

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang Masalah

Perkembangan terjadi sangat pesat pada zaman ini, sehingga manusia beradaptasi untuk mempermudah pola hidup, istilah ini juga berlaku didunia jurnalistik atau pers (Media). Seiring perkembangan zaman media pun ikut berkembang terlebih lagi dengan adanya internet. Perkembangan itu menghasilkan media *online* atau juga disebut dengan media daring. Media *online* merupakan tempat atau forum untuk menyebarkan beberapa media yang ada seperti cetak dan elektronik. minat masyarakat pada perkembangan media *online* semakin hari semakin tinggi dikarenakan kecepatan informasi yang ditawarkan lebih cepat, dan memudahkan untuk menemukan informasi yang lain pada fitur pencarian.

Hari Pers Nasional 2018 jatuh pada tanggal 9 Februari Ketua Dewan Pers Yosep Adi Prasetyo menyatakan, Dari jumlah 47.000 itu terdiri dari 2.000 diantaranya adalah media cetak, 674 radio, 523 televisi termasuk lokal, dan lebihnya media *online*. Hampir 79% merupakan media abal - abal dengan berita yang tidak dapat dipertanggungjawabkan[1]. Akibatnya, masyarakat sering terpapar berita yang tidak benar, selain itu berdampak juga terhadap jumlah artikel atau berita yang di terbitkan, semakin banyak media *online* yang beroperasi maka jumlah artikel atau berita yang diterbitkan juga semakin banyak. Sehingga kita tidak dapat mengolah informasi tersebut dikarenakan jumlahnya .

Berita *online* banyak sekali bermunculan di *internet* dengan jumlah yang sangat ekstrem. Tergambar dari jumlah media *online* yang beroperasi sehingga masyarakat mengalami kesulitan dalam mengelola informasi dan tidak dapat melihat topik apa yang sedang banyak diterbitkan pada media *online*. Salah satu pendekatan yang dapat digunakan untuk mengatur jumlah data yang besar dengan cara pengelompokan atau lebih terkenal dengan Teknik *Clustering*. *Clustering* merupakan proses partisi koleksi objek menjadi beberapa sub-cluster berdasarkan kesamaan isi artikel tersebut. Sehingga kita dapat mengetahui berita yang

berkembang dari cluster tersebut[2] tanpa harus membacanya semua berita pada media *online*.

*Clustering* memiliki beberapa algoritma yang dapat diterapkan dalam kasus ini, salah satunya adalah *Distance-based Clustering Algorithms* merupakan algoritma untuk menentukan *clustering* berdasarkan kedekatan antara data, di mana seiring perkembangan zaman algoritma tersebut memiliki beberapa variasi salah satunya adalah *Distance-based Partitioning Algorithms*[3].

*Distance-based Partitioning Algorithms* merupakan algoritma yang berdasarkan kepada jarak dan melakukan partisi untuk mendapatkan hasil lebih efisien. Dua algoritma *Distance-based Partitioning Algorithms* yang paling sering digunakan yaitu *k-medoid clustering algorithms* dan *k-means clustering algorithms*. [3]

*K-medoid* merupakan algoritma *clustering* data set dari objek ke dalam kelompok  $k$  yang dikenal apriori. Sebuah medoid dapat didefinisikan sebagai objek cluster yang rata-rata perbedaan untuk semua objek dalam cluster minimal yaitu titik paling berlokasi di cluster[3].

*K-means* merupakan algoritma unsupervised machine learning yang digunakan untuk menentukan atau membentuk kelompok dalam data. Satu kelompok data akan mempunyai karakteristik yang sama antara anggota kelompok satu dan anggota kelompok yang lain, di mana jumlah kelompok ditentukan dari nilai  $K$ [4]. Menurut penelitian algoritma *k-medoids* memiliki performa lebih baik dari *k-means*[5], di mana *k-medoids* lebih fleksible terhadap jarak antara data sehingga dapat membentuk *clustering* lebih baik dan *k-medoids* lebih kokoh(*Robustness*) untuk data *outliers* dan *noise* dari pada *k-means* yang meminimalkan jumlah ketidaksamaan berpasangan, sehingga dari kelebihan algoritma *k-medoids* peneliti memutuskan untuk mengimplementasikan algoritma *k-medoid* pada visualisasi untuk memecahkan permasalahan *clustering* terhadap berita online.

Penelitian dengan bidang yang serupa pernah dilakukan sebelumnya, salah satunya berjudul Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak penelitian ini Pengelompokan Wilayah Sebaran Cacat pada Anak[6]. Penelitian tersebut melakukan *clustering* terhadap dataset wilayah dengan persebaran cacat pada anak sehingga dapat

melakukan penanggulangan agar tingkat cacat pada anak menjadi lebih rendah. Penelitian lain dengan judul *Document Clustering using K-Medoids* [7]. Penelitian tersebut melakukan *text clustering* menggunakan berbagai metode yaitu : *K-medoids* Penelitian tersebut menghasilkan *efficient* 18.6% sampai 54% tergantung terhadap dataset yang digunakan oleh peneliti. Penelitian selanjutnya dengan judul , *K-Means and K-Medoids for Indonesian Text Summarization* [8], pada penelitian ini, peneliti melakukan proses *filtering, tokenization, TF-IDF, Similarity, and Clustering in K-Means or K-Medoids*, dengan melakukan percobaan dengan menggunakan 50 data uji yang terdiri dari 806 paragraf, memperoleh hasil *K-Medoids* memiliki hasil yang lebih baik dibandingkan dengan *K-Means* dengan tingkat akurasi 51.6% untuk *K-Means* dan 62.35% untuk *K-Medoids*.

Penelitian berikutnya membahas mengenai *K-medoids algorithm on Indonesian Twitter feeds for clustering trending issue as important terms in news summarization* [9], dalam penelitian ini peneliti membahas penggunaan algoritma *K-Medoids* untuk *clustering* data twitter untuk mendapatkan topic yang sedang *trending* dan ini hampir seperti kata-kata penting dari istilah isu *trending* disalin dari twitter. Penelitian selanjutnya mengenai *An Optimization K-Modes Clustering Algorithm with Elephant Herding Optimization Algorithm for Crime Clustering* [10], penelitian ini membahas mengenai perbandingan algoritma *K-Medoids* dengan algoritma *K-Medoids* dengan melakukan optimasi dengan menggunakan algoritma *Elephant Herding*, penelitian ini menunjukkan hasil algoritma *K-Medoids* dengan optimasi akan meningkatkan *Purity* di mana parameters tersebut dapat meningkatkan akurasi sehingga hasil yang didapat akan lebih sesuai.

Sehingga metode *clustering* ini dapat menyimpulkan *cluster* yang terbentuk dari dataset yang digunakan oleh peneliti, dan memberi gambaran mengenai berita apa yang sedang banyak diberitakan.

PT. Indonesia Indikator merupakan perusahaan yang bergerak dibidang *Big Data and Analytics*. Salah satu data yang dikumpulkan dan di analisa yaitu data berita

*online* dan berita cetak. Maka dari itu menggunakan data yang dimiliki oleh PT. Indonesia Indicator sangat sesuai dengan kebutuhan data pada penelitian ini.

### **1.2. Identifikasi Masalah**

Berdasarkan uraian pada latar belakang, dapat diidentifikasi masalah - masalah sebagai berikut:

1. Jumlah data dari berita *online* sangat banyak, sehingga kita tidak dapat menentukan berita mana yang akan kita baca.
2. Diperlukannya pengelompokan berita *online* berdasarkan *content* beritanya sehingga dapat mengetahui berita yang sedang diperbincangkan tanpa harus membaca semua sumber

### **1.3. Rumusan Masalah**

Rumusan masalah yang didapat oleh peneliti berdasarkan uraian yang dijelaskan dilatar belakang yaitu:

1. Berapa tingkat akurasi dari algoritma *k-medoid* apabila diterapkan pada kasus berita *online*?
2. Bagaimana membuat visualisasi yang dapat menentukan cluster dari berita *online*?

### **1.4. Batasan Masalah**

Penelitian ini agar dapat lebih terfokus pada permasalahan dan sesuai dengan tujuan, maka ditetapkan batasan terhadap masalah yang diteliti. Hal ini dilakukan supaya tidak menyimpang dari tujuan dan permasalahan yang ada, maka Batasan yang dimaksud:

1. Media yang dijadikan Clustering hanya terbatas pada media tertentu seperti berita harian detik, berita harian kompas, berita harian tempo dan lain - lain.
2. Menggunakan data *online news* PT. Indonesia Indikator dengan berita berbahasa Indonesia.

### **1.5. Tujuan Penelitian**

Berdasarkan masalah yang ada, maka penelitian ini memiliki beberapa tujuan, yaitu:

1. Membuat model yang dapat digunakan untuk mendeteksi cluster berita *online*.
2. Membuat visualisasi yang dapat menentukan *cluster* berita *online* berdasarkan model yang dihasilkan.

### **1.6. Manfaat Penelitian**

Berdasarkan tujuan yang hendak dicapai, maka manfaat yang dapat diambil dari penelitian ini adalah sebagai berikut:

1. Mengetahui performa algoritma *k-medoid* terhadap dataset berita *online*.
2. Seseorang dapat melihat pengelompokan berita online tanpa harus membaca semua berita online yang ada.

### **1.7. Kontribusi Penelitian**

Hasil penelitian ini diharapkan dapat memberikan informasi terhadap clustering terhadap berita online yang sangat banyak jumlahnya.

Hasil penelitian ini juga diharapkan dapat memberikan kontribusi pengetahuan walaupun wawasan bagi Lembaga Pendidikan, khususnya Lembaga Pendidikan Teknik informatika, bidang keilmuan yang serumpun, dan pihak yang ingin mengembangkan ilmu mengenai *clustering*.

### **1.8. Sistematika Penelitian**

Agar memudahkan untuk memahami penelitian ini, maka materi - materi dalam tugas akhir ini di kelompokkan ke dalam bagian-bagian sebagai berikut:

#### **1. BAB I Pendahuluan**

Berisi latar belakang, identifikasi masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, kontribusi penelitian dan sistematika penelitian.

#### **2. BAB II Landasan Teori**

Penjelasan definisi/konsep yang akan digunakan dalam penelitian.

#### **3. BAB III Metodologi Penelitian**

Bab ini menjelaskan tentang waktu penelitian, alat pendukung, tahapan penelitian dan metodologi yang digunakan selama penelitian berlangsung.

#### **4. BAB IV Hasil Implementasi dan Pengujian**

Bab ini akan menjelaskan bagai mana sistem ini dikembangkan dari tahap pengumpulan data hingga pengujian pada masing – masing tahap.

## **5. BAB V Kesimpulan dan Saran**

Penyajian hasil penelitian dan analisa dan hasil penelitian (kesimpulan)