

BAB 2

TINJAUAN PUSTAKA

Pada tinjauan pustaka ini akan membahas tentang penelitian berita yang menggunakan *Text Mining*, metode *TF-IDF*, dan *Twitter*. Yang mana penelitian ini akan mengulas secara lengkap tentang pengaplikasian *Text Mining*, metode *TF-IDF*, dan *Twitter*. Mengapa penulis menggunakan metode *TF-IDF* karena dalam penelitian Tugas Akhir (Musa, et al., 2012) menunjukkan bahwa pembobotan *ITF* mampu menghasilkan performansi yang lebih baik jika di bandingkan dengan skema pembobotan *TF-IDF* ketika menggunakan 25%, 50%, dan 75% dataset dari setiap kelas untuk masing-masing topik. Sedangkan pada saat menggunakan 100% dataset, *TF-IDF* menghasilkan performansi yang lebih baik dibandingkan dengan skema pembobotan *ITF*.

Tugas Akhir (Ma'arif, 2015) dalam penelitiannya tentang "Penerapan Algoritma *TF-IDF* Untuk Pencarian Karya Ilmiah" bertujuan untuk menerapkan metode *TF-IDF* yang dapat digunakan untuk mencari karya ilmiah sebagai pengukur tingkat similaritas antara dokumen dengan kata kunci yang didapat dari ekstraksi teks pada dokumen menggunakan *text mining*. Dengan melakukan proses penambangan data yaitu kata pada dokumen lalu didapatkan hasil ekstraksi yang selanjutnya dilakukan pembobotan menggunakan metode *TF-IDF*, maka didapatkan data yang terurut dari yang kemiripannya atau tingkat similaritasnya paling tinggi sehingga pencarian karya

ilmiah menjadi lebih efisien sebagai informasi yang relevan.

Begitu juga dengan penelitian Tugas Akhir (All Farizi, 2015) yang berjudul "Rekomendasi Tag Pada Berita Online Menggunakan TF-IDF dan Collaborative Tagging" penelitian ini bertujuan untuk merekomendasikan *tag* pada suatu artikel berbahasa Indonesia menggunakan metode *TF-IDF* dan *Collaborative tagging*. Dengan perkembangan teknologi internet yang sangat pesat menyebabkan semakin tingginya arus informasi dan semakin meningkatkannya jumlah berita yang terpublikasikan secara *online*.

Dengan teknologi *social tagging* dapat mempermudah proses pencarian berita *online* dan pengaturan arsip berita berdasarkan kesamaan informasi *tag*. Dalam penelitian ini, sistem *autotag* dikembangkan menggunakan metode *TF-IDF* dan *Collaborative tagging* dimana tahapannya diawali dengan *text preprocessing* kemudian *feature selection* dan proses menghitung bobot *TF-IDF*. Kemudian diterapkan metode *Collaborative tagging* yaitu pengukuran kesamaan dengan menghitung persentase kemiripan *tag* hasil ekstraksi *TF-IDF* dengan *tag-tag* berita yang telah dipublikasikan sebelumnya. Penelitian ini menunjukkan bahwa metode *TF-IDF* dapat digunakan untuk memberikan rekomendasi *tag* secara otomatis dengan relevansi *tag* terhadap isi berita yaitu 79,97% dan 80,6% untuk relevansi rata-rata menggunakan *Collaborative Tagging*.

Suatu berita akan bermanfaat jika berita yang di baca sesuai dengan topik berita yang diinginkan oleh pembaca. Berita merupakan sumber informasi yang sangat dinantikan oleh manusia setiap harinya (Ambarwati &

Winarko, 2014). Bisa dikatakan berita telah menjadi kebutuhan sehari-hari bagi manusia untuk mendapatkan informasi tentang perkembangan jaman. Di jaman modern ini berita telah tersedia di berbagai situs berita online maupun di media sosial. Fenomena jejaring sosial yang kian menjamur memudahkan seseorang untuk mengetahui informasi terbaru, berinteraksi dengan banyak orang, dan meningkatkan kualitas diri dalam segi teknologi komunikasi dan informasi (Triyono, 2014).

Salah satu media sosial yang saat ini populer dan banyak penggunaannya adalah *Twitter*, media sosial *Twitter* sekarang semakin banyak penggunaannya termasuk publik figur dan interaksi yang banyak pula di dalamnya sangat beragam isinya (Noorfaik & Novianto, 2014). Termasuk berisi tentang opini maupun berita yang saat ini tersebar di masyarakat. Setiap hari jutaan *tweets* di *share* oleh berbagai pengguna *Twitter* didunia (Setiawan, 2015). Banyaknya kegiatan yang dapat dilakukan menggunakan *Twitter*, maka tidak dipungkiri *Twitter* akan menghasilkan kumpulan data yang besar. Hal tersebut ditunjukkan dari hasil survei bahwa setiap detiknya pada saat kehidupan normal rata-rata jumlah *tweet* yang ada dalam *Twitter* adalah 600 *tweets* menurut (Adiyana & Hakim, 2015), oleh sebab itu diperlukan suatu metode khusus untuk menganalisa data pada *Twitter* sehingga tidak terdapat suatu kondisi yang disebut "*Rich of Data but Poor of Information*".

Menerapkan metode *text mining* pada data *tweets* bisa didapatkan beberapa informasi yang bermanfaat seperti keseringan penggunaan kata-kata menurut aturan asosiasi yang menyertai kata konteks pada berita. Dengan kondisi

tersebut maka *Twitter* dapat digunakan sebagai pendekatan untuk mencari kebutuhan berita yang di butuhkan oleh pengguna *Twitter* dari kebiasaan *tweets* yang dilakukan setiap hari. Pada era teknologi informasi seperti saat ini berita sudah cenderung tersimpan dalam format digital karena mudah dan cepat dalam penyimpanan. Terlalu banyak dokumen teks yang tersimpan dalam komputer maupun internet, membuat pencarian informasi menjadi sulit. Kesulitan dalam pencarian informasi yang sesuai dengan kebutuhan seringkali menjadi penghambat proses pembelajaran dan proses mendapatkan informasi yang aktual (Suliantoro, 2012).

Pengelompokan data menjadi salah satu solusi untuk mengorganisasi dokumen-dokumen berita digital yang berjumlah banyak. Metode pengelompokan dan penambangan data atau biasa disebut *text mining*, dipadukan dengan pembobotan *TF-IDF* yang diimplementasikan ke dalam sebuah aplikasi rekomendasi berita dapat melakukan proses pengelompokan berita secara otomatis sehingga menghemat waktu untuk mencari berita secara manual.

Pada tabel 2.1, dapat dilihat perbandingan penelitian tentang pembangunan aplikasi yang menggunakan *text mining*, metode *TF-IDF*, dan *Twitter*.

Tabel 2.1. Perbandingan Penelitian

No	Pembanding	Ma'arif (2015)	All Farizi (2015)	Adiyana & Hakim (2015)	Surya (2016)
1	Platform	Web	Web	Software R	Web
2	Bahasa Pemrograman	PHP	PHP	PHP	PHP dengan <i>Framework</i> CI

3	Sasaran Pengguna	Mahasiswa	Umum	-	Umum
4	Metode	Text Mining dan TF-IDF	Text Mining dan TF-IDF	Text Mining dan TF-IDF	Text Mining dan TF-IDF
5	Aplikasi	Rekomendasi Tag Berita	Pencarian Karya Ilmiah	Pencarian <i>Twitter</i> Analisis Topik KPK dan Jokowi	Rekomendasi Berita Berbasis Preferensi Pengguna <i>Twitter</i>
6	<i>oAuth API Twitter</i>	X	X	X	✓
7	Aktifitas Log User	X	X	X	✓
8	Share Berita	X	X	X	✓

Beberapa penelitian diatas dapat disimpulkan bahwa *text mining* dan metode pembobotan *TF-IDF* telah banyak di implementasikan dalam pembobotan serta *scoring* kata konteks pada dokumen maupun berita. Oleh sebab itu penulis akan mengimplementasikan kembali *text mining* dan metode pembobotan metode *TF-IDF* sebagai *scoring* kata konteks yang terdapat pada *tweets* sebagai kunci perekomendasi berita. Penelitian ini akan menghasilkan aplikasi baru yang dikembangkan dengan *CodeIgniter* serta menambahkan hal-hal baru yang belum di implementasikan pada penelitian yang sebelum-sebelumnya seperti menggunakan *oAuth API Twitter*. Dengan adanya aplikasi dari hasil penelitian ini diharapkan pembaca berita mendapatkan rekomendasi berita sesuai kebutuhan dan waktu yang dibutuhkan dalam membaca berita lebih efektif tanpa memakan waktu lama untuk mencari berita yang sesuai kebutuhan.