**Penerapan Algoritma KNN untuk Klasifikasi Kategori Berita Berdasarkan Judul (Studi Kasus: Liputan6)**

*Brielt Bella Gracitwo1; Ryan Putranda Kristianto2*

1 FakultasTeknik, Ilmu Informatika, Universitas Katolik Darma Cendika, Surabaya, Indonesia

2FakultasTeknik, Ilmu Informatika, Universitas Katolik Darma Cendika, Surabaya, Indonesia

brielt.bella@student.ukdc.ac.id

***ABSTRACT***

Liputan6 Is a news site that has been registered and a media place for the Indonesian Press Council. Liputan6 generates news articles every hour and the data in real-time. The need to determine the type of news automatically to speed up the work of journalists in making news articles. The K-Nearest Neighbor classification algorithm is used to determine the type of news automatically. The results of this study research to obtain accuracy obtained after knowing the type of news automatically in journalism, the accuracy value obtained by researchers is 87%.

***Keywords:*** *Liputan6, News, Classification, K-Nearest Neighbor*

***ABSTRAK***

Liputan6 Merupakan tempat berita yang sudah terdaftar dan tempat media dewan pers indonesia. Setiap jam Liputan6 menghasilkan artikel berita dan data tersebut bersifat *real time.* Perlunya penentuan jenis berita secara otomatis untuk mempercepat pekerjaan jurnalis dalam pembuatan artikel berita. Digunakan algoritmaklasifikasi *K-Nearest Neighbor* dalam menentukan jenis berita secara otomatis. Hasil dari penelitian ini adalah penelitian mendapatkan akurasi yang didapat setelah mengetahui jenis berita secara otomatis pada jurnalistik, nilai akurasi yang didapat oleh peneliti sebesar 87%*.*

***Kata kunci:*** *Liputan6, Berita, Klasifikasi, K-Nearest Neighbor.*

**1. PENDAHULUAN**

Saat ini teknologi setiap harinya semakin berkembangan bahkan maju. Perkembangan dan kemajuan teknologi saat ini mempengaruhi hampir di seluruh aspek bidang kehidupan. Pengaruh dari perkembangan dan kemajuan teknologi ini juga mempengaruhi hampir di semua peminat membaca berita surat kabar ke membaca secara online melalui *gadget*, laptop atau media lainnya untuk membaca berita. Dengan kemajuan yang ada teknologi pada masa sekarang ini, kita dapat meningkatkan literasi baca bagi para literatur.

Berita merupakan sebuah informasi atas kejadian terkini tertentu, sehingga masyarakat bisa mengetahui akan peristiwa tersebut. Pada era sekarang berita sangat mudah didapatkan karena adanya teknologi internet. Terdapat survei yang dilakukan oleh Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) terdapat 210.026.769 dari 272.682.600 penduduk Indonesia pengguna internet [1]. Dari hasil survei tersebut bahwasanya hamper semua penduduk Indonesia cenderung membaca berita melalui daring. Banyak sekali portal berita daring yang ada pada saat ini salah satunya Liputan6.  Liputan6 merupakan perusahan berita yang bergerak pada portal berita dan bekerjasama dengan stasiun televisi Surya Citra Televisi (SCTV) dalam menyiarkan berita[2]. Namun, terkadang  ada beberapa portal berita yang hanya mencantumkan beritanya saja tanpa memperhatikan suatu hal penting bagi pengguna yaitu penggolongan jenis berita dan berita tersebut menggunakan data *real time,* sehingga hal tersebut membuat pengguna cenderung bingung untuk memilih jenis berita yang ingin  dibaca karena terlihat acak.  Oleh karena itu penelitian ini bertujuan untuk mengetahui judul yang tercantum pada portal berita tersebut termasuk ke dalam jenis berita apa dengan cara melakukan penggolongan dan penerapannya menggunakan klasifikasi pada algoritma *K-Nearest Neighbor*(KNN).

Algoritma KNN Merupakan salah satu algoritma pembelajaran mesin yang bersifat non parametrik pemodelan berbasis machine learning yang digunakan untuk mengetahui klasifikasi terkait penggolongan berita. pada metode Knn klasifikasi terhadap objek yang menggunakan jenis data berita. Metode tersebut sangatlah cocok digunakan bagi peneliti, karena metode itu untuk melakukan penggolongan berita sangatlah akurat.

Beberapa peneliti telah melakukan riset atau penelitian sebelumnya diantarnya peneliti menggunakan metode klasifikasi. Seperti yang dilakukan oleh peneliti (Kurniawan et al., 2017) dengan judul “Klasifikasi Berita Twitter Menggunakan Metode Improved Naïve Bayes”, Dengan hasil penelitian, peneliti mendapatkan hasil nilai precision 0.962961, recall 0.789164[3]. Peneliti selanjutnya dilakukan oleh peneliti (Ulfah and Anam 2020) dengan judul “Analisis Sentimen Hate Speech Pada Portal Berita Online Menggunakan Support Vector Machine (SVM)”[4], peneliti mendapatkan hasil penelitan dengan nilai akurasi sebesar 53,88% berdasarkan 105 komentar. Penelitan selanjutanya dilakukan oleh peneliti (Komalasari, 2018) dengan judul “Klasifikasi Berita Menggunakan Metode K-Nearest Neighbor ”[5]. Peneliti memperoleh hasil dari penelitian dengan nilai akurasi 87%. Peneliti dapat menyimpulkan bawasanya algoritma klasifikasi KNN sangatlah baik digunakan, tipe data yang digunakan dalam penelitian ini data bersifat *real-time*.

Berdasarkan penelitian ini, peneliti melakukan eksperimen dengan menentukan jenis berita. Pada penelitian ini peneliti menggunakan data *real-time*, dengan menerapkan algoritma klasifikasi KNN yaitu untuk menetukan jenis berita secara otomatis. Dengan hasil model klasifikasi penentuan jenis berita yang dilakukan oleh sistem secara otomatis, hal ini dapat membantu kinerja jurnalis Liputan6 dalam meproduksi artikel berita dengan cepat dan tepat tanpa menentukan jenis berita secara manual.

**2. METODE/PERANCANGAN PENELITIAN**

Metode Penelitian merupakan cara yang dilakukan oleh peneliti untuk mencapai tujuan tertentu, dengan cara memecahkan masalah serta menemukan penyelesaian dengan menggunakan prosedur atau peneliti terdahulu. Berdasarkan pemaparan diatas, maka peneliti melakukan perhitungan dengan menggunakan kalasifikasi Knn dalam mengklasifikasi jenis berita yang dilakukan secara otomatis. Dengan menggunakan beberapa metode yaitu : (1) Pengumpulan data; (2) Pengolahan data; (3) TF-IDF;(4) KNN; (5) Hasil Analisa*;*

**Tabel 1.** Alur Penelitian

Pengumpulan Data

Pengelolaan Data

KNN

TF-IDF

Hasil Analisa

**2.1 Pengumpulan Data**

Pengumpulan data merupakan awal dari penelitian untuk melakukan penelitian ini dengan, menggunakan pihak ke tiga yaitu google anyltic.Terdapat 1357 data yang diperoleh dari google amytic berupa file CVS.

**2.2 Pengelolaan Data**

Setelah melakukan pengambilan data, tahapa berikutnya adalah melakukan pengelolaan data. Didalam data tersebut terdapat judul berita, jenis berita, dan jumlah penggunjung. Pada tahap ini data yang akan diambil hanyalah pada judul berita dan jenis berita. Sehingga data jumlah penggunjung akan di sisihkan.

**2.3 *TF-IDF***

Terms Frequency-Inverse Document Frequency atau disingkat TF-IDF merupakan metode yang melakukan untuk menentukan perhitungan nilai dan memberikan bobot pada term didalam dokumen. Metode ini menggabungkan 2 perhitungan yaitu : (1)TF ini akan melakukan perhitungan; (2)IDF melakukan setiap dokumen. Proses Algoritma TF-IDF yang digunakan pada penelitian ini yaitu : (1)Menghitung Nilai TF; (2) Menghitung Nilai IDF; (3) Menghitung tolak ukur TFIDF.

Tabel 2 : Alur Perhitungan TF-IDF

Menghitung nilai TF

Menghitung nilai IDF

Menghitung tolak ukur TFIDF

berikut perhitungan rumus TFIDF.persamaan yang digunakan untuk menghitung nilai Term frequency pada rumus 2.1 sebagai berikut.

Keterangan Rumus :

TF(d,t) = F(d,t) (2.1)

F(d,t) = Kemunculan kata t di dalam dokumen d

Setelah dilakukan perhitungan TF, lakukan perhitungan IDF pada rumus 2.2 sebagai berikut.

Keterangan Rumus :

IDF(t) = log(N/df(t)) (2.2)

n = total dokumen latih

df(t) = dokumen frecuency dari kata t

Terms Frequency-Inverse Document Frequency menjadi tolak ukur untuk melakukan perhitungan statistic dalam proses analisia dokumen hal tersebut pada rumus 2.3 sebagai berikut.

Keterangan Rumus :

wij = tfij x idij (2.3)

wij = bobot term terhadap dokumen

tfij = jumlah kemunculan kata (term) dalam setiap dokumen

idij = jumlah kemunculan kata (term) dalam sebuah dokumen

Deskripsikan secara ringkas mengenai materi dan metode yang digunakan dalam penelitian, meliputi subyek/bahan yang diteliti, alat yang digunakan, rancangan percobaan atau desain yang digunakan, teknik pengambilan sampel, variabel yang akan diukur, teknik pengambilan data, analisis dan model statistik yang digunakan. Kutipan berturut-turut dalam tanda kurung (1)..

Persamaan matematika dinomori dengan angka Arab di dalam tanda kurung buka-tutup pada posisi rata kanan kolom. Penulisan rumus diawali dengan jorokan awal (*first line*) 1,25 cm dan diikuti dengan nomor rumus:

**2.4 *K-Nearest Neighbor***

Deskripsikan secara ringkas mengenai materi dan metode yang digunakan dalam penelitian, meliputi subyek/bahan yang diteliti, alat yang digunakan, rancangan percobaan atau desain yang digunakan, teknik pengambilan sampel, variabel yang akan diukur, teknik pengambilan data, analisis dan model statistik yang digunakan. Kutipan berturut-turut dalam tanda kurung (1)..

Persamaan matematika dinomori dengan angka Arab di dalam tanda kurung buka-tutup pada posisi rata kanan kolom. Penulisan rumus diawali dengan jorokan awal (*first line*) 1,25 cm dan diikuti dengan nomor rumus:

[satu baris kosong]

(1)

Gambar dan tabel dibuat hitam putih dan diletakkan di tengah serta harus diberi nomor urut dan judul. Huruf pertama setiap kata pada judul gambar dan tabel ditulis dengan huruf kapital kecuali kata sambung. Judul gambar diletakkan di bawah gambar sedangkan judul tabel diletakkan di atas tabel. Isi tabel dibuat 1 spasi dengan ukuran huruf 10 pt. Contoh penulisan gambar dan tabel berikut ini.

[satu baris kosong]



**Gambar 1.** Geometri Pencatuan *electromagnetically coupled*

**Tabel 1.** Spesifikasi Substrat yang Digunakan

|  |  |
| --- | --- |
| Jenis Substrat | FR4 (epoxy) |
| Konstanta Dielektrik Relatif () | 4.3 |
| Konstanta Permeabilitas Relatif () | 1 |
| Dielectric Loss Tangent () | 0.09 |
| Ketebalan Substrat (h) | 1.6 mm |
| Konduktifitas Bahan | 5.8 x 107 S/m |

[satu baris kosong]

**3. HASIL DAN PEMBAHASAN** [11 pts/Bold]

Hasil adalah inti dari suatu artikel yang menyajikan data hasil penelitian yang ditemukan dan disusun dalam ilustrasi (tabel, gambar, foto, denah, atau diagram). Pembahasan terhadap hasil penelitian dan pengujian yang diperoleh disajikan dalam bentuk uraian teoritik, baik secara kualitatif maupun kuantitatif [1]. Hasil penelitian sebaiknya ditampilkan dalam berupa grafik atau pun tabel. Untuk grafik dapat mengikuti format untuk diagram dan gambar. Kutipan berturut-turut dalam tanda kurung [2].

Pembahasan berisi penjelasan apa arti hasil dan implikasinya untuk kajian di masa depan, tidak mengulangi apa yang telah dipaparkan dalam kajian pustaka atau hasil. Hubungkan hasilnya dengan pertanyaan yang diajukan di bagian pendahuluan.

**4. KESIMPULAN DAN SARAN** [11 pts/Bold]

Kesimpulan berisi kumpulan dan meringkas hasil yang paling penting dan implikasinya. Isi kesimpulan merupakan jawaban dari tujuan penelitian bukan rangkuman hasil penelitian. Kesimpulan dan saran dibuat secara singkat, jelas, dan padat didasarkan pada hasil penelitian [3]. Kesimpulan dan Saran berupa paragraf tanpa numbering. Kutipan berturut-turut dalam tanda kurung() [4] (Gunakan kutipan menggunakan automatik/aplikasi sitasi)

**UCAPAN TERIMAKASIH** [11 pts/Bold ,jika ada]

Penulis mengucapkan terima kasih kepada Instansi/perusahaan/lembaga yang telah memberi dukungan yang membantu pelaksanaan penelitian dan atau penulisan artikel.

**DAFTAR PUSTAKA** [11 pts/Bold]

Kutipan berturut-turut dalam tanda kurung [1]. Kalimat tanda baca berikut braket [2]. Merujuk hanya untuk nomor referensi, seperti pada [3] -Jangan menggunakan "Ref. [3] "atau" referensi [3]. Minimal daftar pustaka sebanyak 15 Kutipan.

[1] MUHAMMAD ARIF, “Profil Internet Indonesia 2022,” *AJII*, 2022. https://apjii.or.id/content/read/39/559/Laporan-Survei-Profil-Internet-Indonesia-2022 (accessed Nov. 01, 2022).

[2] “Tentang Kami,” *Liputan6*, 2022. https://www.liputan6.com/info/tentang-kami (accessed Nov. 01, 2022).

[3] B. Kurniawan, M. A. Fauzi, and A. W. Widodo, “Klasifikasi Berita Twitter Menggunakan Metode Improved Naïve Bayes,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 1, no. 10, pp. 1193–1200, 2017.

[4] A. N. Ulfah and M. K. Anam, “Analisis Sentimen Hate Speech Pada Portal Berita Online Menggunakan Support Vector Machine (SVM),” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 7, no. 1, pp. 1–10, 2020, doi: 10.35957/jatisi.v7i1.196.

[5] B. P. Komalasari, “Klasifikasi Berita Menggunakan Metode KNearest Neighbor (K-NN),” vol. 5, no. 2, pp. 259–268, 2018, [Online]. Available: https://perpustakaan.ft.unram.ac.id/index.php?p=show\_detail&id=7862

**Referensi harus ditulis dalam gaya IEEE**.. Penyusunan rujukan dalam daftar pustaka berurut berdasarkan urutan sitasi yang digunakan (sekuensial) dan diberi nomor angka arab dalam kurung siku. Penulisan unsur-unsur keterangan pustaka mengikuti kaidah dengan urutan: (1) nama pengarang ditulis dengan urutan nama akhir, nama awal dan nama tengah, tanpa gelar akademik. (2) tahun penerbitan. (3) Judul. (4) tempat penerbitan. (5) nama penerbit.

**LAMPIRAN**

Jika diperlukan, tulisan dapat dilengkapi dengan lampiran.