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

Muhammad Ridwan\*, Lala Septem Riza#, Enjun Junaeti#

Departemen Pendidikan Ilmu Komputer

Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam

Universitas Pendidikan Indonesia

Bandung, Indonesia

Just.muhammadridwan@student.upi.edu, lala.s.riza@upi.edu, enjun@upi.edu

Abstrak—Sistem *Data-to-Text* menjadi salah satu pilihan untuk menerjemahkan data *non-linguistik* kedalam bentuk tekstual. Dengan berkembang

Penelitian ini bertujuan untuk mengembangkan sistem *Data-to-Text* dengan masukan berupa data *Streaming* dalam bentuk *batch* untuk membangkitkan berita secara *general*. Permasalahan utama dalam data *streaming* adalah aliran data yang sangat cepat dan jika data terus menerus disimpan tentu akan menghabiskan sebagian besar memori penyimpanan pada komputer. Untuk mengatasi masalah tersebut diterapkan *Machine Learning* dengan teknik *Piecewise Linear Approximation* menggunakan *Least Square method*. Sistem yang dikembangkan menghasilkan informasi ringkasan data, informasia data terkini dan informasi prediksi. Pengembangan sistem dilakukan dalam bahasa pemrograman R dengan memanfaatkan beberapa *packages* yang tersedia. Eksperimen dilakukan dengan mengukur tingkat *Readibility* dari berita yang dibangkitkan*, Computation Time*, dan membandingkan hasil dengan penelitian terkait. Hasil eksperimen menunjukan bahwa informasi yang dihasilkan terbukti merepresentasikan data yang diberikan dan dapat dipahami oleh tingkat mahasiswa atau diatasnya, serta waktu komputasi cukup baik. Sistem ini mampu menhasilkan informasi berdasarkan data meteorologi, data klimatologi, data keuangan, dan data *time series* lainnya.

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

# Pendahuluan

Seiring perkembangan teknologi saat ini, ketersediaan informasi kian meningkat, terutama informasi berupa data *non-linguistik* atau data *numerik*. Data *non-linguistik* dapat disajikan kedalam bentuk teks atau *linguistik* untuk mempermudah dalam penarikan sebuah informasi [1]. Sehingga dikembangkan sistem *Data-to-Text* (D2T) yang mampu menghasilkan informasi dalam bentuk tekstual dengan masukan berupa data *non-linguistik* atau *data numerik* [1]. Data yang digunakan bisa didapatkan dari berbagai sumber data seperti hasil rekaman sebuah sensor, *event logs,* maupun sumber data lainnya yang dicatat secara berkala.

*Data-to-Text* (D2T) merupakan bagian dari sistem *Natural Language Generation* (NLG) dimana D2T menerjemahkan data ke dalam bentuk teks dengan mengasumsikan bahwa data yang digunakan pada dasarnya benar dan akurat [2]. Sebuah sistem NLG setidaknya terbagi dalam empat bagian utama yaitu (*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* [3].

Arsitektur D2T hampir serupa dengan NLG yang terbagi kedalam empat tahapan 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, denga tujuan agar informasi berupa teks yang disajikan lebih mudah dipahami dibandingkan dengan data *non-linguistik.*

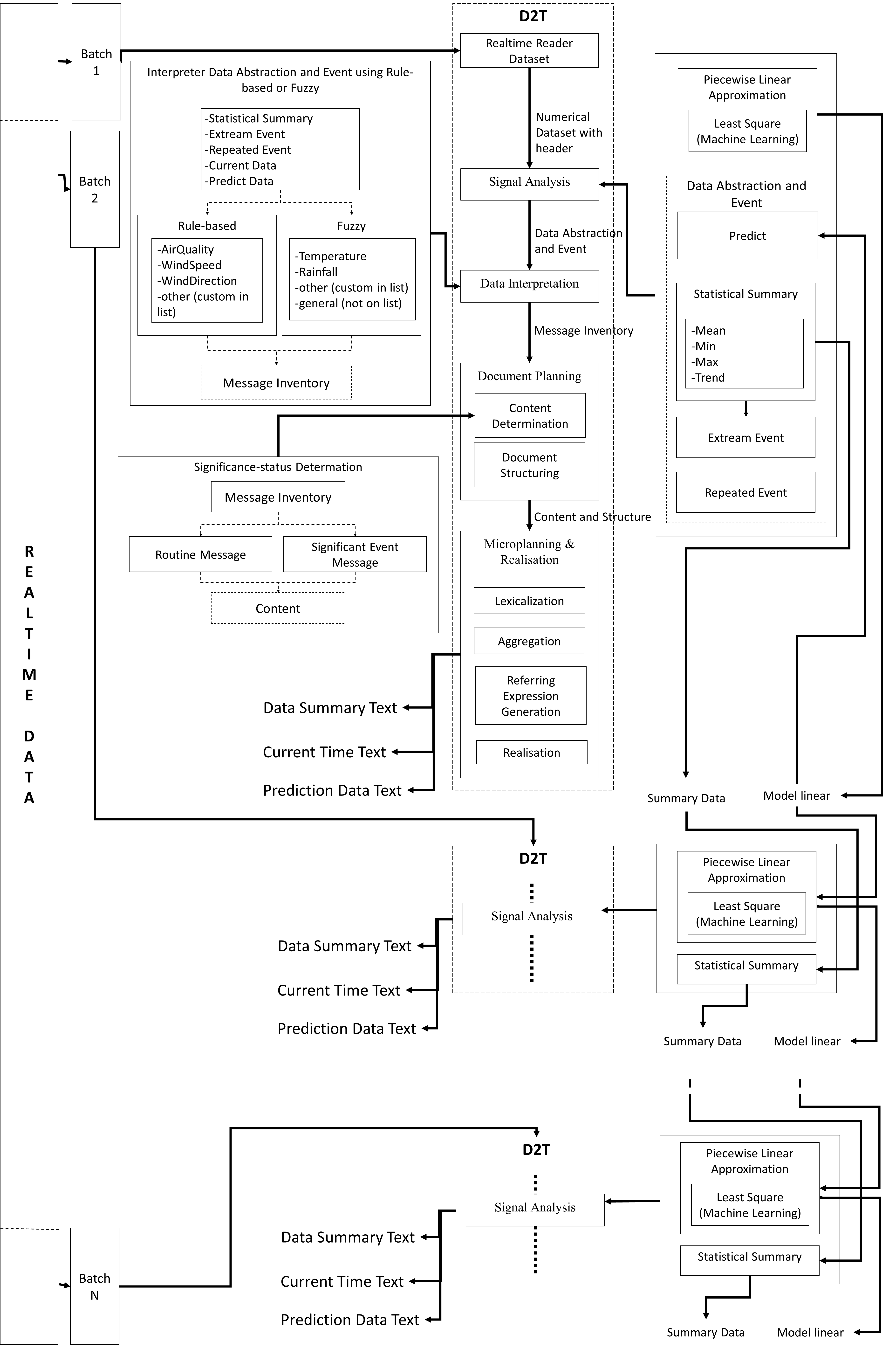
Pada penelitian ini, akan dibahas mengenai pengaplikasian D2T untuk membangkitkan berita berdasarkan data *general* dengan jangka waktu jam, harian, mingguan, bulanan atau bahkan tahunan namun ditabatasi hanya untuk data yang berjenis eksak dan *time series*. Data *general* yang dimaksud adalah data yang tidak terikat pada suatu domain apapun, baik data tersebut memiliki identitas berupa informasi *header*, kategori ataupun tidak, sehingga *corpus* dan keluaran yang dihasilkan bersifat se-*general* mungkin.

Setidaknya ada dua masalah utama yang harus diperhatikan dalam pembangunan sebuah *corpus.* Pertama, beragamnya jenis informasi yang disimpan dalam sebuah *corpus,* sehingga aspek sumber daya menjadi salah satu faktor yang penting dan harus diperhitungkan. Kedua, beragamnya jenis *user,* sehingga *corpus* yang dibangun harus bisa mencakup berbagai kebutuhan dari setiap *user* [4].

Dalam penelitian ini, akan dibangun sebuah sistem D2T yang tidak terikat pada suatu bidang apapun dan dapat menerima masukan berupa data *general* lalu menghasilkan keluaran se-*general* mungkin, agar sistem yang dibangun diharapkan dapat menjadi solusi dari masalah yang sudah disebutkan sebelumnya. Untuk mencapai hal tersebut, pengembangan sistem D2T ini menggunakan bantuan *Machine Learning* seperti *Linear Regession, Knuth-Morris-Pratt* (KMP), *Pearson Correlation, dan Statistical Tools* untuk pengolahan, analisis, dan prediksi data. Selain itu juga, penulis menggunakan memanfaatkan beberapa *packages* yang tersedia dalam R untuk mengefisienkan *Development* *Time*. Berbagai penelitian sudah dilakukan untuk membangun sebuah sistem *Data-to-Text* untuk setiap bidang spesifik tertentu, misalnya pada bidang klimatologi terdapat sistem *Forecast Generator* (FOG) yang dapat mengkonversi peta cuaca menjadi ramalan dalam bentuk kalimat dengan pengolahan bahasa alami [5]*,* laluterdapat sisem *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 [6], selain itu ada *SumTime-Mousam*, aplikasi ini dapat menghasilkan ramalan cuaca laut tekstual untuk rig minyak lepas pantai [7]. Pada bidang kesehatan, sistem *BabyTalk* diperkenalkan dengan tujuan menghasilkan ringkasan teks dari data neonatal selama 45 menit yang nantinya akan 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]. Pada bidang ekonomi, *Knowledge-Based Report Generator* mampu menghasilkan laporan stok berdasarkan data stok produk (*non-linguistik*) suatu pasar [10].

# 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.



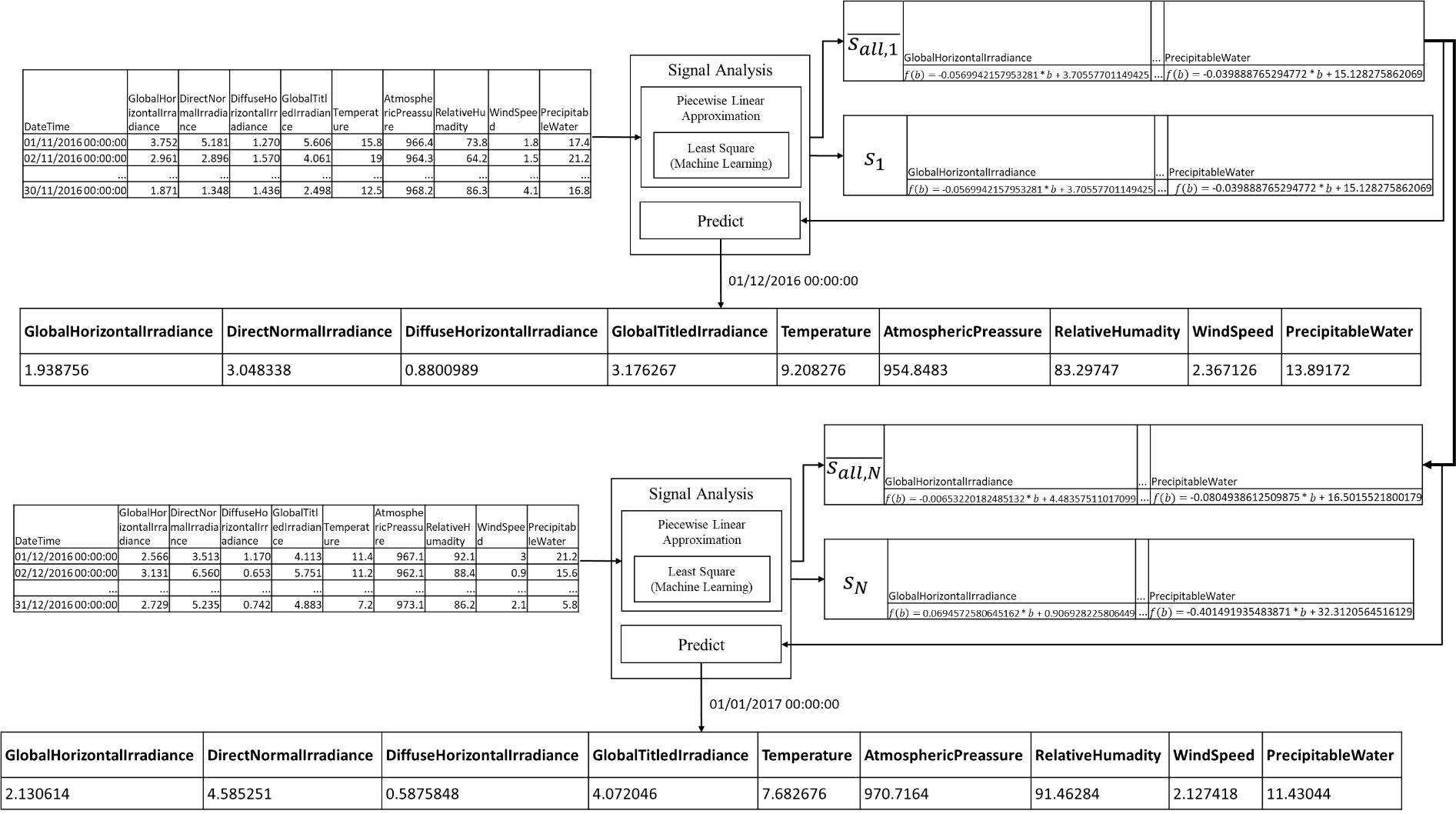
Gambar 1 Model *Data-to-Text* pada *Data Streaming*

1. *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.

1. *Signal Analysis*

Pada proses ini dilakukan penerapan *Machine Learning* untuk mengatasi masalah penyimpanan data dengan teknik PLA [10] menggunakan *Least Square method* [11]. 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* [10].Model ini juga digunakan untuk memprediksi data yang akan datang, dengan model proses seperti pada gambar 2.

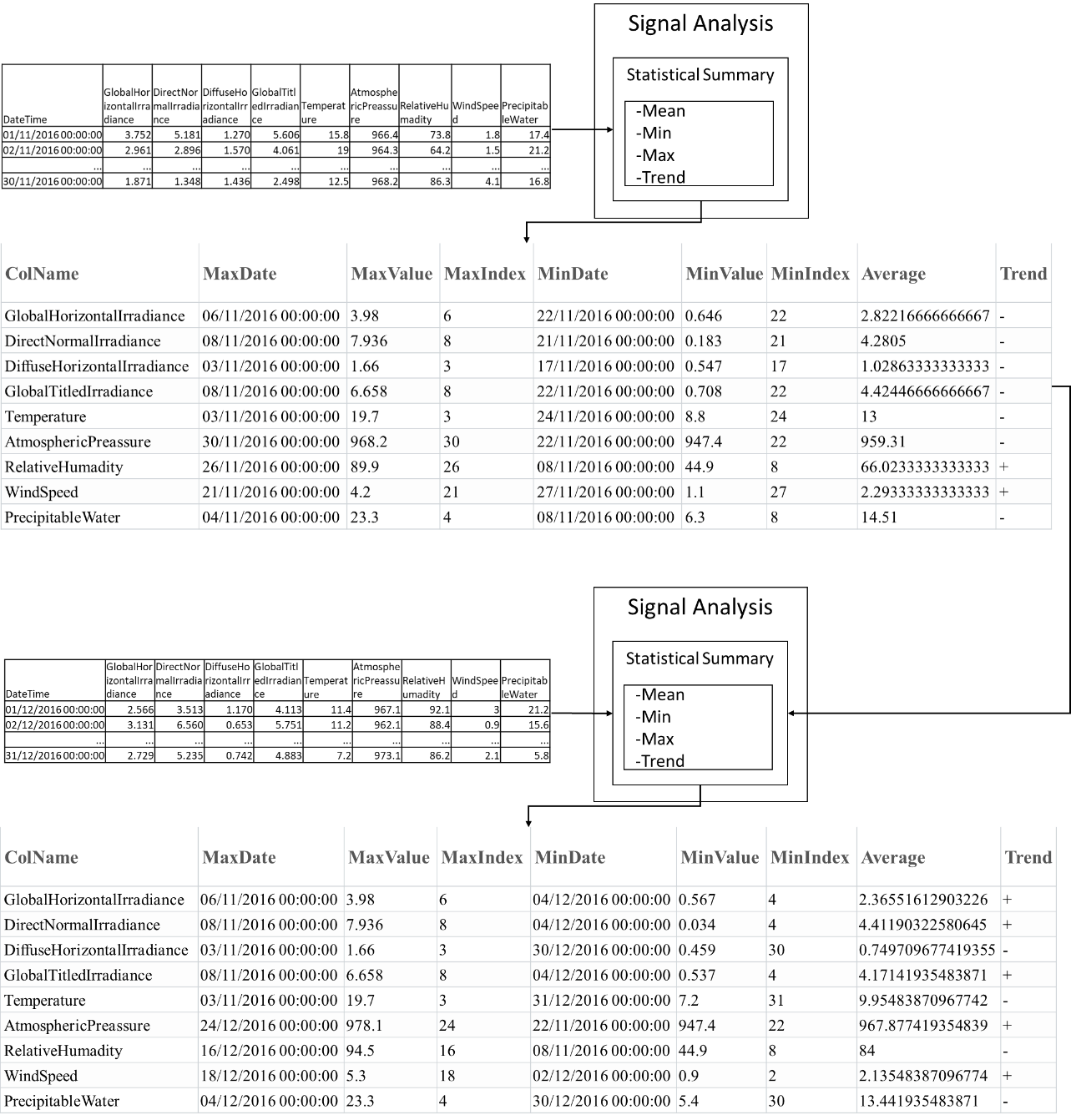
**

Gambar 1 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 , dengan suhu maksimum , maka rentang yang didapat hanya sebatas , padahal pada kenyataannya rentang dapat melebihi .

Dalam proses peringkasan juga terdapat proses penelusuran *statsitical* *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 dan 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.

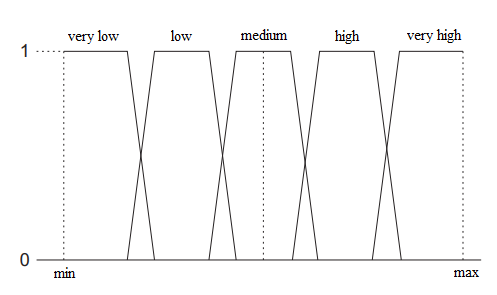


Gambar 2 Model proses ringkas data D2T

1. *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 keanggotan pada folder tersebut dengan format nama [parameter]Adjective.csv.

Pada penelitian ini interpretasi data yang tersedia diantaranya *AirQuality* [12]*, WindSpeed* [13]*, WindDirection* [14]*,* dan *CloudCoverage* [15] menggunakan *Crisp membership function* sedangkan *Temperature* [16]*, Rainfall* [16], dan parameter yang tidak teridentifikasi (*general*) maka akan dilakukan interpretasi data dengan *Fuzzy membership function*. Data general akan di interpretasiberdasarkan *corpus* GeneralAdjective.csv dimana nilai setiap keanggotaan tergantung pada rentang nilai minimum dan maksimun 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* [17].



Gambar 3 *Fuzzy membership function for General Parameter*

1. *Document Planning*

Pada proses ini dilakukan pemilihan konten (*Content Determination*) dan pembentukan struktur teks (*Document Structuring*) [18]. Untuk proses pemilihan konten, dilakukan dengan membagi konten kedalam dua kelompok, yaitu *Routine Message* dan *Significant Event Message* [6], sedangkankan *Document Structuring* dilaukan dengan cara membuat skema berdasarkan *Target Text* yang dibuat [6]. Pada teks ringkasan dan deskripsi data terkini konten dipilih dengan menglompokan kedalam dua kelompok sebelumnya, sedangkan pada informasi prediksi hanya menggunakan kelompok *Routine Message*.

1. *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 anatara perubahan data, misalnya “*increased extremely from*”, “*decrease to*”, dan lain sebagainya [6]. Pada tahap *Aggregation* dilakukan ketika akan menghubungkan beberapa pesan menjadi satu kesatuan dengan menggunakan *Simple Conjunction Referring to Contrast Value* [18]. *Referring Expression Generation* dilakukan dengan cara membangkitkan secara *random* berdasarkan *corpus* yang di buat [6]. *Structure Realisation* dilakukan penerapan dengan menyusun semua konten kedalam struktur yang telah ditentukan [2], kemudian menyimpan teks kedalam JSON untuk ditampilkan didalan situs web.

# Desain Eksperimen

Eksperimen dilakukan dengan cara membandingkan hasil dengan penelitian sebelumnya serta membangkitkan berita dengan 15 *test-case* [19] seperti pada tabel I yang kemudian setiap hasil eksperimen dilakukan pengukuran pada empat aspek yaitu, *Readibility* dan *Computation Time* [19][20]*,* serta perbandingan hasil *statistical resume* dengan *statistical resume* data keseluruhan, juga validasi *representative text*. Pengujian *Readibility* dilakukan dengan menggunakan aplikasi *Readibility 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](http://www.MeteoGalicia.gal), selama satu tahun pada periode 2014-2015 |
| KK2 | 2015-2016 | Situs web [www.MeteoGalicia.gal](http://www.MeteoGalicia.gal), selama satu tahun pada periode 2015-2016 |
| KK3 | 2016-2017 | Situs web [www.MeteoGalicia.gal](http://www.MeteoGalicia.gal), selama satu tahun pada periode 2016-2017 |

# Hasil dan Pembahasan

## 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 |
| --- | --- |
| GNG *Output* #2 | From the dataset, between 7/6/2015 00:00:00 to 25/05/2016 00:00:00, with parameters: Temperature, WindSpeed, WindDirection and Rainfall. It ilustrate that Temperature trend is decreased but the rest is increased.            In today described that CloudCoverage is Mostly cloudy, Temperature is warm, WindSpeed is Light Breeze, WindDirection is South, and Rainfall is low            Regarding the prediction result Temperature will decrease significantly to cold (8.31423200652999), WindDirection, and Rainfall will increase to South West (215.770666282549), and very low (11.3152622397209), CloudCoverage, and WindSpeed will keep stable at Mostly cloudy (59.3566076087131), and Light Breeze (5.631217490117). A conclusion of the predicted result is some variable will increase slowly and some variable will kept stable. |
| GNG *Output* #1 | According to the dataset, from 06/07/2016 00:00:00 to 06/07/2017 00:00:00, with parameters: Temperature, WindSpeed, WindDirection and Rainfall. It can be seen that Temperature trend is decreased but the rest is increased. CloudCoverage, and Temperature parameters are more Lower than a year ago      The Condition of today can be described that CloudCoverage is Mostly cloudy, Temperature is warm, WindSpeed is Light Breeze, WindDirection is North West, and Rainfall is low            Based on prediction result Rainfall will increase to very low (13.03025338626), WindSpeed will keep stable at Light Breeze (2.89752858971176), CloudCoverage, Temperature, and WindDirection will decrease to Partly cloudy (43.9187268745456), cold (6.47740296152383), and South West (234.527931071855). A conclusion of the predicted result is half variable will decrease slowly. |
| [6] *Output* | 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. |
| [19] *Output* | 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 |
| [21] *Output* | Winds northwest 15 diminishingto 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 |
| [22] *Outptut* | -Light rian showers are likely  -Sunny intervals with rain being possible – less likely than not.  -Sunny with rain being unlikely |

Berdasarkan pada tabel I jumlah konten tentu semakin banyak, namun secara tekstual aplikasi ini tidak sebaik *output* DWP [6] 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 [19]. Pada penelitian Goldberg terdapat pesan mengenai angin dan salju secara terpisah-pisah [21] sedangkan dalam penelitian Gkatzia hanya disampaikan terkait keadaan langit. Hal ini menunjukan 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.

## Hasil Eksperimen

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

Tabel III Hasil Pengukuran Aspek *Readibility*

| **Kode Dataset** | ***Flesch Reading Ease Score* (Grammarly)** | ***Flesch Reading Ease Score* (Datayze)** |
| --- | --- | --- |
| **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** | ***Running Time (s)*** |
| **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 |

# Kesimpulan

Pengembangan sistem *Data-to-Text* untuk data *streaming* dengan menggunakan *Mchine 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 [6] 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 memeroleh nilai rata-rata keseluruhan 35.425 pada aspek *Readibility* 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] E. Reiter, “An Architecture for Data-to-Text Systems,” *Comput. Intell.*, vol. 27, no. 1, pp. 23–40, 2011.

[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] S. Muthukrishnan, “Data Streams: Algorithms and Applications,” *Found. Trends® Theor. Comput. Sci.*, vol. 1, no. 2, pp. 117–236, 2005.

[6] B. Putra, L. S. Riza, and Y. Wihardi, “Pengembangan Sistem Data-to-Text untuk Membangkitkan Berita Cuaca dengan Pendekatan Time-Series dalam R,” 2017.

[7] 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.

[8] 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.

[9] K. Kukich, “Design of a knowledge-based report generator,” *Proc. 21st Annu. Meet. Assoc. Comput. Linguist. -*, p. 145, 1983.

[10] 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.

[11] 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.

[12] J. W. Crowder, J. G. Moore, L. DeRose, and W. J. Franek, “Air Pollution Field Enforcement,” no. September 1999, 1999.

[13] R. Rowlett, “Beaufort Scales (Wind Speed),” 2001. [Online]. Available: https://www.unc.edu/~rowlett/units/scales/beaufort.html. [Accessed: 20-May-2018].

[14] 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].

[15] J. Huby, “Cloud Coverage,” 2010. [Online]. Available: http://www.theweatherprediction.com/habyhints/189/. [Accessed: 20-May-2018].

[16] A. Ramos-soto, A. Bugarin, and S. Barro, “Fuzzy Sets Across the Natural Language Generation Pipeline,” vol. c, pp. 1–16, 2016.

[17] 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.

[18] E. Reiter, “Building Natural-Language Generation Systems,” pp. 91–93, 1996.

[19] 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.

[20] A. Belz, “Probabilistic Generation of Weather Forecast Texts,” *Naacl-Hlt*, no. April, pp. 164–171, 2007.

[21] 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.

[22] D. Gkatzia, O. Lemon, and V. Rieser, “Natural Language Generation enhances human decision-making with uncertain information,” 2016.

[23] D. Bertram, “Likert Scales,” *University of Calagary, Department of Computer Science*, 2013. [Online]. Available: http://poincare.matf.bg.ac.rs/~kristina/topic-dane-likert.pdf.