

PERBANDINGAN METODE K-NEAREST NEIGHBOUR DAN NAIVE BAYES CLASSIFICATION UNTUK MEMBENTUK SISTEM REKOMENDASI BERITA OLAHRAGA ONLINE BERBASIS WEB APPLICATION

Dela Rosari Maria Seran¹, Yudi Setyawan², Rokhana Dwi Bekti³

^{1,2,3}Jurusan Statistika, Fakultas Sains Terapan, Institut Sains & Teknologi AKPRIND Yogyakarta
e-mail : ¹delaseran.stat@gmail.com, ²setyawan@akprind.ac.id, ³rokhana@akprind.ac.id

ABSTRACT

Online sports news is news that contains sports content and spread through online media so that it can be accessed easily. With the wide selection of online sports news, a certain method that can establish a reliable recommendation system is needed. Classification is one of the methods to establish text data recommendation system. In this study, primary data in the form of 2,500 sports news titles obtained through web scraping process from several Indonesian sports news sites consisting of 9 categories, namely "bola indonesia", "bola inggris", "bola italia", "bola spanyol", "basket", "badminton", "f1", "motogp", and "esport". Classification is done by K-Nearest Neighbour and Naive Bayes Classification methods. Both of the classification methods use 2,252 training data and 248 testing data. The results show that Naive Bayes Classification is better than K-Nearest Neighbour. Naive Bayes Classification produces an accuracy of 95,161%, while K-Nearest Neighbour only produces an accuracy of 84,677% with the optimum number of nearest neighbours is $K = 10$. To be accessible, the recommendation system is established based on web application using Shiny RStudio. The recommendation system can be accessed via <https://delaseran.shinyapps.io/Sistem-Rekomendasi-Berita-Olahraga-Online/>. This recommendation system is established like a search engine, so users can type in keywords of any sports news topics that they want to read in the search column, then the system will automatically bring up the latest news titles and links from the recommendation results.

Keywords : classification, K-Nearest neighbour, Naive Bayes Classification, sports news recommendation system, web application

INTISARI

Berita olahraga online merupakan berita yang mengandung konten olahraga dan disebarakan melalui media online sehingga dapat diakses dengan mudah. Dengan adanya beragam pilihan berita olahraga online yang tersedia, maka dibutuhkan suatu metode yang dapat membentuk sistem rekomendasi yang andal. Klasifikasi merupakan salah satu metode untuk membentuk sistem rekomendasi data teks. Dalam penelitian ini, data primer berupa 2500 judul berita olahraga diperoleh melalui proses web scraping dari beberapa website berita olahraga Indonesia yang terdiri dari 9 kategori yaitu "bola indonesia", "bola inggris", "bola italia", "bola spanyol", "basket", "badminton", "f1", "motogp", dan "esport". Klasifikasi dilakukan dengan metode K-Nearest Neighbour dan Naive Bayes Classification. Kedua metode klasifikasi tersebut menggunakan 2252 data latih dan 248 data uji. Diperoleh hasil bahwa metode Naive Bayes Classification lebih baik daripada K-Nearest Neighbour. Metode Naive Bayes Classification menghasilkan akurasi sebesar 95,161%, sedangkan metode K-Nearest Neighbour hanya menghasilkan akurasi sebesar 84,677% dengan jumlah tetangga terdekat yang optimal adalah $K = 10$. Agar dapat mudah diakses, sistem rekomendasi dibentuk berbasis web application menggunakan Shiny RStudio. Sistem rekomendasi tersebut dapat diakses melalui <https://delaseran.shinyapps.io/Sistem-Rekomendasi-Berita-Olahraga-Online/>. Sistem rekomendasi ini berbentuk seperti mesin pencarian, sehingga pengguna dapat mengetikkan kata kunci dari topik berita olahraga yang ingin dibaca pada kolom pencarian, lalu sistem akan secara otomatis memunculkan judul berita terbaru sekaligus link dari hasil rekomendasi.

Kata kunci : klasifikasi, K-Nearest neighbour, Naive Bayes Classification, sistem rekomendasi berita olahraga, Web Application

1. PENDAHULUAN

Menurut Sumadiri dalam Rani (2013) berita adalah laporan mengenai fakta atau ide yang benar, menarik, dan/atau penting bagi sebagian besar khalayak yang disebarakan melalui media berkala seperti surat kabar, radio, televisi, atau media *online*. Cara paling praktis untuk menikmati berita adalah melalui media *online*.

Di Indonesia diperkirakan terdapat 43.300 media berita *online*/siber pada tahun 2016 (Dewan Pers, 2016). Dari data tersebut terlihat jelas bahwa terdapat begitu banyak media *online*, ini tentu menunjukkan antusiasme masyarakat yang besar terhadap media-media *online* tersebut.

Salah satu sajian berita yang sangat digemari oleh para pembaca adalah berita olahraga. Konten mengenai olahraga merupakan “menu” bagi setiap media karena mampu menarik pembaca. Ini dapat terlihat dari hampir semua media di Indonesia baik cetak, elektronik maupun *online* yang memiliki rubrik olahraga (Prarsty, 2014). Di Indonesia terdapat banyak website berita olahraga diantaranya adalah bola.com, bolasport.com, sport.detik.com, dan lain-lain. Dengan adanya beragam pilihan berita *online* yang tersedia maka dibutuhkan suatu metode tertentu yang dapat membentuk sistem rekomendasi yang andal. Berdasarkan Scafer, dkk dalam Fadlil dan Mahmudy (2007), sistem rekomendasi adalah suatu program yang melakukan prediksi terhadap sesuatu item, seperti film, musik, buku, berita, produk, dan lain sebagainya yang menarik *user* atau pengguna. Sistem rekomendasi membantu pengguna untuk mengidentifikasi item yang sesuai dengan kebutuhan, kesenangan, dan keinginan mereka. Sistem rekomendasi berita akan membimbing para pembaca untuk menemukan berita yang relevan serta sesuai dengan keinginannya.

Klasifikasi atau kategorisasi teks adalah proses penempatan suatu teks ke suatu kategori atau *class* sesuai dengan karakteristik dari teks tersebut. Sebuah teks dapat dikelompokkan ke dalam kategori tertentu berdasarkan kata-kata yang terdapat dalam teks tersebut (Syarifah dan Muslim, 2015). Ada banyak metode klasifikasi diantaranya adalah Decision Trees, Naive Bayes Classification, Neural Networks, K-Nearest Neighbour, dan lain sebagainya. Pada penelitian ini metode Naive Bayes Classification dipilih karena mampu bekerja dengan baik dalam data latih yang berukuran kecil maupun berukuran besar (Lantz, 2015). Selain itu, metode K-Nearest Neighbour juga dipilih karena tidak mengharuskan data untuk mengikuti suatu distribusi tertentu (Lantz, 2015). Naive Bayes Classification dan K-Nearest Neighbour merupakan metode yang sering digunakan dalam klasifikasi teks dan memberikan hasil yang cukup baik (Azhar *et al.*, 2019). Dengan klasifikasi berita olahraga menggunakan metode K-Nearest Neighbour dan Naive Bayes Classification dapat dibuat suatu sistem rekomendasi berita olahraga *online* berbasis *web application* yaitu suatu aplikasi yang diakses menggunakan penjelajah web melalui jaringan internet.

Pada beberapa penelitian yang telah dilakukan sebelumnya, data yang digunakan tidak terlalu banyak, jumlah masih di bawah 500 data, serta jumlah *class* atau kategori dalam penelitian yang tidak terlalu banyak yaitu kurang dari 5 *class*. Penelitian sebelumnya yang menggunakan Shiny Rstudio tidak mempublikasikan *web application* yang telah dibuat sehingga tidak mudah untuk diakses oleh para pengguna. Oleh karena itu, pada penelitian ini digunakan data yang lebih banyak yaitu sejumlah 2500 data teks serta jumlah *class* atau kategori yang lebih banyak yaitu sejumlah 9 *class* sehingga dapat diketahui kinerja metode Naive Bayes Classification dan K-Nearest Neighbour dalam mengklasifikasikan data teks dengan jumlah data serta *class* yang lebih banyak. Selain itu dalam penelitian ini *web application* yang telah dibuat akan dipublikasikan sehingga dapat diakses oleh para pengguna.

Berdasarkan latar belakang permasalahan di atas maka dapat dibuat rumusan masalah yaitu: Bagaimana gambaran umum berita olahraga *online* yang digunakan dalam pembentukan model klasifikasi, bagaimana hasil klasifikasi berita olahraga *online* menggunakan metode K-Nearest Neighbour, bagaimana hasil klasifikasi berita

olahraga *online* menggunakan metode Naive Bayes Classification, bagaimana hasil perbandingan metode K-Nearest Neighbour dan Naive Bayes Classification dalam mengklasifikasi berita olahraga *online*, bagaimana hasil pembentukan sistem rekomendasi berita olahraga *online* berbasis *web application* menggunakan Shiny RStudio, dan bagaimana hasil evaluasi sistem rekomendasi berita olahraga *online* berbasis *web application*.

2. METODE PENELITIAN

2.1 Metode Pengumpulan Data

Dalam penelitian ini digunakan data primer yang diperoleh dengan melakukan *web scraping*. Data dikumpulkan berdasarkan metode *cluster sampling* bertahap. Berita olahraga *online* diterbitkan setiap harinya dan jika dihitung maka jumlahnya akan sangat banyak. Untuk membatasi jumlah data penelitian maka pada tahap pertama *cluster sampling*, dilakukan pemilihan tanggal untuk proses *web scraping* dan tanggal yang dipilih adalah 14 dan 23 Desember 2019. Ada begitu banyak kategori berita olahraga, maka pada tahap kedua *cluster sampling*, dilakukan pemilihan kategori atau *class* berita olahraga dan kategori yang dipilih adalah bola indonesia, bola inggris, bola italia, bola spanyol, basket, badminton, f1, motogp, serta esport. Kesembilan kategori tersebut dipilih karena merupakan kategori berita olahraga yang paling sering dibahas dalam berbagai *website* berita olahraga.

2.2 Variabel Penelitian

Penelitian yang menggunakan metode klasifikasi ini memiliki dua jenis variabel yaitu variabel input dan variabel target. Pada penelitian ini yang menjadi variabel input adalah judul berita serta variabel target adalah *class* atau kategori berita pada setiap *website* berita olahraga.

2.3 Tahapan Penelitian

Adapun tahapan-tahapan dalam penelitian ini adalah sebagai berikut.

1. Mengumpulkan data dengan metode *cluster sampling* tahap pertama dan kedua. Pengumpulan data dilakukan melalui proses *web scraping* judul berita dan URL berita dari *website* berita olahraga.
2. Membentuk model klasifikasi dengan metode K-Nearest Neighbour dan Naive Bayes Classification. Tahapan ini terdiri atas beberapa bagian, yaitu sebagai berikut.
 - a. Melihat gambaran umum data berita olahraga menggunakan *donut chart*.
 - b. Melakukan *text preprocessing* yang terdiri dari beberapa bagian antara lain *case folding*, *number removal*, *stop character removal*, *stop word removal*, *tokenizing*, dan *rare terms removal*.
 - c. Melakukan *term weighting* dengan metode *binary term frequency* untuk klasifikasi berita olahraga dengan Naive Bayes Classification, serta metode *raw term frequency* untuk klasifikasi berita olahraga dengan K-Nearest Neighbour.
 - d. Melakukan partisi data dengan tujuan untuk membagi data teks menjadi dua bagian yaitu data latih dan data uji. Proses partisi data dilakukan dengan metode *cluster sampling* atau lebih tepatnya metode *cluster sampling* tahap ketiga. Pada *cluster sampling* tahap ketiga akan dipilih secara acak sebanyak 90% judul berita dari setiap *class* berita untuk dijadikan data latih. Judul berita yang tidak terpilih menjadi data latih akan menjadi data uji.
 - e. Mengklasifikasi data dengan metode K-Nearest Neighbour.
 - f. Mengklasifikasi data dengan metode Naive Bayes Classification.

3. Membandingkan tingkat akurasi metode K-Nearest Neighbour dan Naive Bayes Classification. Metode dengan akurasi tertinggi yang akan digunakan untuk membentuk sistem rekomendasi berbasis *web application*.
4. Membentuk sistem rekomendasi berbasis *web application* menggunakan Shiny RStudio.
5. Mengevaluasi sistem rekomendasi.

2.4 K-Nearest Neighbour

Algoritma *K-Nearest Neighbour* adalah algoritma yang digunakan untuk melakukan klasifikasi terhadap suatu objek berdasarkan K buah data latih yang jaraknya paling dekat dengan objek (data uji) tersebut. Syarat nilai K adalah tidak boleh lebih besar dari jumlah data latih (Rivki dan Bachtiar, 2017). *K-Nearest Neighbour* menggunakan ukuran kemiripan untuk membandingkan data uji yang diberikan dengan data latih. Dalam mengklasifikasikan data teks, yang diukur kemiripannya adalah bobot *term* data uji dan bobot *term* data latih. Salah satu ukuran untuk melihat kemiripan antara data uji dan data latih adalah jarak Euclid. Secara matematis jarak Euclid dinyatakan sebagai berikut.

$$d_{p_{iy}, q_{jy}} = \sqrt{\sum_{y=1}^z (p_{iy} - q_{jy})^2} \quad (1)$$

Keterangan:

- $d_{p_{iy}, q_{jy}}$: Jarak Euclid antara data *testing* dan data *training*
- p_i : data uji
- q_j : data latih
- i : indeks data uji
- j : indeks data latih
- z : banyaknya *term*

2.5 Naive Bayes Classification

Naive Bayes Classification merupakan metode pengklasifikasian probabilistik sederhana yang didasarkan pada teorema Bayes dengan asumsi bahwa semua fitur tidak saling terkait satu sama lain atau saling independen (Yuan, 2010). Dalam kaitannya dengan data teks, *Naive Bayes* mengasumsikan setiap kata (*term*) pada setiap kategori (*class*) tidak memiliki ketergantungan satu sama lain (*independent*). Dalam mengklasifikasikan data teks, jika terdapat sejumlah L *class* (C) dan sejumlah n fitur (F), dalam hal ini fitur merupakan kata atau *term* maka perhitungan probabilitas dalam *Naive Bayes Classification* dapat ditulis sebagai berikut (Lantz, 2015).

$$P(C_L | F_1 \cap F_2 \cap \dots \cap F_n) = P(C_L) \prod_{i=1}^n P(F_i | C_L) \quad (2)$$

Dengan demikian hasil klasifikasi merupakan *class* yang menghasilkan nilai probabilitas maksimum atau dapat dinyatakan dalam persamaan berikut.

$$C_{MAP} = \operatorname{argmax}_{C_L \in C} (P(C_L) \prod_{i=1}^n P(F_i | C_L)) \quad (3)$$

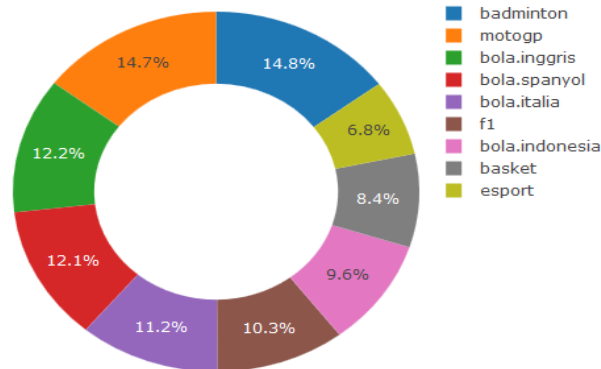
Keterangan:

- $P(C_L | F_1 \cap F_2 \cap \dots \cap F_n)$: *Posterior Probability*
- $P(F_i | C_L)$: *Likelihood*
- $P(C_L)$: *Prior Probability*
- C_{MAP} : *Class dengan Maximum A Posterior Probability*

3. HASIL DAN PEMBAHASAN

3.1 Gambaran Umum Berita Olahraga Online

Dalam penelitian ini digunakan *donut chart* yaitu sebuah grafik yang memiliki kegunaan yang sama dengan *pie chart* yaitu untuk memberikan gambaran umum tentang suatu data kategorik berdasarkan persentasenya. Dalam hal ini, *donut chart* digunakan untuk memberikan gambaran *class* berita olahraga berdasarkan besar persentasenya. Berikut ini ditampilkan hasil eksplorasi *class* berita olahraga tersebut dalam bentuk *donut chart* seperti pada Gambar 1.

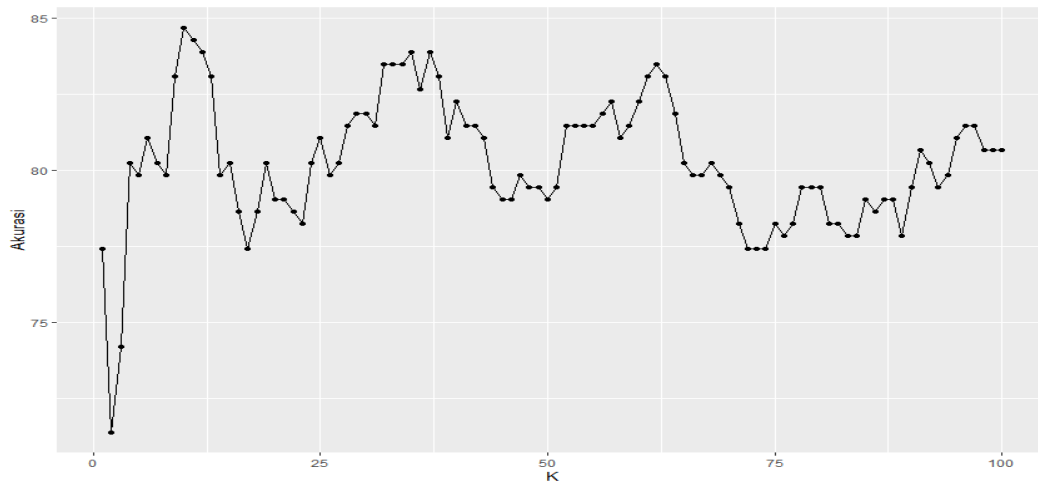


Gambar 1. Donut Chart Class Berita Olahraga

Gambar 1 menunjukkan bahwa berita dari *class* badminton memiliki presentase yang paling banyak yaitu 14,8% atau sebanyak 370 judul berita sedangkan berita dari *class* esport memiliki presentase yang paling sedikit yaitu 6,8% atau sebanyak 170 judul berita.

3.2 Klasifikasi dengan Metode K-Nearest Neighbour

K-Nearest Neighbour adalah algoritma yang digunakan untuk melakukan klasifikasi terhadap suatu objek (data uji), berdasarkan K buah data latih yang jaraknya paling dekat dengan objek (data uji) tersebut. Dalam penelitian ini digunakan jumlah tetangga terdekat K = 1 sampai dengan K = 100. Setelah melakukan klasifikasi serta menghitung tingkat akurasi untuk jumlah tetangga terdekat K = 1 sampai dengan K = 100, selanjutnya akan dibandingkan tingkat akurasi tersebut sehingga dapat diketahui berapa jumlah K tetangga terdekat yang optimal. Hasil perhitungan akurasi klasifikasi ditampilkan pada Gambar 2.



Gambar 2. Akurasi Klasifikasi dengan K-Nearest Neighbour

Dari Gambar 2 terlihat bahwa akurasi klasifikasi mengalami fluktuasi. Tingkat akurasi terendah terjadi ketika jumlah tetangga terdekat sebanyak $K = 2$. Sedangkan tingkat keakuratan tertinggi terjadi ketika jumlah tetangga terdekat sebanyak $K = 10$. Dengan demikian jumlah tetangga terdekat optimal adalah $K = 10$. Tingkat ketepatan klasifikasi berita olahraga menggunakan metode K-Nearest Neighbour dengan jumlah $K = 10$ dapat dilihat melalui confusion matrix berikut ini.

Tabel 1. Confusion Matrix K-Nearest Neighbour K = 10

Predictions	Actual								
	badminton	basket	bola.indonesia	bola.inggris	bola.italia	bola.spanyol	esport	fl	motogp
badminton	33	0	1	0	0	0	0	1	0
basket	1	18	3	0	0	0	1	2	0
bola.indonesia	1	3	16	4	0	2	0	0	0
bola.inggris	0	0	1	25	3	0	1	0	1
bola.italia	1	0	2	0	25	1	1	1	0
bola.spanyol	0	0	0	1	0	27	0	0	0
esport	0	0	0	0	0	0	14	1	0
fl	0	0	0	0	0	0	0	18	1
motogp	1	0	1	0	0	0	0	2	34

Dari Tabel 1 dapat dilihat bahwa *entry-entry* diagonal merupakan hasil klasifikasi yang tepat, sehingga tingkat akurasi klasifikasi dihitung dengan menjumlahkan seluruh *entry* diagonal dibagi dengan total seluruh data uji dikali dengan 100%. Sehingga tingkat akurasi dapat dihitung sebagai berikut.

$$akurasi = \frac{(33 + 18 + 16 + 25 + 25 + 27 + 14 + 18 + 34)}{248} 100\%$$

$$akurasi = 84,677 \%$$

3.3 Naive Bayes Classification

Naive Bayes Classification merupakan metode pengklasifikasian probabilistik sederhana yang didasarkan pada teorema Bayes. Tingkat ketepatan klasifikasi dapat dilihat melalui confusion matrix di Tabel 2.

Tabel 2. Confusion Matrix Naive Bayes Classification

Prediction	Actual								
	badminton	Basket	bola.indonesia	bola.inggris	bola.italia	bola.spanyol	esport	fl	motogp
badminton	36	0	0	0	0	0	0	0	0
basket	0	21	0	0	0	0	0	0	0
bola.indonesia	0	0	22	0	1	0	1	0	0
bola.inggris	0	0	1	28	1	0	0	0	0
bola.italia	0	0	1	1	26	1	1	0	0
bola.spanyol	0	0	0	1	0	29	0	0	0
esport	1	0	0	0	0	0	15	0	0
fl	0	0	0	0	0	0	0	23	0
motogp	0	0	0	0	0	0	0	2	36

Dari Tabel 2 dapat dilihat bahwa *entry-entry* diagonal merupakan hasil klasifikasi yang tepat, sehingga tingkat akurasi klasifikasi dihitung dengan menjumlahkan seluruh *entry* diagonal dibagi dengan total seluruh data uji dikali dengan 100%, sehingga tingkat akurasi dapat dihitung sebagai berikut.

$$akurasi = \frac{(36 + 21 + 22 + 28 + 26 + 29 + 15 + 23 + 36)}{248} 100\%$$

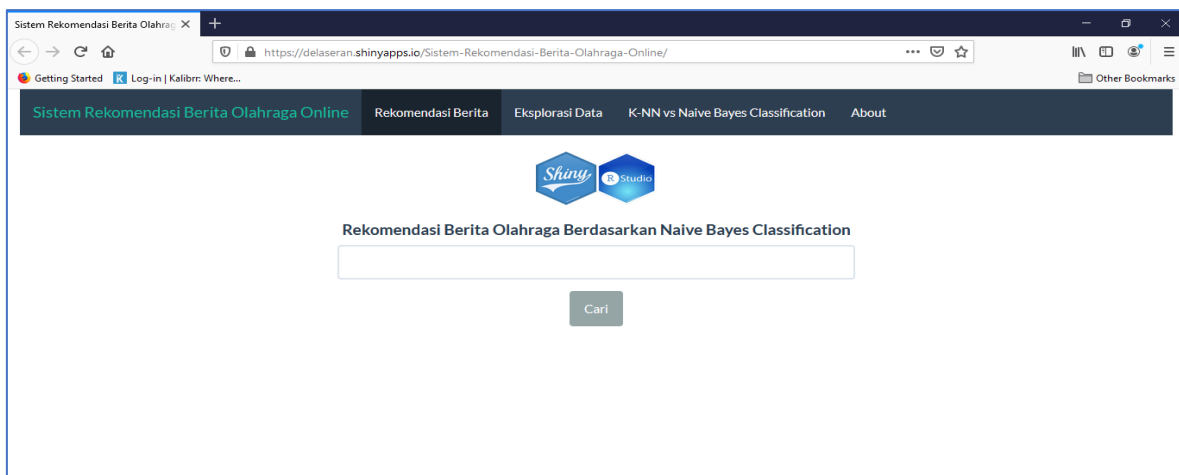
$$akurasi = 95,161 \%$$

3.4 Perbandingan Metode Klasifikasi

Perbandingan metode dilakukan untuk mengetahui metode terbaik diantara K-Nearest Neighbour dan Naive Bayes Classification yaitu yang memberikan tingkat akurasi paling tinggi. Berdasarkan hasil klasifikasi dengan kedua metode diperoleh hasil bahwa Naive Bayes Classification memiliki tingkat akurasi sebesar 95,161%. Hasil tersebut lebih tinggi daripada tingkat akurasi yang diperoleh dari klasifikasi menggunakan K-Nearest Neighbour dengan jumlah tetangga terdekat $K = 10$ yaitu sebesar 84,677%. Ini menunjukkan bahwa metode Naive Bayes Classification lebih baik untuk mengklasifikasikan judul berita olahraga dibandingkan dengan metode K-Nearest Neighbour. Dengan demikian metode yang akan digunakan untuk membentuk sistem rekomendasi adalah metode Naive Bayes Classification.

3.5 Hasil Pembentukan Web Application

Sistem rekomendasi berita olahraga *online* berbasis *web application* pada penelitian ini dibentuk menggunakan Shiny RStudio. *Web application* berupa sistem rekomendasi berita olahraga tersebut dapat diakses melalui <https://delaseran.shinyapps.io/Sistem-Rekomendasi-Berita-Olahraga-Online/>. *Web application* ini memiliki satu tab menu utama yaitu “Rekomendasi Berita” dan tiga tab menu tambahan yaitu “Eksplorasi Data”, “K-NN vs Naive Bayes Classification” dan “About”. Menu utama yaitu “Rekomendasi Berita” berfungsi sebagai mesin pencarian berita olahraga *online*. Pengguna dapat mengetikkan kata kunci berupa topik berita olahraga yang ingin dibaca pada kolom pencarian. Rekomendasi berita didasarkan pada hasil klasifikasi yang dilakukan pada input atau kata kunci yang diketikkan pada kolom pencarian. Kata kunci yang diketikkan tersebut akan secara otomatis mengalami *text preprocessing*, *term weighting*, serta klasifikasi dengan metode Naive Bayes Classification. Berikut adalah tampilan awal *web application*.



Gambar 3. Tampilan Awal *Web Application*

Data latih dan hasil rekomendasi berita olahraga pada *web application* ini merupakan berita terbaru yang diperoleh dari hasil *web scraping* otomatis. Proses *web scraping* dilakukan saat pencarian pertama atau proses eksplorasi data pertama dilakukan setiap kali *web application* ini diakses oleh pengguna. *Web scraping* dilakukan pada tiga website berita olahraga yaitu www.bola.com, www.indosport.com, dan www.ligaolahraga.com. Ketiga website ini dipilih karena memiliki konten berita olahraga yang lengkap dan beragam serta untuk menghemat waktu karena semakin banyak website maka semakin banyak pula waktu yang digunakan untuk melakukan *web scraping*. Pada *web application* ini, proses *web scraping* memakan waktu sekitar 20 – 25 detik dan proses klasifikasi memakan waktu sekitar 3 – 5 detik. Tampilan hasil rekomendasi berita dengan contoh kata kunci “marc marquez” pada Gambar 4.

link	class	n
22 alex marquez bocorkan kondisi terkini marc marquez semakin membaik tapi	motogp	3
1 motogp marc marquez target saya adalah balapan di qatar tetapi rabu 24 februari 2021 2109 wib	motogp	2
3 motogp update kondisi marc marquez pasca operasi ketiga membaik atau makin buruk jumat 12 februari 2021 1857 wib	motogp	2
5 pol espargaro yakin marc marquez tetap berbahaya meski baru alami cedera	motogp	2
6 marc marquez dipastikan absen dalam tes pramusim motogp 2021	motogp	2
7 puig yakin cedera lengan tak akan ubah jatidiri marc marquez	motogp	2

Gambar 4. Tampilan Hasil Rekomendasi pada Menu “Rekomendasi Berita”

Rekomendasi ditampilkan dalam bentuk tabel yang terdiri dari kolom “link” berita, kolom “class” berita, serta kolom “n” yang berfungsi untuk menunjukkan jumlah kata yang sama antara input pada kolom pencarian dengan judul sekaligus link berita olahraga hasil rekomendasi pada kolom “link”.

Tab menu tambahan “Eksplorasi Data” berfungsi untuk menampilkan data latih hasil *web scraping* otomatis dengan visualisasi menggunakan *donut chart* dan *word cloud*. Tab menu “K-NN vs Naive Bayes Classification” berfungsi untuk menampilkan perbandingan hasil klasifikasi metode *Naive Bayes Classification* dan *K-Nearest Neighbour* menggunakan *confusion matrix*. Lalu yang terakhir tab menu “About” berfungsi untuk memberikan penjelasan singkat tentang *web application* ini dan kontak *developer*.

3.6 Evaluasi Web Application

Sistem rekomendasi berita olahraga berbasis *web application* yang dapat diakses melalui <https://delaseran.shinyapps.io/Sistem-Rekomendasi-Berita-Olahraga-Online/> dapat dievaluasi dengan melihat beberapa kelebihan serta kekurangannya. Berikut ini adalah beberapa kelebihannya menurut peneliti.

1. Mudah diakses oleh pengguna karena hanya memerlukan koneksi internet tanpa harus melalui proses instalasi tertentu.

2. Mampu memberikan rekomendasi berita olahraga terbaru kepada pengguna.
3. Efektif untuk mencari topik-topik berita olahraga yang berkaitan erat dengan data latih.
4. Didasarkan pada metode Naive Bayes Classification yang telah terbukti mampu menghasilkan tingkat akurasi yang tinggi.
5. Menyediakan menu tambahan yaitu tab menu Eksplorasi Data yang dapat memberikan informasi secara garis besar kepada pengguna tentang data latih yang digunakan dalam sistem rekomendasi serta tab menu K-NN vs Naive Bayes Classification yang dapat menampilkan perbandingan hasil klasifikasi metode Naive Bayes Classification dan K-Nearest Neighbour menggunakan confusion matrix dan tingkat akurasi.

Setelah membahas beberapa kelebihan dari sistem rekomendasi yang telah dibuat, berikut ini adalah beberapa kekurangannya menurut peneliti.

1. Agar *web application* ini dapat diakses oleh para pengguna maka harus dipublikasikan melalui akun shinyapps.io. Dalam penelitian ini akun shinyapps.io yang digunakan merupakan versi free atau tidak berbayar sehingga terdapat keterbatasan di dalamnya.
2. Sistem rekomendasi belum menyediakan fitur untuk memilih tanggal publikasi berita olahraga.
3. Sistem rekomendasi belum menyediakan fitur *autocorrect* yang akan secara otomatis mengenali dan memperbaiki kesalahan pengetikkan pada kolom pencarian.
4. *Class* atau kategori berita olahraga yang terbatas pada sembilan kategori yaitu bola indonesia, bola inggris, bola italia, bola spanyol, basket, badminton, fl, motogp, serta esport.

4. KESIMPULAN

Gambaran umum dari data yang digunakan untuk melakukan pemodelan klasifikasi memiliki karakteristik yaitu *class* badminton memiliki presentase yang paling banyak yaitu 14,8% atau sedangkan *class* esport memiliki presentase yang paling sedikit yaitu 6,8%. Klasifikasi berita olahraga *online* menggunakan metode K-Nearest Neighbour menghasilkan akurasi sebesar 84,677% dengan jumlah tetangga terdekat optimal adalah $K = 10$. Sementara itu, klasifikasi berita olahraga *online* menggunakan metode Naive Bayes Classification menghasilkan akurasi sebesar 95,161%. Pada perbandingan metode K-Nearest Neighbour dan Naive Bayes Classification dalam mengklasifikasi berita olahraga *online* diperoleh hasil bahwa metode Naive Bayes Classification mampu memberikan tingkat akurasi yang lebih tinggi.

Hasil pembentukan sistem rekomendasi berita olahraga *online* berbasis *web application* menggunakan Shiny RStudio dapat diakses melalui <https://delaseran.shinyapps.io/Sistem-Rekomendasi-Berita-Olahraga-Online/>. Sistem rekomendasi ini berbentuk seperti mesin pencarian (*search engine*) sehingga pengguna dapat mengetikkan kata kunci berupa topik berita olahraga *online* yang ingin dibaca pada kolom pencarian, lalu sistem akan secara otomatis memunculkan judul-judul berita terbaru hasil rekomendasi sekaligus link dari berita-berita tersebut. Bagi peneliti, sistem rekomendasi berita olahraga *online* berbasis *web application* yang telah dibuat ini efektif untuk mencari topik-topik berita olahraga yang berkaitan erat dengan data latih (hasil proses *web scraping* otomatis) serta mampu memberikan rekomendasi berita olahraga terbaru kepada pengguna.

DAFTAR PUSTAKA

- Azhar, Masruroh, S.U., Wardhani, L.K., Okfalisa. (2019). Perbandingan Kinerja Algoritma Naive Bayes dan K-NN Pendekatan Lexicon pada Analisis Sentimen di Media Twitter. *Prosiding Seminar Nasional Fisika Universitas Riau IV*, 3011-1 - 3011-6.
- Dewan Pers. (2016). Catatan Dari Indonesia Untuk World Press Freedom Day. *JURNAL DEWAN PERS*.
- Fadlil, J., Mahmudy, W. F. (2007). Pembuatan Sistem Rekomendasi Menggunakan Decision Tree dan Clustering. *Jurnal Kursor*, 3, 44-56.
- Lantz, B. (2015). *Machine Learning With R*. Birmingham: Packt Publishing Ltd.
- Prarstyia, N. M. (2014). Geliat Surat Kabar Harian Olahraga di Indonesia. *Jurnal Komunikator*, 63-80.
- Rani, N. L. (2013). Persepsi Jurnalis dan Praktisi Humas terhadap Nilai Berita. *Jurnal ILMU KOMUNIKASI*, 10(1), 83-96.
- Rivki, M. Bachtiar, A.M. (2017). Implementasi Algoritma K-Nearest Neighbor dalam Pengklasifikasian Follower Twitter yang Menggunakan Bahasa Indonesia. *Jurnal Sistem Informasi (Journal of Information System)*, 13, 31-37.
- Syarifah, A. dan M. A. Muslim. (2015). pemanfaatan naïve bayes untuk merespon emosi dari kalimat berbahasa indonesia. *UNNES Journal of Mathematics*, 147-156.
- Yuan, L. (2010). An Improved Naive Bayes Text Classification Algorithm In Chinese Information Processing. *Proceedings of the Third International Symposium on Computer Science and Computational Technology*, 267-269.