



Penerapan Algoritma K-Nearest Neighbor untuk Klasifikasi Usaha Masyarakat Berdasarkan Jenis Izin Usaha

Brema Daniel Ginting¹, Yusfrizal², Lina Arliana Nur Kadim³
^{1,2,3} STMIK Kaputama, Indonesia

JL. Veteran, No.4A-9A, Binjai, Sumatera Utara, Indonesia

Email : bremadanielg@gmail.com¹, lina_arliana@yahoo.com²,
yusfrizal80@gmail.com³

Abstract. Business legality is the identity of a business that legalizes a business so that it is recognized by the community. Business legality must be valid according to applicable laws and regulations so that the business can be protected by various documents that are valid in the eyes of the law. One of the supporting factors for the sustainability of a business is influenced by the existence of legal elements of the business being run. Business permits that must be owned by the community are a business establishment deed, business entity NPWP, trade business license (SIUP), company domicile certificate (SKDP) and business registration number (NIB). The increase in community businesses in Sei Bingai District, Langkat Regency has triggered many business permits that are not directly supervised by the local government. Community business permits are important documents in supervising the running of these community businesses. The types of businesses in Sei Bingai District also vary, such as tourism, C mining, trade, factories and so on.

Keywords: Business License, Data Mining, K-Nearest Neighbor Algorithm)

Abstrak. Legalitas usaha merupakan jati diri sebuah usaha yang melegalkan suatu usaha sehingga diakui oleh masyarakat. Legalitas usaha harus sah menurut undang-undang dan hukum yang berlaku sehingga usaha tersebut dapat terlindungi dengan berbagai dokumen yang sah di mata hukum. Salah satu faktor pendukung keberlangsungan suatu usaha dipengaruhi oleh adanya unsur legalitas dari usaha yang di jalani. Izin usaha yang harus dimiliki masyarakat adalah akta pendirian usaha, NPWP badan usaha, surat izin usaha perdagangan (SIUP), surat keterangan domisili perusahaan (SKDP) dan nomor induk berusaha (NIB). Meningkatnya usaha masyarakat di Kecamatan Sei Bingai Kabupaten Langkat memicu banyak izin usaha yang tidak terawasi langsung oleh pemerintah daerah. Surat izin usaha masyarakat merupakan dokumen penting dalam mengawasi berjalannya usaha masyarakat tersebut. jenis usaha di daerah Kecamatan Sei Bingai juga beragam jenisnya seperti pariwisata, galian C, perdagangan, pabrik dan lain sebagainya.

Kata kunci: Izin Usaha, Data mining, Algoritma K-Nearest Neighbor

1. LATAR BELAKANG

Legalitas usaha merupakan jati diri sebuah usaha yang melegalkan suatu usaha sehingga diakui oleh masyarakat. Legalitas usaha harus sah menurut undang-undang dan hukum yang berlaku sehingga usaha tersebut dapat terlindungi dengan berbagai dokumen yang sah di mata hukum. Salah satu faktor pendukung keberlangsungan suatu usaha dipengaruhi oleh adanya unsur legalitas dari usaha yang di jalani. Izin usaha yang harus dimiliki masyarakat adalah akta pendirian usaha, NPWP badan usaha, surat izin usaha perdagangan (SIUP), surat keterangan domisili perusahaan (SKDP) dan nomor induk berusaha (NIB).

Banyaknya ragam usaha masyarakat di Kecamatan Sei Bingai juga akan jenis surat izin usaha beragam juga, sehingga perlu dilakukan klasifikasi surat izin usaha

berdasarkan jenis usaha masyarakat tersebut. Dengan izin usaha tersebut pemerintah daerah dapat mengawasi usaha masyarakat secara langsung serta memastikan bahwa semua jenis usaha mematuhi standar dan regulasi yang berlaku, termasuk aspek kesehatan, keselamatan, dan lingkungan.

Klasifikasi usaha masyarakat merujuk pada pengelompokan atau pengkategorian jenis-jenis usaha yang dilakukan oleh masyarakat berdasarkan karakteristik atau aspek tertentu. Tujuan dari klasifikasi ini adalah untuk memahami dan mengorganisir beragam kegiatan ekonomi yang dilakukan oleh masyarakat dalam suatu wilayah atau sektor tertentu. Dalam konteks penelitian yang mencakup penerapan algoritma *K-Nearest Neighbors* (KNN) untuk klasifikasi usaha masyarakat berdasarkan jenis izin usaha, klasifikasi tersebut dapat mencakup beberapa kriteria seperti, jenis usaha, skala usaha, lokasi geografis dan jenis izin usaha.

Beberapa penelitian telah dilakukan klafikasi dengan algoritma K-NN. Salah satunya adalah jurnal Statistika Aktuaria 2023, ISSN: 2599-2546 “Penerapan Algoritma Klasifikasi K-Nearest Neighbor pada Penyakit Diabetes”. Pada penelitian ini digunakan dataset yang didapat dari Kaggle. Dalam penerapan KNN akan dicari nilai K (jumlah tetangga terdekat) terbaik. Pada penelitian ini didapat nilai 11 merupakan nilai K terbaik dalam rentang 1 hingga 15 dengan nilai akurasi sebesar 76,56%. Sehingga nilai 11 merupakan nilai K yang dapat digunakan dalam penerapan algoritma KNN pada penyakit diabetes. Dengan K = 11 didapati nilai *Accuracy* Sebesar 0,765625, *Precision* sebesar 0,850299401, *Recall* sebesar 0,802259887, *F-1 Score* sebesar 0,825581395 (Azizah et al. 2023)

2. KAJIAN TEORITIS

Dalam adanya permasalahan yang dihadapi, maka perumusan masalah adalah sebagai berikut :

1. Bagaimana cara menyelesaikan klasifikasi usaha masyarakat berdasarkan jenis izin usaha di Kecamatan Sei Bingai menggunakan metode *K-Nearest Neighbor*?
2. Berapa hasil perbandingan akurasi dan presisi dari Algoritma *K-Nearest Neighbor* untuk klasifikasi usaha masyarakat berdasarkan jenis izin usaha?
3. Bagaimana hasil pembuatan sistem dapat membantu Kantor Camat dalam klasifikasi usaha masyarakat berdasarkan jenis izin usaha di Kecamatan Sei Bingai?

Adapun tujuan dari penelitian Skripsi ini adalah sebagai berikut:

1. Untuk merancang dan mengimplementasikan sebuah sistem klasifikasi usaha masyarakat berdasarkan jenis izin usaha dengan metode *K-Nearest Neighbor* (KNN).
2. Untuk mengetahui hasil penerapan klasifikasi usaha masyarakat berdasarkan jenis izin usaha dan hasil perbandingan akurasi dan presisi dari Algoritma K-Nearest Neighbor.
3. Untuk mempelajari dan mengetahui tentang metode *K-Nearest Neighbor* (KNN) dalam klasifikasi

3. METODE PENELITIAN

Data Mining

Menurut Arhami & Nasir (2020, h.1) data *mining* merupakan proses penggalian informasi dan pola yang bermanfaat dari data yang sangat besar. Data *mining* mencakup pengumpulan data, ekstraksi data, analisis data, dan statistik data. Data *mining* juga dikenal sebagai *Knowledge discovery*, *Knowledge extraction*, *data/pattern analysis*, *information harvesting*, dan lain-lain. Dalam menganalisa data salah satu solusi yang dapat digunakan adalah data mining, Data mining akan menjadi patokan atau acuan untuk pengambilan keputusan (Kurniawan et al., 2019).

Menurut Erwin (2017, h.1) data *mining* merupakan salah satu bagian atau proses utama dari *Knowledge Discovery in Database* (KDD) yang bentuk kegiatannya yaitu mengumpulkan dan menggunakan data masa lalu untuk menemukan keteraturan, pola atau hubungan dalam suatu set data yang lebih besar. Secara garis besar KDD meliputi tiga tahapan, yaitu *pre processing*, *process* (data *mining*) dan *post processing*. Kesimpulannya, data *mining* adalah proses penemuan pola-pola tertentu dari sebuah data atau basis data yang berukuran besar untuk memperoleh informasi yang sangat berguna. Jenis atau teknik data *mining* yang diterapkan yaitu *association*, *classification* dan *Clustering*. *Association* mencari pola hubungan yang terdapat pada data atau basis data, *classification* dan *Clustering* lebih cenderung menemukan pola-pola untuk pengelompokan.

K-Nearest Neighbor

K-Nearest Neighbor merupakan sebuah metode untuk melakukan klasifikasi terhadap objek yang baru berdasarkan (K) tetangga terdekatnya. K-NN termasuk algoritman *Supervised Learning* yang mana hasil dari query instance baru, diklasifikasikan berdasarkan mayoritas dari kategori pada KNN. Kelas paling banyak muncul yang akan menjadi kelas hasil klasifikasi (prasetyo, 2014).

Algoritma ini hanya melakukan penyimpanan vektor-vektor fitur dan klasifikasi dari data pembelajaran. Pada fase klasifikasi, fitur-fitur yang sama dihitung untuk data test (yang klasifikasinya tidak diketahui). Jarak dari vektor yang baru ini terhadap seluruh vektor data dihitung, dan sejumlah k buah yang paling dekat diambil. Titik yang baru klasifikasinya diprediksikan termasuk pada klasifikasi terbanyak dari titik-titik tersebut.

Nilai k yang terbaik untuk algoritma ini tergantung pada data. Secara umum, nilai k yang tinggi akan mengurangi efek noise pada klasifikasi, tetapi membuat batasan antara setiap klasifikasi menjadi lebih kabur. Nilai k yang bagus dapat dipilih dengan optimasi parameter, misalnya dengan menggunakan cross-validation. Kasus khusus di mana klasifikasi diprediksikan berdasarkan data pembelajaran yang paling dekat (dengan kata lain, $k = 1$) disebut algoritma k-nearest neighbor. Ketepatan algoritma k-NN ini sangat dipengaruhi oleh ada atau tidaknya fitur-fitur yang tidak relevan, atau jika bobot fitur tersebut tidak setara dengan relevansinya terhadap klasifikasi. Riset terhadap algoritma ini sebagian besar membahas bagaimana memilih dan memberi bobot terhadap fitur agar performa klasifikasi menjadi lebih baik.

K-Nearest Neighbor metode yang bersifat *supervised*, di mana hasil dari *query instance* yang baru diklasifikasikan berdasarkan mayoritas kategori pada KNN. KNN adalah sebuah metode klasifikasi terhadap sekumpulan data berdasarkan pembelajaran data yang sudah terklasifikasikan sebelumnya. Termasuk dalam supervised learning, di mana hasil *query instance* yang baru diklasifikasikan berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam KNN (Melisa et al., 2022).

Sesuai dengan prinsip kerja *K-Nearest Neighbor* yaitu mencari jarak terdekat antara data yang akan dievaluasi dengan k tetangga(neighbor) terdekatnya dalam data pelatihan. Persamaan dibawah ini menunjukkan rumus perhitungan untuk mencari jarak terdekat dengan d adalah jarak dan p adalah dimensi data (Agusta, 2007):

$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2}$$

Dengan keterangan :

x_i : Sampel data

x_i : Data uji

i : Variabel data

d : Jarak

p : Dimensi data

4. HASIL DAN PEMBAHASAN

Data Pendukung Penelitian

Untuk menentukan group dari suatu objek, pertama yang harus dilakukan adalah mengukur jarak *Euclidean Distance* sebagai berikut.

Tabel 1 Transformasi Data

No	Objek	K1	K2	K3	K4	Hasil
1	UD. H	1	2	1	5	1
2	UD. HRP	1	2	1	11	1
3	UT. M	7	1	5	4	2
4	KWJ	5	1	3	12	1
5	UT. TS	7	2	5	13	2
6	DAS	4	2	3	11	2
7	PS	4	1	7	1	2
8	UD. LS	1	2	1	2	1
9	JS	4	1	5	3	2
10	SR	2	2	1	4	1
11	RK	1	3	1	1	1
12	TL	1	2	1	14	1
13	TS	7	3	5	15	2
14	AB	5	1	3	16	2
15	AT	4	3	5	14?

Data Kriteria

Pengkodean data adalah suatu teknik yang dilakukan untuk memberikan penegasan pada proses yang terlibat (data pensinyalan) transmisi data. Dalam proses tersebut perlu diperhatikan. Berikut ini merupakan kode pendukung penelitian ini:

Tabel 2 Jenis Usaha (K1)

Kode	Jenis Usaha
1	Perdagangan
2	Jasa
3	Produksi
Kode	Jenis Usaha
4	Pertambangan
5	Parwisata
6	Teknologi
7	Agribisnis
8	Keuangan

Tabel 3 Skala Usaha (K2)

Kode	Skala Usaha
1	Skala Usaha Mikro
2	Skala Usaha Kecil
3	Skala Usaha Menengah
4	Usaha Besar

Tabel 4 Izin Usaha (K3)

Kode	Izin Usaha
1	Surat izin Usaha Perdagangan (SIUP)
2	Izin Usaha Industri (IUI)
3	Tanda Daftar Usaha Parwisata (TDUP)
4	Nomor Induk Berusaha (NIB)
5	Izin Lokasi
6	Izin Mendirikan Bangunan (IMB)
7	Izin Gangguan (HO)

Tabel 5 Lokasi Usaha (K4)

Kode	Lokasi Usaha
1	Balinten
Kode	Lokasi Usaha
2	Durian Lingga
3	Emplasmen Kwala Mencirim
4	Gunung Ambat
5	Mekar Jaya
6	Namu Ukur Selatan
7	Nama Ukur Utara
8	Pasar IV Namu Terasi
9	Pasar VI Kwala Mencirim
10	Pasar VIII Namu Terasi
11	Pekan Sawah
12	Purwobinangun
13	Rumah Galuh
14	Simpang Kuta Buluh
15	Tanjung Gunung
16	Telagah

Tabel 6 Data Hasil

Keterangan	Kode
Legal	1

Illegal	2
---------	---

Perhitungan Metode K-Nearest Neighbor

Langkah 1 : Menentukan Nilai Parameter K. K=3

Langkah 2 : Menghitung jarak data uji dengan semua data training, menggunakan

Euclidean Distance.

$$d_{x_{\text{train } i}, x_{\text{test, } j}} = \sqrt{\sum_{i,j=1}^n (x_{\text{train } i} - x_{\text{test, } j})^2}$$

Euclidean Distance PX No. 15 : 4, 3, 5, 14

Data 1

$$dis = \sqrt{(4 - 1)^2 + (3 - 2)^2 + (5 - 1)^2 + (14 - 5)^2} = 36,13$$

Data 2

$$ddis = \sqrt{(4 - 1)^2 + (3 - 2)^2 + (5 - 1)^2 + (14 - 11)^2} = 12,40$$

Data 3

$$dis = \sqrt{(4 - 7)^2 + (3 - 1)^2 + (5 - 5)^2 + (14 - 4)^2} = 3,60$$

Data 4

$$dis = \sqrt{(4 - 5)^2 + (3 - 1)^2 + (5 - 3)^2 + (14 - 12)^2} = 4,58$$

Data 5

$$dis = \sqrt{(4 - 7)^2 + (3 - 2)^2 + (5 - 5)^2 + (14 - 13)^2} = 3,16$$

Data 6

$$dis = \sqrt{(4 - 4)^2 + (3 - 2)^2 + (5 - 3)^2 + (14 - 11)^2} = 6,08$$

Data 7

$$dis = \sqrt{(4 - 4)^2 + (3 - 1)^2 + (5 - 7)^2 + (14 - 1)^2} = 26,07$$

Data 8

$$dis = \sqrt{(4 - 1)^2 + (3 - 2)^2 + (5 - 1)^2 + (14 - 2)^2} = 48,10$$

Data 9

$$dis = \sqrt{(4 - 4)^2 + (3 - 1)^2 + (5 - 5)^2 + (14 - 3)^2} = 2$$

Data 10

$$dis = \sqrt{(4 - 2)^2 + (3 - 2)^2 + (5 - 1)^2 + (14 - 4)^2} = 40,06$$

Data 11

$$dis = \sqrt{(4 - 1)^2 + (3 - 3)^2 + (5 - 1)^2 + (14 - 1)^2} = 52,08$$

Data 12

$$dis = \sqrt{(4 - 1)^2 + (3 - 2)^2 + (5 - 1)^2 + (14 - 14)^2} = 3,16$$

Data 13

$$dis = \sqrt{(4 - 7)^2 + (3 - 3)^2 + (5 - 5)^2 + (14 - 15)^2} = 3$$

Data 14

$$dis = \sqrt{(4 - 5)^2 + (3 - 1)^2 + (5 - 3)^2 + (14 - 16)^2} = 4,58$$

Tabel 7 Pengurutan Jarak Terdekat Data Baru Dengan Data Training

No	Objek	K1	K2	K3	K4	Euclidean Distance	Urutan Jarak
1	UD. H	1	2	1	5	36,14	11
2	UD. HRP	1	2	1	11	12,41	9
3	UT. M	7	1	5	4	3,61	5
4	KWJ	5	1	3	12	4,58	6
5	UT. TS	7	2	5	13	3,16	3
6	DAS	4	2	3	11	6,08	8
7	PS	4	1	7	1	26,08	10
8	UD. LS	1	2	1	2	48,10	13
9	JS	4	1	5	3	2,00	1
10	SR	2	2	1	4	40,06	12
11	RK	1	3	1	1	52,09	14
12	TL	1	2	1	14	3,16	4
13	TS	7	3	5	15	3,00	2
14	AB	5	1	3	16	4,58	7

Langkah 3 : Menentukan kategori dari tetangga terdekat.

Dari data dengan kategori mayoritas terdapat kategori Legal dan Illegal. Dari jumlah mayoritas (**Legal > Illegal**) tersebut dapat disimpulkan data baru (data No. 15) (K1=4, K2=3, K3=5, K4=14) termasuk kedalam kategori Illegal.

5. KESIMPULAN DAN SARAN

Kesimpulan

Dengan adanya hasil penerapan algoritma *K-Nearest Neighbor* untuk klasifikasi usaha masyarakat berdasarkan jenis izin usaha, maka penulis dapat mengambil beberapa kesimpulan yaitu sebagai berikut :

1. Dengan dibangunnya sistem ini, pihak kecamatan dapat memudahkan klasifikasi masyarakat berdasarkan jenis izin usaha pada Kecamatan Sei Bingai.

2. Dengan menerapkan data mining penerapan algoritma *K-Nearest Neighbor* untuk klasifikasi usaha masyarakat berdasarkan jenis izin usaha, Algoritma KNN terbukti efektif dalam mengklasifikasikan jenis izin usaha berdasarkan data yang telah diinput. Hasil klasifikasi menunjukkan bahwa metode ini dapat digunakan untuk mengelompokkan usaha dengan akurasi yang memadai, meskipun perlu dilakukan penyesuaian dan evaluasi lebih lanjut untuk mengoptimalkan hasilnya.
3. Hasil analisis menunjukkan bahwa sistem dapat memproses dan menganalisis data dengan baik, menghasilkan klasifikasi yang relevan dan berguna. Namun, hasil ini juga menekankan pentingnya evaluasi berkelanjutan dan pengujian lebih lanjut untuk memastikan bahwa metode KNN memenuhi kebutuhan spesifik dari klasifikasi izin usaha.

Saran

Dari hasil penerapan algoritma *K-Nearest Neighbor* untuk klasifikasi usaha masyarakat berdasarkan jenis izin usaha saran antara lain :

1. Klasifikasi usaha masyarakat berdasarkan jenis izin usaha yang dibuat dapat digunakan di Kecamatan Sei Bingai.
2. Diharapkan pada interface yang dihasilkan, dapat lebih berkembang lagi dengan menampilkan hasil perhitungan untuk mengetahui aturan asosiasi yang akan dibangun.
3. Diharapkan tulisan selanjutnya dapat menambahkan variabel-variabel lain yang berhubungan dengan tulisan ini sesuai dengan data yang ada.

6. DAFTAR REFERENSI

- Arhami, & Nasir. (2020). *Algoritma dan implementasi*. Tangerang: Arhami & Nasir.
- Azizah, N., Firdaus, M. R., Suyaningsih, R., & Indrayatna, F. (2023). Penerapan algoritma klasifikasi K-nearest neighbor. *Jurnal*.
- Baharuddin, M. M., Hasanuddin, T., & Azis, H. (2019). Analisis performa metode K-nearest neighbor untuk identifikasi jenis kaca. *Ilk. J. Ilm.*, 11(28), 269–274.
- Choli, S. R., Handayani, T., Prathivi, R., & Ardianita, T. (2020). Implementasi algoritma klasifikasi K-nearest neighbor (KNN) untuk klasifikasi seleksi penerima beasiswa. *Indonesian Journal on Computer and Information Technology*.
- Erwin. (2017). Data and software engineering (ICoDSE). In *International Conference on Data and Software Engineering* (pp. 1-6).

- Firman, C. E. (2017). Penentuan pola yang sering muncul untuk penjualan pupuk menggunakan algoritma FP-growth. *Jurnal Informatika*, 9(2), 1–8.
- Ida, S. H. B., Faisal, M., Ramadhan, N., & Darniati. (2023). Implementasi algoritma K-nearest neighbor terhadap penentuan risiko kredit usaha mikro kecil dan menengah. *Jurnal Manajemen Informatika dan Komunikasi*.
- Kurniawan, H., Agustin, F., Yusfrizal, & Umami, K. (2018). Implementation of data mining in prediction of sales chip with rough set method. *Jurnal Computer Science, Universitas Potensi Utama*.
- Kusrini, & Emha, T. L. (2009). *Algoritma data mining*. Yogyakarta: Andi Offset.
- Lestari, M. (2014). Penerapan algoritma klasifikasi nearest neighbor (K-NN) untuk mendeteksi penyakit jantung. *Fakt. Tepat*, 7(September), 366–371.
- Melisa, S., Fauzi, A., & Yusfrizal. (2022). Penerapan machine learning untuk klasifikasi tingkat kematangan buah anggur (Vitis) dengan metode K-nearest neighbor. *Jurnal Sistem Informasi, STMIK Kaputama Binjai*.
- Novianti, A. G., & Prasetyo, D. (2017). Penerapan algoritma K-nearest neighbor (KNN) untuk prediksi waktu kelulusan mahasiswa. In *SEMNASSTIKOM* (pp. 108–113).
- Situmorang, A., Yusfrizal, & Ambarita, I. (2023). Pengelompokan data menggunakan algoritma K-means untuk meningkatkan motivasi belajar siswa MTS Alwasliyah 48 Kebun Lada. *Jurnal Sistem Informasi, STMIK Kaputama Binjai*.
- Syahputra, A., Ervina, & Melisa. (2022). Pengaruh modal usaha, lokasi pemasaran dan kualitas produk terhadap pendapatan UMKM. *Journal of Management and Business (JOMB)*.