



---

## Analisis *Big Data* Sentimen Konsumen UMKM Sektor Kuliner Menggunakan *Multi-Label* *K-Nearest Neighbor*

Muhammad Ridho Fauzi<sup>1</sup>, Risky Adit Pratama<sup>2</sup>, Panji Laksono<sup>3</sup>, Puspa Eosina<sup>4\*</sup>

<sup>1,2,3,4\*</sup> Program Studi Teknik Informatika, Fakultas Teknik dan Sains,  
Universitas Ibn Khaldun Bogor, Indonesia

ralhafizhi@gmail.com<sup>1</sup>, risky.aditp06@gmail.com<sup>2</sup>, panjilakso17@gmail.com<sup>3</sup>,  
puspa.eosina@ft.uika-bogor.ac.id<sup>4\*</sup>

### Abstrak

UMKM di Kota Bogor merupakan UMKM yang mengalami perkembangan yang signifikan dari tahun ke tahun. Pada tahun 2019, di Kota Bogor tercatat ada sekitar 30.822 UMKM yang didominasi sektor kuliner 60%, kerajinan 30% dan 10% sektor batik. Dengan adanya pertumbuhan UMKM yang signifikan tersebut, maka akan berpengaruh terhadap persaingan UMKM yang semakin meningkat dan akan semakin sulit untuk melihat peta persaingan pada UMKM, khususnya pada sektor kuliner. Sehingga untuk mengatasi persaingan tersebut salah satu cara yang dapat dilakukan adalah melihat sentimen konsumen terhadap UMKM tersebut dengan menggunakan metode Data Mining. Pada metode Data Mining, terdapat beberapa algoritma populer yang sering digunakan. Salah satu algoritma yang populer adalah *K-Nearest Neighbor* (*K-NN*). *KNN* dapat digunakan pada berbagai kasus penelitian, seperti *multi-class classification*, *binary classification* dan *multi-label classification* yang dapat membantu dalam memprediksi berbagai kasus di kehidupan sehari-hari. Sehingga dalam hal ini, kami mengusulkan sebuah ide baru, yaitu analisis *Big Data* sentimen konsumen UMKM sektor kuliner menggunakan *Multi-Label K-Nearest Neighbor*. Ide tersebut dapat bermanfaat bagi pelaku UMKM sektor kuliner di Kota Bogor dalam mengetahui insight dari hasil analisis sentimen konsumen UMKM dan dapat membantu untuk pengambilan sebuah keputusan bisnis dalam meningkatkan daya saingnya terhadap kompetitor sektor kuliner yang berada di Kota Bogor.

**Kata kunci:** Analisis Sentimen, *Data Mining*, *K-Nearest Neighbor*, UMKM

### Abstract

UMKM in Bogor City are UMKM that have experienced significant development from year to year. In 2019, in the City of Bogor, there were around 30,822 UMKM dominated by the culinary sector 60%, handicrafts 30% and 10% the batik sector. With the significant growth of UMKM, it will affect the competition for UMKM which is increasing and it will be increasingly difficult to see the map of competition in UMKM, especially in the culinary sector. So to overcome this competition, one way that can be done is to see consumer sentiment towards these UMKM by using the *Data Mining* method. In the *Data Mining* method, there are several popular algorithms that are often used. One of the popular

---

*algorithms is K-Nearest Neighbor (K-NN). KNN can be used in various research cases, such as multi-class classification, binary classification and multi-label classification which can help predict various cases in everyday life. So in this case, we propose a new idea, namely Big Data analysis of consumer sentiment of UMKM culinary sector using Multi-Label K-Nearest Neighbor. This idea can be useful for UMKM players in the culinary sector in the City of Bogor in knowing the insights from the results of the UMKM consumer sentiment analysis and can help make business decisions in increasing their competitiveness against culinary sector competitors in Bogor City.*

**Keywords:** *Data Mining, K-Nearest Neighbor, Sentiment Analysis, UMKM*

## PENDAHULUAN

Pada saat ini peranan Usaha, Mikro, Kecil dan Menengah (UMKM) sangat penting dalam perekonomian di Indonesia. Hingga tahun 2017, tercatat sekitar 62.922.617 unit atau 99,99 persen usaha di Indonesia adalah UMKM, sedangkan 0,01 persen lainnya termasuk ke dalam usaha besar dengan jumlah 5.460 unit usaha besar. Sehingga UMKM telah mampu menunjukkan eksistensinya dalam perekonomian di Indonesia (Kementerian Koperasi dan Usaha Mikro, Kecil dan Menengah Republik Indonesia., 2017).

Berdasarkan data perkembangan UMKM di Indonesia di atas, UMKM di Kota Bogor merupakan UMKM yang mengalami perkembangan yang signifikan dari tahun ke tahun. Pada tahun 2019, di Kota Bogor tercatat ada sekitar 30.822 usaha UMKM yang didominasi sektor kuliner 60 persen, kerajinan 30 persen dan 10 persen lainnya adalah sektor batik. Dengan adanya pertumbuhan UMKM yang signifikan tersebut, maka akan berpengaruh terhadap persaingan UMKM yang semakin meningkat dan akan semakin sulit untuk melihat peta persaingan pada UMKM, khususnya pada sektor kuliner. Sehingga untuk mengatasi persaingan tersebut salah satu cara yang dapat dilakukan adalah melihat dan menganalisis sentimen konsumen terhadap UMKM tersebut. Analisis sentimen merupakan metode yang dapat menyelesaikan permasalahan dalam menentukan pengelompokkan berbagai ulasan berdasarkan aspek polaritasnya seperti ulasan positif, negatif, atau netral [1]. Salah satu metode yang dapat digunakan dalam menyelesaikan permasalahan tersebut, adalah metode *Data Mining*.

*Data Mining* merupakan salah satu cabang ilmu dari *Artificial Intelligence* atau kecerdasan buatan yang memiliki seni dan ilmu untuk menemukan pengetahuan, *insight* atau wawasan dan pola dari sekumpulan data baik itu data yang skalanya besar atau kecil [2]. *Data Mining* dapat diterapkan diberbagai bidang mulai dari bidang akademik, bisnis, hingga bidang medis [3]. Pada metode *Data Mining*, terdapat beberapa algoritma populer yang sering digunakan, diantaranya adalah *Decision Tree*, *Artificial Neural Network*, *Support Vector Machine (SVM)*, *K-Nearest Neighbor (K-NN)*, dan *K-Means*. Adapun algoritma yang akan digunakan dan dibahas pada penelitian ini adalah algoritma *K-Nearest Neighbor (K-NN)*.

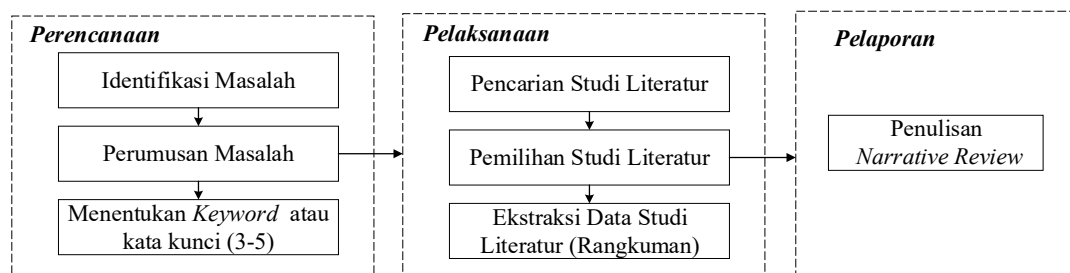
Algoritma *K-Nearest Neighbor* pernah dilakukan dalam penelitian sebelumnya oleh Muhammad Riefky dkk, mereka menerapkan algoritma K-NN untuk mendapatkan dan membandingkan hasil ketepatan klasifikasi sentimen pengguna twitter terhadap SEA Games Filipina pada tahun 2019 [4]. Penelitian lain dilakukan oleh Fikar Cevi Anggian dkk, mereka mengimplementasikan metode *K-Nearest Neighbor* untuk mengklasifikasikan status gunung berapi yang ada di Indonesia [5]. Penelitian yang lain dilakukan oleh Reski Mai Candra dkk dalam mengklasifikasi komentar bullying pada Instagram dengan menggunakan metode *K-Nearest Neighbor*. Penelitian lainnya juga dilakukan oleh Dede Satriawan dkk, yang dilakukan oleh mereka adalah mengklasifikasi tahu putih murni dan tahu putih yang

mengandung formalin dengan perangkat keras tambahan yaitu Arduino Mega beserta sensor input dari Grove-HCHO [6]. Penelitian yang lainnya juga pernah dilakukan oleh Komang Anggada Sugiarta dkk, mereka menggabungkan dua metode antara K-NN dengan *Bat Algorithm* untuk mencari dan mengklasifikasi penyakit ginjal kronis [7].

Berdasarkan uraian tersebut, para peneliti sejauh ini berfokus pada kasus *multiclass classification* dan pemilihan *hyperparameter K* tanpa teknik optimasi *hyperparameter* atau *GridSearchCV* dalam pengaruhnya terhadap hasil akurasi klasifikasi. Maka pada ulasan kali ini, kami akan menjelaskan lebih detail tentang *multiclass classification* dan pemilihan *hyperparameter K*. Penelitian ini mengusulkan sebuah gagasan untuk analisis sentimen konsumen terhadap data ulasan UMKM sektor Kuliner dengan menggunakan metode *Data Mining*. Maka tujuan dari penelitian ini adalah menghasilkan gagasan dalam bentuk artikel ilmiah untuk mengklasifikasikan sentimen konsumen terhadap UMKM sektor kuliner di Kota Bogor menggunakan algoritma K-NN dengan kasus *multi-label classification* dan menggunakan teknik optimasi *hyperparameter* atau *GridSearchCV*.

## METODE PENELITIAN

Berikut adalah prosedur penelitian yang digunakan dalam penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Bagan Prosedur Penelitian

### a. Perencanaan

Beberapa tahapan yang akan dilakukan dalam tahap perencanaan adalah sebagai berikut:

#### 1. Identifikasi Masalah

Tahap ini peneliti akan memahami tujuan dan kebutuhan dari sudut pandang bisnis yang kemudian menterjemahkan pengetahuan ini ke dalam pendefinisian masalah *Data Mining*. Dalam hal ini, peneliti memiliki dua tujuan yaitu mendapatkan gagasan atau ide penelitian baru dari hasil studi literatur mengenai K-NN dan menerapkan algoritma K-NN dalam analisis sentimen konsumen terhadap UMKM kuliner. Oleh karena itu untuk mencapai tujuan tersebut, pada tahap selanjutnya akan ditentukan rencana dan strategi untuk mencapai tujuan tersebut yang dirumuskan dalam bentuk pertanyaan dan penentuan kata kunci.

#### 2. Perumusan Masalah

Tahap ini peneliti merumuskan suatu permasalahan dari hasil identifikasi masalah dalam bentuk pertanyaan. Perumusan masalah ini dikategorikan sebagai pertanyaan utama yang dicari dan yang akan dijawab melalui penelitian yang dalam hal ini adalah *literature review* atau studi literatur.

### 3. Menentukan Keyword atau Kata Kunci

Setelah melakukan identifikasi masalah dan perumusan masalah di atas, maka tahap selanjutnya adalah menentukan kata kunci. Kata kunci ini merupakan kata-kata yang penting dan sering muncul di dalam sebuah teks atau dokumen. Sehingga, kata kunci ini berfungsi untuk mempermudah dalam pencarian informasi agar informasi yang dicari tepat sasaran. Kata kunci yang dipakai adalah *K-Nearest Neighbor*, *UMKM*, Analisis Sentimen, *Klasifikasi*.

#### b. Pelaksanaan

Setelah melakukan perencanaan, maka tahap selanjutnya adalah pelaksanaan. Tahap ini merupakan tahap pengerjaan dari apa yang telah direncanakan, dimana terdapat pencarian, pemilihan, dan ekstraksi studi literatur.

##### Pencarian Studi Literatur

Pada tahap ini adalah melakukan pencarian studi literatur dengan menggunakan kata kunci yang telah ditentukan pada tahap sebelumnya, dengan mencari jurnal ilmiah yang terindeks sebanyak 50 jurnal nasional maupun internasional.

##### Pemilihan Studi Literatur

Pada tahap ini dilakukan pemilihan studi literatur yang akan dipakai dari hasil pencarian jurnal dan tahap selanjutnya akan dilakukan ekstraksi data studi literatur.

##### Ekstraksi Data Studi Literatur

Pada tahap ini dilakukan ekstraksi data studi literature dengan cara merangkuman, melihat kelebihan dan kekurangan setiap jurnal untuk mendapatkan gagasan atau ide penelitian baru.

#### c. Pelaporan

Setelah melakukan perencanaan dan pelaksanaan, maka tahap terakhir adalah pelaporan. Tahap ini merupakan tahap akhir dari penelitian yaitu membuat ulasan studi literatur dalam bentuk artikel *narrative review*.

## HASIL DAN PEMBAHASAN

### a. Algoritma K-Nearest Neighbor (K-NN)

Algoritma *K-Nearest Neighbor* merupakan salah satu algoritma untuk mengklasifikasikan objek data baru terhadap data latih berdasarkan tingkat kemiripan yang paling banyak dan jarak yang terdekat dengan objek data baru tersebut. Maka, dekat atau jauhnya jarak akan menggunakan teknik perhitungan [8], [9]. Sehingga dalam algoritma K-NN ada salah satu *hyperparameter* yang biasa digunakan oleh para peneliti yaitu menentukan jumlah objek terdekat yang diinginkan, yaitu nilai  $k$  objek [8], [10]–[12]. Selain itu, mereka melakukan teknik *k-fold cross validation* hal ini bertujuan untuk mengevaluasi model klasifikasi dengan cara membagi data *training* secara *random* ke dalam beberapa bagian perbandingan yang sama, sehingga akan menghasilkan *error rate* dan seluruh nilai *error rate* akan dihitung untuk memperoleh nilai rata-rata *error rate* tersebut [13]–[15].

### b. Hyperparameter K

Dalam penentuan nilai  $k$ , digunakan bilangan ganjil yang tidak lebih besar dari jumlah data latih [16], [17]. Jika nilai  $k$  bilangan genap, maka hasil klasifikasi akan sulit untuk ditentukan karena masing-masing kelas prediksi bernilai sama. Penentuan nilai  $k$  yang

bervariasi dapat mempengaruhi akurasi model prediksi yang dihasilkan oleh algoritma K-NN [18]. Dalam penelitiannya, [19] menggunakan 6 nilai  $k$  dalam 6 skenario, yaitu 5, 10, 15, 20, 25 dan 30 dan menghasilkan akurasi tertinggi pada nilai  $k=20$  untuk memprediksi masa studi mahasiswa.

Penelitian yang lain dalam kasus sistem rekomendasi *e-commerce* dengan menggunakan nilai  $k$  sebanyak 5 skenario, yaitu 10, 30, 50, 80, dan 100 dengan menghasilkan akurasi tertinggi berada pada nilai  $k=10$  [20]. Penelitian yang lain juga dilakukan dalam kasus pengelompokan musik untuk menentukan suasana hati dengan menggunakan nilai  $k$  sebanyak 16 skenario atau dilakukan pengujian terhadap nilai  $k$  dari 1 sampai 16 dengan menghasilkan akurasi tertinggi berada pada nilai  $k=3$  [21].

Berdasarkan hasil uraian penelitian tersebut, maka dapat dibuktikan bahwa nilai  $k$  sangat berpengaruh terhadap akurasi. Berbagai penelitian lain dapat dilihat pada Tabel 1.

**Tabel 1. Perbandingan Algoritma *K-Nearest Neighbor* dan Aplikasinya**

Penulis	Tahun	Jurnal	Metode	Kasus
(Savera, Suryawan and Setiawan,)[22]	2020	Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)	KNN	Deteksi dini kanker kulit
(Candra and Rozana) [23]	2020	IT Journal Research and Development (ITJRD)	KNN	Klasifikasi komentar <i>bullying</i>
(Satrian) [24]	2020	Jurnal Mahasiswa Aplikasi Teknologi Komputer dan Informasi Vol. 2 No. 1 Thn 2020, Hal 19-23	KNN	Klasifikasi Gamers Usia Muda
(Alghifari and Wibowo) [25]	2019	Jurnal Teknologi & Manajemen Informatika - Vol.5 No.1 2019	KNN	Klasifikasi Kinerja Satpam
(Anggraini and Fadillah) [26]	2019	InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan	KNN	Analisis Deteksi Emosi Manusia Dari Suara Percakapan
(Informa, Indonusa and Issn,) [27]	2018	Jurnal INFORMA Politeknik Indonusa Surakarta ISSN : 2442-7942 Vol. 4 Nomor 2 Tahun 2018	KNN	Memprediksi Varietas Padi untuk lahan pertanian
(YULIANTI and NURDIN) [28]	2018	Jurnal Teknoif Vol. 6 No. 1 April 2018	KNN	Sistem Pendukung Keputusan Penerima Bantuan BSM
(Yudha, Muflikhah and Wihandika) [29]	2018	Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer Vol. 2 No. 2, Februari 2018, hlm. 897-904	Neighbor Weighted KNN-	Klasifikasi Risiko Hipertensi
(Rivki and Bachtiar) [16]	2017	Jurnal Sistem Informasi (Journal of Information Systems). 1/13 (2017), 31-37	KNN	Klasifikasi Follower Twiter yang menggunakan Bahasa Indonesia

### c. K-Fold Cross Validation

Seperti yang sudah dijelaskan di atas mengenai algoritma. Ada salah satu teknik untuk mengevaluasi model prediksi, yaitu *k-fold cross validation*. Berdasarkan pada penelitian-penelitian sebelumnya pada kasus klasifikasi dalam penentuan status gunung berapi menggunakan teknik tersebut, dengan menggunakan *k-fold cross validation* adalah 3 skenario *fold*. Pada percobaan pertama memperoleh akurasi sebesar 78,26%, percobaan ke dua memperoleh akurasi sebesar 82, 61% dan untuk percobaan ke tiga memperoleh akurasi sebesar 78,26%, sehingga didapatkan rata-rata akurasi dari ke 3 *fold* tersebut adalah sebesar

79,71% [13].

Penelitian yang lain dilakukan oleh Ni'mah, Sutojo and Setiadi [14] pada kasus identifikasi tanaman obat herbal, mereka menggunakan 9 skenario *fold cross validation*. Pada percobaan *fold* ke-1 dan ke-4 menghasilkan akurasi sebesar 70%, sedangkan sisanya menghasilkan akurasi yang lebih besar. Akurasi yang tertinggi berada pada *fold* ke-2,5,6,7,8 [14]. Sehingga, berdasarkan uraian ke dua peneliti tersebut, dapat membuktikan bahwa penggunaan *k-fold cross validation* sangat penting dalam mengevaluasi hasil model prediksi dalam K-NN, karena dapat mengurangi waktu komputasi dengan tetap menjaga keakuratan estimasi model dan memvalidasi kinerja model secara lebih akurat.

#### d. Klasifikasi Multi-Kelas

Beberapa tahun terakhir ini algoritma K-NN telah banyak diaplikasikan pada kasus klasifikasi. Klasifikasi tersebut merupakan proses pemberian kelas atau label pada sebuah data baru yang diprediksi oleh algoritma K-NN. Dalam penelitian yang dilakukan oleh Sukanto, dkk [18] dengan kasus prediksi kelompok UKT mahasiswa. Mereka menggunakan sebanyak 6 kelas untuk diprediksi, yaitu UKT1, UKT2, UKT3, UKT4, UKT5, dan UKT 6. Sehingga dalam hal ini proses klasifikasi tersebut dinamakan *multiclass classification*, karena menggunakan lebih dari 2 kelas yang akan diprediksinya.

Penelitian yang lain dilakukan oleh Kartika, dkk [30] dalam kasus penentuan siswa berprestasi, mereka menggunakan 2 kelas prediksi, yaitu kelas berprestasi dan tidak berprestasi. Sehingga, dalam hal ini proses klasifikasi tersebut dinamakan *binary classification*, karena hanya menggunakan dua kelas dalam prediksinya. Selain itu, Putra et al., [31] dalam penelitiannya untuk mendiagnosis gejala dini penyakit stroke. Dalam penelitiannya tersebut, mereka menggunakan *multi-class classification* dengan 3 kelas prediksi, yaitu kelas rendah, sedang dan tinggi. Maka, berdasarkan uraian tersebut secara keseluruhan banyak peneliti yang menerapkan konsep *multi-class classification* pada penelitiannya.

#### e. Pemetaan UMKM (Usaha Mikro Kecil dan Menengah)

UMKM mempunyai peranan yang penting dalam pertumbuhan ekonomi dan industri di golongan menengah ke bawah dan juga suatu negara [32], [33]. Peran strategis UMKM sangat diharapkan oleh pemerintah sebagai penggerak ekonomi yang nyata. Karena jumlah UMKM yang banyak tidak bisa dipungkiri akan melahirkan kompetisi di antara mereka, selain itu memungkinkan terjadinya kompetisi terbuka dengan perusahaan besar di bidang yang sama [32]. Dengan adanya hal tersebut penting sekali pengelompokan UMKM agar dapat membantu pemerintah terkait penetapan strategi pemasaran yang tepat sebagai prioritas utama untuk mengembangkan pasar [34].

Penelitian serupa pernah dilakukan oleh beberapa peneliti yang dapat di lihat pada Tabel 2 dengan hasil uraian penelitian yaitu proses *cluster* atau pengelompokan sangat penting bagi UMKM karena dengan adanya *cluster* UMKM dapat menetapkan strategi pemasaran yang tepat dari hasil pemetaan UMKM sebagai prioritas utama untuk mengembangkan pasar.

**Tabel 2. Perbandingan Pemetaan UMKM dan Aplikasinya**

Penulis	Tahun	Jurnal	Metode	Kasus
(Puntoriza and Fibriani) [34]	2020	Journal of Information System Vol. 5, No. 1, Mei 2020: 86-94	K-Means	Analisis Pesebaran UMKM Di Kota Malang

(Widiastuti and Tamrin) [35]	2020	Jurnal SIMETRIS, Vol. 11 No. 1 April 2020	Location Based Service	Persebaran UMKM Di Kabupaten Jepara
(Widiastuti and Azizah Widiastuti) [36]	2018	Jurnal Sistem Informasi Bisnis (JSINBIS)	K-Means	Pemetaan UMKM Di Kabupaten Jepara
(Napitupulu) [32]	2017	Jurnal Penelitian Pos dan Informatika	K-Means	Klasifikasi UMKM Berdasarkan Tingkat Kesiapan Teknologi
(Irhamni, Khotimah and Sundarwati) [33]	2015	Prosiding Seminar Nasional Manajemen Teknologi XXIII	Ant-KMeans	Cluster IKM

## f. Analisis Sentimen

Dalam beberapa tahun terakhir penelitian mengenai analisis sentimen sangat populer dilakukan. Analisis sentimen telah diakui sebagai salah satu domain penelitian pada bidang *Machine Learning* [37], [38], dalam penelitiannya mengembangkan sistem klasifikasi menggunakan sentimen analisis berdasarkan postingan pada media sosial *twitter*. Dalam hal ini, mereka mengklasifikasikan *tweet* sebagai kelas *ekstremis* atau *non-ekstremis*.

Penelitian lain yang dilakukan oleh Li, Chao and Zhang [39], terkait klasifikasi emosi pada manusia dengan menggunakan dua kelas prediksi, yaitu bahagia dan sedih. Selain Sentimen et al., [4] juga melakukan penelitian yang sama menggunakan data ulasan mengenai “*Sea Games 2019*” yang berasal dari postingan media sosial, yaitu *twitter* dengan menggunakan perbandingan algoritma KNN dan *Support Vector Machine (SVM)*. Sehingga berdasarkan uraian para peneliti tersebut, dapat dibuktikan bahwa penelitian mengenai analisis sentimen sangat populer, khususnya untuk postingan pada sosial media *twitter*.

Penelitian terkait lainnya disebutkan pada Tabel 3 dapat disimpulkan bahwa Metode KNN sangat cocok di gunakan dalam klasifikasi dalam analisis sentiment karena memiliki akurasi yang cukup besar dengan akurasi 80%. Selain itu metode KNN juga memiliki kelebihan mengklasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Data pembelajaran diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data dan KNN memiliki prinsip yang sangat sederhana karena sangat mudah dalam proses penerapannya dalam berbagai kasus salah satunya dalam analisis sentimen.

## g. Peluang dan Tantangan Masa Depan

Pada saat ini, aliran data besar atau *Big Data* telah terjadi diberbagai sektor, salah satunya adalah sektor bisnis [45], [46]. UMKM di Kota Bogor, sebagai salah satu sektor bisnis, mengalami perkembangan yang signifikan, khususnya pada sektor kuliner. Hal ini berpengaruh terhadap meningkatnya data dan informasi yang dihasilkan oleh UMKM tersebut. Aliran data besar terjadi karena beberapa aplikasi atau *website* menghasilkan data terstruktur, tidak terstruktur atau semi-terstruktur yang sangat besar dengan kecepatan tinggi. Hal tersebut akan mempengaruhi pengambilan keputusan di sektor bisnis, khususnya pada UMKM sektor kuliner di Kota Bogor. Pengetahuan terkait *insight* bisnis dapat diperoleh dari data yang bersumber dari data konsumen terkait ulasan atau komentar produk. Di sinilah *Big Data* memainkan perannya. Pada bagian ini akan diuraikan tantangan dan peluang dalam penggunaan *Big Data*, khususnya pada penerapan metode *Data Mining* pada UMKM sektor kuliner di Kota Bogor.

**Tabel 3. Perbandingan Analisis Sentimen dan Tingkat Akurasinya**

Penulis	Tahun	Jurnal	Metode	Kasus	Akurasi
---------	-------	--------	--------	-------	---------

*Analisis Big Data sentimen konsumen UMKM sektor kuliner menggunakan Multi-Label K-Nearest Neighbor*

(Herdiawan) [40]	2018	Jurnal Ilmiah Komputer dan Informatika (KOMPUTA)	KNN	Analisis Terhadap Indihome Berdasarkan Publik	Sentimen Telkom Opini	80%
(Wulandari, Indriati, and Dewi) [41]	2019	Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer Vol. 3, No. 5, Mei 2019, hlm. 5017-5023	BM25 dan KNN	Analisis Ulasan Berbahasa Indonesia	Sentimen pada “Lazada”	89%
(Deviyanto and Wahyudi) [42]	2018	JISKa (Jurnal Informatika Sunan Kalijaga), Vol. 3, No. 1, MEI, 2018, Pp. 1–13 ISSN : 2527-5836 (print)   2528-0074 (online)	KNN	Penerapan Sentimen Pengguna Twitter	Analisis Pada	67,2 %
(Romadloni, Santoso and Budilaksono) [43]	2019	<i>Jurnal IKRA-ITH Informatika Vol 3 No 2 Juli 2019 ISSN 2580-4316</i>	Metode Naive Bayes, KNN dan Decision Tree	Analisis Transportasi Commuter Line	Sentimen Krl	80% (Naïve Bayes)), 80% (KNN), 100%(Decision Tree)
(Fitrianti, Kurniawati and Agustien) [44]	2019	Seminar Nasional Aplikasi Teknologi Informasi (SNATi) 2019 Yogyakarta	KNN	Analisis Review Dengan Teks Bahasa Indonesia	Sentimen Restoran Bahasa	96,61 %

#### h. Tantangan

Penggunaan *Big Data* pada UMKM sektor kuliner, memiliki tantangan yang cukup besar, yaitu keamanan (*privacy*) data, ketersediaan data, dan penambangan data atau *Data Mining*. Masalah yang pertama yaitu keamanan data, terkait dengan pencurian dan penyalah-gunaan informasi pribadi pengguna. Masalah yang kedua adalah ketersediaan data. Data merupakan salah satu kunci untuk melakukan analisis *Big Data* dengan teknik *Data Mining*. Akses terhadap data, seringkali menjadi hambatan. Masalah yang lain adalah proses penambangan data itu sendiri atau *Data Mining*. *Data Mining* akan memberikan solusi prediktif atau deksriptif yang efisien dan terbaik untuk analisis *Big Data*. Namun, jika dilihat dengan perkembangan atau jumlah data yang dihasilkan pada aplikasi atau *website* UMKM serta aplikasi Android [47], sehingga itu menjadi salah satu hambatan pada proses *Data Mining*. Hambatan yang terjadi adalah eksplorasi data dan ekstraksi informasi yang dilakukan akan cukup sulit dengan kompleksnya data dan informasi.

#### i. Peluang

Penggunaan metode *Data Mining* dengan algoritma *K-Nearest Neighbor* telah banyak diaplikasikan diberbagai kasus penelitian (dapat dilihat pada Tabel 1). Algoritma KNN sangat banyak dimanfaatkan dalam kasus *multi-class classification* dan *binary classification*. Selain itu KNN juga dimanfaatkan dalam kasus *multi-label classification*. Namun penerapan pada kasus *multi-label classification* masih jarang dilakukan sehingga menjadi sebuah peluang untuk penelitian. Potensi penggunaan *Big Data* dengan metode *Data Mining* menggunakan algoritma KNN pada UMKM sektor kuliner di Kota Bogor sangat besar, khususnya terkait dengan informasi ulasan atau komentar konsumen terhadap produk UMKM sektor kuliner di Kota Bogor.

## KESIMPULAN



Kesimpulan dari studi literatur ini pertama adalah bahwa algoritma KNN dapat diterapkan pada berbagai kasus seperti pada *multi-class classification*, *binary classification*, dan *multi-label classification* dalam memprediksi berbagai kasus di kehidupan sehari-hari. Kedua, bahwa penggunaan *Big Data* pada UMKM sektor kuliner sangat penting dilakukan, karena jumlah data UMKM sektor kuliner yang semakin meningkat di Kota Bogor. Kesimpulan tersebut, memunculkan ide atau gagasan baru dalam berinovasi. Dengan demikian kami mengusulkan sebuah ide baru, yaitu analisis Big Data sentimen konsumen UMKM sektor kuliner menggunakan *Multi-Label K-Nearest Neighbor* di Kota Bogor. Ide atau gagasan tersebut dapat bermanfaat bagi pelaku UMKM sektor kuliner di Kota Bogor dalam mengetahui *insight* atau wawasan dari hasil analisis sentimen konsumen terhadap UMKM. Sehingga hal ini dapat membantu untuk pengambilan sebuah keputusan bisnis dalam meningkatkan daya saingnya terhadap kompetitor atau UMKM sektor kuliner yang berada di Kota Bogor.

## DAFTAR PUSTAKA

- [1] A. Andreyestha and A. Subekti, "Analisa Sentiment Pada Ulasan Film Dengan Optimasi Ensemble Learning," *J. Inform.*, vol. 7, no. 1, pp. 15–23, 2020, doi: 10.31311/ji.v7i1.6171.
- [2] A. K. Maheswari, Ph.D., *Business Intelligence and Data Mining.*, Business E. LLC, 2015.
- [3] R. Primartha, *Belajar Machine Learning Teori dan Praktik*, Edisi ke-1. Bandung, 2018.
- [4] A. Sentimen, A. Games, S. E. A. Games, M. Riefky, and W. Pramesti, "Sentiment Analysis of Southeast Asian Games ( SEA Games ) in Philippines 2019 Based on Opinion of Internet User of Social Media Twitter with K-Nearest Neighbor and Support Vector Machine Filipina Tahun 2019 berdasarkan Opini Netizen dari Media Sosial Twit," vol. 17, no. 1, pp. 26–41, 2020, doi: 10.20956/jmsk.v.
- [5] F. C. Anggian, N. Hidayat, and M. T. Furqon, "Implementasi Metode Modified K-Nearest Neighbor untuk Klasifikasi Status Gunung Berapi," vol. 3, no. 12, pp. 11027–11033, 2019.
- [6] D. Satriawan, H. Fitriyah, and A. S. Budi, "Sistem Klasifikasi Tahu Putih Murni dan Tahu Putih Mengandung Formalin Menggunakan Metode K-Nearest Neighbor," vol. 3, no. 10, 2019.
- [7] K. A. Sugiarta, I. Cholissodin, and E. Santoso, "Optimasi K-Nearest Neighbor Menggunakan Bat Algorithm Untuk Klasifikasi Penyakit Ginjal Kronis," vol. 3, no. 10, pp. 10301–10308, 2019.
- [8] D. Syahid, J. Jumadi, and D. Nursantika, "Sistem Klasifikasi Jenis Tanaman Hias Daun Philodendron Menggunakan Metode K-Nearest Neighbor (KNN) Berdasarkan Nilai Hue, Saturation, Value (HSV)," *J. Online Inform.*, vol. 1, no. 1, p. 20, 2016, doi: 10.15575/join.v1i1.6.
- [9] R. N. Devita, H. W. Herwanto, and A. P. Wibawa, "Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa Indonesia," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, p. 427, 2018, doi: 10.25126/jtiik.201854773.
- [10] H. Leidiyana, "Penerapan Metode K-Nearest Neighbor Pada Penentuan Grade Dealer," *J. Pengetah. Dan Teknol. Komput.*, vol. 2, no. 2, pp. 108–112, 2017.
- [11] A. Nugroho, "Implementasi Algoritma K-Nearest Neighbor Dalam Memprediksi Potensi Calon Kreditur di KSP Galih Manunggal," *Dasi*, vol. 17, no. 2, pp. 1–6, 2016.
- [12] P. N. Fadila, . I., and D. E. Ratnawati, "Identifikasi Jenis Attention Deficit Hyperactivity

- Disorder (ADHD) Pada Anak Usia Dini Menggunakan Metode Neighbor Weighted K-Nearest Neighbor (NWKNN)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, p. 193, 2016, doi: 10.25126/jtiik.201633195.
- [13] F. Tempola, M. Muhammad, and A. Khairan, "Perbandingan Klasifikasi Antara Knn Dan Naive Bayes Pada Penentuan Status Gunung Berapi Dengan K-Fold Cross Validation Comparison of Classification Between Knn and Naive Bayes At the Determination of the Volcanic Status With K-Fold Cross," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 5, pp. 577–584, 2018, doi: 10.25126/jtiik20185983.
- [14] F. S. Ni'mah, T. Sutojo, and D. R. I. M. Setiadi, "Identifikasi Tumbuhan Obat Herbal Berdasarkan Citra Daun Menggunakan Algoritma Gray Level Co-occurrence Matrix dan K-Nearest Neighbor," *J. Teknol. dan Sist. Komput.*, vol. 6, no. 2, pp. 51–56, 2018, doi: 10.14710/jtsiskom.6.2.2018.51-56.
- [15] G. A. Pradnyana, I. K. A. Suryantara, and I. G. M. Darmawiguna, "Impression Classification of Endek (Balinese Fabric) Image Using K-Nearest Neighbors Method," *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, vol. 3, no. 3, pp. 213–220, 2018, doi: 10.22219/kinetik.v3i3.611.
- [16] M. Rivki and A. M. Bachtiar, "Implementasi Algoritma K-Nearest Neighbor Dalam Pengklasifikasian Follower Twitter Yang Menggunakan Bahasa Indonesia," *J. Sist. Inf.*, vol. 13, no. 1, p. 31, 2017, doi: 10.21609/jsi.v13i1.500.
- [17] C. Gong, Z. Su, P. Wang, and Q. Wang, "Cumulative belief peaks evidential K-nearest neighbor clustering," *Knowledge-Based Syst.*, vol. 200, no. May, p. 105982, 2020, doi: 10.1016/j.knosys.2020.105982.
- [18] S. Sukamto, Y. Adriyani, and R. Aulia, "Prediksi Kelompok UKT Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *JUITA J. Inform.*, vol. 8, no. 1, p. 121, 2020, doi: 10.30595/juita.v8i1.6267.
- [19] S. L. Br. Ginting, Z. Wendi, and D. Astrid, "Teknik Data Mining untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighborhood," *J. Tek. Komput. Unikom – Komputika – Vol. 3, No.2 - 2014 Sist.*, vol. 3, no. 2, pp. 23–28, 2014, [Online]. Available: <http://komputika.tk.unikom.ac.id/jurnal/teknik-data-mining-untuk.10>.
- [20] C. S. D. Prasetya, "Sistem Rekomendasi Pada E-Commerce Menggunakan K-Nearest Neighbor," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 3, p. 194, 2017, doi: 10.25126/jtiik.201743392.
- [21] G. Harsemadi, M. Sudarma, and N. Pramaita, "Implementasi Algoritma K-Nearest Neighbor pada Perangkat Lunak Pengelompokan Musik untuk Menentukan Suasana Hati," *Maj. Ilm. Teknol. Elektro*, vol. 16, no. 1, pp. 14–20, 2017, doi: 10.24843/mite.1601.03.
- [22] T. R. Savera, W. H. Suryawan, and A. W. Setiawan, "Deteksi Dini Kanker Kulit Menggunakan K-Nn Dan Early Detection of Skin Cancer Using K-Nn and Convolutional," vol. 7, no. 2, pp. 373–378, 2020, doi: 10.25126/jtiik.202072602.
- [23] R. M. Candra and A. N. Rozana, "Klasifikasi Komentar Bullying pada Instagram Menggunakan Metode K-Nearest Neighbor," vol. 5, no. 1, pp. 45–52, 2020.
- [24] B. Satrian, "Penerapan Algoritma K-Nn untuk Klasifikasi Gamers Usia Sekolah," vol. 2, no. 1, pp. 19–23, 2020.
- [25] M. R. Alghifari and A. P. Wibowo, "Penerapan Metode K-Nearest Neighbor Untuk Klasifikasi Kinerja Satpam Berbasis Web," *J. Teknol. dan Manaj. Inform.*, vol. 5, no. 1, 2019, doi: 10.26905/jtmi.v5i1.3074.
- [26] N. A. Anggraini and N. Fadillah, "Analisis Deteksi Emosi Manusia dari Suara Percakapan Menggunakan Matlab dengan Metode KNN," *InfoTekJar (Jurnal Nas.*

- Inform. dan Teknol. Jaringan*), vol. 3, no. 2, pp. 176–179, 2019, doi: 10.30743/infotekjar.v3i2.1041.
- [27] J. Informa, P. Indonusa, and S. Issn, “Implementasi Metode K-Nearest Neighbor (Knn) Untuk Memprediksi Varietas Padi Yang Cocok Untuk Lahan Pertanian,” vol. 4, pp. 2–8, 2018.
- [28] E. YULIANTI and Y. A. NURDIN, “Sistem Pendukung Keputusan Penerimaan Bantuan Siswa Miskin (BSM) Berbasis Online Dengan Metode Knn (K-Nearest Neighbor) (Studi kasus : SMPN 1 Koto XI Tarusan),” *J. Teknoif*, vol. 6, no. 1, pp. 12–17, 2018, doi: 10.21063/jtif.2018.v6.1.12-17.
- [29] B. L. Yudha, L. Muflikhah, and R. C. Wihandika, “Klasifikasi Risiko Hipertensi Menggunakan Metode Neighbor Weighted K- Nearest Neighbor ( NWKNN ),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 2, pp. 897–904, 2018.
- [30] J. I. Kartika, E. Santoso, and Sutrisno, “Penentuan Siswa Berprestasi Menggunakan Metode K-Nearest Neighbor dan Weighted Product (Studi Kasus: SMP Negeri 3 Mejayan),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 5, pp. 352–360, 2017.
- [31] D. A. Putra, M. D. S. Sanapiah, A. I. Hanifah, and T. Afirianto, “SEED (Stoke Disease Early Detection Application) - Rancang Bangun Aplikasi Mobile Berbasis Android untuk Mendiagnosis Gejala Dini Penyakit Stroke Menggunakan K-Nearest Neighbor (K-NN),” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 3, p. 287, 2019, doi: 10.25126/jtiik.201963915.
- [32] D. B. Napitupulu, “Clustering SMEs Based on Technology Readiness using K-Means Algorithm,” *J. Penelit. Pos dan Inform.*, vol. 7, no. 2, p. 97, 2017, doi: 10.17933/jppi.2017.070202.
- [33] F. Irhamni, B. K. Khotimah, and T. Sundarwati, “Clustering Industri Kecil Dan Menengah ( Ikm ) Dengan Menggunakan Metode Gabungan,” pp. 1–9, 2015.
- [34] P. Puntoriza and C. Fibriani, “Analisis Persebaran UMKM Kota Malang Menggunakan Cluster K-means,” *JOINS (Journal Inf. Syst.*, vol. 5, no. 1, pp. 86–94, 2020, doi: 10.33633/joins.v5i1.3469.
- [35] N. A. Widiastuti and T. Tamrin, “Penerapan Aplikasi Mobile Location Based Service Untuk Persebaran Usaha Mikro Kecil Menengah Dikabupaten Jepara,” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 11, no. 1, pp. 271–278, 2020, doi: 10.24176/simet.v11i1.4015.
- [36] N. A. Widiastuti and N. A. Azizah Widiastuti, “Teknologi Geolocation Berbasis Android dengan Metode K-Means untuk Pemetaan UMKM di Kabupaten Jepara,” *J. Sist. Inf. Bisnis*, vol. 8, no. 2, p. 218, 2018, doi: 10.21456/vol8iss2pp218-224.
- [37] M. Ghosh and G. Sanyal, “An ensemble approach to stabilize the features for multi-domain sentiment analysis using supervised machine learning,” *J. Big Data*, vol. 5, no. 1, 2018, doi: 10.1186/s40537-018-0152-5.
- [38] S. Ahmad, M. Z. Asghar, F. M. Alotaibi, and I. Awan, “Detection and classification of social media-based extremist affiliations using sentiment analysis techniques,” *Human-centric Comput. Inf. Sci.*, vol. 9, no. 1, 2019, doi: 10.1186/s13673-019-0185-6.
- [39] T. M. Li, H. C. Chao, and J. Zhang, “Emotion classification based on brain wave: a survey,” *Human-centric Comput. Inf. Sci.*, vol. 9, no. 1, 2019, doi: 10.1186/s13673-019-0201-x.
- [40] Herdiawan, “BERDASARKAN OPINI PUBLIK MENGGUNAKAN METODE IMPROVED K-NEAREST NEIGHBOR. Teknik Informatika – Universitas Komputer Indonesia Jurnal Ilmiah Komputer dan Informatika ( KOMPUTA ),” 2018.
- [41] D. Wulandari, Indriati, and C. Dewi, “Analisis Sentimen p ada Ulasan ‘ Lazada ’

- Berbahasa Indonesia Menggunakan BM25 dan K-Nearest Neighbor ( K-NN ) dengan Perbaikan,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 5, pp. 5017–5023, 2019.
- [42] A. Deviyanto and M. D. R. Wahyudi, “Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 3, no. 1, p. 1, 2018, doi: 10.14421/jiska.2018.31-01.
- [43] N. T. Romadloni, I. Santoso, and S. Budilaksono, “Perbandingan Metode Naive Bayes , Knn Dan Decision Tree Terhadap Analisis Sentimen Transportasi Krl,” *J. IKRA-ITH Inform.*, vol. 3, no. 2, pp. 1–9, 2019.
- [44] R. P. Fitrianti, A. Kurniawati, and D. Agusten, “Implementasi Algoritma K - Nearest Neighbor Terhadap Analisis Sentimen Review Restoran Dengan Teks Bahasa Indonesia,” *Semin. Nas. Apl. Teknol. Inf. 2019*, pp. 27–32, 2019.
- [45] T. Kolajo, O. Daramola, and A. Adebisi, “Big data stream analysis: a systematic literature review,” *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0210-7.
- [46] M. Marjani *et al.*, “Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges,” *IEEE Access*, vol. 5, pp. 5247–5261, 2017, doi: 10.1109/ACCESS.2017.2689040.
- [47] S. H. Al Ikhsan, F. Fatimah, R. S. Irawan. Aplikasi Android Sebaran Lokasi UMKM di Kota Bogor Dengan Formula Haversine. *Jurnal Krea-TIF*. Vol 7 No 2, pp 88-102. 2019