

## Analisis Potensi Asal Sekolah pada Jalur Penerimaan Mahasiswa Baru di Universitas Lampung menggunakan Algoritma *K-Means*

Wulan Rahma Izzati<sup>1</sup>, Muhamad Komarudin<sup>2</sup>, Hery Dian Septama<sup>3</sup>, Yessi Mulyani<sup>4</sup>

<sup>1</sup>Program Studi Teknik Informatika

Jurusan Teknik Elektro, Fakultas Teknik, Universitas Lampung

Jalan Soemantri Brojonegoro No.1, Bandar Lampung, 35141

<sup>1</sup>wulan.rahma1040@students.unila.ac.id

<sup>2</sup>m.komarudin@eng.unila.ac.id

<sup>4</sup>yessi.mulyani@eng.unila.ac.id

**Intisari** — Permenristekdikti nomor 126 tahun 2016 mengatur Penerimaan Mahasiswa Baru Program Sarjana pada Perguruan Tinggi Negeri, yakni Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN), Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN), dan Seleksi Mandiri. Di Universitas Lampung (UNILA), asal sekolah calon pendaftar dijadikan salah satu parameter untuk diterimanya calon mahasiswa baru dengan melihat potensi unggul dibidangnya dari setiap sekolah per tahunnya. Pada jalur SNMPTN dan SBMPTN di UNILA akan dilakukan analisis data dengan menggunakan *software tools rapidminer* serta menerapkan teknik *data mining* Algoritma *K-means* untuk melihat potensi serta konsistensi nilai diberbagai asal sekolah pada setiap tahunnya. Dari hasil perhitungan menggunakan Algoritma *K-means* diketahui jalur SBMPTN pada jenis ujian SAINTEK, SOSHUM dan IPC didominasi asal sekolah Bandar Lampung dibagi menjadi 3 kelas dimana *cluster\_0* (program studi Agroteknologi 31%), *cluster\_1* (program studi Teknik Sipil 37%), dan *cluster\_2* (Pendidikan Kedokteran 95). Pada jalur SBMPTN dengan jenis ujian SOSHUM terdiri dari *cluster\_0* (program studi Hukum 48%), *cluster\_1* (program studi Manajemen 44%), dan *cluster\_2* (program studi Akuntansi 47%). Pada jenis ujian Campuran terdiri dari *cluster\_0* (program studi Hukum sebesar 50%), *cluster\_1* (program studi Pendidikan Bahasa Inggris 23%), dan *cluster\_2* (program studi Akuntansi 38%) selama lima tahun terakhir. Pada jalur penerimaan SNMPTN, *cluster\_0* (Hukum 58%), dan *cluster\_1* (Hukum 37%) didominasi oleh asal sekolah Bandar Lampung, sedangkan pada *cluster\_2* (Pendidikan Kedokteran 46%) didominasi oleh asal sekolah kabupaten Lampung Barat selama empat tahun terakhir.

**Kata kunci** — data mining, k-means, penerimaan mahasiswa baru, SBMPTN, SNMPTN

**Abstract** — Permenristekdikti number 126 of 2016 regulates the New Student Admission for Undergraduate Program at State Universities, namely SNMPTN, SBMPTN, and Independent Selection. University of Lampung (UNILA), makes the origin of prospective applicants' schools as one of the parameters for the acceptance of prospective new students by seeing the superior potential in their fields from each school per year. The pattern of potential values based on origin school has not been analyzed in detail for the SNMPTN and SBMPTN pathways at UNILA. Data analysis will be performed using rapidminer software tools and applying K-means Algorithm data mining techniques to see the potential and consistency of values in various origin schools each year. From the calculation results using the K-means Algorithm it is known that the SBMPTN pathway in the SAINTEK, SOSHUM and IPC. dominated by the origin of Bandar Lampung schools divided into 3 classes where *cluster\_0* (Agrotechnology study program 31%), *cluster\_1* (Civil Engineering study program 37%), and *cluster\_2* (Medical Education 95 48%), *cluster\_1* (Management study program 44%), and *cluster\_2* (Accounting study program 47%) .The types of Mixed examinations consist of *cluster\_0* (Law study program by 50%), *cluster\_1* (English Language Education program 23%), and *cluster\_2* (Accounting study program 38%) for the last five years, on the admission path of SNMPTN, *cluster\_0* (Law 58%), and *cluster\_1* (Law 37%) dominated by the origin of Bandar Lampung schools, while in *cluster\_2* (Medical Education 46%) dominated by the origin of the West Lampung district school for the past four years.

**Keywords**— data mining, k-means, acceptance of new students, SBMPTN, SNMPTN

## I. PENDAHULUAN

Pada peraturan Menteri Riset, Teknologi, dan Pendidikan Tinggi nomor 126 tahun 2016 tentang Penerimaan Mahasiswa Baru Program Sarjana pada Perguruan Tinggi Negeri, jalur penerimaan mahasiswa baru dilakukan melalui Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN), Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN), dan Seleksi Mandiri.

Pada seleksi ini calon mahasiswa baru berhak memilih program studi yang diinginkan sesuai dengan penggolongan jurusan yaitu Sains dan Teknologi (SAINTEK), Sosial dan Humaniora (SOSHUM), atau Ilmu Pengetahuan Campuran (IPC). UNILA (Universitas Lampung) setiap tahunnya membuka penerimaan mahasiswa baru diseluruh program studi yang tercatat sebanyak 117.231 pendaftar pada jalur seleksi SBMPTN dan pada jalur SNMPTN sebanyak 107.530 selama lima tahun terakhir. Jumlah lolos seleksi pada jalur SBMPTN sebanyak 12.515 dan pada jalur SNMPTN sebanyak 8.728 selama lima tahun terakhir. Jumlah tersebut menunjukkan, keketatan di Universitas Lampung cukup tinggi.

Keketatan menggambarkan semakin besar peminat pada suatu program studi maka keketatannya semakin tinggi, dikarenakan, persaingan antara calon pendaftar pun semakin besar. Potensi dari setiap sekolah bisa dikenali polanya untuk menemukan potensi unggul dibidangnya setiap tahun. Jika analisis data penerimaan pada jalur SNMPTN dan SBMPTN tersebut dilakukan secara manual akan membutuhkan waktu yang cukup lama.

Sehingga dengan memanfaatkan kemajuan teknologi yang berkembang dapat membantu menganalisis data yang besar (*Big Data*) dengan menggunakan *software tools rapidminer*, serta menerapkan teknik data mining Algoritma *K-means* untuk memberikan sebuah pola potensi nilai berdasarkan asal sekolah mahasiswa baru lolos seleksi pada Universitas Lampung pada tiap tahunnya.

## II. TINJAUAN PUSTAKA

### A. Jalur Penerimaan

Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN) merupakan penerimaan mahasiswa baru program sarjana pada Perguruan Tinggi Negeri (PTN) yang dilaksanakan oleh seluruh Perguruan Tinggi Negeri (PTN) dalam suatu sistem yang terpadu dan serentak secara nasional.

Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN) merupakan sistem ujian saringan masuk perguruan tinggi negeri yang dilaksanakan secara nasional. Ujian tertulis yang diselenggarakan dalam rangka SBMPTN akan dilaksanakan secara serentak dan terpadu pada jam dan hari yang sama, dengan soal yang sama di berbagai wilayah di Indonesia.

### B. Data Mining

*Data Mining* merupakan sebuah cara untuk menemukan sebuah pola yang dimunculkan dalam database dengan ukuran besar dimana cara ini dapat mengambil sebuah keputusan dalam waktu yang cepat. *Data mining* merupakan istilah yang sering dikatakan sebagai suatu metode untuk meminimalisir serta mencari informasi baru didalam suatu *database*. *Data mining* adalah proses pemilihan atau “menambang” pengetahuan dari sekumpulan data dalam jumlah yang banyak [1]. Secara sudut pandang keilmuan *data mining* dimanfaatkan untuk merekam, menganalisis serta penyimpanan data secara *real time* dengan jumlah yang besar.

### C. K-Means

Metode *K-Means* adalah metode yang termasuk dalam algoritma *clustering* berbasis jarak yang membagi data ke dalam sejumlah *cluster* dan algoritma ini hanya bekerja pada atribut numerik [2]. Metode *K-Means* dapat mempartisi data ke dalam *cluster* dimana data yang dimiliki dapat memiliki karakteristik yang berbeda dan dikelompokkan kedalam *cluster* yang lain [3]. Ide dasar algoritma *K-Means* cukup sederhana, yaitu meminimalkan *Sum of Squared Error* (SSE)

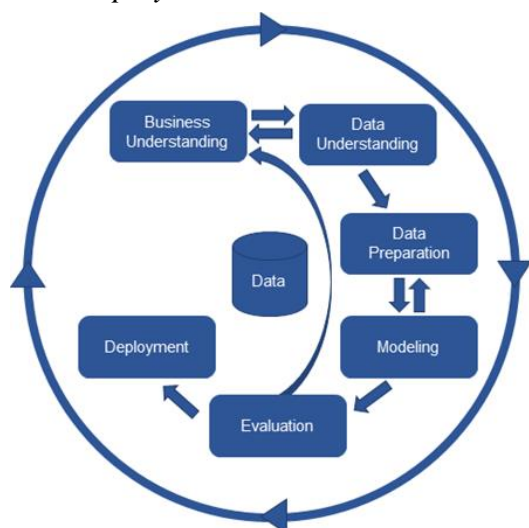
antara objek-objek data dengan jumlah  $k$  *centroid*.

Suatu data akan menjadi anggota dari *cluster* ke- $k$  apabila jarak data tersebut ke pusat *cluster* ke- $k$  bernilai paling kecil jika dibandingkan dengan jarak ke pusat *cluster* lain.

#### D. Cross Industry Standart Process for Data Mining (CRISP-DM)

CRISP-DM merupakan standarisasi data mining yang tersusun pada 3 bagian pengagas data mining yaitu SPSS (ISL), NCR dan Daimler Chrysler (Daimler-Benz) yang dikembangkan tahun 1996 oleh analis dari beberapa industri menetapkan sebagai proses standar strategi pemecahan masalah *data mining* untuk penelitian. Dalam metode ini terdapat 6 tahapan yaitu:

1. *Business Understanding*
2. *Data Understanding*
3. *Data Preparation*
4. *Modeling data*
5. *Evaluation*
6. *Deployment*

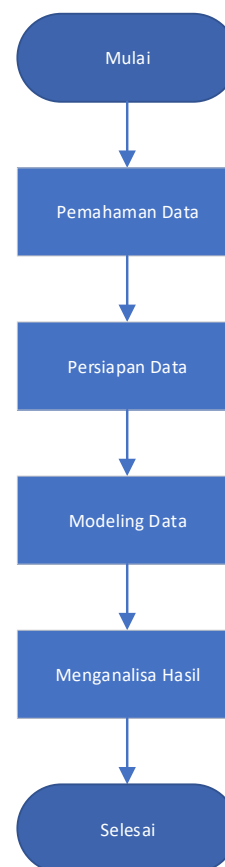


Gbr. 1 Siklus pada metode CRISP-DM

### III. METODOLOGI

Penelitian ini menggunakan metode *Cross Industry Standart Process for Data Mining* (CRISP-DM) pada pemodelan Algoritma K-Means untuk fase pengolahan data, yaitu Pemahaman Data, Persiapan Data,

Modeling Data, dan Menganalisa Hasil seperti pada Gbr. 2 berikut.



Gbr. 2 Diagram Alir Penelitian

#### A. Pemahaman Data

Data yang didapatkan dari BPPMB Universitas Lampung untuk penerimaan mahasiswa baru pada jalur SNMPTN sebanyak 8.728 record data dan *record* data SBMPTN sebanyak 12.515 *record* data selama lima tahun terakhir data siswa yang hanya terima yaitu pada tahun 2013-2017 dengan total 21.243 sebagai bahan uji penelitian.

#### B. Persiapan Data

Kegiatan ini dilakukan meliputi pemilihan data (*record*, *table* dan atribut yang akan dipakai), transformasi data dan pembersihan data (*Data Cleansing*). Variabel yang akan dijadikan acuan adalah asal sekolah (SMA), nilai raport, nilai ujian nasional, serta nilai tes untuk jalur penerimaan SBMPTN. Serta pengelompokan data SAINTEK, SOSHUM,

dan IPC. Pemisahan data yang diterima berdasarkan jalur ilmu pada jalur SBMPTN guna untuk langkah awal dalam melakukan *dataset* awal.

### C. Modeling Data

Fase penerapan metode merupakan fase memulai mengerjakan aktivitas pemilihan teknik dari *data mining* yang akan diterapkan pada *dataset* yang sudah dilakukan sebelumnya, serta mengkalibrasi parameter teknik *data mining* tersebut dengan menggunakan nilai yang optimal. Teknik *data mining* yang digunakan memungkinkan melakukan fase persiapan data secara berulang, hal ini dilakukan untuk mendapatkan kebutuhan yang lebih spesifik. Terdapat 2 pemodelan yang akan dilakukan pada penelitian ini yaitu:

#### 1) Pemodelan data pada jalur SBMPTN



Gbr 3. Pemodelan Data Jalur SNMPTN

Pada **Error! Reference source not found.** merupakan pemodelan data yang akan diterapkan setelah melakukan *preparation data* SBMPTN lalu melakukan penganalisisan pada setiap nilai pada setiap bidang uji mahasiswa baru. Dengan jumlah nilai yang ditentukan, pada bidang ilmu SAINTEK skor rata-rata minimal 520,8 dan skor rata-rata maksimal 724,2. Pada bidang SOSHUM skor rata-rata minimal 522,7 dan skor rata-rata maksimal 689,2.

#### 2) Pemodelan data pada jalur SNMPTN



Gbr. 4 Pemodelan Data SNMPTN

Pada Gbr. merupakan pemodelan data yang akan diuji pada tahap pengujian pada jalur SNMPTN data sekolah yang dijadikan tolak ukur untuk diterimanya calon mahasiswa baru dilihat dari asal sekolah, nilai SBMPTN pada tahun sebelumnya serta nilai Indeks Prestasi (IP) mahasiswa semester satu dan tiga pada mahasiswa terdaftar. Nilai rapor calon mahasiswa semester satu sampai dengan lima, serta piagam prestasi jika ada. Dengan masing-masing persentasi sebagai berikut:



Gbr. 5 Pemodelan Data SNMPTN

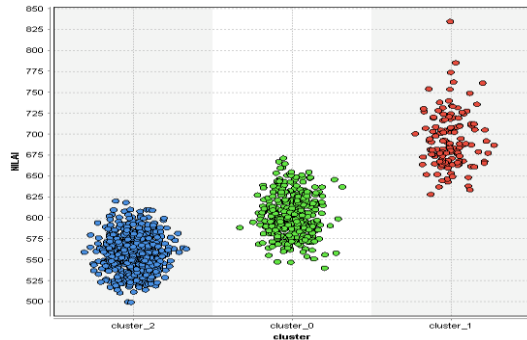
### D. Menganalisa Hasil

Data yang telah terekam dan diambil selanjutnya akan dianalisa dari pola setiap asal sekolah dengan nilai yang ada pada setiap jalur penerimaan di Universitas Lampung. Setelah mendapat wawasan dari data tersebut dapat menjadi saran kepada BPPMB dalam meningkatkan kualitas mahasiswa baru yang akan diterima pada tahun mendatang di Universitas Lampung.

## IV. HASIL DAN PEMBAHASAN

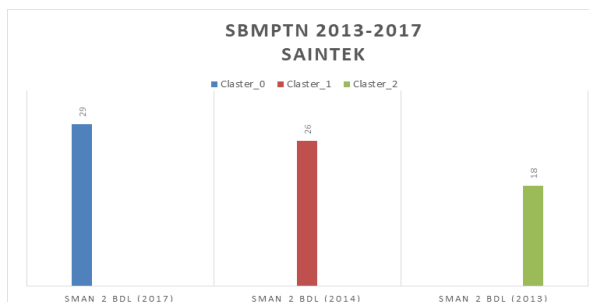
### A. SBMPTN

*Dataset* yang digunakan pada penelitian ini di bagi menjadi 3 bagian yaitu data SAINTEK, SOSHUM, dan IPC dari TAHUN 2013 sampai dengan tahun 2017 pada data penerimaan mahasiswa baru pada jalur SBMPTN.



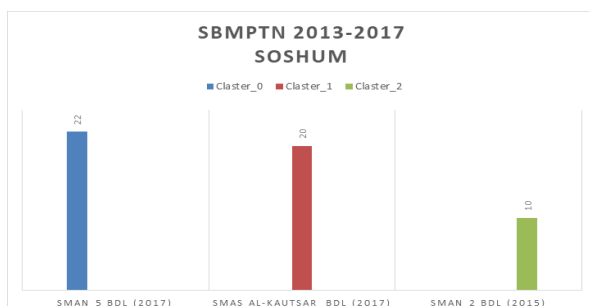
Gbr. 6 Hasil Cluster pada rapidminer

Hasil diatas merupakan hasil pengelompokan nilai berdasarkan titik *centroid*, kemudian dicari nilai yang memiliki tingkat kesamaan yang paling mirip atau paling mendekati, maka didapatkan hasil pengelompokan asal sekolah berdasarkan cluster. Pada hasil ini diambil 3 sekolah disetiap *cluster* yang paling tertinggi atau terbanyak, seperti pada gambar berikut.



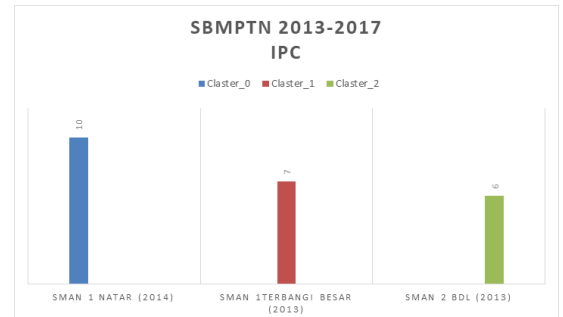
Gbr. 7 Hasil Cluster SBMPTN SAINTEK 2013-2017

Gambar diatas merupakan *highlight* dari setiap *cluster* di jenis ujian SBMPTN dengan jenis ujian SAINTEK, SMAN 2 BDL memiliki 29 siswa terima pada *cluster\_0*, 1 dan 2 pada tahun 2017, 26 siswa pada tahun 2014 dan 21 siswa pada tahun 2013.



Gbr. 8 Hasil Cluster SBMPTN SOSHUM 2013-2017

Gambar diatas merupakan *highlight* dari setiap *cluster* di jenis ujian SBMPTN dengan jenis ujian SOSHUM, SMAN 5 BDL memiliki 22 siswa terima pada *cluster\_0* tahun 2017, SMAS Al-Kautsar 20 siswa terima pada *cluster\_1* tahun 2017, SMAN 2 BDL sebanyak 10 siswa pada tahun 2015 pada *cluster\_2*.



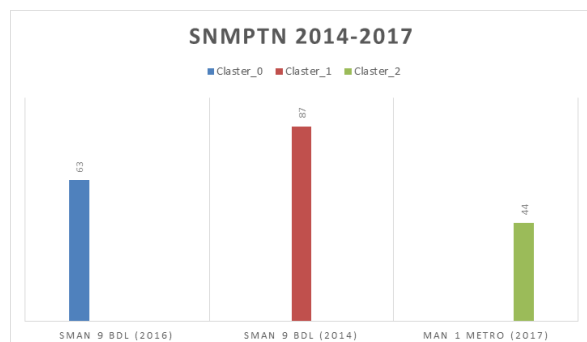
Gbr. 9 Hasil Cluster SBMPTN IPC 2013-2017

Gambar diatas merupakan *highlight* dari setiap *cluster* di jenis ujian SBMPTN dengan jenis ujian IPC, SMAN 1 Natar memiliki 10 siswa terima pada *cluster\_0* tahun 2014, SMAN 1 Terbanggi Besar 7 siswa terima pada *cluster\_1* tahun 2013, SMAN 2 BDL sebanyak 6siswa pada tahun 2013 pada *cluster\_2*.

### B. SNMPTN

Pada jalur SNMPTN menggunakan alur yang sama dengan SBMPTN namun data yang digunakan hanya pada tahun 2014 hingga 2017 hal ini dikarenakan kurangnya sumber daya manusia dalam menangani data pada tahun sebelumnya, pada jalur SNMPTN menggunakan jumlah dari nilai rapor yang terdiri dari mata pelajaran Matematika, Bahasa Indonesia, Bahasa Inggris, Fisika, Biologi dan Kimia, Geografi, Sosiologi, Ekonomi dan Sejarah dimana nilai tersebut dijumlahkan dan diolah untuk mendapatkan hasil.

Pada SNMPTN tetap dibagi pada tiga kelas, yaitu nilai rendah atau *cluster\_0*, nilai sedang untuk *cluster\_1* dan nilai tinggi untuk *cluster\_2*.



Gbr. 10 Hasil Claster SNMPTN 2014-2017

Gambar diatas merupakan *highlight* dari setiap *cluster* di jenis ujian SAINTEK, SMAN 9 BDL memiliki 63 siswa terima pada *cluster\_0* tahun 2016 dan 87 siswa pada tahun 2014 di *cluster\_1*, MAN 1 Metro sebanyak 44 siswa pada tahun 2017 di *cluster\_2*.

## V. KESIMPULAN

Berdasarkan hasil penelitian diatas, pada jalur SBMPTN dibagi menjadi 3 kelas dimana *cluster\_0*, *cluster\_1* dan *cluster\_2* adalah didominasi oleh asal sekolah Bandar Lampung dengan sebaran program studi terbanyak yaitu Agroteknologi (31%), Teknik Sipil (37%) dan Pendidikan Kedokteran (95%) selama lima tahun terakhir pada jenis ujian SAINTEK.

Pada jenis ujian SOSHUM dengan pengelompokan didominasi oleh asal sekolah Bandar Lampung dengan sebaran program studi terbanyak yaitu Hukum (48%), Manajemen(44%), dan Akuntansi (47%)

selama lima tahun terakhir. Dan pada jenis ujian IPC dengan pengelompokan didominasi oleh asal sekolah Bandar Lampung dengan sebaran program studi terbanyak yaitu Hukum (50%), Pendidikan Bahasa Inggris (23%), dan Akuntansi (38%) selama lima tahun terakhir.

Pada jalur penerimaan SNMPTN *cluster* yang dibagi menjadi 3 *class* dimana *cluster\_0* dan *cluster\_1* yang didominasi oleh asal sekolah bandar lampung dengan sebaran progam studi terbanyak yaitu Hukum (58%), Pada *cluster\_2* daerah asal sekolah yang diterima terbanyak pada *class* ini berasal dari daerah kabupaten Lampung Barat dengan sebaran prodi yaitu Pendidikan Kedokteran (46%) selama empat tahun terakhir.

## REFERENSI

- [1] W. Lefebvre-Ulrikson, G. Da Costa, L. Rigutti, and I. Blum, "Data Mining," *Atom Probe Tomogr. Put Theory Into Pract.*, no. June 2002, pp. 279–317, 2016.
- [2] Kusrini and E. Taufiq Luthfi, *Algoritma Data Mining*. Yogyakarta: Andi, 2009.
- [3] Y. S. Nugroho and Setyawan, "Klasifikasi Masa Studi Mahasiswa Fakultas Komunikasi Dan Informatika Universitas Muhammadiyah Surakarta Menggunakan Algoritma C4.5," *KomuniTi*, no. March 2014, pp. 84–91, 2014.