pelanggan Pada Tv Kabel Pt. Aneka Vision Dumai Menggunakan Visual Basic 6.o. *I N F O R M a T I K A*, 9(1), 49.
- Fanani, M. R. (2020). Algoritma Naïve Bayes Berbasis Forward Selection Untuk Prediksi Bimbingan Konseling Siswa. *Jurnal DISPROTEK*, 11(1), 13–22. <https://doi.org/10.34001/jdpt.v11i1.952>
- Jihad, M. A. (2019). Pemanfaatan Metode Technique for Order Preference By Similiarity To Ideal Solution (Topsis) Untuk Menentukan Pelanggan Terbaik. *Jurnal Informasi Dan Komputer*, 7(1), 1–6. <https://doi.org/10.35959/jik.v7i1.117>
- Kurniawan, Y. I. (2018). Perbandingan Algoritma Naive Bayes dan C.45 dalam Klasifikasi Data Mining. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(4), 455. <https://doi.org/10.25126/jtiik.201854803>
- Manalu, E., Sianturi, F. A., & Manalu, M. R. (2017). Penerapan Algoritma Naive Bayes Untuk Memprediksi Jumlah Produksi Barang Berdasarkan Data Persediaan dan Jumlah Pemesanan Pada CV. Papadan Mama Pastries. *Jurnal Mantik Penusa*, 1(2), 16–21. <https://ezp.lib.unimelb.edu.au/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=ffh&AN=2008-10-Aa4022&site=eds-live&scope=site>
- Nofriansyah, D., Erwansyah, K., & Ramadhan, M. (2016). Penerapan Data Mining dengan Algoritma Naive Bayes Clasifier untuk Mengetahui Minat Beli Pelanggan terhadap Kartu Internet XL (Studi Kasus di CV. Sumber Utama Telekomunikasi). *Jurnal Saintikom*, 15(2), 81–92.