

Pengembangan Sistem *Data-To-Text* (D2T) untuk Membangkitkan Berita pada *Data Streaming*

Ahmad Zainal Abidin*, Lala Septem Riza[#], Enjang Ali Nurdin[#]

Departemen Pendidikan Ilmu Komputer
Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam
Universitas Pendidikan Indonesia
Bandung, Indonesia

ahmad.zainal565@student.upi.edu, lala.s.riza@upi.edu, enjang_cs@upi.edu

Abstrak—Penelitian ini bertujuan untuk mengembangkan sistem *Data-to-Text* dengan masukan berupa data *Streaming* dalam bentuk *batch*, untuk membangkitkan berita secara *general*. Pengembangan model sistem *Data-to-Text* dilakukan penerapan *Machine Learning* untuk mengatasi data *Streaming*, dengan teknik *Piecewise Linear Approximation* menggunakan *Least Square method*. Sistem yang dikembangkan menghasilkan informasi ringkasan data, informasi data terkini, dan informasi prediksi. Pengembangan sistem dilakukan dalam bahasa pemrograman R dengan memanfaatkan beberapa *packages* yang tersedia. Eksperimen dilakukan dengan mengukur tingkat *Readability* dari berita yang dibangkitkan, *Computation Time*, dan membandingkan hasil dengan penelitian terkait. Hasil eksperimen menunjukkan bahwa informasi yang dihasilkan terbukti merepresentasikan data yang diberikan, dan dapat dipahami oleh tingkat mahasiswa atau di atasnya, serta waktu komputasi cukup baik. Sistem ini mampu menghasilkan informasi berdasarkan data meteorologi, data klimatologi, dan data keuangan.

Kata Kunci— *Data-to-Text*; *Natural Language Generation*; *Machine Learning*; *streaming*; *Picewise Linear Approximation*; *Least Square Method*; *Time-series*

I. PENDAHULUAN

Seiring perkembangan teknologi saat ini, telah dikembangkan sistem yang mampu menghasilkan informasi dalam bentuk text (*linguistic*) dengan berdasarkan data *non-linguistik* (data *raw* atau mentah yang terukur oleh sensor atau berasal dari serangkaian kejadian) agar mudah dipahami oleh manusia yaitu *Natural Language Generation* (NLG) [1]. 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 *lexicalization* [1].

Data-to-text (D2T) adalah sistem *Natural Language Generation* (NLG) yang mampu menghasilkan teks dari input data *non-linguistik*, seperti data sensor dan serangkaian kejadian [2]. *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 [3].

Arsitektur D2T hampir serupa dengan NLG yang terbagi kedalam empat bagian utama (*signal analysis*, *data*

interpretation, *document planning*, *microplanning* and *realisation*) [2]. D2T merupakan salah satu solusi yang bisa digunakan untuk menerjemahkan data *non-linguistik* 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 ini, akan dibahas mengenai pengaplikasian D2T untuk membangkitkan berita berdasarkan data *streaming* dengan jangka waktu harian, mingguan, bulanan atau bahkan tahunan namun terbatas hanya pada data yang bersifat eksak dan *time series*. Data *stream* adalah urutan kejadian yang hanya dapat dibaca sekali atau beberapa kali menggunakan kemampuan komputasi dan penyimpanan yang terbatas. Sumber data ini dicirikan dengan aliran data dengan kecepatan tinggi, dan dihasilkan oleh distribusi non stasioner dalam lingkungan yang dinamis [4].

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 [5]. 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 [6].

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 terpengaruhi 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 mengefesiesikan *Development Time*, penulis menggunakan memanfaatkan beberapa *packages* yang tersedia dalam R.

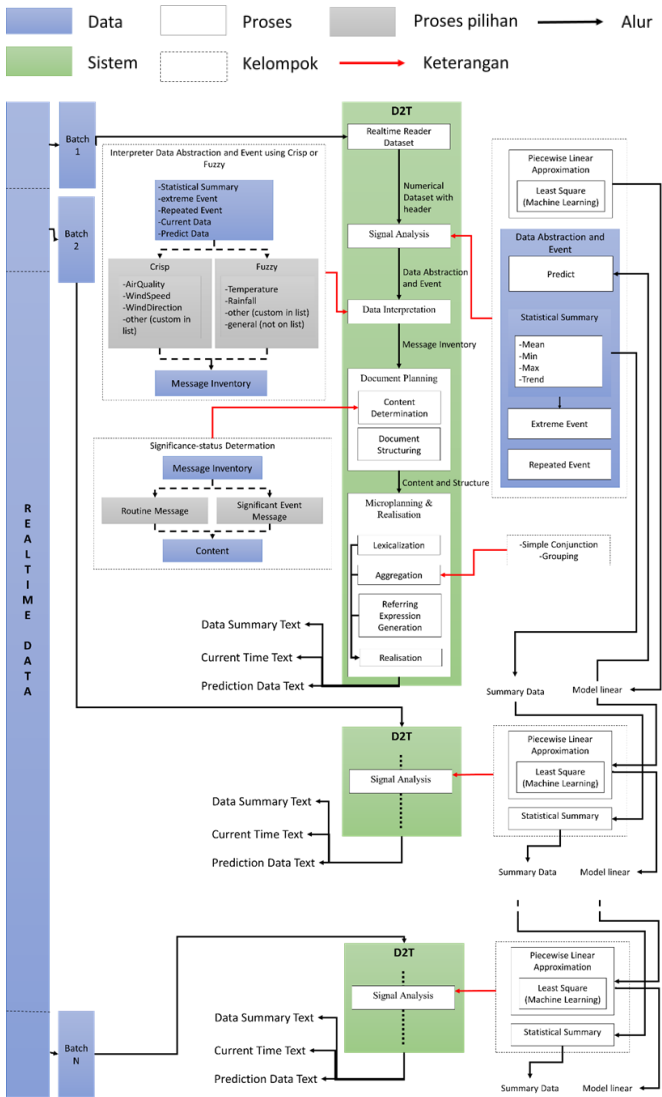
II. PENELITIAN TERKAIT

Berbagai sistem *Data-to-Text* telah dibangun pada penelitian sebelumnya, diantaranya 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 [7], 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 [8], sistem *BT-Nurse* meringkas kejadian selama *shift* keperawatan berlangsung, berdasarkan hasil rekaman medis elektronik pasien [9] dan *Knowledge-Based Report Generator* mampu menghasilkan laporan stok berdasarkan data stok produk (non-linguistik) suatu pasar [10].

III. MODEL SISTEM DATA-TO-TEXT UNTUK DATA STREAMING

Model *Data-to-Text* terbagi kedalam empat bagian utama yaitu *signal analysis*, *data interpretation*, *document planning*, *microplanning* and *realisation* [2]. Pada penelitian ini dilakukan pengembangan model *Data-to-Text* untuk data *streaming* sehingga terdapat satu bagian tambahan yaitu *Realtime Reader Dataset* seperti pada gambar 1. Model ini menggambarkan ringkasan tahap pengembangan sistem.



Gambar 1 Model *Data-to-Text* pada *Data Streaming*

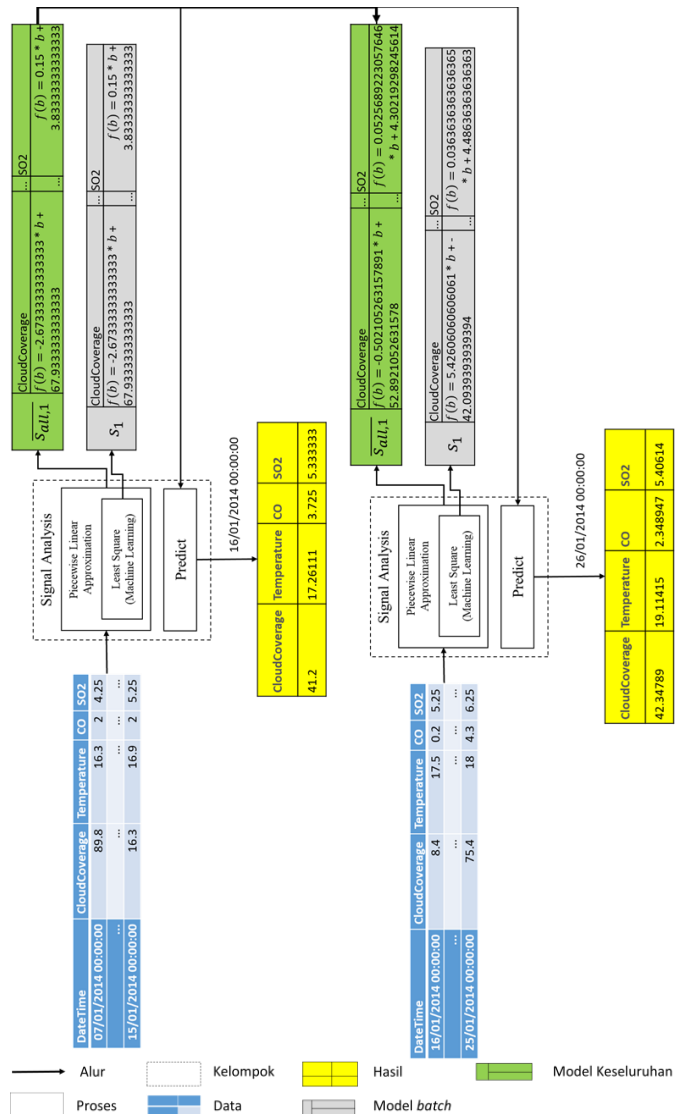
A. Realtime Reader Dataset.

Pada proses ini, dilakukan pengecekan dataset pada folder yang telah ditentukan, jika terdapat dataset baru, maka program *Data-to-Text* dalam bahasa R akan dijalankan. Pada proses ini

dilakukan dengan bahasa pemrograman AJAX untuk mengecek data, kemudian dilakukan pengiriman sinyal kedalam program PHP untuk menjalankan program D2T yang telah dikembangkan. Dengan adanya proses ini, masalah kecepatan aliran data yang sangat cepat mampu teratasi.

B. Signal Analysis

Pada proses ini dilakukan penerapan *Machine Learning* untuk mengatasi masalah penyimpanan data dengan teknik PLA [11] menggunakan *Least Square method* [12]. Penerapan ini dilakukan untuk mendapatkan model dari setiap *batch* data, sehingga jika interpretasi data diperlukan, dapat membangkitkannya menggunakan model yang telah disimpan. Pada setiap *batch* dilakukan penggabungan model yang telah dibuat sebelumnya, sehingga didapatkan satu model yang merepresentasikan keseluruhan *batch* [11]. Model ini juga digunakan untuk memprediksi data yang akan datang, dengan model proses seperti pada gambar 2.



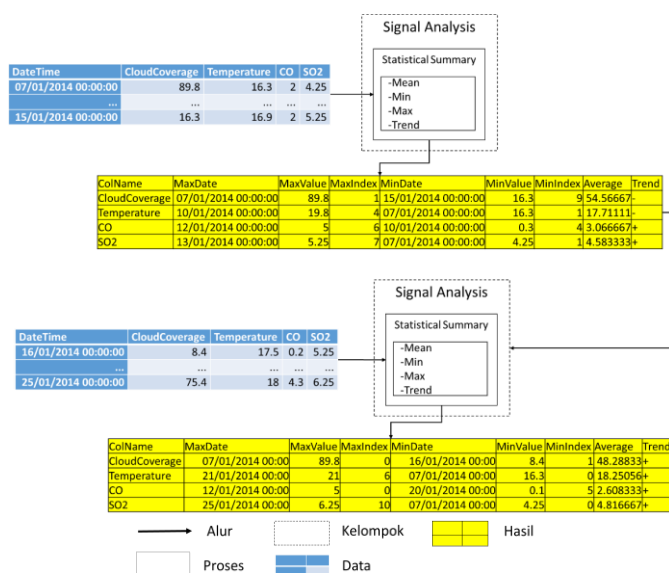
Gambar 2 Model proses PLA dalam D2T dan prediksi data

Proses lainnya yang terjadi dalam *Signal Analysis* ini adalah peringkasan data, dimana setiap ringkasan data yang telah berlalu akan memengaruhi ringkasan data selanjutnya, seperti yang tergambar pada gambar 3. Hal ini dilakukan karena ada kemungkinan pada suatu *batch* terdapat parameter yang rentangnya tidak sesuai dengan yang seharusnya, contohnya suhu minimum pada hari ini 20°C, dengan suhu

maksimum 30°C, maka rentang yang didapat hanya sebatas 10°C, padahal pada kenyataannya rentang dapat melebihi 50°C.

Dalam proses peringkasan juga terdapat proses penelusuran *statistical event*, dimana setiap perubahan rentang nilai maksimum atau minimum, yang diikuti oleh perubahan rata-rata nilai pada suatu parameter akan disimpan dan ditampilkan dalam berita. Dengan berdasarkan kepada hasil ringkasan, maka penelusuran sinyal ekstrim dilakukan. Suatu kejadian dikatakan ekstrim bila nilai kenaikan atau penurunan antara data ke i dan $i + 1$ melebihi 82% nilai rentang data.

Sedangkan untuk suatu data dikategorikan *repeated event* jika jumlah baris data nilai yang sama secara berturut-turut melebihi 10% jumlah baris data secara keseluruhan. Misalnya pada data periode satu tahun, dengan interval data harian (365 baris), data yang masuk dalam kategori *repeated event* hanya data yang nilainya sama secara berturut-turut melebihi 36.5 baris data.



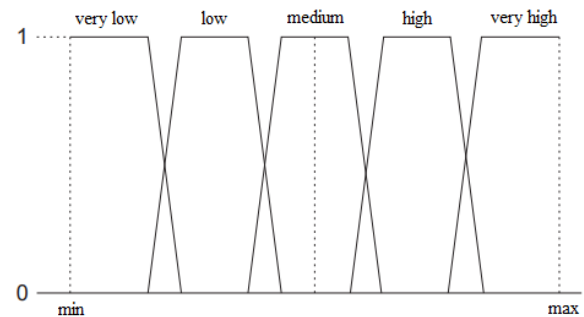
Gambar 3 Model proses ringkas data D2T

C. Data Interpretation

Karena berita yang dibangkitkan berupa berita *general* dimana data apapun dapat menjadi masukan (selama data *time series* dan mengikuti format data masukan), maka pada tahap *Data Interpretasi*, setiap pengguna dapat mengkostumisasi proses ini, namun dengan batasan hanya dapat dilakukan interpretasi data dengan *Fuzzy membership function* dan *Crisp membership function*. Dimana pengguna hanya perlu memasukan parameter kedalam *file* ParameterList.csv pada folder *corpus* dan memasukan *file* fungsi keanggotaan pada folder tersebut dengan format nama [parameter]Adjective.csv.

Pada penelitian ini interpretasi data yang tersedia diantaranya *AirQuality* [13], *WindSpeed* [14], *WindDirection* [15], dan *CloudCoverage* [16] menggunakan *Crisp membership function* sedangkan *Temperature* [17], *Rainfall* [17], dan parameter yang tidak teridentifikasi (*general*) maka akan dilakukan interpretasi data dengan *Fuzzy membership function*. Data general akan diinterpretasi berdasarkan *corpus* GeneralAdjective.csv dimana nilai setiap keanggotaan tergantung pada rentang nilai minimum dan maksimum yang diperoleh dari ringkasan data, hal ini dilakukan hasil dari modifikasi pada *Fuzzy membership function* untuk trend seperti pada Gambar 4, dimana nilai minimum dan maksimum

dari fungsi keanggotaan merupakan nilai minimum dan maksimum pada ringkasan, kemudian membagi keanggotaan sesuai dengan jumlah kategori dalam *corpus general* [18].



Gambar 4 Fuzzy membership function for General Parameter

D. Document Planning

Pada proses ini dilakukan pemilihan konten (*Content Determination*) dan pembentukan struktur teks (*Document Structuring*) [19]. Untuk proses pemilihan konten, dilakukan dengan membagi konten kedalam dua kelompok, yaitu *Routine Message* dan *Significant Event Message* [7], sedangkan *Document Structuring* dilakukan dengan cara membuat skema berdasarkan *Target Text* yang dibuat [7]. Pada teks ringkasan dan deskripsi data terkini konten dipilih dengan menglompokkan kedalam dua kelompok sebelumnya, sedangkan pada informasi prediksi hanya menggunakan kelompok *Routine Message*.

E. Microplanning and Realisation

Pada tahap ini setidaknya ada empat hal yang perlu dilakukan yaitu, *Lexicalisation*, *Aggregation*, *Referring Expression Generation* dan *Structure Realisation*. Pada tahap *Lexicalisation* dilakukan proses representasi anantara perubahan data, misalnya "increased extremely from", "decrease to", dan lain sebagainya [7]. Pada tahap *Aggregation* dilakukan ketika akan menghubungkan beberapa pesan menjadi satu kesatuan dengan menggunakan *Simple Conjunction Referring to Contrast Value* [19]. *Referring Expression Generation* dilakukan dengan cara membangkitkan secara *random* berdasarkan *corpus* yang di buat [7]. *Structure Realisation* dilakukan penerapan dengan menyusun semua konten kedalam struktur yang telah ditentukan [2], kemudian menyimpan teks kedalam JSON untuk ditampilkan didalam situs web.

IV. DESAIN EKSPERIMEN

Eksperimen dilakukan dengan cara membandingkan hasil dengan penelitian sebelumnya serta membangkitkan berita dengan 15 *test-case* [20] seperti pada tabel I yang kemudian setiap hasil eksperimen dilakukan pengukuran pada empat aspek yaitu, *Readability* dan *Computation Time* [20][21], serta perbandingan hasil *statistical resume* dengan *statistical resume* data keseluruhan, juga validasi *representative text*. Pengujian *Readability* dilakukan dengan menggunakan aplikasi *Readability Analyzer* pada situs datayze dan grammarly, untuk *Computation Time* dilakukan dengan menggunakan fungsi *system.time()* dalam R, sedangkan validasi *representative text* dilakukan dengan membandingkan informasi dengan visualisasi data.

Tabel I Test-Case Eksperimen

| Kode Dataset | Dataset | Sumber |
|--------------|------------------|---|
| KB1 | (Buy) Maret 2018 | Situs web Bank Indonesia (https://www.bi.go.id/) kurs beli periode Maret 2018 |
| KB2 | (Buy) April 2018 | Situs web Bank Indonesia (https://www.bi.go.id/) kurs beli periode April 2018 |
| KB3 | (Buy) Mei 2018 | Situs web Bank Indonesia (https://www.bi.go.id/) kurs beli periode Mei 2018 |
| SD1 | Juli 2016 | Data interval harian dari situs web https://solargis.com. Selama satu tahun mulai dari bulan 01 Juli 2016 hingga 31 Juli 2016 |
| SD2 | Agustus 2016 | Data interval harian dari situs web https://solargis.com. Selama satu tahun mulai dari bulan 01 Agustus 2016 hingga 31 Agustus 2016 |
| SD3 | September 2016 | Data interval harian dari situs web https://solargis.com. Selama satu tahun mulai dari bulan 01 September 2016 hingga 30 September 2016 |
| SD4 | Oktober 2016 | Data interval harian dari situs web https://solargis.com. Selama satu tahun mulai dari bulan 01 Oktober 2016 hingga 31 Oktober 2016 |
| SD5 | November 2016 | Data interval harian dari situs web https://solargis.com. Selama satu tahun mulai dari bulan 01 November 2016 hingga 30 November 2016 |
| SD6 | Desember 2016 | Data interval harian dari situs web https://solargis.com. Selama satu tahun mulai dari bulan 01 Desember 2016 hingga 31 Desember 2016 |
| KK1 | 2014-2015 | Situs web www.MeteoGalicia.gal , selama satu tahun pada periode 2014-2015 |
| KK2 | 2015-2016 | Situs web www.MeteoGalicia.gal , selama satu tahun pada periode 2015-2016 |
| KK3 | 2016-2017 | Situs web www.MeteoGalicia.gal , selama satu tahun pada periode 2016-2017 |

V. HASIL DAN PEMBAHASAN

A. Perbandingan *Output* Sistem dengan penelitian terkait

Untuk mempermudah perbandingan *output* dengan penelitian terkait seperti DWP [6], penelitian Ramos [19], dan lainnya, peneliti menggunakan data pada penelitian DWP [6], yaitu data klimatologi dari stasiun MeteoGalicia. Perbandingan *output* dapat dilihat pada tabel II.

Tabel II Perbandingan Output Sistem

| Penelitian | Ouput |
|--------------------|--|
| [22] <i>Output</i> | Winds northwest 15 diminishing to light monday afternoon. Cloudy with occasional light snow. Fog patches. Visibilities 2 to 5 nm in snow. Belle isle. Northeast gulf northeast coast. Gale warning in belle isle and northeast gulf issued. Gale warning in northeast coast continued. Freezing spray warning continued. Winds southwest 15 to 20 knots increasing to west gales 35 |
| [20] <i>Output</i> | With respect to the air quality state, it will be variable although is expected to improve to good, favored by the wind during the coming days |
| [23] <i>Output</i> | -Light rian showers are likely |

| Penelitian | Ouput |
|----------------------|---|
| | -Sunny intervals with rain being possible – less likely than not. -Sunny with rain being unlikely |
| [7] <i>Output</i> | Regarding to the prediction result, tomorrow sky state will be light rain although its covered by partly cloudy sky. Followed by temperature which decreased to warm. According to the air quality state, it will start to change to good. According to the monthly summary result, this month was cooler and wetter than average. With average number of rain days, accordingly the total rain so far is well below the average. There was rain on everyday for 7 days from 02nd to 08th and intense rain was dropped in 06th. The wind for the month was light breeze in average. Average air quality was admissible. Average temperature was increased but 05 th was the coldest day of the month with 13.3 celcius degree temperature. |
| GNG <i>Output</i> #1 | According to the dataset (06 Jul 2016 (00:00:00) - 06 Jul 2017 (00:00:00)), with parameters: Cloud Coverage, Temperature, Wind Speed, Wind Direction and Rainfall. It clear that Temperature trend is decreased and Rainfall trend is constant but the rest is increased Temperature, and Wind Speed parameters are more Fluctuate than last year. The Condition of today can be described that Cloud Coverage in Mostly cloudy condition. Temperature in warm condition. Wind Speed in Light Breeze condition. Wind Direction in North West condition. Rainfall in moderate condition. Based on prediction result Wind Speed, and Rainfall will keep stable at Light Breeze, and moderate, Cloud Coverage, Temperature, and Wind Direction will decrease to Partly cloudy, cold, and South West. A conclusion of the predicted result is half variable will decrease slowly. |
| GNG <i>Output</i> #2 | According to the dataset, between 07 Jun 2015 (00:00:00) to 25 May 2016 (00:00:00), with parameters: Cloud Coverage, Temperature, Wind Speed, Wind Direction and Rainfall. It indicates that Rainfall trend is increased but Temperature trend is decreased and the rest is constant. In today described that Cloud Coverage in Mostly cloudy condition. Temperature in warm condition. Wind Speed in Light Breeze condition. Wind Direction in South condition. Rainfall in moderate condition. Based on the result of prediction Temperature will decrease significantly to cold, Wind Direction will increase to South West, Cloud Coverage, Wind Speed, and Rainfall will keep stable at Mostly cloudy, Light Breeze, and moderate. A conclusion of the predicted result is half variable will kept stable. |

Berdasarkan pada tabel I jumlah konten tentu semakin banyak, namun secara tekstual aplikasi ini tidak sebaik *output* DWP [7] pada penjelasannya, dikarenakan konsep aplikasi ini dibangun untuk data *general* sehingga mampu membangkitkan berita berdasarkan data apapun selama data tersebut mengikuti format data inputan, sedangkan pada penelitian DWP data inputan harus sama dengan yang ada pada penelitian (parameter). Pada penelitian DWP terdapat dua data inputan yaitu klimatologi dan kualitas udara, sehingga konten yang muncul terdapat dua bagian. Sedangkan pada Ramos, teks yang dibangun hanya untuk kualitas udara dan kecepatan angin saja [20]. Pada penelitian Goldberg terdapat pesan mengenai angin dan salju secara terpisah-pisah [22] sedangkan dalam penelitian Gkatzia hanya disampaikan terkait keadaan langit. Hal ini menunjukkan bahwa D2T yang telah dibangun sebelumnya hanya untuk data yang spesifik, tidak *general*, sehingga ada kemungkinan sistem sebelumnya tidak berjalan jika diberikan data *time series* yang lain.

B. Hasil Eksperimen

Pada aspek *Readability* dilakukan penilaian berdasarkan *Flesch Readang Ease Score* yang didapatkan menggunakan *tools Readability Analyzer* pada situs www.datayze.com dan aplikasi Grammarly, sehingga didapatkan hasil pada tabel III.

Tabel III Hasil Pengukuran Aspek *Readability*

| Kode Dataset | <i>Flesch Readang Ease Score (Grammarly)</i> | <i>Flesch Readang Ease Score (Datayze)</i> |
|-----------------------|--|--|
| KB1 | 44 | 42.13 |
| KB2 | 42 | 37.07 |
| KB3 | 33 | 29.11 |
| SD1 | 28 | 19.59 |
| SD2 | 29 | 15.23 |
| SD3 | 30 | 17.3 |
| SD4 | 27 | 15.72 |
| SD5 | 23 | 13.09 |
| SD6 | 28 | 17.29 |
| KK1 | 60 | 57.77 |
| KK2 | 59 | 55.17 |
| KK3 | 63 | 60.39 |
| Rata-rata | 38.83 | 31.66 |
| Rata-rata Keseluruhan | 35.245 | |

Hasil *Computation Time* didapatkan dengan menjalankan fungsi `system.time()` dalam bahasa R, seperti `system.time(source("D2T_Main.R"))`, sehingga mendapatkan hasil seperti pada tabel IV.

Tabel IV Hasil Pengukuran *Computation Time*

| Kode Dataset | <i>Running Time (s)</i> |
|--------------|-------------------------|
| KB1 | 2.01 |
| KB2 | 1.98 |
| KB3 | 2.14 |
| SD1 | 1.84 |
| SD2 | 2.08 |
| SD3 | 2.08 |
| SD4 | 1.98 |
| SD5 | 2.03 |
| SD6 | 2.25 |
| KK1 | 2.44 |
| KK2 | 2.63 |
| KK3 | 2.79 |
| Rata-rata | 2.235 |

VI. KESIMPULAN

Pengembangan sistem *Data-to-Text* untuk data *streaming* dengan menggunakan *Machine Learning* sangat bermanfaat, dimana sistem dapat bekerja tanpa inputan dari *user*, melainkan langsung dari *sensor* atau dari komputer lainnya, dan juga jika sistem digunakan dalam jangka waktu yang sangat lama, penggunaan *hardisk* tidak akan terlalu besar, karena data yang telah di proses akan dihapus, digantikan dengan model yang telah dibuat oleh *Picewise Linear Approximation*.

Penelitian ini menjawab kekurangan penelitian sebelumnya yaitu DWP [7] dimana pembangunan *User Interface* tidak menggunakan *package shinyR*, melainkan kombinasi dengan framework *Codeigniter* dengan bahasa pemrograman PHP, javascript, AJAX dan html dengan perantara output JSON.

Kesimpulan dari keseluruhan hasil eksperimen yang dilakukan, keluaran dari sistem terbukti merepresentasikan data yang diberikan. Penelitian ini memperoleh nilai rata-rata keseluruhan 35.425 pada aspek *Readability* yang artinya keluaran dari sistem ini tergolong dalam kategori dapat dipahami oleh tingkat mahasiswa atau diatasnya. Sedangkan pada aspek *Computation Time* diperoleh rata-rata waktu

komputasi 2.235 detik. Serta akurasi prediksi menggunakan model *linear* hasil dari PLA dengan *Least Square Method* adalah 44%.

Untuk penelitian berikutnya dapat dilakukan Pengembangan *corpus* untuk kasus umum, atau menambahkan *corpus-corpus* untuk kasus khusus, seperti pada *Data Interpretation* serta Menambahkan fitur untuk mendeteksi keterkaitan parameter seperti *Assosiation Rule*. Menggunakan algoritma lain untuk memprediksi data *streaming* untuk dibandingkan dengan penelitian ini.

DAFTAR PUSTAKA

- [1] J. Bateman and M. Zock, "Natural Language Generation," *Oxford Handb. Comput. Linguist.*, no. December 2017, pp. 1–21, 2012.
- [2] E. Reiter, "An Architecture for Data-to-Text Systems," *Comput. Intell.*, vol. 27, no. 1, pp. 23–40, 2011.
- [3] D. Gkatzia, O. Lemon, and V. Rieser, "Data-to-Text Generation Improves Decision-Making Under Uncertainty," *IEEE Comput. Intell. Mag.*, vol. 12, no. 3, pp. 10–17, 2017.
- [4] J. Gama and M. M. Gaber, *Learning from Data Streams: Processing Techniques in Sensor Networks*. 2007.
- [5] P. Domingos and G. Hulten, "Catching up with the data: Research issues in mining data streams," *Work. Res. Issues Data Min. Knowl. Discov.*, vol. 470, 2001.
- [6] S. Muthukrishnan, "Data Streams: Algorithms and Applications," *Found. Trends® Theor. Comput. Sci.*, vol. 1, no. 2, pp. 117–236, 2005.
- [7] B. Putra, L. S. Riza, and Y. Wihardi, "Pengembangan Sistem Data-to-Text untuk Membangkitkan Berita Cuaca dengan Pendekatan Time-Series dalam R," 2017.
- [8] A. Gatt et al., "From data to text in the neonatal intensive care Unit: Using NLG technology for decision support and information management," *AI Commun.*, vol. 22, no. 3, pp. 153–186, 2009.
- [9] J. Hunter et al., "Bt-Nurse: Computer generation of natural language shift summaries from complex heterogeneous medical data," *J. Am. Med. Informatics Assoc.*, vol. 18, no. 5, pp. 621–624, 2011.
- [10] K. Kukich, "Design of a knowledge-based report generator," *Proc. 21st Annu. Meet. Assoc. Comput. Linguist.*, p. 145, 1983.
- [11] T. Palpanas, M. Vlachos, E. Keogh, D. Gunopulos, and W. Truppel, "Online amnesic approximation of streaming time series," *Data Eng. 2004. Proceedings. 20th Int. Conf.*, pp. 339–349, 2004.
- [12] K. R. Cliff and K. M. Billy, "Estimation of the Parameters of a Linear regression System Using the Simple Averaging Method," vol. 13, no. 11, pp. 7749–7758, 2017.
- [13] J. W. Crowder, J. G. Moore, L. DeRose, and W. J. Franek, "Air Pollution Field Enforcement," no. September 1999, 1999.
- [14] R. Rowlett, "Beaufort Scales (Wind Speed)," 2001. [Online]. Available: <https://www.unc.edu/~rowlett/units/scales/beaufort.html>. [Accessed: 20-May-2018].
- [15] J. Zandlo, G. Spoden, P. Bouley, and D. Ruschy, "Wind Direction and Degrees," *University of Minnesota*, 2001. [Online]. Available: <http://snowfence.umn.edu/Components/winddirectionanddegreeswithouttable3.htm>. [Accessed: 20-May-2018].
- [16] J. Huby, "Cloud Coverage," 2010. [Online]. Available: <http://www.theweatherprediction.com/habyhints/189/>. [Accessed: 20-May-2018].
- [17] A. Ramos-soto, A. Bugarin, and S. Barro, "Fuzzy Sets Across the Natural Language Generation Pipeline," vol. c, pp. 1–16, 2016.
- [18] R. Castillo-Ortega, N. Marín, C. Martínez-Cruz, and D. Sánchez, "A proposal for the hierarchical segmentation of time series. Application to trend-based linguistic description," *IEEE Int. Conf. Fuzzy Syst.*, pp. 489–496, 2014.
- [19] E. Reiter, "Building Natural-Language Generation Systems," pp. 91–93, 1996.
- [20] A. Ramos-Soto, A. Bugarín, and S. Barro, "On the role of linguistic descriptions of data in the building of natural language generation systems," *Fuzzy Sets Syst.*, vol. 285, pp. 31–51, 2016.
- [21] A. Belz, "Probabilistic Generation of Weather Forecast Texts," *NaacL-Hlt*, no. April, pp. 164–171, 2007.
- [22] R. I. Kittredge and N. Driedger, "Using Natural-Language Processing

to Produce Weather Forecasts,” *IEEE Expert. Syst. their Appl.*, vol. 9, no. 2, pp. 45–53, 1994.

[23] D. Gkatzia, O. Lemon, and V. Rieser, “Natural Language Generation

enhances human decision-making with uncertain information,” 2016.