

I. PENDAHULUAN

1.1 Latar Belakang

Pembangunan manusia merupakan hal yang bersifat fundamental dalam paradigma pembangunan. Kajian mengenai pembangunan manusia mulai sering diperbincangkan dalam beberapa waktu terakhir di saat pembangunan manusia menjadi isu kebijakan utama dalam negara maju, maupun negara berkembang (Ali, 2010). Hal yang sama juga diutarakan oleh Badan Pusat Statistik (2017) yang menyatakan pembangunan manusia merupakan isu penting dalam berbagai perancangan dan strategi pembangunan yang berkelanjutan. Namun secara garis besar, konsep pembangunan manusia cukup sederhana, yaitu menganggap manusia itu sendiri sebagai tujuan akhir dari suatu pembangunan, bukan sebagai alat dari suatu pembangunan. Pada dasarnya manusia dianggap sebagai kekayaan bangsa yang sesungguhnya (Human Development Report (HDR), 2016). Berbekal dengan konsep tersebut, maka tujuan dari pembangunan manusia adalah dapat menciptakan suatu lingkungan dimana rakyatnya dapat menikmati hidup panjang dan sehat, serta dapat menjalani kehidupan yang produktif. Sehingga pembangunan manusia dapat dijadikan sebagai tolak ukur dari berhasilnya berbagai program pemerintah yang dilakukan untuk meningkatkan kualitas sumber daya manusianya.

United Nations Development Programme (UNDP) merupakan lembaga kemitraan untuk tujuan pembangunan berkelanjutan *development goals*. UNDP telah menerbitkan suatu indikator yaitu Indeks Pembangunan Manusia (IPM). IPM merupakan indeks komposit yang terdiri dari indeks kesehatan dengan komponen Angka Harapan Hidup (AHH), indeks pendidikan dengan komponen Rata-rata Lama Sekolah (RLS) dan Harapan Lama Sekolah (HLS), dan indeks kesejahteraan dengan komponen Pengeluaran Per Kapita (PPP). Berdasarkan indeks komposit tersebut, BPS membentuk indikator IPM menjadi 4 aspek yaitu Angka Harapan Hidup (AHH), Rata-rata Lama Sekolah (RLS), Harapan Lama Sekolah (HLS) dan pengeluaran per kapita (PPP) (Badan Pusat Statistik, 2021) .

Manfaat pentingnya IPM antara lain adalah IPM merupakan indikator penting dalam mengukur keberhasilan dalam membangun kualitas hidup manusia serta IPM dapat menentukan peringkat atau tingkat pembangunan suatu wilayah negara (Badan Pusat Statistik, 2015). Peningkatan IPM tidak hanya melalui pertumbuhan ekonomi saja, tetapi juga harus diikuti dengan pembangunan manusia baik dibidang kesehatan maupun pendidikan. Tingginya

tingkat pendidikan berhubungan dengan tingginya upah yang diperoleh, sehingga akan berpengaruh pada daya beli masyarakat. Pertumbuhan ekonomi dan pembangunan manusia harus selaras dan sejalan melalui pemerataan pembangunan. Saat ini, pemerataan pendidikan guna mempersiapkan masyarakat yang mampu bersaing di era globalisasi dan keterbukaan tak terkecuali di wilayah Sumatera Barat (Purnamasari & Hayati, 2018).

Pertumbuhan IPM merupakan salah satu cara untuk melihat perkembangan pembangunan manusia di suatu wilayah. IPM di Provinsi Sumatera Barat dari tahun 2017- 2021 mengalami pertumbuhan rata-rata 0,54 persen per tahun yang menunjukkan bahwa semakin membaiknya pembangunan manusia secara umum di Provinsi Sumatera Barat. Pada tahun 2021, IPM provinsi Sumatera Barat adalah 72,65, Angka ini meningkat sebesar 0,27 poin atau lebih tinggi sebesar 0,27 persen dibanding tahun 2020 sebesar 72,38. Hal ini merupakan kenaikan kesembilan dalam 10 tahun terakhir dibanding capaian IPM di tahun-tahun sebelumnya. Pada tahun 2020 terjadinya penurunan IPM di Provinsi Sumatera Barat, hal ini dikarenakan Pandemi COVID-19 khususnya pada komponen pengeluaran per kapita. Kemudian ditahun 2021 seiring dengan pemulihan ekonomi sehingga semua komponen IPM menunjukkan kenaikan perbaikan dibanding tahun sebelumnya (BPS, 2021).

Kenaikan nilai IPM yang cukup signifikan setiap tahunnya belum tentu dapat menggambarkan bahwa pembangunan manusia di Sumatera Barat sudah merata, sehingga harus dilakukan sebuah tindakan untuk mengidentifikasi kabupaten/kota mana saja yang mengalami kendala, baik dalam pendidikan, kesehatan maupun kesejahteraan yang mengakibatkan kabupaten/kota tersebut mengalami ketertinggalan dari kabupaten/kota yang lain. Ketertinggalan tersebut sebagai pemicu ketidakmerataan pembangunan manusia di wilayah Sumatera Barat. Pengidentifikasi tersebut dapat dilakukan dengan pengelompokan kabupaten/kota yang ada di wilayah Sumatera Barat berdasarkan indikator IPM.

Secara umum, pengelompokan suatu objek atau dalam hal ini pengelompokan kabupaten/kota di wilayah Sumatera Barat dapat dilakukan dengan menggunakan konsep matematika. Salah satu disiplin ilmu statistika yang dapat digunakan untuk mengelompokkan variabel atau objek adalah *clustering*. *Clustering* merupakan salah satu analisis peubah ganda (*analysis multivariate*) yang dapat digunakan untuk mengelompokkan data observasi ataupun variabel-variabel kedalam kelompok sedemikian rupa sehingga masing-masing kelompok bersifat homogen sesuai dengan variabel yang

digunakan untuk melakukan *clustering* dan heterogen antar kelompok yang terbentuk (Gudono, 2011).

Clustering memiliki jenis yang disebut *clustering* hierarki dan *clustering* non hierarki memiliki kelebihan yaitu (1) Dapat mengelompokkan data observasi dalam jumlah besar dan variabel yang relatif banyak. (2) Dapat dipakai dalam skala ordinal, interval dan rasio. Sedangkan dari segi kelemahan metode hierarki dan non hierarki yaitu (1) Pengelompokan bersifat subjektifitas peneliti karena hanya melihat dari gambar *dendogram*. (2) Data heterogen antara objek penelitian yang satu dengan yang lain akan sulit bagi peneliti untuk menentukan jumlah kelompok yang dibentuk. (3) Metode-metode yang dipakai memberikan perbedaan yang signifikan, sehingga dalam perhitungan biasanya masing-masing metode dibandingkan. (4) Semakin besar observasi, biasanya tingkat kesalahan akan semakin besar (Anggraini & Arum, 2022). Kelemahan yang terdapat pada metode hierarki dengan berbagai metode perbaikan jarak dan non hierarki seperti *k-means* dengan berbagai pemilihan pusat kelompok awal (Yetti dkk (2019). Tantangan untuk menutupi kekurangan dari metode hierarki dan non hierarki mendapatkan solusi *cluster* dengan kualitas yang baik serta adanya keragaman solusi *cluster* yang dihasilkan dari metode yang berbeda merupakan motivasi dikembangkannya *clustering ensemble*.

Clustering ensemble adalah salah satu metode yang digunakan untuk mengombinasikan berbagai solusi *clustering* tanpa melihat karakteristik data awal (Millati et al, 2021). *Clustering ensemble* memiliki keunggulan dibanding metode *cluster* yang lainnya seperti hierarki dan non hierarki yaitu mampu meningkatkan kualitas yang baik dan solusi *cluster* tidak sensitif terhadap *outlier*, dan *noise* terhadap Hasil *clustering* yang dikombinasikan dengan metode sebelumnya dalam penentuan pengamatan (*case*) menggunakan metode non hierarki yaitu *k-means*. *K-means* merupakan metode analisis *cluster* non hierarki yang sederhana, tidak rumit dan memudahkan dalam *generative mechanism* atau menentukan anggota-anggota *clustering* untuk analisis *cluster ensemble* tersebut.

Penelitian relevan yang dilakukan Suhaeni et al (2018) yang membandingkan performa *cluster* berhierarki, *k-means* dan *cluster ensemble* pada studi kasus indikator pelayanan kesehatan ibu hamil. Metode berhierarki yang digunakan adalah *single linkage*, *complete linkage*, *median linkage*, *average linkage* dan *ward linkage*. Sedangkan untuk *ensemble* menggunakan partisi yang diperoleh dari metode berhierarki dan *k-means* sebagai *generative mechanism*. Penelitian tersebut menyatakan bahwa performa *clustering ensemble* lebih baik dibandingkan *clustering* berhierarki maupun *k-means*.

Selanjutnya terdapat penelitian lainnya yang membandingkan performa *clustering ensemble* dengan *generative mechanism* dan fungsi konsensus yang berbeda-beda pada berbagai data yaitu *IS*, *Ionosphere*, *Iris*, *PBC*, *SVS* dan *wine*. Penelitian tersebut dilakukan oleh Wu dkk (2018), yang membandingkan 12 algoritma yang berbeda yaitu *generative mechanism* berupa kriteria algoritma *k-means* dengan subset yang berbeda, *k-means* dengan *subset* yang berbeda, serta membandingkan fungsi konsensus yaitu *single linkage*, *complete linkage*, *average linkage*, *k-means*, *fuzzy C-means* dan *self organizing clustering*. Penelitian tersebut menghasilkan bahwa *clustering ensemble* dengan *generative mechanism* berupa *k-means* dengan inisiasi nilai pusat awal berbeda serta fungsi konsensus berupa *average linkage* adalah algoritma *clustering* terbaik.

Penelitian terkait *clustering ensemble* lainnya juga pernah dilakukan oleh Hadjitodorov et al (2016) dengan membandingkan beberapa model pada 13 *data set* yang berbeda, kemudian diperoleh hasil bahwa *clustering ensemble* berdasarkan *k-means* merupakan solusi *clustering ensemble* terbaik. Pada penelitian tersebut diperoleh juga kesimpulan bahwa fungsi konsensus CSPA lebih baik daripada konsensus pautan tunggal.

Berdasarkan pemaparan masalah sebelumnya, permasalahan yang terjadi terkait dengan perkembangan dan kemajuan dari suatu wilayah yang harus diatasi. Perhitungan IPM merupakan indeks komposit yang terdiri dari 4 indikator yaitu Angka Harapan Hidup (AHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS) dan pengeluaran per kapita riil yang disesuaikan (PPP). Secara tidak langsung dapat diketahui bahwa keempat indikator inilah yang mempengaruhi pencapaian IPM baik nasional, provinsi maupun kabupaten/kota yang dikelompokkan menjadi 4 kriteria yaitu sangat tinggi, tinggi, sedang dan rendah. Oleh karena itu akan lebih tepat jika dilakukan penelitian dengan mengelompokkan wilayah berdasarkan keempat indikator IPM, maka akan diperoleh hasil kelompok wilayah ini pemerintah dapat menentukan daerah mana saja yang akan diprioritaskan mendapatkan bantuan dari pemerintah dan penetapan atau pembaharuan dari kebijakan pemerintah dengan memperhatikan keempat aspek tersebut yang disesuaikan dengan karakteristik wilayah. Oleh karena itu, peneliti ingin menggunakan *clustering ensemble* tersebut karena dapat digunakan pada berbagai kondisi yang tetap *robust* serta lebih efisien dengan performa yang lebih baik. Hal inilah yang membuat penulis ingin menyusunnya dalam sebuah penelitian yang berjudul **“PERBANDINGAN PENGELOMPOKAN PROVINSI SUMATERA BARAT BERDASARKAN INDEKS PEMBANGUNAN MANUSIA MENGGUNAKAN METODE CLUSTERING ENSEMBLE DAN K-MEANS “**.

1.2 Rumusan Masalah

Berdasarkan latar belakang permasalahan, diperoleh rumusan masalah sebagai berikut:

1. Bagaimana hasil pengelompokan kabupaten/kota berdasarkan indeks pembangunan manusia dengan menggunakan metode *k-means* sebagai *generative mechanism* dan *clustering ensemble*?
2. Bagaimana perbandingan hasil *cluster* antara metode *k-means* dan *ensemble*?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah, diperoleh tujuan penelitian sebagai berikut:

1. Mengelompokkan kabupaten/kota di Sumatera Barat menggunakan metode *k-means* sebagai *generative mechanism* dan *clustering ensemble* berdasarkan indeks pembangunan manusia.
2. Mengetahui perbandingan hasil *cluster* antara metode *k-means* dan *ensemble*.

1.4 Batasan Masalah

Adapun batasan permasalahan yang digunakan pada penelitian ini adalah sebagai berikut:

1. Metode untuk menentukan *base cluster* menggunakan metode *k-means*. *K-means* merupakan metode analisis *cluster* non hierarki yang sederhana, tidak rumit dan memudahkan dalam *generative mechanism* atau menentukan anggota-anggota *clustering* untuk analisis *cluster ensemble* tersebut.
2. Fungsi konsensus yang digunakan yaitu CSPA. CSPA adalah salah satu fungsi *consensus* yang efisien dalam melakukan pengelompokan menggunakan *ensemble cluster*.
3. Validasi internal yang digunakan yaitu *silhouette*.

1.5 Manfaat Penelitian

Berdasarkan latar belakang dan tujuan penelitian, maka manfaat yang dapat diperoleh dari penelitian ini yaitu:

1. Mengetahui tentang hasil pengelompokan yang diperoleh dari analisis *cluster ensemble* dengan menggunakan metode *k-means* sebagai *generative mechanism* berdasarkan indeks pembangunan manusia
2. Mengetahui perbandingan antara *clustering ensemble* dan *k-means*.