**BAB I**

**PENDAHULUAN**

* 1. **Latar Belakang**

Dewasa ini, kebutuhan manusia akan informasi semakin tinggi. Perkembangan teknologi dan ilmu pengetahuan menyebabkan ketersediaan informasi meningkat, terutama informasi berupa data non-linguistik atau data numerik. Hal tersebut bertolak belakang dengan kebutuhan user akan informasi yang mudah dipahami dan dimengerti dengan cepat, sehingga waktu yang digunakan user untuk memahami informasi yang berbentuk data numerik relatif lebih lama dibandingkan dengan informasi berupa teks atau berita. Sehingga para peneliti dan pengembang berlomba-lomba untuk mengembangkan aplikasi-aplikasi yang mampu menghasilkan informasi dalam bentuk text dengan *input* data non-linguistik atau data numerik. Salah satunya yaitu aplikasi atau sistem *Data-to-text* (D2T) yang diperkenalkan oleh (Reiter, dkk. 2007).

Sistem D2T ini dapat menerima berbagai masukan beruapa data non linguistik mulai dari data numerik, *event logs,*  maupun data yang dihasilkan dari sensor sekalipun. Dengan kemampuannya yang dapat menerjemahkan data numerik kedalam data berbentuk tekstual secara otomatis, membuat sistem D2T ini menjadi salah satu bagian dari sistem *Natural Language Generation* (NLG). Karena arsitekturnya yang mirip dengan arsitektur sistem NLG, membuat sistem D2T ini sangat erat dengan proses linguistik begitu juga proses analisis data, maka Reiter dkk. (2007) memaparkan bahwa setidaknya ada empat langkah dalam tahapan pembangunan sistem D2T, yaitu: (*signal analysis, data interpretation, document planning, microplanning* dan *realisation*).

Sudah banyak implementasi sistem D2T yang menjadi solusi dalam menyediakan informasi tekstual dalam beberapa bidang. Contohnya pada bidang peramalan cuaca, yaitu aplikasi *Data-to-text Weather Prediction* (DWP) yang dapat menghasilkan ringkasan cuaca selama satu bulan dan prediksi cuaca dalam bentuk berita dari masukan berupa data klimatologi dan data kualiats udara selama satu tahun (Putra, dkk., 2017). Lalu, ada *Forecast Generator* (FOG) yang diperkenalkan oleh (Goldberg, Driedger, & Kitterdge, 1994), aplikasi tersebut dapat mengkonversi peta cuaca menjadi ramalan dalam bentuk kalimat dengan pengolahan bahasa alami. Selain itu, ada *SumTime-Mousam* yang diperkenalkan oleh (Sripada, dkk., 2003), aplikasi ini dapat menghasilkan ramalan cuaca laut tekstual untuk rig minyak lepas pantai. Contoh lainnya yaitu pada bidang 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 peratalatan dan pemberian obat dalam bentuk kalimat. Selain itu, (Hunter, dkk., 2012) memperkenalkan sistem yang dapat menghasilkan ringkasan dari pergantian keperawatan yang berasal dari pencatatan pasien elektronik di Neonatal Intensive Care Unit (NICU)*.* Pada biadang ekonomi, terdapat *Knowledge-Based Report Generator* yang diperkenalkan oleh Kukich (1983) yang dapat mengkonversi data stok produk (non-linguistik) menjadi laporan stok pada suatu pasar. Mengingat luasnya penerapan sistem D2T tersebut, membuat sistem ini menjadi salah satu pilihan utama yang bisa digunakan untuk menerjemahkan data *non-linguistik* agar lebih mudah dipahami oleh masyarakat luas, tanpa menghilangkan makna asli yang terkandung dalam data tersebut.

Seringkali pembangunan sistem D2T ini hanya terdapat untuk satu bidang spesifik seperti yang sudah penulis jelaskan pada bagian sebelumnya. Maka, dalam penelitian ini penulis akan mengembangkan sebuah sistem D2T yang dapat menerima input berupa data *general* dan menerjemahkannya kedalam bentuk tekstual sesuai dengan informasi yang terkandung dalam data tersebut. Hal inilah yang menjadi latar belakang pengembangan sistem D2T pada penelitian ini, sehingga sistem D2T diharapkan mampu menerima masukan data *non-linguistik* dalam domain apapun, dan tidak terbatas pada suatu bidang spesifik.

Sumber data yang digunakan dalam penelitian ini didapatkan langsung dari Badan Pusat Statistik (BPS). Beberapa data yang digunakan, meliputi data inflasi bulanan, data Indeks Harga Konsumen (IHK) per kelompok dan subkelompok, gabungan dari 82 kota, data inflasi menurut kelompok komoditi, data tingkat inflasi gabungan 82 kota, data inflasi umum, data harga yang diatur pemerintah, dan barang bergejolak inflasi indonesia. Karena pada dasarnya di era *Big Data* ini ketersediaan data semakin meningkat, mudah diakses, variatif, dan juga dinamis. Namun jika tidak didampingi dengan sebuah sistem yang dapat mengelola data tersebut sehingga informasi yang diperoleh mudah dipahami maka akan dirasa sangat sulit jika kita harus menganalisis data tersebut secara manual. Maka tidak heran sistem D2T ini bisa menjadi suatu solusi yang dapat mempermudah dalam penyampaian dan analisis suatu informasi.

Dalam pembangunan sistem ini digunakan pendekatan *time-*series untuk mempermudah dalam analisis data, dan juga menggunakan model *Autoregresif Integrated Moving Average* (ARIMA) dalam menentukan hasil prediksi. Selain itu juga untuk mempersingkat waktu pembangunan maka digunakan beberapa package yang sudah tersedia dalam R, sehingga waktu pengembangan dapat dipersingkat.

* 1. **Rumusan Masalah**

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

1. Bagaimana pengembangan model dari sistem *Data-to-Text* untuk membangkitkan berita pada data *general* dengan menggunakan pendekatan *Time Series* dan *Machine Learning*?
2. Bagaimana proses implementasi sistem *Data-to-text* untuk data *general* dalam R?
3. Bagaimana eksperimen dan hasil eksperimen dari sistem *Data-to-text* yang dikembangkan?
   1. **Tujuan Penelitian**

Berdasarkan rumusan masalah di atas, maka tujuan penelitian ini adalah sebagai berikut:

1. Melakukan pengembangan model *Data-to-text* untuk membangkitkan berita pada data *general* dengan menggunakan pendekatan *Time Series* dan *Machine Learning.*
2. Untuk melakukan implementasi model *Data-to-text* untukdata *general* dalammenggunakan bahasa pemrograman R .
3. Untuk mengetahui dan menganalisis kualitas sistem dengan melakukan eksperimen.
   1. **Manfaat Penelitian**

Manfaat yang dapat diperoleh dari penelitian ini adalah:

1. Diharapkan dapat menambahkan pengetahuan tentang sistem *Data-totext* dan *time-series* serta penerapannya dalam membangkitkan bahasa alami untuk mendeskripsikan data *general*.
2. Dapat menjadi salah satu alternatif dan pelengkap dalam menyampaikan hasil analisis data secara otomatis oleh sistem *Data-totext.*
3. Dapat menjadi salah satu referensi dalam pembangunan sistem *Data-to-text* yang memanfaatkan bahasa pemrograman R beserta fiturnya seperti *packages.*
   1. **Batasan Masalah**

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

1. Pembangunan sisten *Data-to-text* denganpendekatan *Time Series* ini hanya didasarkan pada data *numerical* dan *categorical* yang berbentuk eksak.
2. Pembangunan sisten *Data-to-text* ini hanya menggunakan bahasa pemrograman R, HTML, dan JavaScript.
   1. **Sistematika Penulisan**

Sistematika penulisan skripsi ini diuraikan menjadi lima bab, yaitu:

**BAB I PENDAHULUAN**

BAB I teridiri dari latar belakang, rumusan masalah, batasan masalah,

tujuan penelitian yang akan dilakukan, dan sistematikan 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**

BAB IV berisi hasil penelitian serta analisis yang dilakukan selama penelitian.

Bab ini terdiri dari pengumpulan data, pengembangan model, implementasi sistem, eksperimen dan hasil eksperimen.

**BAB V KESIMPULAN DAN SARAN**

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

**LAMPIRAN**

Berisi dokumen-dokumen yang menunjang keabsahan penelitian.