

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dewasa ini, telah dikembangkan banyak aplikasi yang mampu menghasilkan informasi dalam bentuk text dengan *input* non-linguistik atau data numerik. Contohnya seperti yang diperkenalkan oleh (Goldberg, Driedger dan Kitteredge, 1994) yaitu *Forecast Generator (FOG)*, *software* ini dapat mengkonversi peta cuaca menjadi ramalan dalam bentuk kalimat dengan menggunakan pengolah bahasa alami. Contoh lainnya dalam salah satu aplikasi pengontrol kesehatan, yaitu *BABYTALK family System* yang diperkenalkan oleh (Potret, dkk., 2009), aplikasi ini mampu membuat sebuah ringkasan peristiwa yang terjadi selama 45 menit dari sinyal psikologis kontinyu dan diskrit, seperti pengaturan peralatan dan pemberian obat dalam bentuk kalimat. Selain itu, penerapan ini juga telah digunakan untuk menghasilkan ringkasan dari pergantian keperawatan hanya dari sistem pencatatan pasien elektronik, di *Neonatal Intensive Care Unit (NICU)* (Hunter, dkk., 2012). Secara umum, berbagai contoh yang telah dipaparkan tersebut termasuk implementasi dari sistem *Data-to-text*.

Sistem *Data-to-text (D2T)* merupakan bagian dari sistem *Natural Language Generation (NLG)* yang mampu menghasilkan kalimat atau teks dari data numerik secara otomatis (Reiter, E, 2007). Karena sistem perlu mengkonversi dari data raw (data mentah), D2T harus melibatkan analisis data dan proses linguistik. Dengan mengembangkan arsitektur NLG yang dipaparkan oleh (Reiter, dkk., 2005), tahapan pembangunan sistem D2T terdiri dari empat langkah, yaitu: analisis sinyal, interpretasi data, perencanaan dokumen, *microplanning* dan realisasi (Reiter, E, 2007). Tahap ini hampir serupa dengan arsitektur NLG pada umumnya yang megandung tiga modul, diantaranya:

Brahma Putra, 2017

PENGEMBANGAN SISTEM DATA-TO-TEXT UNTUK MEMBANGKITKAN BERITA CUACA DENGAN PENDEKATAN TIME-SERIES DALAM R

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

perencanaan text, perencanaan kalimat, dan realisasi linguistik. Tiga modul tersebut terurai menjadi 6 tahap, yaitu: penentuan konten, penentuan tulisan,

Brahma Putra, 2017

***PENGEMBANGAN SISTEM DATA-TO-TEXT UNTUK MEMBANGKITKAN BERITA CUACA DENGAN
PENDEKATAN TIME-SERIES DALAM R***

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

pengumpulan kalimat, leksikalisasi, pembangkitan ekspresi acuan, dan realisasi linguistik.

Data-to-text telah menjadi sebuah solusi yang dapat dikatakan sukses untuk diterapkan dalam berbagai kasus, seperti pada beberapa referensi yang telah dipaparkan diawal BAB ini. Dalam penelitian ini, akan dibangun sistem *Data-to-text* yang dapat membangkitkan bahasa alami untuk menyampaikan informasi cuaca dan kualitas udara. Hal ini dilakukan untuk melengkapi dan meningkatkan kualitas informasi yang disampaikan agar lebih mudah dipahami oleh manusia, mengingat data yang diperoleh dari stasiun pemantauan merupakan data RAW (data mentah). Data yang diperoleh dari stasiun pemantauan ini, khususnya data kualitas udara merupakan data yang kurang mudah dipahami. Dikarenakan data yang diperoleh bersifat eksak, tidak semua orang dapat memahami data tersebut hanya dengan melihat hasil pemantauannya (angka). Oleh karena itu, dibutuhkan serangkaian proses yang mampu menganalisis data–data yang bersifat *numerik* dari stasiun pemantauan dan menyampaikannya dalam bentuk bahasa yang mudah dipahami oleh manusia. Semua proses yang dibutuhkan ini terliput kedalam satu arsitektur *Data-to-text*. Itulah mengapa *Data-to-text* dapat memberi manfaat yang terasa sangat signifikan untuk mempermudah pemahaman informasi.

Sumber data yang digunakan dari penelitian ini didapatkan dari salah satu stasiun pemantauan meteorologi milik lembaga bidang meteorologi dan geofisika (*GaliWeather*), data ini akan menjadi data *input* yang kemudian akan dibangkitkan hasil analisisnya dalam bahasa alami melalui serangkaian proses dalam arsitektur *Data-to-text*. Data yang digunakan meliputi data kualitas udara (seperti kandungan SO₂, NO, NO₂, NOX, CO, O₃, PM₁₀, dan sebagainya) serta data hasil pemantauan cuaca (meliputi kecepatan angin, arah angin, rata-rata suhu, rata-rata kelembaban relatif, radiasi global matahari, hembusan angin, penyinaran matahari, suhu embun, suhu tanah, suhu udara, curah hujan, tekanan barometrik, dan sebagainya). Sangat masuk akal jika sistem *Data-to-text* dapat menjadi

Brahma Putra, 2017

PENGEMBANGAN SISTEM DATA-TO-TEXT UNTUK MEMBANGKITKAN BERITA CUACA DENGAN PENDEKATAN TIME-SERIES DALAM R

pelengkap dan mempermudah penyampaian informasi, serta mengurangi aktivitas yang biasanya dilakukan oleh manusia. Karena, pada era *Big Data* ini kebutuhan otomatisasi terlihat sangat meningkat dikarenakan jumlah data yang sangat besar, sangat bervariasi, dan pertumbuhannya sangat cepat tidak memungkinkan jika dikelola secara *manual* oleh manusia, apalagi untuk menyimpulkan hasil analisis.

Dalam membangkitkan berita cuaca dan kualitas udara, satu hal yang sangat penting adalah menyampaikan informasi mengenai hasil prediksi. Maka untuk mencapai hal ini, pengembangan sistem *Data-to-text* dalam penelitian ini akan menggunakan pendekatan *time-series* untuk memprediksi data. Yang akan digunakan adalah algoritma *exponential smoothing*. Hal lain yang perlu diperhatikan dalam melakukan pengembangan perangkat lunak adalah dengan memperhatikan *Development Time*, untuk dapat mempersingkat waktu pengembangan, penulis memanfaatkan beberapa *packages* yang tersedia dalam R.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang ada, maka permasalahan dalam skripsi ini dirumuskan sebagai berikut:

1. Bagaimana pengembangan model dari sistem *Data-to-text* untuk pembangkitan berita cuaca dan kualitas udara dengan menggunakan pendekatan *Time Series*?
2. Bagaimana proses implementasi sistem *Data-to-text* dalam R?
3. Bagaimana eksperimen dan hasil eksperimen dari sistem *Data-to-text* yang dikembangkan?

1.3 Batasan Masalah

Dalam penelitian ini, permasalahan dibatasi hal-hal berikut ini:

1. Pembangunan sistem *Data-to-text* dengan pendekatan *Time Series* ini hanya didasarkan pada data kualitas udara dan data cuaca yang diperoleh dari stasiun pemantauan meteorologi *Mabegondo*.

2. Pembangunan sistem *Data-to-text* ini hanya menggunakan bahasa pemrograman R.

1.4 Tujuan

Berdasarkan rumusan masalah diatas, maka tujuan penelitian dalam tugas akhir ini dirumuskan sebagai berikut.

1. Untuk melakukan pengembangan model *Data-to-text* untuk membangkitkan berita cuaca dengan pendekatan *time-series*.
2. Untuk melakukan implementasi model *Data-to-text* dalam bahasa pemrograman R.
3. Untuk mengetahui kualitas sistem dengan melakukan eksperimen.

1.5 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini adalah:

1. Diharapkan dapat menambahkan pengetahuan tentang sistem *Data-to-text* dan *time-series* serta penerapannya dalam membangkitkan bahasa alami untuk mendeskripsikan data kualitas udara dan cuaca.
2. Dapat menjadi salah satu alternatif dan pelengkap dalam menyampaikan hasil analisis data 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*.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini diuraikan menjadi lima bab, yaitu:

BAB I PENDAHULUAN

BAB I terdiri dari latar belakang, rumusan masalah, batasan masalah, tujuan penelitian yang akan dilakukan, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

BAB II terdiri dari beberapa kajian singkat tentang teori-teori dan konsep yang dibutuhkan dalam penelitian. Terdiri dari pembahasan mengenai *Natural Language Processing*, *Natural Language Generation*, *Data-to-text*, *Time-series*, *R Programming*, dan lain-lain.

BAB III METODOLOGI PENELITIAN

BAB III terdiri dari langkah-langkah yang akan dilakukan dalam penelitian. Terdiri dari desain penelitian, alat penelitian, dan bahan penelitian.

BAB IV HASIL PENELITIAN DAN PEMBAHASAN

Berisi hasil penelitian serta analisis yang dilakukan selama penelitian. Yaitu terdiri dari pengembangan model, implementasi sistem, eksperimen dan hasil eksperimen.

BAB V KESIMPULAN DAN SARAN

Berisi kesimpulan yang didapat selama penelitian dan saran-saran dalam meningkatkan kualitas dan kuantitas hasil penelitian.

LAMPIRAN

Berisi dokumen-dokumen yang menunjang keabsahan penelitian.