

BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

Media *Twitter* digunakan guna riset dengan luasnya pemakaian *Twitter* selaku media dialog oleh seluruh golongan. banyak digunakan publik guna mengantarkan komentar secara gambling terpaut bermacam isu [4].

Riset terpaut analisis sentimen dari media *Twitter* telah banyak dicoba lebih dahulu dengan memakai tata cara yang berbeda- beda. Pada riset Sianipar serta Setiawan pada tahun 2015, periset memakai media *Twitter* buat riset bertajuk“ Pendeteksian Kekuatan Sentimen pada Bacaan *Tweet* Berbahasa Indonesia Memanfaatkan *Sentistrength*”. Riset tersebut bertujuan buat menganalisis sentimen pada bacaan *tweet* bahasa Indonesia sekalian memandang dampak dari pemakaian tata cara *Sentistrength*. Opini pada media *Twitter* hendak dikategorikan dalam 3 jenis adalah sentiment positif, sentiment negatif serta sentiment netral. Penilaian tata cara *Sentistrength* yang dicoba merupakan akurasi, uji korelasi, serta uji hipotesis dengan tata cara *t- test*. Nilai akurasi pada riset ini dikira kurang memuaskan sebab kurang lengkapnya terms pada *leksikon* sehingga tidak penuhi kebutuhan bahasa Indonesia yang diperlukan buat mengklasifikasi bahasa *tweet* [5].

Riset Wahid serta Azhari tahun 2016 serta memanfaatkan media *Twitter* dengan judul tulisan“ Peringkasan Sentimen Ekstraktif di *Twitter* Mengenakan *Hybrid TF- IDF* serta *Cosine Similarity*”. Riset ini bertujuan guna mengombinasikan tata cara *SentiStrength*, *Hybrid TF- IDF* serta *Cosine Similarity* guna mengekstraksi ringkasan sentimen positif serta negatif warga terhadap topik selebriti di *Twitter*. Dalam riset ini menyamakan gabungan tata cara *Hybrid TF- IDF* serta *Cosine Similarity* dan gabungan tata cara *SentiStrength*, *Hybrid TF- IDF* serta *Cosine Similarity*. *SentiStrength* digunakan guna mengklasifikasikan *tweet* dengan menganalisis kekuatan kata dalam suatu bacaan *tweet*. Gabungan ini diharapkan menambah keahlian *Hybrid TF- IDF* guna meringkas bacaan/ opini

pada media *Twitter*. Bersumber pada riset tersebut gabungan tata cara *SentiStrength*, *Cosine Similarity* menciptakan tingkatan akurasi yang lebih baik dibanding dengan pemanfaatan *Hybrid TF- IDF* saja. Akurasi pemakaian *Sentistrength* rata- rata 74, 12% serta presisi sebesar 73, 08% [6].

Riset memanfaatkan algoritma *K- Nearest Neighbor* serta dicoba oleh Nurjanah, dkk(2017). Penelitian mengumpulkan informasi *tweet* dari media sosial *Twitter*, diawali dari tahapan *pre- processing*. Pada tahapan *pre- processing*, dicoba tokenisasi, pembersihan informasi, *case folding*, filterisasi, serta *stemming*. Tahapan berikutnya merupakan pembobotan memakai tata cara *TF- IDF*. Hasil pembobotan dihitung jarak kemiripannya memakai *cosine similarity*, setelah itu diolah memakai algoritma *K- Nearest Neighbor*. Sesi berikutnya yakni melaksanakan pembobotan jumlah *Retweet* serta dinormalisasi memakai tata cara min- max. Setelah itu hasil penggabungan tersebut menghasilkan suatu nilai, yang menampilkan hasil klasifikasi dokumen positif ataupun negatif[7].

Wu He, dkk (2013) melakukan penelitian tentang bagaimana data perusahaan di media sosial dapat dianalisis menggunakan metode penambangan teks menggunakan *tools* Nvivo dan SPSS. Dalam penelitian ini, dilakukan penelitian terhadap tiga perusahaan pizza besar. Analisis yang dihasilkan akan membandingkan jumlah *followers* dan jumlah *tweets*, lalu kemudian membagi-bagi *tweets* yang ada menjadi beberapa topik besar. Topik yang dihasilkan dapat menjadi pembanding topik-topik apa yang menjadi bahasan wajib di setiap perusahaan. Nantinya, hasil grafik data dan topik tersebut dapat dijadikan bahan untuk melakukan analisis kompetitif dengan perusahaan lain yang merupakan pesaing[8].

Tong (2016) melakukan analisis text mining menggunakan Latent Dirichlet Allocation (LDA) untuk penerapan pemodelan topik. Dalam penelitian ini, dilakukan analisis pada teks Wikipedia dan Twitter. Dari pengamatan pada data teks di Wikipedia, didapatkan distribusi topik yang ada pada artikel tersebut. Dari pengamatan pada teks di Twitter, didapatkan juga distribusi topik dari data yang di analisis. Dari distribusi topik yang dihasilkan, dapat disimpulkan hal apa yang menjadi topik pembicaraan ataupun topik pada artikel yang dianalisis [9].

Jacobi(2015) melaksanakan kuantitatif analisis dari sekumpulan teks jurnalistik memanfaatkan tata cara pemodelan topik. Dalam penelitiannya, Jacobi(2015) memanfaatkan uraian *Latent Dirichlet Analysis* yang ialah metode unsupervised dari pemodelan topik. Dengan mengaplikasikan tata cara ini, hendak tercipta kelompok kata yang dikelompokkan ke topik dari sekumpulan dokumen. Informasi buat riset ini didapatkan dari postingan New York Times tentang teknologi nuklir. Dalam penjelasannya, riset ini mengkategorikan topik jadi 3 bagian, ialah:(1) Topik yang membentuk pola;(2) Topik yang kurang lebih selalu muncul;(3) Topik yang tidak cocok dengan riset permasalahan [10].

Asmussen (2019) melakukan riset pada sekumpulan literatur menggunakan pemodelan topik. Langkah dari penelitian ini pengumpulan data, pre-processing, pemodelan topik, dan *post-processing*. Dalam penelitian ini, digunakan metode *Latent Dirichlet Allocation* (LDA). Data yang digunakan di penelitian adalah literature berjumlah 650 judul. Hasilnya berbentuk beberapa topik beserta distribusi kata dari setiap topik[11].

Jockers(2013) melaksanakan analisis pada sebagian novel serta literatur dari abad ke- 19 buat menciptakan tema ataupun topik yang timbul dari bermacam literatur tersebut. Riset ini pula mengenali apakah yang menulis literatur tersebut merupakan laki- laki ataupun perempuan. Pengidentifikasi ini, memakai *Bootstrap plot* serta *Permutation plot*. Periset mempraktikkan tata cara pemodelan topik statistik, ialah *Latent Dirichlet Allocation*(LDA). Hasil dari riset ini merupakan sekumpulan distribusi kata yang diinterpretasikan jadi bermacam topik serta gender dari topik tersebut[12].

Evans(2014) melaksanakan metode komputasional buat melaksanakan kualitatif analisis buat informasi tekstual yang besar. Informasi yang digunakan merupakan dari koran kepunyaan Amerika dari tahun 1980 hingga 2012. Metode yang digunakan dalam analisis merupakan topic modelling dengan mempraktikkan *Latent Dirichlet Allocation*(LDA). Sehabis analisis dengan LDA, dihasilkan distribusi kata bersumber pada topik kemudian dicoba pelabelan pada tiap topik secara manual. Setelah itu melaksanakan perbandingan peringkat topik memakai

parameter topik *Dirichlet*. Terus menjadi besar parameter topik *Dirichlet* hingga terus menjadi banyak bagian dokumen yang tercantum ke topik tersebut[13].



Berikut ini adalah tabel analisis yang pernah dilakukan :

Tabel 2.1 Perbandingan Penelitian

Pembanding	Mihovsky, dkk(2017)[14]	Wahid, dkk(2016)[6]	Sianipar, dkk (2015)[5]	Wu He, dkk(2013)[8]	Nurjanah, dkk (2017)[7]	Penulis *)
Topik	<i>Sentiment Strength Detection in Short Informal Text Mike</i>	Peringkasan Sentimen Esktraktif di <i>Twitter</i> Menggunakan <i>Hybrid TF-IDF</i> dan <i>Cosine Similarity</i>	Pendeteksian Kekuatan Sentimen Pada Teks <i>Tweet</i> Berbahasa Indonesia Menggunakan <i>Sentistrength</i>	<i>Social Media Competitive Analysis and Text Mining: A Case Study in the Pizza Industry</i>	Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial <i>Twitter</i>	Analisis Sentimen Publik Terhadap Ibadah <i>Online</i> Di Masa Pandemi Pada Media <i>Twitter</i> Dengan Menggunakan Metode <i>Sentistrength</i>
Metode	<i>SentiStrength</i> , <i>support vector IDF</i> dan <i>Cosine Similarity machines</i>	<i>Sentistrength</i> , <i>Hybrid TF-SentiStrength</i>	<i>Sentistrength</i>	Text Mining menggunakan tools Nvivo dan SPSS	<i>KNearest Neighbor</i>	<i>Sentistrength</i>

Data	Data diambil dari komentar berbentuk teks pendek	Data <i>Twitter</i> yang dikumpulkan sebanyak 7421	Data diambil dari 6 akun masing-masing 100 <i>tweet</i> secara acak	Melakukan analisis kompetitif	<i>Tweet</i> dari media sosial <i>Twitter</i>	Data <i>Twitter</i> yang dikumpulkan berbentuk teks dari komentar
Hasil Penelitian	Positif 60,6% dan negatif 72,8%.	Menghasilkan ringkasan dengan akurasi yang lebih baik.	Metode <i>sentistrength</i> akurasi sebesar 57,33%	Menemukan kelemahan dan kesempatan baru	Metode Knearest dapat menghasilkan <i>retweet</i> min-max	meringkas <i>tweet</i> yang menyinggung tentang beribadah <i>online</i>

NB : *) Sedang dalam proses penelitian