

**PENERAPAN ALGORITMA *K-MEANS* UNTUK PENGELOMPOKAN
PEMBANGUNAN JALAN PADA DINAS PEKERJAAN UMUM DAN
PENATAAN RUANG**

***IMPLEMENTATION K-MEANS ALGORITHM FOR CLUSTERING OF
ROAD CONSTRUCTION AT THE DEPARTMENT OF PUBLIC WORKS
AND SPATIAL PLANNING***

PROPOSAL SKRIPSI

Dibuat sebagai syarat pelaksanaan penelitian mahasiswa program sarjana

Oleh :

Hari Ilham Nur Akbar

NIM : 1806129



**JURUSAN ILMU KOMPUTER
PROGRAM STUDI TEKNIK INFORMATIKA
INSTITUT TEKNOLOGI GARUT**

2022

LEMBAR PENGESAHAN

PENERAPAN ALGORITMA *K-MEANS* UNTUK PENGELOMPOKAN PEMBANGUNAN JALAN PADA DINAS PEKERJAAN UMUM DAN PENATAAN RUANG

IMPLEMENTATION K-MEANS ALGORITHM FOR CLUSTERING OF ROAD CONSTRUCTION AT THE DEPARTMENT OF PUBLIC WORKS AND SPATIAL PLANNING


PROPOSAL SKRIPSI

Oleh :

Hari Ilham Nur Akbar

NIM : 1806129

Proposal Skripsi ini telah diperiksa dan disetujui oleh:

Acc 2/4-2022 Pembimbing I


Dr. Dede Kurniadi., S.Kom., M.Kom.

NIDN: 0402098301

Pembimbing II



Yoga Handoko Agustin, S.Kom., M.Kom.

NIDN: 0424089002

RINGKASAN

Jalan sebagai media yang mendukung bidang ekonomi, sosial dan budaya. Jalan yang layak dan baik pembangunannya sebagai penghubung antar wilayah guna menaikkan perekonomian bergerak lebih baik. Dinas (PUPR) Pekerjaan Umum dan Penataan Ruang sebagai penyedia jasa untuk melaksanakan pembangunan proyek konstruksi jalan, panduan pelaksanaan, monitoring serta peninjauan pelaksanaan kegiatan terhadap ketentuan-ketentuan dan persyaratan yang ditetapkan. Pembangunan jalan di daerah memiliki panjang dan nominal Pembangunan yang beragam berdasarkan kebutuhannya. Dalam musyawarah rencana pembangunan desa pada PUPR, desa mengajukan pembangunan jalan dengan panjang dan nominal harga yang berbeda, namun pengelompokan pembangunan jalan di lakukan dengan pencocokan jumlah dan nominal dengan proses yang masih manual. Hal ini menyebabkan kurang efektif baik dalam segi mencapai tujuan pembangunan jalan maupun dalam segi waktu rekap menentukan kelompok pembangunan jalannya. Penelitian ini bertujuan untuk membuat pengelompokan pembangunan Jalan Pada Dinas PUPR serta hasil berupa mendapatkan kelompok pembangunan jalan yang diajukan oleh desa sesuai dengan karakteristik dan proses yang lebih efektif sehingga mengetahui prioritas pembangunannya. Metode yang digunakan dalam penelitian ini menggunakan Metode *Cross-Industry Standard Process For Data Mining* (CRISP-DM) dengan menggunakan perhitungan algoritma *K-Means* dengan alur dalam algoritma *K-Means* dengan mengelompokan data berdasarkan jarak antara data terhadap titik *centroid* klaster yang didapatkan melalui proses berulang. Analisis perlu menentukan jumlah K sebagai input algoritma dan evaluasi menggunakan *Davies-Bouldin Indeks* (DBI). Adapun data yang digunakan merupakan data pembangunan jalan di Dinas PUPR kabupaten Garut. Hasil dari penelitian mendapatkan model pengelompokan pembangunan jalan yang diajukan oleh desa yang lebih efektif dan sebagai bahan pengambil kebijakan pengelompokan pembangunan jalan di Dinas PUPR karena mengetahui prioritas pembangunan jalan berdasarkan kelompoknya.

KATA PENGANTAR

Assalamu'alaikum Warrahmatullahi Wabarakatuh

Alhamdulillahirobbil'alamiin, Segala puji beserta syukur penulis panjatkan kehadirat Allah SWT karena atas Rahmat, Karunia serta Hidayah-Nya penulis dapat menyelesaikan Proposal Skripsi serta dapat menyelesaikan tepat pada waktunya tanpa ada halangan yang berarti. Shalawat serta salam tak lupa penulis curahkan kepada Nabi Muhammad SAW.

Dalam penyusunan Proposal Skripsi ini penulis menyampaikan rasa terima kasih dan penghargaan yang setinggi-tingginya kepada semua pihak yang telah memberikan dorongan, bimbingan, dan penghargaan selama ini, antara lain kepada yang terhormat :

1. Bapak Dr. Hilmi Aulawi, S.T., M.T., selaku Rektor Institut Teknologi Garut;
2. Bapak Dr. Dede Kurniadi, S.Kom., M.Kom., selaku selaku Dosen Pembimbing I yang telah memberikan arahan serta masukan sampai dapat menyelesaikan Proposal Skripsi ini dan juga selaku Ketua Jurusan Ilmu Komputer Institut Teknologi Garut;
3. Bapak Ridwan Setiawan, S.T., M.Kom., selaku Ketua Program Studi Teknik Informatika Institut Teknologi Garut;
4. Yoga Handoko Agustin, S.Kom., M.Kom., selaku Dosen Pembimbing II yang telah memberikan arahan serta masukan sampai dapat menyelesaikan Proposal Skripsi; dan
5. Seluruh dosen pengajar dan staf Institut Teknologi Garut.

Dalam penyusunan proposal ini, penulis menyadari bahwa masih banyak kekurangan. Kritik dan saran yang membangun sangatlah diperlukan. Akhir kata, penulis berharap propal ini dapat berguna bagi pembaca dan bagi penulis.

Garut,
Penulis

DAFTAR ISI

RINGKASAN	i
KATA PENGANTAR	ii
DAFTAR ISI.....	iii
DAFTAR GAMBAR	iv
DAFTAR TABEL.....	v
DAFTAR LAMPIRAN.....	vi
BAB I PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	5
1.3. Tujuan Penelitian.....	5
1.4. Batasan Masalah	5
1.5. Manfaat Penelitian	6
1.6. Keaslian Penelitian	6
BAB II TINJAUAN PUSTAKA.....	8
2.1. Penelitian Terkait.....	8
2.3. Definisi Jalan	14
2.3. <i>Data Mining</i>	15
2.4. Pengelompokan (<i>Clustering</i>).....	16
2.5. Algoritma <i>K-Means</i>	17
2.6. <i>Davies-Bouldin Indeks</i> (DBI)	19
2.7. <i>Cross-Industry Standard Process for Data mining</i> (CRISP-DM).....	20
2.8. <i>RapidMiner</i>	22
BAB III METODE PENELITIAN.....	24
3.1. Kerangka Penelitian.....	24
3.2. <i>Diagram Activity</i>	25
3.3. Sumber Daya Penelitian	26
3.4. Sumber Data	27
JADWAL RENCANA KEGIATAN PENELITIAN	29
DAFTAR PUSTAKA	30

DAFTAR GAMBAR

Gambar 1.1. Data Pembangunan Jalan Di Dinas PUPR.....	2
Gambar 2.1. Ilustrasi Proses <i>Clustering</i>	16
Gambar 2.2. Enam Siklus Hidup Pengembangan Data Mining	21
Gambar 2.3. Tampilan Lembar Kerja <i>RapidMiner</i>	23
Gambar 3.1. Kerangka Penelitian	24
Gambar 3.2. <i>Diagram Activity</i>	25

DAFTAR TABEL

Tabel 2.1. Tabel Kesenjangan Penelitian	11
Tabel 3.1. Sumber Daya Penelitian	27
Tabel 3.2. Sampel Data Pembangunan Jalan Oleh Dinas PUPR.....	28
Tabel Jadwal Rencana Kegiatan Penelitian	29

DAFTAR LAMPIRAN

Lampiran A : Presensi Bimbingan.....	32
Lampiran B : Surat Bukti Penelitian.....	33

BAB I

PENDAHULUAN

1.1. Latar Belakang

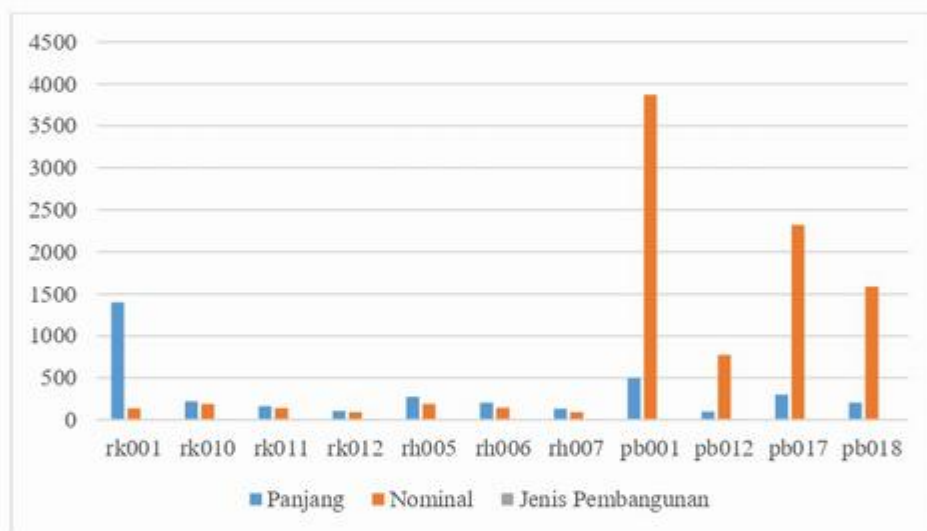
Jalan sebagai bagian sistem transportasi nasional yang memiliki peran penting dalam mendukung bidang ekonomi, sosial dan budaya. Tersedianya jalan yang layak, menjadi penghubung antar wilayah guna akhirnya membantu menaikkan perekonomian bergerak lebih baik (Yuono, 2017).

Berdasarkan Undang-undang No. 38 Tahun 2004 Tentang Jalan, pemerintah pada dasarnya bertujuan mewujudkan perkembangan antar daerah yang seimbang untuk tercapainya hasil penanganan jalan yang memberikan pelayanan optimal. Dan salah satu program Pemerintah untuk bertujuan mewujudkan perkembangan antar daerah dengan pembangunan jalan yaitu oleh Dinas (PUPR) Pekerjaan Umum dan Penataan Ruang sebagai penyedia jasa untuk melaksanakan pembangunan proyek konstruksi jalan dan digunakan sebagai panduan pelaksanaan, monitoring serta peninjauan terhadap pelaksanaan kegiatan terhadap ketentuan-ketentuan dan persyaratan yang telah ditetapkan dalam perencanaan program (Jatnika et al., 2020)

PUPR No. 21/PRT/M/2019 Tentang Manajemen Keselamatan Konstruksi, Seringkali Rencana Mutu Pelaksanaan Konstruksi (RMPK) yang direncanakan oleh penyedia jasa belum sepenuhnya sesuai dengan pekerjaan sehingga adanya perbedaan hasil proyek. Maka dari itu perlunya penyesuaian rencana dan standarisasi yang telah dikeluarkan oleh PUPR untuk mengupayakan standar sesuai dengan aturan dari PUPR (Jatnika et al., 2020).

Pembangunan jalan yang dilaksanakan di daerah sebagai bagian sistem transportasi dengan kualitas pembangunan yang beragam berdasarkan pertimbangan penggunaan jalan untuk pejalan kaki, kendaraan bermotor atau bermobil. Berdasarkan hasil observasi data dari Dinas PUPR kabupaten Garut, setiap tahunnya desa mengajukan pembangunan jalan dengan panjang dan nominal yang berbeda-beda dalam musyawarah rencana pembangunan desa pada PUPR, namun setelah pembangunan selesai dan dilakukan survei nyatanya masih menerapkan anggaran pada pelaksanaan dengan ketidak sesuaian, dan penentuan

prioritas pembangunan jalan masih dilakukan dengan pencocokan panjang dan nominal oleh pegawai PUPR dengan melihat pengajuan panjang, nominal dan keterangan kegiatan pembangunan, dimana pembangunan pendek dengan nominal yang tinggi ditentukan menjadi kelompok pembangunan yang besar, dan pengajuan dengan ukuran yang panjang dengan nominal yang kecil ditentukan menjadi kelompok pembangunan yang sedang atau kecil tanpa perhitungan sebelumnya. Hal ini dapat menyebabkan proses pengelompokan pembangunan jalan pada pembangunan dirasa kurang efektif baik dalam segi keabsahan, mencapai tujuan maupun dalam segi waktu rekap pengelompokan. Berikut merupakan data pembangunan jalan oleh Dinas PUPR yang ditunjukkan pada Gambar 1.1.



Gambar 1.1. Data Pembangunan Jalan Di Dinas PUPR

(Sumber : (PUPR, 2021))

Dimana dapat dilihat bahwasanya desa dengan keterangan yang dibuat menjadi kode-kode mengajukan pembangunan jalan dengan panjang, nominal juga jenis pembangunan yang berbeda, seperti jenis kegiatan yaitu pembangunan jalan 500 meter dengan nominal uang sekitar 3 miliar 800 juta. Maka dari itu dengan permasalahan tersebut dibutuhkan suatu algoritma yang dapat mempermudah untuk pengelompokan pembangunan jalan sehingga mempermudah menentukan kelompok pembangunan jalan dalam pengajuan pembangunan jalan dari desa.

Berdasarkan penelitian yang dilakukan oleh Listiani et al. (2019) penelitian ini menerapkan algoritma *K-Means Clustering* untuk rekomendasi pekerjaan berdasarkan pengelompokan data penduduk dengan nilai akurasi mencapai 79 persen. Penelitian kedua yang dilakukan oleh Vernanda et al. (2021) penelitian ini membahas pemetaan daerah rawan kecelakaan lalu lintas menghasilkan 3 *Cluster*, adapun 3 *Cluster* tersebut yaitu Sangat Rawan(SR), Rawan(R), dan Cukup Rawan (CR) dengan algoritma *K-Means* untuk *Clustering* dimana pada aturan lift ratio, jika hasil akhir bernilai lebih dari 1 maka aturan tersebut kuat. Penelitian ke tiga yang dilakukan oleh Nabila et al. (2021) penelitian ini menghasilkan pengelompokan yang berbeda pada Kasus Covid-19 Di Provinsi Lampung dengan *Algoritma K-Means* dengan hasil dari kedua perhitungan per akhir bulan sama-sama mendekati 0 yang berarti klaster yang dievaluasi menghasilkan klaster yang baik. Penelitian ke empat dilakukan oleh Mulyani et al. (2018) penelitian ini melakukan pengelompokan rekomendasi bimbingan belajar berdasarkan segmentasi akademik siswa dengan hasil 2 kelompok siswa berdasarkan nilai rata-rata raport dari semester 1-5 untuk bahan pertimbangan pemilihan mata pelajaran pada pengayaan dalam rekomendasi bimbingan belajar berdasarkan segmentasi akademik siswa menggunakan algoritma *K-Means* untuk *Clustering* Dan *Fp-growth* untuk Asosiasi. Penelitian kelima yang dilakukan oleh Kurniadi et al. (2018) bertujuan untuk menganalisis, memprediksi, dan mengklasifikasikan mahasiswa yang berpotensi mendapatkan beasiswa di perguruan tinggi dengan menggunakan algoritma *K-Nearest Neighbor* dengan hasil akurasi 95,83 persen, penelitian ini memprediksi peluang terbesar untuk mendapatkan beasiswa PPA (Peningkatan Prestasi Akademik) yang dimiliki siswa. Penelitian keenam yang dilakukan oleh Puspitasari et al. (2017) bertujuan untuk Mengelompokkan data jalan yang rusak kedalam empat kategori prioritas perbaikan jalan. Untuk memberikan rekomendasi DBMP dalam menentukan prioritas perbaikan jalan di kota Samarinda. Melakukan pengelompokkan data ke dalam beberapa kelompok. Apabila kemiripan data yang satu dengan yang lain yang sangat besar dengan data lain, maka data berada pada cluster yang sama. Penelitian ketujuh yang dilakukan oleh Asmiatun & Wakhidah (2018) mengolah data dengan menggunakan algoritma

k-means dengan tujuan untuk mengelompokkan ke beberapa kecamatan tujuan dari pengelompokkan untuk mengetahui kecamatan mana yang paling rawan rusak. Hasilnya dari 7 kecamatan terdapat satu kecamatan yang paling rawan jalan rusak yaitu kecamatan Gayamsari. Penelitian kedelapan yang dilakukan oleh Asmiatun et al. (2020) bertujuan untuk memfokuskan penerapan metode untuk pengolahan data dengan mengelompokkan data-data yang didapat dari sensor accelerometer menggunakan metode *K-medoids*. Dari data yang digunakan jumlah 638 data telah berhasil dikelompokkan menjadi 4 cluster yang dapat menunjukkan tingkat frekuensi jalan kondisi baik, sedang, rusak ringan dan rusak berat.

Banyak algoritma yang bisa digunakan seperti *K-Nearest Neighbor* untuk sistem prediksi, *Fp-growth* untuk mengasosiasikan data atau dalam pengelompokan data menggunakan *K-Means*. Berdasarkan rujukan penelitian, tujuan dari penelitian ini untuk pengelompokan pembangunan jalan Pada Dinas PUPR sehingga menggunakan algoritma *K-Means* menjadi alternatif pemilihan dikarenakan berdasarkan penelitian yang dilakukan (Asmiatun & Wakhidah, 2018) algoritma *K-Means* cocok digunakan untuk pengelompokan yang menghasilkan klaster untuk pengelompokan kondisi permukaan jalan, dengan ini pun dapat di terapkan pada kasus pengelompokan pembangunan jalan. Serta penelitian ini menggunakan atribut kuantitas, harga juga jenis kegiatan pembangunan, dan menghasilkan pengelompokan pembangunan jalan yang diajukan oleh desa sesuai dengan kriteria juga dengan proses yang lebih efektif. Penelitian dilakukan dengan pengclusteran nilai atribut, kemudian dibuat kelompok pembangunan jalan kemudian evaluasi *Clustering* dengan (DBI) *Davies-Bouldin Index* metode evaluasi internal yang mengukur evaluasi cluster pada suatu metode pengelompokan yang didasarkan pada nilai kohesi dan separasi, serta *deployment* berupa visualisasi pengetahuan algoritma pengelompokan pembangunan jalan yang diajukan oleh desa bagi Dinas PUPR yang mampu memberikan gambaran yang baik untuk implementasikan dan dikembangkan dalam sebuah aplikasi.

1.2. Rumusan Masalah

Berdasarkan permasalahan pada latar belakang dapat diambil rumusan masalah yang akan menjadi pembahasan penelitian yaitu :

1. Bagaimana penerapan algoritma *K-Means* dalam pengelompokan pembangunan jalan di Dinas PUPR?
2. Bagaimana karakteristik pembangunan jalan di Dinas PUPR berdasarkan hasil klasterisasi ?

1.3. Tujuan Penelitian

Berdasarkan masalah tersebut, tujuan dari penelitian ini adalah;

1. Untuk mendapatkan model klasterisasi yang menghasilkan pengetahuan yang dapat membantu pengelompokan pembangunan jalan di Dinas PUPR.
2. Untuk mengetahui karakteristik pembangunan jalan berdasarkan klasterisasi untuk membantu pengelompokan prioritas pembangunan jalan selanjutnya di Dinas PUPR.

1.4. Batasan Masalah

Batasan masalah pada permasalahan diatas, diantaranya :

1. Data yang digunakan dalam klastering pembangunan jalan ini adalah data di Dinas PUPR tahun 2015 sd 2021 dengan jumlah 263 *record*;
2. Data atau atribut yang akan di klaster yaitu data atribut kuantitas, harga juga jenis kegiatan pembangunan jalan;
3. Pengelompokan diawali dari 3 kelompok berdasarkan diskusi dengan Dinas PUPR, yang apabila menghasilkan nilai DBI yang buruk maka diujicoba dengan membuat jumlah kelompok yang selanjutnya dengan ketentuan nilai DBI yang baik berdasarkan hasil dan diskusi peneliti dengan tempat studi kasus yaitu Dinas PUPR;
4. Algoritma yang digunakan dalam penelitian ini adalah *K-Means*;
5. Metode yang digunakan dalam penelitian ini adalah CRISP-DM;
6. Pada tahap evaluasi menggunakan *Davies-Bouldin Indeks* (DBI);dan
7. Visualisasi pengetahuan hasil dari perhitungan algoritma *K-Means*.

1.5. Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini meliputi manfaat teoritis dan manfaat praktis. Manfaat praktis merupakan manfaat langsung yang dirasakan oleh praktisi pengguna hasil penelitian. Sedangkan manfaat teoritis merupakan manfaat yang didapatkan oleh pembaca penelitian. Adapun manfaat tersebut dijabarkan sebagai berikut :

1. Manfaat Teoritis

- a. Bagi mahasiswa diharapkan dapat mengetahui suatu gambaran bagaimana proses pengelompokan kelas pembangunan jalan yang lebih efektif; dan
- b. Bagi PUPR diharapkan mendapatkan pemahaman pengelompokan kelas pembangunan jalan yang diajukan oleh desa dapat lebih efektif dengan penerapan data mining.

2. Manfaat Praktis

Bagi PUPR diharapkan mendapatkan model pengelompokan pembangunan jalan yang diajukan oleh desa yang lebih efektif dan sebagai bahan pengambil kebijakan pengelompokan pembangunan jalan di Dinas PUPR sehingga mengetahui prioritas pembangunan jalan pada berdasarkan kelompoknya.

1.6. Keaslian Penelitian

Berdasarkan rujukan pada latar belakang dan observasi sebelumnya, skripsi dengan topik Penerapan Algoritma *K-Means* Untuk Pengelompokan Pembangunan Jalan Pada Dinas PUPR belum pernah dilakukan pada penelitian sebelumnya. Adapun keaslian pada penelitian ini dapat teridentifikasi sebagai berikut:

1. Lokasi studi kasus di Dinas PUPR dengan kantor yang bertempat di Jl. Raya Samarang No.117, Sukagalih, Kec. Tarogong Kidul, Garut-Jawa Barat (44151);
2. Penggunaan algoritma *K-Means* untuk pengelompokan pada penelitian sebelumnya diterapkan pada kasus rekomendasi pekerjaan, kasus covid-19, bimbingan belajar, daerah rawan kecelakaan dan peluang mendapatkan

beasiswa sementara penelitian ini memfokuskan pengklasteran pada pengelompokan untuk pembangunan jalan di Dinas PUPR Garut;

3. Evaluasi *Clustering* menggunakan metode *Davies-Bouldin Indeks* (DBI);
4. Menggunakan 3 atribut yaitu atribut kuantitas, harga juga jenis kegiatan pembangunan dari data pengajuan pembangunan jalan oleh desa pada dinas PUPR pada tahun 2015 sd 2021 dengan 263 *record*;
5. Menerapkan pendekatan metodologi CRISP-DM untuk melaksanakan tahapan penelitian; dan
6. Menghasilkan visualisasi pengetahuan hasil perhitungan algoritma pengelompokan pembangunan jalan yang diajukan oleh desa bagi Dinas PUPR.

BAB II

TINJAUAN PUSTAKA

2.1. Penelitian Terkait

Penelitian yang dilakukan oleh Listiani et al. (2019) menerapkan algoritma *K-Means* untuk pengelompokan data penduduk untuk rekomendasi jenis pekerjaan berdasarkan faktor pendidikan, umur dan jenis kelamin. Pada penelitian ini, dataset 651 *record* tahun kelahiran 1969 sampai 1999 dan lulusan SLTA/ sederajat, terdapat 17 atribut yaitu No.KK, NIK, nama, jenis kelamin, tempat lahir, tanggal dan tahun lahir, golongan darah, agama, status, status dalam keluarga, pendidikan terakhir, pekerjaan, nama orang tua, alamat, Rt dan Rw. Menghasilkan 4 kelompok jenis pekerjaan berdasarkan jenis kelamin, pendidikan, umur dan jenis pekerjaan dengan nilai akurasi yang diperoleh yaitu sebesar 79% dan diklasifikasikan sebagai fair atau cukup pada klasifikasi nilai AUC.

Penelitian kedua yang dilakukan oleh Vernanda et al. (2021) dengan tujuan memberikan informasi kepada masyarakat mengenai titik-titik daerah yang rawan kecelakaan agar masyarakat dapat lebih waspada dalam berkendara, untuk menjaga diri sendiri dan orang lain. Selain itu membantu pihak Satlantas dalam mengevaluasi wilayah-wilayah yang memiliki kerawanan laka lantas yang tinggi dengan algoritma *K-Means*. Sehingga wilayah tersebut segera di tangani dengan cepat dan dapat menekan angka kecelakaan lalu lintas di wilayah Kota Malang. Dengan dataset 223 kejadian dari tahun 2019-2020 dengan dibagi menjadi 3 klaster Sangat Rawan(SR), Rawan(R), dan Cukup Rawan (CR).

Penelitian ketiga yang dilakukan oleh Nabila et al. (2021) dengan Menganalisis data kasus Covid-19 agar dapat mengetahui pengelompokan pada masalah tersebut di Provinsi Lampung menggunakan metode *Clustering* dengan *algoritma K-Means*. Dengan dataset sampel data kasus Covid-19 sebanyak 15 Kabupaten/Kota di Provinsi Lampung. Dengan 6 atribut Kabupaten/Kota, suspek, probable, konfirmasi positif, selesai isolasi, dan kematian yang digunakan dalam proses perhitungan dan membagi data ke dalam 4 *Cluster* yang dikategorikan sebagai zona merah, zona orange, zona kuning dan zona hijau. Menghasilkan pengelompokan yang berbeda

dikarenakan jumlah pada atribut suspek, probable, konfirmasi positif, selesai isolasi, dan kematian pada setiap Kabupaten/Kota yang tidak sama. Berdasarkan hasil pengujian mencari nilai *Davies-Bouldin Index* (DBI) menggunakan perhitungan secara manual dan menggunakan bantuan *tools RapidMiner* memiliki perbedaan. Perhitungan manual memiliki hasil yang lebih baik jika dibandingkan dengan bantuan *tools RapidMiner*, tetapi hasil dari kedua perhitungan per akhir bulan samasama mendekati 0 yang berarti klaster yang dievaluasi menghasilkan klaster yang baik.

Penelitian keempat yang dilakukan oleh Mulyani et al. (2018) penelitian yang dapat dimanfaatkan oleh sekolah untuk mengatur jadwal mata pelajaran pengayaan sesuai dengan kebutuhan kelompok siswa tersebut. Dengan dataset 190 Jumlah *Cluster* ada $k=2$ dan maksimal itemset adalah 3 itemset. Menghasilkan 2 kelompok siswa berdasarkan nilai rata-rata raport dari semester 1-5 untuk bahan pertimbangan pemilihan mata pelajaran pada pengayaan dengan kelompok 1 pola yang dihasilkan adalah Inggris -> Kimia dan kelompok 2 pola yang dihasilkan adalah Kimia -> Inggris dan Biologi.

Penelitian kelima yang dilakukan oleh Kurniadi et al. (2018) mengimplementasikan model algoritma *K-Nearest Neighbor* (K-NN) dalam menganalisis, memprediksi, dan mengklasifikasikan siswa dengan potensi mendapatkan beasiswa. Dengan jumlah dataset untuk pengujian 1018 data mahasiswa dan 24 calon penerima beasiswa. Atribut yang digunakan pada proses prediksi yaitu semester, IPK, jumlah tanggungan dan penghasilan orang tua. Perhitungan jarak nilai dari atribut pengujian pada setiap atribut pelatihan menggunakan persamaan *Euclidean Distance* dan untuk uji nilai akurasi model dihitung menggunakan *Confusion Matrix*. Hasil dari simulasi model prediksi memperlihatkan bahwa faktor penentu data pelatihan dari jumlah dan variasi nilai yang berbeda dapat meningkatkan kinerja algoritma K-NN dengan tingkat akurasi terbaik 95,83 persen dalam memprediksi siswa yang memiliki peluang terbesar untuk mendapatkan beasiswa.

Penelitian keenam yang dilakukan oleh Puspitasari et al. (2017) mengimplementasikan model algoritma Metode *Fuzzy C-Means* dalam

Mengelompokkan data jalan yang rusak kedalam empat kategori prioritas perbaikan jalan. Untuk memberikan rekomendasi DBMP dalam menentukan prioritas perbaikan jalan di kota Samarinda. Dataset pengujian berupa data jalan yang rusak di Kota Samarinda selama satu tahun yaitu tahun 2015-2016. Berdasarkan hasil pengujian dan analisis system penentuan prioritas perbaikan jalan, maka dapat disimpulkan bahwa sistem ini mampu memberikan rekomendasi keputusan yang lebih baik dalam proses penentuan prioritas perbaikan jalan di kota Samarinda dengan menggunakan metode *Fuzzy C-Means* sehingga keputusan yang diperoleh lebih tepat, terkomputerisasi dan mencegah pengambilan keputusan yang bersifat subjektif.

Penelitian ketujuh yang dilakukan oleh Asmiatun & Wakhidah (2018) mengimplementasikan model algoritma Metode *K-Means Clustering* dalam Tujuan dari pengelompokkan tersebut untuk mengetahui kecamatan mana yang paling rawan rusak. Pengumpulan dataset menggunakan aplikasi accelerometer. Aplikasi tersebut memanfaatkan sensor accelerometer dan GPS yang ada di android untuk merekam jalan yang tidak rata. Data yang direkam meliputi nama kelurahan, latitude, dan longitude. Hasilnya dari 7 kecamatan terdapat satu kecamatan yang paling rawan jalan rusak yaitu kecamatan Gayamsari. Untuk penelitian selanjutnya diharapkan dapat menguji akurasi dari penerapan algoritma k-means. Sehingga jika akurasi yang dihasilkan sudah baik maka dapat diimplementasikan ke dalam aplikasi.

Penelitian kedelapan yang dilakukan oleh Asmiatun et al. (2020) Memfokuskan penerapan metode untuk pengolahan data dengan mengelompokkan data-data yang didapat dari sensor accelerometer menggunakan metode *K-medoids*. Dataset yang telah dihasilkan dari aplikasi roadroid yang memanfaatkan sensor accelerometer berfungsi sebagai perekam kondisi jalan. Data ini diambil pada kisaran bulan januari tahun 2019 oleh pihak Dinas Pekerjaan Umum Kota Semarang. Dari data yang digunakan jumlah 638 data telah berhasil dikelompokkan menjadi 4 cluster yang dapat menunjukkan tingkat frekuensi jalan kondisi baik, sedang, rusak ringan dan rusak berat.

Adapun penelitian ini merujuk kepada penelitian sebelumnya yang mendapatkan persamaan dan kesenjangan dari tiap rujukan, sebagaimana yang terdapat pada Tabel 2.1. berikut :

Tabel 2.1. Tabel Kesenjangan Penelitian

No	Peneliti	Tujuan Penelitian	Metode yang digunakan	Hasil Penelitian
1.	(Listiani et al., 2019)	Mengelompokkan data penduduk untuk rekomendasi jenis pekerjaan berdasarkan faktor pendidikan, umur dan jenis kelamin menggunakan data mining <i>algoritma K-Means Cluster</i> .	<i>K-Means Clustering</i>	1. Hasil : Menghasilkan 4 kelompok jenis pekerjaan berdasarkan jenis kelamin, pendidikan, umur dan jenis pekerjaan dengan nilai akurasi yang diperoleh yaitu sebesar 79%. 2. Kesenjangan : Untuk cakupan data dan wilayah yang belum luas sehingga jumlah k masih kurang
2.	(Vernanda et al., 2021)	Memberikan informasi mengenai titik-titik daerah rawan kecelakaan agar lebih waspada dalam berkendara, membantu pihak Satlantas dalam mengevaluasi wilayah yang memiliki kerawanan laka lantas tinggi. Sehingga di tangani dengan cepat dan menekan	<i>K-Means Clustering</i>	1. Hasil : Menghasilkan 3 <i>Cluster</i> yaitu Sangat Rawan(SR), Rawan(R), dan Cukup Rawan (CR). 2. Kesenjangan : Hasil perhitungan manual mengnampilkan kecamatan yang berbeda.

No	Peneliti	Tujuan Penelitian	Metode yang digunakan	Hasil Penelitian
3.	(Nabila et al., 2021)	<p>angka kecelakaan lalu lintas.</p> <p>Menganalisis data kasus Covid-19 untuk pengelompokan pada masalah kasus Covid-19 di Provinsi Lampung.</p>	<i>K-Means Clustering</i>	<p>1. Hasil :</p> <p>Menghasilkan pengelompokan yang berbeda dikarenakan jumlah pada atribut Suspek, Probable, Konfirmasi Positif, Selesai Isolasi, dan Kematian pada setiap Kabupaten/Kota yang tidak sama.</p> <p>2. Kesenjangan : Penelitian dilakukan dengan data per Kabupaten/Kota sehingga hasil kurang detail dan pengelompokan yang berbeda.</p>
4.	(Mulyani et al., 2018)	Mengelompokan data untuk mengatur jadwal mata pelajaran pengayaan sesuai dengan kebutuhan kelompok siswa.	<i>K-Means Clustering</i>	<p>1. Hasil :</p> <p>Menghasilkan 2 Kelompok siswa berdasarkan nilai rata-rata raport, Kelompok 1 pola yang dihasilkan adalah Inggris -> Kimia dan Kelompok 2 pola yang dihasilkan adalah Kimia -> Inggris dan Biologi.</p> <p>2. Kesenjangan : Untuk jumlah k</p>

No	Peneliti	Tujuan Penelitian	Metode yang digunakan	Hasil Penelitian
				yang kecil sehingga pengelompokan belum akurat dan sesuai
5.	(Kurniadi et al., 2018)	Menganalisis, memprediksi, dan mengklasifikasikan siswa yang memiliki potensi untuk mendapatkan beasiswa di universitas.	Algoritma K-Nearest Neighbor	1. Hasil : Memiliki tingkat akurasi mencapai 95,83 persen. 2. Kesenjangan : <i>Clustering</i> pengelompokan pembangunan jalan akan menggunakan metode <i>K-Means</i>
6	(Puspitasari et al., 2017)	Mengelompokkan data jalan yang rusak kedalam empat kategori prioritas perbaikan jalan. Untuk memberikan rekomendasi DBMP dalam menentukan prioritas perbaikan jalan di kota Samarinda.	Metode <i>Fuzzy C-Means</i>	1. Hasil : Melakukan pengelompokkan data ke dalam beberapa kelompok. Apabila kemiripan data yang satu dengan yang lain yang sangat besar dengan data lain, maka data berada pada cluster yang sama. 2. Kesenjangan : <i>Clustering</i> pengelompokan pembangunan jalan akan menggunakan metode <i>K-Means</i>
7	(Asmiatun & Wakhidah, 2018)	Tujuan dari pengelompokkan untuk mengetahui kecamatan mana	<i>K-Means Clustering</i>	1. Hasil : Hasilnya dari 7 kecamatan terdapat satu

No	Peneliti	Tujuan Penelitian	Metode yang digunakan	Hasil Penelitian
		yang paling rawan rusak.		kecamatan yang paling rawan jalan rusak yaitu kecamatan Gayamsari. 2. Kesenjangan : Diharapkan dapat menguji akurasi dari penerapan algoritma <i>k-means</i> . Sehingga jika akurasi yang dihasilkan sudah baik maka dapat diimplementasikan ke dalam aplikasi.
8	(Asmiatun et al., 2020)	Memfokuskan penerapan metode untuk pengolahan data dengan mengelompokkan data-data yang didapat dari sensor accelerometer menggunakan metode <i>K-medoids</i>	metode <i>K-medoids</i>	1. Hasil : Dari data yang digunakan jumlah 638 data telah berhasil dikelompokkan menjadi 4 cluster yang dapat menunjukkan tingkat frekuensi jalan kondisi baik, sedang, rusak ringan dan rusak berat. 2. Kesenjangan : <i>Clustering</i> pengelompokan pembangunan jalan akan menggunakan metode <i>K-Means</i>

2.3. Definisi Jalan

Jalan merupakan sarana transportasi utama untuk mencapai suatu tujuan dari tempat satu ke tempat lain bagi setiap lalu lintas yang melewatinya. Oleh karena itu, kondisi jalan sangat berpengaruh bagi keamanan dan keselamatan setiap

pengguna jalan. Jalan adalah prasarana transportasi darat yang meliputi segala bagian jalan, termasuk bangunan pelengkap dan perlengkapannya yang diperuntukkan bagi lalu lintas, yang berada pada permukaan tanah, di atas permukaan tanah, di bawah permukaan tanah dan atau air, serta di atas permukaan air, kecuali jalan kereta api, jalan lori, dan jalan kabel. Jalan Umum adalah jalan yang diperuntukkan bagi lalu lintas umum. Jalan Khusus adalah jalan yang dibangun oleh pemerintah atau instansi, badan usaha, perseorangan, atau kelompok masyarakat untuk kepentingan sendiri (Nizar & E.Purba, 2018).

Jalan sebagai bagian sistem transportasi nasional punya peran penting dalam mendukung bidang ekonomi, sosial dan budaya. Dengan tersedianya jalan, dapat menjadi penghubung antar wilayah yang pada akhirnya perekonomian dapat bergerak lebih baik. Jalan dipedesaan berfungsi sebagai penghubung antar desa atau ke lokasi pemasaran, sebagai penghubung hunian, serta penghubung desa ke pusat kegiatan yang lebih tinggi tingkatannya (kecamatan/kabupaten) (Yuono, 2017).

2.3. Data Mining

Data Mining merupakan proses penggalian informasi dan pola yang bermanfaat dari data yang besar, dimana mencakup pengumpulan data, ekstraksi data, dan statistik data. Dimana data mining dikenal sebagai *knowledge discovery*, *knowledge extraction*, *data/pattern analysis*, *information harvesting* (Arhami & Nasir, 2020).

Data mining bertujuan untuk menemukan pola yang sebelumnya tidak diketahui. Jika pola-pola tersebut telah diperoleh maka dapat digunakan untuk menyelesaikan berbagai macam permasalahan, Data mining saat ini telah menjadi suatu teknologi baru yang kuat dengan potensi besar untuk membantu perusahaan fokus pada informasi paling penting dalam data yang telah mereka kumpulkan tentang perilaku pelanggan dan pelanggan potensial mereka. Melalui data mining, perusahaan dapat menemukan informasi dalam data yang begitu besar melalui pengolahan yang tepat dan efektif dengan berbagai metode yang ada dalam data mining sehingga secara sederhana data mining dapat digambarkan sebagai suatu

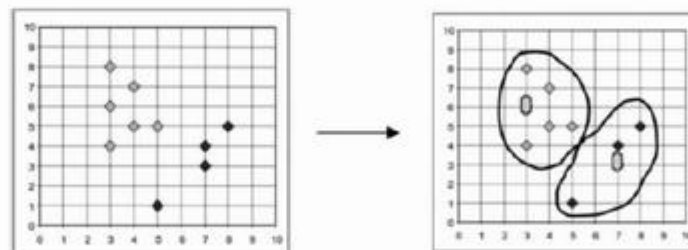
pola atau model atau kaidah atau pengetahuan yang dihasilkan dari data mining (Arhami & Nasir, 2020).

Salah satu penggunaan data mining yaitu dalam bisnis adalah saat data mining digunakan untuk menemukan pola dan hubungan dalam data dan membantu membuat keputusan bisnis yang lebih baik. Data mining dapat membantu mengenali tren penjualan, mengembangkan kampanye pemasaran yang lebih cerdas, dan secara akurat memprediksi loyalitas pelanggan (Arhami & Nasir, 2020).

2.4. Pengelompokan (*Clustering*)

Pengelompokan atau *clustering* merupakan salah satu metode pembelajaran tidak terawasi (*unsupervised learning*), yang dalam permasalahan *Clustering* data-data atau nilai-nilai yang ada belum mempunyai target atau belum mempunyai label kelasnya, sehingga perlu diprediksi ke dalam kelas mana nantinya suatu objek itu akan masuk berdasarkan kesamaan polanya atau kesamaan kelompoknya (Arhami & Nasir, 2020).

Jika diberikan beberapa *record* misalnya data, sampel, objek, observasi, dan lain sebagainya untuk mengorganisasikan mereka ke dalam klaster, *record-record* tersebut diklasterkan melalui proses pengelompokan fisik atau abstrak objek-objek tersebut ke dalam kelompok atau kelas yang memiliki kesamaan dan kedekatan (Arhami & Nasir, 2020).



Gambar 2.1. Ilustrasi Proses *Clustering*

(Sumber : (Arhami & Nasir, 2020))

Suatu klaster dapat digambarkan sebagai bagian himpunan dari objek-objek yang memiliki kesamaan ciri, pola atau kedekatan. Suatu himpunan bagian dari objek merupakan jarak antar tiap dua objek yang ada dalam klater (Arhami & Nasir, 2020).

2.5. Algoritma *K-Means*

K-Means merupakan salah satu algoritma *Clustering*. Tujuan algoritma ini yaitu untuk membagi data menjadi beberapa kelompok. Algoritma ini menerima masukan berupa data tanpa label kelas. Hal ini berbeda dengan K-NN (dan algoritma *supervised learning* lainnya) yang menerima masukan berupa vektor (Primartha, 2018).

Pada algoritma *K-Means*, komputer mengelompokkan sendiri data-data yang menjadi masukannya tanpa mengetahui terlebih dulu target kelasnya. Masukan yang diterima adalah data atau objek dan k buah kelompok (*Cluster* yang diinginkan). Algoritma ini akan mengelompokkan data atau objek ke dalam k buah kelompok tersebut. Pada setiap *Cluster* terdapat titik pusat (*centroid*) yang merepresentasikan *Cluster* tersebut. Secara sederhana dapat dijelaskan bahwa *K-Means* adalah algoritma *machine learning* yang digunakan untuk menyelesaikan permasalahan *Clustering* (Primartha, 2018).

K-Means ditemukan oleh beberapa orang yaitu Lloyd (1957, 1982), Forgey (1965), Friedman and Rubin (1967), and McQueen (1967). Ide dari *Clustering* pertama kali ditemukan oleh Lloyd pada tahun 1957, namun hal tersebut baru dipublikasi pada tahun 1982. Pada tahun 1965, Forgey juga mempublikasi teknik yang sama sehingga terkadang dikenal sebagai Lloyd-Forgey pada beberapa sumber. *K-Means* merupakan metode klasterisasi yang paling terkenal dan Queen pada tahun 1967 (Primartha, 2018).

Menurut (Arhami & Nasir, 2020) Algoritma ini merupakan salah satu algoritma *Clustering* yang sangat umum dalam mengelompokkan data sesuai dengan kesamaan karakteristik atau ciri-ciri yang serupa. Kelompok data yang dihasilkan disebut sebagai klaster. Algoritma *K-Means* juga mudah untuk diimplementasikan dan memiliki kompleksitas waktu dan ruang yang relatif kecil. Algoritma ini juga merupakan algoritma yang cukup efisien dalam komputasinya dan memberikan hasil yang cukup baik dan memuaskan jika klasternya compact, hyperspherical dalam shape dan mampu memisahkan fitur-fitur ruangnya dengan baik. Menurut Hasn & Kamber, algoritma *K-Means* bekerja dengan membagi data

ke dalam k buah klaster yang telah ditentukan. Langkah-langkah dasar untuk algoritma *K-Means* adalah:

1. Tentukan nilai k klaster sesuai dengan yang diinginkan;
2. Pilih titik-titik atau sampel yang menjadi anggota klaster secara acak;
3. Tentukan nilai centroid atau titik tengah dari klaster tersebut dengan rumus:

$$M_k = \left(\frac{1}{n_k} \right) \sum_{i=1}^{n_k} X_{ik} \quad (2.1)$$

Dimana X_{ik} = sampel yang ada dalam cluster C_k

4. Hitung *square error* untuk tiap klaster C_k yang merupakan jumlah kuadrat dari jarak *Euclidean* antara tiap sample C_k dan titik tengahnya (*centroid*). Error ini dikenal juga dengan nama *within Cluster variation* (WCV), yaitu:

$$e_k^2 = \sum_{i=1}^{n_k} (X_{ik} - M_k)^2 \quad (2.2)$$

5. Selanjutnya jumlah dari keseluruhan error dari *K-Cluster* juga dihitung dengan rumus:

$$E_k^2 = \sum_{k=1}^k e_k^2 \quad (2.3)$$

6. Kelompokkan kembali semua sampel berdasarkan jarak minimum dari masing-masing pusat M_1, M_2, \dots, M_k , sehingga diperoleh distribusi baru dari sampel sesuai klasternya. Untuk memperoleh distribusi sampel baru tersebut dapat dilakukan dengan menghitung jarak masing-masing titik pusat dengan keseluruhan sampel $d(M_1, X_1) \dots d(M_k, X_k)$. Perhitungan jarak dari masing-masing titik tersebut dapat menggunakan beberapa metode, contohnya:

Euclidean Distance

Jarak antara dua titik dalam satu, dua, tiga atau sampai dimensi n dapat dihitung sebagai berikut:

$$d(p, q) = \sqrt{((p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2)} \quad (2.4)$$

7. Tuliskan hasil anggota klaster baru sesuai dengan hasil yang diperoleh pada langkah ke-5; dan
8. Ulangi langkah ke-3 sampai beberapa iterasi sehingga nantinya ditemukan nilai total square error turun secara signifikan.

Adapun karakteristik dari algoritma *K-Means* salah satunya adalah sangat

sensitif pengelompokan titik pusat awal kluster karena *K-Means* membangkitkan titik pusat kluster awal secara random. Pada saat pembangkitan awal titik pusat yang random tersebut mendekati solusi pembangkitan awal titik pusat yang random tersebut mendekati solusi akhir pusat kluster, *K-Means* mempunyai kemungkinan yang tinggi untuk menemukan titik pusat kluster yang tepat. Sebaliknya, jika awal titik pusat tersebut jauh dari solusi akhir pusat kluster, maka besar kemungkinan ini menyebabkan hasil pengklasteran yang tidak tepat. Akibatnya *K-Means* tidak menjamin hasil pengklasteran yang unik. Inilah yang menyebabkan metode *K-Means* sulit untuk mencapai optimum global, akan tetapi hanya minimum lokal. Selain itu, algoritma *K-Means* hanya bisa digunakan untuk data yang atributnya bernilai numerik (Primartha, 2018).

2.6. Davies-Bouldin Indeks (DBI)

Menurut (Nabila et al., 2021) Metode yang digunakan dalam menentukan evaluasi cluster dapat menggunakan *Davies-Bouldin Index* (DBI). Dimana *Davies-Bouldin Index* merupakan salah satu metode evaluasi internal yang mengukur evaluasi *cluster* pada suatu metode pengelompokan yang didasarkan pada nilai kohesi dan separasi. Dalam suatu pengelompokan, kohesi didefinisikan sebagai jumlah dari kedekatan data terhadap *centroid* dari *cluster* yang diikuti. Sedangkan separasi didasarkan pada jarak antar *centroid* dari clusternya.

Sum of square within cluster (SSW) merupakan persamaan yang digunakan untuk mengetahui matrik kohesi dalam sebuah cluster ke-*i* yang dirumuskan sebagai berikut:

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i) \quad (2.5)$$

Keterangan:

m = Jumlah data dalam cluster ke-*i*

c = *Centroid cluster* ke-*i*

d = Jarak

Sum of square between cluster (SSB) merupakan persamaan yang digunakan untuk mengetahui separasi antar *cluster* yang dihitung menggunakan persamaan:

$$SSB_{i,j} = d(c_i, c_j) \quad (2.6)$$

Keterangan:

d = Jarak

c = *Centroid cluster* ke-i

Setelah nilai kohesi dan separasi diperoleh, kemudian dilakukan pengukuran rasio (R_{ij}) untuk mengetahui nilai perbandingan antara cluster ke-i dan cluster ke-j. *Cluster* yang baik adalah *cluster* yang memiliki nilai kohesi sekecil mungkin dan separasi yang sebesar mungkin. Nilai rasio dihitung menggunakan persamaan sebagai berikut:

$$R_{i,j} = \frac{SSW_I + SSW_J}{SSB_{i,j}} \quad (2.7)$$

Nilai rasio yang diperoleh tersebut digunakan untuk mencari nilai *davies-bouldin index* (DBI) dari persamaan berikut:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (2.8)$$

Dari persamaan tersebut, k merupakan jumlah cluster yang digunakan. Semakin kecil nilai DBI yang diperoleh (non-negatif ≥ 0), maka semakin baik cluster yang diperoleh dari pengelompokan *K-Means* yang digunakan.

2.7. Cross-Industry Standard Process for Data mining (CRISP-DM)

Menurut Chapman dalam (Wardani, 2021) CRISP-DM (*Cross-Industry Standard Process for Data mining*) merupakan suatu konsorsium perusahaan yang didirikan oleh Komisi Eropa pada Tahun 1996 dan telah ditetapkan sebagai proses standar dalam data mining yang dapat diaplikasikan di berbagai sektor industri.

4. *Modeling*

Dalam tahap ini akan dilakukan pemilihan dan penerapan berbagai teknik pemodelan dan beberapa parameternya akan disesuaikan untuk mendapatkan nilai yang optimal. Secara khusus, ada beberapa teknik berbeda yang dapat diterapkan untuk masalah data mining yang sama. Di pihak lain ada teknik pemodelan yang membutuhkan format data khusus. Sehingga pada tahap ini masih memungkinkan kembali ke tahap sebelumnya;

5. *Evaluation*

Pada tahap ini, model sudah terbentuk dan diharapkan memiliki kualitas baik jika dilihat dari sudut pandang analisa data. Pada tahap ini akan dilakukan evaluasi terhadap keefektifan dan kualitas model sebelum digunakan dan menentukan apakah model dapat mencapai tujuan yang ditetapkan pada fase awal (*business understanding*). Kunci dari tahap ini adalah menentukan apakah ada masalah bisnis yang belum dipertimbangkan. Di akhir dari tahap ini harus ditentukan penggunaan hasil proses data mining; dan

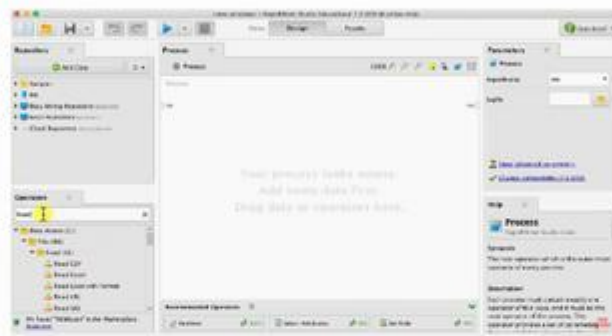
6. *Deployment*

Pada tahap ini, pengetahuan atau informasi yang telah diperoleh akan diatur dan dipresentasikan dalam bentuk khusus sehingga dapat digunakan oleh pengguna. Tahap *deployment* dapat berupa pembuatan laporan sederhana atau mengimplementasikan proses data mining yang berulang dalam perusahaan. Dalam banyak kasus, tahap deployment melibatkan konsumen, di samping analisis data, karena sangat penting bagi konsumen untuk memahami tindakan yang harus dilakukan untuk menggunakan model yang telah dibuat.

2.8. *RapidMiner*

RapidMiner merupakan perangkat lunak yang bersifat terbuka (*open source*). *RapidMiner* adalah sebuah solusi untuk melakukan analisis terhadap *data mining*, *text mining* dan analisis prediksi. *RapidMiner* menggunakan berbagai teknik

deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. *RapidMiner* memiliki kurang lebih 500 operator *data mining*, termasuk operator untuk *input*, *output*, *data preprocessing* dan visualisasi. *RapidMiner* merupakan software yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri. *RapidMiner* ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi (Siregar & Puspabhuana, 2017).



Gambar 2.3. Tampilan Lembar Kerja *RapidMiner*

(Sumber : (Siregar & Puspabhuana, 2017))

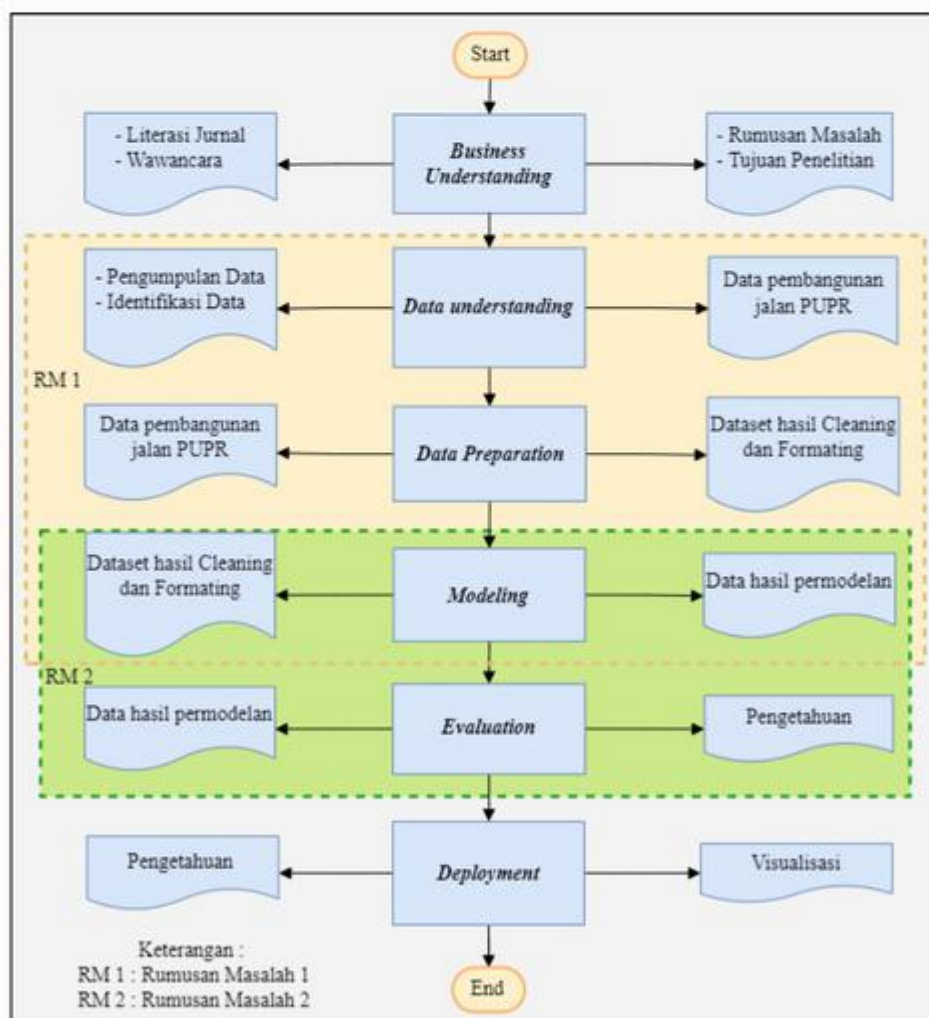
BAB III

METODE PENELITIAN

Metodologi yang digunakan dalam penerapan *K-Means* pada pengelompokan Pembangunan jalan di dinas Pekerjaan Umum dan Penataan Ruang (PUPR) menggunakan pendekatan metodologi *CRISP-DM*.

3.1. Kerangka Penelitian

Kerangka penelitian dalam pengelompokan Pembangunan jalan ini menerapkan pendekatan *CRISP-DM* sebagaimana disajikan pada Gambar 3.1 berikut :

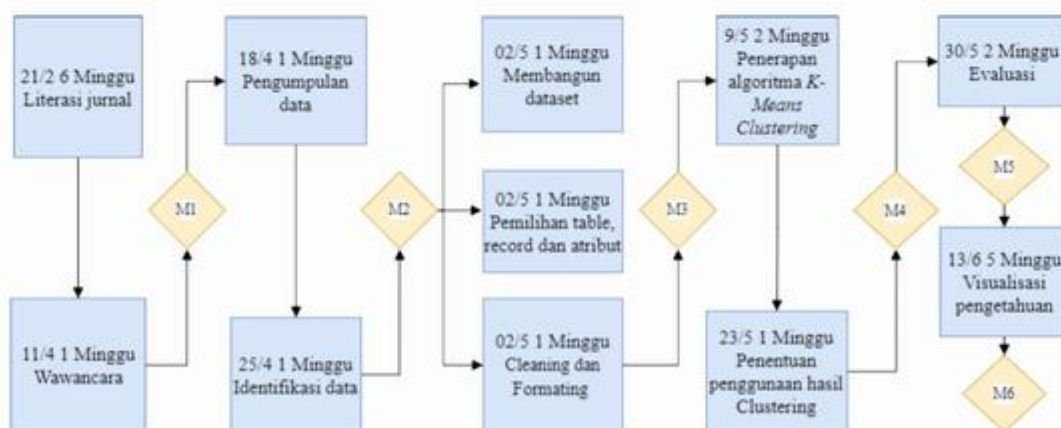


Gambar 3.1. Kerangka Penelitian

Penelitian ini menerapkan pendekatan *CRISP-DM* dimana kegiatan awal dimulai dengan *business understanding* dengan kegiatan memahami kebutuhan PUPR rencana dan strategi penyelesaiannya, *data understanding* dengan melakukan pengumpulan data dari dinas PUPR dan dianalisis kualitas datanya, *data preparaton* membuat dataset beserta atribut datanya, kemudian di *cleaning* dan di *forming*, *cleaning* dilakukan dengan menghapus *missing value* pada attribut yang memiliki data yang kosong atau *null* agar data yang dihitung lebih akurat karena sudah tidak ada *missing value*. Data yang sudah bersih, memasuki tahap *forming* data, data berjenis non nominal dilakukan proses inialisasi kedalam bentuk angka/numerikal, *modeling* pemilihan penerapan teknik pada dataset yang telah dibangun, pada tahap ini *Clustering* dan *K-Means* menjadi pilihan untuk pengolahan data dari dinas PUPR, *evaluation* melakukan perhitungan DBI dan menganalisis hasil apakah sesuai dengan tujuan utama atau melihat apabila ada permasalahan yang belum terpikirkan sebelumnya juga mempertimbangkan hasil akhir dari proses *Clustering* ini dan *deployment* yaitu tahap penerapan hasil proses data mining, pada hal ini hasil akhirnya adalah berupa visualisasi pengetahuan.

3.2. Diagram Activity

Berdasarkan kerangka penelitian dengan pendekatan CRISP-DM yang telah dibuat, maka diagram alur aktivitas dapat di lihat pada Gambar sebagai berikut :



Gambar 3.2. Diagram Activity

Untuk diagram aktivitas dibuat agar alur pelaksanaan dapat berlangsung secara lebih efektif dimana kegiatan diawali dengan analisis tujuan dari literasi dari

peneliti sebelumnya dan observasi, kemudian analisa kebutuhan pada intansi yang memiliki suatu masalah, kemudian melakukan pengumpulan data dan identifikasi kualitas data sehingga akhirnya dapat dibuat dataset yang telah dibangun dari hasil pembangunan pada PUPR yang kemudian di identifikasi kualitasnya dan menentukan table, *record* dan atributnya kemudian di *cleaning* dan di *formating* dengan menghapus *missing value* pada attribut yang memiliki data yang kosong atau *null*, tahap *formating* data berjenis non nominal dilakukan proses inisialisasi kedalam bentuk angka. Setelah itu dilakukan penerapan algoritma *K-Means* sampai dengan menghasilkan kelompok-kelompok pada pembangunan jalan di dinas PUPR kemudian evaluasi *clustering* dengan *Davies-Bouldin Index* (DBI) dengan alur hasil perhitungan pada tahap modeling di lanjutkan dengan perhitungan DBI sampai dengan nilainya minimum sesuai dengan tujuan sebelumnya di analisis kebutuhan yaitu menghasilkan kelompok pembangunan jalan yang baik sehingga menghasilkan nilai dari perhitungan DBI dan kemudian di tentukan bersama Dinas PUPR untuk menerapkan hasil perhitungan *clustering* tersebut lalu tahapan setelah pengelompokan hasil *clustering* seluruhnya sehingga dengan hasil observasi sebelumnya pengelompokan ini dibuatkan visualisasi pengetahuan sehingga dinas PUPR atau pihak terkait yang terbatu mengetahui kelompok pembangunan jalan yang diajukan oleh desa pada Dinas PUPR.

3.3. Sumber Daya Penelitian

Sumber daya penelitian baik tim peneliti dan tempat studi kasus juga perangkat yang digunakan dalam penelitian pengelompokan pembangunan jalan di dinas PUPR disajikan dengan sebagaimana tampak pada Tabel 3.1 berikut :

Tabel 3.1. Sumber Daya Penelitian

No	Aktivitas	Manusia	Perangkat
1	Literasi jurnal	Tim Peneliti	<i>Smartphone</i> , Laptop, Buku
2	Wawancara	Hari Ilham Nur Akbar, Dinas PUPR	<i>Smartphone</i> , Buku
3	Pengumpulan data	Hari Ilham Nur Akbar, Dinas PUPR	<i>Flashdisk</i> , <i>Smartphone</i>
4	Identifikasi kualitas data	Tim Peneliti	Laptop
5	Membangun dataset	Tim Peneliti	Laptop
6	Pemilihan table, record dan atribut	Tim Peneliti	Laptop
7	<i>Cleaning</i> dan <i>formating</i>	Tim Peneliti	Laptop
8	Penerapan Algoritma <i>K-Means</i>	Tim Peneliti	Laptop
9	Penentuan penggunaan hasil <i>Clustering</i>	Tim Peneliti	Laptop
10	Evaluasi	Tim Peneliti	Laptop
11	Visualisasi Pengetahuan	Tim Peneliti	Laptop

3.4. Sumber Data

Data yang digunakan untuk penerapan *K-Means* pada penelitian ini merupakan data pembangunan jalan oleh Dinas PUPR dari tahun 2015 sampai dengan 2021 yang berisikan 263 *record* data dengan keterangan no, kode pembangunan, jenis pembanguna dengan keterangan angka dari rehabilitasi berupa angka 1, rekontruksi angka 2 dan pembangunan dengan angka 3, kemudian panjang pembangunan dalam ukuran meter dan nominal pada rupiah. Berikut merupakan tabel sampel data pembanguna jalan yang ditampilkan dari total 263 data pada Tabel 3.2 :

Tabel 3.2. Sampel Data Pembangunan Jalan Oleh Dinas PUPR

No	Kode Pembangunan	Jenis Pembangunan	Panjang	Nominal
1	rk001	2	1400	1353660000
2	rk010	2	221	190.502.000
3	rk011	2	163	140.506.000
4	rk012	2	105	90.510.000
5	rh005	1	274	191.416.400
6	rh006	1	205	143.213.000
7	rh007	1	132	92.215.200
8	pb001	3	500	3.868.949.000
9	pb012	3	100	773.789.800
10	pb017	3	300	2.321.369.400
11	pb018	3	205	1.586.269.100
12	rh083	1	131	91.516.600
13	rh084	1	131	91.516.600
14	rh085	1	132	92.215.200
15	rk081	2	105	90.510.000
16	rk082	2	335	191.452.000
17	rk083	2	160	91.440.000
18	rk084	2	208	190.611.200
19	rk085	2	208	190.611.200
20	rk086	2	335	191.452.000
21	rk087	2	232,51	198.271.040
22	rk088	2	232,51	198.271.040
23	rh086	1	131	91.516.600
24	rh087	1	131	91.516.600
25	rh088	1	73	91.688.000
26	rh089	1	73	91.688.000
27	rh090	1	125	190.487.500
28	rh091	1	125	190.487.500
29	pb040	3	100	773.789.800
30	pb041	3	100	773.789.800
31	pb042	3	100	773.789.800
32	pb043	3	100	773.789.800
33	pb044	3	100	773.789.800
34	rk089	2	160	91.440.000
35	rk090	2	232,51	198.271.040
36	rk091	2	232,51	198.271.040
37	rk092	2	208	190.611.200
38	rk093	2	335	191.452.000
39	rk094	2	160	91.440.000

JADWAL RENCANA KEGIATAN PENELITIAN

Jadwal kegiatan penelitian ini menggambarkan kegiatan yang dilaksanakan pada penerapan *K-Means* untuk pengelompokan Pembangunan jalan di Dinas PUPR berikut kegiatan disajikan pada Tabel berikut :

Tabel Jadwal Rencana Kegiatan Penelitian


NO	KEGIATAN/AKTIVITAS	BULAN DAN MINGGU KE-																											
		FEBRUARI				MARET				APRIL				MEI				JUNI				JULI				AGUSTUS			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	<i>Business understanding</i>																												
	<i>a. Literatur jurnal</i>																												
	<i>b. Wawancara</i>																												
2	<i>Data understanding</i>																												
	<i>a. Pengumpulan data</i>																												
	<i>b. Identifikasi kualitas data</i>																												
3	<i>Data Preparation</i>																												
	<i>a. Membangun data set</i>																												
	<i>b. Pemilihan table, record dan atribut</i>																												
	<i>c. Cleaning dan Formating</i>																												
4	<i>Modeling</i>																												
	<i>a. Penerapan Algoritma K-Means</i>																												
	<i>b. Penggunaan dan penentuan hasil clustering</i>																												
5	<i>Evaluation</i>																												
	<i>a. Evaluasi</i>																												
6	<i>Deployment</i>																												
	<i>a. Visualisasi</i>																												
7	Proposal Skripsi																												
8	Seminar Proposal Skripsi																												
9	Laporan Skripsi																												
10	Jurnal Skripsi																												
11	Poster Skripsi																												
12	Sidang Skripsi																												
13	Revisi Laporan Skripsi																												

DAFTAR PUSTAKA

- Arhami, M., & Nasir, M. (2020). *Data Mining Algoritma dan Implementasi*.
- Asmiatun, S., & Wakhidah, N. (2018). Identifikasi Pengelompokan Kondisi Permukaan Jalan Menggunakan Algoritma K-Means. *Jurnal Pengembangan Rekayasa Dan Teknologi*, 14(1), 17. <https://doi.org/10.26623/jprt.v14i1.1215>
- Asmiatun, S., Wakhidah, N., & Putri, A. N. (2020). Penerapan Metode K-Medoids Untuk Pengelompokan Kondisi Jalan Di Kota Semarang. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 6(2), 171–180. <https://doi.org/10.35957/jatisi.v6i2.193>
- Jatnika, F. A., Akbardin, J., Natawidjana, R., Nanang, D., & Herman, D. (2020). *Perancangan Petunjuk Pelaksanaan (Sop) Berdasarkan Rencana Pembangunan Jalan Di Provinsi Banten*. 229.
- Kurniadi, D., Abdurachman, E., Warnars, H. L. H. S., & Suparta, W. (2018). The prediction of scholarship recipients in higher education using k-Nearest neighbor algorithm. *IOP Conference Series: Materials Science and Engineering*, 434(1). <https://doi.org/10.1088/1757-899X/434/1/012039>
- Listiani, L., Agustin, Y. H., & Ramdhani, M. Z. (2019). Implementasi algoritma k-means cluster untuk rekomendasi pekerjaan berdasarkan pengelompokan data penduduk. *SENSITif: Seminar Nasional Sistem Informasi Dan Teknologi Informasi*, 761–769.
- Mulyani, E. D. S., Susanto, Agustin, Y. H., & Surgawi, N. M. (2018). Implementasi Algoritma K-Means Dan Fp- Growth Untuk Rekomendasi Bimbingan Belajar Berdasarkan Segmentasi Akademik Siswa. *IT Journal*, 6(2), 160–173.
- Nabila, Z., Rahman Isnain, A., & Abidin, Z. (2021). Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means. *Jurnal Teknologi Dan Sistem Informasi (JTSI)*, 2(2), 100. <http://jim.teknokrat.ac.id/index.php/JTSI>
- Nizar, H. A., & E.Purba, V. (2018). Evaluasi Jalan Rabat Beton baru pada STA 2.000 -STA 3.000 di Jalan Jambura Nagori Buntu Bayu Kecamatan Hatanduhan Kabupaten Simalungun. *Angewandte Chemie International Edition*, 6(11), 951–952., 10–27.
- Primartha, R. (2018). *Belajar Machine Learning: Teori dan Praktik*.
- PUPR, D. (2021). *Data Pembangunan Jalan Di Dinas PUPR*.
- Puspitasari, N., Rosmasari, R., & Stefanie, S. (2017). Penentuan Prioritas Perbaikan Jalan Menggunakan Fuzzy C-Means : Studi Kasus Perbaikan Jalan

- Di Kota Samarinda. *Jurnal Teknologi Dan Sistem Komputer*, 5(1), 7.
<https://doi.org/10.14710/jtsiskom.5.1.2017.7-14>
- Siregar, A. M., & Puspabhuana, A. (2017). *DATA MINING: Pengolahan Data Menjadi Informasi dengan RapidMiner*.
- Vernanda, A. A., Faisol, A., & Vendyansyah, N. (2021). Penerapan Metode K-Means Clustering Untuk Pemetaan Daerah Rawan Kecelakaan Lalu Lintas Di Kota Malang Berbasis Website. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 5(2), 836–844. <https://doi.org/10.36040/jati.v5i2.3791>
- Wardani, N. W. (2021). *Penerapan Data Mining Dalam Analytic CRM*.
- Yuono, T. (2017). *Evaluasi Kuat Tekan Jalan Beton Yang Pola Pembangunannya Dengan Pemberdayaan Masyarakat*. 1–5.

Lampiran A : Presensi Bimbingan



**JURUSAN ILMU KOMPUTER
PROGRAM STUDI TEKNIK INFORMATIKA**

BERITA ACARA BIMBINGAN SKRIPSI

A DATA MAHASISWA			
1	Nama Mahasiswa	Hari Ilham Nur Akbar	
2	Nomor Induk	1806129	
3	Kontrak Skripsi	I] Ganjil [^{1/2} ; Gasap, tahun akademik 2021 / 2022	
4	Judul Skripsi	PENERAPAN ALGORITMA K-MEANS UNTUK PENGELompokAN PEMBANGUNAN JALAN PADA DINAS PEKERJAAN UMUM DAN PENATAAN RUANG	
B DATA PEMBIMBING			
1	Pembimbing I (P1)	Dr. Dede Iqurniandi, S.Kom, M.Kom.	
	Nomor Induk:	0402098301	
	Jabatan	Lektor	
2	Pembimbing II (P2)	Yuga Hantoko Agustin, S.Kom, M.Kom.	
	Nomor Induk:	0424088002	
	Jabatan	Asisten Ahli	
C DATA BIMBINGAN/WENGGU:			
NO	KEGIATAN BIMBINGAN	TANGGAL/JAM:	PARAF
1	Judul & Bab I	11/3 - 2022	P1 <input checked="" type="checkbox"/> P2 <input checked="" type="checkbox"/>
	Judul & Bab I	9/3 - 2022	P1 <input checked="" type="checkbox"/> P2 <input checked="" type="checkbox"/>
2	Bab I, II, III revisi	29/3 - 2022	P1 <input checked="" type="checkbox"/> P2 <input checked="" type="checkbox"/>
	Bab I - III Revisi	24/03 - 2022	P1 <input checked="" type="checkbox"/> P2 <input checked="" type="checkbox"/>
3	Bab III revisi	28/3 - 2022	P1 <input checked="" type="checkbox"/> P2 <input checked="" type="checkbox"/>
	Bab III Revisi	28/03 - 2022	P1 <input checked="" type="checkbox"/> P2 <input checked="" type="checkbox"/>
4	Bab I, II, III rapikan	31/3 - 2022	P1 <input checked="" type="checkbox"/> P2 <input checked="" type="checkbox"/>
	Rapikan Laporan	31/03 - 2022	P1 <input checked="" type="checkbox"/> P2 <input checked="" type="checkbox"/>
5	Bab I, II, III Acc Proposal	2/4 - 2022	P1 <input checked="" type="checkbox"/> P2 <input checked="" type="checkbox"/>
	ACC Promosal	02/04 - 2022	P1 <input checked="" type="checkbox"/> P2 <input checked="" type="checkbox"/>


Berita Acara Bimbingan Skripsi 1/2

Lampiran B : Surat Bukti Penelitian

	PEMERINTAH KABUPATEN GARUT BADAN KESATUAN BANGSA DAN POLITIK Jalan Patriot No. 10 A Telp. (0262) 2247473 Garut 44151
<hr/>	
Nomor : 072/251-Bakesbangpol/III/2022	Garut, 17 Maret 2022
Lampiran : 1 (Satu) lembar	Kepada :
Perihal : <i>Pemohonan Penelitian</i>	Yth, Kepala Dinas PUPR Kabupaten Garut
	di
	Tempat

Dalam rangka membantu Mahasiswa/i Institut Teknologi Garut bersama ini terlampir
Rekomendasi Penelitian Nomor : 072/251-Bakesbangpol/III/2022 Tanggal 17 Maret 2022, **HARI**
ILHAM NUR AKBAR yang akan melaksanakan Penelitian dengan mengambil lokasi Dinas PUPR
Kabupaten Garut. Demi kelancaran Penelitian dimaksud, mohon bantuan dan kerjasamanya untuk
membantu Kegiatan tersebut.

Demikian atas perhatiannya, kami haturkan terima kasih.


Kepala Badan Kesatuan Bangsa dan Politik
Kabupaten Garut
Drs. H. NORRODHIN, M.Si.
Pembina Tk. I, IV/b
NIP. 19661019 199203 1 005

Tembusan, disampaikan kepada :

1. Yth. Kepala Bappeda Kabupaten Garut;
2. Yth. Rektor ITG;
3. Arsip



PEMERINTAH KABUPATEN GARUT
BADAN KESATUAN BANGSA DAN POLITIK

Jalan Patriot No. 10 A Telp. (0262) 2247473 Garut 44151

REKOMENDASI PENELITIAN

Nomor :072/251-Bakesbangpol/III/2022

- a. Dasar :
1. Peraturan Menteri Dalam Negeri Nomor 41 Tahun 2010 tentang Organisasi dan Tata Kerja Kementerian Dalam Negeri (Berita Negara Republik Indonesia Tahun 2010 Nomor 316), sebagaimana telah diubah dengan Peraturan Menteri Dalam Negeri Nomor 14 Tahun 2011 tentang Perubahan Atas Peraturan Menteri Dalam Negeri Nomor 41 Tahun 2010 tentang Organisasi dan Tata Kerja Kementerian Dalam Negeri (Berita Negara Republik Indonesia Tahun 2011 Nomor 168);
 2. Peraturan Menteri Dalam Negeri Nomor 3 Tahun 2018 tentang Penerbitan Surat Keterangan Penelitian.

Memperhatikan : Surat dari, Institut Teknologi Garut Nomor: 319/ITG/A.5/B/III/2022 Tanggal 15 Maret 2022

KEPALA BADAN KESATUAN BANGSA DAN POLITIK KABUPATEN GARUT, memberikan Rekomendasi kepada :

- | | | |
|-------------------------------------|---|--|
| 1. Nama / NPM / NIM/NIDN | : | HARI ILHAM NUR AKBAR/1806129 |
| 2. Alamat | : | Kp.Cisanca Kaler RT/Rw 003/007 Ds.Mekarwangi
Kec.Tarogong Kaler Kab.Garut |
| 3. Tujuan | : | Penelitian |
| 4. Lokasi/ Tempat | : | Dinas PUPR Kabupaten Garut |
| 5. Tanggal/ Lama Penelitian | : | 17 Maret 2022 s/d 17 Juni 2022 |
| 6. Bidang/ Status/ Judul Penelitian | : | - |
| 7. Nama Penanggung jawab | : | Dr.Hilmi Aulawi,S.T.,M.T. |
| 8. Anggota | : | - |
1. Melaporkan hasil Penelitian ke Badan Kesatuan Bangsa dan Politik Kabupaten Garut;
 2. Menjaga dan menjunjung tinggi norma atau adat istiadat dan Kebersihan, Ketertiban, Keindahan (K3) masyarakat setempat dilokasi Penelitian;
 3. Tidak melakukan hal-hal yang bertentangan dengan hukum dan atas dasar adat istiadat di lokasi Penelitian atau sesuatu yang dapat meresahkan masyarakat dan disintegrasi bangsa.

Demikian rekomendasi ini dibuat untuk digunakan seperlunya.

Kepala Badan Kesatuan Bangsa dan Politik
Kabupaten Garut

Dr. H. ERRODIN, M.Si.

Pembina Tk.I, IV/b

NIP. 19661019 199203 1 005

Tembusan, disampaikan kepada:

1. Yth. Kepala Bappeda Kabupaten Garut;
2. Yth. Rektor ITG;
3. Arsip.