

BAB I

PENDAHULUAN

1.1 Latar Belakang

Seiring perkembangan teknologi saat ini, telah dikembangkan sistem yang mampu menghasilkan informasi dalam bentuk text (*linguistic*) dengan berdasarkan data *non-linguistic* (data *raw* atau mentah yang terukur oleh sensor atau berasal dari serangkaian kejadian) agar mudah dipahami oleh manusia yaitu *Natural Language Generation* (NLG) (Bateman & Zock, 2012). Data tersebut bisa didapatkan dari hasil survey, transaksi, atau rekaman sensor, seperti data statistik keuangan, data cuaca, data transaksi jual beli, dan lainnya. Arsitektur NLG terbagi dalam empat bagian utama (*macroplanning*, *microplanning*, *linguistic realization* dan *presentation*), dimana setiap bagian memiliki sub bagian sendiri, seperti pada *macroplanning* terdapat sub bagian *content planning*, *text planning*, dan *Rhetorical structure theory* (RST) dan pada *microplanning* terdapat *lexicalisation* (Bateman & Zock, 2012).

Data-to-Text (D2T) merupakan bagian dari sistem *Natural Language Generation* (NLG) dimana D2T menerjemahkan data ke dalam teks dengan mengasumsikan bahwa data yang digunakan pada dasarnya benar dan akurat (Gkatzia, Lemon, & Rieser, 2017). Arsitektur D2T hampir serupa dengan NLG yang terbagi kedalam empat bagian utama (*signal analysis*, *data interpretation*, *document planning*, *microplanning and realisation*) (Reiter, 2011). Beberapa contoh penerapan D2T seperti pada sistem *Data-to-text Weather Prediction* (DWP) yang mampu menghasilkan ringkasan berita klimatologis dan cuaca selama satu bulan serta memberikan informasi prediksi untuk satu hari berikutnya (Putra et al, 2017), sistem *BabyTalk* menghasilkan ringkasan teks dari data neonatal selama 45 menit kemudian ringkasan tersebut digunakan sebagai bahan pendukung keputusan presentasi modalitas yang terjadi saat itu (Gatt et al., 2009), sistem *BT-Nurse* meringkas kejadian selama *shift* keperawatan berlangsung, berdasarkan hasil rekaman medis elektronik pasien (Hunter et al., 2011) dan *Knowledge-Based Report Generator* mampu menghasilkan laporan stok berdasarkan data stok produk (*non-linguistic*) suatu pasar (Kukich, 1983). D2T merupakan salah satu solusi yang bisa digunakan untuk menerjemahkan data *non-linguistic* kepada masyarakat tanpa menghilangkan makna yang terdapat didalam data tersebut, tentu dengan demikian justru masyarakat akan sangat mudah dalam memahami makna dari data tersebut.

Pada penelitian sebelumnya telah dikembangkan model *Data-to-Text* untuk membangkitkan berita cuaca yang mencakup informasi ringkasan cuaca dan prediksi cuaca. Model yang telah dikembangkan kemudian diterapkan pada data yang tersedia pada suatu waktu saja, contohnya data cuaca tahunan 2014-2015,

2015-2016 dan 2016-2017. Pembangkitan berita cuaca dengan menggunakan data tahun 2015-2016 hanya menghasilkan teks berdasarkan data tersebut, tidak ada keterkaitan antar data tahunan yang ada. Sistem dengan *input* data *streaming* dituntut untuk dapat membangun model yang merepresentasikan setiap data yang diproses pada suatu waktu, karena data lampau tidak akan mungkin dapat ditelusuri kembali, serta mampu membangun model yang merepresentasikan keseluruhan data (Domingos & Hulten, 2001). Hal ini lah yang menjadi latar belakang dilakukannya pengembangan model *Data-to-Text* pada data *streaming* dengan diterapkannya *Machine Learning* yang diharapkan dapat menyelesaikan permasalahan tersebut.

Data *streaming* bisa berasal dari sensor (cuaca, kualitas udara, kualitas air), survey berkala (pertumbuhan penduduk, pengunjung tempat wisata), atau data rutinitas (keuangan, transaksi jual beli), data otomatis dapat diolah sebagai data *input*, karena komputer yang memasukkan data ke komputer lain atau secara langsung dari sensor tanpa harus menunggu manusia untuk memasukan data (Muthukrishnan, 2005). Untuk mengolah data yang bersifat eksak tersebut secara manual sulit dilakukan dan tentu saja dibutuhkan sumber daya manusia yang kemampuan analisa yang baik dan kemampuan untuk merubahnya menjadi sebuah kalimat (Hashem et al., 2015; Sudrina, 2016), namun dengan adanya D2T ini sumber daya tersebut bisa kita minimalisir, karena D2T diterapkan untuk menganalisis data tersebut yang kemudian menyampaikannya dalam bentuk kalimat yang semua masyarakat dapat memahaminya dengan mudah. Hal ini dilakukan untuk memudahkan masyarakat secara umum dalam memahami makna data tersebut serta efisiensi penggunaan sumber daya.

Data *stream* (dalam bentuk *batch*) akan menjadi data *input* sistem, yang kemudian akan dilakukan serangkaian proses D2T untuk dijadikan sebuah berita. Data yang digunakan meliputi data historis rupiah terhadap mata uang lain (kurs), *solar radiation and meteorological*, dan data klimatologi. Bila hanya melihat kedalam data tersebut makna yang didapat hanya nilai data saat itu saja, sedangkan untuk memahami *trend* atau kesimpulan dari data perlu analisis yang lebih lanjut, sehingga sangat tepat untuk menerapkan sistem D2T. Selain menyampaikan informasi yang mudah dipahami, secara tidak sadar bahwa D2T dapat ini mengurangi aktifitas yang dilakukan manusia dalam menggali informasi yang ada pada data tersebut. Karena tentu data yang di analisis bukan hanya satu atau dua data saja, melainkan dengan jumlah data yang sangat besar, mengingat kita telah memasuki era *Big Data*.

Dalam pembangunan sistem D2T ini, tidak hanya informasi mengenai ringkasan data *streaming* yang akan dibangkitkan sebagai berita, namun dengan tambahan informasi data terkini dan prediksi data *streaming* di masa yang akan datang, serta prediksi tersebut terpengaruh oleh model yang didapatkan dari setiap *batch streaming*. Maka untuk mencapai hal tersebut, pengembangan sistem D2T pada penelitian ini menggunakan pendekatan *Machine Learning* untuk memprediksi dengan metode *Piecewise Linear Approximation* (PLA) dengan *Least Square*. Untuk mengefisienkan *Development Time*, penulis menggunakan memanfaatkan beberapa *packages* yang tersedia dalam R.

1.2 Rumusan Masalah Penelitian

Rumusan masalah dalam penelitian ini adalah:

1. Bagaimana pengembangan model dari sistem *Data-to-Text* untuk pembangkitan berita pada data *streaming* dengan menggunakan pendekatan *Machine Learning*?
2. Bagaimana proses implementasi sistem *Data-to-text* dalam R serta proses implementasi sistem *Data-to-text* untuk data *streaming*?
3. Bagaimana eksperimen dan hasil eksperimen dari sistem *Data-to-text* yang dikembangkan?

1.3 Tujuan Penelitian

Tujuan dilakukannya penelitian ini adalah sebagai berikut:

1. Untuk melakukan pengembangan model *Data-to-text* untuk membangkitkan berita pada data *streaming* dengan pendekatan *Machine Learning*.
2. Untuk melakukan implementasi model *Data-to-text* dalam bahasa pemrograman R dan implementasi model *Data-to-text* berdasarkan data *streaming*.
3. Untuk menganalisis kualitas sistem dengan melakukan eksperimen.

1.4 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut:

1. Diharapkan dapat menambahkan pengetahuan tentang sistem *Data-to-Text* dan *Machine Learning* serta penerapannya dalam membangkitkan bahasa alami untuk mendeskripsikan data yang diberikan.
2. Dapat menjadi salah satu alternatif dalam menyampaikan hasil analisis data yang disampaikan sebagai berita secara otomatis oleh sistem *Data-to-Text*.

3. Dapat menjadi salah satu referensi dalam pembangunan sistem *Data-to-text* yang memanfaatkan bahasa pemrograman R beserta fiturnya seperti *packages*.
4. Dapat menjadi salah satu referensi dalam pembangunan sistem *Data-to-text* pada data *streaming* dengan R, PHP, javascript, AJAX dan HTML.

1.5 Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah:

1. Pembangunan sistem *Data-to-Text* dengan pendekatan *Machine Learning* ini hanya didasarkan pada data *streaming* dalam bentuk *batch* yang bersifat eksak dan *time series*.
2. Pembangunan sistem *Data-to-Text* ini menggunakan bahasa pemrograman R serta PHP, javascript, AJAX dan HTML dalam framework CodeIgniter.
3. Data yang digunakan seolah-olah sudah tersinkronisasi dengan sensor, yang mengirim data secara berturut-turut kedalam sistem. Tidak dibahas bagaimana sensor mengirimkan data kedalam sistem.

1.6 Sistematika Penulisan

Pada bagian sistematika penulisan ini akan diuraikan mengenai penjelasan tiap bab.

BAB I PENDAHULUAN

BAB I menjelaskan mengenai latar belakang dilakukannya penelitian, dimana pengembangan dilakukan dikarenakan model sistem *Data-to-Text* penelitian sebelumnya yaitu DWP (Putra et al, 2013), tidak memenuhi persyaratan pada sistem dengan masukan data *Streaming* (Domingos & Hulten, 2001). Bab ini terdiri dari latar belakang, rumusan masalah, batasan masalah, tujuan penelitian yang akan dilakukan, manfaat penelitian dan sistematikan penulisan.

BAB II KAJIAN PUSTAKA

BAB II berisi tentang kajian pustaka yang dilakukan penulis yang ditujukan untuk menggali pengetahuan mengenai ilmu yang terkait dengan penelitian dan berujung pada pemahaman yang cukup untuk melakukan penelitian. Bab ini terdiri dari beberapa kajian singkat tentang teori-teori dan konsep yang dibutuhkan dalam penelitian. Terdiri dari pembahasan mengenai *Natural Language*, *Natural Language Processing*, *Natural Language Generation*, *Data-to-text*, *Machine Learning*, *R Programming*, dan lain-lain.

BAB III METODOLOGI PENELITIAN

BAB III menjelaskan mengenai langkah-langkah dalam melakukan penelitian mulai dari perumusan masalah hingga penarikan kesimpulan, serta mendeskripsikan alat dan bahan yang digunakan dalam penelitian. Bab ini terdiri dari langkah-langkah yang akan dilakukan dalam penelitian. Terdiri dari desain penelitian, alat penelitian, dan bahan penelitian.

BAB IV HASIL PENELITIAN DAN PEMBAHASAN

BAB IV menjabarkan hasil penelitian yang telah dilakukan dan analisis kualitas sistem yang telah dibuat, dengan mengukur tingkat *Readability*, *Computation Time*, analisis kualitas sistem dengan membandingkan *Representative Text* dengan grafis, dan perbandingan dengan penelitian terkait. Semua pertanyaan mengenai masalah yang diangkat dalam tema skripsi dibahas di sini. Bab ini terdiri dari pengumpulan data, pengembangan model, implementasi sistem, desain eksperimen, dan hasil dan analisa hasil eksperimen.

BAB V KESIMPULAN DAN SARAN

BAB V menjelaskan mengenai penarikan kesimpulan dari hasil penelitian, serta memaparkan beberapa saran untuk penelitian selanjutnya. Bab ini berisi kesimpulan dan saran.