



Comparison of Levels of Alcoholic Drinks in the World Using the Naïve Bayes Algorithm and K-Nearest Neighbor

Perbandingan Tingkat Kadar Minuman Beralkohol di Dunia Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbor

Aldi Saputra¹, Rony Paringga Hasibuan², Renaldi³, Rahmaddeni⁴

^{1,2,3,4}Program Studi Teknik Informatika, STMIK Amik Riau, Pekanbaru

E-Mail: ¹2110031802037@sar.ac.id, ²2110031802068@sar.ac.id,
³2110031802060@sar.ac.id, ⁴rahmaddeni@sar.ac.id

Corresponding Author: Rahmaddeni

Abstract

Alcohol is a substance contained in liquor that is widely consumed by most countries with cold climates, but it is also possible for countries with tropical climates. There are many alcoholic drinkers from all over the world, the level of alcohol content varies from country to country. There are several types of alcoholic beverages that are commonly drunk and circulated in the world, one of which is beer, spirit, and wine. In terms of the level of alcohol content of these types of drinks, there are limits on the levels that can be consumed by everyone to minimize the adverse effects of consuming alcohol in excess. This study aims to produce information about the comparison of high levels of alcoholic beverages and low levels of alcohol. in every country. This study obtained an information on the comparison of alcohol levels to minimize excessive consumption of alcoholic beverages in that country. Therefore, a comparative study was conducted using 2 algorithms to obtain information that has the maximum level of accuracy. In the comparative study, the algorithm adopted is using the Naive Bayes algorithm and K-NN with modeling using rapid miner applications. By using these 2 algorithms, it will produce a comparison of the types of alcoholic beverages consumed throughout the world based on the level of alcohol content contained in each drink. The data used in this study were 192 data consisting of 30 test data and 172 training data. The accuracy results obtained from the Naive Bayes method are 97,44% and 100% from the KNN method.

Keywords: Alcohol, KNN, Naive bayes, Rapid miner

Abstrak

Alkohol merupakan zat yang terkandung dalam minuman keras yang banyak di konsumsi oleh kebanyakan negara beriklim dingin tetapi tidak menutup kemungkinan juga untuk negara-negara yang beriklim tropis. Ada banyak peminum alkohol dari seluruh dunia, tingkat kandungan alkohol bervariasi dari satu negara ke negara lain. Ada beberapa jenis minuman beralkohol yang umum diminum dan beredar di dunia, salah satunya adalah bir, spirit, dan wine. Dalam hal tingkat kadar alkohol jenis minuman tersebut terdapat batasan tingkat kadar yang dapat dikonsumsi oleh setiap orang untuk meminimalisir dampak buruk dalam mengkonsumsi alkohol secara berlebihan, Penelitian ini bertujuan untuk menghasilkan sebuah informasi tentang perbandingan kadar jenis minuman beralkohol yang tinggi dan memiliki tingkat kadar alkohol yang rendah di setiap negara. Penelitian ini memperoleh sebuah informasi perbandingan tingkat kadar alkohol untuk meminimalisir konsumsi minuman beralkohol secara berlebihan pada negara tersebut. Oleh karena itu, maka dilakukanlah sebuah studi banding yang menggunakan 2 algoritma untuk mendapat sebuah informasi yang memiliki tingkat keakuratan yang maksimal. Pada studi banding, algoritma yang di angkat yaitu menggunakan algoritma Naive Bayes dan K-NN dengan pemodelan menggunakan aplikasi rapid miner. Dengan menggunakan 2 algoritma ini akan menghasilkan perbandingan jenis minuman beralkohol yang banyak dikonsumsi di seluruh dunia berdasarkan tingkat kadar alkohol yang terkandung dalam setiap minuman. Data yang dipakai terhadap penelitian ini yaitu 192 data yang terdiri dari 30 data uji dan 172 data latih. Hasil akurasi yang diperoleh dari metode Naive Bayes adalah 97,44% dan 100% dari metode KNN.

Kata Kunci: Alkohol, KNN, Naïve bayes, Rapid Miner

1. PENDAHULUAN

Minuman beralkohol merupakan minuman yang diminati oleh semua kalangan di dunia. Alkohol merupakan zat yang terkandung dalam minuman keras yang banyak di konsumsi oleh kebanyakan negara beriklim dingin tetapi tidak menutup kemungkinan juga untuk negara-negara yang beriklim tropis. Dari semua jenis minuman yang terkandung zat alkohol terdapat jenis resmi di edarkan di dunia sesuai dengan tingkat kadar yang ditentukan. Minuman yang lazim di minum oleh seluruh dunia adalah jenis bir, spirit dan wine. Dalam hal tingkat kadar alkohol jenis minuman tersebut terdapat batasan tingkat kadar yang dapat dikonsumsi oleh setiap orang untuk meminimalisir dampak buruk dalam mengkonsumsi alkohol secara berlebihan. Dampak buruk dari berlebihan mengkonsumsi alkohol yang bisa mengakibatkan kondisi suatu negara menjadi tidak aman karena tingkat kriminalitas yang tinggi akibat pengaruh buruk dari ketergantungan mengkonsumsi minuman beralkohol. Maka dari itu penelitian ini untuk menghasilkan sebuah informasi tentang perbandingan kadar jenis minuman beralkohol yang tinggi dan memiliki tingkat kadar alkohol yang rendah disetiap negara. Dengan demikian, dari hasil penelitian ini peneliti memperoleh sebuah informasi perbandingan tingkat kadar alkohol untuk meminimalisir konsumsi minuman beralkohol secara berlebihan pada negara tersebut. maka dilakukanlah sebuah studi banding yang menggunakan 2 algoritma untuk mendapat sebuah informasi yang memiliki tingkat keakuratan yang maksimal. Penelitian ini bertujuan menentukan hasil perbandingan jenis minuman alkohol yang banyak dikonsumsi oleh setiap negara menggunakan metode naïve bayes dan knn yang pemodelannya di terapkan pada aplikasi rapid miner.

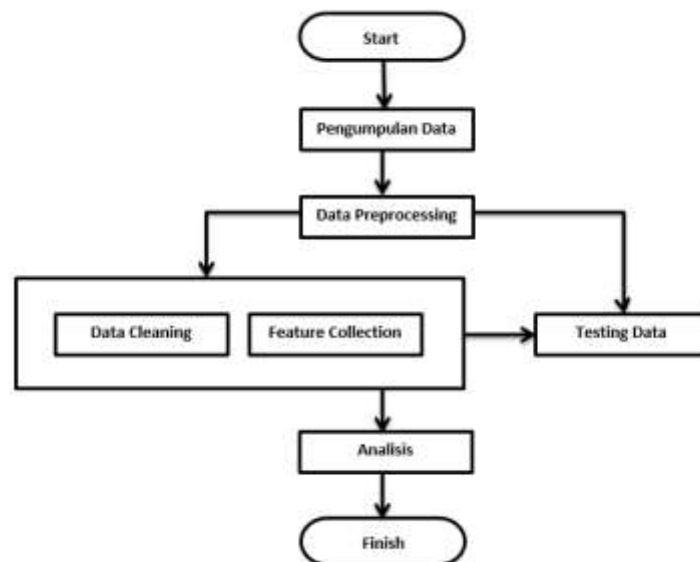
Berdasarkan penelitian sebelumnya juga telah dilakukan penelitian yaitu yang berjudul “Perbandingan kinerja metode naïve bayes dan K-NN untuk klasifikasi artikel berbahasa indonesia” pada [1] hasil penelitian tersebut, hasil dari akurasi metode Naïve Bayes lebih tinggi dari K-NN dengan menghasilkan sebuah akurasi klasifikasi yang tidak bisa mencapai hasil sempurna.

Perbandingan 2 metode dalam klasifikasi perbandingan telah dilakukan sebelumnya oleh yang berjudul “Analisa perbandingan metode naïve bayes classifier dan K-NN terhadap klasifikasi data” dimana [2] hasil dari K-NN lebih tinggi di dibandingkan dari metode Naïve bayes.

Dalam proses penelitian ini di butuhkan tahap *lifecycle* di antaranya tahap pengumpulan dataset, tahap preprocessing, training model dan pemodelan algoritma pada rapid miner.

2. METODE PENELITIAN

Dalam penelitian ini, maka terdapat sebuah proses yang mencakup seluruh aspek kegiatan dengan tujuan untuk mendapatkan hasil akurasi yang didapatkan pada penelitian yang dilakukan. Berikut proses metodologi penelitian ini dalam bentuk gambar di bawah ini :



Gambar 1. Metodologi penelitian

2.1. Pengumpulan data

Dataset yang dipakai pada penelitian ini di dapat dari <https://www.kaggle.com/datasets/satoshidatamoto/beer-consumption-and-life-expectancy-a-surprise> yang terdapat 192 record data.

2.2. Tahap preprocessing

Merupakan tahapan dalam membersihkan data, sehingga mendapatkan data yang bagus dan tidak rusak. Tahap preprocessing diartikan sebagai tahap awal untuk membersihkan kata-kata kalimat yang tidak digunakan [2]. Proses keseluruhan dilakukan menggunakan program python3, sehingga dilakukan secara otomatis. Terdapat beberapa tahapan yang digunakan pada tahap preprocessing ini yaitu :

- a. Data cleaning
Merupakan pembersihan data untuk menghilangkan data yang salah input, duplikat, dan tidak konsisten.
- b. Feature selection
Merupakan fitur untuk mengurangi jumlah data yang tidak relevan atau yang berlebihan.

2.3 Konsumsi

Merupakan pembelanjaan atas barang-barang dan jasa-jasa yang dilakukan oleh rumah tangga dengan tujuan untuk memenuhi kebutuhan dari orang yang melakukan pembelanjaan [3].

2.4 Metode Naïve bayes

Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan [4]. Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu [5].

2.5 Metode K-Nearest Neighbor

Algoritma KNN merupakan metode yang menggunakan algoritma supervised. Algoritma supervised learning tujuannya untuk mendapatkan pola baru sedangkan unsupervised learning tujuannya untuk mendapatkan pola dalam sebuah data [6]. Algoritma ini digunakan karena memiliki akurasi yang cukup tinggi seperti yang dihasilkan dari beberapa penelitian terdahulu [7].

2.6 Rapid Miner

Rapid Miner adalah aplikasi data mining berbasis open-source yang terkemuka dan ternama. Didalamnya terdapat aplikasi yang berdiri sendiri untuk analisis data dan sebagai mesin data mining seperti untuk loading data, transformasi data, pemodelan data, dan metode visualisasi data. Pada tahun 2007 akhirnya diganti namanya menjadi RapidMiner [4].

3. HASIL DAN PEMBAHASAN

3.1 Datasheet

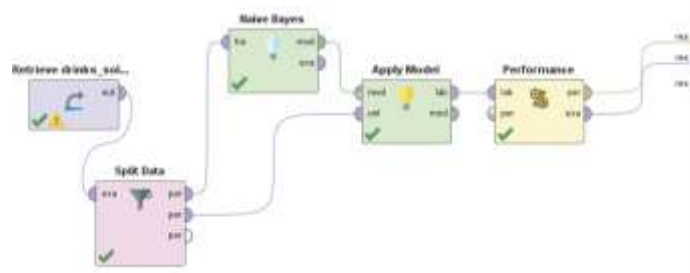
Data yang didapat memiliki 193 record dan memiliki 7 atribut. Datasheet dapat dilihat pada gambar dibawah ini :

	A	B	C	D	E	F	G
1	index	country	beer_servspirit_serwine_ser	total_litres_of_pure_alcohol			Kategori
2	0	Afghanistan	0	0	0		0 Medium Alcohol
3	1	Albania	89	132	54		4.9 High Alcohol
4	2	Algeria	25	0	14		0.7 Medium Alcohol
5	3	Andorra	245	138	312		12.3 High Alcohol
6	4	Angola	217	57	45		5.7 High Alcohol
7	5	Antigua & Barbuda	102	128	45		4.9 High Alcohol
8	6	Argentina	195	25	221		7.8 High Alcohol
9	7	Armenia	21	179	11		3.7 Medium Alcohol
10	8	Australia	261	72	212		9.7 High Alcohol
11	9	Austria	279	75	191		9.7 High Alcohol
12	10	Azerbaijan	21	46	5		1.3 Medium Alcohol
13	11	Bahamas	122	176	49		6.2 High Alcohol
14	12	Bahrain	42	68	7		2 Medium Alcohol
15	13	Bangladesh	0	0	0		0 Medium Alcohol
16	14	Barbados	143	173	36		6.2 High Alcohol
17	15	Belarus	142	373	42		9.9 High Alcohol
18	16	Belgium	295	84	212		10.5 High Alcohol
19	17	Belize	263	114	8		6.8 High Alcohol

Gambar 2. Datasheet

3.2 Pemodelan naïve bayes

- a. Implementasi algoritma naïve bayes pada rapid miner



Gambar 3. Pemodelan naïve bayes di rapid miner

b. Hasil akurasi

accuracy: 97.44%

	true Medium Alcohol	true High Alcohol	class precision
pred. Medium Alcohol	20	0	100.00%
pred. High Alcohol	1	18	94.74%
class recall	95.24%	100.00%	

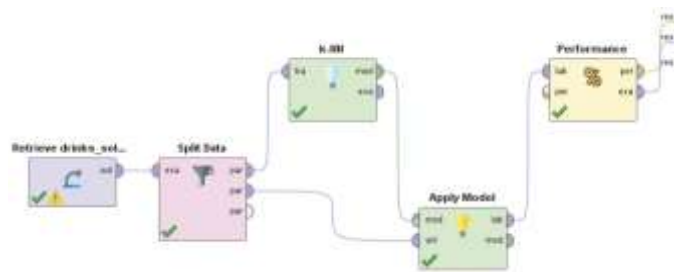
Gambar 4. Hasil akurasi menggunakan metode naïve bayes

Dari hasil proses yang dilakukan, hasil akurasi yang didapatkan dalam menggunakan metode naïve bayes adalah sebesar 97,44%.

3.3 Pemodelan K-Nearest neighbor

a. Implementasi algoritma K-Nearest neighbor pada rapid miner

b.



Gambar 5. Pemodelan K-Nearest neighbor di rapid miner

c. Hasil akurasi

accuracy: 100.00%

	true Medium Alcohol	true High Alcohol	class precision
pred. Medium Alcohol	21	0	100.00%
pred. High Alcohol	0	18	100.00%
class recall	100.00%	100.00%	

Gambar 6. Hasil akurasi menggunakan metode K-Nearest neighbor

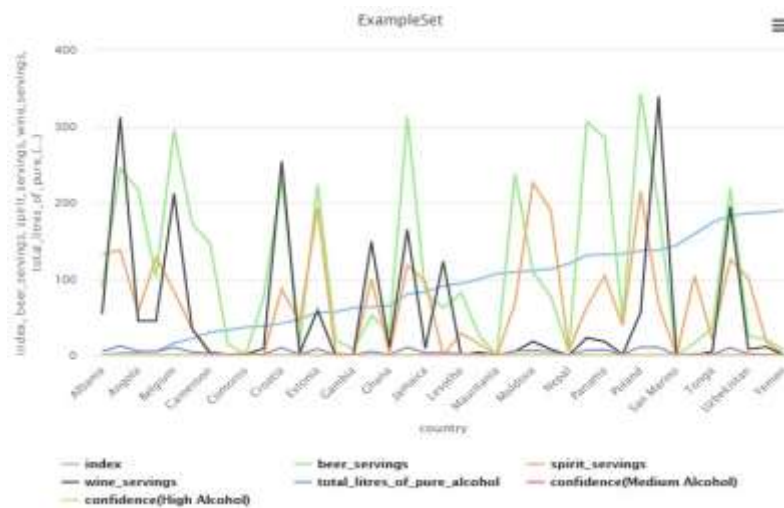
Dari hasil proses yang dilakukan, hasil akurasi yang didapatkan dalam menggunakan metode K-Nearest neighbor adalah sebesar 100%.

3.4 Hasil splitting data

Row No.	country	Kategori	prediction(K...	confidence(...	confidence(...	index	beer_servin...	spirit_servin...	wine
1	Albania	High Alcohol	High Alcohol	0	1	1	89	132	54
2	Andorra	High Alcohol	High Alcohol	0	1	3	245	138	312
3	Angola	High Alcohol	High Alcohol	0.405	0.595	4	217	57	45
4	Artigua & Bar...	High Alcohol	High Alcohol	0	1	5	102	128	45
5	Belgium	High Alcohol	High Alcohol	0	1	16	295	84	212
6	Botswana	High Alcohol	High Alcohol	0.430	0.570	22	173	35	35
7	Cameroon	Medium Alco...	Medium Alco...	0.622	0.378	30	147	1	4
8	Chad	Medium Alco...	Medium Alco...	1	0	33	15	1	1
9	Comoros	Medium Alco...	Medium Alco...	1	0	37	1	3	1
10	Congo	Medium Alco...	Medium Alco...	1	0	38	76	1	9
11	Croatia	High Alcohol	High Alcohol	0	1	42	230	87	254
12	Djibouti	Medium Alco...	Medium Alco...	1.000	0	47	15	44	3
13	Estonia	High Alcohol	High Alcohol	0	1	56	224	194	59

Gambar 7. Hasil splitting data

3.5 Hasil data dalam bentuk grafik



Gambar 8. Hasil data dalam bentuk grafik

4. KESIMPULAN

Dari hasil penelitian yang dilakukan, maka disimpulkan bahwa dari data penelitian adalah 193 record yang terdiri dari 39 data uji dan 154 data latih, diperoleh:

1. Hasil akurasi yang didapat dari metode naïve bayes ini adalah 97,44%.
2. Hasil akurasi yang didapat dari metode k-nearest neighbor adalah 100%.
3. Hasil perbandingan kedua algoritma tersebut menunjukkan bahwa akurasi algoritma k-nearest neighbor lebih tinggi dari naïve bayes.
4. Pada algoritma naïve bayes membuktikan bahwa negara Poland paling banyak memiliki jenis bir dengan kadar tertinggi sebesar 343.
5. Pada algoritma k-nearest neighbor membuktikan bahwa negara Poland paling banyak memiliki jenis bir dengan kadar tertinggi sebesar 343.

REFERENSI

- [1] R. N. Devita, H. W. Herwanto, and A. P. Wibawa, "Perbandingan kinerja metode naïve bayes dan k-nearest neighbor untuk klasifikasi artikel berbahasa indonesia," *JTIK (Jurnal Teknol. Inf. dan Ilmu Komputer, Vol, 5*, 2018.
- [2] R. Rahmadden, M. K. Anam, Y. Irawan, S. Susanti, and M. Jamaris, "Comparison of Support Vector Machine and XGBSVM in Analyzing Public Opinion on Covid-19 Vaccination," *Ilk. J. Ilm.*, vol. 14, no. 1, 2022.

- [3] E. Elvina, “Pengaruh Pendapatan Dan Konsumsi Terhadap Perilaku Konsumen Di Kabupaten Labuhanbatu,” *ECOBISMA (Jurnal Ekon. Bisnis dan Manajemen)*, vol. 5, no. 2, pp. 120–129, 2018.
- [4] A. P. Natasuwarna, “Seminar Pendekatan Data Mining Memprediksi Profil Sosial Masyarakat Menggunakan Aplikasi RapidMiner,” in *SNPMas: Seminar Nasional Pengabdian pada Masyarakat*, 2019, pp. 38–44.
- [5] R. Nofitri and N. Irawati, “Integrasi Metode Neive Bayes Dan Software Rapidminer Dalam Analisis Hasil Usaha Perusahaan Dagang,” *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 6, no. 1, pp. 35–42, 2019.
- [6] M. Nanja and P. Purwanto, “Metode k-nearest neighbor berbasis forward selection untuk prediksi harga komoditi lada,” *Pseudocode*, vol. 2, no. 1, pp. 53–64, 2015.
- [7] D. Efriadi, R. Rahmaddeni, A. Agustin, and J. Junadhi, “Prediksi Penambahan Piutang Iuran Jaminan Sosial Ketenagakerjaan menggunakan Algoritma K-Nearest Neighbor,” *Edumatic J. Pendidik. Inform.*, vol. 6, no. 1, pp. 49–57, 2022.