**Penerapan Algoritma SGD 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 yang data tersebut. Data ini digunakan untuk mengklasifi jenis berita. Pada penelitian ini menggunakan *software system* informasi yaitu google anyltic. Pada data ini Terdapat 2287 data yang diperoleh dari google anyltic berupa file CSV.

**2.2 Pengelolaan Data**

Setelah melakukan pengambilan data, tahapan berikutnya adalah melakukan pengelolaan data. data tersebut terdapat judul berita, jenis berita, dan jumlah penggunjung. Pada tahap ini data yang sudah didapat, data tersebut dilakukan pengelolaan data. Data yang setelah dilakukan penglolaan terdapat data judul berita dan jenis berita. Sehingga data tersebut dapat dilakukan penyisian pada pengunjung.

**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

**2.4 *K-Nearest Neighbor***

Algoritma k-nearest neighbor merupakan algoritma yang digunakan untuk pembelajaran terawasi, yang memprediksi bahwa data termasuk dalam kelas dengan mayoritas kelas tetangga terdekat. Tentukan berapa banyak K dari tetangga terdekat untuk data prediksi, dan pertimbangkan *class* mayoritas dari K tetangga. Tujuan dari algoritma ini adalah untuk mengklasifikasikan objek baru berdasarkan atribut dan sample-sample dari training data. Berikut rumus KNN.

𝑑(𝑥, 𝑦) = √∑ (𝑥𝑖 − 𝑦𝑖) 𝑛 2 𝑖=1

(1)

Ket:

d(x,y) = Jarak data uji dengan data latih

n = Banyaknya data latih

x = Data latih

y = Data uji

**2.5 Hasil Analisa**

hasil Analisa yang digunakan pada penelitian ini untuk menentukan jenis berita di liputan 6 dengan membuat pemodelan klasifikasi jenis berita yang menggunakan metode knn, sehingga metode yang digunakan peneliti tersebut dapat membantu jurnalis liputan 6 dalam memproduksi artikel jenis berita dengan cepat dan tidak menentukan jenis berita secara manual.

**3. HASIL DAN PEMBAHASAN**

**3.1 Pengumpulan Data**

Penelitan ini mengambil data realtime dengan menggunakan pihak ketiga yaitu Google Analytics. Sebelumnya peneliti mendapatkan akses melalui program magang antara Universitas Katolok Darma Cendika dengan PT.Liputan6. Data yang diambil berupa file dengan format CSV, didalamnya terdapat data judul berita, jenis berita, dan jumlah pengunjung website. Jumlah keselurahn data sebesar 2287, berikut contoh data 5 teratas.

|  |  |  |
| --- | --- | --- |
| **Judul** | **Jenis Berita** | **User Aktif** |
| Wajah Istri Sultan Kelantan yang Pernah Bersengketa dengan Miss Moscow Terlihat untuk Pertama Kali | Lifestyle | 7 |
| Nonton Video Porno Bareng Pasangan Beri Dampak Positif atau Justru Negatif? | Health | 7 |
| Cuek dengan Isu Perselingkuhan Behati Prinsloo Tertangkap Kamera Temani Adam Levine Manggung | Health | 5 |
| Skrining Hormon Tiroid pada Bayi Baru Lahir Bisa Cegah Gangguan Tumbuh Kembang Anak | Health | 5 |
| Wajah Istri Sultan Kelantan yang Pernah Bersengketa dengan Miss Moscow Terlihat untuk Pertama Kali | Lifestyle | 5 |

**3.2 Pengelolaan Data**

Setelah dilakukan pengumpulan data, data akan dilakukan uji coba untuk pemilihan file apa saja yang idibutuhkan dalam kalasifikasi jenis berita menggunakan TF-IDF. Hal yang penting dalam penelitian ini untuk klasifikasi jenis berita menggunkan 2 file yaitu Judul Berita dan Jenis Berita. Berikut 2 file yang digunakan dalam klasifikasi jenis berita pada tabel 2.

|  |  |
| --- | --- |
| **Jenis Berita** | **User Aktif** |
| Health | 9 |
| Hot | 7 |
| Islami | 7 |
| Crypto | 6 |
| Regional | 6 |

**3.3 *TF-IDF***

**Perhitungan ini nanti aja mas sama”**

**3.4 *K-Nearest Neighbor***

**Perhitungan ini nanti aja mas sama”**

**3.5 Hasil Analisa**

**Ini juga**

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

Berdasarkan hasil perhitungan diatas, maka dapat ditarik sebuah kesimpulan penelitian yang tercantum pada poin-poin berikut:

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]

Terima kasih disampaikan kepada Liputan 6 atas telah mengizinkan untuk dilakukan penelitian ini, serta terimakasih juga kepada Bapak ... yang telah membantu pengambilan data liputan 6, serta kepada Ryan Putrandra Kristanto terimakasih sudah menjadi dosen pendamping dan mmebantu pelaksanaan penelitian ini.

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.