

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pendidikan merupakan proses kemampuan serta keahlian diri yang akan terus tumbuh berkembang secara terus-menerus. Peningkatan kualitas pendidikan juga merupakan investasi dan aspek yang vital dalam meningkatkan peningkatan ekonomi, kualitas sumber daya manusia, serta menurunkan tingkat kemiskinan dan pengangguran di suatu daerah. Bekal dasar dalam menempuh pendidikan adalah memiliki kemampuan menulis, membaca, berkomunikasi dan berhitung secara efisien. Buta huruf adalah salah satu permasalahan dalam menempuh pendidikan dan menghambat proses pencapaian tujuan peningkatan kualitas pendidikan, sedangkan yang dimaksud dengan buta huruf yaitu ketidakmampuan seseorang dalam menulis atau membaca kalimat sederhana dalam bahasa Indonesia atau berbagai bahasa lainnya. Dampak dari tidak meratanya pendidikan salah satunya merupakan buta huruf yang menyebabkan rendahnya kualitas sumber daya manusia (SDM).

Permasalahan buta huruf atau buta aksara adalah salah satu pekerjaan rumah bagi pemerintahan Indonesia dan tantangan untuk dunia pendidikan. Berdasarkan data Badan Pusat Statistik (BPS), angka buta huruf rentan usia 15-44 tahun yang mencapai nilai tertinggi dari tahun 2015 hingga tahun 2019 berada pada provinsi Papua dengan jumlah nilai tertinggi pada tahun 2015 total jumlah persentase dari penduduk buta huruf sebesar 28,47%, tahun 2016 memiliki nilai persentase sebesar 28,21%, tahun 2017 memiliki nilai persentase sebesar 24,66%, tahun 2018 memiliki nilai persentase sebesar 21,2%, dan pada tahun 2019 memiliki nilai persentase sebesar 20,21%. Tingginya tingkat buta huruf disebabkan oleh tingkat kemiskinan yang cukup tinggi sehingga warga kurang mempunyai kesempatan belajar dan tidak memfasilitasi dirinya untuk belajar. Untuk memperkecil tingkat buta huruf di masyarakat maka diperlukan suatu pendekatan yang sesuai dengan karakter dan kultur, sehingga mereka memiliki motivasi untuk belajar.

Berdasarkan penjelasan di atas terdapat metode yang dapat menyelesaikan permasalahan tersebut secara kompleks melalui salah satu cabang ilmu komputer seperti *data mining*. Istilah *data mining* digunakan untuk menjabarkan penemuan pengetahuan di dalam daftar data untuk mengambil dan mengidentifikasi informasi yang bermanfaat dari berbagai daftar yang besar. Pada *data mining* terdapat beberapa metode seperti *classification* dan *clustering*. Secara umum *classification* merupakan pemrosesan untuk menemukan sebuah model atau fungsi yang menjelaskan dan merincikan konsep atau kelas data untuk kepentingan tertentu. Sedangkan metode *clustering* digunakan untuk mengelompokkan data berdasarkan karakteristik yang sama ke dalam satu *cluster* dan data yang berbeda masuk ke dalam *cluster* yang lainnya. Pada metode *clustering* terdapat beberapa metode algoritma di antaranya adalah algoritma *Fuzzy C-Means* dan algoritma *K-Means*. Kedua algoritma tersebut akan mengelompokkan data ke dalam bentuk satu atau lebih *cluster*, sehingga data yang memiliki karakteristik sama akan dikelompokkan ke dalam *cluster* yang sama dan data yang mempunyai karakteristik berbeda akan dikelompokkan ke dalam *cluster* yang lain.

Sebelumnya penelitian telah dilakukan oleh (Rahayu, 2019) dengan studi kasus “Analisis *K-Medoids* Dalam Pengelompokan Penduduk Buta Huruf Menurut Provinsi” menggunakan *dataset* pada tahun 2009 sampai 2017. Penelitian tersebut dilakukan untuk mengimplementasikan algoritma *K-Medoids* sebanyak 3 *cluster* dengan kategori *cluster* tingkat rendah, sedang dan tinggi untuk melihat sejauh mana tingkat buta huruf yang ada di Indonesia. Dari penelitian tersebut menghasilkan data analisa tingkat buta huruf berdasarkan provinsi di Indonesia yang dibagi menjadi 3 *cluster* dengan menggunakan *tools rapidminer*. Hasil dari penelitian tersebut diperoleh 3 *cluster* yakni 1 provinsi sebagai *cluster* tertinggi, 12 provinsi sebagai *cluster* sedang, dan 21 Provinsi sebagai *cluster* terendah. Pada penelitian yang lain dilakukan oleh (Agustina dan Prihandoko, 2018) dengan studi kasus “Perbandingan Algoritma *K-Means* Dengan Algoritma *Fuzzy C-Means* Untuk *Clustering* Tingkat Kedisiplinan Kinerja Karyawan”. Pada penelitian tersebut, tujuan yang ingin dicapai peneliti

yaitu melakukan perbandingan antara algoritma *K-Means* dengan algoritma *Fuzzy C-Means* untuk mengetahui kelebihan dari setiap algoritma dalam melakukan *clustering* sehingga memperoleh hasil yang terbaik. Hasil dari penelitian tersebut nilai validasi dari algoritma *Fuzzy C-Means* lebih optimal dalam menghasilkan metode yang lebih baik, dengan nilai validasi sebesar 0,758 dikarenakan nilai validasi lebih mendekati nilai 1, dibandingkan dengan algoritma *K-Means* dengan nilai validasi sebesar 0,528.

Berdasarkan penjelasan di atas, penulis bertujuan untuk mencari *cluster* optimum untuk algoritma *Fuzzy C-Means* dan algoritma *K-Means* menggunakan metode *Dunn Index* dengan skenario 3 sampai 10 *cluster*, kemudian dilakukan perbandingan antara kedua algoritma tersebut menggunakan uji validitas. Penulis melakukan perbandingan algoritma *Fuzzy C-Means* dan *K-Means* untuk melihat algoritma yang lebih baik dalam melakukan *clustering*. Uji validitas pada algoritma *K-Means* akan dilakukan dengan menggunakan metode *Silhouette Index*, jika nilai *Silhouette Index* semakin besar nilai validitasnya atau mendekati nilai 1, maka semakin baik algoritma tersebut. Sedangkan pada algoritma *Fuzzy C-Means* akan dilakukan dengan menggunakan metode *Partition Coefficient Index*, jika semakin besar nilai validitas atau mendekati nilai 1 nilai *Partition Coefficient Index* maka dapat disimpulkan bahwa algoritma tersebut semakin baik. Oleh karena itu, penelitian ini dibuat dengan judul “Perbandingan *Fuzzy C-Means* Dan *K-Means* Untuk Mengelompokkan Tingkat Buta Huruf Berdasarkan Provinsi di Indonesia”.

1.2 Rumusan Masalah

Rumusan masalah yang diselesaikan dari latar belakang di atas, yaitu:

1. Berapa *cluster* optimum pada algoritma *Fuzzy C-Means* dan algoritma *K-Means* menggunakan metode *Dunn Index* ?
2. Bagaimana perbandingan algoritma *Fuzzy C-Means* dan algoritma *K-Means* menggunakan metode *Partition Coefficient Index* dan metode *Silhouette Index* ?

1.3 Batasan Masalah

Batasan masalah yang terdapat pada penelitian ini yang perlu diketahui, antara lain:

1. Data yang dipakai berasal dari Badan Pusat Statistik yaitu “Persentase Penduduk Buta Huruf Menurut Kelompok Umur”.
2. Data yang digunakan terdiri dari 33 Provinsi berdasarkan 5 tahun terakhir dari tahun 2015 sampai 2019 dengan kelompok umur 15-44 tahun, jumlah data sebanyak 165 data.
3. Skenario penentuan *cluster* optimum menggunakan 3 sampai 10 *cluster*.
4. Pengukuran *cluster* optimum menggunakan metode uji validitas *Dunn index* dan perbandingan algoritma menggunakan metode uji validitas *Partition Coefficient Index* dan *Silhouette Index*.
5. *Tools* yang digunakan untuk *clustering* adalah *software* R Studio.

1.4 Tujuan Penelitian

Tujuan penelitian berdasarkan rumusan masalah di atas adalah:

1. Menentukan *cluster* optimum pada algoritma *Fuzzy C-Means* dan *K-Means*.
2. Membandingkan algoritma *Fuzzy C-Means* dan algoritma *K-Means* untuk menentukan algoritma yang lebih baik dalam melakukan *clustering*.

1.5 Manfaat Penelitian

1. Menjadi salah satu alternatif metode penelitian dalam mengelompokkan data tingkat buta huruf berdasarkan provinsi di Indonesia.
2. Menjadi referensi perbandingan algoritma *Fuzzy C-Means* dan algoritma *K-Means* dengan mencari validitas algoritma menggunakan metode *Partition Coefficient Index* dan *Silhouette Index*.
3. Informasi yang diperoleh sebagai alat bantu dalam memberikan informasi kepada pemerintah untuk bahan pertimbangan dalam mengambil keputusan.