

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan kecanggihan teknologi yang ada pada saat ini terus mengalami perubahan termasuk dalam media *online*, terlihat bahwa banyak media cetak mulai beralih ke media *online* portal berita. Portal berita digunakan untuk menyampaikan informasi yang biasanya terdapat pada media cetak, salah satu manfaat perubahan ke media *online* karena mendapatkan pangsa pasar yang lebih besar dan tentu pengunjung akan merasa dimudahkan dengan pencarian informasi yang diberikan oleh portal berita karena lebih cepat dipublikasi.

Portal berita memberikan beberapa navigasi yang akan menjadi topik berita yang ingin dicari. Navigasi seperti ekonomi, peluang usaha, dan solusi ukm merupakan navigasi untuk mempermudah pembaca dalam mencari informasi tentang perkembangan dalam bidang ekonomi dan bisnis atau usaha. Informasi pada portal berita tentang perkembangan ekonomi sering kali didapatkan dari masalah lapangan pekerjaan serta bisnis atau usaha pada masyarakat Indonesia. Pertumbuhan ekonomi tidak sepadan dengan jumlah lapangan kerja, di sisi lain jumlah pengusaha muda di Indonesia dinilai masih sedikit jumlahnya karena minat berwirausaha masyarakat Indonesia dinilai masih minim. Keinginan berwirausaha sering kandas karena calon pengusaha takut untuk mengalami kegagalan. Salah satu permasalahan yang menjadi penyebab kegagalan dalam berwirausaha adalah produk yang kurang menjual, ke tidak sungguhan dalam menjalankan usaha seperti halnya produk yang akan ditawarkan kepada pembeli bukan merupakan produk

yang dibutuhkan oleh masyarakat namun hanya berdasarkan pada keinginan pribadi semata, karena calon pengusaha memang tidak mengetahui kebutuhan masyarakat. Keterbatasan ide dalam menentukan bisnis adalah kendala besar yang sering dialami oleh calon pengusaha yang menyebabkan produk kurang diminati di kalangan masyarakat.

Data Mining merupakan proses yang menerapkan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi serta pengetahuan dari berbagai sumber yang besar. *Clustering* merupakan cabang dari data mining, algoritma K-Means merupakan salah satu metode *clustering non hierarki* yang memartisi data ke dalam *cluster* sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda di kelompokkan ke dalam *cluster* yang lain. *Text mining* dapat menambang data yang berupa teks di mana sumber data dari teks judul dan tag berita pada portal berita, sehingga kata-kata mewakili makna dari judul berita tersebut dan dapat dilakukan analisa keterhubungannya.

Navigasi tema bisnis pada portal berita membantu untuk mengelompokkan jenis berita bertema bisnis tetapi pada navigasi tersebut tidak ada pengelompokan jenis bisnis yang sering dibahas. Algoritma K-Means dapat mengelompokkan macam-macam ide bisnis yang ada pada portal berita berdasarkan *cluster* yang di dapatkan dari *text mining*, sehingga dapat dilakukan analisa untuk mengetahui jenis bisnis yang sering di informasikan pada portal berita.

Beberapa penelitian telah dilakukan dengan menggunakan algoritma K-Means di antaranya :

Penelitian yang diteliti oleh Eko Yulian yang berjudul "*Text Mining* dengan *K-Means Clustering* pada Tema LGBT dalam Arsip Tweet Masyarakat Kota Bandung". Tujuan penelitian ini untuk mengelompokkan komentar-komentar di twitter tentang LGBT di Kota Bandung ke dalam *cluster-cluster* yang dapat dibedakan. Penelitian ini dipublikasikan di Jurnal Online Fakultas Sains dan Teknologi UIN Sunan Ampel Surabaya, dimuat pada jurnal Matematika "MANTIK", Vol. 04 No. 01 pada Mei 2018. P-ISSN: 2527-3159 E-ISSN: 2527-3167.

Penelitian yang diteliti oleh Brillian Aristyo Rahadian, Diva Kurnianingtyas, Dyan Putri Mahardika, Tusty Nadia Maghfira, dan Imam Cholissodin yang berjudul "*Analisis Judul Majalah Kawanku Menggunakan Clustering K-Means dengan Konsep Imulasi Big Data Pada Hadoop Multi Node Cluster*". Tujuan pengujian ini dilakukan untuk membuat variasi data set yang berbeda ukuran atau banyaknya data yang digunakan dari keseluruhan data dipecah-pecah menjadi beberapa bagian, dari variasi data yang digunakan selanjutnya dilihat waktu eksekusi yang diperlukan. Penelitian ini dipublikasikan di Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK), dimuat pada jurnal Vol. 4, No. 2, hlm. 75-80 pada Juni 2017. P-ISSN: 2355-7699 E-ISSN: 2528-6579.

Penelitian yang diteliti oleh Dwi Smaradahana Indraloka, dan Budi Santosa yang berjudul "*Penerapan Text Mining untuk Melakukan Clustering Data Tweet Shopee Indonesia*". Tujuan penelitian ini untuk penentuan jenis konten dan perhitungan rata-rata jumlah *retweet* pada tiap klaster sehingga Pelaku bisnis Shopee Indonesia dapat mengetahui jenis konten *tweet* yang banyak dilakukan *retweet* oleh *followers* untuk dapat menggunakan jenis konten *tweet* tersebut

sebagai sarana untuk melakukan *advertising* kepada pengguna Twitter. Penelitian ini dipublikasikan di Publikasi Online ITS (POMITS), dimuat pada jurnal Sains dan Seni Its Vol. 6, No. 2 (2017) 2337-3520(2301-928x *Print*) pada tahun 2017.

Penelitian yang diteliti oleh Alfian Futuhul Hadi, Dimas Bagus C. W., dan Moh. Hasan yang berjudul "*Text Mining* Pada Media Sosial Twitter Studi Kasus: Masa Tenang Pilkada Dki 2017 Putaran 2". Tujuan penelitian ini untuk menggali informasi dari media sosial Twiter yang dapat digunakan sebagai gambaran tentang apa saja yang terjadi selama proses pilkada (khususnya pada masa tenang). Selain itu, peneliti menduga bahwa ada keterkaitan antara media sosial dengan hasil dari pencoblosan. Penelitian ini dipublikasikan di Jurnal Online Departemen Matematika Universitas Airlangga Surabaya, dimuat pada jurnal Seminar Nasional Matematika dan Aplikasinya, pada 21 Oktober 2017.

Penelitian yang diteliti oleh Sulastris, dan Dwi Agus Diartono yang berjudul "*Analisa Jejaring Sosial Twitter Menggunakan Klastering Kmeans dan Hirarki Agglomeratif*". Tujuan penelitian ini bertujuan untuk menganalisa data tweet yang berhubungan dengan kata kunci atau topik Piala AFF Suzuki 2016 dan Pemilihan Kepala Daerah 2017 yang sedang hangat dibicarakan di berbagai media. Penelitian ini dipublikasikan di Jurnal Online SINTAK Universitas Stikubank, dimuat pada jurnal Prosiding SINTAK ISBN: 978-602-8557-20-7, pada tahun 2017.

Penelitian yang diteliti oleh Miftachur Robani, dan Achmad Widodo yang berjudul "*Algoritma K-Means Clustering Untuk Pengelompokan Ayat Al Quran Pada Terjemahan Bahasa Indonesia*". Penelitian ini bertujuan untuk menghasilkan suatu sistem informasi yang dapat melakukan pengelompokan ayat Al Quran

dengan metode K-Means. Penelitian ini dipublikasikan di JSINBIS (Jurnal Sistem Informasi Bisnis) Universitas Diponegoro , dimuat pada jurnal JSINBIS (Jurnal Sistem Informasi Bisnis); Vol 6, No 2 (2016): Volume 6 Nomor 2 Tahun 2016 *Digital Object Identifier* 10.21456/vol6iss2pp164-176.

Penelitian yang diteliti oleh Frando Christo Wulur, dan Irwan Sembiring yang berjudul "Klasifikasi *Alert* pada *Intrusion Detection System* Menggunakan Algoritma K-Means". Tujuan penelitian ini untuk dapat mengklasifikasikan serangan pada sistem IDS. Penelitian ini dipublikasikan di Institutional Repository Satya Wacana *Christian University* pada Oktober 2015.

Penelitian yang diteliti oleh Rendy Handoyo, R. Rumani M, dan Surya Michrandi Nasution yang berjudul "Perbandingan Metode *Clustering* Menggunakan Metode Single Linkage dan K-Means pada Pengelompokan Dokumen". Tujuan penelitian ini untuk melihat pengaruh jumlah dokumen, jumlah *cluster*, dan metode *clustering* dalam mengelompokkan dokumen. Penelitian ini dipublikasikan di E-Jurnal Mikorskil, dimuat pada JSM STMIK Mikorskil Vol 15, No 2, Oktober 2014. ISSN. 1412-0100.

Berdasarkan permasalahan yang dikemukakan di atas, maka penulis tertarik menganalisis untuk mengetahui bisnis yang populer sering dibahas pada portal berita, karena dengan adanya hasil analisis tersebut dapat diketahui bahwa bisnis apa saja yang sedang populer dan memenuhi kebutuhan yang di butuh kan oleh masyarakat pada umumnya karena informasi bisnis yang disampaikan pada portal berita adalah bisnis yang terjadi pada masyarakat, sehingga calon pengusaha dapat lebih maksimal dalam pemilihan bisnis atau usaha seperti dalam menentukan produk yang dijual. Calon pengusaha dapat mengambil ide dari bisnis yang sedang

populer dengan menambah ide kreativitas untuk inovasi produknya ataupun mengambil ide bisnis yang tidak terlalu populer tetapi sesuai dengan kebutuhan masyarakat dilingkungannya.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan, maka permasalahan yang dapat dirumuskan adalah :

1. Bagaimana mengelompokkan beberapa objek bisnis dari portal berita yang serupa ke dalam sebuah *cluster* dan menentukan jumlah *cluster* terbesar sampai terkecil menggunakan algoritma K-Means?
2. Bagaimana menguji kekuatan anggota *cluster* terhadap *cluster* menggunakan pengujian *Silhouette Coefficient* ?

1.3 Batasan Masalah

Supaya hasil pengelompokan yang dihasilkan dapat digunakan secara maksimal maka dalam penelitian ini diberikan batasan masalah sebagai berikut :

1. Kata yang diproses dalam analisis penelitian ini dari teks judul dan tag artikel bisnis di beberapa portal berita Indonesia antara lain *cncindonesia*, *peluang detik*, *okezone*, *peluang usaha kontan*, dan *bisnis tempo*.
2. Penelitian dilakukan menggunakan data teks judul dan tag artikel bisnis pada tahun 2018.
3. Bahasa yang diolah pada pemrosesan kata hanya bahasa Indonesia baku menurut Kamus Besar Bahasa Indonesia (KBBI).

4. Penelitian ini tidak memperhatikan kesalahan ejaan atau penulisan pada dokumen, serta tidak memperhatikan sinonim atau persamaan kata.
5. Bahasa pemrograman yang digunakan dalam sistem ini adalah PHP untuk *backend* serta HTML, CSS dan Javascript untuk *frontend* dan menggunakan database MySQL.

1.4 Maksud dan Tujuan Penelitian

Maksud dan tujuan dari penelitian antara lain:

1.4.1 Maksud Penelitian

1. Menganalisis dan menentukan *cluster* dari teks judul dan tag artikel bisnis pada portal berita.
2. Mencari jumlah *cluster* terbesar sampai terkecil menggunakan algoritma K-Means.
3. Membuat visualisasi kata yang sering muncul disetiap *cluster* menggunakan *Word Cloud*.
4. Menguji kekuatan anggota *cluster* terhadap *cluster* menggunakan pengujian *Silhouette Coefficient*.

1.4.2 Tujuan Penelitian

1. Memberikan informasi kepada masyarakat luas tentang bisnis yang paling populer yang sering dibahas pada portal berita.
2. Memaparkan hasil kepopuleritasan bisnis sehingga dapat menjadi ide bisnis bagi masyarakat ataupun calon pengusaha melalui *Word Cloud* yang terbentuk di setiap *cluster*.

1.5 Manfaat Penelitian

1. Bagi Penulis di Portal Berita

Sebagai gagasan untuk lebih membahas bisnis dari hasil analisis tersebut dan membahas bisnis baru pada kalangan masyarakat, karena peran penyampaian informasi jenis bisnis sangat penting sebagai ide-ide peluang bisnis.

2. Bagi Masyarakat dan Calon Pengusaha

Sebagai acuan untuk mengetahui bisnis apakah yang sedang populer di kalangan masyarakat sehingga dapat menjadi ide gagasan bisnis.

1.6 Metode Penelitian

Peneliti menjabarkan cara-cara memperoleh data-data yang digunakan untuk kebutuhan penelitian.

1. Metode Pengumpulan Data

Data yang digunakan pada penelitian ini menggunakan Data Primer. Data yang dipakai pada sistem ini dari teks judul dan tag artikel bisnis di beberapa portal berita Indonesia. Serta menggunakan data sekunder, data sekunder adalah data yang diperoleh dari hasil studi literatur dan dokumen yang terkait.

2. Metode Analisa

Metode analisa yang digunakan pada penelitian ini adalah analisa kebutuhan fungsional dan non fungsional. Kebutuhan fungsional adalah kebutuhan yang berisi proses-proses yang nantinya harus disediakan oleh

sistem. Kebutuhan non fungsional digunakan untuk mengetahui spesifikasi kebutuhan oleh sistem.

3. Metode Perancangan

Sistem yang dirancang dengan algoritma K-Means menggunakan metode flowchart. Flowchart adalah alur proses atau bagan dengan simbol-simbol yang menunjukkan urutan dan hubungan antar proses beserta perintahnya. Simbol-simbol tersebut memiliki makna tersendiri dengan proses tertentu.

4. Metode *Testing*

Sistem yang telah dirancang akan diuji coba dengan menggunakan metode pengujian yang digunakan untuk mengukur kualitas *cluster* menggunakan *Silhouette Coefficient*. *Silhouette Coefficient* digunakan untuk melihat kualitas dan kekuatan *cluster*, seberapa baik suatu objek ditempatkan dalam suatu *cluster*.

1.7 Sistematika Penulisan

Sistematika penulisan bertujuan untuk mempermudah dalam membuat laporan skripsi. Adapun sistematika penulisan pada laporan implementasi algoritma K-Means ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi mengenai informasi pokok permasalahan dan pemaparan penelitian secara umum, adapun hal-hal yang dibahas adalah latar belakang, rumusan masalah, batasan masalah, maksud dan tujuan penelitian, manfaat penelitian, metode penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini berisi mengenai konsep dasar teori-teori yang berhubungan dengan pembahasan penelitian dari sumber pustaka dan referensi yang menjadi landasan dasar dalam perancangan, analisa kebutuhan sampai dengan implementasi dan pengujian sistem.

BAB III ANALISIS DAN PERANCANGAN PENELITIAN

Bab ini berisi mengenai analisa kebutuhan dalam pembangun sistem, analisa sistem yang sedang berjalan pada sistem ini sesuai dengan metode implementasi perangkat lunak yang digunakan.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi mengenai tahapan implementasi dan pengujian yang merupakan tahapan yang dilakukan dalam mengimplementasikan dari hasil penelitian, analisa dan perancangan yang telah diidentifikasi untuk mengimplementasikan dan menguji sistem.

BAB V PENUTUP

Bab ini berisikan kesimpulan yang diambil dari hasil penulisan dan saran yang dapat menjadi masukan bagi penelitian selanjutnya.