

BAB I

PENDAHULUAN

1.1 Latar Belakang

Salah satu tujuan negara Indonesia adalah memajukan kesejahteraan rakyat, seperti yang tertuang dalam alinea IV Pembukaan UUD 1945. Kesejahteraan rakyat berkaitan dengan pembangunan yang dilakukan oleh pemerintah. Dalam melakukan pembangunan pada suatu daerah, pemerintah harus mengetahui kondisi kesejahteraan rakyat pada daerah tersebut dikarenakan kesejahteraan rakyat di setiap daerah berbeda-beda.

Kesejahteraan pada setiap daerah dapat dilihat berdasarkan indikator-indikator yang mampu mengukur kesejahteraan rakyat dan dapat dijadikan sebagai ukuran tercapai atau tidaknya pembangunan di suatu daerah. Indikator kesejahteraan rakyat cukup luas sehingga suatu taraf kesejahteraan rakyat hanya dapat dinilai melalui indikator-indikator terukur dari berbagai aspek pembangunan.

Dalam publikasi data Badan Pusat Statistik disebutkan bahwa tingkat kesejahteraan rakyat Indonesia dapat dilihat dari berbagai aspek kesejahteraan diantaranya kependudukan, kesehatan, pendidikan, ketenagakerjaan dan kemiskinan [1]. Dari berbagai aspek kesejahteraan tersebut, terdapat indikator-indikator yang mencerminkan tingkat kesejahteraan rakyat. Indikator ini selanjutnya dapat digunakan untuk melihat kemiripan kondisi kesejahtera-

an rakyat di suatu daerah dengan daerah lain sehingga dapat dilakukan pengelompokan daerah guna membantu pemerintah dalam menyusun dan menentukan prioritas pembangunan.

Dalam mengelompokkan daerah atau provinsi di Indonesia dapat menggunakan analisis *cluster*. Analisis *cluster* merupakan analisis statistika yang digunakan dalam pengelompokan objek. Tujuan analisis ini yaitu mengelompokkan objek-objek pengamatan ke dalam beberapa *cluster* berdasarkan peubah yang diamati, sehingga objek-objek yang berada dalam *cluster* yang sama lebih mirip dibandingkan dengan objek pada *cluster* yang berbeda.

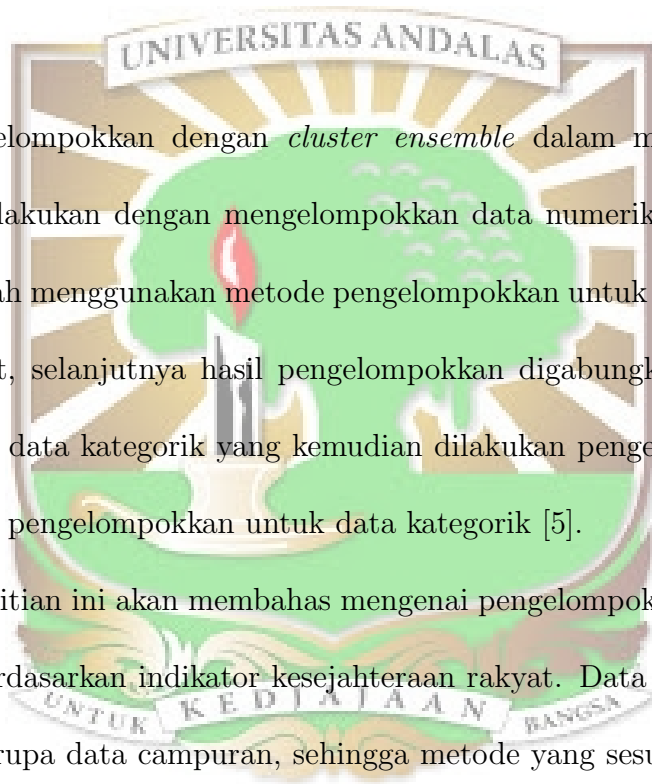
Untuk data numerik, pengelompokan objek dapat menggunakan metode hirarki dan non hirarki. Metode hirarki digunakan bila peneliti tidak menentukan jumlah *cluster* yang diinginkan, sedangkan metode non hirarki digunakan bila jumlah *cluster* ditentukan oleh peneliti. Untuk pengelompokan data kategorik, metode hirarki dan non hirarki juga dapat digunakan, selain itu metode lain yang dapat digunakan yaitu metode ROCK (*Robust Clustering Using Links*). Metode ROCK dikembangkan oleh Guha, Rastogi, dan Shim pada tahun 1999. Metode ini menggunakan suatu ukuran kemiripan yang disebut *link* dalam membentuk *cluster*[3].

Metode lain yang juga dapat digunakan dalam pengelompokan objek yaitu *cluster ensemble* yang pertama kali diperkenalkan oleh Strehl dan Gosh pada tahun 2002. Prinsip metode ini yaitu menggabungkan hasil pengelompokan dari berbagai metode pengelompokan yang berbeda sehingga memperoleh hasil pengelompokan akhir yang lebih baik [9].

Permasalahan yang sering ditemui dalam analisis *cluster* adalah data yang digunakan bertipe campuran (numerik dan kategorik). Metode yang seringkali digunakan dalam pengelompokan data campuran adalah dengan mentransformasikan data kategorik menjadi numerik atau sebaliknya. Selain metode tersebut pada tahun 2005, He Xu dan Deng mengembangkan metode *cluster ensemble* yang dapat digunakan untuk mengelompokkan data campuran [5].

Pengelompokkan dengan *cluster ensemble* dalam menangani data campuran dilakukan dengan mengelompokkan data numerik dan kategorik secara terpisah menggunakan metode pengelompokkan untuk masing-masing data tersebut, selanjutnya hasil pengelompokkan digabungkan dan dipandang sebagai data kategorik yang kemudian dilakukan pengelompokkan dengan metode pengelompokkan untuk data kategorik [5].

Penelitian ini akan membahas mengenai pengelompokkan provinsi di Indonesia berdasarkan indikator kesejahteraan rakyat. Data indikator kesejahteraan berupa data campuran, sehingga metode yang sesuai untuk digunakan yaitu metode *cluster ensemble* dengan pengelompokkan data numerik menggunakan metode hirarki, sedangkan pengelompokkan data kategorik dilakukan dengan metode ROCK. Metode ROCK ini selanjutnya akan digunakan lagi pada tahap penggabungan.



1.2 Rumusan Masalah

Berdasarkan latar belakang, maka rumusan masalah pada penelitian ini adalah bagaimana menerapkan analisis *cluster ensemble* dengan metode ROCK untuk mengelompokkan provinsi di Indonesia berdasarkan indikator kesejahteraan rakyat.

1.3 Batasan Masalah

Pada penelitian ini data indikator kesejahteraan rakyat yang digunakan merupakan data tahun 2017 dengan indikator yang digunakan yaitu persentase penduduk miskin, tingkat pengangguran terbuka, angka melek huruf, angka partisipasi sekolah, angka harapan hidup, angka morbiditas, kepadatan penduduk dan indeks gini.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah menerapkan analisis *cluster ensemble* dengan metode ROCK untuk mengelompokkan provinsi di Indonesia berdasarkan indikator kesejahteraan rakyat.

1.5 Sistematika Penulisan

Penulisan ini terdiri dari lima bab. BAB I Pendahuluan yang memuat latar belakang, rumusan masalah, batasan masalah, tujuan penulisan, dan sistematika penulisan. BAB II Landasan Teori yang menjelaskan teori-teori



mengenai analisis *cluster*, metode pengelompokan, *cluster ensemble*, dan evaluasi hasil pengelompokan. BAB III Metode Penelitian yang berisi sumber data, peubah yang digunakan dalam penelitian, dan teknik analisis data. BAB IV Hasil dan Pembahasan yang berisi gambaran umum data dan hasil pengelompokan data. BAB V Penutup yang berisi kesimpulan dan saran dari penelitian.

