

KLASIFIKASI BERITA ONLINE PADA LIPUTAN 6 MENGGUNAKAN METODE KNN

Andika Tri Saputra¹, Yasin Aril Mustofa², Yulianty Lasena^{3*}, Irma Surya Kumala Idris⁴

^{1,2,3,4}Fakultas Ilmu Komputer, Teknik Informatika, Universitas Ichsan Gorontalo, Gorontalo, Indonesia

Email: andikatrisaputra14@gmail.com, arieldcc@gmail.com, yuliantylasena86@gmail.com, mhaladp@gmail.com

Abstrak - Penelitian ini bertujuan untuk menguji efektivitas metode klasifikasi KNN dalam mengklasifikasi berita online pada Liputan 6. Data yang digunakan dalam penelitian ini adalah data berita online dari Liputan 6 yang diterbitkan pada tanggal 19 November 2023 s/d tanggal 25 November 2023 sebanyak 788 judul berita. Data tersebut meliputi judul berita, tanggal berita, link berita dan kelas berita. Hasil penelitian menunjukkan bahwa metode klasifikasi KNN dapat mengklasifikasi berita online pada Liputan 6 dengan akurasi sebesar 91% untuk 6 kelas berita yang berbeda, yaitu kelas Bisnis, Pemilu, Islami, Showbiz, Bola dan Tekno. Hasil penelitian ini menunjukkan bahwa metode klasifikasi KNN dapat digunakan untuk melakukan klasifikasi berita online dengan baik dan benar.

Kata Kunci: data mining, klasifikasi, KNN, TF-IDF, berita online, Liputan 6

Abstract This research aims to test the effectiveness of the KNN classification method in classifying online news on Liputan 6. The data employed are online news data from Liputan 6 published from November 19, 2023, to November 25, 2023, namely 788 news titles. The data includes news titles, dates, links, and news classes. The results indicate that the KNN classification method can classify online news on Liputan 6 with an accuracy of 91% for six distinguished news classes, namely Business, Election, Islamic, Showbiz, Ball, and Techno classes. The research results indicate that the KNN classification method can be used to classify online news properly and correctly

Keywords: data mining, classification, KNN, TF-IDF, online news, Liputan 6

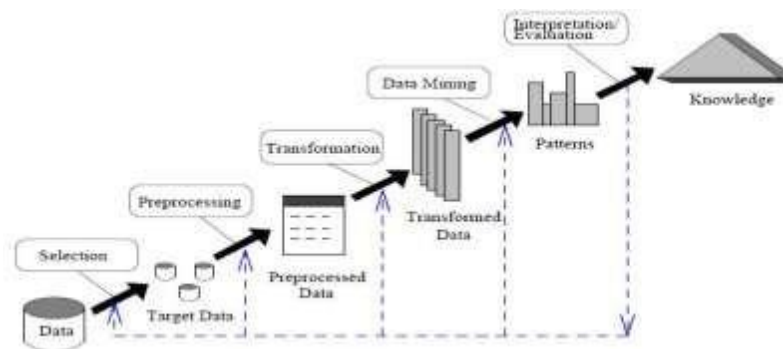
1. PENDAHULUAN

Berita merupakan sebuah informasi *public* yang turut ambil bagian bagi perkembangan peradaban manusia. Begitu banyaknya berita yang akan diterbitkan setiap harinya membuat tim editor kewalahan dalam menentukan kategori berita yang akan diterbitkan, tentu saja kondisi tersebut membutuhkan waktu yang cukup lama untuk menentukan kategori tiap berita yang masuk. [1] Keunggulan dari metode KNN dibuktikan dari peneliti sebelumnya yaitu penelitian yang dilakukan oleh Andi Ahmad Irfa 2018, dalam penelitiannya, peneliti menggunakan metode KNN untuk melakukan klasifikasi Topik Berita yang Berbahasa Indonesia hasil penelitian yang diperoleh yaitu mendapatkan *performa micro average f1-measure* sebesar 69,9%. [2], penelitian yang dilakukan oleh Nengah Widya Utami, 2022 menganalisis mengenai sentiment pembelajaran daring pada masa covid 19 dengan algoritma KNN, hasil akurasi yang diperoleh sebesar 82.48%. [3] dan penelitian yang dilakukan oleh Retno Sari, 2020 membahas tentang penerapan metode KNN dalam menganalisis sentimen pada Review Objek Wisata Dunia Fantasi, nilai akurasi yang diperoleh dengan nilai $k=7$ yaitu akurasi 77.01, *precision* 92.38, *recall* 61.56 dan nilai AUC 0.894. [4]

2. TINJAUAN PUSTAKA

2.1 Data Mining

Menurut Daryl Pregibon menyatakan bahwa "Data mining adalah campuran statistik, kecerdasan buatan, dan riset basis data" [5]. Proses *Knowledge Discovery in Database* (KDD) merupakan bagian dari data mining. KDD adalah proses terorganisir untuk mengidentifikasi pola yang valid, baru, berguna, dan dapat dimengerti dari sebuah dataset yang besar dan kompleks [6].



Gambar 2. 1 Proses *Knowledge Discovery in Database* (KDD) [6]

2.2. Fungsi Data Mining

Penambangan data sebenarnya mempunyai banyak sekali manfaat atau fungsi. Namun, secara umum, terdapat 2 fungsi data *mining* yaitu: [7]

1. Deskriptif: Kapasitas yang berbeda mengacu pada kapasitas dalam memahami informasi yang terkonsentrasi lebih lanjut.
2. Prediktif: Fungsi Prediktif dari data mining adalah elemen interaksi yang pada akhirnya akan mengungkap contoh informasi tertentu.

2.3 Text Mining

Text mining ialah cara yang dapat diterapkan untuk mengekstraksi sebuah data untuk diteliti yang mana data tersebut bersumber dari data dalam bentuk teks. salah satu disiplin ilmu yang diterapkan pada informasi *retrival*, *linguistic* komputasi, *machine learning*, *statistic* dan data mining adalah *Text Mining*. [8] untuk melakukan information extraction, klasifikasi, clustering dan information retrieval [9]. *Text mining* adalah upaya untuk mendapatkan informasi supayapengguna mampu berinteraksi terhadap sekumpulan dokumen dari masa ke masa dengan menerapkan perangkat analisis [10].

2.3.1 Algoritma KNN

Algoritma KNN adalah sebuah metode klasifikasi terhadap sekumpulan data berdasarkan pembelajaran data yang sudah terklasifikasikan sebelumnya. Termasuk dalam *supervised learning*, dimana hasil *query instance* yang baru diklasifikasikan berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam KNN. [11]

2.3.2 Penerapan algoritma KNN

Contoh penerapan algoritma KNN (*K-Nearest Neighbors*) dalam *text mining*:

1. Menentukan dataset yang akan digunakan. Misalnya, kita memiliki dataset berikut ini:

Tabel 2. 2 Contoh Data *Training*

	<i>Text</i>
	Aku suka makan nasi goreng
	Anjing tersebut sangat lucu
	Dia adalah teman terbaikku

- Mengubah teks menjadi vektor *numerik*. Misalnya, menggunakan metode TF- IDF (*Term Frequency-Inverse Document Frequency*) untuk menghasilkan vektor *numerik* dari setiap teks dalam dataset.

Tabel 2. 3 Hasil Pengolahan TF-IDF Data *Training*

	Text
	0.7, 0.3, 0, 0, 0
	0, 0, 0.5, 0.5, 0
	0, 0, 0, 0, 0.7

Pada contoh ini, kita menggunakan 5 fitur *numerik*, dengan masing-masing fitur merepresentasikan kata-kata yang muncul dalam dataset

- Menentukan dokumen baru yang akan diklasifikasikan. Misalnya, kita memiliki dokumen baru dengan teks berikut:

Tabel 2. 4 Data *Testing*

	Text
	Aku suka anjing

- Mengubah dokumen baru menjadi vektor *numerik* menggunakan metode yang sama seperti langkah sebelumnya.

Tabel 2. 5 Hasil TF_IDF Data *Testing*

	Text
	0.5, 0.2, 0, 0.3, 0

- Menentukan parameter K, yaitu jumlah tetangga terdekat yang akan digunakan. Misalnya, kita menggunakan K = 1.
- Menghitung jarak antara dokumen baru dan setiap dokumen dalam dataset menggunakan metode yang sesuai, misalnya jarak *Euclidean*.
 - Jarak *NewText* dengan *Text1*: $\sqrt{((0.7-0.5)^2 + (0.3-0.2)^2 + (0-0)^2 + (0-0.3)^2 + (0-0)^2)} = 0.2$
 - Jarak *NewText* dengan *Text2*: $\sqrt{((0-0.5)^2 + (0-0.2)^2 + (0.5-0)^2 + (0.5-0.3)^2 + (0-0)^2)} = 0.854$
 - Jarak *NewText* dengan *Text3*: $\sqrt{((0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0.7-0)^2)} = 0.7$
- Mengurutkan jarak secara menaik dan memilih K dokumen terdekat. Dalam contoh ini, kita memilih 1 dokumen terdekat.

- Dokumen terdekat : *Text 1*

- Melabeli dokumen terdekat berdasarkan mayoritas kelas tetangga terdekat. Misalnya, jika *Text1* memiliki label kelas "positif", maka dokumen baru juga akan diberi label kelas "positif".

Dalam contoh ini, kita dapat mengklasifikasikan dokumen baru "Aku sukaanjing" menjadi kelas "positif" berdasarkan KNN dengan $K = 1$.

2.3.3 Confusion Matrix

Confusion matrix adalah sebuah tabel yang digunakan untuk mengevaluasi performa dari suatu sistem klasifikasi. Tabel ini berisi empat jenis hasil klasifikasi yang mungkin terjadi dalam sebuah percobaan klasifikasi, yaitu *true positive (TP)*, *false positive (FP)*, *true negative (TN)*, dan *false negative (FN)*.

True positive (TP) adalah jumlah data positif yang benar diprediksi sebagai positif oleh sistem klasifikasi. *False positive (FP)* adalah jumlah data negatif yang salah diprediksi sebagai positif. *True negative (TN)* adalah jumlah data negatif yang benar diprediksi sebagai negatif. *False negative (FN)* adalah jumlah data positif yang salah diprediksi sebagai negatif.

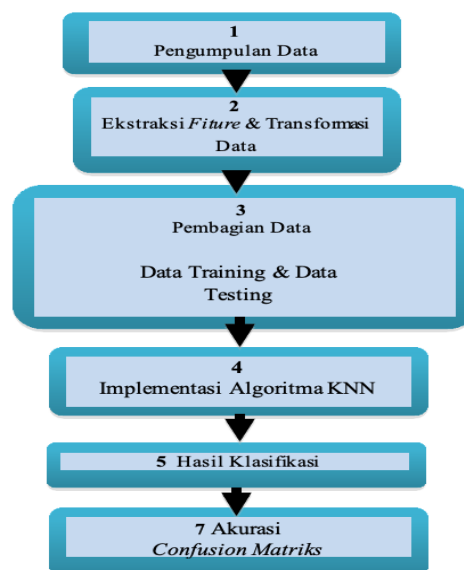
3. METODE PENELITIAN

3.1 Jenis, Metode, Subjek, Objek, Waktu dan Lokasi Penelitian

Penelitian ini menggunakan metode penelitian studi kasus media berita. Dengan demikian jenis penelitian ini adalah kuantitatif. Subjek penelitian ini adalah Analisa pengelompokan berita di Indonesia Menggunakan Metode KNN. Penelitian ini dimulai dari Agustus 2023 – November 2023 mencakup pengumpulan dan pengolahan data beserta bimbingan.

3.1.1 Model Yang Diusulkan

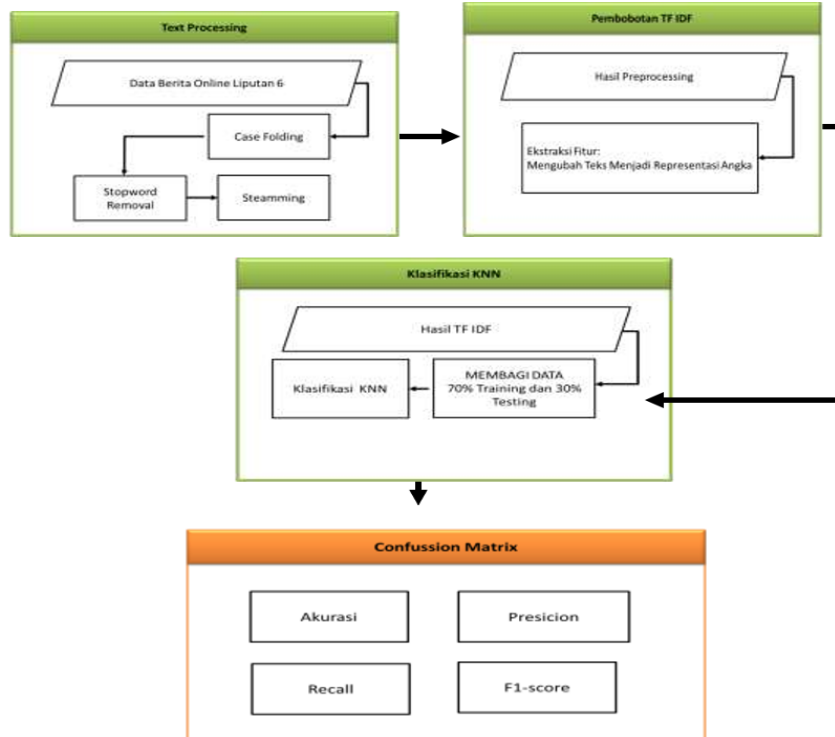
Berikut adalah tahapan penelitian yang akan dilakukan peneliti.



Gambar 2.2 Model Penelitian

3.2 Pemodelan

Penulis Membuat flowchart untuk model yang dibangun:



3.2.1 Pembobotan TF-IDF

Peneliti membuat *flowchart* untuk pembobotan TF-IDF berikut ini.



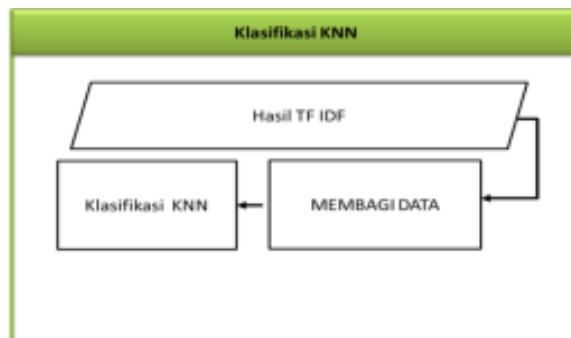
Gambar 4. 4 Pembobotan TF-IDF

Tabel 4. 12 Hasil Ekstraksi Fitur TF-IDF Berita online

HASIL EKSTRAKSI FITUR TF-IDF BERITA ONLINE	
(0, 814)	0.31080953051052124
(0, 261)	0.33403159063840226
(0, 922)	0.29433321664134726
(0, 887)	0.20509414949439395
(0, 1121)	0.27111115651346623
(0, 798)	0.29433321664134726
(0, 1042)	0.33403159063840226
(0, 942)	0.33403159063840226
:	:
(248, 603)	0.48973689538951704
(248, 769)	0.38454277967496103
(248, 978)	0.4315335428378243
(248, 841)	0.4315335428378243

3.2.2 Klasifikasi KNN

Proses Klasifikasi *KNN* dilakukan Setelah pembobotan menggunakan TF-IDF berhasil dilakukan. Pada tahapan ini peneliti melakukan klasifikasi menggunakan algoritma *KNN classifier* dengan *google colab*. Dibawah ini adalah *flowchart* untuk proses klasifikasi *KNN*



Gambar 4. 5 Klasifikasi KNN

3.2.3 Confusion Matrix

Untuk mengetahui kualitas dari hasil penerapan metode diperlukan evaluasi metode , dalam evaluasi metode atau algoritma peneliti menggunakan pengujian atau evaluasi *confusion matrix*.

Confusion matrix digunakan untuk mengevaluasi sistem klasifikasi yang berdasarkan kemampuan parameter prediktif suatu model. parameter prediktif adalah parameter yang mengukur kualitas set uji statistik yang diberikan berdasarkan atribut yang ada dalam sampel.

3.3 Hasil Pengumpulan Data

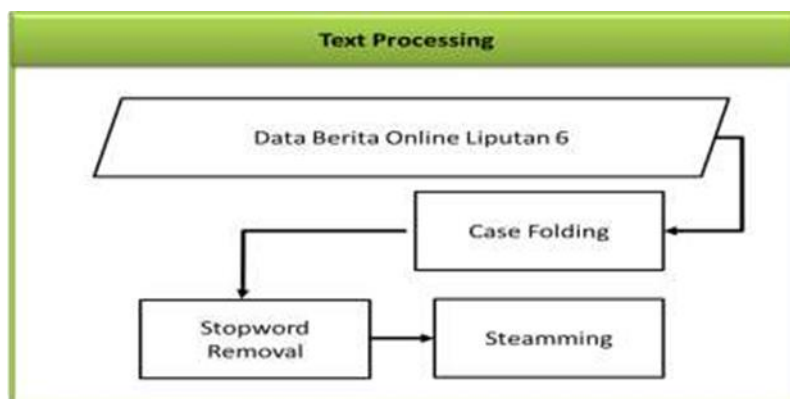
Dataset berita online yang digunakan pada penelitian ini adalah dataset yang dikumpulkan dari laman <https://www.liputan6.com/> dengan menggunakan teknik *Web scraping*. Teknik ini digunakan untuk mengambil data dari halaman web secara otomatis. Teknik *Web scraping* dilakukan sesuai jumlah jenis berita pada laman tersebut yang banyak diakses oleh pengunjung atau masyarakat. Berita online yang diunduh adalah berita yang diterbitkan dari tanggal 19 November 2023 s/d 25 November 2023. Adapun jumlah dataset berita online yang sudah dikumpulkan berdasarkan jenis berita dapat dilihat pada Tabel 4.1 berikut ini :

Tabel 4. 1 Jumlah Dataset Berdasarkan Jenis Berita

No	Jenis Berita	Jumlah Dataset
1	Pemilu	140
2	Bisnis	140
3	Bola	140
4	Showbiz	140
5	Tekno	121
6	Islami	106
Total Dataset		787

3.4 Preprocessing Data

Peneliti membuat *flowchart* mengenai *preprocessing*.



Gambar 4.2 Preprocessing

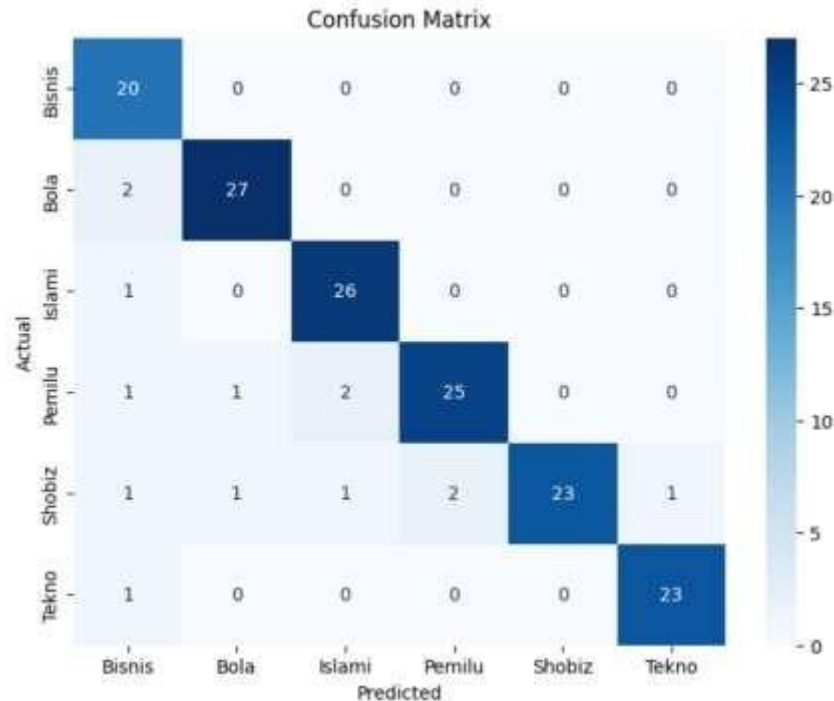
Pada tahapan *preprocessing* peneliti menggunakan *framework google colab* berbasis *python*, untuk melakukan proses *Case folding*, *stopword removal* dan *stemming*.

4. HASIL DAN PEMBAHASAN

4.1 Pembahasan Kinerja Model

Klasifikasi berita online pada Liputan 6 dengan menggunakan model yang sudah dilatih yaitu metode KNN dilakukan pengukuran kinerja model dengan menggunakan metode *Confusion Matriks* pada data testing sebanyak 25 data. Adapun hasil pengukuran kinerja model tersebut dalam

melakukan klasifikasi berita online antaradata aktual dan hasil klasifikasi dapat dilihat pada Gambar 5.1 berikut :



Gambar 5. 1 Confusion Matriks Data Testing

4.2 Pembahasan Hasil Pemodelan

Model KNN dalam klasifikasi berita online yang paling optimal atau memiliki akurasi paling tinggi perlu dilakukan beberapa percobaan, untuk metode KNN harus dicari nilai k yang paling baik dan persentase pembagian data training dan data testing apakah digunakan perbandingan 80:20 atau 70:30. Hal ini bisa dilakukan dengan cepat dengan menggunakan teknik Elbow dalam penentuan nilai k terbaik

KESIMPULAN

Berdasarkan hasil penelitian yang diuraikan di atas mengenai klasifikasi berita online pada liputan6 dengan metode KNN dapat ditarik kesimpulan bahwa :

1. Penerapan metode KNN dalam klasifikasi berita online pada liputan6 dapat diterapkan dengan baik untuk melakukan klasifikasi berita kedalam 6 kelas berita yang berbeda yaitu : Bisnis, Pemilu, Islami, Showbiz, Bola dan Tekno.
2. Metode klasifikasi KNN dapat mengklasifikasi berita online pada liputan6 dengan akurasi sebesar 91% untuk 6 kelas berita yang berbeda

DAFTAR PUSTAKA

- [1] D. Ariadi and K. Fithriasari, "Klasifikasi Berita Indonesia Menggunakan Metode Naive Bayesian Classification dan Support Vector Machine dengan Confix Stripping Stemmer," *J. SAINS DAN SENI ITS Vol. 4, No.2*, vol. 4, no. 2, pp. 248–253, 2015.
- [2] G. Tika, A. Adiwijaya, and N. D. W. Cahyani, "Klasifikasi Topik Berita Berbahasa Indonesia

- Menggunakan Multilayer Perceptron,” *eProceedings Eng.*, vol. 6, no. 1, pp. 2137–2143, 2019.
- [3] 2021 Herfandi, Arman Diansyah, Eri Sasmita Susanto, “JINTEKS (Jurnal Informatika Teknologi dan Sains) ISSN 2686-3359 (Online) JINTEKS (Jurnal Informatika Teknologi dan Sains) ISSN 2686-3359 (Online),” (*Jurnal Inform. Teknol. dan Sains*, vol. 3, no. 1, pp. 301–307, 2021.
- [4] R. Sari, “Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi Menggunakan Algoritma K-Nearest Neighbor (K-Nn),” *EVOLUSI J. Sains dan Manaj.*, vol. 8, no. 1, pp. 10–17, 2020.
- [5] A. Bengnga and R. Ishak, “Prediksi Jumlah Mahasiswa Registrasi Per Semester Menggunakan Linier Regresi Pada Universitas Ichsan Gorontalo,” *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 136–143, 2018.
- [6] I. A. Ahmad Sabri, M. Man, W. A. W. Abu Bakar, and A. N. Mohd Rose, “Web Data Extraction Approach for Deep Web using WEIDJ,” *Procedia Comput. Sci.*, vol. 163, pp. 417–426, 2019.
- [7] “Data Mining: Pengertian, Fungsi, Penerapan, dan Metode Pengambilannya - Cermati.com.”
- [8] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, “Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM),” *Jambura J. Electr. Electron. Eng.*, vol. 5, no. 1, pp. 32–35, 2023.
- [9] A. Firdaus and W. I. Firdaus, “Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan),” *J. JUPITER*, vol. 13, no. 1, p. 66, 2021.
- [10] R. Feldman and J. Sanger, *The Text Mining Handbook*. 2006.
- [11] M. Learning, “Learning basic concept,” pp. 1–39, 2019.