BIG DATA: PEMETAAN SEKOLAH UNTUK KUNJUNGAN PROMOSI PENERIMAAN MAHASISWA BARU MENGGUNAKAN METODE K-NEAREST NEIGHBOR(KNN) DAN K-MEANS CLUSTERING

Proposal ini dibuat untuk memenuhi persyaratan kelulusan matakuliah Program Internship II dan Pra Tugas Akhir

Dibuat Oleh,

1.16.4.060 Yogi Aditya Saputra

PROGRAM DIPLOMA IV TEKNIK INFORMATIKA

POLITEKNIK POS INDONESIA

BANDUNG

2020

LEMBAR PENGESAHAN

BIG DATA: PEMETAAN SEKOLAH UNTUK KUNJUNGAN PROMOSI PENERIMAAN MAHASISWA BARU MENGGUNAKAN METODE K-NEAREST NEIGHBOR(KNN) DAN K-MEANS CLUSTERING

Yogi Aditya Saputra 1.16.4.060

Laporan Program Internship II ini telah diperiksa, disetujui dan disidangkan
di Bandung, 12 Mei 2020

Oleh :

Pembimbing,

Syafrial Fachri Pane, S.T., M.T.I.,EBDP.
NIK: 117.88.233

Menyetujui,

Ketua Program Studi DIV Teknik Informatika,

M. Yusril Helmi Setyawan, S.Kom., M.Kom.
NIK: 113.74.163

LEMBAR PENGESAHAN

BIG DATA: PEMETAAN SEKOLAH UNTUK KUNJUNGAN PROMOSI PENERIMAAN MAHASISWA BARU MENGGUNAKAN METODE K-NEAREST NEIGHBOR(KNN) DAN K-MEANS CLUSTERING

Yogi Aditya Saputra 1.16.4.060

Laporan Program Internship II ini telah diperiksa, disetujui dan disidangkan
di Bandung, 12 Mei 2020

Oleh :

Penguji Utama,	Penguji Pendamping,
Syafrial Fachri Pane, S.T., M.T.I.,EBDP.	Rolly Maulana Awangga, S.T., M.T.
NIK: 117.88.233	NIK: 117.86.219

Mengetahui,
Koordinator Internship II,

Noviana Riza, S.Si., M.T.
NIK: 103.78.065

ABSTRAK

Penerapan Big Data pada penelitian ini adalah akan menerapkan konsep Big Data dalam pemetaan sekolah untuk kunjungan promosi penerimaan mahasiswa baru. Proses pemetaan sekolah ini menggunakan metode *K-Means Clustering* dan *K-Nearest Neighbor* dengan Python. Hasil dari penelitian ini diharapkan akan membantu tim PMB dalam menentukan strategi promosi penerimaan mahasiswa baru berdasarkan hasil pemetaan sekolah tersebut.

Kata Kunci: *Big Data*, Pemetaan Sekolah, *K-Means Clustering*, *K-Nearest Neighbor*, Promosi.

ABSTRACT

The application of Big Data in this research is to apply the concept of Big Data in school mapping for new student acceptance promotion visits. This school mapping process uses the K-Means Clustering and K-Nearest Neighbor method in Python. The results of this study are expected to assist the PMB team in determining the strategy for promoting new student admissions based on the results of the school mapping.

Keywords: *Big Data, School Mapping, K-Means Clustering, K-Nearest Neighbor, Promotion.*

KATA PENGANTAR

Assalamualaikum warahmatullahi wabarakatuh. Segala puji bagi Allah SWT yang telah memberikan kemudahan sehingga dapat menyelesaikan laporan Internship II dan Pra Tugas Akhir ini, tanpa pertolongan-Nya mungkin penulis tidak akan sanggup menyelesaikannya dengan baik. Shalawat dan salam semoga terlimpah curahkan kepada Nabi Muhammad SAW beserta sahabat dan keluarga Beliau.

Laporan ini disusun untuk memenuhi kelulusan matakuliah Internship II dan Pra Tugas Akhir pada Program Studi DIV Teknik Informatika. Proses Internship II dan Pra Tugas Akhir ini juga tidak terlepas dari bantuan berbagai pihak. Oleh karena itu, pada kata pengantar ini penulis menyampaikan terimakasih kepada :

1. Syafrial Fachri Pane, S.T., M.T.I.,EBDP. selaku Pembimbing Internship II dan Pra Tugas Akhir ini.
2. Rolly Maulana Awangga, S.T., M.T. selaku Penguji Pendamping dalam penyusunan laporan Internship II dan Pra Tugas Akhir ini.
3. Noviana Riza, S.Si., M.T. selaku Koordinator Internship II dan Pra Tugas Akhir Tahun Akademik 2019/2020.
4. M. Yusril Helmi Setyawan, S.Kom., M.Kom. selaku Ketua Program Studi DIV Teknik Informatika Tahun Akademik 2019/2020.

Penulis telah membuat laporan ini dengan sebaik-baiknya, diharapkan memberikan kritik dan saran dari semua pihak yang bersifat membangun, terimakasih.

Bandung, 12 Mei 2020

Yogi Aditya Saputra

DAFTAR ISI

DAFTAR GAMBAR

BAB I

PENDAHULUAN

1.1 Latar Belakang

Setiap bergantinya tahun akademik baru, perguruan tinggi swasta selalu bersaing dengan perguruan tinggi negeri dan perguruan tinggi swasta lain untuk mendapatkan jumlah mahasiswa baru.[1] Calon mahasiswa baru merupakan sumber utama pendapatan bagi perguruan tinggi swasta dan faktor terpenting yang harus dijadikan perhatian serius, oleh karena itu upaya meraih hati calon mahasiswa baru harus dapat dilakukan oleh perguruan tinggi.[2] Agar bisa bertahan, setiap perguruan tinggi swasta harus melakukan berbagai macam hal dalam memikat daya tarik calon mahasiswa baru untuk memilih perguruan tinggi tertentu.Untuk itulah perguruan tinggi swasta memerlukan kegiatan promosi dalam penerimaan mahasiswa baru. Promosi adalah sebuah tindakan atau upaya untuk membujuk pelanggan untuk melakukan pembelian sehingga tecapai tujuan perusahaan yang menciptakan permintaan.[1] Dalam hal ini, pelanggan yang dimaksud adalah calon mahasiswa baru, perusahaan sebagai perguruan tinggi serta permintaan sebagai mahasiswa di perguruan tinggi tersebut.

Kegiatan promosi tersebut dalam perguruan tinggi biasanya terdapat pada kegiatan Penerimaan Mahasiswa Baru(PMB). Kegiatan PMB yang sudah rutin dilaksanakan secara tidak langsung menghasilkan banyak data mahasiswa baru. Hal ini perlu diperhatikan oleh perguruan tinggi untuk mengelola data tersebut yang nantinya menjadi informasi penting. Informasi itu salah satunya mengenai pemetaan sekolah untuk kunjungan promosi penerimaan mahasiswa baru.[3]

Permasalahan yang terjadi dalam perekrutan mahasiswa adalah banyaknya jumlah calon mahasiswa yang tidak melakukan registrasi ulang untuk setiap tahunnya padahal peserta seleksi telah dinyatakan lulus. Hal ini menjadi masalah yang harus diselesaikan oleh pihak kampus, yang berarti kuota mahasiswa baru untuk setiap program studi belum tercapai. Salah satu penyebab terjadinya penurunan jumlah mahasiswa baru pada tahun ini adalah kurang dilakukannya pengolahan data mahasiswa secara tepat berdasarkan data historis oleh pihak admisi perguruan tinggi.Hal tersebut dapat mempengaruhi pengambilan keputusan dalam menentukan wilayah promosi yang tepat sasaran. Pengolahan data mahasiswa seharusnya dilakukan agar

dapat menentukan wilayah promosi yang tepat sasaran sehingga tidak terjadi penurunan jumlah mahasiswa pada tahun berikutnya.[4]

Big Data adalah teknologi baru untuk mengelola, menganalisis, dan memvisualisasikan data yang berkembang pesat yang dihadapi dalam perusahaan dan masyarakat.[5] Konsep Big Data yaitu 3V yaitu Volume(ukuran data), Velocity(kecepatan mengolah data) serta Variety(variasi tipe data).[6] Penelitian ini menerapkan Big Data dengan menggunakan metode K-Means untuk mengelompokkan sekolah-sekolah yang akan menjadi prioritas kunjungan promosi kegiatan penerimaan mahasiswa baru, setelah itu K-Nearest Neighbor untuk memberikan rekomendasi sekolah lain berdasarkan koordinat terdekat.

Clustering atau biasa disebut pengelompokkan adalah suatu alat bantu pada data mining yang bertujuan mengelompokkan objek-objek kedalam cluster-cluster.[7] K-Means adalah salah satu metode clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster atau kelompok.[8] Klasifikasi adalah proses penemuan model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui.[9] K-Nearest Neighbour (k-nn) adalah algoritma klasifikasi data sederhana dimana penghitungan jarak terpendek dijadikan ukuran untuk mengklasifikasikan suatu kasus baru berdasarkan ukuran kemiripan.[10]

Dengan menerapkan konsep Big Data pada pemetaan sekolah untuk kunjungan promosi penerimaan mahasiswa baru berdasarkan hasil pengolahan data mahasiswa dari tahun sebelumnya. Informasi yang dihasilkan pada penelitian ini dapat membantu tim PMB dalam memetakan sekolah untuk kunjungan promosi penerimaan mahasiswa baru tercapai.

1.2 Rumusan Masalah

Berdasarkan latar belakang sebelumnya, dapat disimpulkan bahwa rumusan masalah sebagai berikut :

1. Bagaimana cara mengimplemtasikan konsep Big Data dalam pemetaan sekolah untuk kunjungan promosi penerimaan mahasiswa baru menggunakan metode K-Means Clustering dan K-Nearest Neighbor(KNN).

1.3 Ruang Lingkup

Agar penelitian ini berjalan dengan baik dan terarah, maka penelitian ini perlu adanya batasan masalah. Berikut adalah batasan masalah dalam penelitian ini :

1. Pada penelitian ini, Hadoop digunakan sebagai tempat menyimpan dataset.
2. Pada penelitian ini, Spark digunakan sebagai untuk melakukan proses pengolahan data. Dimana Spark diintegrasikan dengan Jupyter Notebook.
3. Pada penelitian ini, Virtual Hadoop digunakan sebagai nodes. Dimana nantinya jika terjadi permasalahan pada nodes atau yang lainnya, masih memiliki backup data pada nodes master. Sehingga kemungkinan kehilangan data tidak terjadi.
4. Penelitian ini menggunakan data penerimaan mahasiswa baru dari tahun 2017 - 2019.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah sebelumnya, dapat disimpulkan bahwa :

1. Dapat mengimplemtasikan konsep Big Data dalam pemetaan sekolah untuk kunjungan promosi penerimaan mahasiswa baru menggunakan metode K-Means Clustering dan K-Nearest Neighbor(KNN).

1.5 Manfaat Penelitian

Manfaat dari penelitian ini sebagai berikut :

1. Bagi Perguruan Tinggi:
    - Dapat membantu tim penerimaan mahasiswa baru(PMB) dalam mengambil keputusan untuk melakukan kunjungan promosi penerimaan mahasiswa baru dengan cepat dan efisien.
2. Bagi peneliti:
    - Dapat mengimplemtasikan konsep Big Data dalam pemetaan sekolah untuk kunjungan promosi penerimaan mahasiswa baru menggunakan metode K-Means Clustering dan K-Nearest Neighbor(KNN).

BAB II

LANDASAN TEORI

2.1 Teori Umum

2.1.1 Konsep *Data, Information, Knowledge*

Konsep *Data, Information, Knowledge* biasa sering disebut konsep hirarki data wisdom sehingga menjadi DIKW(*Data, Information, Knowledge, Wisdom.* DIKW adalah sebuah model yang dikonsep oleh Russell Lincoln Ackoff. Ackoff adalah seorang konsultan manajemen dan mantan profesor dibidang manajemen di Wharton School yang mengkhususkan diri dalam riset operasi dan teori organisasi. Artikelnya merumuskan apa yang sekarang biasa disebut DIKW untuk pertama kali pada tahun 1988 sebagai pidato pada *International Society for General Systems Research.* Jonatan Hey (2004) mengatakan Asal usul hirarki DIKW (Data, Informasi, Pengetahuan, Kebijaksanaan) pertama kali tampil pada domain Manajemen Pengetahuan dan Sains Informasi. Meskipun referensi ke DIKW hierarki dibuat oleh Zeleny (1987) dan Ackoff (1989) dalam bidang Manajemen Pengetahuan, referensi terdekat ada pada petunjuk asli dari artikel T.S. Eliot yang muncul di Futuris oleh Cleveland (1982). Ada hal yang menarik, pada Eliot atau Harland. Sebelumnya data tidak ada dalam hirarki informasi, pengetahuan, kebijaksanaan, tetapi ditambahkan oleh yang lain. Sejak orang-orang mengajukan data (D) penambahan pada hirarhi (IKW), Ackoff memasukkan pemahaman (dan beberapa menggunakan intelijensia) dalam level sebelum mencapai kebijaksanaan (wisdom), dan Zeleny mengajukan pencerahan (enlighment) sebagai tahap akhir yang melampaui kebijaksanaan (wisdom).

Pada konsep DIKW fakta-fakta yang terjadi diformulasikan menjadi sebuah data. Formulasi fakta ini dicatat dan direkam dalam berbagai bentuk data seperti teks, angka, gambar, suara, video, dan symbol. Hasil pencatatan data ini dimaknai dalam berbagai konteks untuk kemudian menjadi informasi. Saat informasi bersinggungan dengan pengalaman dan gagasan dari penggunanya maka informasi ini berubah menjadi pengetahuan yang nantinya akan mempengaruhi keputusan. Tercapainya istilah pengetahuan dan kebijaksanaan, perlu dicatat bahwa hal itu bergantung pada data dan informasi (yaitu pengetahuan adalah kumpulan dari data atau informasi), dan bahwa kebijaksanaan harus membantu orang untuk membuat keputusan yang baik. [11]

Gambar 2.1: Konsep DIKW

2.1.2 Sumber Data

Berdasarkan sumber data dapat dibedakan menjadi dua, yaitu :

1. Data primer adalah data yang diperoleh atau dikumpulkan oleh orang yang melakukan penelitian atau yang bersangkutan yang memerlukannya. Data primer disebut juga data asli atau data baru.

2. Data sekunder adalah data yang diperoleh atau dikumpulkan dari sumber-sumber yang telah ada. Data itu biasanya diperoleh dari perpustakaan atau laporan-laporan, dokumen peneliti yang terdahulu. Data skunder disebut juga data tersedia. [12]

2.1.3 *Big Data*

Big Data adalah teknologi baru untuk mengelola, menganalisis, dan memvisualisasikan data yang berkembang pesat yang dihadapi dalam perusahaan dan masyarakat. Big Data Analytics (BDA) mengacu pada teknologi dan kerangka kerja yang dirancang untuk dengan cepat menyimpan, mengkonversi, mentransfer, dan menganalisis sejumlah besar data yang terus diperbarui, langsung bervariasi, terstruktur dan tidak terstruktur untuk keuntungan komersial dan sosial. BDA kini telah berevolusi dari sistem manajemen basis data besar ke layanan cloud untuk memproses dan menganalisis data untuk membuatnya lebih ekonomis, lebih efektif, dan lebih mudah bagi pengguna untuk memanipulasi. Perusahaan vendor utama data global antara lain IBM, Oracle, SAP, EMC, Teradata, dan SAS. Solusi yang saat ini ditawarkan oleh vendor ini meliputi Gudang Data, Penambangan Data, Analisis Bisnis, Intelijen Bisnis, Data Visualisasi, Pendukung Keputusan, Antarmuka Otomasi, dan sejenisnya.

Big Data memiliki tiga fitur: volume, kecepatan, dan variasi. Sebagian besar diskusi di masa lalu berfokus pada cara menyimpan volume data. Kecepatan dan variasi sangat penting dalam diferensiasi kompetitif. Ragam mengacu pada berbagai format data. Data dapat berupa data terstruktur yang dapat diurutkan atau data non-terstruktur, seperti gambar, musik, video, esai, dan diskusi. Dibandingkan

dengan data terstruktur, data non-terstruktur memberikan refleksi yang lebih baik dari kenyataan untuk membuat keputusan penting. Fitur lainnya adalah kecepatan. Dalam lingkungan bisnis di mana setiap hitungan detik, bisnis harus mengumpulkan dan menganalisis data secara tepat waktu untuk membuat keputusan penting lebih cepat dibandingkan dengan pesaing mereka. Dengan memproses volume besar informasi yang terus berubah yang harus diproses segera, bisnis dapat mengubah massa data yang tampaknya tidak berguna menjadi nilai ekonomi. [5]

2.1.4 *Python*

Python adalah bahasa pemrograman tingkat tinggi yang dewasa ini telah menjadi standar dalam komputasi ilmiah. Python merupakan bahasa open source multiplatform yang dapat digunakan pada berbagai macam sistem operasi (Windows, Linux, dan MacOS). Selain itu, python juga merupakan bahasa pemrograman yang fleksibel dan mudah untuk dipelajari. Program yang ditulis dalam Python umumnya lebih mudah dibaca dan jauh lebih ringkas dibandingkan penulisan program dalam bahasa C atau Fortran.Python juga memiliki modul standar yang menyediakan sejumlah besar fungsi dan algoritma, untuk menyelesaikan pekerjaan seperti mengurai data teks, memanipulasi dan menemukan file dalam disk, membaca menuliskan file terkompresi, dan mengunduh data dari server web. Dengan menggunakan Python, para programmer juga dapat dengan mudah menerapkan teknik komputasi tingkat lanjut, seperti pemrograman berorientasi objek.

Bahasa pemrograman Python dalam banyak hal berbeda dengan bahasa pemrograman prosedural, seperti C; C++; dan Fortran. Dalam Fortran, C, dan C++, file source code harus dikompilasi ke dalam bentuk executable file sebelum dijalankan. Pada Python, tidak terdapat langkah kompilasi, sebagai gantinya source code ditafsir kan secara langsung baris demi baris. Keunggulan utama dari suatu bahasa pemrograman terinterpretasi seperti Python adalah tidak membutuhkan pendeklarasian variabel, sehingga lebih fleksibel dalam penggunaannya. Namun, terdapat kelemahan yang mencolok, yaitu program – program numerik yang dijalankan pada Python lebih lambat ketimbang dijalankan menggunakan bahasa pemrograman terkompilasi. Kelemahan ini tentu membuat kita berpikir apakah Python cocok untuk digunakan dalam komputasi ilmiah? Meskipun bekerja dengan agak lambat, Python memiliki banyak fungsi – fungsi sederhana yang dapat menjalankan hal – hal yang umumnya dikerjakan dengan subroutine rumit dalam C dan/atau Fortran. Sehingga Python merupakan pilihan tepat dalam komputasi ilmiah dewasa ini.

Untungnya, banyak routine numerik dan matematis umum yang telah disusun sebelumnya, yang dikelompokan ke dalam dua buah paket (NumPy dan SciPy) yang dapat diimpor secara mudah ke dalam Python. Paket NumPy (Numerical Python) menyediakan banyak routine dasar guna memanipulasi array dan matriks numerik dalam skala besar. Paket SciPy (Scientific Python) memperluas kegunaan NumPy dengan kumpulan algoritma ilmiah yang sangat berguna, seperti minimalisasi; transformasi Fourier; regresi; dan banyak teknik – teknik aplikasi matematis lainnya. Karena bersifat open source, kedua paket ini sangat populer di kalangan ilmuwan. Dengan adanya paket NumPy dan SciPy, Python dapat berdiri sejajar (bahkan di atas) bersama program komputasi ilmiah Matlab dalam penggunaannya sebagai alat bantu sains dewasa ini. [13]

2.1.5 *Hadoop*

Hadoop adalah Suatu software framework (kerangka kerja perangkat lunak) open source berbasis Java di bawah lisensi Apache untuk aplikasi komputasi data besar secara intensif.

Hadoop File System dikembangkan menggunakan desain sistem file yang terdistribusi. Tidak seperti sistem terdistribusi, HDFS sangat faulttolerant dan dirancang menggunakan hardware low-cost. Atau dalam arti lain, Hadoop adalah Software platform (platform perangkat lunak) sebagai analytic engine yang memungkinkan seseorang dengan mudah untuk melakukan pembuatan penulisan perintah (write) dan menjalankan (run) aplikasi yang memproses data dalam jumlah besar, dan di dalamnya terdiri dari:

- HDFS - *Hadoop Distributed File System*
- *MapReduce - Offline Computing Engine*

Dalam komputasi, platform menggambarkan semacam (hardware architecture) arsitektur perangkat keras atau (software framework) kerangka kerja perangkat lunak (termasuk kerangka kerja aplikasi), yang memungkinkan perangkat lunak dapat berjalan.

Ciri khas dari platform meliputi arsitekturnya komputer, sistem operasi, bahasa pemrograman dan runtime libraries atau GUI yang terkait. Apa yang ada pada Hadoop dari sudut pandang:

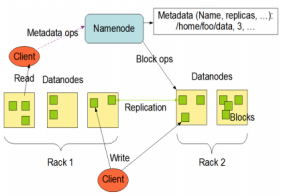
- Platform : Komputer sebagai node.
- Framework : HDFS Explorer [6]

2.1.5.1 *Hadoop Distributed File System*(HDFS)

Hadoop terdiri dari HDFS *(Hadoop Distributed file System)* dan Map Reduce. HDFS sebagai direktori di komputer dimana data hadoop disimpan. Untuk pertama kalinya, direktori ini akan di "format" agar dapat bekerja sesuai spesifikasi dari Hadoop. HDFS sebagai file system, tidak sejajar dengan jenis file system dari OS seperti NTFS, FAT32. HDFS ini menumpang diatas file system milik OS baik Linux, Mac atau Windows.

Data di Hadoop disimpan dalam cluster.

Cluster biasanya terdiri dari banyak node atau komputer/server. Setiap node di dalam cluster ini harus terinstall Hadoop untuk bisa jalan. [6]



Gambar 2.2: HDFS Arsitektur

2.1.6 *Spark*

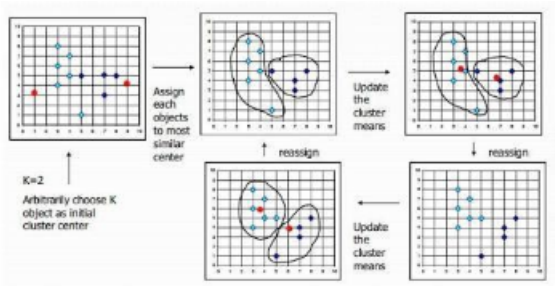
Apache Spark adalah sumber terbuka, sistem komputasi tujuan umum yang digunakan untuk data besar analitik. Spark mampu menyelesaikan pekerjaan secara substansial lebih cepat dari alat data besar sebelumnya (mis. Apache Hadoop) karena caching dalam memori, dan eksekusi query yang dioptimalkan. Spark menyediakan API pengembangan dengan Python, Java, Scala, dan R. Di atas kerangka komputasi utama, Spark menyediakan pembelajaran mesin, SQL, analisis grafik, dan perpustakaan streaming.

API Python Spark dapat diakses melalui paket PySpark. Instalasi untuk eksekusi lokal atau koneksi jarak jauh ke kluster yang ada dapat dilakukan dengan perintah conda atau pip. [14]

```
# PySpark installation with conda
conda install -c conda-forge pyspark
# PySpark installation with pip
pip install pyspark
```

2.1.7 *Clustering*

Clustering atau klasterisasi adalah suatu alat bantu pada data mining yang bertujuan mengelompokkan objek-objek kedalam cluster-cluster. Cluster adalah sekelompok atau kumpulan objek-objek data yang memiliki kemiripan karakteristik satu sama lain dalam cluster yang sama dan berbeda karakteristik terhadap objek-objek yang berbeda cluster. [7]



Gambar 2.3: Metode Clustering

2.1.8 K Means

Metode ini adalah salah satu metode non hierarchi yang umum digunakan. Metode ini termasuk dalam teknik partisi yang membagi atau memisahkan objek ke kelompok daerah bagian yang terpisah. Pada K-Means, setiap objek harus masuk dalam kelompok tertentu, tetapi dalam satu tahapan proses tertentu, objek yang sudah masuk dalam satu kelompok, pada satu tahapan berikutnya akan pindah ke kelompok lain.

Hasil cluster dengan metode K-Means sangat bergantung pada nilai pusat kelompok awal yang diberikan. Pemberian nilai awal yang berbeda bisa menghasilkan kelompok yang berbeda. Ada beberapa cara memberi nilai awal misalnya dengan mengambil sampel awal dari objek, lalu mencari nilai pusatnya, memberi nilai awal secara random, menentukan nilai awalnya atau menggunakan hasil dari kelompok hirarki dengan jumlah kelompok yang sesuai.

K Means adalah suatu metode penganalisaan data atau metode Data Mining yang melakukan proses pemodelan tanpa supervisi (unsupervised) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Metode K-Means berusaha mengelompokan data yang ada kedalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu dengan yang lainnya dan mempunyai karakteristik yang berbeda dengan data yang berada pada

kelompok yang lain. Dengan kata lain, metode ini berusaha untuk meminimalkan variasi antar data yang berada didalam suatu kelompok dan memaksimalkan variasi dengan data yang berada pada kelompok lainnya. [7]

2.1.9 Klasifikasi

Klasifikasi adalah proses penemuan model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui.

Klasisifkasi data terdiri dari dua langkah proses, yang pertama adalah proses learning (fase training) dimana algoritma klasifikasi dibuat untuk menganalisa data training direpresentasikan dalam bentuk rule klasifikasi, proses kedua adalah klasifikasi dimana data tes digunakan untuk memperkirakan akurasi dari rule klasifikasi.

Proses klasifikasi didasarkan pada empat komponen yaitu :

1. Kelas

   variabel dependen yang berupa kategorikal yang merepresentasikan label yang terdapat pada objek.

2. *Predictor*

   Variabel independen direpresentasikan oleh karakteristik atribut data.

3. Data *Testing*

   Berisi data baru yang akan diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasi dievaluasi.

4. Data *Training*

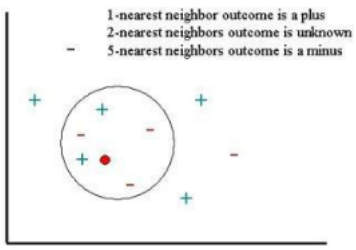
   Satu set data yang berisi dari kedua komponen diatas yang digunakan untuk menentukan kelas yang cocok berdasarkan predictor.

Terdapat beberapa algoritma yang sering digunakan pada teknik klasifikasi yaitu algoritma k-nearest neighbor classification, pohon keputusan (decision tree), naive bayesian classification, dan support vector machines. [9]

2.1.10 K-Nearest Neighbor

Algoritma k-nearest neighbour (k-nn) adalah algoritma klasifikasi data sederhana dimana penghitungan jarak terpendek dijadikan ukuran untuk mengklasifikasikan suatu kasus baru berdasarkan ukuran kemiripan. Algoritma k-nn tergolong dalam algoritma supervised yaitu proses pembentukan algoritma diperoleh melalui proses pembelajaran (learning) pada record-record lama yang sudah terklasifikasi dan hasil pembelajaran tersebut dipakai untuk mengklasifikasikan record baru dengan output yang belum diketahui.

Dalam algoritma k-nn sebuah data baru diklasifikasikan berdasarkan jarak data baru dengan tingkat kemiripan data baru terdekat terhadap data pola. Jumlah data tetangga terdekat ditentukan dan dinyatakan dengan k. Misalkan ditentukan k=1, maka kasus ini hanya diklasifikasikan untuk satu data dari tetangga terdekat. Jika nilai k didefinisikan berbeda oleh user, misal k=5 maka kasus dengan 5 jarak terpendek dipilih, kemudian diklasifikasi berdasarkan instance kelas target dimana kasus dengan jumlah mayoritas instance kelas target ditentukan sebagai klasifikasi untuk kasus baru. Representasi k-NN dengan nilai k=1, k=2 dan k=5 dapat dilihat pada Gambar 2.4.



Gambar 2.4: Ilustrasi K-NN

Keterangan:

- Jika 1-nearest neighbour maka hasil +
- Jika 2-nearest neighbour maka hasil tidak diketahui
- Jika 5-nearest neighbour maka hasil -

Untuk lebih jelas melihat hubungan antara data mining Penentuan nilai k terbaik tergantung pada data. Nilai k yang tinggi bisa mengurangi efek noise pada klasifikasi, tetapi membuat batasan antara setiap kelas menjadi kabur. Sedangkan penentuan nilai k=1 belum tentu bisa menjawab permasalahan data mining dalam hal ini tingkat

validitas. Nilai k terbaik dapat dipilih dengan optimasi parameter, misalnya dengan menggunakan k-fold cross validation. Untuk membedakan nilai k pada cross validation dengan nilai k pada k-NN, maka digunakan n-fold cross validation untuk mengacu kepada istilah yang sama yaitu k-fold cross validation. [10]

Berikut ini langkah – langkah klasifikasi metode k-nearest neighbor :

1. Menentukan parameter k (jumlah tetangga paling dekat).
2. Menghitung jarak antara data yang akan dievaluasi dengan semua pelatihan.
3. Mengurutkan jarak yang terbentuk.
4. Menentukan jarak terdekat sampai urutan k.
5. Memasangkan kelas yang bersesuaian.
6. Mencari jumlah kelas dari tetangga yang terdekat dan menetapkan kelas tersebut sebagai kelas data yang akan dievaluasi. [9]

2.1.11 Pengumpulan Data

Pengumpulan data merupakan salah satu tahapan sangat penting dalam penelitian. Teknik pengumpulan data yang benar akan menghasilkan data yang memiliki kredibilitas tinggi, dan sebaliknya.[15] Didalam pengumpulan data tersebut terdapat beberapa teknik yang sering digunakan, seperti observasi, wawancara, dan kuesioner. Berikut penjelasan terkait teknik pengumpulan data tersebut.

2.1.11.1 Observasi

Observasi selain sebagai salah satu tahap dalam pelaksanaan PTK sekaligus juga berfungsi sebgai alat untuk pengumpulan data. Metode ini sangat sesuai untuk merekam aktivitas yang bersifat proses. Misalnya kegiatan siswa selama melakukan praktikum di laboratorium, interaksi siswa selama kegiatan pembelajaran, atau saat mereka sedang melakukan diskusi. Dalam istilah assessment, kegiatan observasi merupakan bagian dari informal assessment (authentic assessment) yang bersifat langsung (direct assessment).

Dilihat dari sudut pelaksanaannya, kegiatan observasi bisa bersifat langsung (partiscipatif observation) maupun tidak langsung (non-participatif observation). Dalam observasi tidak langsung, peneliti tidak terlibat secara langsung dalam proses pembelajaran (tidak berinteraksi langusung dengan objek yang diteliti), namun hanya merekam segala aktivitas sesuai fokus atau indikator yang diinginkan.

Observasi langsung dilakukan dengan adanya keterlibatan secara langsung oleh peneliti dalam proses pembelajaran yang dilakukan bersama guru dan siswa, atau bahkan peneliti sekaligus sebagai guru. Sebenarnya kondisi seperti inilah yang diharapkan nanti. Artinya ke depan guru harus berfungsi sebagai peneneliti di kelasnya sendiri (sebagai participant observer).

Dilihat dari teknik pelaksanaannya, observasi dapat dibedakan menjadi observasi terbuka, terfokus, terstruktur, dan sistematis. Observasi terbuka biasa dikenal dengan kegiatan observasi yang dilakukan dengan membuat catatan bebas tentang segala aktivitas yang berkaitan langsung dengan objek yang diteliti. Misalnya peneliti ingin merekam segala aktivitas yang dianggap penting selama anak sedang melakukan kegiatan diskusi.

Observasi terfokus dilaksanakan dengan merekam segala sesuatu yang maksud dan tujuanmya telah ditentukan atau direncanakan sebelumnya, termasuk alat bantu yang akan digunakan. Observasi ini digunakan untuk mengamati atau merekam baik aktivitas yang dilakukan oleh guru maupun siswa selama kegiatan belajar mengajar berlangsung. Untuk menghindari subjektivitas observer, maka perlu dilengkapi dengan pedoman observasi yang begitu rinci, sehingga observer tinggal merekam sasaran dengan memberikan coding pada lembar pengamatan seseuai kesepakatan yang telah ditetapkan sebelumnya.

Observasi terstruktur dilaksanakan dengan dibuatnya suatu lembar atau pedoman observasi yang berisi indikator-indikator yang mungkin muncul. Dalam hal ini observer tinggal memberi tanda ceklist pada gejala yang muncul selama proses pengamatan. Observasi model ini untuk menghindarkan subjektivitas dari pengamat. Melalui pengamatan model ini akan teridentifikasi suatu pola atau kecenderungan interaktif baik antara siswa dengan siswa atau antara siswa dengan guru.

Observasi sistematis berupa suatu pedoman yang bersifat standart atau baku, sehingga mampu mendapatkan data kuantitatif dalam jumlah dan kualitas yang memadai. Namun kelemahan observasi seperti ini dianggap kurang informatif.

2.1.11.2 Kuesioner

Self report dapat berbentuk angket atau kuesioner yang diberikan kepada para peserta didik untuk mengungkap tentang wawasan, pandangan atau aspek kepribadian, yang jawabannya dapat diberikan secara tertulis. Keuntungan menggunakan metode angket, yaitu bisa digunakan untuk kelas yang besar, dan membutuhkan waktu yang relatif singkat.

Dilihat dari cara menjawabnya, angket dapat dibedakan menjadi angket terbuka dan tertutup. Angket terbuka bila pihak yang ingin mengisi diberikan kesempatan untuk menjawab sesuai perasaan dan pengalaman mereka. Sedangkan pada angket tertutup, pihak penjawab tidak diberi kebebasan untuk menjawab pertanyaan sesuai pengalaman dan perasaan mereka. Sebab pada kuesioner jenis ini sudah diberikan alternatif jawaban mulai dari kategori sangat senang sampai pada kategori tidak senang, atau dari setuju hingga tidak setuju.

2.1.11.3 Wawancara

Kegiatan wawancara dilakukan untuk mendapatkan informasi yang mendalam tentang persepsi, pandangan, wawasan, atau aspek kepribadian para peserta didik yang diberikan secara lesan dan spontan. Kegiatan wawancara agar lebih terarah, biasanya dilengkapi dengan pembuatan pedoman wawancara.

Wawancara yang baik adalah yang bersifat mendalam. Artinya dengan menginterpretasi jawaban siswa akan diperoleh banyak informasi, yang mungkin tidak bisa ditemukan pada penggunaan metode lainnya. [16]

2.2 Tinjauan Pustaka

1. Syaliman, Khairul Umam, Adli Abdillah Nababan, and Nadia Widari Nasution. "Pembentukan Prototype Data Dengan Metode K-Means Untuk Klasifikasi dalam Metode K-Nearest Neighbor (K-NN). (2017)" Hasil dari penelitian tersebut adalah membandingan tingkat akurasi setiap nilai K pada metode K-Means Clustering dan K-Nearest Neighbor [17]. Perbedaan dengan penelitian ini adalah melakukan pengelompokkan data berdasarkan beberapa parameter serta memberikan rekomendasi sekolah berdasarkan koordinat lokasi sekolah lain.

2. Anwar, Ade Muchlis Maulana, Prihastuti Harsani, and Aries Maesya. "PENENTUAN DAERAH PRIORITAS PELAYANAN AKTA KELAHIRAN DENGAN METODE K-NN DAN K-MEANS. (2020)". Hasil dari penelitian tersebut adalah menerapkan metode K-NN terlebih dahulu untuk mengklasifikasikan data, lalu mengelompokkan data dengan K-Means clustering [18]. Perbedaan dengan penelitian ini adalah menerapkan K-Means clustering untuk mengelompokkan data berdasarkan lokasi sekolah tersebut lalu menerapkan metode K-NN untuk memberikan rekomendasi sekolah yang terdekat dengan menggunakan koordinat sekolah tersebut.

3. Nurmahaludin, Nurmahaludin, and Gunawan Rudi Cahyono. "Klasifikasi Kualitas Air PDAM Menggunakan Algoritma KNN Dan K-Means. (2019)". Hasil dari penelitian tersebut adalah membandingkan kinerja K-NN dan K-Means dalam mengklasifikasikan kualitas air di PDAM Bandarmasih [19]. Perbedaan dengan penelitian ini adalah menerapkan K-Means clustering untuk mengelompokkan data berdasarkan lokasi sekolah tersebut lalu menerapkan metode K-NN untuk memberikan rekomendasi sekolah yang terdekat dengan menggunakan koordinat sekolah tersebut.

4. ebrinanto, Falih Gozi, Candra Dewi, and Anang Tri Wiratno. "Implementasi Algoritme K-Means Sebagai Metode Segmentasi Citra Dalam Identifikasi Penyakit Daun Jeruk.(2018)". Hasil dari penelitian tersebut adalah melakukan segmentasi daun untuk mengidentifikasi penyakit pada daun jeruk menggunakan K-Means serta penerapan K-NN untuk mengklasifikasikan penyakit yang ada pada daun jeruk tersebut [20]. Perbedaan dengan penelitian ini adalah menerapkan K-Means clustering untuk mengelompokkan data berdasarkan lokasi sekolah tersebut lalu menerapkan metode K-NN untuk memberikan rekomendasi sekolah yang terdekat dengan menggunakan koordinat sekolah tersebut.

5. Setiawan, Rony. "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru (Studi Kasus: Politeknik Lp3i Jakarta)." Jurnal Lentera Ict 3.1 (2017). Hasil dari penelitian ini adalah penerapan algoritma K-Means untuk menghasilkan kelompok mahasiswa dengan rata-rata uang pembayaran tiap kelompok. Sehingga dapat menentukan strategi promosi di Politeknik LP3I Jakarta yang didominasi masyarakat ekonomi rendah dan menengah [21]. Perbedaan dengan penelitian ini adalah menerapkan K-Means clustering untuk mengelompokkan data berdasarkan lokasi sekolah tersebut lalu menerapkan metode K-NN untuk memberikan rekomendasi sekolah yang terdekat dengan menggunakan koordinat sekolah tersebut.

6. Budiman, Ramdani. "Penerapan Data Mining Untuk Menentukan Lokasi Promosi Penerimaan Mahasiswa Baru Pada Universitas Banten Jaya (Metode K-Means Clustering)." ProTekInfo (Pengembangan Riset dan Observasi Teknik Informatika) 6.1 (2019). Hasil dari penelitian ini adalah penerapan K-Means clustering untuk mengelompokkan data persebaran mahasiswa, dimana hasil cluster terbagi menjadi 3 kelompok. Kelompok tinggi, sedang dan rendah.

Strategi promosi yang dilakukan yang tepat sasaran disetiap cluster yang telah terbentuk [22]. Perbedaan dengan penelitian ini adalah menerapkan K-Means clustering untuk mengelompokkan data berdasarkan lokasi sekolah tersebut lalu menerapkan metode K-NN untuk memberikan rekomendasi sekolah yang terdekat dengan menggunakan koordinat sekolah tersebut.

7. Handayanto, Agung, et al. "Analisis dan Penerapan Algoritma Support Vector Machine (SVM) dalam Data Mining untuk Menunjang Strategi Promosi." JUITA: Jurnal Informatika 7.2 (2019). Hasil dari penelitian ini adalah memprediksi jumlah mahasiswa yang akan mendaftar ulang pada prodi Informatika berdasarkan data sebelumnya menggunakan algoritma SVM. Hasil akurasi dari penerapan metode ini sebesar 73,6% [23]. Perbedaan dengan penelitian ini adalah menerapkan K-Means clustering untuk mengelompokkan data berdasarkan lokasi sekolah tersebut lalu menerapkan metode K-NN untuk memberikan rekomendasi sekolah yang terdekat dengan menggunakan koordinat sekolah tersebut.

8. Sugiarti, Sri, et al. "Sistem Pendukung Keputusan Penentuan Kebijakan Strategi Promosi Kampus Dengan Metode Weighted Aggregated Sum Product Assesment (WASPAS)." JURIKOM (Jurnal Riset Komputer) 5.2 (2018). Hasil dari penelitian ini adalah penerapan algoritma Weighted Aggregated Sum Product Assesment (WASPAS) pada sistem pendukung keputusan berdasarkan tingkat akurasi sebagai dasar pertimbangan untuk menentukan kebijakan strategi promosi kampus dengan tepat [24]. Perbedaan dengan penelitian ini adalah menerapkan K-Means clustering untuk mengelompokkan data berdasarkan lokasi sekolah tersebut lalu menerapkan metode K-NN untuk memberikan rekomendasi sekolah yang terdekat dengan menggunakan koordinat sekolah tersebut.

9. Ilyas, Ilyas, Fitri Marisa, and Dwi Purnomo. "Implementasi Metode Trend Moment (Peramalan) Mahasiswa Baru Universitas Widyagama Malang." JOINTECS (Journal of Information Technology and Computer Science) 3.2 (2018). Hasil dari penelitian ini adalah tingkat akurasi dengan penerapan metode Trend Moment pada kasus ini mencapai akurasi sebesar 98,25% berdasarkan data 2 tahun yaitu tahun 2016 dan 2017 [25]. Perbedaan dengan penelitian ini adalah menerapkan K-Means clustering untuk mengelompokkan data berdasarkan lokasi sekolah tersebut lalu menerapkan metode K-NN untuk memberikan rekomendasi sekolah yang terdekat dengan menggunakan koordinat sekolah tersebut.

10. Anggreini, Novita Lestari. "Teknik Clustering Dengan Algoritma K-medoids Untuk Menangani Strategi Promosi Di Politeknik Tedc Bandung." Jurnal Tekno logi Informasi dan Pendidikan 12.2 (2019). Hasil dari penelitian ini adalah penerapan algoritma K-Medoids dapat menghasilkan informasi yang dapat diusulkan ke direktur untuk kedepannya untuk mendukung promosi penerimaan mahasiswa baru menjadi lebih efisien dan efektif [26]. Perbedaan dengan penelitian ini adalah menerapkan K-Means clustering untuk mengelompokkan data berdasarkan lokasi sekolah tersebut lalu menerapkan metode K-NN untuk memberikan rekomendasi sekolah yang terdekat dengan menggunakan koordinat sekolah tersebut.

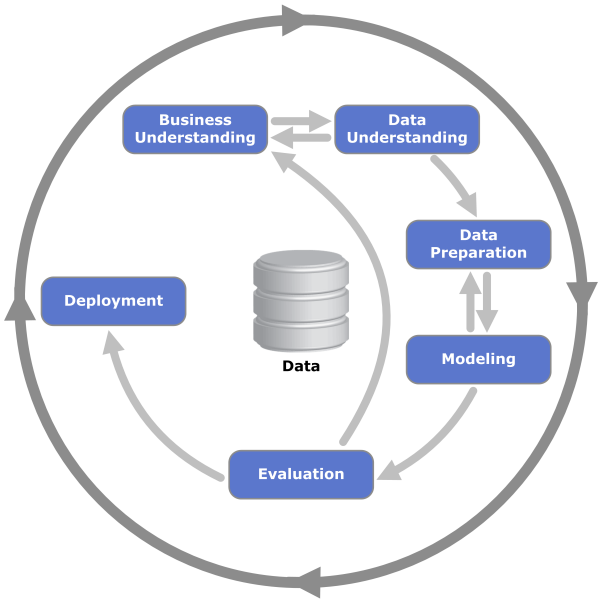
BAB III

METODOLOGI PENELITIAN

3.1 Diagram Alur Metodologi Penelitian

Metodologi penelitian merupakan cara ilmiah yang digunakan untuk mendapatkan data-data yang nantinya dapat dianalisis untuk keperluan tertentu.[27] Sedangkan penelitian adalah langkah sistematis dalam upaya memecahkan masalah untuk mengambil keputusan. [28]

Pada penelitian ini peneliti melakukan implementasi konsep big data dalam pemetaan sekolah untuk kunjungan promosi penerimaan mahasiswa baru menggunakan metode *K-Means Clustering* dan *K-Nearest Neighbor*. Untuk menyelesaikan masalah tersebut perlu diterapkannya sebuah metodologi penelitian, yaitu CRISP-DM (*Cross-Industry Standard Process for Data Mining*). Metodologi CRISP-DM adalah suatu metodologi data mining yang disusun oleh konsorsium perusahaan yang didirikan oleh Komisi Eropa pada tahun 1996 dan telah ditetapkan sebagai proses standar dalam data mining.[21]



Gambar 3.1: Diagram Alur Metodologi Penelitian

3.2 Tahapan-Tahapan Diagram Alur Metodologi Penelitian

Dalam CRISP-DM, terdapat beberapa tahapan. Namun pada penelitian ini, peneliti hanya melakukan sampai tahap *evaluation*. Karena tujuan dari penelitian ini adalah untuk mengetahui informasi terkait pemetaan sekolah untuk kunjungan promosi penerimaan mahasiswa baru (PMB).

3.2.1 *Business Understanding*

Fase pertama ini adalah proses untuk memahami tujuan dan kebutuhan dari sudut pandang bisnis, kemudian menterjemakan pengetahuan ini ke dalam pendefinisian masalah dalam data mining.[21] Pada fase ini yang dilakukan pada penelitian ini adalah bagaimana cara mengimplemtasikan konsep Big Data dalam pemetaan sekolah untuk kunjungan promosi penerimaan mahasiswa baru menggunakan metode K-Means Clustering dan K-Nearest Neighbor(KNN).

3.2.2 *Data Understanding*

Fase kedua ini dikenal sebagai fase pemahaman data, dimana dengan dimulainya proses pengumpulan data yang kemudian akan dilanjutkan dengan proses untuk mendapatkan pemahaman yang mendalam tentang data, mengidentifikasi masalah kualitas data, atau untuk mendeteksi adanya bagian yang menarik dari data yang dapat digunakan untuk hipotesa untuk informasi yang tersembunyi.[21] Terkait dengan data yang digunakan pada penelitian ini adalah data mahasiswa penerimaan mahasiswa baru dari periode 2017 sampai periode 2019 dengan atribut tahun ajaran, nama mahasiswa, jenis kelamin, penghasilan orang tua, nama perguruan tinggi, nama program studi, asal sekolah, nama wilayah, nama provinsi, koordinat X, koordinat Y.

3.2.3 *Data Preparation*

Fase ini meliputi semua kegiatan untuk membangun dataset akhir (data yang akan diproses pada tahap pemodelan/*modeling*) dari data mentah.Tahap ini dapat diulang beberapa kali. Pada tahap ini juga mencakup pemilihan tabel, record, dan atribut-atribut data, termasuk proses pembersihan dan transformasi data untuk kemudian dijadikan masukan dalam tahap pemodelan (*modeling*).[21] Pada fase ini, peneliti melakukan persiapan data seperti melakukan perhitungan jumlah mahasiswa berdasarkan provinsi, wilayah serta masing-masing sekolah.

3.2.4 *Modeling*

Fase ini akan dilakukan pemilihan dan penerapan berbagai teknik pemodelan dan beberapa parameternya akan disesuaikan untuk mendapatkan nilai yang optimal. [21] Pada fase ini, peneliti melakukan *modeling* data menggunakan aplikasi Spark yang diintegrasikan dengan jupyter notebook dan hadoop, serta dimasukkannya juga metode K-Means Clustering dan K-Nearest Neighbor.

3.2.5 *Evaluation*

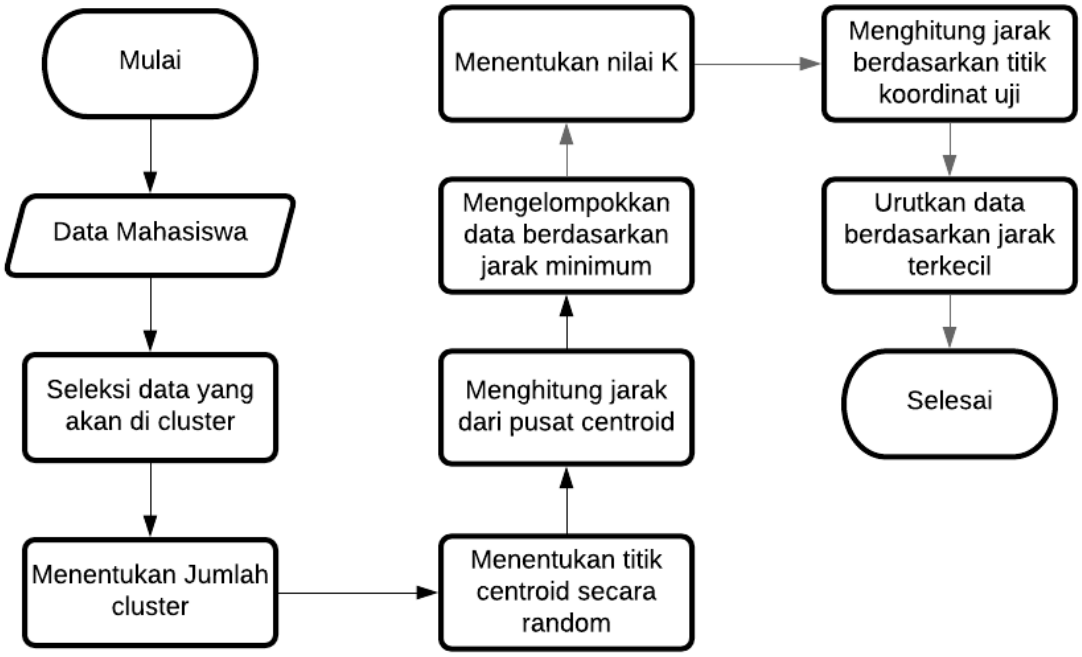
Fase ini akan dilakukan evaluasi terhadap keefektifan dan kualitas model sebelum digunakan dan menentukan apakah model dapat mencapat tujuan yang ditetapkan pada fase awal (Business Understanding).[21] Pada fase ini, peneliti melakukan proses interpretasikan data kemudian diperoleh hasil pemetaan sekolah untuk kunjungan promosi penerimaan mahasiswa baru.

3.3 Penerapan Metode K-Means Clustering dan K-Nearest Neighbor

Pada penelitian ini menggunakan metode K-Means Clustering dan K-Nearest Neighbor. Penerapan metode K-Means Clustering pada penelitian ini untuk memetakan sekolah-sekolah yang akan dikunjungi untuk melakukan promosi penerimaan mahasiswa baru. Dimana proses pemetaan ini terdapat tiga jenis, yaitu :

- Berdasarkan Provinsi, dengan atribut yang digunakan seperti nama provinsi serta jumlah mahasiswa yang diterima.
- Berdasarkan Wilayah, dengan atribut yang digunakan seperti nama provinsi, nama wilayah, serta jumlah mahasiswa yang diterima.
- Berdasarkan Sekolah, dengan atribut yang digunakan seperti nama provinsi, nama wilayah, nama sekolah, serta jumlah mahasiswa yang diterima.

Sedangkan metode K-Nearest Neighbor pada penelitian ini untuk memberikan rekomendasi sekolah terdekat dari sekolah yang akan dikunjungi berdasarkan titik koordinat dari sekolah yang akan dikunjungi serta atribut yang digunakan adalah nama provinsi, nama wilayah, nama sekolah, koordinat X, koordinat Y.



Gambar 3.2: Diagram Alur Pemetaan Sekolah

Pada gambar 3.2 merupakan tahapan dalam melakukan pemetaan sekolah untuk kunjungan promosi menggunakan algoritma K-Means Clustering dan K-Nearest Neighbor sebagai berikut :

1. Data Mahasiswa

   Pada tahap ini, adalah tahapan pengumpulan dan persiapan data mahasiswa yang akan di proses pemetaan.

2. Seleksi data yang akan di cluster

   Pada tahap ini, adalah tahapan menentukan atribut data yang akan digunakan.

3. Menentukan jumlah cluster

   Pada tahap ini, adalah tahapan menentukan jumlah cluster atau nilai K yang diinginkan.

4. Menentukan titik centroid secara random

   Pada tahap ini, adalah tahapan menentukan titik centroid secara random pada setiap cluster.

5. Menghitung jarak dari pusat centroid

   Pada tahap ini, adalah tahapan menghitung jarak minimum dari titik centroid.

6. Mengelompokkan data berdasarkan jarak minimum

   Pada tahap ini, adalah tahapan mengelompokkan data berdasarkan hasil perhitungan jarak paling minimum.

7. Menentukan Nilai K

   Pada tahap ini, adalah tahapan menentukan nilai K untuk menentukan berapa tetangga terdekat dari titik koordinat uji.

8. Menghitung jarak berdasarkan titik koordinat uji

   Pada tahap ini, adalah tahapan menghitung jarak berdasarkan titik uji koordinat.

9. Urutkan data berdasarkan jarak terkecil

   Pada tahap ini, adalah tahapan mengurutkan data berdasarkan jarak paling kecil dari titik uji koordinat.

DAFTAR PUSTAKA

[1] S. Saifuddin, I. Setiawati, and E. Ujianto, "Penerapan metode analytical hierarchy process (ahp) dalam penentuan tempat pemasangan media promosi penerimaan mahasiswa baru di universitas tunas pembangunan," 2019.

[2] M. Jamaris *et al.*, "Aplikasi analisa faktor pca dan cfa mempengaruhi minat calon mahasiswa masuk stmik amik riau untuk menentukan strategi promosi," *SATIN-Sains dan Teknologi Informasi*, vol. 5, no. 2, pp. 82–89, 2019.

[3] N. Yahya and A. Jananto, "Komparasi kinerja algoritma c. 45 dan naive bayes untuk prediksi kegiatan penerimaanmahasiswa baru (studi kasus: Universitas stikubank semarang)," 2019.

[4] I. Kurniawati, R. E. Indrajit, and M. Fauzi, "Peran bussines intelligence dalam menentukan strategi promosi penerimaan mahasiswa baru," *IKRA-ITH INFORMATIKA: Jurnal Komputer dan Informatika*, vol. 1, no. 2, pp. 70–79, 2017.

[5] S. Sutandi, "Pengaruh big data dan teknologi blockchain terhadap model bisnis sektor logistik dengan pendekatan business model canvas," *Jurnal Logistik Indonesia*, vol. 2, no. 1, pp. 9–20, 2018.

[6] I. Cholissodin and E. Riyandani, "Analisis big data," *Fakultas Ilmu Komputer (Filkom), Universitas Brawijaya (UB), Malang*, 2016.

[7] M. Hariyanto and R. T. Shita, "Clustering pada data mining untuk mengetahui potensi penyebaran penyakit dbd menggunakan metode algoritma k-means dan metode perhitungan jarak euclidean distance," *SKANIKA*, vol. 1, no. 1, pp. 117–122, 2018.

[8] G. A. Pradnyana and A. A. J. Permana, "Sistem pembagian kelas kuliah mahasiswa dengan metode k-means dan k-nearest neighbors untuk meningkatkan kualitas pembelajaran," *JUTI: Jurnal Ilmiah Teknologi Informasi*, vol. 16, no. 1, pp. 59–68, 2018.

[9] R. K. Niswatin, "Sistem pendukung keputusan penempatan jurusan mahasiswa baru menggunakan metode k-nearest neighbor," *Cogito Smart Journal*, vol. 1, no. 1, pp. 55–67, 2016.

[10] A. Nugroho, "Implentasi algoritma k-nearest neighbor dalam memprediksi potensi calon kreditur di ksp galih manunggal," *Data Manajemen dan Teknologi Informasi (DASI)*, vol. 17, no. 2, pp. 1–6, 2016.

[11] N. Grataridarga, "Konsep data, information, knowledge dan wisdom (dikw) hierarchy pada manajemen kearsipan," *JIPI (Jurnal Ilmu Perpustakaan dan Informasi)*, vol. 4, no. 1, pp. 117–127, 2019.

[12] R. Ananda and M. Fadhli, "Statistik pendidikan: teori dan praktik dalam pendidikan," 2018.

[13] S. H. S. Herho, "Tutorial pemrograman python 2 untuk pemula," 2018.

[14] A. Spark, "Apache spark," *Retrieved January*, vol. 17, p. 2018, 2018.

[15] M. Rahardjo, "Metode pengumpulan data penelitian kualitatif," 2011.

[16] B. H. Purnomo, "Metodedan teknik pengumpulan data dalam penelitian tindakan kelas (classroomaction research)," *Jurnal Pengembangan Pendidikan*, vol. 8, no. 1, 2011.

[17] K. U. Syaliman, A. A. Nababan, and N. W. Nasution, "Pembentukan prototype data dengan metode k-means untuk klasifikasi dalam metode k-nearest neighbor (k-nn)," in *Semantika (Seminar Nasional Teknik Informatika)*, vol. 1, no. 1, 2017, pp. 185–190.

[18] A. M. M. Anwar, P. Harsani, and A. Maesya, "Penentuan daerah prioritas pelayanan akta kelahiran dengan metode k-nn dan k-means," *Komputasi: Jurnal Ilmiah Ilmu Komputer dan Matematika*, vol. 17, no. 1, pp. 319–328, 2020.

[19] N. Nurmahaludin and G. R. Cahyono, "Klasifikasi kualitas air pdam menggunakan algoritma knn dan k-means," in *Seminar Nasional Riset Terapan*, vol. 4, 2019, pp. B01–B07.

[20] F. G. Febrinanto, C. Dewi, and A. T. Wiratno, "Implementasi algoritme k-means sebagai metode segmentasi citra dalam identifikasi penyakit daun jeruk," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN*, vol. 2548, p. 964X, 2018.

[21] R. Setiawan, "Penerapan data mining menggunakan algoritma k-means clustering untuk menentukan strategi promosi mahasiswa baru (studi kasus: Politeknik lp3i jakarta)," *Jurnal Lentera Ict*, vol. 3, no. 1, pp. 76–92, 2017.

[22] R. Budiman *et al.*, "Penerapan data mining untuk menentukan lokasi promosi penerimaan mahasiswa baru pada universitas banten jaya (metode k-means clustering)," *ProTekInfo (Pengembangan Riset dan Observasi Teknik Informatika)*, vol. 6, no. 1, pp. 6–14, 2019.

[23] A. Handayanto, K. Latifa, N. D. Saputro, and R. R. Waliansyah, "Analisis dan penerapan algoritma support vector machine (svm) dalam data mining untuk menunjang strategi promosi," *JUITA: Jurnal Informatika*, vol. 7, no. 2, pp. 71–79, 2019.

[24] S. Sugiarti, D. K. Nahulae, S. Syafrizal, T. E. Panggabean, and M. Sianturi, "Sistem pendukung keputusan penentuan kebijakan strategi promosi kampus dengan metode weighted aggregated sum product assesment (waspas)," *JURIKOM (Jurnal Riset Komputer)*, vol. 5, no. 2, pp. 103–108, 2018.

[25] I. Ilyas, F. Marisa, and D. Purnomo, "Implementasi metode trend moment (peramalan) mahasiswa baru universitas widyagama malang," *JOINTECS (Journal of Information Technology and Computer Science)*, vol. 3, no. 2, pp. 69–74, 2018.

[26] N. L. Anggreini *et al.*, "Teknik clustering dengan algoritma k-medoids untuk menangani strategi promosi di politeknik tedc bandung," *Jurnal Teknologi Informasi dan Pendidikan*, vol. 12, no. 2, pp. 1–7, 2019.

[27] V. H. Kristanto, *Metodologi Penelitian Pedoman Penulisan Karya Tulis Ilmiah:(KTI)*. Deepublish, 2018.

[28] S. Juliansyah Noor *et al.*, *Metodologi Penelitian: Skripsi, Tesis, Disertasi & Karya Ilmiah*. Prenada Media, 2016.