**PENGEMBANGAN SISTEM *DATA-TO-TEXT* (D2T)UNTUK MEMBANGKITKAN BERITA PADA DATA *STREAMING***

**SKRIPSI**

Diajukan untuk Memenuhi Bagian Dari

Syarat Memperoleh Gelar Sarjana Komputer

Program Studi Ilmu Komputer



Oleh:

Ahmad Zainal Abiddin (1404862)

**PROGRAM STUDI ILMU KOMPUTER**

**DEPARTEMEN PENDIDIKAN ILMU KOMPUTER**

**FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS PENDIDIKAN INDONESIA**

**BANDUNG**

**2018**

**PENGEMBANGAN SISTEM *DATA-TO-TEXT* (D2T)UNTUK MEMBANGKITKAN BERITA PADA DATA *STREAMING***

Oleh

Ahmad Zainal Abiddin

Sebuah skripsi yang diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Komputer pada Fakultas Matematika dan Ilmu Pengetahuan Alam

© Ahmad Zainal Abiddin 2018

Universitas Pendidikan Indonesia

Agustus 2018

Hak Cipta dilindungi undang-undang.

Skripsi ini tidak boleh diperbanyak seluruhya atau sebagian,

dengan dicetak ulang, difoto kopi, atau cara lainnya tanpa ijin dari penulis.

LEMBAR PENGESAHAN

**PENGEMBANGAN SISTEM *DATA-TO-TEXT* (D2T) UNTUK MEMBANGKITKAN BERITA PADA DATA *STREAMING***

Oleh:

Ahmad Zainal Abidin

1404862

DISETUJUI DAN DISAHKAN OLEH:

Pembimbing I

**Lala Septem Riza, M.T., Ph.D.**

NIP.197809262008121001

Pembimbing II

**Dr. H. Enjang Ali Nurdin, M. Kom.**

NIP.196711211991011001

Mengetahui,

Kepala Departemen Pendidikan Ilmu Komputer

**Prof. Dr. H. Munir, M.IT.**

NIP.196603252001121001

**PERNYATAAN**

Dengan ini saya menyatakan bahwa skripsi saya dengan judul “Pengembangan Sistem *Data-to-Text* (D2T) untuk membangkitkan berita pada data *streaming*” ini beserta seluruh isinya adalah benar-benar karya saya sendiri. Saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika ilmu yang berlaku dalam masyarakat keilmuan. Atas pernyataan ini saya siap menanggung resiko/sanksi apabila dikemudian hari ditemukan adanya pelanggaran etika keilmuan atau klaim dari pihak lain terhadap keaslian karya saya ini.

Bandung, Agustus 2018

Ahmad Zainal Abiddin

# KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT karena hanya dengan kehendak, rahmat serta karunia-Nya lah penyusun skripsi yang berjudul “Pengembangan Sistem *Data-to-Text* (D2T) untuk membangkitkan berita pada data *streaming*” ini dapat terselesaikan.

Penyusun skripsi ini ditujukan untuk memenuhi dan melengkapi salah satu syarat untuk mendapatkan gelar sarjana komputer atas jenjang studi S1 pada Program Studi Ilmu Komputer, Departemen Pendidikan Ilmu Komputer, Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam, Universitas Pendidikan Indonesia.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat banyak kekurangan dan keterbatasan yang perlu disempurnakan. Oleh karena itu, penulis sangat mengharapkan saran maupun kritik yang membangun agar tidak terjadi kesalahan yang sama dikemudian hari.

Bandung, Agustus 2018

Penulis,

Ahmad Zainal Abiddin

# UCAPAN TERIMA KASIH

Allhamdulillah, segala puji bagi Allah Subhanahu wa Ta’ala yang telah mengkaruniakan berkah dan kasih sayang-Nya sehingga atas izin-Nya penulis dapat menyelesaikan skripsi ini yang berjudul “Pengembangan Sistem *Data-to-Text* (D2T) untuk membangkitkan berita pada data *streaming*” dengan penuh ketercapaian lainnya.

Penulis menyadari bahwa terealisasikannya skripsi ini tidak lepas dari pernanan, dukungan, dan bantuan dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Keluarga dan saudara-saudara tercinta yang selalu memberikan berbagai dukungan dan selalu mendo’akan penulis dalam menyelesaikan skripsi untuk meraih gelar sarjana komputer ini.
2. Bapak Lala Septem Riza, MT. Ph.D. selaku pembimbing I, atas seluruh jasanya dengan sabar memberikan bimbingan terbaik dan seluruh ilmu bermanfaat yang diberikan kepada penulis sehingga terwujud apa yang diharapkan dari penelitian ini.
3. Bapak Dr. H. Enjang Ali Nurdin, M. Kom. Selaku pembimbing II, atas seluruh jasanya yang telah memberikan bimbingan terbaik dan seluruh ilmu bermanfaat yang diberikan kepada penulis sehingga terwujud apa yang diharapkan dari penelitian ini.
4. Bapak Prof. Dr. Munir, M.IT. sebagai Ketua Departemen Program Studi Ilmu Komputer, yang mendukung serta mendoakan segala yang terbaik untuk penulis.
5. Bapak Dr. H. Wawan Setiawan, M.Kom. selaku pembimbing akademik yang telah mengarahkan penulis selama mengampu pendidikan di Ilmu Komputer UPI.
6. Seluruh staf dan dosen program Ilmu Komputer yang telah membantu, memberi ilmu, mendukung serta mendoakan segala yang terbaik untuk penulis.
7. Seluruh rekan seperjuangan mahasiswa ilmu komputer angkatan 2014, rekan angkatan, departemen dan fakultas yang telah memberikan dukungan, motivasi, inspirasi dan bantuan lainnya sehingga dalam menyelesaikan penelitian ini penulis dapat dengan semangat dapat menyelesaikannya.
8. Kakak dan adik tingkat Ilmu Komputer dan Pendidikan Ilmu Komputer yang tidak bias penulis sebutkan satu persatu. Terima kasih atas doa dan dukungannya.
9. Seluruh elemen yang terlibat didalam proses perkuliahan hingga dapat menyelesaikan segala proses perkuliahan.

**PENGEMBANGAN SISTEM *DATA-TO-TEXT* (D2T) UNTUK MEMBANGKITKAN BERITA PADA DATA *STREAMING***

# 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 permasalahan 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 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 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*

***DATA-TO-TEXT SYSTEM DEVELOPMENT TO GENERATE NEWS ON STREAMING DATA***

# *ABSTRACT*

*The study aims to develop a Data-to-Text system with input streams data in the form of batches, to generate general news. Development of a Data-to-Text system model is done by applying Machine Learning to resolve data streaming problem, with Piecewise Linear Approximation technique using Least Square Method. This system produces a summary text of data, current data description, and predict information. System development is done in R programming language by utilizing several available packages. Experiments were carried out by measuring the level of Readability of news that was raised, Computation Time, and compared with related research. The experiment results show that generated news proven to represent data provided, can be understood by students with college level or above, and Computation Time is good. This system can generate a variety of news based on given data, like climates data, meteorology data, and exchange rate of money.*

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

# DAFTAR ISI

[LEMBAR PENGESAHAN i](#_Toc522784331)

[KATA PENGANTAR i](#_Toc522784332)

[UCAPAN TERIMA KASIH ii](#_Toc522784333)

[ABSTRAK iv](#_Toc522784334)

[*ABSTRACT* v](#_Toc522784335)

[DAFTAR ISI vi](#_Toc522784336)

[DAFTAR GAMBAR viii](#_Toc522784337)

[DAFTAR TABEL xii](#_Toc522784338)

[BAB I PENDAHULUAN 1](#_Toc522784339)

[1.1 Latar Belakang 1](#_Toc522784340)

[1.2 Rumusan Masalah Penelitian 3](#_Toc522784343)

[1.3 Tujuan Penelitian 3](#_Toc522784347)

[1.4 Manfaat Penelitian 3](#_Toc522784348)

[1.5 Batasan Masalah 4](#_Toc522784349)

[1.6 Sistematika Penulisan 4](#_Toc522784350)

[BAB II KAJIAN PUSTAKA 7](#_Toc522784351)

[2.1 Pengertian *Natural Language* 7](#_Toc522784352)

[2.2 Pengertian *Natural Language Processing* 7](#_Toc522784359)

[2.3 Pengertian *Natural Language Generation* 10](#_Toc522784363)

[2.4 Arsitektur *Data-to-Text* 11](#_Toc522784369)

[2.5 Penelitian Terkait *Data-to-Text* 20](#_Toc522784370)

[2.6 *Machine Learning* 22](#_Toc522784377)

[2.6.1. *Supervised Learning* 23](#_Toc522784378)

[2.6.2. *Unsupervised Learning* 27](#_Toc522784379)

[2.6.3. *Semi Supervised Learning* 28](#_Toc522784380)

[2.6.4. *Reinforcement Learning* 28](#_Toc522784381)

[2.7 Data *Time Series* 28](#_Toc522784382)

[2.8 Data *Streaming* 30](#_Toc522784383)

[2.9 Pemrograman R 32](#_Toc522784384)

[2.9.1. Model Data dalam R 34](#_Toc522784385)

[2.9.2. Contoh Kode Program Bahasa R 34](#_Toc522784386)

[2.9.3. *Package* dalam Bahasa R 37](#_Toc522784387)

[BAB III METODE PENELITIAN 38](#_Toc522784388)

[3.1 Desain Penelitian 38](#_Toc522784390)

[3.2 Metode Penelitian 40](#_Toc522784393)

[3.3 Alat dan Bahan Penelitian 40](#_Toc522784394)

[3.4 Data Penelitian 41](#_Toc522784397)

[3.4.1. Data Masukan 41](#_Toc522784398)

[3.4.2. Data Keluaran 41](#_Toc522784399)

[BAB IV HASIL PENELTIAN DAN PEMBAHASAN 43](#_Toc522784400)

[4.1. Pengumpulan Data 43](#_Toc522784402)

[4.1.1. Data Historis Rupiah 43](#_Toc522784403)

[4.1.2. *Solar Radiation and Meteorological Data* 44](#_Toc522784404)

[4.1.3. Data Klimatologi dan Kualitas Udara 47](#_Toc522784405)

[4.2 Pengembangan Model *Data-to-text Streaming* dengan *Machine Learning* 49](#_Toc522784406)

[4.2.1 Model Proses *Realtime Reader Dataset* 51](#_Toc522784407)

[4.2.2 Model Proses *Signal Analysis* 52](#_Toc522784408)

[4.2.3 Model Proses *Data Interpretation* 60](#_Toc522784409)

[4.2.4 Model Proses *Document Planning* 68](#_Toc522784410)

[4.2.5 Model Proses *Microplanning and Realisation* 75](#_Toc522784411)

[4.2.6 Model Antarmuka Perangkat Lunak 78](#_Toc522784412)

[4.3 Implementasi Sistem *Data-to-Text* 79](#_Toc522784413)

[4.3.1 Implementasi Proses *Realtime Reader* 79](#_Toc522784414)

[4.3.2 Implementasi Proses *Signal Analysis* 82](#_Toc522784415)

[4.3.3 Implementasi Proses *Data Interpretation* 92](#_Toc522784416)

[4.3.4 Implementasi Proses *Document Planning* 99](#_Toc522784417)

[4.3.5 Implementasi Proses *Microplanning and Realisation* 103](#_Toc522784418)

[4.4 Hasil Implementasi 108](#_Toc522784424)

[4.4.1 *Output* Sistem 108](#_Toc522784425)

[4.4.2 Perbandingan *Output* Sistem dengan penelitian terkait 110](#_Toc522784426)

[4.5 Panduan Penggunaan Aplikasi 112](#_Toc522784427)

[4.6 Rancangan Eksperimen 113](#_Toc522784428)

[4.7 Analisis dan Hasil Eksperimen 116](#_Toc522784429)

[4.7.1. Eksperimen dengan Data Kurs Beli Rupiah 116](#_Toc522784430)

[4.7.2. Eksperimen dengan Data Radiasi dan Meteorologi 128](#_Toc522784432)

[4.7.3. Eksperimen dengan Data Klimatologi dan Kualitas Udara 140](#_Toc522784433)

[4.7.4 Analisis Perbandingan dengan Penelitian Sebelumnya 149](#_Toc522784434)

[BAB V KESIMPULAN DAN SARAN 152](#_Toc522784435)

[5.1 Kesimpulan 152](#_Toc522784436)

[5.2 Saran 152](#_Toc522784437)

[DAFTAR PUSTAKA 154](#_Toc522784438)

# DAFTAR GAMBAR

[Gambar 2.1 Arsitektur sistem Data-to-Text (Reiter, 2011) 12](#_Toc521256016)

[Gambar 2.2 Arsitektur DWP (Putra et al., 2017) 13](#_Toc521256017)

[Gambar 2.3 Contoh implementasi Signal Analysis (Putra et al., 2017) 14](#_Toc521256018)

[Gambar 2.4 Contoh implementasi data interpretation Rainfall DWP (Putra et al., 2017) 15](#_Toc521256019)

[Gambar 2.5 Contoh Content Determination Significant Event Message DWP (Putra et al., 2017) 16](#_Toc521256020)

[Gambar 2.6 Contoh Target Text DWP (Putra et al., 2017) 17](#_Toc521256021)

[Gambar 2.7 Contoh skema dalam bentuk tree berdasarkat Target Text DWP (Putra et al, 2017) 17](#_Toc521256022)

[Gambar 2.8 Contoh implementasi Lexicalisation tren DWP (Putra et al., 2017) 18](#_Toc521256023)

[Gambar 2.9 Contoh Simple Conjunction Referring to Contrast Value (Putra et al., 2017) 19](#_Toc521256024)

[Gambar 2.10 Contoh Referring Expression Generation (Putra et al., 2017) 19](#_Toc521256025)

[Gambar 2.11 Contoh implementasi Structure Realisation DWP (Putra et al., 2017) 20](#_Toc521256026)

[Gambar 2.12 Contoh *Supervised Learning* pada pengenalan koin 24](#_Toc521256027)

[Gambar 2.13 Kombinasi dua garis linier (Palpanas et al., 2004). 24](#_Toc521256028)

[Gambar 2.14 Contoh *Unsupervised Learning* dalam pengenalan koin. 27](#_Toc521256029)

[Gambar 2.15 Logo bahasa pemrograman R. 32](#_Toc521256030)

[Gambar 2.16 Antarmuka R Graphical User Interface (RGui). 33](#_Toc521256031)

[Gambar 2.17 Model data dalam pemrograman R (Budiharto, 2013). 34](#_Toc521256032)

[Gambar 2.18 Operator *concantenate* dalam R. 34](#_Toc521256033)

[Gambar 2.19 Menampilkan data pertama hingga ke-dua dalam R. 34](#_Toc521256034)

[Gambar 2.20 Penggunaan fungsi *sum* dalam R. 35](#_Toc521256035)

[Gambar 2.21 Penggunaan fungsi *concantenate* untuk *string* dalam R 35](#_Toc521256036)

[Gambar 2.22Pembuatan matriks dalam R*.* 35](#_Toc521256037)

[Gambar 2.23Contoh *Visualisasi* grafis dalam R. 36](#_Toc521256038)

[Gambar 2.24Contohperulangan dalam R*.* 36](#_Toc521256039)

[Gambar 2.25Contohimplementasi *decission* dalam R*.* 37](#_Toc521256040)

[Gambar 2.26Contohfungsi dalam R. 37](#_Toc521256041)

[Gambar 3.1 Desain Penelitian 38](#_Toc521256042)

[Gambar 3.2 Model Linear *Sequential Model* (Pressman, 2001) 40](#_Toc521256043)

[Gambar 4.1 Model *Data-to-text* pada *Data Streaming* dengan pendekatan *Machine Learning* 50](#_Toc521256044)

[Gambar 4.2Model proses *Signal Analysis* prediksi dengan PLA dan *Least Square* 54](file:///D:\data\kuliah\Skripsi\perjalanan\2018-07-30-v_Prasidang\draft_template_anita.docx#_Toc521256045)

[Gambar 4.3 Model proses *Signal Analysis statistical summary* 56](#_Toc521256046)

[Gambar 4.4 *Fuzzy membership function Temperature* (Ramos-soto et al., 2016) 65](#_Toc521256047)

[Gambar 4.5 *Fuzzy membership function Rainfall* (Ramos-soto et al., 2016) 66](#_Toc521256048)

[Gambar 4.6 *Linguistic variable for trend description* (Castillo-Ortega et al., 2014) 66](#_Toc521256049)

[Gambar 4.7 *General Fuzzy membership function* 67](#_Toc521256050)

[Gambar 4.8 *Routine Message* untuk ringkasan data 68](#_Toc521256051)

[Gambar 4.9 *Significant Event Message* untuk ringkasan data 69](#_Toc521256052)

[Gambar 4.10 *Target Text* untuk ringkasan data 70](#_Toc521256053)

[Gambar 4.11 *Structure text* ringkasan 71](#_Toc521256054)

[Gambar 4.12 pohon struktur teks ringkasan 72](#_Toc521256055)

[Gambar 4.13 *Routine Message* untuk data terakhir 72](#_Toc521256056)

[Gambar 4.14 *Significance Event Message* untuk data terakhir 73](#_Toc521256057)

[Gambar 4.15 *Target Text* data terakhir 73](#_Toc521256058)

[Gambar 4.16 *Structure text* data terakhir 73](#_Toc521256059)

[Gambar 4.17 Pohon struktur teks data terakhir 74](#_Toc521256060)

[Gambar 4.18 *Routine Message* prediksi 74](#_Toc521256061)

[Gambar 4.19 *Target Text* untuk prediksi 75](#_Toc521256062)

[Gambar 4.20 *Structure text* prediksi 75](#_Toc521256063)

[Gambar 4.21 pohon struktur prediksi 75](#_Toc521256064)

[Gambar 4.22 *Template* *Lexicalisation* *Comparasion Predict* 76](#_Toc521256065)

[Gambar 4.23Struktur *file* applikasi D2T 78](#_Toc521256066)

[Gambar 4.24Antarmuka sistem *General News Generator* (GNG) 79](#_Toc521256067)

[Gambar 4.25 *Realtime animation for check file* 80](#_Toc521256068)

[Gambar 4.26 *Execute* D2T\_Main.R 81](#_Toc521256069)

[Gambar 4.27 *Read and Remove Dataset* *in* R 81](#_Toc521256070)

[Gambar 4.28 *Write Result* JSON and csv *in* R 81](#_Toc521256071)

[Gambar 4.29 *Get* JSON *in* AJAX 82](#_Toc521256072)

[Gambar 4.30 *PLAStream function* 83](#_Toc521256073)

[Gambar 4.31 *Statistical Summary function* 84](#_Toc521256074)

[Gambar 4.32 *Trend Analysis function* 85](#_Toc521256075)

[Gambar 4.33 *TypeEventResume function* 87](#_Toc521256076)

[Gambar 4.34Merubah *Statistical Summary* karena adanya *Statistical Event* 88](#_Toc521256077)

[Gambar 4.35 *ResumeEventExtreme function* 90](#_Toc521256078)

[Gambar 4.36 *ResumeEventRepeat function* 91](#_Toc521256079)

[Gambar 4.37 *Corpus for Data Interpretation* 93](#_Toc521256080)

[Gambar 4.38 *DataInterpreterAdjective function* 94](#_Toc521256081)

[Gambar 4.39 *MembershipFuzzy function* 95](#_Toc521256082)

[Gambar 4.40 *MembershipCrisp function* 95](#_Toc521256083)

[Gambar 4.41 *AirQualityCalculation function* (Putra et al, 2017) 98](#_Toc521256084)

[Gambar 4.42 *Content Determination ResumeEvent function* 99](#_Toc521256085)

[Gambar 4.43 Impelementasi *Event Determination Summary Event* 100](#_Toc521256086)

[Gambar 4.44 Implementasi *Event Determination Repeated Event* 101](#_Toc521256087)

[Gambar 4.45 Implementasi *Event Determination Extreme Event* 102](#_Toc521256088)

[Gambar 4.46 Implementasi *Event Determination Current Text* 103](#_Toc521256089)

[Gambar 4.47 Implementasi *Lexicalisation Extreme Event* 104](#_Toc521256090)

[Gambar 4.48 Implementasi *Aggregation* pada *Trend Description* 105](#_Toc521256091)

[Gambar 4.49 Implementasi *Referring Expression Generation* untuk *Intro* 105](#_Toc521256092)

[Gambar 4.50 Implementasi *Referring Expression Generation* untuk *Time Descritption* 106](#_Toc521256093)

[Gambar 4.51 Implementasi *Structure Realisation* 108](#_Toc521256094)

[Gambar 4.52 *News* *from first batch* 109](#_Toc521256095)

[Gambar 4.53 *News* *from second batch* 109](#_Toc521256096)

[Gambar 4.54 Eksperimen pertama dengan data kurs beli Maret 2018 116](#_Toc521256097)

[Gambar 4.55 Eksperimen ke-dua dengan data kurs beli April 2018 117](#_Toc521256098)

[Gambar 4.56 Eksperimen ke-tiga dengan data kurs beli Mei 2018 118](#_Toc521256099)

[Gambar 4.57 Plot perbandingan *Representative Text* tentang Australian Dollar (Buy) 121](#_Toc521256100)

[Gambar 4.58 Plot perbandingan *Representative Text* tentang Lao Kip (Buy) 122](#_Toc521256101)

[Gambar 4.59 Plot perbandingan *Representative Text* tentang Yen Jepang (Buy) 123](#_Toc521256102)

[Gambar 4.60 Plot perbandingan *Representative Text* tentang Baht Thailand (Buy) 124](#_Toc521256103)

[Gambar 4.61 Plot perbandingan *Representative Text* tentang Peso Filipina (Buy) 125](#_Toc521256104)

[Gambar 4.62 Plot perbandingan *Representative Text* tentang Ringgit Malaysia (Buy) 126](#_Toc521256105)

[Gambar 4.63 Eksperimen ke-empat dengan data radiasi sinar matahari dan meteorologi Juli 2016 128](#_Toc521256106)

[Gambar 4.64 Eksperimen ke-lima dengan data radiasi sinar matahari dan meteorologi Agustus 2016 129](#_Toc521256107)

[Gambar 4.65 Eksperimen ke-enam dengan data radiasi sinar matahari dan meteorologi September 2016 130](#_Toc521256108)

[Gambar 4.66 Eksperimen ke-tujuh dengan data radiasi sinar matahari dan meteorologi Oktober 2016 130](#_Toc521256109)

[Gambar 4.67 Eksperimen ke-delapan dengan data radiasi sinar matahari dan meteorologi November 2016 131](#_Toc521256110)

[Gambar 4.68 Eksperimen ke-sembilan dengan data radiasi sinar matahari dan meteorologi Desember 2016 131](#_Toc521256111)

[Gambar 4.69 Plot perbandingan *Representative Text* tentang Diffuse Horizontal Irradiance (DIF) 134](#_Toc521256112)

[Gambar 4.70 Plot perbandingan *Representative Text* tentang Air Pressure (AP) 135](#_Toc521256113)

[Gambar 4.71 Plot perbandingan *Representative Text* tentang Relative Humidity (RH) 136](#_Toc521256114)

[Gambar 4.72 Plot perbandingan *Representative Text* tentang Precipitable Water (PWAT) 137](#_Toc521256115)

[Gambar 4.73 Plot perbandingan *Representative Text* tentang Temperature 138](#_Toc521256116)

[Gambar 4.74 Eksperimen ke-sepuluh dengan data klimatologi dan kualitas udara 2014-2015 140](#_Toc521256117)

[Gambar 4.75 Eksperimen ke-sebelas dengan data klimatologi dan kualitas udara 2015-2016 141](#_Toc521256118)

[Gambar 4.76 Eksperimen ke-duabelas dengan data klimatologi dan kualitas udara 2016-2017 142](#_Toc521256119)

[Gambar 4.77 Plot perbandingan *Representative Text* tentang Temperature 145](#_Toc521256120)

[Gambar 4.78 Plot perbandingan *Representative Text* tentang Cloud Coverage (CloudCoverage) 146](#_Toc521256121)

[Gambar 4.79 Plot perbandingan *Representative Text* tentang Wind Speed (WindSpeed) 147](#_Toc521256122)

[Gambar 4.80 Plot perbandingan *Representative Text* tentang Rainfall 148](#_Toc521256123)

# DAFTAR TABEL

[Tabel 1 Penelitian terkait D2T dan NLG 20](#_Toc522703384)

[Tabel 2.2 Contoh data penjualan dengan periode bulanan 26](#_Toc522703385)

[Tabel 2.3 Perhitungan metode *least square* 26](#_Toc522703386)

[Tabel 4.1 Kutipan data Kurs Beli Rupiah 44](#_Toc522703387)

[Tabel 4.2 Kutipan data radiasi sinar matahari dan meteorologi interval waktu harian 46](#_Toc522703388)

[Tabel 4.3 Kutipan data klimatologi 48](#_Toc522703389)

[Tabel 4.4 Contoh kasus dataset *batch* pertama 51](#_Toc522703390)

[Tabel 4.5 Contoh kasus dataset *batch* ke-dua 51](#_Toc522703391)

[Tabel 4.6 Hasil *Extreme Event* contoh kasus *batch* pertama 58](#_Toc522703392)

[Tabel 4.7 Hasil *Extreme Event* contoh kasus *batch* ke-dua 58](#_Toc522703393)

[Tabel 4.8 Hasil *Repeated Event* contoh kasus *batch* pertama 59](#_Toc522703394)

[Tabel 4.9 Hasil *Repeated Event* contoh kasus *batch* ke-dua 60](#_Toc522703395)

[Tabel 4.10 *Crisp membership function Wind Speed* (Rowlet, 2001) 60](#_Toc522703396)

[Tabel 4.11 *C*risp *membership function Wind Direction* (Zandlo et al., 2001) 61](#_Toc522703397)

[Tabel 4.12 *C*risp *membership function Cloud Coverage* (Huby, 2010) 62](#_Toc522703398)

[Tabel 4.13 *Crisp membership function* PSI dan *breakpoint* sub-indeks PSI (Crowder, Moore, DeRose, & Franek, 1999) 63](#_Toc522703399)

[Tabel 4.14 Keterangan rumus PSI 64](#_Toc522703400)

[Tabel 4.15 Hasil *Data Interpretation* contoh kasus *batch* pertama 67](#_Toc522703401)

[Tabel 4.16 Hasil *Data Interpretation* contoh kasus *batch* ke-dua 67](#_Toc522703402)

[Tabel 4.17 Hasil *Content Determination Summary* contoh kasus *batch* pertama 69](#_Toc522703403)

[Tabel 4.18 Hasil *Content Determination Summary* contoh kasus *batch* ke-dua 69](#_Toc522703404)

[Tabel 4.19 Hasil *Content Determination Current Data* contoh kasus *batch* pertama 73](#_Toc522703405)

[Tabel 4.20 Hasil *Content Determination Current Data* contoh kasus *batch* ke-dua 73](#_Toc522703406)

[Tabel 4.21 Hasil *Content Determination Predict* contoh kasus *batch* pertama 74](#_Toc522703407)

[Tabel 4.22 Hasil *Content Determination Predict* contoh kasus *batch* ke-dua 75](#_Toc522703408)

[Tabel 4.23 Hasil akhir contoh kasus *batch* pertama 77](#_Toc522703409)

[Tabel 4.24 Hasil akhir contoh kasus *batch* ke-dua 77](#_Toc522703410)

[Tabel 4.25 Hasil prediksi data radiasi sinar matahari dan meteorologi 84](#_Toc522703411)

[Tabel 4.26 Hasil ringkasan data radiasi sinar matahari dan meteorologi 86](#_Toc522703412)

[Tabel 4.27 Hasil fungsi *TypeEventResume* 87](#_Toc522703413)

[Tabel 4.28 Hasil ringkasan data radiasi sinar matahari dan meteorologi *batch* ke-n 89](#_Toc522703414)

[Tabel 4.29 Hasil extreme event analysis 91](#_Toc522703415)

[Tabel 4.30 Hasil repeat event analysis 92](#_Toc522703416)

[Tabel 4.31 Daftar Parameter dalam *file* csv bernama ParameterList 92](#_Toc522703417)

[Tabel 4.32 *Crisp membership value Wind Speed* 96](#_Toc522703418)

[Tabel 4.33 *Crisp membership value Wind Direction* 96](#_Toc522703419)

[Tabel 4.34 *Crisp membership value Cloud Coverage* 97](#_Toc522703420)

[Tabel 4.35 *Crisp membership value Air Quality* 98](#_Toc522703421)

[Tabel 4.36 *Fuzzy membership value Temperature* 98](#_Toc522703422)

[Tabel 4.37 *Fuzzy membership value Rainfall* 99](#_Toc522703423)

[Tabel 4.38 Representasi Waktu untuk *Referring Expression Generation* 107](#_Toc522703424)

[Tabel 4.39 *Output model and statistical summary* 110](#_Toc522703425)

[Tabel 4.40Perbandingan *output* dengan penelitian terkait 110](#_Toc522703426)

[Tabel 4.41  *Flesch Reading Ease* 114](#_Toc522703427)

[Tabel 4.42Dataset Eksperimen 115](#_Toc522703428)

[Tabel 4.43 Hasil evaluasi *Readability* dengan data kurs beli rupiah 118](#_Toc522703429)

[Tabel 4.44 Hasil evaluasi *Computation Time* dengan data kurs beli rupiah 119](#_Toc522703430)

[Tabel 4.45 Hasil perbandingan *Statistical Summary* data kurs beli rupiah 119](#_Toc522703431)

[Tabel 4.46 Analisis *trend* dan *event* dengan data kurs beli rupiah 127](#_Toc522703432)

[Tabel 4.47 Analisis prediksi dengan data kurs beli rupiah 128](#_Toc522703433)

[Tabel 4.48 Hasil evaluasi *Readability* dengan data radiasi sinar matahari dan meteorologi 132](#_Toc522703434)

[Tabel 4.49 Hasil evaluasi *Computation Time* dengan data radiasi sinar matahari dan meteorologi 132](#_Toc522703435)

[Tabel 4.50 Hasil perbandingan *Statistical Summary* data radiasi sinar matahari dan meteorologi 133](#_Toc522703436)

[Tabel 4.51 Analisis *trend* dan *event* dengan data radiasi sinar matahari dan meteorologi 139](#_Toc522703437)

[Tabel 4.52 Analisis prediksi dengan data radiasi sinar matahari dan meteorologi 139](#_Toc522703438)

[Tabel 4.53 Hasil evaluasi *Readability* dengan data klimatologi dan kualitas udara 143](#_Toc522703439)

[Tabel 4.54 Hasil evaluasi *Computation Time* dengan data klimatologi dan kualitas udara 143](#_Toc522703440)

[Tabel 4.55 Hasil perbandingan *Statistical Summary* data klimatologi dan kualitas udara 144](#_Toc522703441)

[Tabel 4.56 Analisis *trend* dan *event* dengan data klimatologi dan kualitas udara 149](#_Toc522703442)

[Tabel 4.57 Analisis prediksi dengan data klimatologi dan kualitas udara 149](#_Toc522703443)

[Tabel 4.58 Perbandingan dengan penelitian sebelumnya 151](#_Toc522703444)

# BAB I PENDAHULUAN

## 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 manusiayaitu *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 kemapuan 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 mengefesiensikan *Development Time*, penulis menggunakan memanfaatkan beberapa *packages* yang tersedia dalam R*.*



## 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 mengggunakan 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?

## 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.

## 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.

## 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.

## 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 teridiri 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 penelitan 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*, analisi 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.

# BAB II KAJIAN PUSTAKA

## Pengertian *Natural Language*

*Natural Language* atau bahasa alami adalah bahasa yang digunakan dalam kehidupan sehari-hari dengan tujuan mempermudah komunikasi sesama manusia. *Natural Language* merupakan bahasa yang digunakan manusia dalam berkomunikasi seperti bahasa Indonesia, bahasa Inggris, dan bahasa lainnya sesuai letak geografisnya (Putra et al., 2017). Banyak peneliti yang berlomba-lomba untuk menciptakan teknologi yang mampu menghasilkan bahasa alami, agar komputer dengan manusia dapat saling berkomunikasi, salah satu contoh teknologi tersebut adalah *Natural Language Processing* (NLP). Ada tiga aspek utama pada *Natural Language*:

1. *Syntax*: menjelaskan bentuk dari bahasa. Syntax biasa dispesifikasikan oleh sebuah *grammar*. *Natural language* jauh lebih daripada formal *language* yang digunakan untuk logika kecerdasan buatan dan program komputer.
2. *Semantics*: menjelaskan arti dari kalimat dalam satu bahasa. Meskipun teori semantik secara umum sudah ada, ketika membangun sistem *natural language understanding* untuk aplikasi tertentu, akan digunakan representasi yang paling sederhana.
3. *Pragmatics*: menjelaskan bagaimana pernyataan yang ada berhubungan dengan dunia. Untuk memahami bahasa, agen harus mempertimbangan lebih dari hanya sekedar kalimat. Agen harus melihat lebih ke dalam konteks kalimat, keadaan dunia, tujuan dari speaker dan listener, konvensi khusus, dan sejenisnya.

## Pengertian *Natural Language Processing*

*Natural Language Processing* (NLP) adalah teknik-teknik komputasi yang didorong secara teoritis untuk menganalisa dan merepresentasikan bahasa alami pada tingkat analisis linguistik untuk mencapai pemrosesan bahasa seperti manusia (Liddy, 2001). NLP merupakan salah satu bidang ilmu yang termasuk kedalam *Artificial Intelligence* (AI) yang hanya berfokus pada pemrosesan bahasa alami. Tujuan dari NLP adalah mempermudah komunikasi antara komputer dengan manusia ataupun manusia dengan komputer. Ada banyak aplikasi yang telah menerapkan NLP, salah satu diantaranya adalah Chatbot. Aplikasi Chatbot dapat membuat user seolah-olah mampu berkomuniasi langsung dengan komputer dalam bahasa manusia, contoh aplikasi Chatboth adalah sim-simi. Contoh lainnya adalah aplikasi *Stemming* atau *Lemmatization*, aplikasi ini dapat melakukan pemotongan kata dalam bahasa tertentu menjadi bentuk dasar pengenalan fungsi setiap kata dalam kalimat. Contoh lainnya adalah aplikasi *Summarization* yang dapat melakukan peringkasan terhadap sebuah berita, serta aplikasi Translation Tools yang mampu menerjemahkan bahasa. NLP bekerja ketika komputer membaca bahasa alami, yang kemudian dilakukan serangkaian proses sehingga menjadi data terstruktur.

*Natural Language Processing* (NLP) mempelajari tentang bagaimana mengubah bahasa alami (dokumen atau suara) kedalam bentuk yang lebih terstruktur (simbol atau numerik) agar lebih mudah untuk dimanipulasi atau diproses oleh program komputer. Lebih jauh lagi NLP bertujuan untuk memahami makna informasi yang benar dan memproses bahasa alami yang mempertimbangkan batasan semantik, gramatikal dan konteks.

Beberapa istilah dalam penerapan NLP (Liddy, 2001):

1. *Part-of-speech tagging*:Sangatlah sulit untuk menandai istilah-istilah dalam suatu teks yang terkait dengan bagian tertentu dari suatu naskah (misalnya kata benda, kata kerja, kata sifat atau kata keterangan), karena bagian dari naskah tidak hanya bergantung pada definisi istilah tetapi juga pada konteks dimana teks digunakan.
2. *Text segmentation*: Beberapa bahasa tulisan, seperti bahasa mandarin, jepang, dan thai, tidak memiliki batasan kata. Dalam contoh ini, tugas *text-parsing* memerlukan idetifikasi terhadap batasan kata, yang seringkali merupakan tugas yang sangat sulit. Tantangan serupa dalam segmentasi naskah muncul ketika menganalisa bahasa verbal, karena suara menyajikan rangkaian huruf dan kata yang bercampur satu sama lain.
3. *Word sense disambiguation*:Banyak kata yang memiliki lebih dari satu arti. Memilih arti yang paling masuk akal hanya bisa dicapai dengan mempertimbangkan konteks di mana kata digunakan.
4. *Syntactic ambiguity*:Tata bahasa dalam bahasa alami seringkali ambigu yang artinya, ada berbagai struktur kalimat yang memungkinakan yang perlu dipertimbangkan. Memilih struktur yang paling tepat biasanya memerlukan paduan informasi kontekstual dan semantik.
5. *Imperfect or irregular input*:Aksen asing atau lokal dan berbagai hambatan vokal dalam pidato dan kesalahan ketik dan tata bahasa dalam teks-teks menyebabkan pengolahan bahasa bahkan lebih sulit.
6. *Speech acts*:Suatu kalimat seringkali dianggap sebagai suatu aksi oleh si pembicara. Struktur kalimatnya sendiri mungkin tidak berisi cukup informasi untuk mendefinisikan tindakan ini.

Aplikasi yang paling sering digunakan menggunakan NLP adalah sebagai berikut (Liddy, 2001):

1. *Information Retrieval*: Ilmu untuk melakukan pencarian terhadap berbagai dokumen yang relevan, menemukan informasi tertentu didalamnya, dan menghasilkan metadata untuk isinya.
2. *Information Extraction* (IE): Sejenis ‘*information retrieval’* yang tujuannya adalah untuk mengekstrak secara otomatis informasi terstruktur, seperti data yang sudah terdefinisi dengan baik secara semantik dan secara kontekstual yang sudah terkelompok dari domain tertentu, dengan menggunakan berbagai dokumen tak-terstruktur yang bisa terbaca oleh mesin.
3. *Question-Answering*: Pekerjaan menjawab secara otomatis suatu pertanyaan yang diajukan dalam bahasa alami; yaitu, menghasilkan jawaban bahasa manusia ketika diberi pertanyaan bahasa manusia. Untuk mendapatkan jawaban terhadap pertanyaan, program computer bisa menggunakan baik database pra-terstruktur atau kumpulan dokumen bahasa alami (suatu ‘*text corpus*’ seperti *world wide web*).
4. *Summarization*: Tingkat NLP yang lebih lanjut, menghasilkan sebuah rangkuman singkat dari sebuah dokumen (dengan teks dalam jumlah besar) dengan waktu yang cepat.
5. *Machine Translation*: Penerjemahan otomatis dari satu bahasa manusia ke bahasa manusia yang lain.
6. *Dialogue Systems*: Berkomunikasi dengan komputer layaknya dengan manusia, yang memungkinkan dikembangkan lebih lanjut mengingat potensi yang dimiliki sistem ini sangat besar.
7. *Speech Recognition*: Mengubah kata-kata verbal menjadi *input* yang bisa terbaca oleh mesin. Dengan adanya sound clip dari orang yang sedang berbicara, sistem menghasilkan dikte dari teks.
8. *Natural Language Generation*: Sistem mengubah informasi dari database komputer (simbolik atau numerik) menjadi bahasa manusia yang bisa dibaca.
9. *Natural Language Understanding*: Sistem mengekstrak informasi yang mewakili makna dari suatu sumber teks (dokumen atau rangkuman).
10. *Speech Synthesis*: Perangkat yang mampu berbicara atau membaca teks.



## Pengertian *Natural Language Generation*

*Natural Language Generation* (NLG) adalah sistem yang mampu menghasilkan informasi dalam bentuk text (*linguistic*) dengan berdasarkan data *non-linguistic* (data *raw* atau mentah yang terukur atau berasal dari serangkaian kejadian)agar mudah dipahami oleh manusia(Bateman & Zock, 2012). NLG merupakan salah satu tipe dari NLP dan merupakan studi penentuan pilihan untuk menyampaikan informasi (Reiter, 2010), sebagai contohnya terdapat pilihan keluaran kalimat:

1. Saya membawa apel dari rumah ke sekolah. Saya memakannya saat jam istirahat.
2. Saya membawa apel dari rumah ke sekolah. Saya memakan apel tersebut saat jam istirahat.

Pilihan a dan b tentu memiliki makna yang sama, hanya saja panjang kalimat a dan b berbeda, a akan lebih cepat untuk dibaca dan dipahami oleh manusia, namun a akan sulit dipahami kembali oleh mesin, karena terdapat ‘nya’ yang menunjuk sesuatu, dan hal tersebut perlu ditelusuri kembali oleh mesin, sedangkan pilihan b memang lebih lama untuk dibaca oleh manusia bila dibandingkan dengan a, namun tetap mudah dipahami oleh manusia dan mudah pula dipahami oleh mesin, dengan demikian NLG akan memilih b. Berbeda dengan NLP, NLG bekerja dengan *input* data terstruktur yang kemudian dilakukan serangkaian proses untuk membangkitkan bahasa alami.

Dalam membangun sistem NLG yang dapat dikatakan sukses memerlukan pengetahuan teknik (menguraikan, merepresentasikan, dan mengatur proses informasi dari *input*) serta pengetahuan tentang keadaan dan kendala dari pengguna (prosesor informasi yang diterima) (Bateman & Zock, 2012).

*Natural Language Generation* (NLG) mempelajari tentang bagaimana menyampaikan informasi dari data terstruktur (simbolik atau numerik) kedalam bahasa alami agar lebih mudah untuk dipahami oleh manusia secara umum. NLG terurai menjadi tiga jenis definisi, yaitu:

1. NLG sebagai *mapping problem.*
2. NLG sebagai *problem of choice.*
3. NLG sebagai *planning problem*

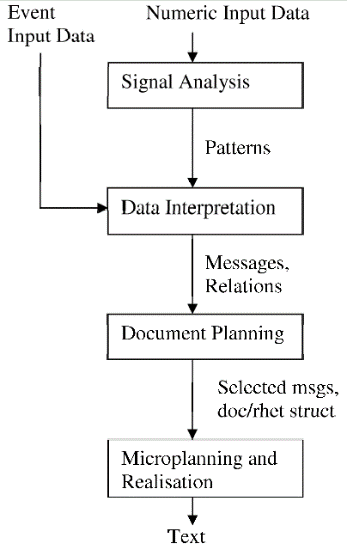
Dimana setiap definisi mewakili setiap layer utama dari arsitekturnya. Beberapa istilah dalam penerapan NLG:

1. *Macroplanning*: Memilih konten dan menyusun rencana teks dari sumber atau data mentah sesuai dengan tujuan pengaplikasian.
2. *Microplanning*: Mendeskripsikan kejadian dengan mengelompokkan materi yang terkait erat sehingga dapat membangun informasi yang lebih terintegrasi dan ringkas.
3. *Surface realization*: Kontruksi gramatikal yang dipilih (pemilihan sintaksis), dengan memerhatikan batasan tata bahasa, kompleksitas kalimat, sehingga menjadi bentuk akhir dari kata-kata (morfologi).

## Arsitektur *Data-to-Text*

*Data-to-text* (D2T) adalah sistem *Natural Language Generation* (NLG) yang mampu menghasilkan teks dari *input* data *non-linguistic*, seperti data sensor dan serangkaian kejadian (Reiter, 2011). Teks yang dihasilkan mencakup makna dari data *input* sehingga tidak mengurangi tujuan data tersebut dibuat, bahkan teks tersebut memudahkan manusia untuk memahaminya, karena disesuaikan dengan bahasa alami. D2T tidak hanya menghasilkan teks yang mampu dipahami dengan mudah oleh manusia, namun harus dapat menyampaikan *knowledge* kepada pembaca.

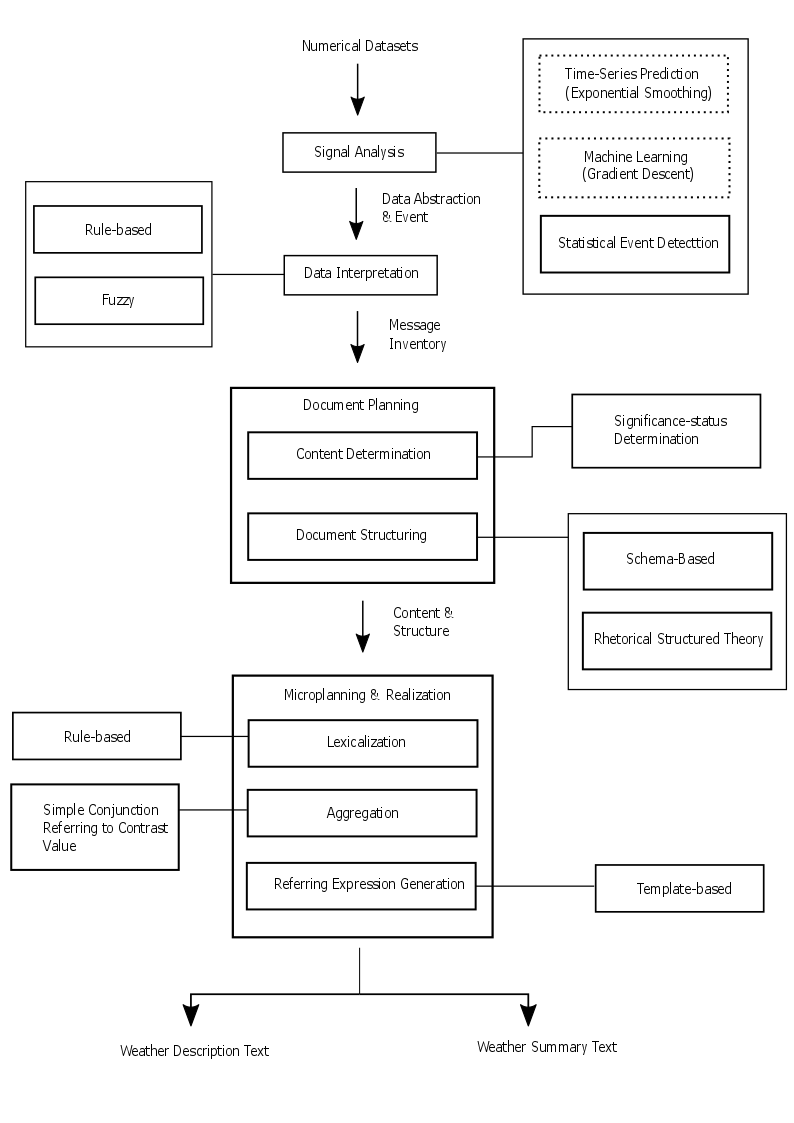
D2T telah diaplikasikan pada berbagai bidang, seperti *Data-to-text* *Weather Prediction* (DWP) yang mampu menghasilkan ringkasan berita klimatologis dan cuaca selama satu bulan serta memberikan informasi prediksi untuk satu hari berikutnya (Putra, Riza, & Wihardi, 2017),sistem *BabyTalk* menghasilkan 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). Beberapa contoh tersebut membuktikan bahwa penerapan D2T memberi manfaat dalam berbagai bidang, seperti klimatologis, kesehatan, ekonomi dan lainnya*.*



Gambar 2.1 Arsitektur sistem Data-to-Text (Reiter, 2011)

Gambar 2.1 merupakan arsitektur utama D2T (Reiter, 2011), yang terdiri dari empat tahapan utama, yaitu *Signal Analysis*, *Data Interpretation*, *Document Planning*, dan *Microplanning and Realisation*. Dalam penelitian DWP, dilakukan pengembangan model sehingga menjadi seperti pada digambar 2.2, dimana terlihat bahwa *output* yang dihasilkan dibagi menjadi dua kategori yaitu *Weather Description Text* dan *Weather Summary Text* yang artinya model tersebut hanya berlaku pada data cuaca.

Pada gambar 2.2 dijelaskan bahwa terdapat data masukan berupa dataset numerik, yang kemudian diproses pada tahap *Signal Analysis*, didalam proses *Signal Analysis* terdapat penerapan pendekatan *Time Series* dengan menggunakan *Exponetial Smoothing*, serta penerapan *Machine Learning* dengan menggunakan *Gradient Descent*, dan juga terdapat proses *Statistical Event Detection*. Proses *Signal Analysis* menghasilkan *Data Abstraction and Event* yang menjadi *input* bagi proses *Data Interpretation*. Dalam proses *Data Interpretation* menggunakan dua cara, yaitu *Rule-based* dan *Fuzzy* yang menghasilkan *Message Inventory*. Pada proses *Document Planning* terdapat dua proses utama yaitu *Content Determination* dan *Document Structuring*. *Content Determination* dilakukan dengan cara mengelompokan *Message* yang ada berdasarkan *Significance Status Determination*. Struktur dokumen dibangun berdasarkan *Schema-based* dan *Rhetorical Structured Theory*. Pada proses *Document Planning* dihasilkan *Content* dan *Structure* yang kemudian diproses pada tahap terakhir yaitu *Microplanning* dan *Realization* yang pada akhirnya menghasilkan text berbentuk paragraf terdiri dari *Weather Description Text* dan *Weather Summary Text*.



Gambar 2.2 Arsitektur DWP (Putra et al., 2017)

1. *Signal Analysi*s

*Signal Analysis* merupakan hal pertama yang dilakukan dalam membangun sistem D2T seperti yang telah diimpementasikan dalam *automatic generation textual summaries* (Portet et al., 2009), yaitu mencoba untuk mendeteksi adanya pola sederhana dari *input* data numerik. Tujuan utama dari analisis sinyal adalah untuk mengganti data numerik tersebut menjadi sebuah pola diskrit. Hal ini menjadikan sistem dapat menggunakan pola secara simbolis (diskrit) dari pada dalam bentuk numerik.

Intinya, proses analisis sinyal merupakan proses menganalisis data *input* numerik untuk menghasilkan *output* berupa informasi yang akan disampaikan. Contohnya dalam kasus pembangkitan berita cuaca, *input* yang digunakan merupakan seluruh data cuaca yang terjadi selama satu bulan, lalu data tersebut dianalisis sehingga didapatkan sinya-sinyal dari sekumpulan data tersebut seperti hujan terbesar, hari terkering, dan lain-lain.

Pada penelitian DWP (Putra et al., 2017), dilakukan dengan pendekatan *Time-Series* digunakan karena setiap parameter data yang ada merupakan hasil dari sensor yang direkam pada waktu yang sama, pada tahap ini juga dilakukan pencarian pesan dasar (numerik) serta *Analysis* *Event* untuk setiap pesanyang terjadi selama satu bulan. Contoh implementasi pada tahap ini adalah melakukan peringkasan secara statistik (*mean*, *min*, *max, sum*) seperti pada gambar 2.3 dalam bahasa R.

|  |
| --- |
| ...  **for(**i **in** i**:**n**){**  max\_amt**[**i**]** **<-** max**(**LM**[**i**])** #Max value in parameter indeks i  min\_amt**[**i**]** **<-** min**(**LM**[**i**])** #Min value in parameter indeks i  sum\_amt**[**i**]** **<-** sum**(**LM**[**,i**])** #Total value in parameter indeks i  ....  **}**  LMmean\_result **<-** colMeans**(**xLM**)** #Last month average of all parameter  .... |

Gambar 2.3 Contoh implementasi Signal Analysis (Putra et al., 2017)

1. *Data Interpretation*

Langkah ke-dua setelah mendapatkan sinyal-sinyal dari proses analisis sinyal, yang harus dilakukan kemudian adalah menerjemahkan sinyal-sinyal yang telah didapatkan tersebut kedalam pesan dan menganalisis apakah ada relasi antara pesan-pesan yang didapatlan. Jadi, tujuan utama dari *Data Interpretation* ini adalah untuk memetakan pola dan *event* dasar menjadi pesan dan relasi dimana manusia membutuhkannya.

Sebagai contoh, misalnya terdapat data suhu udara hari ini senilai 40oC. Maka dengan melalui serangkaian proses interpretasi data ini, angka 40oC diiterpretasikan menjadi pesan *“very hot”*, seperti implementasi pada DWP (Putra et al., 2017) pada gambar 2.4 dengan berdasarkan pada *fuzzy membership function* (Ramos-soto, Bugarin, & Barro, 2016).

|  |
| --- |
| membership\_check **<-** **function(**partition,v,pname,oname**){**  i**=**1; n**=**length**(**partition**)**  membership\_value **<-** c**(**1**)**  **for(**i **in** i**:**n**){**  **....**  **if((**v**<**a**)||(**v**>**d**)){**  membership\_value**[**i**]** **<-** 0  **}**  **if((**v**>=**a**)&&(**v**<=**b**)){**  membership\_value**[**i**]** **<-** **(** **(**v**-**a**)** **/** **(**b**-**a**)** **)**  **}**  **if((**v**>**b**)&&(**v**<=**c**)){**  membership\_value**[**i**]** **<-** 1  **}**  **if((**v**>**c**)&&(**v**<=**d**)){**  membership\_value**[**i**]** **<-** **(** **(**d**-**v**)** **/** **(**d**-**c**)** **)**  **}**  **}**  i**=**1; biggest**=**0; part**<-**"a"  **for(**i **in** i**:**n**){**  **....**  **if(**biggest**<=**membership\_value**[**i**]){**  biggest **<-** membership\_value**[**i**]**  part**<-**pname**[**i**]**  **}**  **}**  return **(**part**)**  **}**  ....  Temperature\_partition **<-** c**(**"very cold.","cold.","warm.","hot.","very hot."**)**  Temperature\_interval **<-** list**(**very\_cold**=**c**(**a**=**0,b**=**0,c**=**5,d**=**10**)**,cold**=**c**(**a**=**5,b**=**10,c**=**15,d**=**20**)**,  warm**=**c**(**a**=**15,b**=**20,c**=**25,d**=**30**)**, hot**=**c**(**a**=**25,b**=**30,c**=**35,d**=**40**)**,very\_hot**=**c**(**a**=**35,b**=**40,c**=**45,d**=**50**))**  Temperature\_interval **<-** membership\_partition**(**Temperature\_interval,"Temperature"**)**  InterpretationResult\_temperature **<-** membership\_check**(**Temperature\_interval,as.double**(**climatePredictionResult**[**"Average.Temperature"**])**,Temperature\_partition," "**)**  ..... |

Gambar 2.4 Contoh implementasi data interpretation Rainfall DWP (Putra et al., 2017)

1. *Document Planning*

Langkah ke-tiga yang dilakukan dalam arsitektur ini adalah menentukan *event* mana yang akan disebutkan didalam teks, dan juga didalam struktur dokumen. Analisis sinyal dan *Data Interpretation* dapat menghasilkan sejumlah pesan, pola, dan *event* yang banyak, tetapi teks biasanya terbatas untuk mendeskripsikan sebagian kecil pesan. Perencanaan dokumen harus menentukan pesan mana yang sebenarnya dapat dikomunikasikan dalam bentuk teks, pilihan ini didasarkan pada genre dan domain. Dalam langkah ini juga harus direncanakan bagaimana pesan disebutkan dalam sebuah teks yang berkaitan antara satu dengan yang lainnya.

Menurut Reiter (2011) bahwa serangkaian proses *Document Planning* ini diantaranya adalah membagi tugas menjadi beberapa bagian berikut:

1. *Content Determintation*

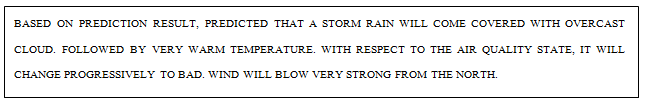
Tahap ini melakukan pemilihan *event* atau pesan yang didapatkan, idenya adalah membagi status pesan menjadi *Routine Message* dan *Significant Event Message*. *Routine Message* merupakan pesan-pesan yang akan selalu disampaikan disetiap pembangkitan kalimat, sedangkan *Significant Event Message* adalah pesan-pesan yang hanya akan disampaikan jika dan hanya jika indikasi pembangkitan dipenuhi. Artinya, *Significant Event Message* hanya disampaikan saat kondisi tertentu. DWP menerapkan *Significant Event Message* dalam menentukan event suatu hujan terjadi berturut turut secara *extreme* atau tidak seperti pada gambar 2.5.

|  |
| --- |
| RainExtremeMessage\_function **<-** **function(**Dataset**){**  **if(**interpreterRainfall **==** "no rain" **||** interpreterRainfall**==**"light rain" **||** interpreterRainfall**==**"moderate rain"**){**  return**(**"x"**)** #artinya tidak extreme  **}else{**  **.....** #get repeated date  **return (date)**  **}**  } |

Gambar 2.5 Contoh Content Determination Significant Event Message DWP (Putra et al., 2017)

1. *Documenet Structuring*

*Document Structuring* adalah proses penentuan bagaimana struktur pesan yang akan disampaikan. Urutan pesan-pesan ditentukan sesuai dengan relasinya masing-masing. Ada beberapa cara untuk membuat struktur dokumen, salah satunya adalah dengan menggunakan skema. Skema tersebut dibuat berdasarkan *Target Text* yang ingin dicapai atau berdasarkan penalaran (Reiter, 1996). Contoh “*Target Text*” dalam DWP dapat dilihat pada gambar 2.6.

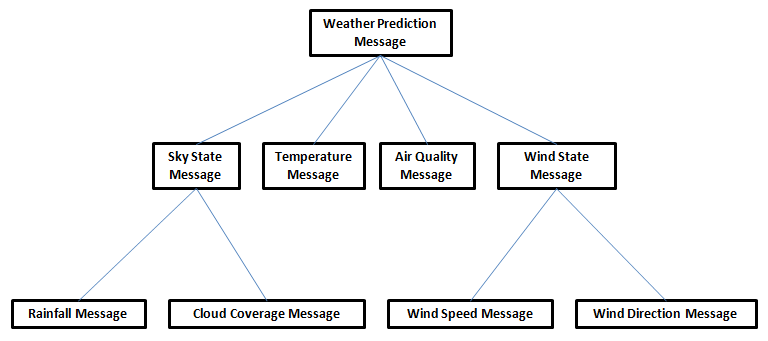


Gambar 2.6 Contoh Target Text DWP (Putra et al., 2017)

Reiter (1996) menjelaskan bahwa dengan berdasarkan *Target Text* pada gambar 4.10 dilakukan pembuatan skema dengan cara sebagai berikut:

* + - 1. Ambil contoh sejumlah teks yang sama dalam *Target Text*.
      2. Identifikasi terhadap pesan-pesan yang ada, tentukan bagaimana setiap pesan dapat dibangun berdasarkan data.
      3. Mengusulkan aturan atau struktur yang menjelaskan mengapa pesan “x” ada dalam teks A tetapi tidak ada didalam teks B. Penentuan ini lebih mudah jika disusun dalam bentuk seperti taksonomi atau pohon.
      4. Diskusikan hasil analisis bersama pakar.

Dengan berdasarkan tahapan diatas, maka didapat skema dalam bentuk pohon seperti pada gambar 2.7, dimana *Weather Prediction Message* terdiri dari empat komponen, yaitu *Sky State Message*, *Temperature Message*, *Air Quality Message*, dan *Wind State Message*.



Gambar 2.7 Contoh skema dalam bentuk tree berdasarkat Target Text DWP (Putra et al, 2017)

1. *Microplaning and Realisation*

Langkah ke-empat adalah membangkitkan bahasa alami dalam bentuk teks didasarkan pada konten dan struktur yang dipilih pada tahap perencanaan dokumen. Tahap *Microplanning* dan realisasi harus menentukan bagaimana sebenarnya mengekspresikan apa yang telah disusun pada tahap-tahap sebelumnya (*signal analysis, data interpretation,* dan *document planning*).

Dalam proses *Microplanning*, pesan-pesan yang disampaikan akan melalui serangkaian proses berikut:

1. *Lexicalisation*

Proses *lexicaisation* adalah bagaimana melakukan pemilihan kata atau frase yang akan digunakan dalam mendeksripsikan segala hal, contohnya mendeskripsikan relasi, tren, dan kemungkinan. DWP menjelaskan tren yang terjadi pada kualitas udara dengan membandingkan dua baris data, yakni data ke-*n* dan data ke-(*n-1*), dimana *n* merupakan jumlah baris data, seperti pada gambar 2.8.

|  |
| --- |
| ....  TrendDesc\_template **<-** **function** **(**IVL,data**){**  **if((**IVL**[**1**]==**"0"**)&&(**IVL**[**2**]==**"0"**)){**  TrendDesc **<-** change\_word\_bank\_AQ**(**"stable"**)**  **}**  **if(((**IVL**[**1**]==**"+"**)&&(**IVL**[**2**]==**"-"**))||((**IVL**[**1**]==**"-"**)&&(**IVL**[**2**]==**"+"**))){**  TrendDesc **<-** change\_word\_bank\_AQ**(**"mediumChange"**)**  **}**  **....**  return**(**TrendDesc**)**  **}**  .... |

Gambar 2.8 Contoh implementasi Lexicalisation tren DWP (Putra et al., 2017)

1. *Aggregation*

Proses *aggregation* adalah bagaimana setiap kata digabungkan menjadi frase, bagaimana frase dihubungkan menjadi kalimat, dan bagaimana kalimat digabungkan menjadi paragraf. Intinya, proses *Aggregation* adalah menghubungkan pesan yang didapat dengan menggunakan beberapa teknik. Ada beberapa teknik yang dapat dilakukan untuk proses *aggregation,* salah satu diantaranya adalah dengan menggunakan *simple conjunction*, seperti dalam pada gambar 2.9.

|  |
| --- |
| ....  Contrast\_lexicalisation1 **<-** **function(**msg1,msg2**){**  **if(**msg1**[**2**]==**msg2**[**2**]){**  return**(**"and"**)**  **}else{**  return**(**"but"**)**  **}**  **}**  .... |

Gambar 2.9 Contoh Simple Conjunction Referring to Contrast Value (Putra et al., 2017)

1. *Referring Expression Generation*

Proses ini berisi mengenai bagaimana sistem dapat merujuk informasi tertentu kepada sebuah subjek. Contohnya: “Suhu hari ini tergolong sangat panas”, sistem dikondisikan agar dapat menyampaikan bahwa informasi “sangat panas” adalah penjeasan informasi dari subjek “Suhu”. Salah satu contoh penerapan dengan cara *hardcode* pada kode program seperti pada gambar 2.10, dimana “*The wind for the month was*” merupakan subjek dari kalimat yang akan ditampilkan.

|  |
| --- |
| ....  MonthlyMsg4\_aggregation**<-function(**msg1,msg2**){**  **if(**msg2**==**"false"**){**  msg**<-**paste**(**"The wind for the month was",msg1,"in average."**)**  **}**  **else{**  msg**<-**paste**(**"The wind for the month was",msg1,"in average, but",msg2**)**  **}**  return**(**msg**)**  **}**  .... |

Gambar 2.10 Contoh Referring Expression Generation (Putra et al., 2017)

1. *Structure Realisation*

Pada proses ini, setiap struktur yang telah dibuat dalam proses *dokumen planning* direalisasikan sehingga menghasilkan teks dalam bentuk aktual (Reiter, 1996). Contohnya, merealisasikan struktur teks dalam bahasa pemrograman menjadi teks aktual dalam HTML, LaTeX, RTF, SABLE, dan lain-lain, seperti pada gambar 2.11, yang merealisasikan sturktur dalam bahasa R dengan menggabungkan hasil seleksi konten, kemudian menampilkan hasil kedalam HTML dengan bantuan *package* Shiny.

|  |
| --- |
| Structure\_Realization\_predict **<-** **function(){**  ....  Sky\_State **<-** Sky\_Agg**(**Rain\_State,Cloud\_State**)**  Sky\_Intro **<-** Prediction\_Intro**()**  Sky\_Sentence **<-** paste**(**Sky\_Intro,Sky\_State**)**  Temperature\_Intro **<-** Temperature\_Intro**()**  Temperature\_Sentence **<-** paste**(**Temperature\_Intro, Temperature\_State**)**  ....  **}**  ....  library**(**shiny**)**  .....  tags**$**div**(**class**=**"col-lg-8",  tags**$**h2**(**"Weather Prediction News"**)**,  tags**$**p**(**align**=**"justify",HTML**(**'&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;'**)**,  Prediction\_Result  **)**,  tags**$**hr**(**class**=**"style13"**)**,  tags**$**p**(**align**=**"justify",HTML**(**'&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;'**)**,    MonthlyMsg  **)**  **)**  ..... |

Gambar 2.11 Contoh implementasi Structure Realisation DWP (Putra et al., 2017)

## Penelitian Terkait *Data-to-Text*

Penelitian terkait dengan sistem *Data-to-text* akhir-akhir ini telah menjadi perhatian tertentu bagi para peneliti, ditunjukan dengan banyaknya penelitian baru terkait dengan bidang ini (D2T dan NLG). Beberapa penelitian sejauh ini mengenai *Data-to-text* dapat dilihat pada Tabel 2.1.



Tabel 1 Penelitian terkait D2T dan NLG

| Aplikasi / Penulis | Metode *Content Selection* | Domain | Sumber Data |
| --- | --- | --- | --- |
| (Kukich, 1983) | Rule-Based | Market | Database |
| (Boyd, 1998) | No Content Selection | Weather | Database |
| (S.G. Sripada, Reiter, Hunter, & Yu, 2001) | Two Stage model: (1) Domain Reasoner; (2) communicative reasoner | Weather, Oil Rigs | Sensor data, Numerical Data |
| (Somayajulu G. Sripada, Reiter, Hunter, & Yu, 2003) | Gricean Maxims | Weather, Gas Turbins, Heatlh | Sensor data |
| (Hallett, Power, & Scott, 2006) | Rule-Based | Heatlh | Database |
| (Yu, Reiter, Hunter, & Mellish, 2007) | Rules derived from corpus analysis and main knowledge | Gas Tourbines | Sensor |
| (Somayajulu G. Sripada & Gao, 2007) | Decompression Models | Dive | Sensor |
| (Turner, Sripada, Reiter, & Davy, 2008) | Decision Tree | Georeferenced Data | Database |
| (Gatt et al., 2009) | Rule-Based | Health | Sensor |
| (Thomas, Sripada, & Noordzij, 2012) | Document Schema | Georeferenced Data | Datatbase |
| (Demir, Carberry, & McCoy, 2012) | Rule-based | Domain Indepentent | Graph-database |
| (Reddington & Tintarev, 2011) | Threshold-based rules | Assitive Technology | Sensor |
| (Banaee, Ahmed, & Loutfi, 2013) | Rule-based | Health | Grid of sensor |
| (Schneider, Mort, Mellish, Reiter, & Wilson, 2013) | Rule-based | Health | Sensor |
| (Ramos-Soto, Bugarín, & Barro, 2016) | *Fuzzy*-sets | Weather | Database |
| (Gkatzia, Lemon, & Rieser, 2016) | Rule-based | Weather | Numerical data with assigned probabilities |
| (Putra et al., 2017) | Rule-Based and Fuzzy | Weather | Numerical data |

## *Machine Learning*

*Machine Learning* termasuk dalam bagian dari ilmu komputer yang dapat membelajarkan komputer sehingga memiliki kemampuan untuk belajar tanpa diprogram secara eksplisit (Samuel, 1959). *Machine Learning* merupakan bagian dari kecerdasan buatan yang berfokus dalam mempelajari, mendesain, dan membuat sebuah algoritma yang memiliki kemampuan untuk belajar dari data yang ada. Agar sebuah perangkat memiliki kecerdasan, maka komputer atau mesin tersebut harus dapat belajar. Dengan kata lain, *Machine Learning* berisi tentang keseluruhan proses pembelajaran komputer atau mesin menjadi cerdas dan dapat belajar dari data. *Machine Learning* sudah ada dan mulai digunakan sejak 50 tahun yang lalu dan sudah banyak digunakan di berbagai bidang. Contohnya pada bidang ekonomi, keilmuan, industri dan sebagainya*.*

Salah satu implementasi *Machine Learning* yang pernah dilakukan oleh Arthur Samuel sekitar 59 tahun yang lalu yaitu pembuatan permainan catur dengan computer (Samuel, 1959). Catur dipilih karena permainan sangat mudah tetapi memerlukan strategi yang bagus. Samuel membuat permainan catur ini berdasarkan pohon penyelesaian. Pencarian penyelesain dilakukan dengan menyusuri pohon permasalahan sampai mendapatkan solusinya*.*

Awal ditemukannya *Machine Learning* yaitu pada tahun 1914, seorang ilmuan dari Spanyol, Torres y Quevedo, membuat sebuah mesin catur yang dapat mengalahkan atau melakukan skakmat pada raja lawan dengan sebuah ratu dan raja (Shannon, 1950). Perkembangan secara sistematis kemudian dimulai segera setelah diketemukannya komputer digital.

Artikel ilmiah pertama tentang Kecerdasan Buatan ditulis oleh Alan Turing pada tahun 1950 (Turing, 1950), dan kelompok riset pertama dibentuk tahun 1954 di Carnegie Mellon University oleh Allen Newell and Herbert Simon. Namun bidang Kecerdasan Buatan baru dianggap sebagai bidang tersendiri di konferensi Dartmouth tahun 1956, di mana 10 peneliti muda memimpikan mempergunakan komputer untuk memodelkan bagaimana cara berfikir manusia. Mereka berhipotesis bahwa mekanisme berfikir manusia dapat secara tepat dimodelkan dan disimulasikan pada komputer digital.

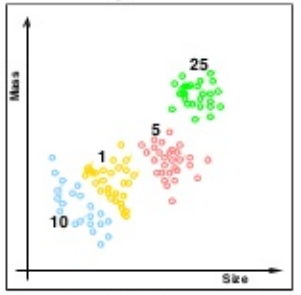
*Machine Learning* memiliki beberapa tipe dengan proses pembelajaran yang berbeda, tipe-tipe tersebut akan dijelaskan pada sub-bab berikutnya

### *Supervised Learning*

Tugas dari *Supervised Learning* terdiri dari pembangunan model yang memetakan nilai *input* pada nilai ouput dimana *data* training tersedia (Riza, 2015). *Supervised Learning* adalah *Machine Learning* yang membutuhkan label sebagai tujuan dari pelatihan data atau *data training* (Mohri, Rostamizadeh, & Talwalkar, 2012). *Supervised Learning* merupakan suatu pembelajaran yang terawasi, dimana jika *output* yang diharapkan telah terdapat pada daftar yang diketahui sebelumnya. Pada metode ini, setiap pola yang diberikan kedalam model *Machine Learning* telah diketahui *output*nya. Contoh algoritma dari salah satu bagian dari *Machine Learning* yaitu jaringan saraf tiruan yang mernggunakan metode *Supervised Learning* adalah hebbian (hebb rule), perceptron, adaline, boltzman, hapfield, dan backpropagation*.*

Berikut ini adalah beberapa contoh penerapan tipe *Machine Learning*, *Supervised Learning*:

1. Klasifikasi: sebuah metode untuk menyusun data secara sistematis menurut aturan-aturan yang telah ditetapkan sebelumnya (Athoillah, Irawan, M., & Imah, Elly, 2015). Dengan melakukan klasifikasi, dari data yang telah ada dapat dibuat sebuah model prediksi dengan *output* kelas. Beberapa algoritma klasifikasi yang cukup terkenal adalah k-Means, SVM, EM, Naïve Bayes, dan kNN.
2. Regresi: Analisis regresi adalah salah satu metode statistik untuk memprediksi nilai dari satu atau lebih variabel respon/dependen dari satu set variabel prediktor/independen (Härdle & Simar, 2007).

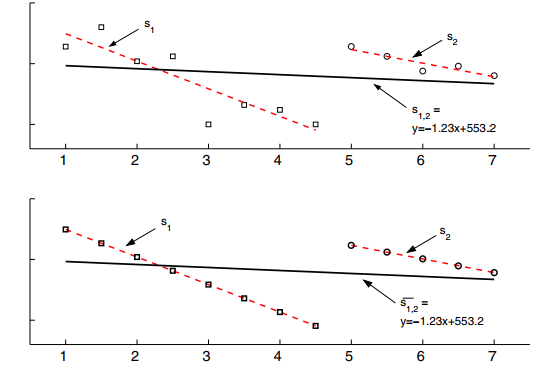


Gambar 2.12 Contoh *Supervised Learning* pada pengenalan koin

Pada Gambar 2.12, diperlihatkan bagaimana klasifikasi dari pengenalan koin, terlihat sangat jelas lokasi bagian dari tiap kelas, seperti koin dengan nilai sepuluh terpisah dipaling bawah dengan warna biru, koin dengan nilai satu yang berwarna kuning tidak bercampur dengan yang lainnya, dan seterusnya. Beberapa contoh penerapan *Supervised Learning* yaitu pada kasus klasifikasi sentimen (Ye, Zhang, & Law, 2009), prediksi virus hepatitis B (Hospital, 2003), dan *Gait Event Detection* (Williamson & Andrews, 2000). Pada penelitian ini juga diterapkan teknik *Piecewise Linear Approximation* dengan *Least Square Method* yang termasuk dalam kategori regresi.

* + - 1. ***Piecewise Linear Approximation***

*Piecewise Linear Approximation* (PLA)adalah teknik yang digunakan untuk merepresentasikan data dalam skala besar dengan cara memecah data menjadi beberapa segmen. Skema PLA memungkinkan untuk digunakan dalam proses yang berkelanjutan untuk mencari sebuah solusi (Palpanas, Vlachos, Keogh, Gunopulos, & Truppel, 2004).



Gambar 2.13 Kombinasi dua garis linier (Palpanas et al., 2004).

Gambar 2.13 menjelaskan bahwa diasumsikan ada *N* titik data *time series*, T[*i*], 1 ≤ *i ≤ N*, kemudian kita gunakan untuk menyesuaikan data tersebut kedalam dua segmen *linear* (dapat menggunakan *Linear Square* atau metode lain). Segmen 1 untuk data ke-1 sampai *n*, segmen 2 untuk data ke-(*n*+1) sampai *N*. Kemudian kita lakukan penyesuaian data untuk ke-dua segmen kedalam sebuah garis *linear* hingga didapat dan untuk garis segmen 1 dan 2. Bayangkan jika data asli yang kita miliki hilang atau kita hapus untuk selamanya dan yang kita miliki hanya dan , kemudian kita ingin memperkirakan semua titik data aslinya hanya dengan satu garis *linear*. Sebuah garis *linear* yang kita dapatkan dari dan kita sebut dan telah dibuktikan garis tersebut sama dengan garis aslinya seperti pada gambar 2.13 (Palpanas et al., 2004).

* + - 1. ***Least Square Method***

*Least Square* adalah metode untuk menentukan fungsi pendekatan polynomial y = f(x) yang paling mendekati data (x1, y1) sampai (xn, yn) (Cliff & Billy, 2017). Rumus untuk mendapatkan model linear dengan metode kuadrat terkecil (*least square method*) yaitu:

(1)

Dimana untuk mencari nilai a dan b diperoleh berdasarkan:

(2)

(3)

(4)

(5)

Dimana *y’* menunjukan nilai taksiran Y pada nilai t tertentu. Sedangkan *a* adalah nilai *intercept* dari Y. Kemudian *b* adalah nilai *slope*, artinya besar kenaikan nilai *y’* pada setiap nilai *t*. Nilai *t* sendiri adalah nilai tertentu yang menunjukan periode waktu, dan *n* adalah jumlah data.

Tabel 2.2 Contoh data penjualan dengan periode bulanan

| Bulan | Penjualan |
| --- | --- |
| Januari  Februari  Maret  April  Mei  Juni  Juli | 38  23  29  49  57  140  33 |
| Total | 369 |

Tabel 2.3 Perhitungan metode *least square*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Bulan | y | t |  | 2 |  |
| Januari  Februari  Maret  April  Mei  Juni  Juli | 38  23  29  49  57  140  33 | 1  2  3  4  5  6  7 | -3  -2  -1  0  1  2  3 | 9  4  1  0  1  4  9 | -14,71  -29,71  -23,71  -3,71  4,29  87,29  -19,71 |
| Σ(total) | 369 |  | 0 | 29 | 0,03 |

Contoh kasus, bila kita memiliki data penjualan dengan periode bulanan, seperti pada tabel 2.2 kemudian kita ingin merepresentasikan data tersebut kedalam sebuah garis *linear* menggunakan metode *least square*. Perhitungan berdasarkan data tabel 2.3 dengan persamaan (2), persamaan (3), persamaan (4), dan persamaan (1) didapat bahwa:

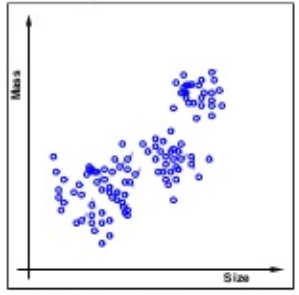
|  |
| --- |
|  |

Dengan demikian kita dapat merepresentasikan data dengan model *linear* yang telah didapat dengan melakukan perulangan.

### *Unsupervised Learning*

*Unsupervised Learning* terdiri dari pembangunan model dari *data training* dengan tidak mengandung nilai *output* (Riza, 2015). *Unsupervised Learning* merupakan pembelajan yang tidak terawasi dimana tidak memerlukan target *output*. Teknik ini menggunakan prosedur yang berusaha untuk mencari partisi dari sebuah pola. *Unsupervised Learning* mempelajari bagaimana sebuah sistem dapat belajar untuk merepresentasikan pola *input* dalam cara yang menggambarkan struktur statistikal dari keseluruhan pola *input*. Berbeda dari *Supervised Learning*, *Unsupervised Learning* tidak memiliki target *output* yang eksplisit atau tidak ada pengklasifikasian *input.*

Dalam *Machine Learning*, teknik *Unsupervised* sangat penting. Hal ini dikarenakan cara kerjanya mirip dengan cara bekerja otak manusia. Dalam melakukan pembelajaran, tidak ada informasi dari contoh yang tersedia. Oleh karena itu, *Unsupervised Learning* menjadi esensial. Pada metode ini tidak dapat ditentukan hasil seperti apa yang diharapkan selama proses pembelajaran, nilai bobot yang disusun dalam proses range tertentu tergantung pada nilai *output* yang diberikan. Tujuan metode *Unsupervised Learning* ini agar kita dapat mengelompokkan *Unit*-*Unit* yang hampir sama dalam satu area tertentu. Pembelajaran ini biasanya sangat cocok untuk klasifikasi pola. Contoh algoritma jaringan saraf tiruan yang menggunakan metode *Unsupervised* ini adalah competitive, hebbian, kohonen, *Learning Vector Quantizatio (*LVQ), dan neocognitron.



Gambar 2.14 Contoh *Unsupervised Learning* dalam pengenalan koin.

Salah satu contoh dari *Unsupervised Learning* adalah clustering, sistem diharapkan mampu untuk memisahkan data serupa ke dalam kelompoknya masing-masing, seperti pada Gambar 2.14, belum diketahui kelas dari masing-masing data, mesinlah yang menentukan berdasarkan kedekatannya. Beberapa contoh penerapan *Unsupervised Learning* diantaranya sistem pendeteksi intrusi (Zanero & Savaresi, 2004), menemukan komunitas pengguna di Internet sesuai dengan kriteria (Paliouras, Papatheodorou, Karkaletsis, & Spyropoulos, 2002), dan pengembangan strategi pengendalian manufaktur (Bowden & Bullington, 1996).

### *Semi Supervised Learning*

*Semi Supervised Learning* adalah penggabungan dari *Supervised* dan *Unsupervised Learning*. Dimana hasil keluaran sistem ada yang termasuk dalam kategori yang sudah ditetapkan namun ada juga yang tidak. Beberapa contoh penerapannya yaitu kasus representasi kata (Turian, J., Ratinov, L., & Bengio, 2010), *Co-Tracking* (Tang, Brennan, Zhao, & Tao, 2007), dan identifikasi peptida (Käll, Canterbury, Weston, Noble, & MacCoss, 2007)*.*

### *Reinforcement Learning*

Pada *Reinforcement Learning* model yang dihasilkan terus berkembang seiring pemakaian oleh pengguna, dimana model terus menerus diperbaiki sesuai kondisi penerapan. Beberapa contoh penerapannya yaitu pada simulasi sepak bola dalam *RoboCup* (Stone, Sutton, & Kuhlmann, 2005) dan penerbangan helikopter terbalik secara otonom (Ng et al., 2006)*.*

## Data *Time Series*

Kumpulan data yang tercatat dalam periode waktu mingguan, bulanan, kuartalan, atau tahunan (Mishra & Jain, 2014). Ada 4 faktor yang mempengaruhi data *Time Series*. Dalam data ekonomi biasanya didapatkanadanya fluktuasi atau variasi dari waktu ke waktu atau disebut dengan variasi *Time Series*.  Variasi ini biasanya disebabkan oleh adanya faktor *Trend* (*trend factor*), Fluktuasi siklis (*cyclical fluktuation*), Variasi musiman (*seasonal variation*), dan pengaruh *random* (*irregular atau random influences*)*.* Beberapa contoh data *time series* diantarnya kurs mata uang, data klimatologi, data pertumbuhan penduduk, dan lainnya.

*Trend* adalah keadaan yang merepresentasikan kecenderungan data yang menaik atau menurun dari waktu ke waktu.  Contoh yang menunjukkan trend menaik yaitu pendapatan per-kapita, jumlah penduduk, sedangkan trend yang menurun adalah volume minyak bumi dari waktu ke waktu. Variasi musiman adalah fluktuasi yang muncul secara reguler setiap tahun yang biasanya disebabkan oleh iklim, kebiasaan (mempunyai pola tetap dari waktu ke waktu).  Contoh yang menunjukan variasi musiman seperti penjualan pakaian atau kue kering akan meningkat pada saat hari raya ‘idul fitri*.*

Variasi siklis muncul ketika data dipengaruhi oleh fluktuasi ekonomi jangka panjang, variasi siklis ini bisa terulang setelah jangka waktu tertentu. Variasi siklis biasanya akan kembali normal setiap 10 atau 20 tahun sekali, bisa juga tidak terulang dalam jangka waktu yang sama.  ini yang membedakan antara variasi siklis dengan musiman. Gerakan siklis tiap komoditas mempunyai jarak waktu muncul dan sebab yang berbeda-beda, yang sampai saat ini belum dapat dimengerti.  Contoh yang menunjukkan variasi siklis seperti industri konstruksi bangunan mempunyai gerakan siklis antara 15-20 tahun sedangkan industri mobil dan pakaian gerakan siklisnya lebih pendek lagi.  Variasi *random* adalah suatu variasi atau gerakan yang tidak teratur (*irregular*).  Variasi ini pada kenyataannya sulit diprediksi. Contoh variasi ini dalam data *Time Series* karena adanya perang, bencana alam dan sebab-sebab unik lainnya yang sulit diduga.  Total variasi dalam data *Time Series* adalah merupakan hasil dari keempat faktor tersebut yang mempengaruhi secara bersama-sama.  Dalam tulisan ini hanya akan dianalisa dua variasi pertama, sedangkan dua variasi terakhir tidak dianalisa karena memang pola variasi tersebut tidak tersistem dengan baik selain membutuhkan waktu yang sangat lama untuk mendapatkan data yang panjang.

Model *Time Series* adalah suatu peramalan nilai-nilai masa depan yang didasarkan pada nilai-nilai masa lampau suatu parameter dan atau kesalahan masa lampau. Model *Time Series* biasanya lebih sering digunakan untuk suatu peramalan/prediksi. Dalam teknik peramalan dengan *Time Series* ada dua kategori utama yang perlu dilakukan pengujian, yaitu pemulusan (*smoothing*) dan dekomposisi (*decomposition*). Metode pemulusan mendasarkan ramalannya dengan prinsip rata-rata dari kesalahan masa lalu (*averaging smoothing past errors*) dengan menambahkan nilai ramalan sebelumnya dengan persentase kesalahan (*percentage of the errors*) antara nilai sebenarnya (*actual value*) dengan nilai ramalannya (*forecasting value*).  Metoda dekomposisi mendasarkan prediksinya dengan membagi data *Time Series* menjadi beberapa komponen dari Trend, Siklis, Musiman dan pengaruh *Random*. Kemudian mengkombinasikan prediksi dari komponen-komponen tersebut (kecuali pengaruh *random* yang sulit diprediksi). Pendekatan lain untuk peramalan adalah metoda causal atau yang lebih dikenal dengan sebutan regresi.

## Data *Streaming*

Data *streaming* adalah urutan urutan kejadian yang hanya dapat dibaca sekali atau beberapa kali (Babcock, Babu, Datar, Motwani, & Widom, 2002; Domingos & Hulten, 2001; Gaber, Zaslavsky, & Krishnaswamy, 2005) 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. Artinya data ini akan sulit untuk kita simpan kedalam sebuah program, untuk dilakukan perhitungan mengingat ukuran data akan terus bertambah besar, dan penyimpanan akan sangat terkuras habis oleh data yang menumpuk dari waktu ke waktu (Gama & Gaber, 2007)*.*

Data *streaming* dihasilkan oleh sensor yang aktif merekam serangkaian kejadian. Pengaplikasian sensor tergolong dalam beberapa jenis, yaitu pemantauan, perekaman jejak, dan kontrol (Gama & Gaber, 2007). Dengan berdasarkan definisi diatas tentu sangat mudah untuk menemukan data *streaming* dalam lingkungan kita, seperti pada telepon genggam, sensor pada mobil, sensor keamanan, televisi, *cctv*, dan lain sebagainya. Tentu kita pernah melihat atau bahkan menggunakan peralatan tersebut, namun kita tidak pernah melihat secara langsung pergerakan data yang ada didalamnya.

Beberapa model data *stream* berdasarkan aliran data diantaranya (Muthukrishnan, 2005):

1. *Time Series Model*: Aliran data ini berdasarkan waktu, baik itu dalam milidetik, detik, menit, jam, hingga tahunan.
2. *Cash Register Model*: Aliran data ini berdasarkan iterasi pada setiap transaksi yang dilakukan, semakin banyak alat kasir dan transaksi yang terjadi, semakin cepat pula pergerakan data*.*
3. *Turnstile Model*: Aliran data ini seperti pintu geser, data terkadang dapat dimasukan atau dikeluarkan secara dinamis, model ini tergolong sulit untuk mengimplementasikannya

Perbedaan data *streaming* dengan data lainnya adalah 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 menggali pengetahuan dari berbagai aliran data yang terdistribusi, adalah salah satu tantangan paling signifikan yang kita hadapi saat ini, bahkan hingga algoritma konvensional untuk mengolah data statsis menjadi diragukan untuk diimplementasikan, karena dalam pengembangan sebuah sistem untuk data *streaming* kita perlu menggali informasi dengan penerapan algoritma yang dapat bekerja secara efisien (Gama & Gaber, 2007).

Beberapa contoh aplikasi yang menggunakan data *stream* (Babcock et al., 2002):

1. *Traderbot*: Mesin pencari keuangan berbasis web yang mengevaluasi perintah melalui data keuangan *streaming real-time* seperti ticker saham dan bahan dasar suatu berita.
2. *iPolicy Networks*: Sistem keamanan terintegrasi yang menyediakan layanan seperti dukungan *firewall* dan deteksi intrusi melalui aliran paket jaringan multi-gigabit.
3. *Yahoo*: Situs web besar memonitor log web (*clickstream*) secara *online* untuk mengaktifkan aplikasi seperti personalisasi, pemantauan kinerja, dan *load-balancing*.

Untuk pengiriman pemrosesan dan penggalian informasi data *streaming* dapat dilakukan dengan dua pendekatan, yaitu *Centralized Approach* dan *Distributed Approach* (Gama & Gaber, 2007). *Centralized Approach* yaitu data dikirim ke sebuah server pusat yang menjalankan sebuah sistem penggalian data, hal ini menjadi solusi paling sederhana, namun tentu akan memerlukan memori yang besar dan pemrosesan yang tidak sebentar. *Distributed Approach* dilakukan pada sensor itu sendiri, dimana sebuah sensor bukanlah sekedar penangkap data, namun sensor bekerja dengan saling berkomunikasi dengan sensor lainnya untuk melakukan pemrosesan data, sehingga penggalian informasi tersebar pada setiap sensor.

Dalam pengembangan sistem yang mengimplementasikan *Machine Learning* atau *Data Mining*, disarankan memenuhi kriteria (Domingos & Hulten, 2001):

* Memiliki *Computation Time* yang sangat baik, hal ini dikarenakan bila sistem bekerja dengan waktu yang lama maka akan terjadi tumpukan data untuk diproses selanjutnya, dan akan terus menumpuk yang mengakibatkan sistem akan berhenti bekerja karena memori yang digunakan terlampau besar.
* Pemrosesan dilakukan pada memori utama, tanpa memandang jumlah data yang akan di proses.
* Dapat membangun model yang merepresentasikan setiap data yang diproses pada suatu waktu, karena data lampau tidak akan mungkin dapat ditelusuri kembali.
* Dapat membangun model yang merepresentasikan keseluruhan data.
* Model yang dibuat harus selalu mutakhir, tetapi juga harus mencakup semua informasi dari data masa lalu.

## Pemrograman R

Bahasa R merupakan sebuah proyek yang dirancang sebagai bahasa pemrograman yang gratis, *open source*, yang dapat digunakan sebagai pengganti dari bahasa pemrograman Splus, pada mulanya dikembangkan sebagai bahasa S di *AT&T Bell Labs*, dan sekarang dipasarkan oleh *Insightful Corporation of Seattle,* di Washington. R adalah sistem untuk komputasi statistik dan grafik. Sebagai sebuah sistem, R memiliki banyak sekali fitur. Sebagai bahasa pemrograman, R memiliki visualisasi grafik yang *high level,* antarmuka ke bahasa pemrograman lain, dan fasilitas *debugging* (Spector, 2004)*.* Logo dari bahasa pemrograman R sendiri dapat dilihat pada Gambar 2.15*.*



Gambar 2.15 Logo bahasa pemrograman R, sumber: https://www.r-project.org.

Berikut adalah kelebihan dari penggunaan bahasa R (Ihaka & Gentleman, 2012):

1. Serba guna (*versatile*)

R adalah bahasa pemrograman, sehingga tidak ada batasan bagi pengguna untuk memakai prosedur yang hanya terdapat pada paket-paket yang standar. Bahkan pemrograman R adalah berorientasi obyek dan memiliki banyak library yang sangat bermanfaat yang dikembangkan oleh kontributor. Pengguna bebas menambah dan mengurangi *library* tergantung kebutuhan. R juga memiliki antarmuka pemrograman C, phyton, bahkan java yang tentu saja berkat usaha serta kerja keras para kontributor aktif proyek R. Jadi selain bahasa R ini cukup pintar, penggunaannya pun bisa menjadi lebih pintar dan kreatif. Beberapa analisis yang membutuhkan fungsi lanjutan memang ada yang belum tersedia dalam R. Tidak berarti R tidak menyediakan fasilitas tersebut, namun lebih karena faktor waktu. Jadi hanya menunggu waktu saja *package* lanjutan tersebut tersedia

1. Interaktif (*interactive*)

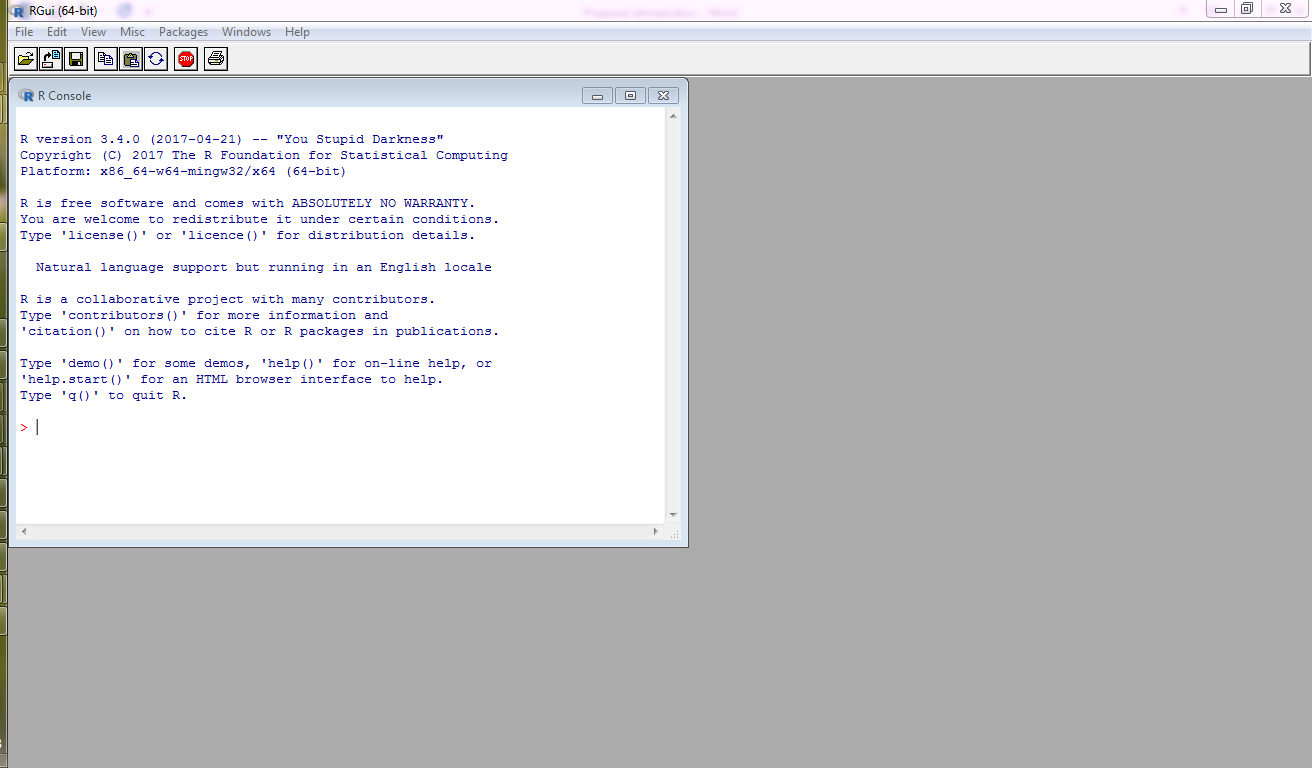
Pada saat ini analisis data membutuhkan pengoperasian yang interaktif. Apalagi jika data yang dianalisis adalah data yang bergerak. R dilengkapi dengan konektivitas ke database server, olap, maupun format data web service seperti XML, spreadsheet dan sebagainya. Sehingga apabila data set berubah hasil analisis pun dapat segera ikut berubah (*real time*).

1. Berbasis S yaitu turunan dari tool statitistik komersial S-Plus

R hampir seluruhnya kompatibel dengan S-Plus. Artinya sebagian besar kode program yang dibuat oleh S dapat dijalankan di S-plus kecuali fungsi-fungsi yang sifatnya *add-on packages* atau tambahan yang dibuat oleh kontributor proyekR.

1. Populer

Secara umum SAS adalah *software* statistika komersial yang populer, namun demikian R atau S adalah bahasa yang paling populer digunakan oleh peneliti di bidang statistika. Beberapa tulisan berupa jurnal statistika mengkonfirmasi kebenaran hal ini. R juga populer untuk aplikasi kuantitatif dibidang keuangan.

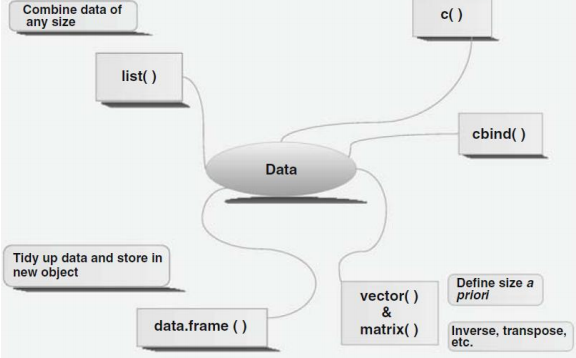


Gambar 2.16 Antarmuka R Graphical User Interface (RGui).

RGui merupakan tools dalam pemrograman R, antar muka tools ini dapat dilihat pada Gambar 2.16, diperlihatkan bahwa dalam antarmuka RGui terdapat layar *console* yang berfungsi untuk memasukan perintah, terdapat *menu-bar*, *tool-bar*dan lain-lain sesuai dengan fungsinya masin-masing.

### Model Data dalam R

Pada bahasa R, data dipandang sebagai suatu objek yang memiliki suatu atribut dan berbagai fungsionalitas (Budiharto, 2013). Sifat data ditentukan oleh type data dan mode data. Ada berbagai type data yang dikenal oleh R, antara lain vektor, matriks, list, data frame, *array*, *factor* dan fungsi *built in*. Berikut ini beberapa model data yang umum digunakan serta contoh penerapan fungsi *built in*. Untuk menyimpan data di R ada berbagai metode seperti menggunakan fungsi c(), list(), cbind() dan data.frame() seperti Gambar 2.17.



Gambar 2.17 Model data dalam pemrograman R (Budiharto, 2013).

### Contoh Kode Program Bahasa R

Berikut adalah beberapa contoh kode program yang dapat dilakukan oleh bahasa pemrograman R.

1. Menggabungkan data

Untuk menggabungkan data dalam bahasa R, dapat menggunakan fungsi *concantenate* (c). contoh dari penggunaan fungsi ini dapat dilihat pada Gambar 2.18.

|  |
| --- |
| x **<-** c**(**1,2,3,4,5**)**  y **<-** c**(**6,7,8,9**)**  z **<-** c**(**x,y**)**  z  # output  # [1] 1 2 3 4 5 6 7 8 9 |

Gambar 2.18 Operator *concantenate* dalam R.

Pada Gambar 2.19 berikut adalah contoh untuk menampilkan dua data pertama dalam vektor z yang telah dibuat.

|  |
| --- |
| x**[**1**:**2**]**  # output  # [1] 1 2 |

Gambar 2.19 Menampilkan data pertama hingga ke-dua dalam R.

Selain itu, untuk menampilkan jumlah dari seluruh elemen, dapat digunakan fungsi sum. Implementasi dari fungsi sum ini dapat dilihat pada Gambar 2.20.

|  |
| --- |
| sum**(**x**)**  # output  # [1] 15 |

Gambar 2.20 Penggunaan fungsi *sum* dalam R.

Contoh lainnya, untuk memasukan data string, dapat dilihat pada Gambar 2.21.

|  |
| --- |
| Z **<-** c**(**"ahmad", "zainal", "abidin"**)**  Z  # output  # [1]"ahmad" "zainal" "abidin" |

Gambar 2.21 Penggunaan fungsi *concantenate* untuk *string* dalam R

1. Membuat matriks

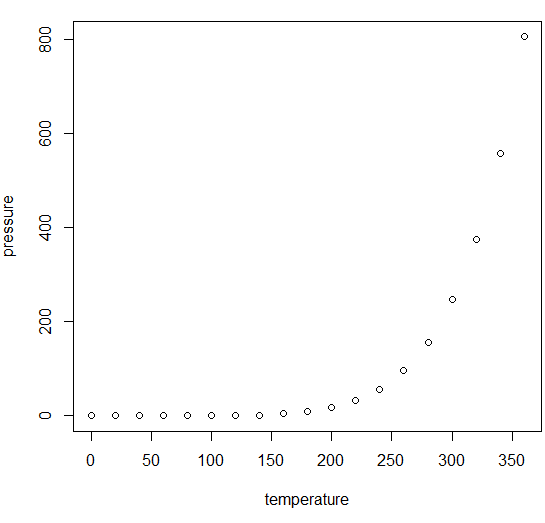
Matriks adalah data dua dimensi dimana sebagian besar fungsi‐fungsi statistik dalam R dapat dianalisis dengan menggunakan bentuk matriks. Bentuk matriks ini juga banyak digunakan pada operasi fungsi‐fungsi built‐in untuk aljabar *linear* dalam R, seperti untuk penyelesaian suatu persamaan *linear*. Argumen yang diperlukan adalah elemen‐elemen dari matriks, dan argumen optional yaitu banyaknya baris dan banyaknya kolom. Berikut contohnya ada pada Gambar 2.22.

|  |
| --- |
| M **<-** matrix**(**c**(**1**:**6**)**,nrow**=**2,ncol**=**3**)**  m  # output  # [,1] [,2] [,3]  # [1,] 1 3 5  # [2,] 2 4 6 |

Gambar 2.22Pembuatan matriks dalam R*.*

1. Membuat *visualisasi* data

Salah satu keunggulan dalam bahasa pemrograman R adalah visualisasi data dapat disajikan dengan mudah. Data yang berhasil dientri atau diimport dari aplikasi lain selayaknya divisualisasikan pada grafik untuk analisa. Sebagai contoh, kita dapat menggunakan data dari R yaitu variabel pressure, dengan command “*plot(pressure)”* maka akan menghasilkan grafik seperti pada Gambar 2.23.



Gambar 2.23Contoh *Visualisasi* grafis dalam R.

1. Membuat perulangan

Salah satu cara yang paling populer hampir diseluruh bahasa pemrograman dalam melakukan perulangan adalah fungsi FOR. Contoh implementasi fungsi FOR dalam bahasa R dapat dilihat pada Gambar 2.24.

|  |
| --- |
| i **<-** 1  n **<-** 10  **for(**i **in** i**:**n**){**  print**(**"Hello World!"**)**  **}**  # output  # [1] "Hello World!"  # [1] "Hello World!"  # [1] "Hello World!"  # [1] "Hello World!"  # [1] "Hello World!"  # [1] "Hello World!"  # [1] "Hello World!"  # [1] "Hello World!"  # [1] "Hello World!"  # [1] "Hello World!" |

Gambar 2.24Contohperulangan dalam R*.*

1. Membuat *decision*

Membuat *decision* dalam dunia pemrograman adalah hal yang paling utama. Dalam bahasa R, membuat decision identik dengan bagaimana bahasa C melakukannya. Dapat dilihat pada Gambar 2.25.

|  |
| --- |
| i **<-** 1  n **<-** 10  **if(** n **>** i**){**  print**(**"Hello World!"**)**  **}**  # output  # [1] "Hello World!" |

Gambar 2.25Contohimplementasi *decission* dalam R*.*

1. Membuat fungsi

Dalam pemrograman terstruktur, salah satu hal yang penting adalah membuat fungsi. Dalam bahasa R, contoh pembuatan fungsi dapat dilihat pada Gambar 2.26.

|  |
| --- |
| Penjumlahan **<-** **function(**a,b**){**  X **<-** a**+**b  return**(**X**)**  **}**  Penjumlahan**(**1,2**)**  # output  # [1] 3 |

Gambar 2.26Contohfungsi dalam R.

### *Package* dalam Bahasa R

Secara konseptual, R *package* adalah kumpulan fungsi, objek data, dan dokumentasi yang secara koheren mendukun operasi analisis data. R adalah bahasa pemrograman *open-source* dan lingkungan analisis yang mengandung lebih dari 8000 *packages* untuk statistik, *bio-informatics,* visualisasi, *Machine Learning,* ekonomi, dan lain-lain (Ihaka & Gentleman, 2012). Bahkan, sampai 30 Juli 2018 banyak *packages* yang terpublish dalam *cran-r project* lebih dari 12800 *packages*. Agar mudah digunakan dan untuk menjaga kualitasnya serta untuk terus mempertahankannya, kebanyakan R *package* disimpan di repositori berikut: Jaringan Arsip R Komprehensif (CRAN, http://cran.r-project.org/) dan Bioconductor Project. (Http://www.bioconductor.org/) (Riza et al., 2016).

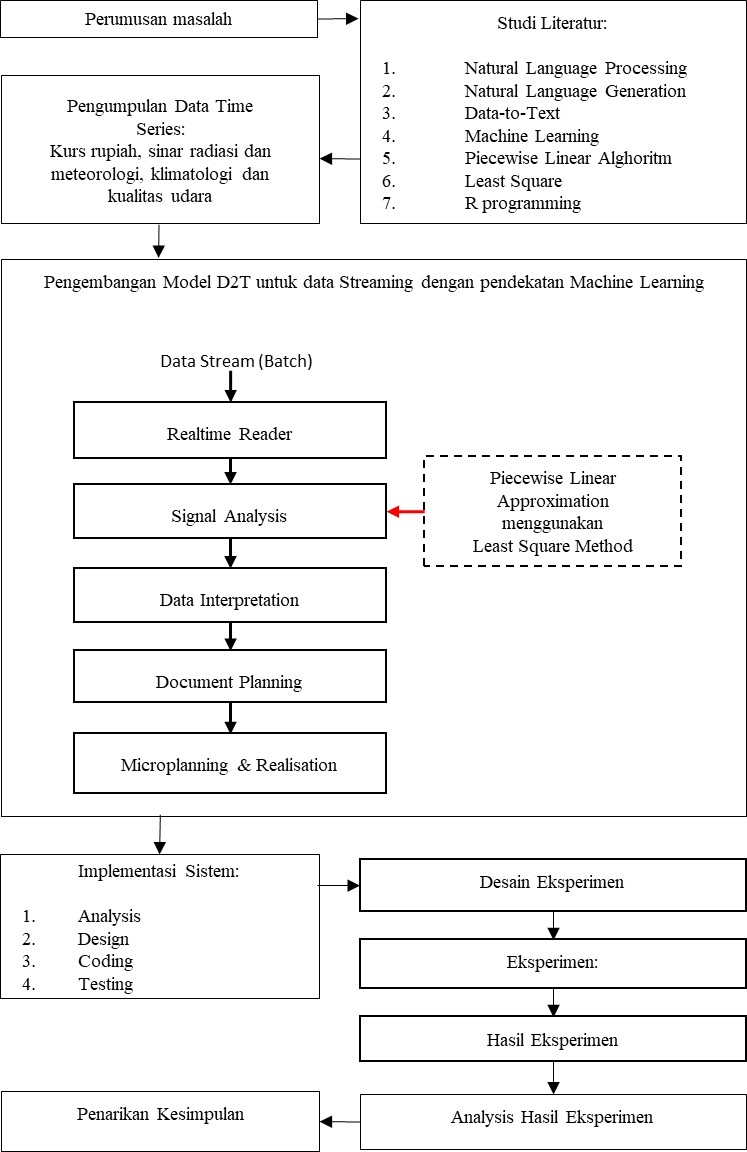
# BAB III METODE PENELITIAN

Pada bab ini akan dijelaskan mengenai metodologi penelitian, mulai dari desain penelitian, metode penelitian, alat penelitian dan data penelitian.



## Desain Penelitian

Desain penelitian adalah kerangka kerja yang digunakan untuk melakukan penelitian. Pada bagian ini penulis akan memaparkan kerangka kerja dari mulai penelitian sampai selesai. Desain penelitian yang digunakan dalam pembangunan sistem *Data-to-Text* berdasarkan data *Streaming* dengan pendekatan *Machine Learning* digambarkan pada gambar 3.1.



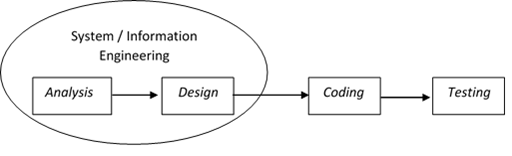
Gambar 3.1 Desain Penelitian

Langkah-langah penelitian yang dilakukan meliputi:

1. Tahap Persiapan adalah tahap awal dari penelitian, tahap ini dimulai dari identifikasi masalah, kemudian merumuskan masalah, lalu menentukan algoritma atau metode yang akan digunakan untuk menyelesaikan masalah tersebut.
2. Studi literatur merupakan kegiatan dimana peneliti melakukan tahap pembelajaran materi yang terkait dengan *Natural Language Generation*, *Data-to-Text*, data *Time-Series*, data *Streaming*, pendekatan *Machine Learning* pada *Data-to-Text*, penggunaan bahasa pemrograman R, serta penelitian terkait dengan topik ini. Dalam mempelajari tentang bahasan di atas penulis mempelajari dari beberapa sumber, seperti buku, jurnal, juga internet, ataupun bahan bacaan lainnya yang didapat dari berbagai sumber.
3. Pengumpulan data dilakukan dengan mendownload data nilai tukar rupiah terhadap mata uang lain (kurs) di situs Bank Indonesia, *solar radiation and meteorological*dan klimatologi.
4. Pengembangan Sistem merupakan tahap pembuatan program dalam bahasa R untuk mengolah data yang telah dikumpulkan mulai dari *signal analysis*, *data interpretation*, *document planning*, hingga *microplanning & realisation*. Pengembangan sistem ini dilakukan dalam beberapa tahap sesuai dengan metode pengembangan *sekuensial linear*. Tahap pertama yang dilakukan adalah *analysis*, pada tahap ini peneliti menganalisis bagaimana program dibuat. Tahap selanjutnya adalah *design*, pada tahap ini peneliti membuat desain program sehingga user dapat dengan mudah menggunakannya. Tahap ke-tiga *coding*, tahap ini mulai dilakukan implementasi *analysis* dan *design* yang telah dilakukan sebelumnya dan terakhir adalah tahap *testing* program yang telah dibuat.
5. Setelah program dibuat, tahap selanjutnya adalah tahap eksperimen. Pada tahap ini program diuji coba sesuai dengan tujuannya dan dilakukan penilaian berdasarkan aspek *Readability*, *Computation Time*, perbandingan teks dengan grafis, serta perbandingan dengan penelitian terkait.
6. Tahap ini hasil program akan dianalisis dan dilakukan validasi apakah berita sesuai dengan makna data yang digunakan.

## Metode Penelitian

Dalam penelitian ini, dilakukan pengembangan perangkat lunak menggunakan model *Linear Sequential*. *Linear Sequential* mengusulkan sebuah pendekatan kepada pengembangan perangkat lunak yang sistematik dan sekuensial yang dimulai pada tingkat dan kemajuan sistem pada seluruh analis, desain, kode, pengujian, dan pemeliharan. Berikut adalah proses gambaran dari *Linear Sequential Model* gambar 3.2.



Gambar 3.2 Model Linear *Sequential Model* (Pressman, 2001)

1. *System / Informartion engineering* Merupakan bagian dari sebuah sistem terbesar yang mana dalam pengerjaannya dimulai dengan menetapkan berbagai kebutuhan dari semua elemen yang diperlukan sistem dan mengalokasikannya ke dalam pembentukan perangkat lunak.
2. Analisis perangkat lunak merupakan tahap menganalisis hal-hal yang diperlukan dalam pembentukan sebuah perangkat lunak.
3. Desain merupakan beberapa langkah proses yang berfokus pada empat buah atribut yang berbeda dari program, yakni struktur data, arsitektur perangkat lunak, representasi antarmuka, dan sebuah algoritma.
4. *Coding* dilakukan untuk menerjemahkan pembuatan disain ke dalam bentuk yang bisa dimengerti oleh mesin. Sehingga komputer bisa merepresentasikan ke dalam bentuk perangkat lunak.
5. Tes merupakan langkah paling akhir yang dikerjakan, sebuah pengetesan pada perangkat lunak yang sudah melalui beberapa tahap dan dapat dipakai oleh user, dalam tes juga dapat di lakukan pengecekan apakah perangkat lunak yang dibuat sudah sesuai.

## Alat dan Bahan Penelitian

Penelitian ini menggunakan seperangkap laptop yang dilengkapi perangkat lunak pendukung, dengan spesifikasi perangkat keras sebagai berikut:

1. Prosesor Intel® Core™ i5-4200M CPU @ 2.50 GHz
2. Kartu Grafis Intel® HD *Graphics* 4600
3. *Random Access Memory* (RAM) 8 GB
4. *Hard Disk Drive* 500 GB
5. Monitor 14 inci dengan resolusi 1366x768 piksel

Adapun perangkat lunak yang digunakan adalah:

1. *Sistem Operasi Mocrosoft Windows* 10 64-bit
2. *Sublime* 64-bit
3. PHP 5.6.3
4. *RStudio*
5. *R i386 v3.4.0*
6. Xampp v3.2.1
7. Google Chrome 64-bit v67.0.3396.99
8. *Microsoft Excel* 2013

Alat-alat penelitian tersebut digunakan untuk mengembankan aplikasi yang nantinya akan digunakan untuk melakukan penelitian dan eksperimen.



## Data Penelitian

Data yang akan digunakan dalam penelitian ini diperoleh dengan pencarian *file* original yang cukup memadai untuk digunakan, pengambilan *file* yang merupakan berekstensi .csv akan digunakan dalam penelitian ini.

### Data Masukan

Data yang digunakan dalam penelitian ini adalah data historis rupiah terhadap mata uang lain (kurs), *solar radiation and meteorological* Almeria (Spanyol), data *electricity consumption* (*Load*), *wind power generation and solar power generation prices* di Negara Austria dan data klimatologi milik lembaga meteorologi dan klimatologi galicia yaitu MeteoGalicia yang berlokasi di kota Mabegondo, Provinsi A Coruna, daerah komunitas otonom Galicia, Spanyol dalam penelitian sebelumnya, semua data bersifat eksak.

### Data Keluaran

Data keluaran dari eksperimen berupa berita dari data yang dimasukan yang mencakup *summary text, current time description*, dan *predict information*, ringkasan dalam bentuk data statistik, serta model *linear* dari setiap parameter dalam data.

# BAB IV HASIL PENELTIAN DAN PEMBAHASAN

Pada bab ini penulis menguraikan hasil dari implementasi dan uji coba sistem. Bab ini terdiri dari pengumpulan data, pengembangan model, implementasi sistem, hasil implementasi, panduan penggunaan aplikasi, rancangan eksperimen, analisis dan hasil eksperimen.



## Pengumpulan Data

Penelitian menggunakan data historis rupiah terhadap mata uang lain (kurs), *solar radiation and meteorological*, serta data klimatologi dan kualitas udara. Pemilihan data dilakukan berdasarkan sifat dari data yaitu *time series* serta beberapa data dari penelitian sebelumnya, dengan tujuan untuk membandingkan hasil dari sistem yang dikembangkan dengan sistem pada penelitian sebelumnya. Untuk pembahasan lebih detail mengenai data yang digunakan terdapat pada sub-bab selanjutnya.

### Data Historis Rupiah

Data diperoleh dari situs web Bank Indonesia (https://www.bi.go.id/). Data yang digunakan sebanyak 62 baris kurs beli (menukar mata uang asing ke rupiah) terhitung dari tanggal 01 Maret 2018 sampai dengan 31 Mei 2018, kemudian penulis membagi data tersebut menjadi data bulanan (3 *batch*). Untuk mengakses data yang tersedia, penulis mengunjungi situs Bank Indonesia dengan bahasa Indonesia, kemudian pada menu Moneter penulis memilih sub menu dari Informasi Kurs yaitu Kurs Transaksi BI, kemudian penulis memilih mata uang apa saja yang akan ditampilkan dalam data (https://www.bi.go.id/en/moneter/informasi-kurs/transaksi-bi/Default.aspx).

Setelah melakukan serangkaian proses mengumpulkan data, maka data yang digunakan merupakan data nilai tukar mata uang asing terhadap rupiah untuk setiap nilai 1 mata uang asing:

* + - * Dolar Australia (AUD)
      * Yuan Renminbi Cina (CNY)
      * Krone Denmark (DKK)
      * Euro Uni Eropa (EUR)
      * Poundsterling Inggris (GBP)
      * Dolar Hongkong (HKD)
      * Yen Jepang (JPY)
      * Kip Laos (LAK)
      * Ringgit Malaysia (MYR)
      * Peso Filipina (PHP)
      * Baht Thailand (THB)
      * Dolar Amerika (USD)

Data secara lengkap dapat dilihat pada lampiran, sedangkan pada tabel 4.1 merupakan kutipan data Kurs Beli Rupiah dengan parameter DateTime, AUDBuy, CNYBuy, DKKBuy, EURBuy, GBPBuy, HKDBuy, JPYBuy, LAKBuy, MYRBuy, PHPBuy, THBBuy serta USDBuy.

Tabel 4.1 Kutipan data Kurs Beli Rupiah

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| DateTime | AUDBuy | CNYBuy | DKKBuy | EURBuy | ... | USDBuy |
| 01/03/2018 00:00:00 | 10604.53 | 2166.31 | 2247.04 | 16733.67 | ... | 13724 |
| 02/03/2018 00:00:00 | 10606.51 | 2159.5 | 2253.29 | 16784.41 | ... | 13677 |
| 05/03/2018 00:00:00 | 10595.03 | 2155.26 | 2263.07 | 16850.87 | ... | 13671 |
| 06/03/2018 00:00:00 | 10639.71 | 2158.36 | 2268.67 | 16896.04 | ... | 13681 |
| 07/03/2018 00:00:00 | 10675.84 | 2163.55 | 2281.88 | 16998.36 | ... | 13694 |
| ... | ... | ... | ... | ... | ... | ... |
| 24/05/2018 00:00:00 | 10688.13 | 2214.81 | 2221.17 | 16543.85 | ... | 14134 |
| 25/05/2018 00:00:00 | 10664.28 | 2206.93 | 2214.73 | 16498.2 | ... | 14095 |
| 28/05/2018 00:00:00 | 10595.61 | 2188.02 | 2199.02 | 16378.35 | ... | 13995 |
| 30/05/2018 00:00:00 | 10451.95 | 2174.53 | 2161.77 | 16085.62 | ... | 13962 |
| 31/05/2018 00:00:00 | 10492.65 | 2164.04 | 2174.27 | 16181.08 | ... | 13881 |

### *Solar Radiation and Meteorological Data*

Data mengenai radiasi sinar surya dan meteorolgi dari Plataforma Solar de Almeria (PSA), Spanyol berdasarkan data satelit Meteosat MSG dan MFG ((c) 2017 EUMETSAT) dan dari data atmosfer ((c) 2017 ECMWF dan NOAA) dengan metode Solargis. Data sebenarnya terdapat lebih dari 201600 baris yang dimulai dari tahun 1994 hingga tahun 2016, namun data yang digunakan merupakan sebagian kecil dari data harian, terhitung enam bulan dari 1 Juli 2016 sampai 31 Desember 2016 sebanyak 184 baris data yang dibagi kedalam 6 *batch*. mulai dari tahun Untuk mengakses data yang tersedia, penulis mengunjungi situs web https://solargis.com, memilih menu *Products,* konten *Time series and TMY data* kemudian sub menu *Usefull Resource* (https://solargis.com/products/time-series-and-tmy-data/useful-resources/).

Data yang disediakan oleh Solargis terdiri dari data berikut:

* + - * Tanggal Data.
      * *Global Horizontal Irradiance* (GHI) yaitu jumlah radiasi langsung (*direct*) dan tersebar (*diffuse*) yang diterima pada bidang horizontal, dengan satuan (Watt per meter persegi).
      * *Direct Normal Irradiance* (DNI) yaitu komponen yang terlibat dalam termal (*concentrating solar power*, CSP) dan teknologi konsentrasi fotovoltaik (*concentrated photovoltaic*, CPV), dengan satuan (Watt per meter persegi).
      * *Diffuse Horizontal Irradiance* (DIF) yaitu jumlah radiasi matahari yang diterima per satuan luas permukaan normal tidak secara langsung terkena sinar matahari tetapi sinar telah tersebar oleh molekul dan partikel di atmosfer, dengan satuan (Watt per meter persegi).
      * *Global Titled Irradiance* (GTI) yaitu radiasi total yang diterima oleh permukaan (dengan kemiringan tertentu) dan azimut secara tetap atau ssuai pergerakan matahari, dengan satuan (Watt per meter persegi).
      * Suhu udara pada jarak 2 meter (TEMP), dengan satuan oC (*Celcius*).
      * Kelembaban relatif (RH) (persentase).
      * Kecepatan angin pada jarak 10 meter (WS) dengan satuan meter per detik ().
      * Kandungan air (PWAT) dengan satuan kilogram per meter persegi ().

Berdasarkan hasil akuisisi dan analisis, penulis menggunakan semua parameter yang ada, penulis mengelompokan data menjadi data bulanan (*batch*), kutipan data terdapat pada tabel 4.2. Data tersebut terdiri dari parameter DateTime, GHI,DNI, DIF, GTI, Temperature, AP, RH, WindSpeed dan PWAT, Sedangkan untuk data lengkap terdapat pada lampiran.

Tabel 4.2 Kutipan data radiasi sinar matahari dan meteorologi interval waktu harian

| DateTime | GHI | DNI | DIF | GTI | Temperature | AP | RH | WindSpeed | PWAT |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 01/07/2016 00:00:00 | 7.84 | 6.285 | 3.055 | 7.32 | 24.9 | 960.9 | 57.2 | 1.7 | 27.3 |
| 02/07/2016 00:00:00 | 7.978 | 6.619 | 3.033 | 7.426 | 26.2 | 961.1 | 48.1 | 2.2 | 27.9 |
| 03/07/2016 00:00:00 | 8.252 | 7.481 | 2.725 | 7.666 | 27.2 | 960.9 | 32.4 | 4 | 18.6 |
| 04/07/2016 00:00:00 | 8.248 | 7.966 | 2.406 | 7.656 | 24.6 | 960.5 | 51.5 | 2.8 | 24.4 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 01/08/2016 00:00:00 | 7.595 | 6.935 | 2.594 | 7.496 | 27.6 | 961.6 | 35.1 | 2.8 | 18.7 |
| 02/08/2016 00:00:00 | 7.523 | 6.842 | 2.62 | 7.442 | 25.8 | 962.9 | 52.1 | 2.2 | 22.7 |
| 03/08/2016 00:00:00 | 7.586 | 7.213 | 2.447 | 7.529 | 26.6 | 960.3 | 50.6 | 2 | 23.7 |
| 04/08/2016 00:00:00 | 8.006 | 9.253 | 1.718 | 7.91 | 29 | 958 | 33.5 | 1.9 | 15.4 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 01/09/2016 00:00:00 | 6.613 | 6.85 | 2.126 | 7.336 | 23.7 | 964.8 | 62.3 | 2.8 | 27.2 |
| 02/09/2016 00:00:00 | 6.462 | 6.425 | 2.199 | 7.179 | 24.4 | 964.8 | 51.7 | 2.1 | 26.2 |
| 03/09/2016 00:00:00 | 5.926 | 5.014 | 2.725 | 6.477 | 25.5 | 964.3 | 46.3 | 1.9 | 27.4 |
| 04/09/2016 00:00:00 | 6.471 | 6.457 | 2.214 | 7.261 | 27.9 | 964.6 | 34.3 | 1.9 | 26.7 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 01/10/2016 00:00:00 | 5.462 | 7.262 | 1.234 | 7.135 | 19.5 | 958.7 | 64.5 | 1.6 | 20.4 |
| 02/10/2016 00:00:00 | 5.103 | 6.142 | 1.498 | 6.586 | 19.9 | 959.5 | 63.1 | 1.3 | 19 |
| 03/10/2016 00:00:00 | 5.419 | 7.607 | 1.225 | 7.187 | 19.9 | 963.3 | 60.6 | 2.2 | 17 |
| 04/10/2016 00:00:00 | 5.302 | 7.195 | 1.305 | 7.039 | 20.9 | 963.5 | 59.6 | 1.8 | 21.1 |
| 05/10/2016 00:00:00 | 5.081 | 6.225 | 1.658 | 6.663 | 20.4 | 961.5 | 67.7 | 2.1 | 22 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 01/11/2016 00:00:00 | 3.752 | 5.181 | 1.27 | 5.606 | 15.8 | 966.4 | 73.8 | 1.8 | 17.4 |
| 02/11/2016 00:00:00 | 2.961 | 2.896 | 1.57 | 4.061 | 19 | 964.3 | 64.2 | 1.5 | 21.2 |
| 03/11/2016 00:00:00 | 3.404 | 3.6 | 1.66 | 4.801 | 19.7 | 962 | 48.3 | 1.8 | 17.1 |
| 04/11/2016 00:00:00 | 1.722 | 0.351 | 1.582 | 1.834 | 18.7 | 960.5 | 52.1 | 1.5 | 23.3 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 01/12/2016 00:00:00 | 2.566 | 3.513 | 1.17 | 4.113 | 11.4 | 967.1 | 92.1 | 3 | 21.2 |
| 02/12/2016 00:00:00 | 3.131 | 6.56 | 0.653 | 5.751 | 11.2 | 962.1 | 88.4 | 0.9 | 15.6 |
| 03/12/2016 00:00:00 | 1.567 | 1.004 | 1.268 | 2.009 | 11.5 | 956.5 | 90.6 | 1.8 | 21.8 |
| 04/12/2016 00:00:00 | 0.567 | 0.034 | 0.558 | 0.537 | 12.1 | 954.5 | 91.7 | 2.4 | 23.3 |

### Data Klimatologi dan Kualitas Udara

Data ini berasal dari stasiun pemantauan milik lembaga meteorologi dan klimatologi galicia yaitu MeteoGalicia yang berlokasi di kota Mabegondo, Provinsi A Coruna, daerah komunitas otonom Galicia, Spanyol. Data yang digunakan pada penelitian sebelumnya yaitu DWP yang telah dibahas pada bab 2 sebelumnya. Data yang digunakan merupakan data harian dalam jangka waktu satu tahun sebanyak 3 *batch* (07 Januari 2014-06 Januari 2015, 07 Juni 2015-25 Mei 2016, dan 06 Juli 2016-06 Juli 2017), sebanyak masing masing 365 baris data, dengan parameter:

Waktu data

Curah hujan rata-rata per hari dengan satuan L/m2 (Liter per meter kuadrat)

Cakupan awan rata-rata per hari dengan satuan % (Persen)

Rata-rata suhu rata-rata per hari dengan satuan oC (derajat celcius)

Kecepatan angin rata-rata per hari dengan satuan km/h (Kilometer per jam)

Arah angin rata-rata per hari dengan satuan o (derajat)

Data Karbon Monoksida (CO) per hari dengan satuan ppm (Part per Million)

Data Nitrogen Monoksida (NO) per hari dengan satuan ppm (Part per Million)

Data Nitrogen Dioksida (NO2) per hari dengan satuan ppm (Part per Million

Data Nitrogen Oksida (NOX) per hari dengan satuan ppm (Part per Million)

Data Ozone (O3) per hari dengan satuan ppm (Part per Million)

Data Particulate Matter 10 mikronmeter per hari

Data Particulate Matter 25 mikronmeter per hari

Data Sulfur Dioksida (SO2) per hari dengan satuan ppm (Part per Million)

Parameter tersebut merupakan parameter hasil analisis pada penelitian DWP (Putra et al, 2017). Kutipan data terdapat pada tabel 4.3 dengan parameter DateTime, CloudCoverage, Temperature, WindSpeed, WindDirection dan Rainfall, data secara lengkap terdapat pada lampiran.

Tabel 4.3 Kutipan data klimatologi

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| DateTime | CloudCoverage | Temperature | WindSpeed | ... | PM25 | SO2 |
| 07/01/2014 00:00:00 | 89.8 | 16.3 | 7.74 | ... | 13.25 | 4.25 |
| 08/01/2014 00:00:00 | 49.5 | 17.8 | 6.59 | ... | 10.25 | 4.25 |
| 09/01/2014 00:00:00 | 19.5 | 18.4 | 7.2 | ... | 11.25 | 4.25 |
| 10/01/2014 00:00:00 | 61.4 | 19.8 | 6.73 | ... | 10.25 | 4.25 |
| 11/01/2014 00:00:00 | 78.4 | 19.2 | 14.69 | ... | 14.25 | 4.25 |
| ... | ... | ... | ... | ... | ... | ... |
| 01/01/2015 00:00:00 | 19.2 | 17.5 | 5.62 | ... | 11.25 | 7.25 |
| 02/01/2015 00:00:00 | 24.2 | 19.1 | 6.88 | ... | 8.25 | 7.25 |
| 03/01/2015 00:00:00 | 35.7 | 19.6 | 6.91 | ... | 9.25 | 9.25 |
| 04/01/2015 00:00:00 | 65.3 | 19.1 | 6.48 | ... | 17.25 | 8.25 |
| 05/01/2015 00:00:00 | 23.6 | 22.5 | 4.61 | ... | 13.25 | 7.25 |
| ... | ... | ... | ... | ... | ... | ... |
| 01/01/2016 00:00 | 83.4 | 11.3 | 7.7 | ... | 41 | 25 |
| 02/01/2016 00:00 | 100 | 14.2 | 12.28 | ... | 38 | 25 |
| 03/01/2016 00:00 | 98.4 | 13.3 | 15.48 | ... | 41 | 25 |
| 04/01/2016 00:00 | 88.5 | 7.9 | 17.75 | ... | 39 | 26 |
| 05/01/2016 00:00 | 38.9 | 7.7 | 9.65 | ... | 41 | 26 |
| ... | ... | ... | ... | ... | ... | ... |
| 01/07/2017 00:00:00 | 45.6 | 15.4 | 6.01 | ... | 27 | 8 |
| 02/07/2017 00:00:00 | 12.6 | 18.9 | 7.34 | ... | 20 | 7 |
| 03/07/2017 00:00:00 | 13.3 | 21.8 | 4.86 | ... | 14 | 9 |
| 04/07/2017 00:00:00 | 18.7 | 24 | 6.59 | ... | 14 | 9 |
| 05/07/2017 00:00:00 | 81.1 | 19.2 | 8.35 | ... | 12 | 7 |
| 06/07/2017 00:00:00 | 58.3 | 17.9 | 6.48 | ... | 12 | 7 |

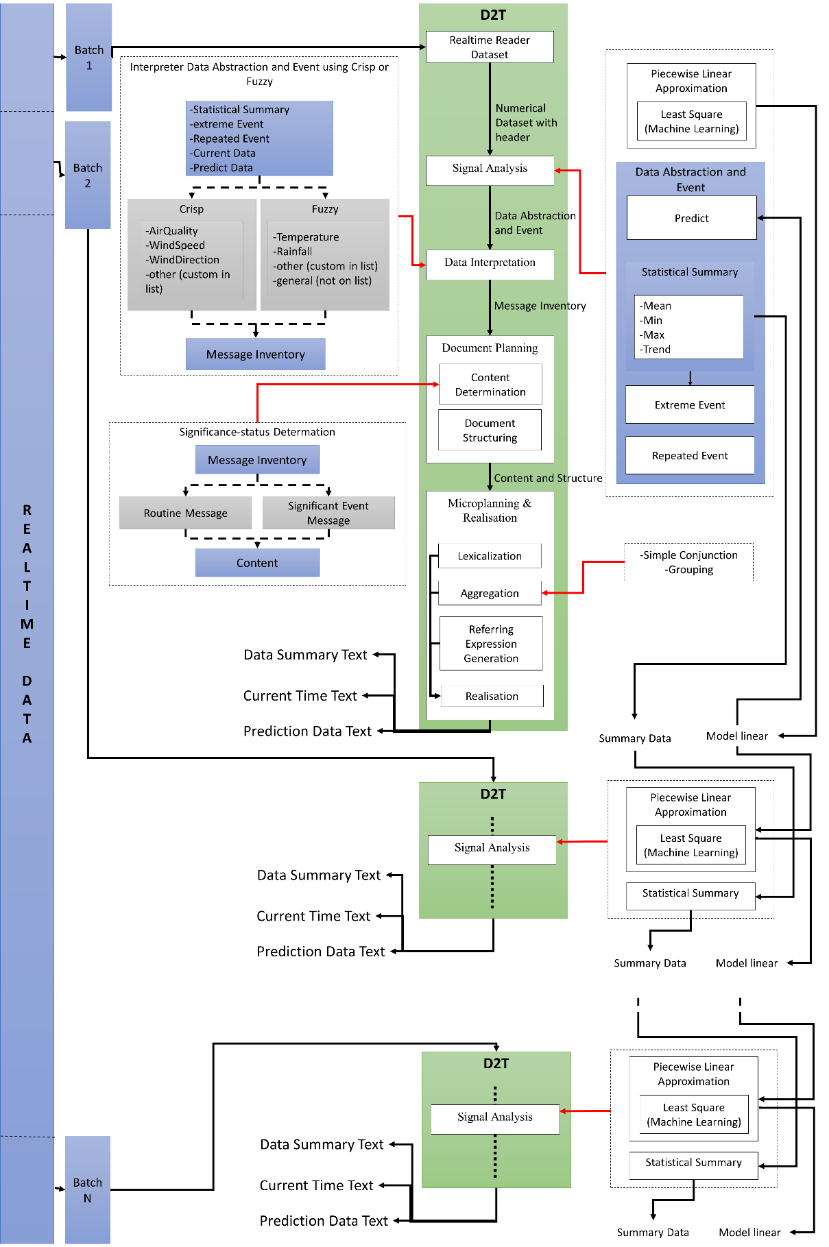
## Pengembangan Model *Data-to-text Streaming* dengan *Machine Learning*

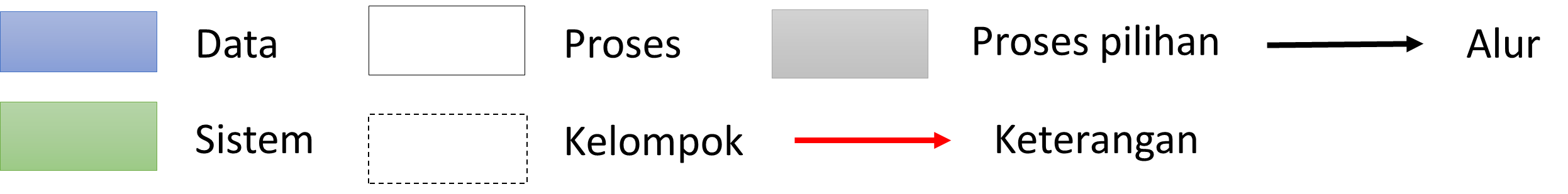
Model sistem *Data-to-Text* yang dibangun pada penelitian ini mengacu pada model *Data-to-Text* yang diusulkan oleh Reiter (2007), dengan melanjutkan pengembangan sistem DWP yang dilakukan oleh Putra (2017) menjadikan model *Realtime* serta menggunakan *Machine Learning* pada proses *Signal Analysis* sehingga setiap data *batch* yang di proses memengaruhi data yang akan diproses selanjutnya meskipun data asli sudah tidak ada.

Pada penelitian ini berita yang dibangkitkan terdiri dari tiga bagian, bagian pertama mengenai deskripsi hasil ringkasan data selama periode *t* (bisa bulanan, harian, tahunan, atau periode waktu lainnya), bagian ke-dua deskripsi data ke *n* pada setiap *batch*, dan ke-tiga informasi hasil prediksi untuk data *n+1*.

Model penelitian pada gambar 4.1 memperilahtkan bahwa terdapat perbedaan pada data *input* yang dilakukan oleh Realtime Reader Dataset, yang secara terus menerus melakukan pengecekan adanya dataset, bila terdapat dataset maka mulai membaca dataset dan lakukan *Signal Analysis* setelah membaca data, fungsi Realtime Reader Dataset ini tidak berhenti bekerja, perbedaan lainnya terdapat pada *Sygnal Analysis* terdapat dua output baru yakni *Model Linear* dari *batch* dataset dan Statistical Data. Pembahasan mengenai model tersebut akan dipaparkan pada sub-bab berikutnya.

Pada gambar 4.1 digambarkan bahwa pertama kali data muncul adalah *batch-1*, yang kemudian dibaca oleh *Realtime Reader Dataset*, kemudian dilanjutkan dengan proses *Signal Analysis* dengan *input* data numerik dan output *Data Abstraction and Event*, pada *Signal Analysis* terdapat beberapa proses untuk mengumpulkan *Data Abstraction and Event*, didalamnya terdapat beberapa proses yang tidak berurutan, namun ada juga yang berurutan, dimana *Extreme Event* bisa dideteksi setelah *Statistical Summary* didapat. Pada proses berikutnya *Data Abstraction and Event* digunakan sebagai *input* dari proses *Data Interpretation*, dimana terdapat pilihan cara untuk menginterpretasi data, yaitu dengan *Crisp* atau *Fuzzy*, proses dilanjutkan hingga mendapat teks. Sedangkan satu rangkaian utuh proses *Data-to-Text* secara utuh dikelompokkan dalam kotak berwarna hijau.





Gambar 4.1 Model *Data-to-text* pada *Data Streaming* dengan pendekatan *Machine Learning*

### Model Proses *Realtime Reader Dataset*

Proses pengecekan adanya dataset untuk dilakukan serangkaian proses D2T ini sebenarnya terpisah dari pemrograman bahasa R, penulis menggunakan bahasa pemrograman Javascript, AJAX dan html untuk pengecekan dataset, bila terdapat dataset makan fungsi ini akan memanggil controller pada CodeIgniter (PHP) untuk menjalankan sistem D2T yang dibangun dengan menggunakan bahasa pemrograman R, fungsi ini akan terus berjalan selama aplikasi yang dibangun dijalankan.

Tabel 4.4 Contoh kasus dataset *batch* pertama

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| DateTime | CloudCoverage | Temperature | CO | SO2 |
| 07/01/2014 00:00:00 | 89.8 | 16.3 | 2 | 4.25 |
| 08/01/2014 00:00:00 | 49.5 | 17.8 | 4.8 | 4.25 |
| 09/01/2014 00:00:00 | 19.5 | 18.4 | 1 | 4.25 |
| 10/01/2014 00:00:00 | 61.4 | 19.8 | 0.3 | 4.25 |
| 11/01/2014 00:00:00 | 78.4 | 19.2 | 4.4 | 4.25 |
| 12/01/2014 00:00:00 | 42.9 | 16.9 | 5 | 4.25 |
| 13/01/2014 00:00:00 | 60.3 | 17 | 4.7 | 5.25 |
| 14/01/2014 00:00:00 | 73 | 17.1 | 3.4 | 5.25 |
| 15/01/2014 00:00:00 | 16.3 | 16.9 | 2 | 5.25 |

Misalkan terdapat dataset seperti pada tabel 4.4 untuk *batch* pertama kemudian data masuk sebagai file csv dengan nama Dataset kedalam *folder* DatasetsRealTime, maka proses *Realtime Reader Dataset* akan mengirimkan sinyal untuk menjalankan D2T dalam bahasa R dan kemudian dataset dibaca dan melakukan proses *Signal Analysis*.

Tabel 4.5 Contoh kasus dataset *batch* ke-dua

| DateTime | CloudCoverage | Temperature | CO | SO2 |
| --- | --- | --- | --- | --- |
| 16/01/2014 00:00:00 | 8.4 | 17.5 | 0.2 | 5.25 |
| 17/01/2014 00:00:00 | 32.7 | 17.9 | 1.9 | 5.25 |
| 18/01/2014 00:00:00 | 13.4 | 18.1 | 3.7 | 5.25 |
| 19/01/2014 00:00:00 | 79.8 | 18.4 | 0.9 | 4.25 |
| 20/01/2014 00:00:00 | 36.9 | 19.6 | 0.1 | 4.25 |
| 21/01/2014 00:00:00 | 25.1 | 21 | 2.9 | 5.25 |
| 22/01/2014 00:00:00 | 13 | 21 | 2.7 | 5.25 |
| 23/01/2014 00:00:00 | 73.7 | 19 | 3.7 | 5.25 |
| 24/01/2014 00:00:00 | 61.7 | 17.4 | 1.1 | 4.25 |
| 25/01/2014 00:00:00 | 75.4 | 18 | 4.3 | 6.25 |

Begitu juga dengan dataset *batch* ke-dua berupa *file* csv dengan nama Dataset yang masuk kedalam *folder* DatasetsRealTime, yang kemudian pada model ini dikirim sinyal untuk menjalankan sistem D2T yang telah dibangun, untuk membaca dataset dan melakukan proses *Signal Analysis*.

### Model Proses *Signal Analysis*

Proses *Signal Analysis* merupakan tahapan awal sistem D2T, dimana sistem melakukan analisis untuk setiap dataset yang diberikan, mulai dari meringkas dataset selama periode *t* menjadi *Statistical Summary* seperti yang dilakukan oleh Potret pada pengembangan sistem *BABYTALK Family System* yang meringkas kejadian selama 45 menit (Portet et al., 2009) dan pada sistem DWP yang meringkas keadaan klimatologis dan kualitas udara selama satu bulan (Putra et al, 2017). Pada tahap ini juga sistem melakukan analisis untuk data ke-*n* (baris terakhir dari dataset), *Extreme Event*, *Repeated Event,* dan prediksi untuk *n+1* data menggunakan model *linear* yang telah didapat dari serangkaian proses *Piecewise Linear Approximation*. Semua data hasil tersebut digabungkan kedalam *Data Abstraction and Event*. Sehingga *input*, proses, dan output pada tahap ini adalah sebagai berikut:

* *Input*: *Numerical Dataset with header*
* Proses: *Signal Analysis*
* *Output*: *Data Abstraction and Event*

Sub-proses pada tahap ini akan dibahas pada sub-bab berikutnya.

* + - 1. **Proses *Signal Analysis* Untuk Prediksi Data**

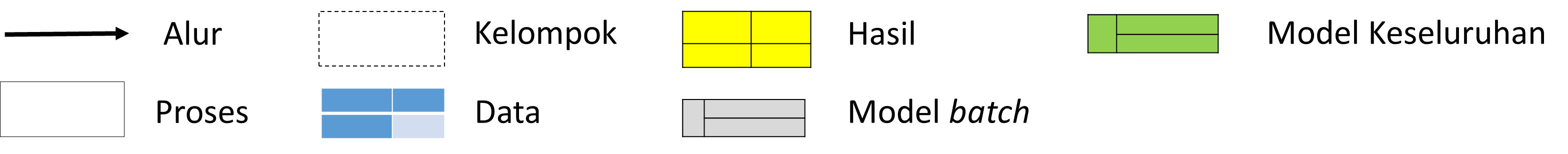
Prediksi data dilakukan dengan teknik *Piecewise Linear Approximation* (PLA) dengan metode *Least Square* (Palpanas et al., 2004). Pada proses ini dihasilkan model *linear* dari setiap *batch* data dan model *linear* untuk keseluruhan data (*batch* pertama hingga *batch* ke-*N*), hal ini dilakukan karena data yang digunakan merupakan data *Time Series* sehingga antara *batch* satu dan lainnya memiliki keterikatan. Penerapan PLA dilakukan untuk merepresentasikan data *streaming* lampau dan menggabungkannya menjadi sebuah garis *linear*, mengingat data yang digunakan hanya dalam satu kali proses dan selanjutnya data terhapus untuk mengatasi masalah *storage* berlebihan dikarenakan kemunculan data sangat cepat dan memiliki ukuran yang tidak sedikit bila kita menggabungkan setiap data *stream*.

Ilustrasi dari tahapan ini terdapat pada gambar 4.2 dimana pada *batch* pertama hanya menghasilkan satu model *linear* () didapat dari penggunaan metode *Least Square* dan model tersebut digunakan sebagai model keseluruhan data yang ada untuk memprediksi data *n+1*, sedangkan untuk iterasi selanjutnya model dihasilkan satu model *linear* (), model tersebut hanya merepresentasikan data *batch* ke-*N*. Untuk mendapatkan model yang merepresentasikan keseluruhan *batch* maka dilakukan penggabungan antara dua model *linear* menjadi satu model *linear*, dengan cara membangkitkan sebanyak *n* data dengan model dan kemudian menggunakan metode *Least Square* untuk mendapatkan satu model *linear* yang merepresentasikan keseluruhan data .

Dengan menggunakan dataset pada tabel 4.4 dan dataset pada tabel 4.5 maka didapatkan hasil seperti pada gambar 4.2, dimana data pertama berwarna biru pada bagian kiri atas gambar (kotak berwarna biru) menjadi masukan pada proses *Signal Analysis* untuk prediksi data, kemudian data tersebut diolah menjadi sebuah model *linear* (kotak berwarna abu) dan menggabungkan model *linear* dengan model *linear* keseluruhan data sebelumnya, karena tidak terdapat model *linear* sebelumnya pada *batch* pertama maka model *linear* yang dibuat digunakan juga sebagai model keseluruhan data (kotak berwarna hijau). Pada pemrosesan data *batch* ke-dua maka selain dataset *batch* ke-dua yang menjadi masukan sistem, ada juga masukan berupa model *linear* keseluruhan data sebelumnya, yang kemudian dilakukan proses penggabungan model *linear* dengan teknik PLA (kotak hijau paling bawah), sedangkan *output* prediksi terlihat pada kotak berwarna kuning. Penerapan PLA dengan *Least Square Method* termasuk dalam proses *Signal Analysis* ditandai dengan kotak dengan garis putus-putus.

**

Gambar 4.2Model proses *Signal Analysis* prediksi dengan PLA dan *Least Square*

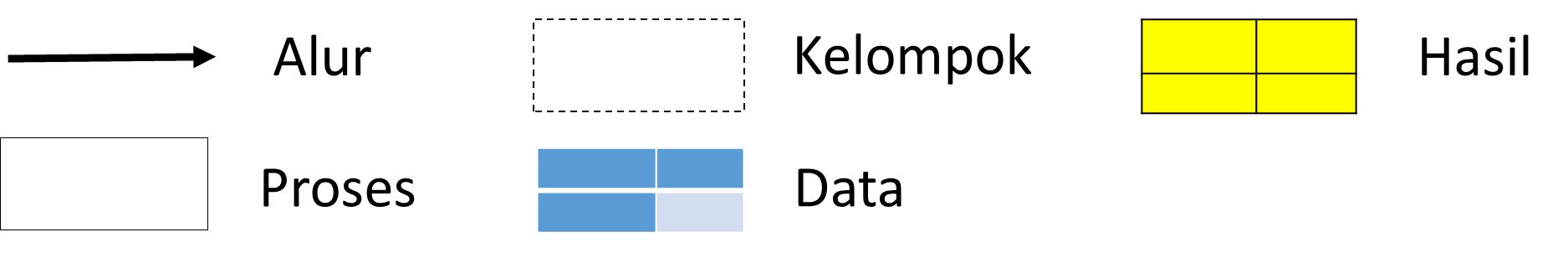
****

* + - 1. **Proses *Signal Analysis* Untuk Ringkasan Data**

Ringkasan data ini berfungsi untuk menentukan pesan yang akan disampaikan pada *Summary Text*, seperti parameter pada data dan trend data. Pada penelitian DWP setidaknya ada 3 point utama yang di dapat dalam proses ini yaitu sinyal pesan bulanan dari setiap parameter (*Monthly\_Message*), *repeat event* untuk parameter *Rainfall* serta *extreme event* untuk parameter *Rainfall, WindSpeed* dan *Temperature* (Putra et al, 2017). Pada penelitian ini deteksi sinyal *extreme event* dilakukan pada proses yang berbeda, ringkasan ini juga memengaruhi *Extreme Event* yang akan dideteksi, karena untuk menentukan suatu *event* bersifat *extreme* memerlukan *range* antara nilai maksimal dan minimal data. Pada Gambar 4.3 ringkasan data setiap *batch* akan dipengaruhi oleh ringkasan ringkasan sebelumnya. Hal ini dilakukan agar meningkatkan kemampuan sistem untuk mengiterpretasi data pada proses selanjutnya, mengingat adanya kemungkinan rentang pada setiap *batch* bukanlah rentang aslinya, terkecuali untuk parameter interpretasi data yang sudah disediakan oleh sistem atau oleh user serta untuk melakukan perbandingan kondisi pada *batch* selanjutnya.

Sebagai contoh sederhana, jika pada suatu *batch* untuk parameter Temperature pada dataset tabel 4.4 nilai minimum saat ini adalah 16.3, nilai maksimum saat ini adalah 19.8, maka untuk saat ini interpretasi data tersebut dibagi kedalam 5 bagian *Fuzzy* (bagian ini akan dibahas pada sub-bab berikutnya) hanya dengan rentang 3.5 *point* saja, sedangkan pada dataset tabel 4.5 nilai maksimum Temperature ada pada titik 21, sehingga rentang asli data tersebut berubah menjadi 4.7 *point*. Maka dengan adanya pengaruh antara ringkasan, nilai minimum dan maksimum data akan terus berkembang seiring banyaknya data yang di proses dan tingkat interpretasi data akan lebih tepat.

Data statistik yang dicari pada proses ini diantaranya data minimal (nilai, tanggal, index data), data maksimal (nilai, tanggal, index data), nilai rata-rata serta representasi *trend* untuk setiap parameter yang ada, sehingga jika ada *n* parameter, maka sinyal yang didapat sebanyak *n\*6.*



Gambar 4.3 Model proses *Signal Analysis statistical summary*

Dengan menggunakan dataset pada tabel 4.4 dan 4.5 maka didapat hasil ringkas data seperti pada gambar 4.3, dimana pada *batch* pertama *input* proses hanya berdasarkan dataset, sedangkang pada *batch* ke-dua *input* proses terdiri dari dataset *batch* ke-dua dan *Summary Data* sebelumnya.

* + - 1. **Proses *Signal Analysis* Untuk *Extreme Event***

Pada proses ini dilakukan pencarian nilai kenaikan dan penurunan tertinggi untuk setiap parameter, untuk menghitung kenaikan atau penurunan tersebut termasuk pada kategori *extreme* diperlukan *range* nilai dari setiap parameter (*max-min*). Ketika nilai kenaikan atau penurunan melebihi 60% maka kenaikan atau penurunan tersebut termasuk dalam kategori *significance*, sedangkan jika kenaikan atau penurunan lebih dari 75% dari nilai *range*, maka kenaikan atau penurunan tersebut termasuk dalam kategori *extreme*. Contohnya ketika cuaca sedang sangat panas tiba tiba sehari kemudian berubah menjadi sangat dingin, itulah *extreme* *event* yang dimaksudkan.

Berikut beberapa contoh kasus pada penentuan *extreme event*:

|  |
| --- |
| Bila pada parameter X hasil ringkasan yang didapat adalah nilai minimum 0 dan maksimal 100, kenaikan tertinggi terdapat pada titik *t* sebesar 50 *point*, sedangkan penurunan tertinggi terdapat pada titik *t+10* sebesar 30 *point*, maka pada tahap ini sinyal yang di hasilkan adalah “0”, artinya tidak ada *extreme event* pada parameter ini, karena kriteria *extreme event* membutuhkan lebih dari 60 *point*.  Bila pada parameter X hasil ringkasan yang didapat adalah nilai minimum 0 dan maksimal 1, kenaikan tertinggi terdapat pada titik *t* sebesar 0.3 *point*, sedangkan penurunan tertinggi terdapat pada titik *t-2* sebesar 0.63 *point*, maka pada tahap ini sinyal yang di hasilkan adalah “-”, artinya ada *extreme event* pada parameter ini yaitu penurunan yang signifikan, karena kriteria *extreme event* lebih dari 0.6 *point*.  Bila pada parameter X hasil ringkasan yang didapat adalah nilai minimum 10 dan maksimal 20, kenaikan tertinggi terdapat pada titik *t* sebesar 2 *point*, sedangkan penurunan tertinggi terdapat pada titik *t+2* sebesar 10 *point*, maka pada tahap ini sinyal yang di hasilkan adalah “--”, artinya ada *extreme event* pada parameter ini yaitu penurunan yang sangat drastis dari nilai tertinggi menjadi nilai terendah, karena nilai penurunan 100% dari nilai range data.  Bila pada parameter X hasil ringkasan yang didapat adalah nilai minimum 0 dan maksimal 30, kenaikan tertinggi terdapat pada titik *t* sebesar 20 *point*, sedangkan penurunan tertinggi terdapat pada titik *t+2* sebesar 10 *point*, maka pada tahap ini sinyal yang di hasilkan adalah “+”, artinya ada *extreme event* pada parameter ini yaitu kenaikan yang signifikan, karena nilai kenaikan melebihi 60% dari nilai range data.  Bila pada parameter X hasil ringkasan yang didapat adalah nilai minimum 0 dan maksimal 1, kenaikan tertinggi terdapat pada titik *t* sebesar 1 *point*, sedangkan penurunan tertinggi terdapat pada titik *t-2* sebesar 0.5 *point*, maka pada tahap ini sinyal yang di hasilkan adalah “++”, artinya ada *extreme event* pada parameter ini yaitu kenaikan yang sangat drastis, karena nilai kenaikan sama dengan nilai range data.  Bila pada parameter X hasil ringkasan yang didapat adalah nilai minimum 0 dan maksimal 2, kenaikan tertinggi terdapat pada titik *t* sebesar 1.3 *point*, sedangkan penurunan tertinggi terdapat pada titik *t-2* sebesar 1.3 *point*, maka pada tahap ini sinyal yang di hasilkan adalah “+-”, artinya ada *extreme event fluctuate* pada parameter ini yaitu kenaikan dan penurunan yang signifikan, karena nilai kenaikan dan penurunan sama dengan nilai range data.  Bila pada parameter X hasil ringkasan yang didapat adalah nilai minimum 0 dan maksimal 2, kenaikan tertinggi terdapat pada titik *t* sebesar 2 *point*, sedangkan penurunan tertinggi terdapat pada titik *t-2* sebesar 2 *point*, maka pada tahap ini sinyal yang di hasilkan adalah “++--”, artinya ada *extreme event fluctuate* pada parameter ini yaitu kenaikan dan penurunan yang sangat drastis, karena nilai kenaikan dan penurunan sama dengan nilai range data. |

Dengan menggunakan data pada tabel 4.4 dan dilakukan proses deteksi *Extreme Event* didapatkan hasil pada tabel 4.6, dimana pada parameter CloudCoverage terdeteksi penurunan yang sangat drastis pada data ke-delapan, dengan nilai penurunan sebesar 56.7 *point*, begitu pun pada parameter lainnya dengan perbedaan sinyal yang didapat.

Tabel 4.6 Hasil *Extreme Event* contoh kasus *batch* pertama

| ColName | extremeInc | indexExInc | extremeDec | indexExDec | summaryExtreme |
| --- | --- | --- | --- | --- | --- |
| CloudCoverage | 41.9 | 3 | 56.7 | 8 | -- |
| Temperature | 1.5 | 1 | 2.3 | 5 | - |
| CO | 4.1 | 4 | 3.8 | 2 | ++-- |
| SO2 | 1 | 6 | 0 | 1 | ++ |

Sedangkan dengan menggunakan data *batch* ke-dua pada tabel 4.5 didapatkan hasil pada tabel 4.7, dimana pada parameter Temperature tidak terdapat *Extreme Event,* sedangkan pada parameter lainnya terjadi kenaikan yang signifikan dan sangat drastis karena nilai minimum dan maksimum yang digunakan dalam *Summary Data* hanya berdasarkan tabel 4.4 dan tabel 4.5 saja.

Tabel 4.7 Hasil *Extreme Event* contoh kasus *batch* ke-dua

| ColName | extremeInc | indexExInc | extremeDec | indexExDec | summaryExtreme |
| --- | --- | --- | --- | --- | --- |
| CloudCoverage | 66.4 | 3 | 42.9 | 4 | ++ |
| Temperature | 1.4 | 5 | 2 | 7 | 0 |
| CO | 3.2 | 9 | 2.8 | 3 | + |
| SO2 | 2 | 9 | 1 | 3 | ++ |

* + - 1. **Proses *Signal Analysis* Untuk *Repeated Event***

Proses ini sama seperti mencari sinyal *extreme event* namun penelusuran dilakukan untuk mencari parameter dengan nilai sama berturut turut lebih dari 10% jumlah data. Berikut beberapa contoh kasus pada penentuan *repeated event*:

|  |
| --- |
| * + 1. Pada parameter X dalam data Y dengan jumlah 30 baris data (bulanan), terdapat nilai yang sama secara berturut-turut pada baris 5 sampai baris 10 (5 baris), maka sinyal tersebut merupakan sinyal *repeated event* karena 5 baris lebih besar dari 10% \* 30 baris.     2. Pada parameter X dalam data Z dengan jumlah 366 baris data (tahunan), terdapat 20 baris data dengan nilai yang sama secara berturut-turut, karena 20 baris tidak lebih besar dari 10% \* 366 baris maka sinyal tersebut bukan *repeated event*. |

Dengan menggunakan data pada tabel 4.4 dan dilakukan proses deteksi *Repeated Event* didapatkan hasil pada tabel 4.8, dimana pada parameter SO2 terdeteksi adanya *Repeated Event* sebanyak 6 data dimulai dari data dengan indeks pertama, dengan nilai berulang 4.25.

Tabel 4.8 Hasil *Repeated Event* contoh kasus *batch* pertama

|  |  |  |  |
| --- | --- | --- | --- |
| ColName | value | index | numRepeat |
| CloudCoverage | 0 | 0 | 0 |
| Temperature | 0 | 0 | 0 |
| CO | 0 | 0 | 0 |
| SO2 | 4.25 | 1 | 6 |

Sedangkan dengan menggunakan data pada tabel 4.5 dan dilakukan proses deteksi *Repeated Event* didapatkan hasil pada tabel 4.9, dimana pada parameter Temperature dan SO2 terdeteksi adanya *Repeated Event*.

Tabel 4.9 Hasil *Repeated Event* contoh kasus *batch* ke-dua

|  |  |  |  |
| --- | --- | --- | --- |
| ColName | value | index | numRepeat |
| CloudCoverage | 0 | 0 | 0 |
| Temperature | 21 | 6 | 2 |
| CO | 0 | 0 | 0 |
| SO2 | 5.25 | 1 | 3 |

### Model Proses *Data Interpretation*

Pada tahap ini dilakukan penerjemahan sinyal-sinyal yang didapat kedalam pesan-pesan untuk merepresentasikan sinyal yang didapat kedalam bahasa manusia dengan bentuk kata atau frasa (Reiter, 2011). Proses penerjemahan ini menggunakan *Fuzzy* atau *Rule-based* seperti yang dilakukan pada penelitian DWP sebelumnya (Putra et al, 2017). Pada proses interpretasi data dengan metode *Fuzzy* dilakukan dengan mengacu pada *Fuzzy membership function* setiap parameter, sedangkan pada proses metode *Rule-based* mengacu pada *crisp* *membership function* yang ditentukan oleh *expert* untuk parameter terkait. Semua pesan tersebut akan menjadi *input* dalam proses berikutnya, yaitu *document planning*. Sehingga *input*, proses, dan output pada tahap ini adalah sebagai berikut:

* *Input*: *Data Abstraction and Event*
* Proses: *Data Interpretation*
* *Output*: *Message Inventory*

Beberapa acuan *corpus* interpretasi data pada sistem yang dibangun adalah sebagai berikut:

* + - 1. **Interpretasi Kecepatan Angin**

Berdasarkan yang ditetapkan oleh Russ Rowlet (2001) dari *University of North California*, kecepatan angin tergolong pada 13 kategori seperti pada tabel 4.10.

Tabel 4.10 *Crisp membership function Wind Speed* (Rowlett, 2001)

| No | Speed (km/h) | Class |
| --- | --- | --- |
| 1 | x < 2 | Calm |
| 2 | 2 <= x < 5 | Light Air |
| 3 | 6 <= x < 11 | Light Breeze |
| 4 | 12 <= x < 19 | Gentle Breeze |
| 5 | 20 <= x < 29 | Moderate Breeze |
| 6 | 30 <= x < 39 | Fresh Breeze |
| 7 | 40 <= x < 50 | Strong Breeze |
| 8 | 51 <= x < 61 | Near Gale |
| 9 | 62 <= x < 74 | Gale |
| 10 | 74 <= x < 87 | Strong Gale |
| 11 | 88 <= x < 102 | Storm |
| 12 | 103 <= x < 118 | Violent Storm |
| 13 | 119 <= x < 130 | Hurricane |

* + - 1. **Interpretasi Arah Angin**

Pada interpretasi data arah angin penulis menggunakan *crisp membership function* yang dibuat oleh Zandlo, Bouley dan Ruschy dengan 16 kategori seperti pada tabel 4.11 (Zandlo et al., 2001).

Tabel 4.11 *C*risp *membership function Wind Direction* (Zandlo, Spoden, Bouley, & Ruschy, 2001)

| No | Degree | Class |
| --- | --- | --- |
| 1 | 348.75-11.25 | North |
| 2 | 11.25-33.75 | North North East |
| 3 | 33.75-56.25 | North East |
| 4 | 56.25-78.75 | East North East |
| 5 | 78.75-101.25 | East |
| 6 | 101.25-123..75 | East South East |
| 7 | 123.75-146.25 | South East |
| 8 | 146.25-168.75 | South South East |
| 9 | 168.75-191.25 | South |
| 10 | 191.25-213.75 | South South West |
| 11 | 213.75-236.25 | South West |
| 12 | 236.25-258.75 | West South West |
| 13 | 258.75-281.25 | West |
| 14 | 281.25-303.75 | West North West |
| 15 | 303.75-326.25 | North West |
| 16 | 326.25-348.75 | North North West |

* + - 1. **Interpretasi Cakupan Awan**

Peneliti menggunakan data interpretasi untuk cakupan awan yang didefinisikan oleh Jeff Huby seorang ahli meteorologis (Huby, 2010). Kategori cakupan awan terdapat pada tabel 4.12.

Tabel 4.12 *C*risp *membership function Cloud Coverage* (Huby, 2010)

|  |  |  |
| --- | --- | --- |
| NO | Cloud Coverage (%) | Class |
| 1 | 0-10 | Clear |
| 2 | 10-20 | Foggy |
| 3 | 20-30 | Mostly sunny |
| 4 | 40-60 | Partly cloudy |
| 5 | 70-80 | Mostly cloudy |
| 6 | 80-90 | Broken |
| 7 | 90-100 | overcast |

* + - 1. **Interpretasi Kualitas Udara**

Untuk menginterpretasi data kualitas udara penulis merujuk pada kategori yang ditentukan oleh Crowder, Moore, DeRose, dan Franek (Crowder, Moore, DeRose, dan Franek, 1999) dengan menentukan nilai *Polutant Standard Index* (PSI) seperti pada penelitian DWP (Putra et al, 2017). PSI ditentukan dengan membagi konsentrasi masing-masing lima polutan udara utama sehingga mendapatkan angka antara 0 dan 500, kelima pulutan tersebut adalah *Particle Matter* (PM10), partikel halus (PM2.5), sulfur dioksida (SO2), karbon monoksida (CO), ozon (O3) dan nitrogen dioksida (NO2) untuk jangka waktu tertentu yang ditetapkan oleh *National Ambient Air Quality Standard* (NAAQS) seperti pada tabel 4.13.

Tabel 4.13 *Crisp membership function* PSI dan *breakpoint* sub-indeks PSI (Crowder, Moore, DeRose, & Franek, 1999)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| PSI *Index* | PSI | PM25 | PM10 | SO2 | CO | O3 | NO2 |
| Good | 0-50 | 0-12 | 0-50 | 0-80 | 0-5 | 0-118 | 0 |
| Moderate | 51-100 | 13-55 | 51-15 | 81-365 | 5.1-10 | 119-157 | 0 |
| Unhealthy | 101-200 | 56-150 | 151-350 | 366-800 | 10.1-17.0 | 158-235 | 1130 |
| Very Unhealthy | 201-300 | 151-250 | 351-420 | 801-1600 | 17.1-34.0 | 236-785 | 1131-2260 |
| Hazzardous | 301-500 | 251-500 | 421-600 | 1601-2620 | 34.1-57.5 | 786-1180 | 2261-3750 |

Untuk menghitung PSI menggunakan fungsi linier tersegmentasi yang mengubah konsentrasi ambien ke skala yang berkisar dari 0 sampai 500. Setiap sub-indeks *i*, dihitung dengan menggunakan fungsi linier tersegmentasi yang berhubungan dengan konsentrasi polutan, *Xi* terhadap nilai sub-indeks, *Ii*. Fungsi linier tersegmentasi terdiri dari segmen garis lurus yang menggabungkan koordinat diskrit (yaitu *breakpoint* yang terdapat pada tabel 4.12). Untuk polutan *i* dan segmen *j*, koordinat titik *jth* break diwakili oleh nilai sub-indeks *Ii, j* dan konsentrasi *Xi, j* memberikan pasangan yang dipesan (*Xi, j, Ii, j*) (Putra et al, 2017). Jika konsentrasi yang diamati adalah *Xi* nilai sub-indeks yang sesuai *Ii* perhitungan dilakukan dengan menggunakan persamaan (keterangan rumus terdapat pada tabel 4.14):

(5)

Untuk mengimplementasikan persamaan diatas kedalam sebuah program, dapat dilakukan dengan cara (Putra et al, 2017):

1. Definisikan tiap *breakpoint* bagi setiap variabel polutan (PM25, PM10, SO2, CO, O3, NO2. Nilai *breakpoint* dari setiap variabel ini akan digunakan saat mengimplementasikan persamaan.
2. Definisikan tiap *breakpoint* untuk *Index* PSI.
3. Hitung nilai PSI bagi tiap variabel, dengan mengimplementasikan persamaan diatas.
4. Setelah didapatkan nilai PSI dari tiap variabel, cari nilai PSI terbesar dari.
5. Setelah nilai PSI terbesar didapatkan, cari interval dari PSI *Index* yang sesuai dengan nilai ini dari setiap indeks yang tersedia.
6. Setelah nilai PSI *Index* didapatkan, lalu ubah kedalam kategori sistem.

Tabel 4.14 Keterangan rumus PSI

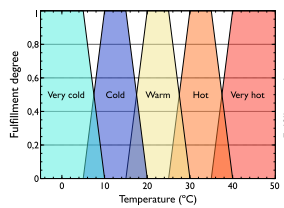
| NO | Variabel | Keterangan |
| --- | --- | --- |
| 1 | Xi | Nilai konsentrasi pollutant (yang diamati) ke-i. |
| 2 | Ii,j | Nilai PSI untuk polutan ke-i dan *breakpoint* ke-j (seperti yang diberikan pada Tabel). |
| 3 | Ii,j+1 | Nilai PSI untuk polutan ke-i dan breakpoint ke-(j + 1) (seperti yang diberikan pada Tabel). |
| 4 | Xi,j | Konsentrasi untuk polutan dan *breakpoint* seperti yang diberikan pada Tabel. |

Berikut contoh perhitungan PSI:

|  |
| --- |
| Misalkan diketahui nilai polutan PM25 adalah 25 , maka:   * + - * 1. Berdasarkan interval *breakpoint* nilai 25 pada PM25 berada diantara 12-55.         2. Maka perhitungan dilakukan pada segmen pertama (j = 1). Nilai dari *X1,1* = 12, *X1,2* = 55, *I1,1* = 50, dan *I1,2* = 100.7777         3. Sehingga kita dapat menghitung dengan menggunakan persamaan (5):         4. Maka nilai dari sub-indeks PSI (*I1*) PM25 adalah , sedangkan kita misalkan nilai sub-indeks PSI(*I2*) PM10 adalah 46.1412344, (*I3*) SO2 adalah 43.135112, (*I4*) CO adalah 45.166162, (*I5*) O3 adalah 15.136222 dan (*I6*) NO2 adalah -\*.         5. Cari nilai maksimal dari semua sub-indeks.         6. Diperoleh nilai 65.11627907, kemudian dipetakan kedalam tabel 4.9.   “Moderate”. |

* + - 1. **Interpretasi Suhu**

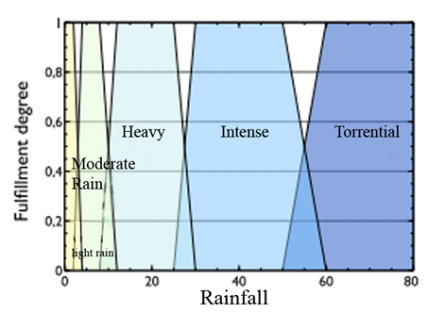
Interpretasi data suhu menggunakan *Fuzzy membership function* Ramos-soto, Alejandro, Bugarin, dan Barro yang terdiri dari *Very Cold*, *Cold*, *Warm*, *Hot*, dan *Very Hot* (Ramos-soto, Alejandro, Bugarín, dan Barro, 2016). Fungsi keanggotaan dapat dilihat pada gambar 4.4.



Gambar 4.4 *Fuzzy membership function Temperature* (Ramos-soto et al., 2016)

* + - 1. **Interpretasi Curah Hujan**

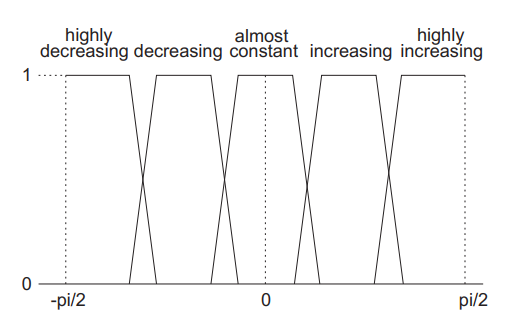
Interpretasi data curah hujan juga menggunakan *Fuzzy membership function* Ramos-soto, Alejandro, Bugarin, dan Barro yang terdiri dari *Light Rain, Moderate Rain, Heavy Rain, Intense Rain*,dan *Torential Rain* . Fungsi keanggotaan dapat dilihat pada gambar 4.5.



Gambar 4.5 *Fuzzy membership function Rainfall* (Ramos-soto et al., 2016)

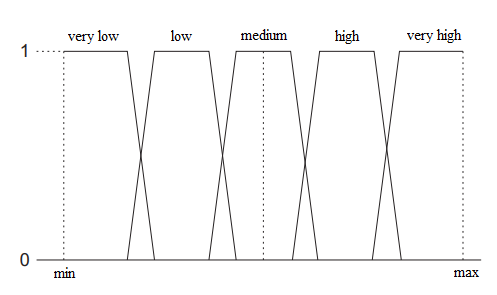
* + - 1. **Interpretasi Data Umum**

Pada interpretasi data umum, penulis mengacu pada pendefinisian trend oleh Castillo berdasarkan nilai pi (nilai rata-rata data), kemudian memasukan data pada *Fuzzy membership function* seperti pada gambar 4.6 (Castillo-Ortega, Marín, Martínez-Cruz, & Sánchez, 2014).



Gambar 4.6 *Linguistic variable for trend description* (Castillo-Ortega et al., 2014)

Dengan modifikasi pada interpretasi dan batas *min-max* fungsi keanggotaan, dikarenakan parameter yang digunakan secara *general* sehingga *Fuzzy membership function* untuk data *general* menjadi seperti gambar 4.7, dengan *corpus* *very low, low, medium, high,* dan *very high* (Fallah-Ghalhary, Mousavi-Baygi, & Nokhandan, 2009).



Gambar 4.7 *General Fuzzy membership function*

Dengan menggunakan *Data Abstraction and Event* dengan data pada tabel 4.4 didapatkan hasil interpretasi data seperti pada tabel 4.15.

Tabel 4.15 Hasil *Data Interpretation* contoh kasus *batch* pertama

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | CloudCoverage | Temperature | CO | SO2 |
| Resume Data | Mostly cloudy | warm | medium | low |
| Current Data | Foggy | cold | low | very high |
| Predict Data | Partly cloudy | cold | high | very low |
| ... | ... | ... | ... | ... |
| Trend | decrease | decrease | increase | increase |

Kemudian dilanjutkan dengan menggunakan *Data Abstraction and Event* berdasarkan data pada tabel 4.5 didapatkan hasil interpretasi data seperti pada tabel 4.16.

Tabel 4.16 Hasil *Data Interpretation* contoh kasus *batch* ke-dua

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | CloudCoverage | Temperature | CO | SO2 |
| Resume Data | Mostly cloudy | warm | medium | low |
| Current Data | Foggy | cold | low | very high |
| Predict Data | Partly cloudy | cold | high | very low |
| ... | ... | ... | ... | ... |
| Trend | decrease | decrease | increase | increase |

### Model Proses *Document Planning*

Pada proses ini dilakukan pemilihan konten (*Content Determination*), dan pembentukan struktur teks (*Document Structure*) yang akan ditampilkan (Reiter, 1996). Konten dipilih berdasarkan *Message Inventory* yang didapat, struktur teks dibangun berdasarkan *schema*-*based* (Turner et al., 2008). Sehingga *input*, proses, dan output pada tahap ini adalah sebagai berikut:

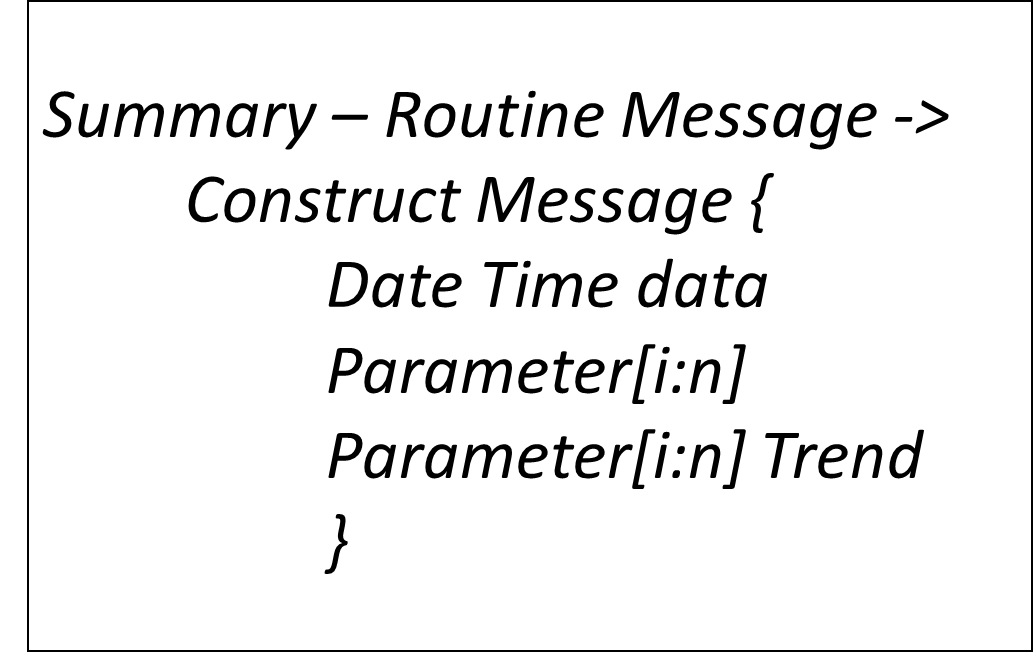
* *Input*: *Message Inventory*
* Proses: *Document Planning*
* *Output*: *Content and Structure*

Pemilihan konten dan pembentukan struktur untuk setiap paragraf akan dijelaskan pada sub-bab berikutnya.

* + - 1. **Perencanaan Dokumen untuk Ringkasan**

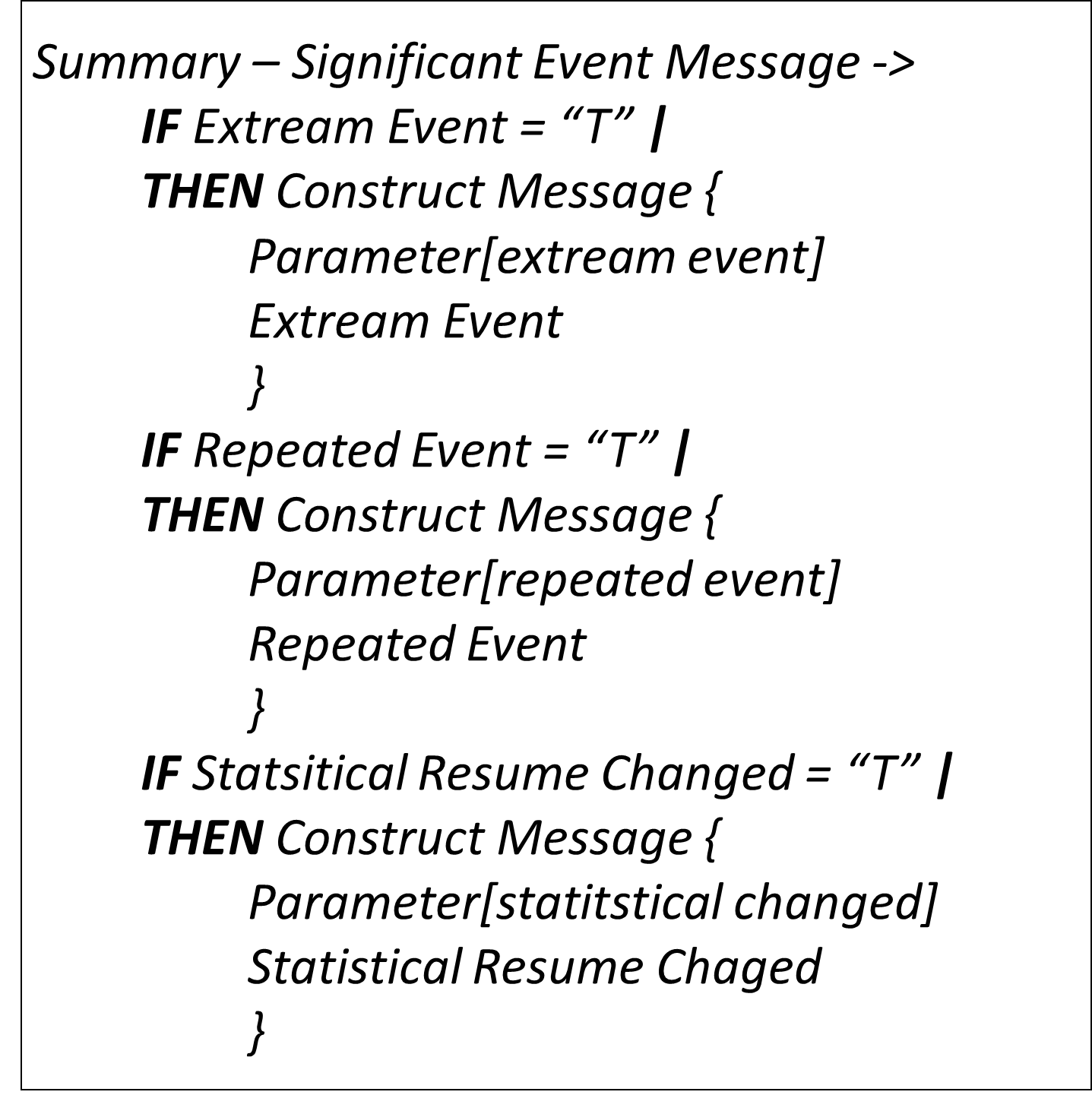
1. *Content Determination*

Pada proses untuk memilih konten ringkasan, penulis mengacu pada penelitian DWP (Putra et al, 2017). Dimana konten dikelompokkan kedalam dua bagian yaitu *Routine Message* dan *Significant Event Message. Routine Message* merupakan pesan umum yang selalu muncul dalam teks yang akan ditampilkan, sedangkan *Significant Event Message* adalah pesan yang muncul ketika persyaratan pada *event* tersebut terpenuhi, pada penelitian ini persyaratan *Significant Event Message* adalah ketika terdapat *Extreme Event* atau *Repeated Event* atau ketika *Statistical* *Summary* data ke *N* lebih tinggi atau lebih rendah dibandingkan *Statistical* *Summary* keseluruhan data. Untuk gambaran *Routine Message* pada gambar 4.8, terlihat bahwa terdapat variabel parameter[i:n], yang artinya dimulai dari parameter ke-i hingga parameter ke-n.



Gambar 4.8 *Routine Message* untuk ringkasan data

*Significant Event Message* dapat dilihat pada gambar 4.9, dimana pada gambar 4.9 terdapat variabel parameter[*extreme event*] atau sejenisnya, artinya paramter dimana *extreme event* tersebut terjadi.



Gambar 4.9 *Significant Event Message* untuk ringkasan data

Sebagai contoh dengan menggunakan data pada tabel 4.4 maka akan dihaslikan konten seperti pada tabel 4.17.

Tabel 4.17 Hasil *Content Determination Summary* contoh kasus *batch* pertama

|  |  |
| --- | --- |
| Routine Message | It can be seen that Cloud Coverage, Temperature trend is decreased but the rest is increased. |
| Significance Event Message | Cloud Coverage decreased extremely from 14 Jan 2014 to 15 Jan 2014 (decreased 56.7 points), Temperature decreased significantly from 11 Jan 2014 to 12 Jan 2014 (decreased 2.3 points), CO fluctuate extremely (decreased 3.8 points and increased 4.1 points),There was repeated value in SO2 from 07 Jan 2014 to 13 Jan 2014 with value 4.25. |

Kemudian dilanjutkan dengan menggunakan data pada tabel 4.5 maka akan dihaslikan konten seperti pada tabel 4.18.

Tabel 4.18 Hasil *Content Determination Summary* contoh kasus *batch* ke-dua

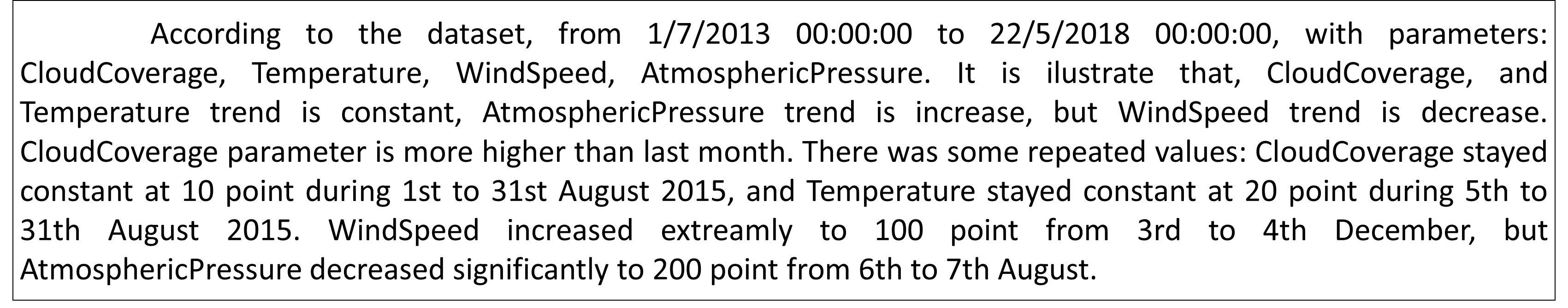
|  |  |
| --- | --- |
| Routine Message | It can be seen that Trend of all variable is increased. |
| Significance Event Message | Temperature, and SO2 parameters are more Higher than last week. but Cloud Coverage, and CO parameters are more Lower than last week. Cloud Coverage increased extremely from 18 Jan 2014 to 19 Jan 2014 (increased 66.4 points), CO increased significantly from 24 Jan 2014 to 25 Jan 2014 (increased 3.2 points) |

1. *Document Structuring*

Penentuan struktur dokumen dapat dilakukan dengan beberapa cara, diantaranya membangun skema dokumen atau penalaran eksplisit (Reiter dan Dale, 2000). Teks yang dihasilkan diharapkan sesuai dengan pola umum atau konvensional, maka dari itu diperlukan contoh teks dengan pola umum dan digunakan sebagai *Target Text*. *Target Text* yang dibuat untuk mendeskripsikan ringkasan data dapat dilihat pada gambar 4.10.

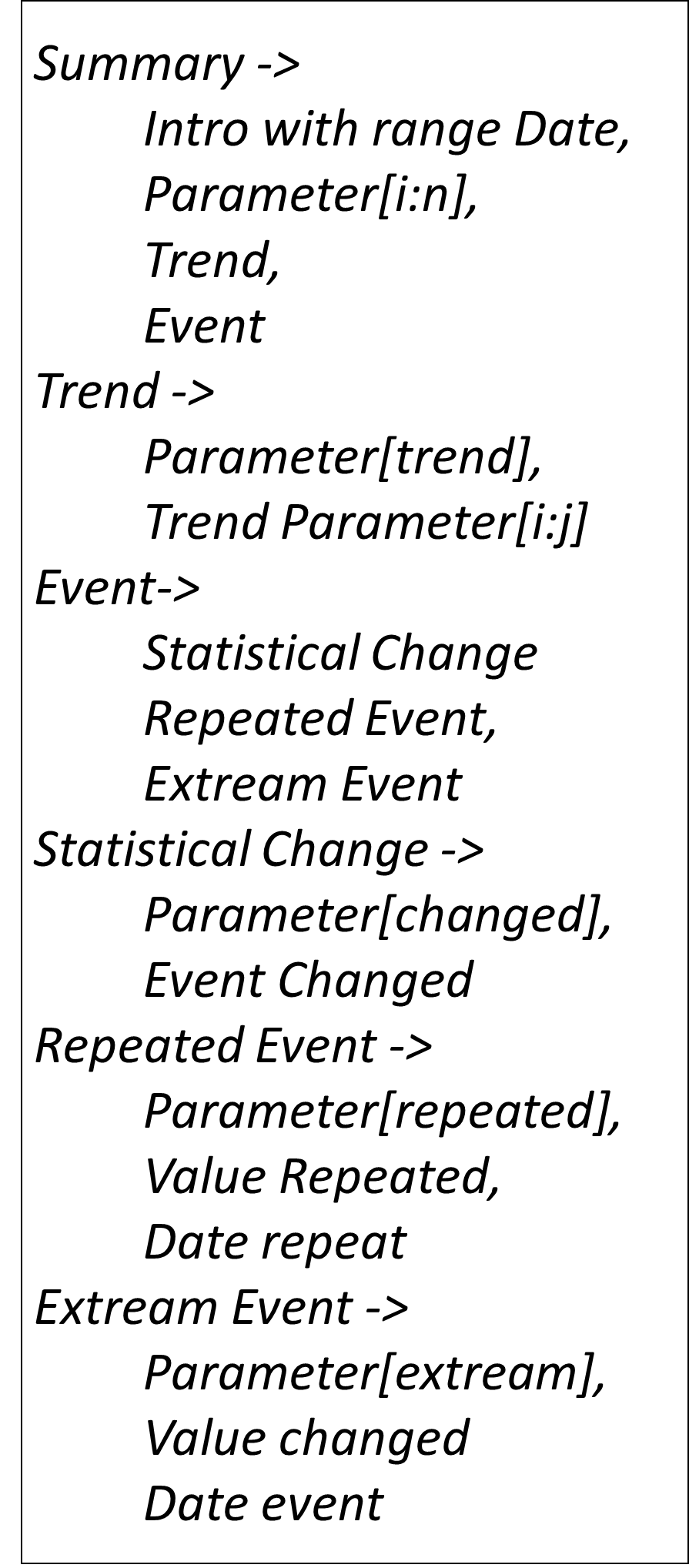
Reiter (1996) menjelaskan bahwa dengan berdasarkan *Target Text* pada gambar 4.10 dilakukan pembuatan skema dengan cara sebagai berikut:

* + - 1. Ambil contoh sejumlah teks yang sama dalam *Target Text*.
      2. Identifikasi terhadap pesan-pesan yang ada, tentukan bagaimana setiap pesan dapat dibangun berdasarkan data.
      3. Mengusulkan aturan atau struktur yang menjelaskan mengapa pesan “x” ada dalam teks A tetapi tidak ada didalam teks B. Penentuan ini lebih mudah jika disusun dalam bentuk seperti taksonomi.
      4. Diskusikan hasil analisis bersama pakar.



Gambar 4.10 *Target Text* untuk ringkasan data

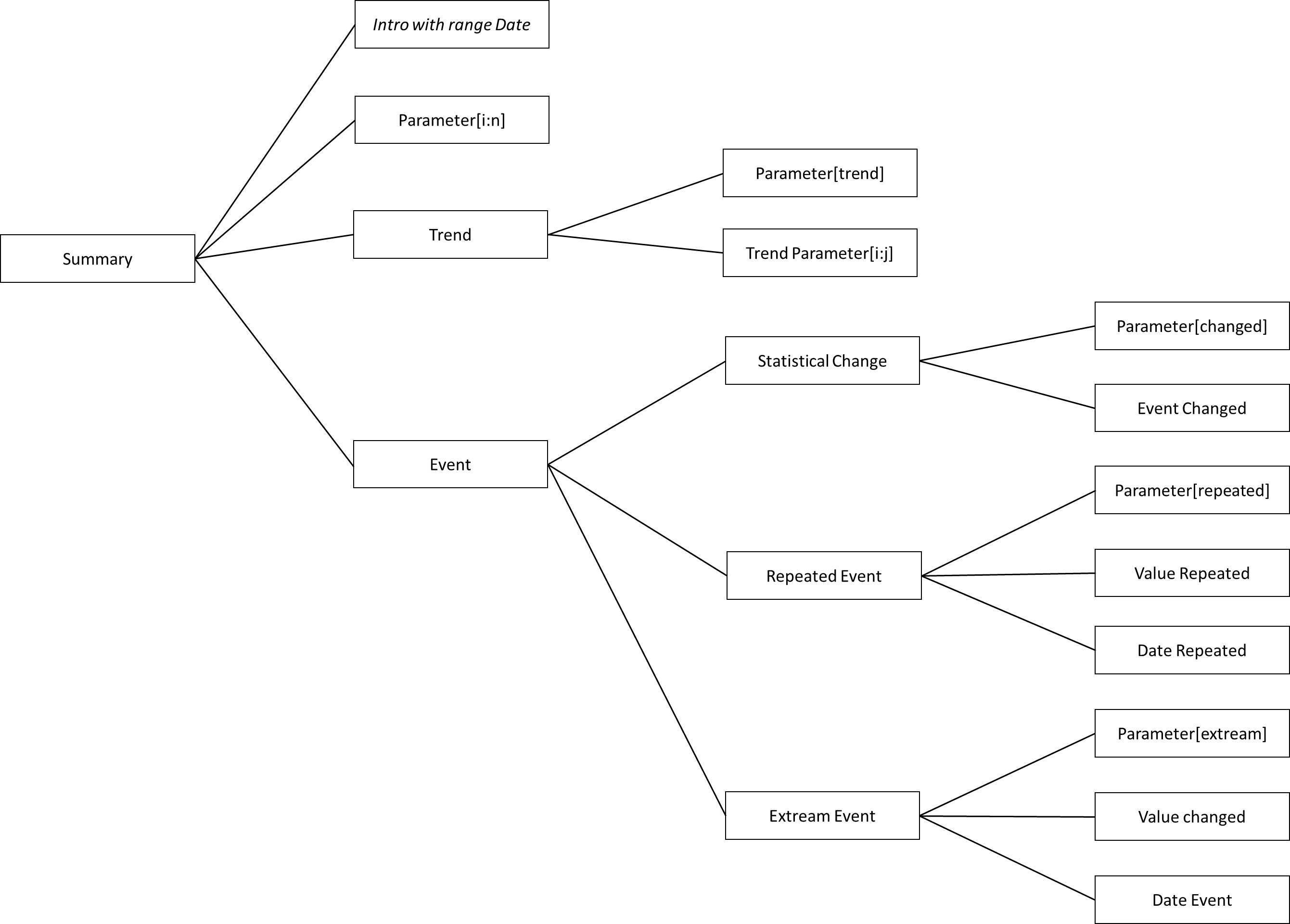
Dengan mengikuti langkah tersebut, penulis membuat skema struktur seperti pada gambar 4.11 berdasarkan *Target Text* pada gambar 4.10.



Gambar 4.11 *Structure text* ringkasan

Sehingga didapat pula pohon struktur dokumen seperti pada gambar 4.12. Pada struktur dokumen terdapan variabel i, j dan n, variabel i menunjukan iterasi dimulai dari i, sedangkan j dan n merupakan batas iterasi, n menunjukan jumlah keseluruhan parameter, sedangkan j jumlah keseluruhan jenis trend.

Pada gambar 4.12 didapat bahwa struktur dokumen pada ringkasan data terdiri dari *Intro with Range Date*, Parameter ke-i hingga ke-n, *Trend*, dan *Event*. Dimana *Trend* merupakan *Routine Message*, sedangkan *Event* merupakan *Significance Event Message*.

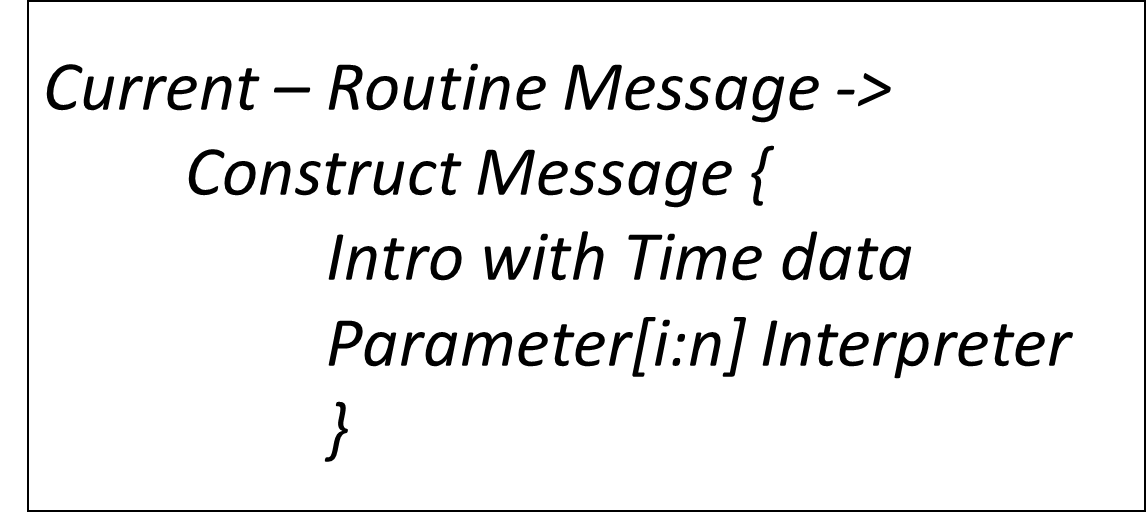


Gambar 4.12 pohon struktur teks ringkasan

* + - 1. **Perencanaan Dokumen untuk Data Terakhir**

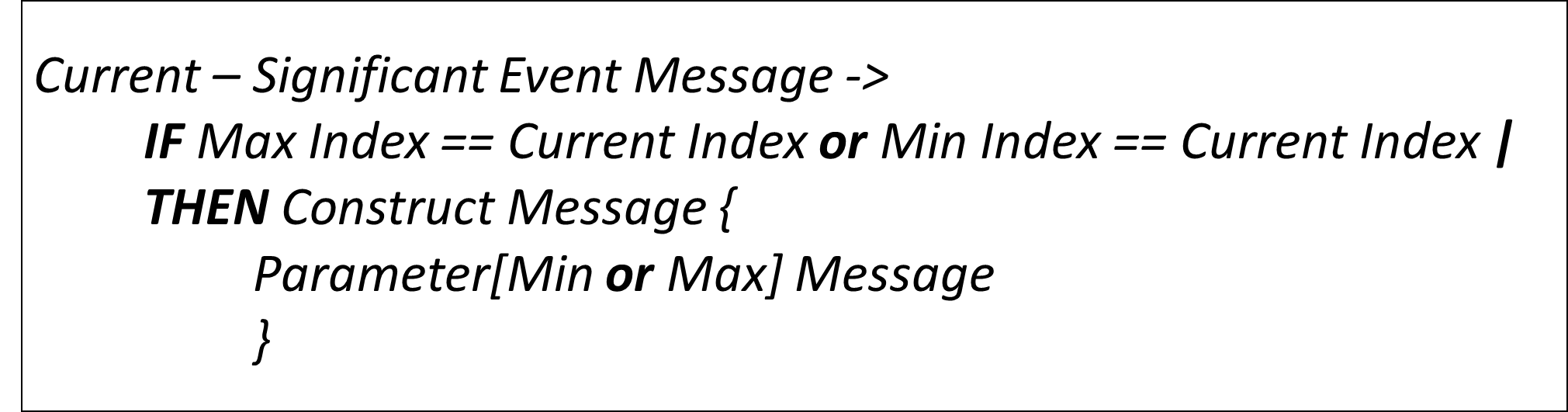
1. *Content Determination*

Pada pemilihan konten untuk data terakhir tetap menggunakan pembagian pesan kedalam dua kategori *Routine Message* dan *Significance Event Message*. *Routine Message* dapat dilihat pada gambar 4.13.

****

Gambar 4.13 *Routine Message* untuk data terakhir

Sedangkan *Significance Event Message* berdasarkan persyaratan jika data terakhir melebihi atau sama dengan nilai statistik maksimum atau minimum data *statistical summary* batch saat ini dapat dilihat pada gambar 4.14.



Gambar 4.14 *Significance Event Message* untuk data terakhir

Dengan menggunakan data pada tabel 4.4, maka didapat hasil seperti pada tabel 4.19.

Tabel 4.19 Hasil *Content Determination Current Data* contoh kasus *batch* pertama

|  |  |
| --- | --- |
| Routine Message | Cloud Coverage in Foggy condition. Temperature in cold condition. CO in low condition. SO2 in very high condition. |
| Significance Event Message | Cloud Coverage (at the lowest value so far) |

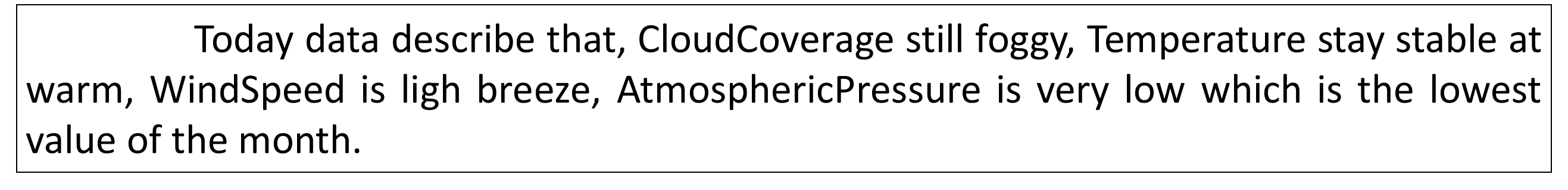
Diteruskan dengan menggunakan data pada tabel 4.5, maka didapat hasil seperti pada tabel 4.20.

Tabel 4.20 Hasil *Content Determination Current Data* contoh kasus *batch* ke-dua

|  |  |
| --- | --- |
| Routine Message | Cloud Coverage in Broken condition. Temperature in warm condition. CO, and SO2 in very high condition. |
| Significance Event Message | SO2 (at the highest value of so far) |

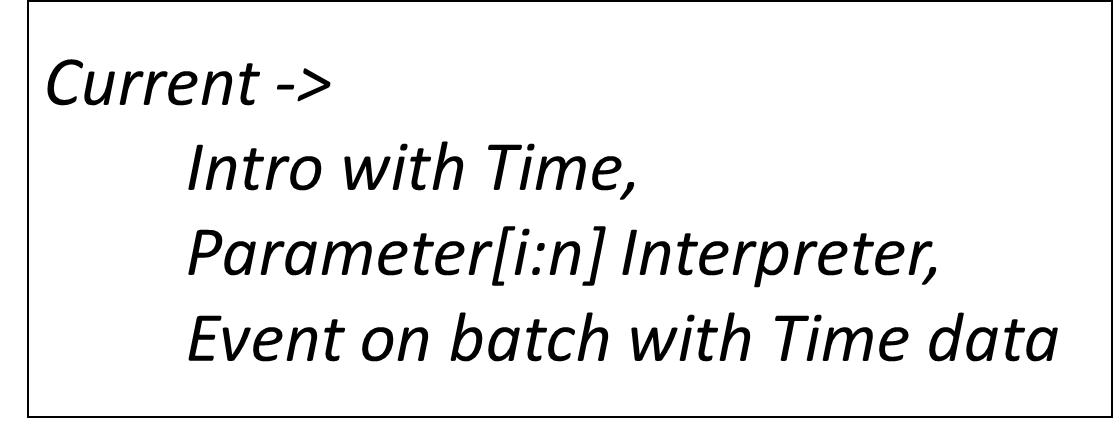
1. *Document Structuring*

*Target Text* untuk data terakhir pada suatu batch, dapat dilihat pada gambar 4.15.



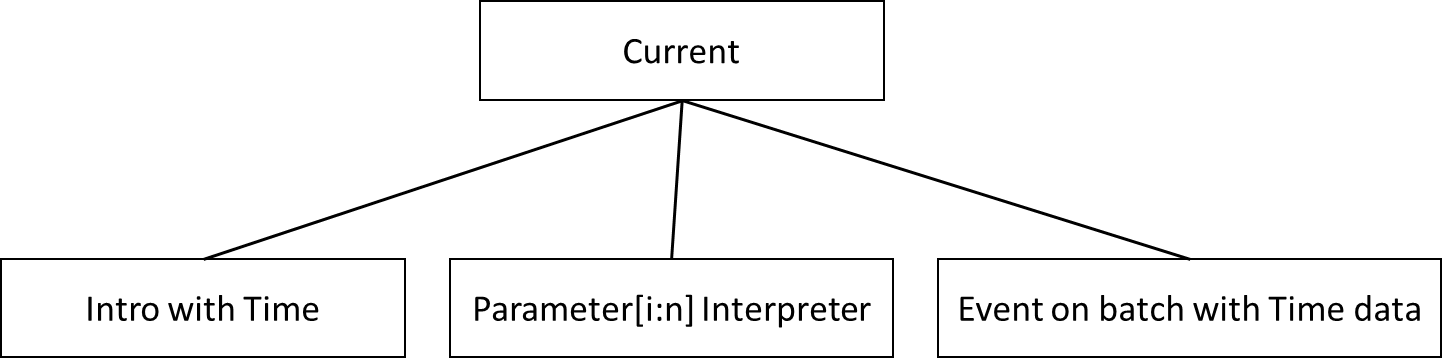
Gambar 4.15 *Target Text* data terakhir

Dengan berdasarkan *Target Text* pada gambar 4.15 dilakukan pembuatan skema dengan cara yang sama seperti pembuatan skema pada ringkasan, sehingga didapatkan hasil seperti pada gambar 4.16.



Gambar 4.16 *Structure text* data terakhir

Sedangkan untuk pohon struktur dapat dilihat pada gambar 4.17. Variabel i dan n merupakan iterasi dari i hingga n parameter, dimana i dimulai dari 1, waktu pada *intro* didapat dari interval *DateTime* antar data sedangkan waktu pada *event on batch* didapat dari interval *DateTime* data awal hingga data akhir. Pada gambar 4.17 terlihat bahwa struktur dokumen *Current* terdiri dari *Intro with Time*, interpretasi dari setiap parameter, dan *Event*.

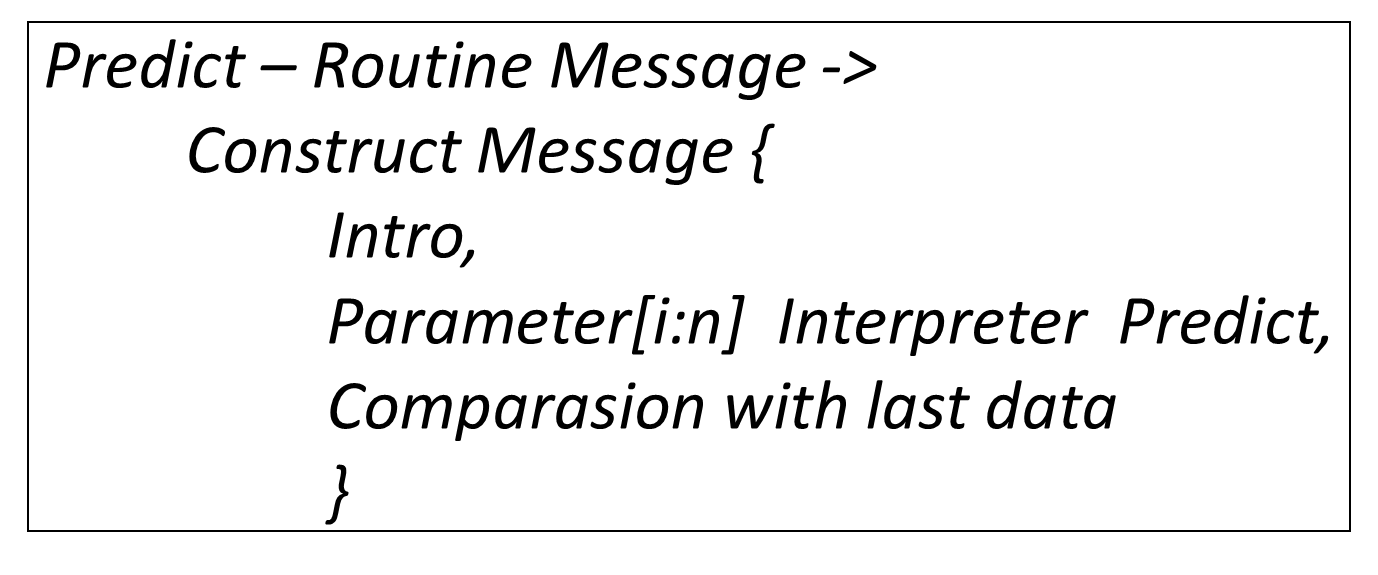


Gambar 4.17 Pohon struktur teks data terakhir

* + - 1. **Perencanaan Dokumen untuk Prediksi**

1. *Content Determination*

Pada pemilihan konten untuk prediksi hanya menggunakan *Routine Message*, dengan penambahan konten perbandingan dengan data terakhir. *Routine Message* dapat dilihat pada gambar 4.18.



Gambar 4.18 *Routine Message* prediksi

Dengan menggunakan data pada tabel 4.4, maka didapat hasil seperti pada tabel 4.21.

Tabel 4.21 Hasil *Content Determination Predict* contoh kasus *batch* pertama

|  |  |
| --- | --- |
| Routine Message | SO2 will increase extremely to very low, Temperature will keep stable at cold, Cloud Coverage, and CO will increase to Partly cloudy, and high. A conclusion of the predicted result is half variable will increase slowly. |

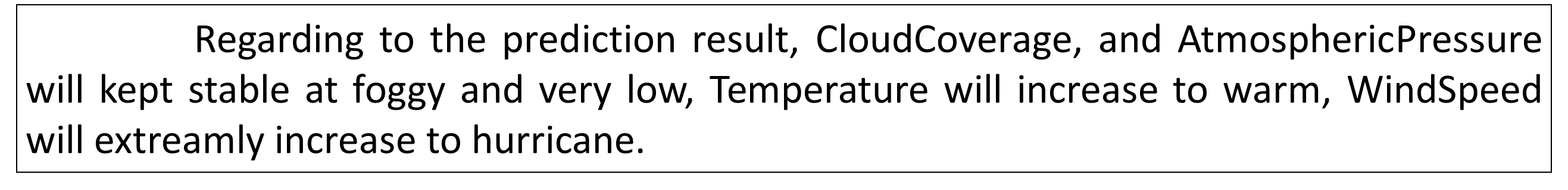
Diteruskan dengan menggunakan data pada tabel 4.5, maka didapat hasil seperti pada tabel 4.22.

Tabel 4.22 Hasil *Content Determination Predict* contoh kasus *batch* ke-dua

|  |  |
| --- | --- |
| Routine Message | Temperature will keep stable at warm, Cloud Coverage, CO, and SO2 will decrease to Partly cloudy, medium, and medium. A conclusion of the predicted result is overall variable will decrease slowly. |

1. *Document Structuring*

*Target Text* untuk prediksi dapat dilihat pada gambar 4.19, dimana hanya menyajikan informasi prediksi dan perbandingan kenaikan atau penurunan dengan data terakhir.

**

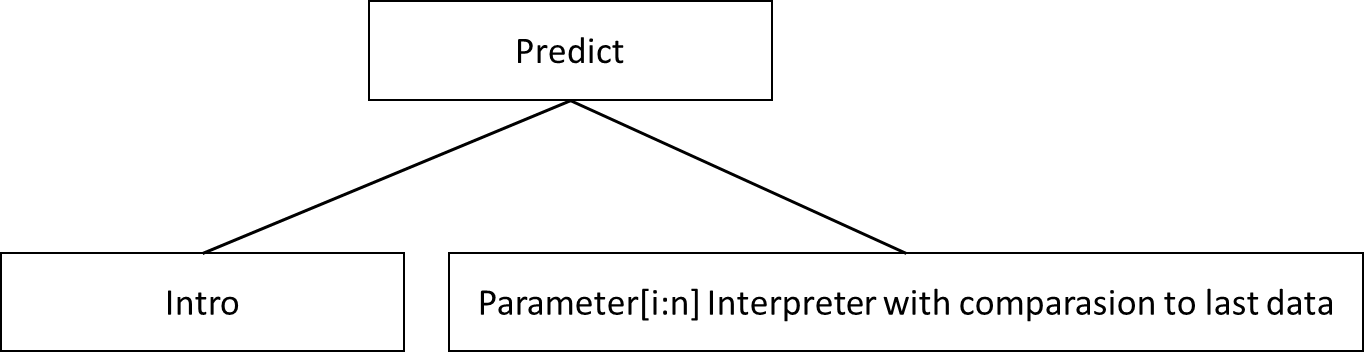
Gambar 4.19 *Target Text* untuk prediksi

Dengan berdasarkan *Target Text* pada gambar 4.19 dilakukan pembuatan skema dengan cara yang sama seperti pembuatan skema pada ringkasan, sehingga didapatkan hasil seperti pada gambar 4.20.



Gambar 4.20 *Structure text* prediksi

Dengan pohon struktur yang dibuat dapat dilihat pada gambar 4.21, dimana terdiri dari *Intro* dan prediksi dari setiap parameter.



Gambar 4.21 pohon struktur prediksi

### Model Proses *Microplanning and Realisation*

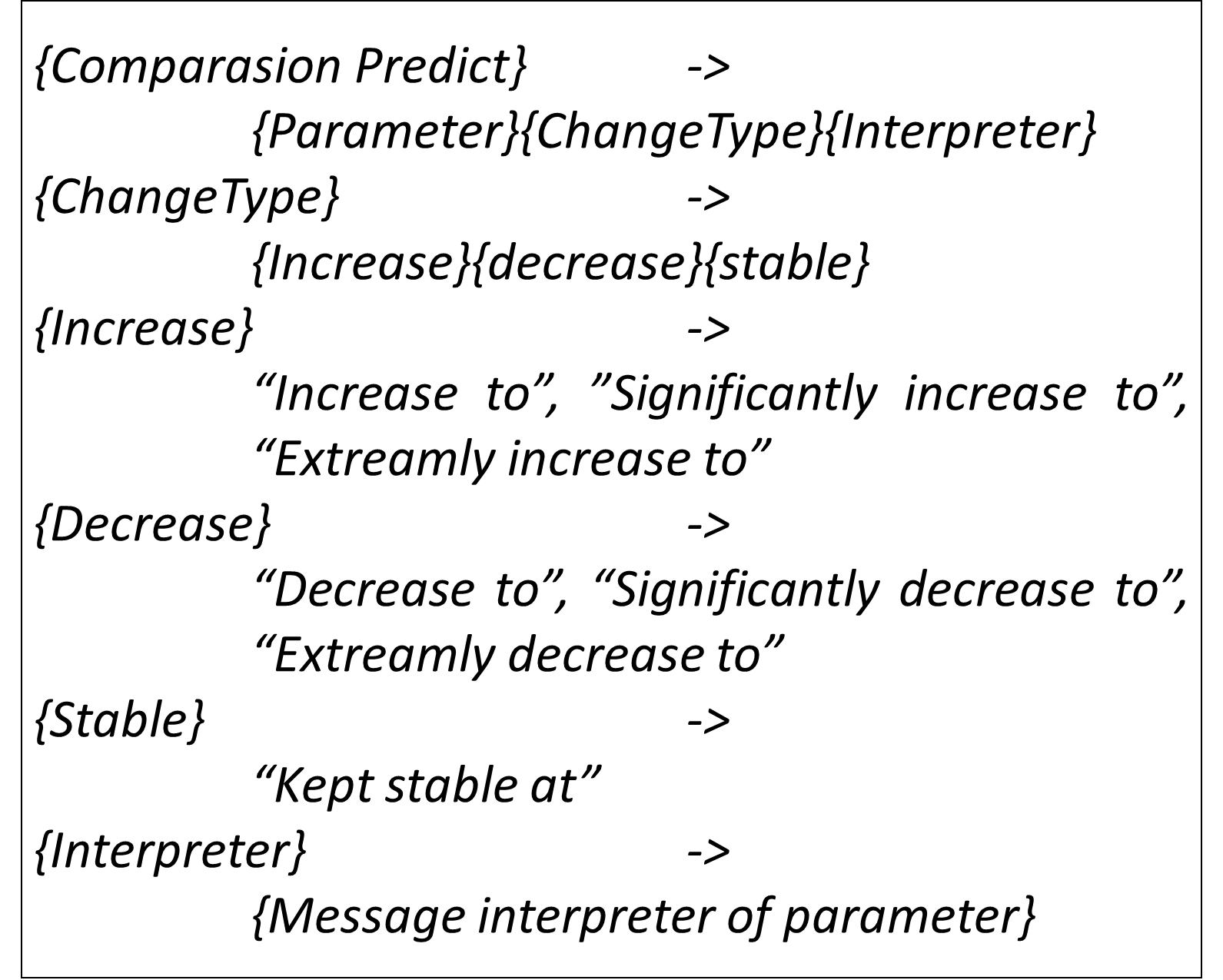
Pada proses *Microplanning* terbagi menjadi tiga proses yaitu, *Lexicalisation*, *Aggregation*, dan *Referring Expression Generation*, setelah selesai melakukan serangkaian proses tersebut maka teks dapat direalisasikan berdasarkan struktur yang ada. Dimana *input*, proses, dan output pada tahap ini adalah sebagai berikut:

* *Input*: *Content and Structure*
* Proses: *Microplanning and Realisation*
* *Output*: *Summary Text, Current Text,* dan *Predict Text*

Berikut penjelasan sub-proses *Microplanning and Realisation*.

* + - 1. ***Lexicalisation***

Pada tahap ini perubahan data pada parameter akan direpresentasikan kedalam bahasa manusia, seperti pada teks ringkasan yaitu “*increased extremely from*”, ” *decreased significantly from*” atau pada teks prediksi seperti “*will be decrease extremely to*”, “*will be increase significantly to*”, “*will be decrease to*”, dan lainnya. Proses ini merujuk pada representasi kualitas udara yang diterapkan dalam DWP (Putra et al, 2017), namun dengan perubahan penerapan tidak hanya dilakukan untuk parameter tertentu, tetapi secara umum. Maka dari itu peneliti membuat template untuk *lexicalisation* seperti pada gambar 4.22.



Gambar 4.22 *Template* *Lexicalisation* *Comparasion Predict*

* + - 1. ***Aggregation***

*Aggregation* dilakukan untuk menghubungkan beberapa pesan terkait menjadi satu kesatuan. Pada penelitian ini teknik yang digunakan hanya *Simple Conjunction,* yaitu menggunakan penghubung “*and*”atau “*but*” untuk membentuk sebuah kalimat yang diterapkan pada bagian *trend* didalam teks ringkasan (Reiter, 1996). Penghubung “*and*”diterapkan dalam kondisi dimana *trend* antara parameter tidak bertolak belakang, seperti “*CloudCoverage trend is stable, and the rest is increase.*” Sedangkan penghubung “*but*” diterapkan dalam kondisi *trend* yang saling bertolak belakang, seperti “*CloudCoverage trend is increase, but the rest is decrease.*”

* + - 1. ***Referring Expression Generation***

Pada tahap ini, dilakukan pemilihan kata atau ungkapan untuk merepresentaiskan entitas. Tahap ini relatif sederhana, karena dapat diterapkan secara *hard code* pada implementasi (Reiter, 1996), namun peneliti menerapkan apa yang telah diterapkan pada penelitian DWP, yaitu dengan membangkitkan secara *random* berdasarkan *corpus* yang dibuat (Putra et al, 2017).

* + - 1. ***Structure Realisation***

Pada tahap ini teks direalisasikan kedalam bentuk aktual berdasarkan struktur yang telah dibuat pada saat proses *document planning* (Reiter, 2011). Realisasi teks dapat dilakukan dengan menampilkan teks dalam bahasa pemrograman HTML, LaTeX, RTF, atau lainnya.

Sehingga hasil akhir dengan data pada tabel 4.4 terdapat pada tabel 4.23.

Tabel 4.23 Hasil akhir contoh kasus *batch* pertama

|  |
| --- |
| Regarding the dataset (07 Jan 2014 (00:00:00) - 15 Jan 2014 (00:00:00)), with parameters: Cloud Coverage, Temperature, CO and SO2. It shows that Cloud Coverage, Temperature trend is decreased but the rest is increased. Cloud Coverage decreased extremely from 14 Jan 2014 to 15 Jan 2014 (decreased 56.7 points), Temperature decreased significantly from 11 Jan 2014 to 12 Jan 2014 (decreased 2.3 points), CO fluctuate extremely (decreased 3.8 points and increased 4.1 points),There was repeated value in SO2 from 07 Jan 2014 to 13 Jan 2014 with value 4.25.  Based on the data provided, today can be described that Cloud Coverage (at the lowest value so far) in Foggy condition. Temperature in cold condition. CO in low condition. SO2 in very high condition.  Based on prediction result SO2 will increase extremely to very low, Temperature will keep stable at cold, Cloud Coverage, and CO will increase to Partly cloudy, and high. A conclusion of the predicted result is half variable will increase slowly. |

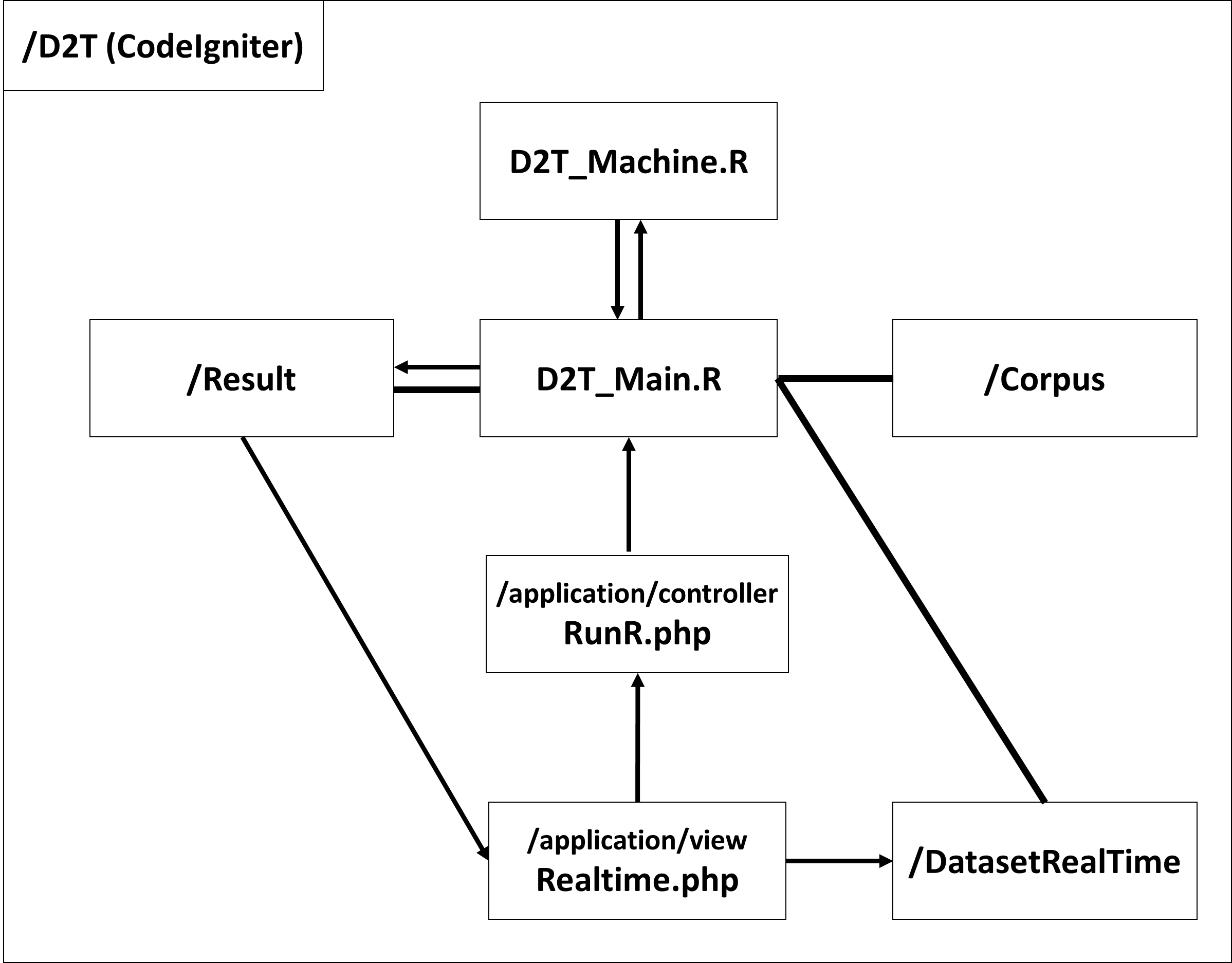
Sedangkan hasil akhir dengan data pada tabel 4.5 sebagai data lanjutan *batch* pertama terdapat pada tabel 4.24.

Tabel 4.24 Hasil akhir contoh kasus *batch* ke-dua

|  |
| --- |
| According to the dataset, between 16 Jan 2014 (00:00:00) to 25 Jan 2014 (00:00:00), with parameters: Cloud Coverage, Temperature, CO and SO2. It can be seen that Trend of all variable is increased. Temperature, and SO2 parameters are more Higher than last week. but Cloud Coverage, and CO parameters are more Lower than last week. Cloud Coverage increased extremely from 18 Jan 2014 to 19 Jan 2014 (increased 66.4 points), CO increased significantly from 24 Jan 2014 to 25 Jan 2014 (increased 3.2 points)  The Condition of today can be explained that Cloud Coverage in Broken condition. Temperature in warm condition. CO, and SO2 (at the highest value of so far) in very high condition.  Based on the result of prediction Temperature will keep stable at warm, Cloud Coverage, CO, and SO2 will decrease to Partly cloudy, medium, and medium. A conclusion of the predicted result is overall variable will decrease slowly. |

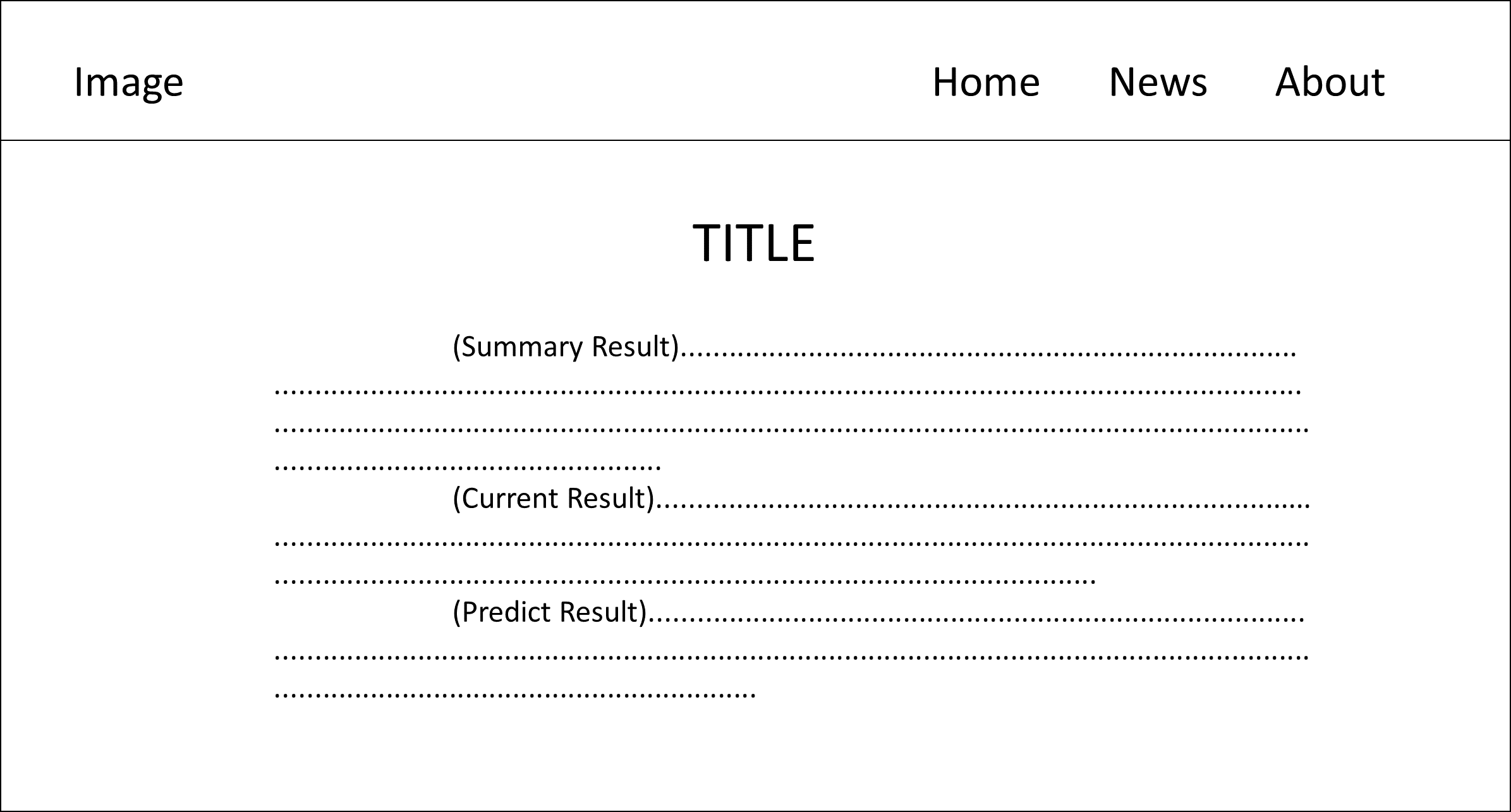
### Model Antarmuka Perangkat Lunak

Antarmuka sistem dibangun berbasis web menggunakan bahasa pemrograman HTML, Javascript, AJAX, dan PHP dalam framework CodeIgniter. Untuk mengintegrasikan D2T yang diimplemetasikan dalam bahasa pemrograman R, peneliti menggunakan perantara pengiriman data melalui JSON.



Gambar 4.23Struktur *file* applikasi D2T

Pada gambar 4.23 terlihat bahwa sistem D2T dibangun dalam framework CodeIgniter, yang sistem utamanya terdiri dari D2T\_Machine.R dimana semua implementasi *Data-to-Text* dalam bentuk fungsi berada disana, kemudian D2T\_Main.R yang didalamnya terdapat *read* data baik dataset, data hasil sebelumnya ataupun *corpus*, *write* hasil dalam bentuk CSV dan JSON serta pemanggilan fungsi D2T yang dibutuhkan. Proses *read* data dilakukan kedalam tiga folder, yaitu /DatasetRealTime, /Corpus, dan /Result. Proses *write* dilakukan kedalam folder /Result. Sedangkan file Realtime.R berfungsi sebagai pengecek bilamana ada dataset baru, kemudian mengirim sinyal terhadap RunR.php untuk menjalankan D2T\_Main.R, ketika proses selesai Realtime.R akan mengambil hasil yang ada di folder /Result dan menampilkannya. Gambar 4.24 menunjukan desain antarmuka sitem yang dibangun, dengan menggunakan framework css *bootstrap*.



Gambar 4.24Antarmuka sistem *General News Generator* (GNG)

## Implementasi Sistem *Data-to-Text*

Pada sub-bab ini akan dipaparkan secara menyeluruh mengenai implementasi sistem *Data-to-Text* secara teknis dengan model pada sub-bab 4.2.

### Implementasi Proses *Realtime Reader*

Untuk mengimplemetasikan proses ini penulis menggunakan Javascript animation dan AJAX untuk mengecek dataset baru (fungsi checkFile), bila terdapat dataset baru, maka sinyal akan dikirimkan melalui fungsi RunR() dan kemudian D2T\_Main.R akan dijalankan. Pada gambar 4.25 terlihat animasi dijalankan setiap 1000 *milisecond* untuk mengecek adanya file dataset atau tidak, jika terdapat dataset, maka akan dikirim sinyal “*run*” kepada contrlorrer RunR dengan metode *post*, jika tidak terdapat dataset baru, maka cek konten yang ditampilkan apakah sesuai dengan konten terakhir yang dihasilkan oleh D2T\_Main.R, jika terdapat perbedaan, maka rubah konten, perubahan ini tidak memerlukan reload page.

|  |
| --- |
| setTimeout**(**checkFile**,**1000**);** //Animation Checker  ***function*** checkFile**(){**  // Checking Dataset  $**.**ajax**({**  type**:** 'HEAD'**,**  url**:** 'http://localhost/D2T/DatasetsRealTime/Dataset.csv'**,**  success**:** ***function*()** **{**  // This is a new dataset  // Send signal to controller RunR for running D2T\_Main  $**.**ajax**({**  url**:**"http://localhost/D2T/RunR"**,** //the page containing php script  type**:** "post"**,** //request type,  dataType**:** 'json'**,**  data**:** **{**exec**:** "run"**}**  **});**  setTimeout**(**checkFile**,** 5000**);**  **},**  error**:** ***function*()** **{**  // No dataset found  // Check last update file content  $**.**ajax**({**  type**:** 'HEAD'**,**  url**:** 'http://localhost/D2T/Result/tempTime.json'**,**  success**:** ***function*()** **{**  // Validation Content in Web  checktempTime**();**  **},**  error**:** ***function*()** **{**  console**.**log**(**"nothing change"**);**  **}**  **});**  setTimeout**(**checkFile**,**1000**);**  **}**  **});**  **}** |

Gambar 4.25 *Realtime animation for check file*

Sedangkan pada gambar 4.26 terlihat bahwa sinyal “*run*” diproses untuk menjalankan fungsi yang dapat mengeksekusi D2T\_Main.R untuk membaca dataset.

|  |
| --- |
| **public** **function** executeR(){  // Execute D2T\_Main.R  exec ("\".\\R-3.4.0\\bin\\Rscript.exe\" .\\D2T\_Main.R 2>&1", $output);  // echo '<pre>', join("\r\n", $output), "</pre>\r\n";  }  **public** **function** Index()  {  **if**($\_POST["exec"] == "run"){  $this->executeR();  }  } |

Gambar 4.26 *Execute* D2T\_Main.R

Setelah proses membaca *file* selesai, maka dataset dihapus untuk meminimalisir penggunaan *hardisk* yang besar seperti pada gambar 4.27.

|  |
| --- |
| # Read data  dataset **<-** read.table**(**file**=**"DatasetsRealTime/Dataset.csv", sep**=**",", header**=TRUE)**  fn **<-** "DatasetsRealTime/Dataset.csv"  # Delete data  **if** **(**file.exists**(**fn**))** file.remove**(**fn**)** |

Gambar 4.27 *Read and Remove Dataset* *in* R

Sedangkan untuk menyimpan hasil digunakan library “jsonlite” dengan menggunakan fungsi toJSON() kemudian menuliskan data pada file dengan fungsi write() dan write.csv() seperti pada gambar 4.28.

|  |
| --- |
| library**(**jsonlite**)**  timeInterval **<-** toJSON**(**timeInterval**)**  resumeResult **<-** toJSON**(**resumeResult**)**  currentResult **<-** toJSON**(**currentResult**)**  predictResult **<-** toJSON**(**predictResult**)**  columnName **<-** toJSON**(**columnName**)**  write.csv**(**statisticalResume, file **=** "Result/statisticalResume.csv",row.names**=FALSE)**  write**(**columnName, file**=**'Result/columnName.json'**)**  write**(**timeInterval, file**=**'Result/timeInterval.json'**)**  write**(**now, file**=**'Result/tempTime.json'**)**  write**(**resumeResult, file**=**'Result/resumeResult.json'**)**  write**(**currentResult, file**=**'Result/currentResult.json'**)**  write**(**predictResult, file**=**'Result/predictResult.json'**)** |

Gambar 4.28 *Write Result* JSON and csv *in* R

Kemudian untuk menampilkan data pada web menggunakan AJAX $.getJSON() dan menampilkan hasilnya dengan append() seperti pada gambar 4.29.

|  |
| --- |
| $**.**getJSON**(**'http://localhost/D2T/Result/tempTime.json'**,** ***function*(**data**)** **{**  $**(**'#tempTime'**).**val**(**data**);**  **});**  $**.**getJSON**(**'http://localhost/D2T/Result/resumeResult.json'**,** ***function*(**data**)** **{**  $**(**'#newsResume'**).**append**(**"<p>&nbsp&nbsp&nbsp&nbsp&nbsp&nbsp&nbsp&nbsp&nbsp&nbsp"**+**data**+**"</p>"**);**  **});**  $**.**getJSON**(**'http://localhost/D2T/Result/currentResult.json'**,** ***function*(**data**)** **{**  $**(**'#newsCurrent'**).**append**(**"<p>&nbsp&nbsp&nbsp&nbsp&nbsp&nbsp&nbsp&nbsp&nbsp&nbsp"**+**data**+**"</p>"**);**  **});**  $**.**getJSON**(**'http://localhost/D2T/Result/predictResult.json'**,** ***function*(**data**)** **{**  $**(**'#newsPredict'**).**append**(**"<p>&nbsp&nbsp&nbsp&nbsp&nbsp&nbsp&nbsp&nbsp&nbsp&nbsp"**+**data**+**"</p>"**);**  **});** |

Gambar 4.29 *Get* JSON *in* AJAX

### Implementasi Proses *Signal Analysis*

* + - 1. **Prediksi Data**

Prediksi data dilakukan dengan membangun fungsi yang merepresentasikan teknik *Piecewise Linear Approximation* (PLA) dengan metode *Least Square* yang tersedia dalam bahasa R. Gambar 4.30 menjelaskan alur kerja fungsi PLAStream(), dimana pertama akan dibuat model linear dengan *Least Square* (fungsi lsfit()) untuk setiap parameter, jika batch bukan yang pertama, maka dilakukan pembacaan model sebelumnya dan dilakukan penggabungan model menjadi satu model.

Hasil prediksi menggunakan model keseluruhan dimasukkan kedalam sebuah dataframe, prediksi yang dilakukan hanya memprediksi data ke-*n+1*, kemudian data dimasukan kedalam dataframe dalam R untuk dilakukan proses selanjutnya yaitu *Data Interpretation.*

|  |
| --- |
| # PLA in stream  PLAStream **<-** **function(**dataset2**){**  col **<-** 2 #start at parameter 2, parameter 1 is DateTime  predicted **<-** matrix**()** #temporary predict result  **for(**col **in** col**:**length**(**dataset2**)){**  ....  resultLS **<-** lsfit**(**x,y**)** #create model linear using Least Square  #result model  Intercept **<-** resultLS**$**coefficients**[**1**]**  X **<-** resultLS**$**coefficients**[**2**]**  ....  **if(**stat **==** 0**){**  #if this is batch 1, then model linear batch 1 ~ model for all batch  write.csv**(**resultLSframe, file **=** resName,row.names**=FALSE)**  write.csv**(**resultLSframe, file **=** modName,row.names**=FALSE)**  #predict using model of this batch  predicted**[**col**-**1**]** **<-** X **\*** as.numeric**(**lengthCol**+**1**)** **+** Intercept  names**(**predicted**)[**col**-**1**]<-**paste**(**colnames**(**dataset2**[**col**]))**  **}else{**  #this is not batch 1, so merge linear result with model all (before this batch)  .....  #recreate data from past  mergeData **<-** 0  **for(**i **in** i**:(**endStream **-** lengthCol**)){**  mergeData**[**j**]** **<-** as.numeric**(**modelAll**[**"X"**])** **\*** as.numeric**(**i**)** **+** as.numeric**(**modelAll**[**"Intercept"**])**  j **<-** j**+**1  **}**  i **<-** startStream  #continue merge with recreate data this batch  **for(**i **in** i**:**endStream**){**  mergeData**[**j**]** **<-** resultLS**$**coefficients**[**2**]** **\*** as.numeric**(**i**)** **+** resultLS**$**coefficients**[**1**]**  j **<-** j**+**1  **}**  mergeData **<-** na.omit**(**mergeData**)**  #create new model linear from data  y **<-** mergeData  x **<-** c**((**startStream **-** lengthCol**):**endStream**)**  resultLSS **<-** lsfit**(**x,y**)**  ....  #write model all to csv  modelAll **<-** data.frame**(**Segment,X,Intercept,Ndata,row.names **=** **NULL)**  write.csv**(**modelAll, file **=** modName,row.names**=FALSE)**  #predict with model all  predicted**[**col**-**1**]** **<-** as.numeric**(**modelAll**[**"X"**])** **\*** as.numeric**(**j**)** **+** as.numeric**(**modelAll**[**"Intercept"**])**  names**(**predicted**)[**col**-**1**]<-**paste**(**colnames**(**dataset2**[**col**]))**  **}**  **}**  predicted**<-**t**(**predicted**)**  predicted**<-**data.frame**(**predicted**)**  return**(**predicted**)**  **}** |

Gambar 4.30 *PLAStream function*

Contoh hasil prediksi data radiasi sinar matahari dan meteorologi pada tabel 4.2 bulan Juli dapat dilihat pada tabel 4.25.

Tabel 4.25 Hasil prediksi data radiasi sinar matahari dan meteorologi

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| GHI | DNI | ... | WindSpeed | PWAT |
| 3.049937 | 5.263696 | ... | 4.223878 | 8.352567 |

* + - 1. **Ringkasan Data**

Implementasi untuk meringkas data, penulis menggunakan fungsi *min*, *max* dan *mean* dalam R seperti pada gambar 4.31.

|  |
| --- |
| StatisticalAnalysis **<-** **function(**dataset2**){**  dataset2WithoutDate **<-** dataset2**[** , colnames**(**dataset2**)** **!=** "DateTime"**]**  ColName **<-** Average **<-** MaxValue **<-** MaxIndex **<-** MaxDate **<-** MinValue **<-** MinIndex **<-** MinDate **<-** Trend **<-** c**(**""**)**    i**=**1  n**=**length**(**dataset2WithoutDate**)**  **for(**i **in** i**:**n**){**  ColName**[**i**]** **<-** colnames**(**dataset2WithoutDate**[**i**])**  #MAX  MaxValue**[**i**]** **<-** max**(**dataset2WithoutDate**[**i**])**  max\_index2 **<-** as.integer**(**which**(**dataset2WithoutDate**[**i**]==**max**(**dataset2WithoutDate**[**i**])))**  MaxIndex**[**i**]** **<-** max\_index2**[**1**]**  max\_index0 **<-** max\_index2**[**1**]**  MaxDate**[**i**]** **<-** as.character**(**dataset2**$**DateTime**[**max\_index0**])**  #MIN  MinValue**[**i**]** **<-** min**(**dataset2WithoutDate**[**i**])**  min\_index2 **<-** as.integer**(**which**(**dataset2WithoutDate**[**i**]==**min**(**dataset2WithoutDate**[**i**])))**  MinIndex**[**i**]** **<-** min\_index2**[**1**]**  min\_index0 **<-** min\_index2**[**1**]**  MinDate**[**i**]** **<-** as.character**(**dataset2**$**DateTime**[**min\_index0**])**  #AVERAGE  Average**[**i**]** **<-** mean**(**dataset2WithoutDate**[**,i**])**  Trend**[**i**]** **<-** TrendAnalysis**(**ColName**[**i**]**,dataset2WithoutDate**[[**i**]]**,as.double**(**MaxValue**[**i**])**, as.double**(**MinValue**[**i**])**, nrowDataset**)**  **}**  dataset2Statistical **<-** data.frame**(**ColName, MaxDate, MaxValue, MaxIndex, MinDate, MinValue, MinIndex, Average, Trend**)**;  return**(**dataset2Statistical**)**  **}** |

Gambar 4.31 *Statistical Summary function*

Untuk *trend* peneliti menggunakan pendekatan linear dengan fungsi lm() untuk menentukan kecenderungan data naik atau turun dan menyimpannya sebagai *file* *png* seperti pada gambar 4.32.

|  |
| --- |
| TrendAnalysis **<-** **function(**ColName,dataset2,max,min,nrowDataset**){**  **if(**length**(**unique**(**dataset2**))** **==** 1**){**  result **<-** "0"  **}else{**  x **=** c**(**1**:**length**(**dataset2**))**  reg **=** lm**(**dataset2**~**x**)**  df **<-** reg**$**coefficients**[**2**]** **+** reg**$**coefficients**[**1**]**  dl **<-** reg**$**coefficients**[**2**]\***nrowDataset **+** reg**$**coefficients**[**1**]**  range **<-** dl**-**df  stat **<-** 0  **if(**range **<** 0**){**  range **<-** range **\*** **(-**1**)**  stat **<-** 1  **}**  rangeReal **<-** max**-**min  **if(**range **>** 0.05**\***rangeReal**){**  **if(**stat **==** 0**){**  result **<-** "+"  **}else{**  result **<-** "-"  **}**  **}else{**  result **<-** "0"  **}**  i **<-** 1  newData **<-** matrix**()**  **for(**i **in** i**:**length**(**dataset2**)){**  newData**[**i**]** **<-** as.double**(**reg**$**coefficients**[**2**])\***i **+** as.double**(**reg**$**coefficients**[**1**])**  **}**  png**(**filename**=**paste0**(**ColName,".png"**)**, width**=**500, height**=**500**)**  plot**(**ts**(**dataset2,1**))**  lines**(**ts**(**newData,1**)**,col**=**"green"**)**  dev.off**()**  **}**    return**(**result**)**  **}** |

Gambar 4.32 *Trend Analysis function*

Hasil dari proses ringkasan data tabel 4.2 bulan Juli dapat dilihat pada tabel 4.26.

Tabel 4.26 Hasil ringkasan data radiasi sinar matahari dan meteorologi

| **ColName** | **MaxDate** | **MaxValue** | **MaxIndex** | **MinDate** | **MinValue** | **MinIndex** | **Average** | **Trend** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| GHI | 30/07/2016 00:00 | 3.887 | 30 | 19/07/2016 00:00 | 0.718 | 19 | 2.96970967741935 | + |
| DNI | 28/07/2016 00:00 | 7.61 | 28 | 19/07/2016 00:00 | 0 | 19 | 5.10735483870968 | + |
| DIF | 22/07/2016 00:00 | 1.415 | 22 | 13/07/2016 00:00 | 0.631 | 13 | 0.932516129032258 | - |
| GTI | 28/07/2016 00:00 | 6.564 | 28 | 19/07/2016 00:00 | 0.655 | 19 | 4.98758064516129 | + |
| Temperature | 03/07/2016 00:00 | 12.8 | 3 | 18/07/2016 00:00 | 1.2 | 18 | 7.41935483870968 | - |
| AP | 27/07/2016 00:00 | 974.6 | 27 | 06/07/2016 00:00 | 943.2 | 6 | 963.987096774194 | + |
| RH | 22/07/2016 00:00 | 76.3 | 22 | 28/07/2016 00:00 | 31 | 28 | 56.6870967741935 | + |
| WindSpeed | 18/07/2016 00:00 | 10.5 | 18 | 27/07/2016 00:00 | 1.6 | 27 | 4.3258064516129 | - |
| PWAT | 12/07/2016 00:00 | 13.8 | 12 | 05/07/2016 00:00 | 5.5 | 5 | 8.56451612903226 | - |

Jika *batch* bukan *batch* pertama, maka akan dideteksi pula *statistical event* yang terjadi dengan fungsi TypeEventResume() pada gambar 2.33.

|  |
| --- |
| TypeEventResume **<-** **function(**statisticalResume, statisticalResumePast**){**  # Perbandingan resume now dengan max min resume past  .....  n**=**nrow**(**statisticalResume**)**  **for(**i **in** i**:**n**){**  .....  **if(**as.double**(**statisticalResumePast**[**statisticalResumePast**$**ColName **==** ColName**[**i**]**,"MinValue"**])** **>** as.double**(**statisticalResume**[**statisticalResume**$**ColName **==** ColName**[**i**]**,"MinValue"**])){**  MinEvent**[**i**]** **<-** "Higher"  **}else** **if(**as.double**(**statisticalResumePast**[**statisticalResumePast**$**ColName **==** ColName**[**i**]**,"MinValue"**])** **<** as.double**(**statisticalResume**[**statisticalResume**$**ColName **==** ColName**[**i**]**,"MinValue"**])){**  MinEvent**[**i**]** **<-** "Lower"  **}else{**  MinEvent**[**i**]** **<-** "Same"  **}**  ....  **if(**AverageEvent**[**i**]** **==** "Higher" **&&** MaxEvent**[**i**]** **==** "Higher"**){**  MaxDate**[**i**]** **<-** as.character**(**statisticalResumePast**[**statisticalResumePast**$**ColName **==** ColName**[**i**]**,"MaxDate"**])**  **}else** **if(**AverageEvent**[**i**]** **==** "Lower" **&&** MaxEvent**[**i**]** **==** "Higher"**){**  MaxEvent**[**i**]** **<-** "+"  **}**  ...  eventNow **<-** data.frame**(**ColName, AverageEvent, MaxEvent, MinEvent, MaxDate, MinDate**)**  return**(**eventNow**)**  **}** |

Gambar 4.33 *TypeEventResume function*

Sehingga dengan data pada tabel 4.2 bulan Juli didapat hasil seperti pada tabel 4.27.

Tabel 4.27 Hasil fungsi *TypeEventResume*

| ColName | AverageEvent | MaxEvent | MinEvent | MaxDate | MinDate |
| --- | --- | --- | --- | --- | --- |
| GHI | Higher | Higher |  | 30/07/2016 00:00:00 |  |
| DNI | Higher | Lower |  |  |  |
| DIF | Higher | Lower |  |  |  |
| GTI | Higher | Lower |  |  |  |
| Temperature | Lower | Lower |  |  |  |
| AP | Lower | Lower | Higher |  | 06/07/2016 00:00:00 |
| RH | Lower | Lower | Higher |  | 28/07/2016 00:00:00 |
| WindSpeed | Lower | Lower |  |  |  |
| PWAT | Lower | Lower |  |  |  |

Jika terdapat *statistical event* maka ringkasan pada *batch* saat ini ringkasan saat ini akan berubah terpengaruh oleh ringkasan masa lampau yang tergambar pada gambar 2.34.

|  |
| --- |
| **...**  **for(**i **in** i**:**nrow**(**statisticalResume**)){**  **if(**as.double**(**as.character**(**statisticalResume**[**statisticalResume**$**ColName **==** as.character**(**statisticalResume**[**i,"ColName"**])**,"MaxValue"**]))** **<=** as.double**(**as.character**(**statisticalResumePast**[**statisticalResumePast**$**ColName **==** as.character**(**statisticalResume**[**i,"ColName"**])**,"MaxValue"**]))){**  statisticalResume**$**MaxDate **<-** as.character**(**statisticalResume**$**MaxDate**)**  statisticalResume**$**MaxValue **<-** as.character**(**statisticalResume**$**MaxValue**)**  statisticalResume**$**MaxIndex **<-** as.character**(**statisticalResume**$**MaxIndex**)**  statisticalResume**[**statisticalResume**$**ColName **==** as.character**(**statisticalResume**[**i,"ColName"**])**,"MaxValue"**]** **<-** as.double**(**as.character**(**statisticalResumePast**[**statisticalResumePast**$**ColName **==** as.character**(**statisticalResume**[**i,"ColName"**])**,"MaxValue"**]))**  statisticalResume**[**statisticalResume**$**ColName **==** as.character**(**statisticalResume**[**i,"ColName"**])**,"MaxIndex"**]** **<-** 0  statisticalResume**[**statisticalResume**$**ColName **==** as.character**(**statisticalResume**[**i,"ColName"**])**,"MaxDate"**]** **<-** as.character**(**statisticalResumePast**[**statisticalResumePast**$**ColName **==** as.character**(**statisticalResume**[**i,"ColName"**])**,"MaxDate"**])**  statisticalResume**$**MaxDate **<-** as.factor**(**statisticalResume**$**MaxDate**)**  statisticalResume**$**MaxValue **<-** as.factor**(**statisticalResume**$**MaxValue**)**  statisticalResume**$**MaxIndex **<-** as.factor**(**statisticalResume**$**MaxIndex**)**  **}**  **....**  **}**  *...* |

Gambar 4.34Merubah *Statistical Summary* karena adanya *Statistical Event*

Dengan menggunakan data pada tabel 4.2 bulan November didapat hasil seperti pada tabel 4.28, dimana terdapat waktu dengan periode yang berbeda tergantung pada seberapa jauh nilai data statistik yang melebihi statistik sebelumnya.

Tabel 4.28 Hasil ringkasan data radiasi sinar matahari dan meteorologi *batch* ke-n

| ColName | MaxDate | MaxValue | MaxIndex | MinDate | MinValue | MinIndex | Average | Trend |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| GHI | 06/11/2016 00:00 | 3.98 | 6 | 22/11/2016 00:00 | 0.646 | 22 | 2.82216666666667 | - |
| DNI | 08/11/2016 00:00 | 7.936 | 8 | 19/07/2016 00:00 | 0 | 0 | 4.2805 | - |
| DIF | 03/11/2016 00:00 | 1.66 | 3 | 17/11/2016 00:00 | 0.547 | 17 | 1.02863333333333 | - |
| GTI | 08/11/2016 00:00 | 6.658 | 8 | 19/07/2016 00:00 | 0.655 | 0 | 4.42446666666667 | - |
| Temperature | 03/11/2016 00:00 | 19.7 | 3 | 18/07/2016 00:00 | 1.2 | 0 | 13 | - |
| AP | 27/07/2016 00:00 | 974.6 | 0 | 06/07/2016 00:00 | 943.2 | 0 | 959.31 | - |
| RH | 26/11/2016 00:00 | 89.9 | 26 | 28/07/2016 00:00 | 31 | 0 | 66.0233333333333 | + |
| WindSpeed | 18/07/2016 00:00 | 10.5 | 0 | 27/11/2016 00:00 | 1.1 | 27 | 2.29333333333333 | + |
| PWAT | 04/11/2016 00:00 | 23.3 | 4 | 05/07/2016 00:00 | 5.5 | 0 | 14.51 | - |

* + - 1. ***Extreme Event***

Untuk mendeteksi *extreme event* peneliti menggunakan fungsi ResumeEventExtreme(), dengan *input* dataset, banyak data, banyak parameter, dan *statistical summary* dari hasil ringkas data. Kenaikan atau penurunan data yang tergolong *extreme event* jika dan hanya jika interval kenaikan atau penurunan melebihi 82% interval data (82% \* (max-min)) seperti yang terlihat pada gambar 4.35.

|  |
| --- |
| ResumeEventExtreme **<-** **function(**dataset2, lengthDataset2, nrowDataset, statisticalResume**){**  extremeInc **<-** matrix**()**  indexExInc **<-** matrix**()**  extremeDec **<-** matrix**()**  indexExDec **<-** matrix**()**  ColName **<-** matrix**()**    i **<-** 2  **for(**i **in** i**:**lengthDataset2**){**  ColName**[**i**-**1**]** **<-** colnames**(**dataset2**[**i**])**  extremeDec**[**i**-**1**]** **<-** min**(**diff**(**dataset**[**,i**]))**  extremeInc**[**i**-**1**]** **<-** max**(**diff**(**dataset**[**,i**]))**  indexExInc**[**i**-**1**]** **<-** which.max**(**diff**(**dataset**[**,i**]))**  indexExDec**[**i**-**1**]** **<-** which.min**(**diff**(**dataset**[**,i**]))**  **}**  i **<-** 1  summaryExtreme **<-** matrix**()**  **for(**i **in** i**:(**lengthDataset2**-**1**)){**  extremeInc**[**i**]** **<-** as.double**(**extremeInc**[**i**])**  extremeDec**[**i**]** **<-** as.double**(**extremeDec**[**i**])**  summaryExtreme**[**i**]** **<-** "0"  maxValue **<-** as.double**(**as.character**(**statisticalResume**[**i,"MaxValue"**]))**  minValue **<-** as.double**(**as.character**(**statisticalResume**[**i,"MinValue"**]))**  rangeReal **<-** **(**maxValue **-** minValue**)**  rangeExtreme **<-** rangeReal **-** **(**rangeReal**\***2**/**11**)**    **if(**extremeInc**[**i**]** **>** rangeExtreme **&&** extremeDec**[**i**]** **<** rangeExtreme**){**  rangeExtreme **<-** **(**maxValue **-** minValue**)**  **if(**extremeInc**[**i**]** **>=** rangeExtreme**){**  summaryExtreme**[**i**]** **<-** "++"  **}else{**  summaryExtreme**[**i**]** **<-** "+"  **}**    **}else** **if(**extremeDec**[**i**]** **>** rangeExtreme **&&** extremeInc**[**i**]** **<** rangeExtreme**){**  rangeExtreme **<-** **(**maxValue **-** minValue**)**  **if(**extremeDec**[**i**]** **>=** rangeExtreme**){**  summaryExtreme**[**i**]** **<-** "--"  **}else{**  summaryExtreme**[**i**]** **<-** "-"  **}**  ....  **}**  **}**  result **<-** data.frame**(**ColName,extremeInc,indexExInc,extremeDec,indexExDec,summaryExtreme,row.names **=** **NULL)**  return**(**result**)**  **}** |

Gambar 4.35 *ResumeEventExtreme function*

Dengan contoh hasil dengan data tabel 4.2 bulan November seperti pada tabel 4.29, dimana terdapat dua ekstrim *event* yaitu kenaikan ekstrim pada parameter GHI dan penurunan sangat ekstring pada parameter Temperature.

Tabel 4.29 Hasil extreme event analysis

| ColName | extremeInc | indexExInc | SummaryExtreme |
| --- | --- | --- | --- |
| GHI | 2.091 | 19 | 0 |
| DNI | 6.716 | 19 | + |
| DIF | 0.658 | 4 | 0 |
| GTI | 4.831 | 19 | 0 |
| Temperature | 60.100 | 16 | -- |
| AP | 8.100 | 5 | 0 |
| RH | 8.400 | 12 | 0 |
| WindSpeed | 2.700 | 16 | 0 |
| PWAT | 6.200 | 2 | 0 |

* + - 1. **Repeated Event**

Deteksi sinyal berulang secara berturut-turut dilakukan dengan menggunakan fungsi ResumeEventRepeat() dimana didalamnya terdapat fungsi *run length encoding* (rle()), yaitu menghitung jumlah nilai yang sama secara berurutan dalam satu vektor. Pada penelitian ini dipilih nilai dengan jumlah perulangan paling banyak, dan diseleksi hanyak nilai dengan jumlah perulangan lebih dari 10% jumlah data jika jumlah data lebih dari 19 baris, sedangkan kurang dari itu minimal 2 baris data sama berturut-turut yang masuk dalam kategori ini seperti pada gambar 4.36.

|  |
| --- |
| ResumeEventRepeat **<-** **function(**dataset2, ColName, nrowDataset**){**  temp **<-** rle**(**dataset2**)**  ....  **if(**nrowDataset **>** 19**){**  **if(**numRepeat **<** nrowDataset**/**10**){**  value **<-** 0  numRepeat **<-** 0  **}**  **}else{**  **if(**numRepeat **<** 2**){**  value **<-** 0  numRepeat **<-** 0  **}**  **}**  result **<-** data.frame**(**ColName,value,index,numRepeat,row.names **=** **NULL)**  return**(**result**)**  **}** |

Gambar 4.36 *ResumeEventRepeat function*

Dengan hasil seperti pada tabel 4.30, dimana terdapat sebuah *repeated event* pada parameter PWAT dengan nilai 21.2 sebanyak 12 baris mulai dari index pertama.

Tabel 4.30 Hasil repeat event analysis

| ColName | extremeInc | indexExInc | SummaryExtreme |
| --- | --- | --- | --- |
| GHI | 0 | 0 | 0 |
| DNI | 0 | 0 | 0 |
| DIF | 0 | 0 | 0 |
| GTI | 0 | 0 | 0 |
| Temperature | 0 | 0 | 0 |
| AP | 0 | 0 | 0 |
| RH | 0 | 0 | 0 |
| WindSpeed | 0 | 0 | 0 |
| PWAT | 21.2 | 1 | 12 |

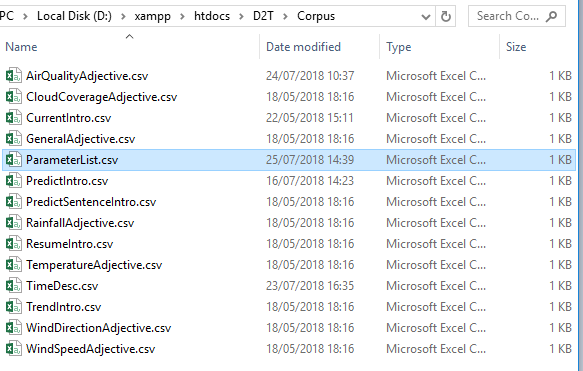
### Implementasi Proses *Data Interpretation*

Pada tahap ini impelentasi dilakukan dengan cara membuat daftar parameter beserta cara menginterpretasikan, apakah dalam bentuk *Fuzzy membership function* atau *Crisp membership function* yang disimpan dalam file */Corpus/ParameterList.csv* seperti pada tabel 4.31.

Tabel 4.31 Daftar Parameter dalam *file* csv bernama ParameterList

|  |  |
| --- | --- |
| Parameter | Interpreter |
| AirQuality | Crisp |
| WindSpeed | Crisp |
| WindDirection | Crisp |
| CloudCoverage | Crisp |
| Temperature | Fuzzy |
| Rainfall | Fuzzy |

Beserta fungsi keanggotaan dari parameter tersebut kedalam file csv (*separator coma*) dengan format nama *file* [*parameter*]*Adjective* seperti pada gambar 4.37*,* sehingga sistem dapat bekerja secara *general* untuk data apapun sesuai kebutuhan pengguna.



Gambar 4.37 *Corpus for Data Interpretation*

Proses interpretasi ini dilakukan dengan memanggil fungsi DataInterpreterAdjective(), fungsi ini bekerja dengan cara menyamakan nama kolom pada data dengan daftar parameter yang ada, jika terdapat dalam daftar maka lakukan data interpretasi sesuai fungsinya, jika tidak terdaftar maka lakukan dengan cara *general* seperti pada gambar 4.38.

Cara kerja data interpretasi untuk parameter *general* adalah dengan membuat sebanyak *n* (didapat dari banyak kategori dalam file GenneralAdjective.csv) *Fuzzy membership function* dengan *interval* nilai sesuai dengan nilai minimum dan maksimum dari *statistical summary* untuk parameter tersebut, kemudian memasukkan nilai kedalam fungsi pada gambar 4.38.

|  |
| --- |
| DataInterpreterAdjective **<-** **function(**value, type**=**"General",statisticalResume**){**  #read daftar list  parameters **<-** read.table**(**file**=**paste0**(**"Corpus/ParameterList.csv"**)**, sep**=**",", header**=TRUE)**  ....  **if(**statFound **==** 1**){**  #parameter on list  corpus **<-** read.table**(**file**=**paste0**(**"Corpus/",type,"Adjective.csv"**)**, sep**=**",", header**=TRUE)**  **if(**rule **==** "Fuzzy"**){**  result **<-** MembershipFuzzy**(**value, corpus**)**;  **}else{**  result **<-** MembershipCrisp**(**value, corpus**)**;  **}**  **}else{**  #general  ...  **if(**minRange **==** maxRange**){**  result **<-** "Constant"  **}else{**  n **=** nrow**(**corpus**)**  node **=** **(**2**\***n**)+**n**-**1  ...  rangenode **=** **(**maxRange**-**minRange**)/**node  i**=**1  j**=**0  membershipValue **<-** c**()**  **for** **(**i **in** i**:**n**)** **{**  **if(**i **==** 1**){**  ...  **}else{**  v1**<-**minRange**+(**j**)\***rangenode;  v2**<-**minRange**+(**j**+**1**)\***rangenode;  v3**<-**minRange**+(**j**+**3**)\***rangenode;  v4**<-**minRange**+(**j**+**4**)\***rangenode;    ##/ ¯ \ <- 1st area, 2nd area, 3rd area  #first area  **if((**value**>=**v1**)&&(**value**<=**v2**)){**  membershipValue**[**i**]** **<-** **(** **(**value**-**v1**)** **/** **(**v2**-**v1**)** **)**;  #second area (optimum)  **}else** **if((**value**>**v2**)&&(**value**<=**v3**)){**  membershipValue**[**i**]** **<-** 1;  #third area  **}else** **if((**value**>**v3**)&&(**value**<=**v4**)){**  membershipValue**[**i**]** **<-** **(** **(**v4**-**value**)** **/** **(**v4**-**v3**)** **)**;  #fourth, default condition (outside)  **}else{**  membershipValue**[**i**]** **<-** 0;  **}**  j **<-** j**+**3  **}**  **}**  result **<-** corpus**[**which.max**(**membershipValue**)**, "Category"**]**  **}**  **}**  return**(**result**)**  **}** |

Gambar 4.38 *DataInterpreterAdjective function*

Untuk implementasi *Fuzzy membership function* terdapat pada gambar 4.39.

|  |
| --- |
| MembershipFuzzy **<-** **function(**value, corpus**){**  i **<-** 1;  n **<-** nrow**(**corpus**)**;  m **<-** length**(**corpus**)**;    membershipValue **<-** c**()**  **for(**i **in** i**:**n**){**  v1**<-**corpus**[**i, "v1"**]**;  v2**<-**corpus**[**i, "v2"**]**;  v3**<-**corpus**[**i, "v3"**]**;  v4**<-**corpus**[**i, "v4"**]**;    ##/ ¯ \ <- 1st area, 2nd area, 3rd area  #first area  **if((**value**>=**v1**)&&(**value**<=**v2**)){**  membershipValue**[**i**]** **<-** **(** **(**value**-**v1**)** **/** **(**v2**-**v1**)** **)**;  #second area (optimum)  **}else** **if((**value**>**v2**)&&(**value**<=**v3**)){**  membershipValue**[**i**]** **<-** 1;  #third area  **}else** **if((**value**>**v3**)&&(**value**<=**v4**)){**  membershipValue**[**i**]** **<-** **(** **(**v4**-**value**)** **/** **(**v4**-**v3**)** **)**;  #fourth, default condition (outside)  **}else{**  membershipValue**[**i**]** **<-** 0;  **}**  **}**    #check highest membership result  membershipResult **<-** corpus**[**which.max**(**membershipValue**)**, "Category"**]**  return **(**membershipResult**)**  **}** |

Gambar 4.39 *Fuzzy Membership function*

Sedangkan untuk *Crisp membership function* terdapat pada gambar 4.40.

|  |
| --- |
| MembershipCrisp **<-** **function(**value, corpus**){**  return **(**sapply**(**value, **function(**v**)** corpus**[**v **>=** corpus**[**"Lower"**]** **&** v **<** corpus**[**"Upper"**]**,"Category"**]))**  **}** |

Gambar 4.40 *Crisp Membership function*

* + - 1. **Kecepatan Angin**

Implementasi *rule-based* untuk kecepatan angin seperti pada tabel 4.32 yang disimpan dalam *file* *WindSpeedAdjective.csv* dengan tipe *Crisp*.

Tabel 4.32 *Crisp membership value Wind Speed*

| **Category** | **Lower** | **Upper** |
| --- | --- | --- |
| Calm | 0 | 3 |
| Light Air | 3 | 6 |
| Light Breeze | 6 | 12 |
| Gentle Breeze | 12 | 20 |
| Moderate Breeze | 20 | 30 |
| Fresh Breeze | 30 | 40 |
| Strong Breeze | 40 | 51 |
| Near Gale | 51 | 62 |
| Gale | 62 | 74 |
| Strong Gale | 74 | 88 |
| Storm | 88 | 103 |
| Violent Storm | 103 | 119 |
| Hurricane | 119 | 130 |

* + - 1. **Arah Angin**

Implementasi *rule-based* untuk arah angin seperti pada tabel 4.33 yang disimpan dalam *file* *WindDirectionAdjective.csv* dengan tipe *Crisp*.

Tabel 4.33 *Crisp membership value Wind Direction*

| **Category** | **Lower** | **Upper** |
| --- | --- | --- |
| North | 0 | 11.25 |
| North North East | 11.25 | 33.75 |
| North East | 33.75 | 56.25 |
| East North East | 56.25 | 78.75 |
| East | 78.75 | 101.25 |
| East South East | 101.25 | 123.75 |
| South East | 123.75 | 146.25 |
| South South East | 146.25 | 168.75 |
| South | 168.75 | 191.25 |
| South South West | 191.25 | 213.75 |
| South West | 213.75 | 236.25 |
| West South West | 236.25 | 258.75 |
| West | 258.75 | 281.25 |
| West North West | 281.25 | 303.75 |
| North West | 303.75 | 326.25 |
| North North West | 326.25 | 348.75 |
| North | 348.75 | 360 |

* + - 1. **Cakupan Awan**

Implementasi *rule-based* untuk cakupan awan seperti pada tabel 4.34 yang disimpan dalam *file* *CloudCoverageAdjective.csv* dengan tipe *Crisp*.

Tabel 4.34 *Crisp membership value Cloud Coverage*

|  |  |  |
| --- | --- | --- |
| **Category** | **Lower** | **Upper** |
| Clear | 0 | 11 |
| Foggy | 11 | 21 |
| Mostly sunny | 21 | 31 |
| Partly cloudy | 31 | 51 |
| Mostly cloudy | 51 | 71 |
| Broken | 71 | 81 |
| Overcast | 81 | 100 |

* + - 1. **Kualitas Udara**

Perhitungan PSI untuk kualitas udara terdapat pada fungsi AirQualityCalculation (Putra et al, 2017) seperti pada gambar 4.41 dengan modifikasi pengecekan parameter yang ada, sehingga tidak semua data dapat menghasilkan interpretasi kualitas udara.

|  |
| --- |
| AirQualityCalculation **<-** **function** **(**dataset2**){**  ColName **<-** unique**(**names**(**dataset2**))**  i **<-** 1  stat **<-** 0  **for(**i **in** i**:**length**(**ColName**)){**  **if(**ColName**[**i**]** **==** "PM25" **||** ColName**[**i**]** **==** "PM10" **||** ColName**[**i**]** **==** "CO" **||** ColName**[**i**]** **==** "SO2" **||** ColName**[**i**]** **==** "O3"**){**  stat **<-** stat**+**1  **}**  **}**  **if(**stat **== 5){**  #lakukan perhitungan parameter  ...  PSI\_data **<-** c**(**PM25\_PSI\_value,PM10\_PSI\_value,SO2\_PSI\_value,CO\_PSI\_value,O3\_PSI\_value  PSI\_value **<-** as.integer**(**max**(**PSI\_data**))**  **}**  return**(**PSI\_value**)**;  **}** |

Gambar 4.41 *AirQualityCalculation function* (Putra et al, 2017)

Implementasi *rule-based* untuk kualitas udara seperti pada tabel 4.35 yang disimpan dalam *file* *AirQualityAdjective.csv* dengan tipe *Crisp*.

Tabel 4.35 *Crisp membership value Air Quality*

|  |  |  |
| --- | --- | --- |
| **Category** | **Lower** | **Upper** |
| Good | 0 | 51 |
| Moderate | 51 | 101 |
| Unhealthy | 101 | 201 |
| Very Unhealthy | 201 | 301 |
| Hazzardous | 301 | 501 |

* + - 1. **Suhu**

Implementasi *rule-based* untuk suhu seperti pada tabel 4.36 yang disimpan dalam *file* *TemperatureAdjective.csv* dengan tipe *Fuzzy*.

Tabel 4.36 *Fuzzy membership value Temperature*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Category** | **v1** | **v2** | **v3** | **v4** |
| very cold | 0 | 0 | 5 | 10 |
| Cold | 5 | 10 | 15 | 20 |
| Warm | 15 | 20 | 25 | 30 |
| Hot | 25 | 30 | 35 | 40 |
| very hot | 35 | 40 | 45 | 50 |

* + - 1. **Curah Hujan**

Implementasi *rule-based* untuk curah hujan seperti pada tabel 4.37 yang disimpan dalam *file* *RainfallAdjective.csv* dengan tipe *Fuzzy*.

Tabel 4.37 *Fuzzy membership value Rainfall*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Category** | **v1** | **v2** | **v3** | **v4** |
| light | 0 | 0 | 2.5 | 2.75 |
| moderate | 2.5 | 2.75 | 7.5 | 11.25 |
| heavy | 7.5 | 11.25 | 15 | 30 |
| intense | 15 | 30 | 50 | 60 |
| torential | 50 | 60 | 70 | 80 |

### Implementasi Proses *Document Planning*

Berdasarkan penjelasan pada sub-bab model proses *Document Planning* yang telah dipaparkan sebelumnya, implementasi dilakukan pemilihan konten yang akan ditampilkan diantaranya pada pemilihan fungsi ResumeEvent yang digunakan, jika *batch* saat ini merupakan *batch* pertama, maka tentu tidak akan ada *statistical event*, sehingga ResumeEvent yang digunakan hanya untuk memilih konten *extreme event* dan *repeated* *event* sedangkan bila bukan *batch* pertama maka konten yang dipilih termasuk *statistical event* seperti pada gambar 4.42.

|  |
| --- |
| **if(**stateType **==** 0**){**  resumeEvent **<-** ResumeEvent**(**dataset**[**,1**]**, typeEventResume, repeatedEvent, extremeEvent**)**  **}else{**  resumeEvent **<-** ResumeEvent**(**dataset**[**,1**]**, typeEventResume, repeatedEvent, extremeEvent,type**=**"0"**)**  **}** |

Gambar 4.42 *Content Determination ResumeEvent function*

ResumeEvent mengumpulkan semua *event* kemudian dipilih data mana saja yang memenuhi kriteria, yang selanjutnya direpresentasikan kedalam bahasa manusia seperti implementasi pada gambar 4.43 mengenai *statistical event*.

|  |
| --- |
| freq **<-** table**(**typeEventResume**[**"MaxEvent"**])**  freq2 **<-** table**(**typeEventResume**[**"MinEvent"**])**  stateHigh **<-** 0  fluktuatif **<-** as.character**(**typeEventResume**[**typeEventResume**[**"MaxEvent"**]==**"Higher" **&&** typeEventResume**[**"MinEvent"**]==**"Higher"**])**  **if(**length**(**fluktuatif**)** **>** 0**){**  i **<-** 1  **if(**length**(**fluktuatif**)** **==** 1**){**  result **<-** paste0**(**result," and ",fluktuatif**[**i**]**," parameters is more Fluctuate than @TimeDesc"**)**  result **<-** ChangeTimeDesc**(**result, dfdataset2, type**=**"past", typeEventResume**[**typeEventResume**$**ColName **==** fluktuatif**[**i**]**,"MaxDate"**])**  **}else{**  **...**  **}**  .......  **if(!**is.na**(**freq2**[**"Higher"**])){**  **if(**freq2**[**"Higher"**]** **==** 1**){**  stateHigh **<-** 1  temp **<-** as.character**(**typeEventResume**[**typeEventResume**$**MinEvent **==** "Higher","ColName"**])**    result **<-** paste0**(**result,temp**[**1**]**," parameter is more Lower than @TimeDesc"**)**  result **<-** ChangeTimeDesc**(**result, dfdataset2, type**=**"past", typeEventResume**[**typeEventResume**$**ColName **==** temp**[**1**]**,"MinDate"**])**    **}**  .....  **}**    **}** |

Gambar 4.43 Impelementasi *Event Determination Summary Event*

Implementasi *Content Determination Repeated Event* pada gambar 4.44, dimana terlihat pengecekan *event repeated*.

|  |
| --- |
| **if(**nrow**(**repeatedEvent**[**repeatedEvent**[**, "numRepeat"**]** **!=** 0,**])** **==** 1**){**  value **<-** 0  numRepeat **<-** 0  datetime **<-** 0  ColName **<-** 0  stat **<-** 0  i **<-** 1  **while(**stat **==** 0**){**  **if(**repeatedEvent**[**i,"numRepeat"**]** **>** 0**){**  stat **<-** 1  value **<-** repeatedEvent**[**i,"value"**]**  numRepeat **<-** repeatedEvent**[**i,"numRepeat"**]**  dateStart **<-** dataset2**[**repeatedEvent**[**i,"index"**]]**  dateEnd **<-** dataset2**[**repeatedEvent**[**i,"index"**]+**numRepeat**]**  ColName **<-** repeatedEvent**[**i,"ColName"**]**  **}**  i **<-** i**+**1  **}**  result **<-** paste0**(**"There was repeated value in ",ColName," from ",dateStart, " to ", dateEnd, " with value ", value, "."**)**  **}else** **if(**nrow**(**repeatedEvent**[**repeatedEvent**[**, "numRepeat"**]** **!=** 0,**])** **>** 2**){**  end **<-** nrow**(**repeatedEvent**[**repeatedEvent**[**, "numRepeat"**]** **!=** 0,**])**  result **<-** "There was some repeated value: "  value **<-** 0  numRepeat **<-** 0  datetime **<-** 0  ColName **<-** 0  stat **<-** 0  i **<-** 1  l **<-** 1  **while(**stat **<** end **&&** i **<** nrow**(**repeatedEvent**)){**  **if(**repeatedEvent**[**i,"numRepeat"**]** **>** 0**){**  stat **<-** stat **+** 1  value **<-** repeatedEvent**[**i,"value"**]**  numRepeat **<-** repeatedEvent**[**i,"numRepeat"**]**  dateStart **<-** dataset2**[**repeatedEvent**[**i,"index"**]]**  dateEnd **<-** dataset2**[**repeatedEvent**[**i,"index"**]+**numRepeat**]**  ColName **<-** repeatedEvent**[**i,"ColName"**]**  **if(**stat **==** end**){**  **if(**l **==** 1**){**  result **<-** paste0**(**"There was repeated value in ",ColName," from ",dateStart, " to ", dateEnd, " with value ", value, "."**)**  **}else{**  result **<-** paste0**(**result," and ",ColName," from ",dateStart, " to ", dateEnd, " with value ", value, "."**)**  **}**  **}else{**  l **<-** l**+**1  result **<-** paste0**(**result,ColName," from ",dateStart, " to ", dateEnd, " with value ", value, ", "**)**  **}**  **}**  i **<-** i**+**1  **}**  **}** |

Gambar 4.44 Implementasi *Event Determination Repeated Event*

Sedangkan untuk *Content Determination Extreme Event* dapat dilihatpada gambar 4.45.

|  |
| --- |
| **if(**nrow**(**extremeEvent**[**extremeEvent**[**, "summaryExtreme"**]** **!=** 0,**])** **==** 1**){**  value **<-** 0  numRepeat **<-** 0  datetime **<-** 0  ColName **<-** 0  stat **<-** 0  i **<-** 1  **while(**stat **==** 0**){**  **if(**extremeEvent**[**i,"summaryExtreme"**]** **==** "-"**){**  stat **<-** 1  value **<-** extremeEvent**[**i,"extremeDec"**]**  dateStart **<-** dataset2**[**extremeEvent**[**i,"indexExDec"**]]**  dateEnd **<-** dataset2**[**extremeEvent**[**i,"indexExDec"**]+**1**]**  ColName **<-** extremeEvent**[**i,"ColName"**]**  result **<-** paste0**(**result," ",ColName," decreased significantly from ",dateStart, " to ", dateEnd, " (decreased ", value, " points)."**)**  **}else** **if(**extremeEvent**[**i,"summaryExtreme"**]** **==** "+"**){**  stat **<-** 1  value **<-** extremeEvent**[**i,"extremeInc"**]**  dateStart **<-** dataset2**[**extremeEvent**[**i,"indexExInc"**]]**  dateEnd **<-** dataset2**[**extremeEvent**[**i,"indexExInc"**]+**1**]**  ColName **<-** extremeEvent**[**i,"ColName"**]**  result **<-** paste0**(**result," ",ColName," increased significantly from ",dateStart, " to ", dateEnd, " (increased ", value, " points)."**)**  **}else** **if(**extremeEvent**[**i,"summaryExtreme"**]** **==** "++"**){**  stat **<-** 1  value **<-** extremeEvent**[**i,"extremeInc"**]**  dateStart **<-** dataset2**[**extremeEvent**[**i,"indexExInc"**]]**  dateEnd **<-** dataset2**[**extremeEvent**[**i,"indexExInc"**]+**1**]**  ColName **<-** extremeEvent**[**i,"ColName"**]**  result **<-** paste0**(**result," ",ColName," increased extremely from ",dateStart, " to ", dateEnd, " (increased ", value, " points)."**)**  **}else** **if(**extremeEvent**[**i,"summaryExtreme"**]** **==** "--"**){**  stat **<-** 1  value **<-** extremeEvent**[**i,"extremeDec"**]**  dateStart **<-** dataset2**[**extremeEvent**[**i,"indexExDec"**]]**  dateEnd **<-** dataset2**[**extremeEvent**[**i,"indexExDec"**]+**1**]**  ColName **<-** extremeEvent**[**i,"ColName"**]**  result **<-** paste0**(**result," ",ColName," decreased extremely from ",dateStart, " to ", dateEnd, " (decreased ", value, " points)."**)**    **}**  i **<-** i**+**1  **}**  **....**  **}** |

Gambar 4.45 Implementasi *Event Determination Extreme Event*

Implementasi *Content Determination* pada *Current Text* pada pemilihan *event* dapat dilihat pada gambar 4.46.

|  |
| --- |
| **if(**as.character**(**statisticalResume**[**statisticalResume**$**ColName **==** colnames**(**interpreterNow**[**i**])**,"MaxIndex"**])** **==** as.character**(**nrow**(**dataset2**))){**  result **<-** paste0**(**result," which is the highest value of @TimeDesc (", statisticalResume**[**statisticalResume**$**ColName **==** colnames**(**interpreterNow**[**i**])**,"MaxValue"**]**," points)"**)**  result **<-** ChangeTimeDesc**(**result, dataset2**[**"DateTime"**]**, type**=**"this\_range"**)**  **}else** **if(**as.character**(**statisticalResume**[**statisticalResume**$**ColName **==** colnames**(**interpreterNow**[**i**])**,"MinIndex"**])** **==** as.character**(**nrow**(**dataset2**))){**  result **<-** paste0**(**result," which is the lowest value of @TimeDesc (", statisticalResume**[**statisticalResume**$**ColName **==** colnames**(**interpreterNow**[**i**])**,"MinValue"**]**," points)"**)**  result **<-** ChangeTimeDesc**(**result, dataset2**[**"DateTime"**]**, type**=**"this\_range"**)**  **}** |

Gambar 4.46 Implementasi *Event Determination Current Text*

Implementasi *Document Structuring* dilakukan dalam hal analisa struktur dokumen pada pembahasan sub-bab sebelumnya, untuk merealisasikan struktur yang dibuat ada pada sub-bab *Structure Realisation*.

### Implementasi Proses *Microplanning and Realisation*

Pada tahap ini implementasi dilakukan pada setiap tahap perbandingan data, contoh implementasi pada tahap *Lexicalisation* seperti yang telah dijelaskan pada sub-bab sebelumnyaterdapat pada gambar 4.47.

|  |
| --- |
| freq **<-** table**(**summaryPredict**)**  freq **<-** sort**(**freq**)**  summaryDF **<-** data.frame**(**ColName,summaryPredict**)**  result **<-** ""  **if(**length**(**freq**)** **==** 1**){**  temp **<-** ""  **for(**j **in** j**:**nrow**(**summaryDF**)){**  **if(**j **!=** nrow**(**summaryDF**)){**  temp **<-** paste0**(**temp,interpreterPredict**[**summaryDF**$**ColName**[**j**]]**," (",datasetPredicted**[**summaryDF**$**ColName**[**j**]]**,"), "**)**  result **<-** paste0**(**result,summaryDF**$**ColName**[**j**]**,", "**)**  **}else{**  temp **<-** paste0**(**temp,"and ",interpreterPredict**[**summaryDF**$**ColName**[**j**]]**," (",datasetPredicted**[**summaryDF**$**ColName**[**j**]]**,")."**)**  **if(**summaryDF**$**summaryPredict**[**j**]** **==** "+++"**){**  result **<-** paste0**(**result,"and ",interpreterPredict**[**summaryDF**$**ColName**[**j**]]**," will increase extremely to ",temp**)**  **}else** **if(**summaryDF**$**summaryPredict**[**j**]** **==** "---"**){**  result **<-** paste0**(**result,"and ",interpreterPredict**[**summaryDF**$**ColName**[**j**]]**," will decrease extremely to ",temp**)**  **}**  **...**  **}**  **.....**  **}** |

Gambar 4.47 Implementasi *Lexicalisation Extreme Event*

Implementasi *Aggregation* dengan *Simple Conjunction Referring to Contrast Value* dapat dilihat pada gambar 4.48 dengan fungsi ResumeTrend().

|  |
| --- |
| ResumeTrend **<-** **function(**statisticalResume**){**  freq **<-** table**(**statisticalResume**[**"Trend"**])**  result **<-** ""    #FIRST CONDITION  **if(!**is.na**(**freq**[**"0"**])** **&&** **!**is.na**(**freq**[**"+"**])** **&&** **!**is.na**(**freq**[**"-"**])){**  **if(**freq**[**"+"**]** **<=** freq**[**"-"**]** **&&** freq**[**"+"**]** **<=** freq**[**"0"**]){**  **if(**freq**[**"-"**]** **<** freq**[**"0"**]){**  #1ST GROUP  listTrend **<-** statisticalResume**[**statisticalResume**$**Trend **==** "+", **]**  i**<-**1  n **<-** nrow**(**listTrend**)**  **for(**i **in** i**:**n**){**  **if(**i**==**1**){**  result **<-** paste0**(**result,listTrend**[**i,"ColName"**])**    **}else{**  result **<-** paste0**(**result,", ",listTrend**[**i,"ColName"**])**  **}**  **}**  result **<-** paste**(**result, "trend is increased", "but "**)**  #2ND GROUP  .....  }  return**(**result**)**  **}** |

Gambar 4.48 Implementasi *Aggregation* pada *Trend Description*

Untuk contoh *Referring Expression Generation* terdapat pada gambar 4.49 dimana implementasi dilakukan pada pemilihan kalimat *intro* secara *random* berdasarkan *corpus* yang ada.

|  |
| --- |
| ReadIntro **<-** **function(**source**=**"Data", type**=**"General"**){**  type  **...**  as.matrix**(**read.table**(**file**=**paste0**(**"Corpus/",type,"Intro.csv"**)**, header**=FALSE**, sep**=**';'**))**  # print(corpus)  n **<-** length**(**corpus**)**  random\_value **<-** as.integer**(**runif**(**1,1,n**+**0.5**))**  ...  # return("Woops no data intro!");  **}**  ReadResumeIntro **<-** **function(**dataset2, ColName, source**=**"dataset2"**){**  corpus **<-** as.matrix**(**read.table**(**file**=**paste0**(**"Corpus/", "ResumeIntro.csv"**)**, header**=FALSE**, sep**=**';'**))**    #Randoming corpus  n **<-** length**(**corpus**)**  random\_value **<-** as.integer**(**runif**(**1,1,n**))**  result **<-** corpus**[**random\_value**]**  .....  return **(**result**)**  **}** |

Gambar 4.49 Implementasi *Referring Expression Generation* untuk *Intro*

Penentuan waktu data dalam bahasa manusia terdapat pada gambar 4.50 serta representasi dalam bentuk *file* dengan nama TimeDesc.csv yang berisi data pada tabel 4.22.

|  |
| --- |
| ChangeTimeDesc **<-** **function(**source, dataset2, type **=** "0", datePast **=** ""**){**  n **<-** nrow**(**dataset2**)**  **if(**is.null**(**nrow**(**dataset2**))){**  n **<-** length**(**dataset2**)**  **}**  timeFirst **<-** as.character**(**dataset2**[**n**-**1,'DateTime'**])**  timeLast **<-** as.character**(**dataset2**[**n,'DateTime'**])**  dayData **<-** ""  **if(**grepl**(**"/", timeFirst**)){**  timeFirst **<-** strptime**(**timeFirst, "%d/%m/%Y %H:%M:%OS"**)**  timeLast **<-** strptime**(**timeLast, "%d/%m/%Y %H:%M:%OS"**)**  dayData **<-** timeLast  dayData **<-** strptime**(**dayData, "%d/%m/%Y"**)**  **}else{**  timeFirst **<-** strptime**(**timeFirst, "%Y-%m-%d %H:%M:%OS"**)**  timeLast **<-** strptime**(**timeLast, "%Y-%m-%d %H:%M:%OS"**)**  dayData **<-** timeLast  dayData **<-** strptime**(**dayData, "%Y-%m-%d"**)**  **}**  now **<-** as.character**(**Sys.time**())**  now **<-** strptime**(**now, "%Y-%m-%d"**)**    same **<-** 0  **if(**type **==** "0"**){**  same **<-** 1  **}......**  difTime **<-** as.numeric**(**timeLast**-**timeFirst,units**=**"secs"**)**  corpus **<-** read.table**(**file**=**paste0**(**"Corpus/TimeDesc.csv"**)**, sep**=**",", header**=TRUE)**  timeDesc **<-** as.character**(**corpus**[**corpus**$**SecMin **<** difTime **&** corpus**$**SecMax **>=** difTime **&** corpus**$**Same **==** same,"Desc"**])**  result **<-** source  **if(**grepl**(**"@TimeDesc", source**)){**  result **<-** gsub**(**"@TimeDesc",timeDesc,source**)**  **}**    return **(**result**)**  **}** |

Gambar 4.50 Implementasi *Referring Expression Generation* untuk *Time Descritption*

Data pada tabel 4.38 merupakan pemilihan representasi waktu berdasarkan rentang waktu detik, tentu isi dari data tersebut dapat berubah sesuai kebutuhan pengguna.

Tabel 4.38 Representasi Waktu untuk *Referring Expression Generation*

| **Desc** | **SecMin** | **SecMax** | **Same** |
| --- | --- | --- | --- |
| this minute | -1 | 3000 | 1 |
| this hour | 3000 | 80000 | 1 |
| today | 80000 | 550000 | 1 |
| this week | 550000 | 2400000 | 1 |
| this month | 2400000 | 29000000 | 1 |
| this year | 29000000 | 66290000 | 1 |
| unknown range date | 66290000 | 1E+11 | 1 |
| unknown range date | -1E+09 | -1 | 1 |
| last minute | -1 | 3000 | 0 |
| last hour | 3000 | 80000 | 0 |
| last day | 80000 | 550000 | 0 |
| last week | 550000 | 2400000 | 0 |
| last month | 2400000 | 29000000 | 0 |
| last year | 29000000 | 66290000 | 0 |
| more than a year ago | 66290000 | 1E+11 | 0 |
| unknown range date | -1E+09 | -1 | 0 |
| that minute | -1 | 3000 | 2 |
| that hour | 3000 | 80000 | 2 |
| that day | 80000 | 550000 | 2 |
| that week | 550000 | 2400000 | 2 |
| that month | 2400000 | 29000000 | 2 |
| that year | 29000000 | 66290000 | 2 |
| more than a year ago | 66290000 | 1E+11 | 2 |
| unknown range date | -1E+09 | -1 | 2 |
| next minute | -1 | 3000 | 3 |
| next hour | 3000 | 80000 | 3 |
| tomorrow | 80000 | 550000 | 3 |
| next week | 550000 | 2400000 | 3 |
| next month | 2400000 | 29000000 | 3 |
| next year | 29000000 | 66290000 | 3 |
| more than a next year | 66290000 | 1E+11 | 3 |
| unknown range date | -1E+09 | -1 | 3 |

Sedangkan *Structure Realisation* implementasi dilakukan seperti pada akhir proses pada gambar 4.51

|  |
| --- |
| resumeResult **<-** paste**(**resumeIntro, resumeTrend, resumeEvent**)**  currentResult **<-** paste**(**currentIntro, currentDesc, descAQ**)**  predictResult **<-** paste**(**predictIntro, predictContent**)** |

Gambar 4.51 Implementasi *Structure Realisation*

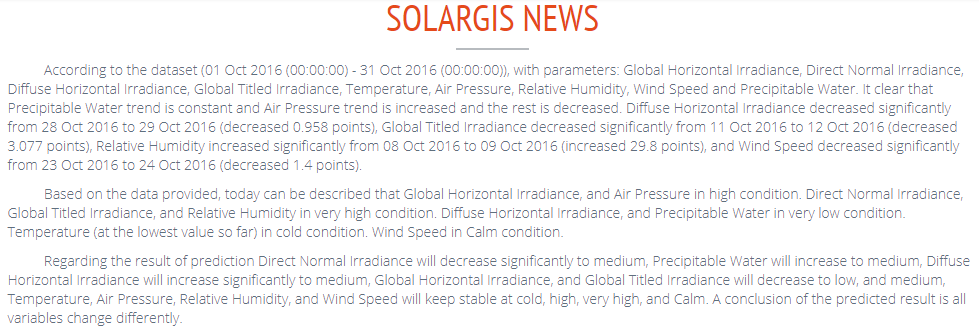


## Hasil Implementasi

Setelah selesai melakukan berbagai proses bada sub-bab sebelumnya, maka hasil dari sistem D2T yang telah peneliti buat dapat ditampilkan kedalam situs web dengan *framework* CodeIgniter dan css Bootstrap. Teks yang dihasilkan terdiri dari tiga paragraf, dimana paragraf pertama menjelaskan tentang ringkasan data dan *event* yang terjadi, paragraf ke-dua menjelaskan tentang keadaan saat ini, dan paragaf terakhir mendeskripsikan hasil prediksi yang dilakukan dengan teknik PLA dan menggunakan metode *Least Square*. Pada sub-bab ini akan dipaparkan mengenai hasil keluaran sistem, perbandingan hasil keluaran dengan sistem pada penelitian terkait, dan validasi keluaran sistem.

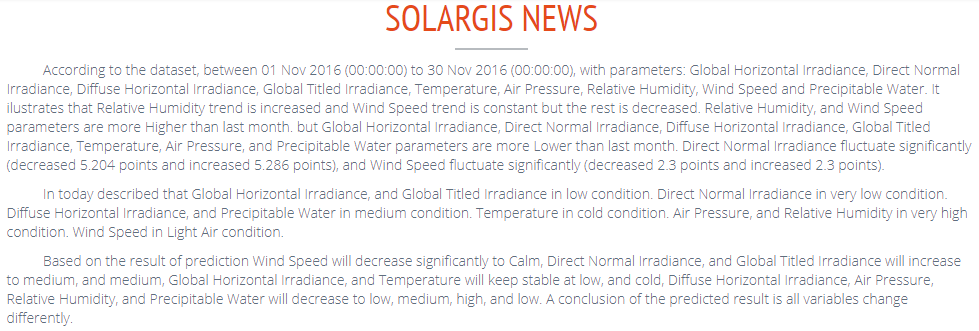
### *Output* Sistem

*Output* yang dihasilkan sistem ini berupa berita secara umum untuk data apa saja, model *linear* untuk setiap *batch* serta untuk keseluruhan *batch* dan *statistical summary*. Beberapa hasil keluaran teks sistem dengan *input* data radiasi sinar matahari dan meteorologi periode waktu harian dan setiap *batch* merupakan data dalam satu bulan dapat dilihat pada gambar 4.52, sedangkan output model dan *statistical summary* pada tabel 4.25.



Gambar 4.52 *News* *from first batch*

Dari berita pada gambar 4.52 dijelaskan bahwa berdasarkan pada dataset pada bulan tersebut *trand* parameter Atmospheric Pressure, Relative Humidity, Wind Speed dan Precipitable Water turun, sedangkan parameter lainnya naik dalam hal ini penghubung antara dua hal yang kontras menggunakan “*but*” yang merupakan implementasi dari *Aggregation refferrence to contrast value*. Pada paragraf ke-dua terdapat kata “*In Today explained that*” yang merupakan implementasi dari *Referring Expression Generation* terdapat *event* bahwa pada hari itu Temperature dalam kondisi paling rendah di bulan itu. Untuk interpretasi dari setiap nilai jelas terpapar pada paragraf deskripsi dan juga terlihat proses *Lexicalisation* pada “*decrease significantly*”,”*increase to*”, “*increase significantly*”, dan lainnya.



Gambar 4.53 *News* *from second batch*

Sedangkan berita pada gambar 4.53 bagian paragraf pertama mulai menunjukan adanya *statistical event* yang merepresentasikan adana kenaikan suku serte penurunan nilai pada beberapa parameter. Sedangkan pada paragraf ke-dua, tidak ada *event* yang berlangsung, dikarenakan syarat tidak terpenuhi. Sedangkan pada Tabel 4.39 dengan data tabel 4.2 bulan Oktober terlihat model yang didapatkan dengan menggunakan *Least Square Method* untuk dan , serta hasil dari penggunaan teknik PLA dengan *Least Square* yaitu yang digunakan untuk prediksi data.

Tabel 4.39 *Output model and statistical summary*

| **ColName** | **MaxDate** | **MaxValue** | **...** |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| GHI | 01/10/2016 00:00 | 5.462 | ... |  |  |  |
| DNI | 30/10/2016 00:00 | 8.299 | ... |  |  |  |
| DIF | 27/10/2016 00:00 | 2.113 | ... |  |  |  |
| GTI | 03/10/2016 00:00 | 7.187 | ... |  |  |  |
| Temperature | 08/10/2016 00:00 | 22.4 | ... |  |  |  |
| AP | 28/10/2016 00:00 | 971.7 | ... |  |  |  |
| RH | 26/11/2016 00:00 | 89.9 | ... |  |  |  |
| WindSpeed | 21/11/2016 00:00 | 4.2 | ... |  |  |  |
| PWAT | 18/10/2016 00:00 | 31.8 | ... |  |  |  |

### Perbandingan *Output* Sistem dengan penelitian terkait

Untuk mempermudah perbandingan *output* dengan penelitian terkait seperti DWP (Putra, 2017), penelitian Ramos-soto (2015), dan lainnya, peneliti menggunakan data pada penelitian DWP (Putra, 2017), yaitu data klimatologi dari stasiun MeteoGalicia. Perbandingan *output* dapat dilihat pada tabel 4.40.

Tabel 4.40Perbandingan *output* dengan penelitian terkait

| Penelitian | Output |
| --- | --- |
| (Kittredge & Driedger, 1994)  *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 |
| (Ramos-Soto et al., 2015) *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 |
| (Gkatzia et al., 2016)  *Output* | -Light rian showers are likely  -Sunny intervals with rain being possible – less likely than not.  -Sunny with rain being unlikely |
| DWP (Putra, 2017) *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. |
| GNG Output #1 | 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. |
| GNG Output #2 | 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. |

Berdasarkan pada tabel 4.40 jumlah konten tentu semakin banyak, namun secara tekstual aplikasi ini tidak sebaik *output* DWP (Putra, 2017) pada penjelasannya, dikarenakan konsep aplikasi ini dibangun untuk data *general* sehingga mampu membangkitkan berita berdasarkan data apapun selama data tersebut mengikuti format data masukan, sedangkan pada penelitian DWP data masukan harus sama dengan yang ada pada penelitian (parameter). Pada penelitian DWP terdapat dua data masukan yaitu klimatologi dan kualitas udara, sehingga konten yang muncul terdapat dua bagian. Sedangkan pada Ramos (2016), teks yang dibangun hanya untuk kualitas udara dan kecepatan angin saja. Pada penelitian Kittredge (1994) terdapat pesan mengenai angin dan salju secara terpisah-pisah sedangkan dalam penelitian Gkatzia (2016) hanya disampaikan terkait keadaan langit. Hal ini menunjukan bahwa D2T yang telah dibangun sebelumnya hanya untuk data yang spesifik, tidak *general*.

## Panduan Penggunaan Aplikasi

Berikut tahapan untuk penggunaan aplikasi:

* + - 1. *Extrack* file D2T.rar kedalam htdocs.
      2. Jalankan Apache pada xampp.
      3. Kunjungi url localhost/D2T/Main/Admin.
      4. Klik Realtime.
      5. Masukan judul berita, dan centang yang diperlukan.
      6. Hasil akan muncul jika dataset telah diberikan sebelumnya.
      7. Masukkan dataset kedalam folder /D2T/DatasetRealTime dengan nama Dataset.csv (lebih tepat jika data berasal dari sensor dan secara otomatis mengirim data kedalam folder ini dengan nama file tersebut). Berita akan terupdate secara otomatis.

## Rancangan Eksperimen

Untuk melakukan evaluasi pembangkitan kalimat bahasa alami oleh suatu sistem diperlukan serangkaian proses diantaranya evaluasi dengan menggunakan *NIST* dan *BLEU*, evaluasi oleh *expert* pada bidang data terkait, evaluasi komputasi, dan evaluasi waktu pembangunan sistem (Belz, 2007). Ada pula cara evaluasi dengan melakukan 20 kali pembangkitan dengan *input* yang berbeda-beda, kemudian setiap hasil pembangkitan dievaluasi oleh *expert* dan dilakukan penilaian. Nilai tersebut didapatkan berdasarkan pertanyaan terkait mengenai *relavance* dan *truthfullnes*, setelah semua hasil dinilai, maka lakukan perhitungan rata-rata sehingga dapat dilakukan penarikan kesimpulan (Ramos-Soto et al., 2016).

Dengan demikian, peneliti memutuskan untuk membuat skenario eksperimen dengan mempertimbangan dua aspek, yaitu:

1. *Readability*, yaitu mengevaluasi kualitas teks dengan menggunakan penilaian *Flesch Reading Ease*. Peneliti memutuskan menggunakan Grammarly (https://www.Grammarly.com) dan *Readability Analyzer* (https://datayze.com/readability-analyzer.php). Penilaian dilakukan untuk semua hasil experimen, kemudian menghitung rata-rata . Hasil dari rata-rata tersebut kemudian diterjemahkan dengan memasukan nilai pada tabel 4.41.
2. *Computation Time*, yaitu mengevaluasi waktu komputasi sistem dengan menggunakan perintah System.time(), sebagai contoh system.time(source(“D2T\_Main.R”)). Perhitungan dilakukan untuk semua hasil experimen, kemudian menghitung rata-rata. Perlu diperhatikan bahwa sebelum melakukan menjalankan system.time() pastikan semua komponen yang dibutuhkan *source code* tersedia, dikarenakan ketika terjadi sedikit saja kesalahan, maka fungsi tersebut tidak akan menghasilkan apapun.

Tabel 4.41  *Flesch Reading Ease*

| Score | School Level | Notes |
| --- | --- | --- |
| 100.0-90.0 | 5th grade | *Very easy to read. Easily understood by an average 11-year-old student.* |
| 90.0–80.0 | 6th grade | *Easy to read. Conversational English for consumers.* |
| 80.0–70.0 | 7th grade | *Fairly easy to read.* |
| 70.0–60.0 | 8th & 9th grade | *Plain English. Easily understood by 13- to 15-year-old students.* |
| 60.0–50.0 | 10th to 12th grade | *Fairly difficult to read.* |
| 50.0–30.0 | College | *Difficult to read.* |
| 30.0–0.0 | College Graduate | *Very difficult to read. Best understood by university graduates.* |

Selain menilai ke-dua aspek tersebut, peneliti juga melakukan analisis terhadap kebenaran informasi yang dibangkitkan dengan cara sebagai berikut:

1. Membandingkan hasil *statistical summary* dari sistemdengan data statistik keseluruhan dengan *microsoft exceli*.
2. Membandingkan *Representative Text* dengan kondisi real, baik itu *trend*, prediksi, atau lainnya.

Kemudian peneliti juga melakukan perbandingan fitur dengan penelitian-penelitian terkait.

Terdapat tiga jenis data yang digunakan, yaitu data kurs beli rupiah (Bank Indonesia), radiasi sinar matahari dan meteorologi interval data satu hari (Plataforma Solar de Almeria) serta klimatologi dan kualitas udara (MeteoGalicia) dan pada eksperimen ini dilakukan prebandingan hasil prediksi dengan penelitian DWP (Putra et al, 2017), terdapat 12 *test-case* yang diberkan kepada sistem seperti pada tabel 4.42.

Tabel 4.42Dataset 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 |

## Analisis dan Hasil Eksperimen

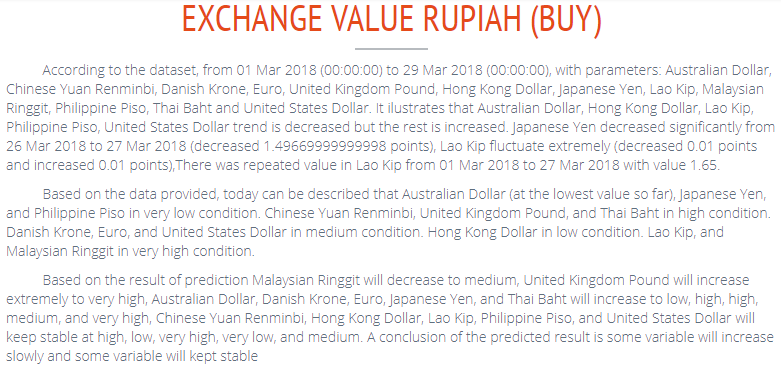
Pada sub-bab ini, akan dijelaskan mengenai hasil eksperimen yang telah dilakukan, serta menganalisis hasil eksperimen tersebut, mulai dari eksperimen dengan data kurs beli rupiah, kurs jual rupiah, sinar radiasi matahari dan meteorologi dalam periode tahunan, sinar radiasi matahari dan meteorologi dalam periode bulanan, hingga dengan menggunakan data klimatologi dan kualitas udara.

### Eksperimen dengan Data Kurs Beli Rupiah

Eksperimen menggunakan data kurs nilai beli rupiah yang didapat dari situs Bank Indonesia yang dibagi menjadi tiga *batch*, yaitu *batch* Maret 2018, *batch* April 2018, dan *batch* Mei 2018.

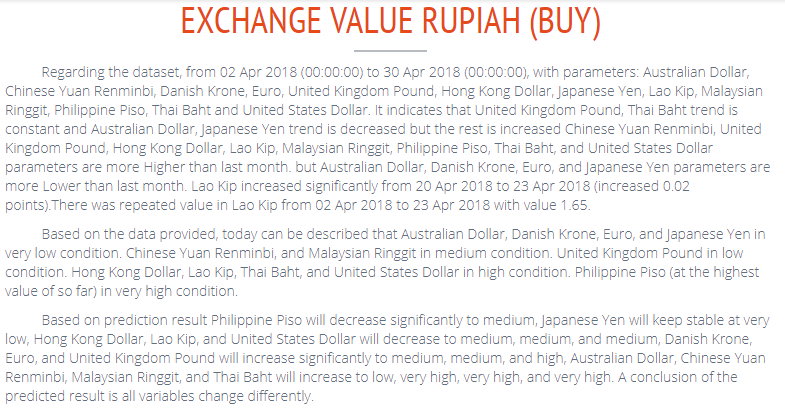
* + - 1. **Hasil Eksperimen**

Hasil *batch* pertama dengan data kurs nilai beli rupiah pada bulan Maret 2018 (KB1) terdapat pada gambar 4.54.



Gambar 4.54 Eksperimen pertama dengan data kurs beli Maret 2018

Dimana dalam informasi pada gambar 4.54 dijelaskan bahwa *trend* Australian Dollar, Hong Kong Dollar, Lao Kip, Philippine Piso, United States Dollar menurun sedangkan parameter lainnya naik. Terdapat beberapa *event* dalam data periode Maret 2018, dimana Japanese Yen turun secara signifikan sebanyak 1.4967 point pada tanggal 26 Maret, terjadi fluktuasi pada Lao Kip secara ekstrim yaitu turun 0.01 point dan naik 0.01 point, *repeated event* yang terjadi pada parameter Lao Kip pada tanggal 1 Maret 2018 hingga 27 Maret 2018, yang artinya nilai tukar rupiah terhadap kip stabil selama satu bulan. Pada paragraf ke-dua terlihat dskripsi setiap parameter, dan terdapat *event* pada parameter Australian Dollar, dimana pada hari ini merupakan nilai tukar paling rendah untuk kategori kurs jual sejauh data yang pernah dianalisis. Pada paragraf ke-tiga disimpulkan bahwa beberapa parameter akan mengalami kenaikan secara perlahan serta beberapa parameter akan tetap stabil seperti kondisi saat ini.



Gambar 4.55 Eksperimen ke-dua dengan data kurs beli April 2018

Untuk *batch* ke-dua dengan data kurs nilai beli rupiah pada bulan April 2018 (KB2) dengan hasil pada gambar 4.55, ada kata baru dalam *trend* yaitu United Kingdom Pound, Thai Baht tetap *constant*, serta beberapa parameter berada dalam kondisi yang lebih tinggi dibandingkan bulan kemarin, tetapi sebagian yang lain berada pada kondisi lebih rendah dibandingkan bulan kemarin. Serta ada kenaikan yang cukup signifikan pada parameter Lao Kip tanggal 20 April, *repeated event* untuk Laos Kip tetap terjadi, artinya selama dua bulan berturut-turut nilai tukar rupiah dan kip stabil. Serta terdapat *event* pada teks data saat ini, yaitu pada saat ini merupakan titik tertinggi nilai tukar piso terhadap rupiah.



Gambar 4.56 Eksperimen ke-tiga dengan data kurs beli Mei 2018

Hasil *batch* ke-tiga dengan data bulan Mei 2018 (KB3) yang terdapat pada gambar 4.56, terlihat bahwa terdapat *Significance Event Message* pada parameter ringgit Malaysia dan baht Thailand, serta nilai beberapa parameter pada bulan ini lebih tinggi dari bulan kemarin tetapi beberapa parameter lainnya lebih kecil dari bulan kemarin. Beberapa parameter pada *Current Text* merupakan nilai terendah pada bulan ini.

* + - 1. **Analisis Hasil Eksperimen**

Analisis eksperimen dengan menggunakan data kurs beli bulan Maret-Mei 2018 adalah sebagai berikut:

1. Analisis aspek *Readability*

Evaluasi dilakukan dengan menggunakan metode *Flesh Reading Ease* dengan menggunakan *Readability Analyzer* yang terdapat pada situs www.datayze.com dan aplikasi Grammarly. Hasil pengujian aspek *Readabilty* dapat dilihat pada tabel 4.43. Dimana rata-rata yang didapatkan adalah 37.885 yang kemudian berdasarkan tabel 4.41 didapat bahwa informasi dapat dipahami oleh kalangan mahasiswa keatas. Untuk hasil lebih lengkap terdapat didalam lampiran.

Tabel 4.43 Hasil evaluasi *Readability* dengan data kurs beli rupiah

| **No** | **Kode Dataset** | ***Flesch Reading Ease Score* (Grammarly)** | ***Flesch Reading Ease Score* (Datayze)** |
| --- | --- | --- | --- |
| **1** | **KB1** | 44 | 42.13 |
| **2** | **KB2** | 42 | 37.07 |
| **3** | **KB3** | 33 | 29.11 |
| **Rata-rata** | | 39.67 | 36.1 |
| **Rata-rata Keseluruhan** | | 37.885 | |

1. Analisis aspek *Computation Time*

Hasil perhitungan *Computation Time* pada data kurs beli rupiah terdapat pada tabel 4.44. Terlihat bahwa rata-rata *running time* sistem ini 2.043 detik. Untuk hasil lebih lengkap terdapat didalam lampiran.

Tabel 4.44 Hasil evaluasi *Computation Time* dengan data kurs beli rupiah

| **No** | **Kode Dataset** | ***Running Time* (*s*)** |
| --- | --- | --- |
| **1** | **KB1** | 2.01 |
| **2** | **KB2** | 1.98 |
| **3** | **KB3** | 2.14 |
| **Rata-rata** | | 2.043 |

1. Analisis kesamaan *Statistical Summary*

Karena data berupa file csv, peneliti melakukan pengecekan data statistik dengan fungsi MAX(), MIN(), dan AVERAGE() dalam *microsoft excel* dengan hasil seperti pada tabel 4.45. Terlihat bahwa nilai minimun dan maksimum data sama dengan aslinya, hanya saja terjadi perbedaan pada rata-rata nilai keseluruhan, hal ini disebabkan pembagian data kedalam *batch* sehingga hasil rata-rata berbeda.

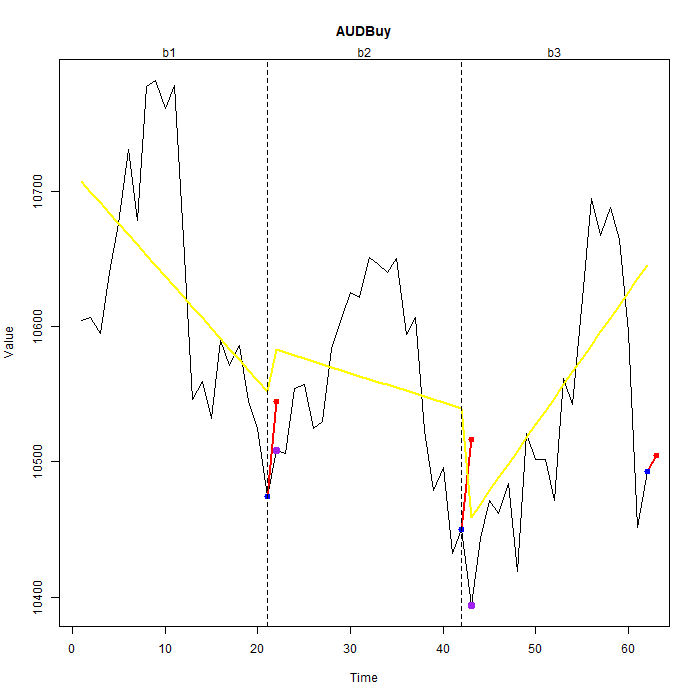
Tabel 4.45 Hasil perbandingan *Statistical Summary* data kurs beli rupiah

| ColName | MaxDate | MaxValue | Max Value Excel | MinDate | MinValue | Min Value Excel | Average | Average Excel | Selisih Average |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| AUDBuy | 13/03/2018 00:00 | 10782.04 | 10782.04 | 02/05/2018 00:00 | 10393.95 | 10393.95 | 10563.93 | 10575.09 | -11.16 |
| CNYBuy | 24/05/2018 00:00 | 2214.81 | 2214.81 | 05/03/2018 00:00 | 2155.26 | 2155.26 | 2183.593 | 2179.673 | 3.92 |
| DKKBuy | 07/03/2018 00:00 | 2281.88 | 2281.88 | 30/05/2018 00:00 | 2161.77 | 2161.77 | 2244.058 | 2251.998 | -7.94 |
| EURBuy | 07/03/2018 00:00 | 16998.36 | 16998.36 | 30/05/2018 00:00 | 16085.62 | 16085.62 | 16714.05 | 16773.41 | -59.36 |
| GBPBuy | 17/04/2018 00:00 | 19641.75 | 19641.75 | 31/05/2018 00:00 | 18458.95 | 18458.95 | 19043.32 | 19110.61 | -67.29 |
| HKDBuy | 24/05/2018 00:00 | 1800.81 | 1800.81 | 27/03/2018 00:00 | 1738.05 | 1738.05 | 1765.136 | 1759.026 | 6.11 |
| JPYBuy | 23/03/2018 00:00 | 130.8052 | 130.8052 | 02/05/2018 00:00 | 126.3762 | 128.393 | 127.98 | 122.509096774194 | 5.470903 |
| LAKBuy | 21/05/2018 00:00 | 1.69 | 1.69 | 27/03/2018 00:00 | 1.64 | 1.64 | 1.663702 | 1.659516 | 0.004186 |
| MYRBuy | 23/05/2018 00:00 | 3551.56 | 3551.56 | 14/05/2018 00:00 | 3486.96 | 3486.96 | 3522.734 | 3521.15 | 1.584 |
| PHPBuy | 22/05/2018 00:00 | 269.67 | 269.67 | 27/03/2018 00:00 | 261.33 | 261.33 | 265.4631 | 264.6802 | 0.7829 |
| THBBuy | 23/04/2018 00:00 | 440.57 | 440.57 | 31/05/2018 00:00 | 433.37 | 433.37 | 437.8685 | 437.9737 | -0.1052 |
| USDBuy | 24/05/2018 00:00 | 14134 | 14134 | 27/03/2018 00:00 | 13639 | 13639 | 13850.49 | 13801.19 | 49.3 |

1. Analisis *Representative Text*

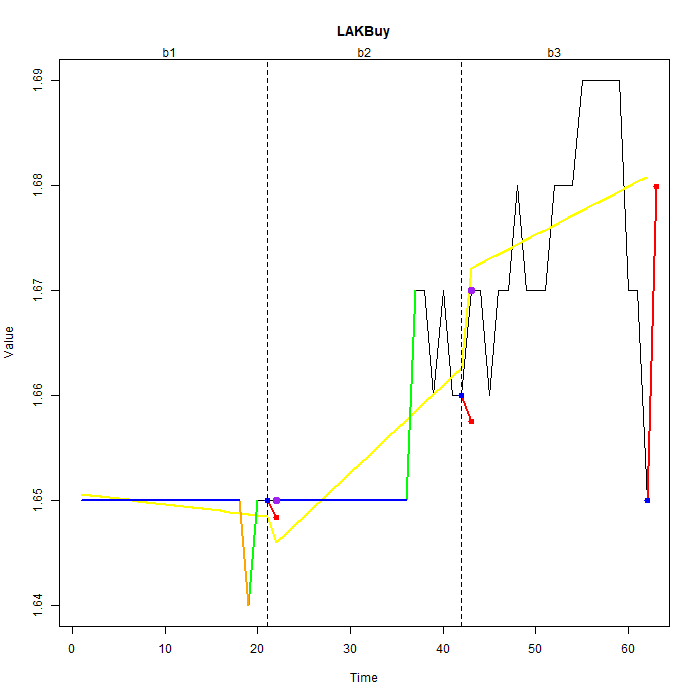
Pada analaisis *Representative Text* dilakukan pembahasan hanya pada beberapa parameter saja, parameter yang dibahas adalah AUDBuy, LAKBuy, JPYBuy, THBBuy, PHPBuy, dan MYRBuy, dengan melihat pada *event* yang terjadi pada parameter tersebut.

Untuk membandingkan *Representative Text* pada informasi dilakukan dengan cara membuat plot untuk data asli (berwarna hitam), plot *trend* dengan model *linear square* setiap *batch* (berwarna kuning), data terakhir/*current* diberi tanda biru, data prediksi *n+1* berwarna merah, data real n+1 dengan warna ungu, *Significance Event* *Message* untuk *extreme event* kenaikan berwarna hijau, penurunan berwarna jingga, *repeated event* dengangarisberwarna biru. Untuk keseluruhan plot terdapat pada lampiran.



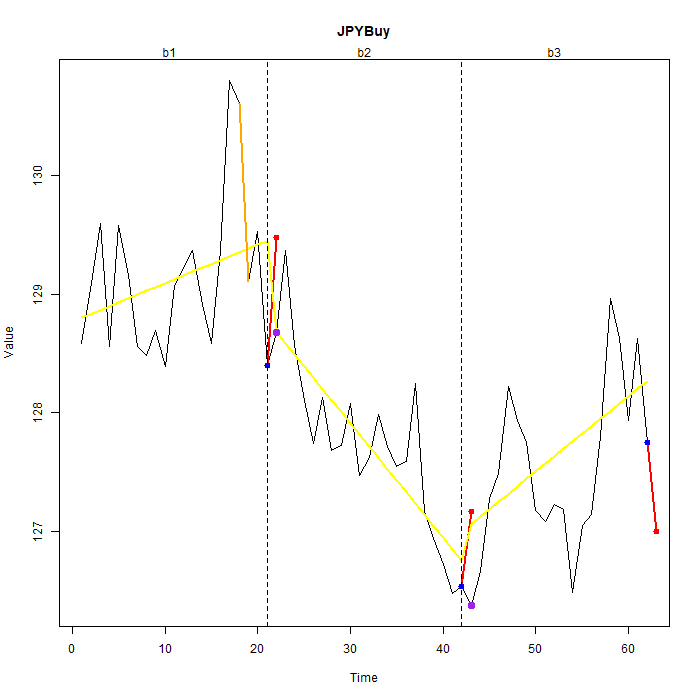
Gambar 4.57 Plot perbandingan *Representative Text* tentang Australian Dollar (Buy)

Dalam teks pada *batch* pertama (bulan Maret 2018) dikatakan bahwa *trend* dolarAustralia menurun hal tersebut dibuktikan dengan *linear* *model* pada batch pertama menurun, serta dikatakan bahwa pada data terakhir *batch* pertama nilai Australian Dollar merupakan data dengan nilai terendah pada sejauh ini (*at the lowest value*), dibuktikan dengan lingkaran biru pada gambar 4.57, karena data perbandingan hanya pada *batch* pertama. Pada *batch* ke-dua dikatakan bahwa dolar Australia menurun dan tetap pada kondisi *very low* namun bukan yang terendah seperti yang terlihat pada gambar 4.57. Pada *batch* ke-tiga dikatakan bahwa dollar Australia mengalami kenaikan, pada paragraf ke-dua dolar Australia dikatakan dalam kondisi *low*.



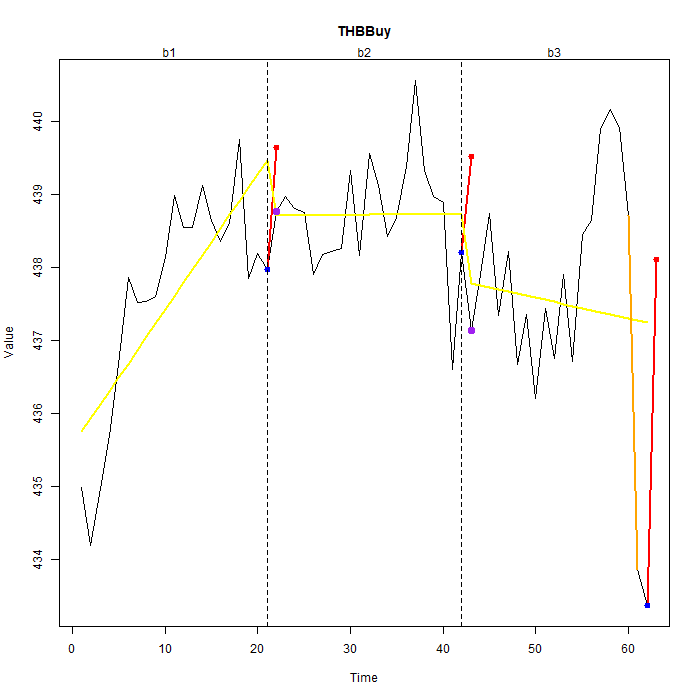
Gambar 4.58 Plot perbandingan *Representative Text* tentang Lao Kip (Buy)

Pada parameter kip Laos, hasil *batch* pertama menunjukan bahwa *trend* menurun, terlihat pada gambar 4.58 memang ada sedikit penurunan pada *batch* pertama, dan terjadi *extreme event* fluktuatif, dikarenakan terdapat *extreme event* naik (garis hijau) dan turun (garis jingga) dalam satu periode waktu, serta *repeat event* yang memenuhi persyaratan sehingga ditampilkan pada teks dengan ditunjukannya garis datar berwarna biru pada beberapa waktu. Pada *batch* ke-dua dikatakan bahwa kip Laos mengalami kenaian, terlihat pada gambar 4.58 memang terjadi kenaikan dan nilai pada *batch* ke-dua lebih tinggi dari *batch* sebelumnya sehingga dalam teks disebutkan *more Higher than last month*, serta masih terjadi *repeated event*. *Batch* ke-tiga menunjukan bahwa tetap terjadi kenaikan dan lebih tinggi dari nilai bulan lalu, dengan ditunjukannya *trend increase* dan kata *more higher than last month*, terlihat pada gambar 4.58.



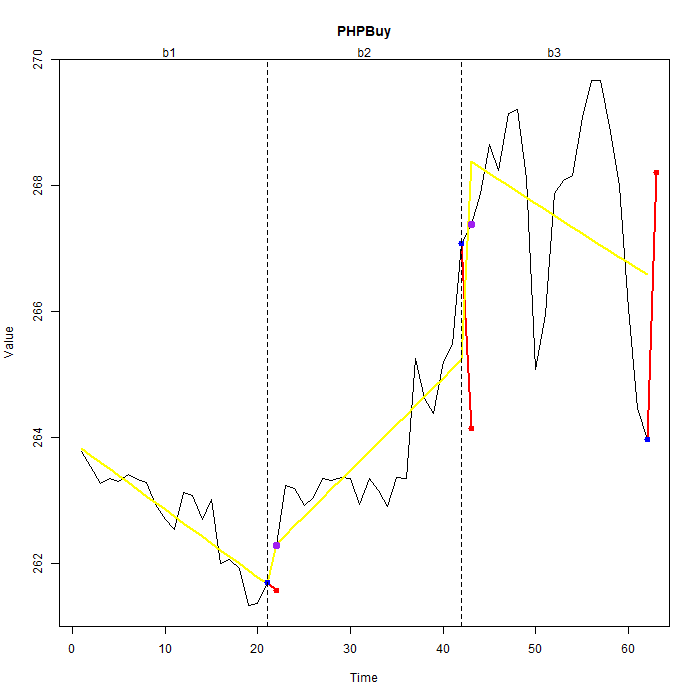
Gambar 4.59 Plot perbandingan *Representative Text* tentang Yen Jepang (Buy)

Terlihat bawha pada *batch* pertama terbukti terdapat penurunan cukup drastis pada tanggal 26 Maret, terlihat ada kenaikan yang cukup tinggi sebelumnya namun tidak dikatakan sebagai *Significance Event Message* karena data naik secara bertahap, *trend* yang dikatakan *increased* sesuai dengan apa yang terdapat pada gambar 4.59. Kemudian pada *batch* ke-dua dikatakan bahwa parameter yen Jepang lebih rendah dari bulan lalu, hal itu dibuktikan dengan posisi grafis *batch* ke-dua lebih rendah dibandingkan *batch* pertama. Pada paragraf ke-tiga dikatakan bahwa *trend* yen Jepang naik, terbukti pada gambar 4.59, namun kenaikan tersebut belum merubah status yen Jepang yang lebih rendah dari bulan lalu, seperti yang terlihat pada grafis *batch* ke-dua dan ke-tiga.



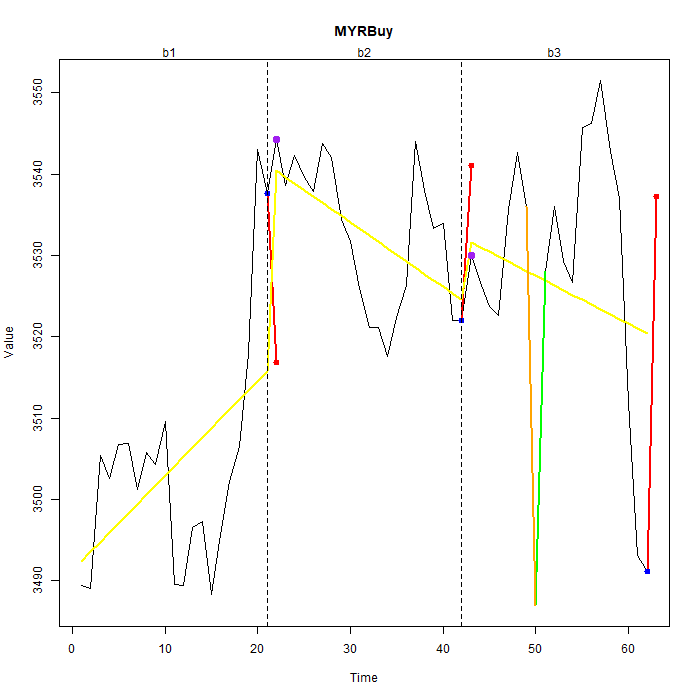
Gambar 4.60 Plot perbandingan *Representative Text* tentang Baht Thailand (Buy)

Dalam teks *batch* pertama dikatakan bahwa *trend* baht Thailand meningkat seperti yang terlihat dalam gambar 4.60, dan tidak ada *event* yang terjadi. Pada paragraf ke-dua dikatakan *trend* baht Thailand *constant* artinya nilai stabil antara naik dan turun seperti pada gambar 4.60 *batch* ke-dua, sedangkan pada paragraf ke-tiga dikatakan *trend* menurun dan nilai baht Thailand lebih rendah dari bulan lalu (*batch* ke-dua) seperti yang terlihat pada gambar 4.60, dan terdaji penurunan yang yang signifikan menjelang akhir bulan, dan pada paragraf ke-dua dikatakan bahwa nilai tukar baht Thailand merupakan nilai terendah sejauh ini, dibuktikan dengan *point* biru pada akhir *batch* ke-tiga pada gambar 4.60.



Gambar 4.61 Plot perbandingan *Representative Text* tentang Peso Filipina (Buy)

*Trend* peso Filipina pada *batch* pertama dikatakan menurun sesuai dengan gambar 4.61, pada *batch* ke-dua *trend* peso Filipina naik, dan pada paragraf ke-dua dikatakan nilai tukar peso Filipina berada pada posisi tertinggi sejauh ini, terbukti dengan melihat pada gambar 4.61 dengan menghilangkan gambar *batch* ke-tiga, karena data *batch* ke-tiga tidak diperhitungkan. Pada *batch* ke-tiga kondisi peso Filipina mengalami penurunan kembali, namun tidak ada penurunan yang termasuk dalam *Significance Event Message*.



Gambar 4.62 Plot perbandingan *Representative Text* tentang Ringgit Malaysia (Buy)

Pada parameter ringgit Malaysia *statistical* *event* terjadi pada *batch* ke-tiga, dimana nilai minimum dan maksimum melebihi *batch* sebelumnya, sehingga dikatakan *more fluctuate than last month* dan terjadi fluktuasi yang ekstrim pada *batch* ke-tiga, seperti pada gambar 4.62.

Analisis keseluruhan mengenai kebenaran *Representative Text* untuk *trend* dan *event* terdapat pada tabel 4.46, untuk *Trend* ke-n dengan nilai “T” sebagai benar, dan “F” menunjukan salah dalam representasi data sedangkan untuk *Event* ke-n dengan nilai *x* sebagai jumlah event yang terjadi pada satu waktu (*extreme event*, *repeated event*, *statistical event¸* dan *highest/lowest point*), dengan nilai maksimum 4 dan minimum 0.

Tabel 4.46 Analisis *trend* dan *event* dengan data kurs beli rupiah

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | AUDBuy | CNYBuy | DKKBuy | EURBuy | GBPBuy | HKDBuy | JPYBuy | LAKBuy | MYRBuy | PHPBuy | THBBuy | USDBuy |
| T1 | T | T | T | T | T | T | T | T | T | T | T | T |
| E1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 0 |
| T2 | T | T | T | T | T | T | T | T | T | T | T | T |
| E2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 3 | 1 | 2 | 1 | 1 |
| T3 | T | T | T | T | T | T | T | T | T | T | T | T |
| E3 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 2 | 1 | 3 | 1 |

Sedangkan untuk hasil prediksi dilakukan hanya pada *batch* pertama dan ke-dua, karena pada *batch* ke-tiga tidak ada data real untuk dibandingkan. Pada parameter dolar Australia *batch* pertama dikatakan bahwa nilai dolar Australia akan naik menjadi *low* sesuai dengan yang ada pada gambar 4.57, namun pada *batch* ke-dua dikatakan bahwa dolar Australia akan naik kembali ke posisi *low*, tapi ternyata data real mengatakan bahwa dolar Australia menurun ke posisi *very low*. Untuk parameter kip Laos pada *batch* pertama diprediksi tetap bertahan pada posisi *very high*, namun karena terjadi perubahan satatistik sehingga merubah interpretasi data menjadi *low*. Pada parameter yen Jepang *batch* pertama dikatakan bahwa akan naik pada posisi *medium*, terlihat bahwa pad kondisi real memang terjadi kenaikan yang masih dalam kategori *medium*.

Analisis keseluruhan prediksi terdapat pada tabel 4.47, dengan adalah prediksi pada *batch* ke-n, dan adalah kondisi real padadata yang diprediksi, dengan tanda warna merah sebagai prediksi yang tidak tepat. Sehingga keseluruhan prediksi pada *batch* pertama dan ke-dua terdapat total 10 benar dan 14 salah.

Tabel 4.47 Analisis prediksi dengan data kurs beli rupiah

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | AUDBuy | CNYBuy | DKKBuy | EURBuy | GBPBuy | HKDBuy | JPYBuy | LAKBuy | MYRBuy | PHPBuy | THBBuy | USDBuy |
| P1 | low | high | High | high | very high | low | medium | very high | medium | very low | very high | Medium |
| R1 | low | high | Medium | medium | medium | low | medium | low | very high | very low | high | low |
| P2 | low | very high | medium | medium | high | medium | very low | medium | very high | medium | very high | Medium |
| R2 | very low | low | Medium | high | low | medium | very low | medium | high | high | Medium | Medium |

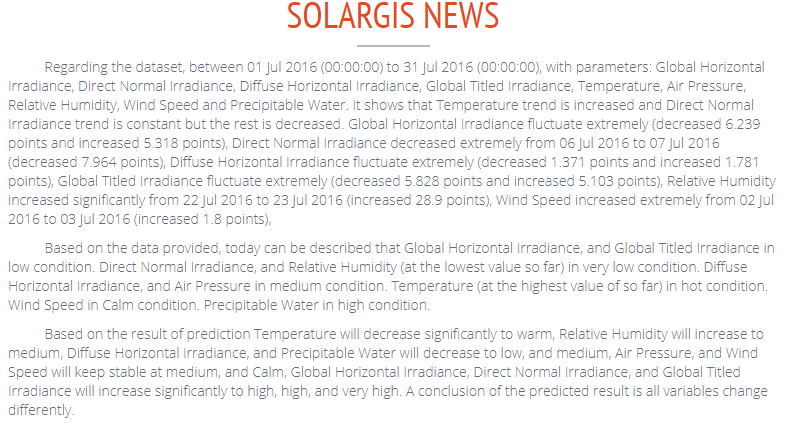


### Eksperimen dengan Data Radiasi dan Meteorologi

Eksperimen menggunakan data radiasi sinar matahari dan meteorologi dari Plataforma Solar de Almeria (PSA) yang didapat melalui situs solargis, dengan mengambil data dari bulan Juli 2016 hingga Desember 2016, kemudian membaginya menjadi 6 *batch* (perbulan).

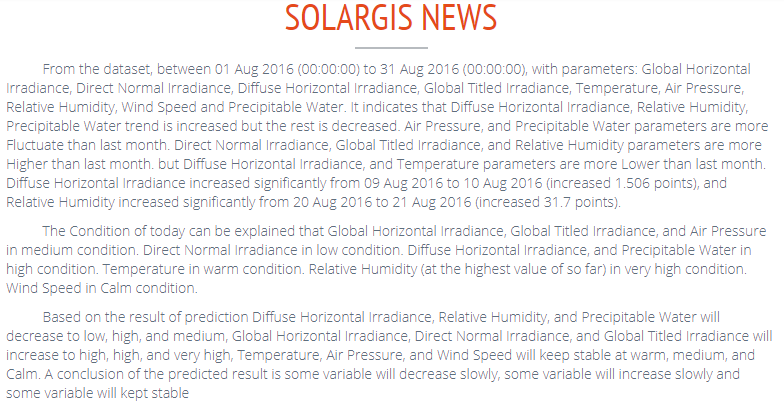
* + - 1. **Hasil Eksperimen**

Hasil *batch* pertama dengan data radiasi sinar matahari dan meteorologi pada bulan Juli 2016 (SD1) terdapat pada gambar 4.63.

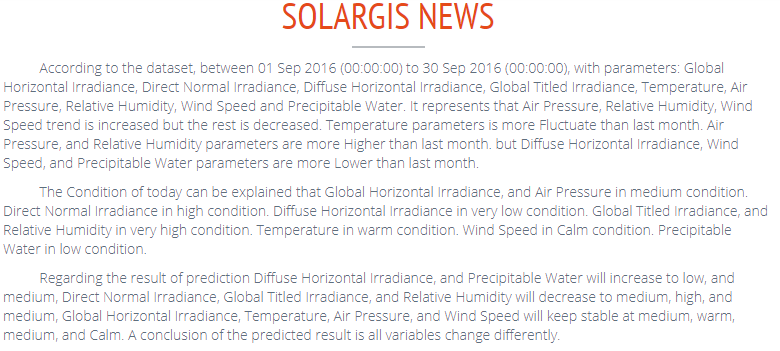


Gambar 4.63 Eksperimen ke-empat dengan data radiasi sinar matahari dan meteorologi Juli 2016

Pada gambar 4.63 dikatakan bahwa *trend* dari parameter suhu naik, pada parameter DNI konstan, sedangkan yang lainnya menurun. Dapat dilihat juga tergapat banyak event yang terjadi, diantaranya fluktuasi yang ekstrim, penurunan yang ekstrim, kenaikan yang signifikan, kenaikan yang ekstrim, nilai terendah pada data terakhir parameter RH, dan nilai tertinggi dari suhu.

Gambar 4.64 Eksperimen ke-lima dengan data radiasi sinar matahari dan meteorologi Agustus 2016

*Batch* ke-dua pada gambar 4.64 tertulis bahwa *trend* DIF, RH, dan PWAT naik, tetapi yang lainnya turun, terlihat juga *Statistical Event* yang terjadi, yaitu dikatakan beberapa parameter dalam kondisi *more fluctuate than last month*, yang artinya nilai minimum dan maksimum *Statistical Summary* berubah. *Event* lainnya yang tertulis adalah parameter DNI, GTI, dan RH dalam kondisi yang lebih tinggi dibandingkan bulan kemarin, tetapi DIF, dan Temperature dalam kondisi yang lebih rendah dibanding bulan kemarin, hal menarik terjadi dimana dikatakan *trend* DIF naik namun lebih rendah dari bulan kemarin. Pada paragraf ke-dua terdapat kondisi dimana RH pada titik tertinggi sejauh ini.

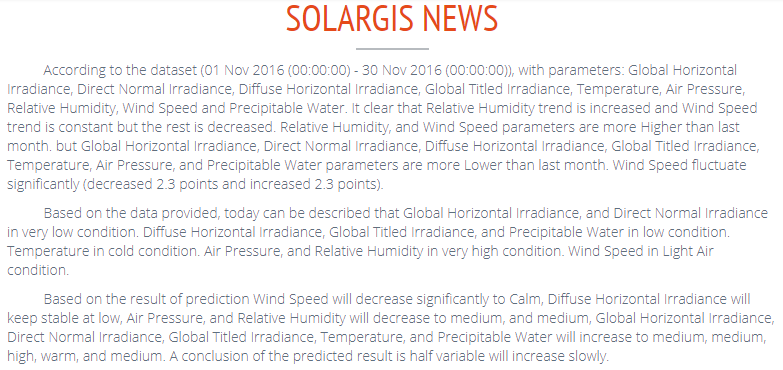
Gambar 4.65 Eksperimen ke-enam dengan data radiasi sinar matahari dan meteorologi September 2016

Pada *batch* ke-tiga pada gambar 4.65, terlihat *event-event* ekstrim menghilang, mungkin penyebabnya adalah nilai minimum dan maksimum yang semakin menjauh sehingga syarat *event* menjadi lebih tinggi. *Trend* AP, RH, dan WindSpeed dikatakan naik, namun parameter lainnya turun. Beberapa parameter dikatakan lebih tinggi dari bulan sebelumnya, dan beberapa parameter juga dalam kondisi lebih rendah dari bulan sebelumnya.



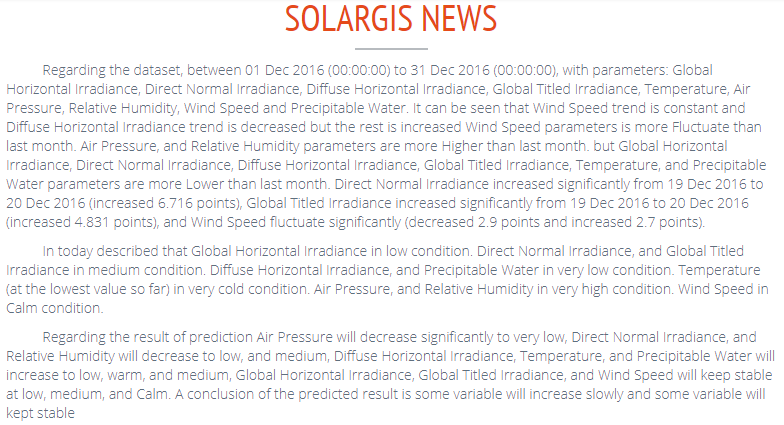
Gambar 4.66 Eksperimen ke-tujuh dengan data radiasi sinar matahari dan meteorologi Oktober 2016

Pada gambar 4.66 terlihat bahwa hal yang sama terjadi pada *batch* ke-empat, dimana *Significance Event Message* tidak ada, namun terjadi *Statistical Event* pada semua parameter, dimana AP mengalami fluktuasi, RH mengalami kenaikan, dan parameter lainnya mengalami penurunan. Pada paragraf ke-dua diceritakan bahwa parameter suhu pada kondisi terendah sejauh ini.



Gambar 4.67 Eksperimen ke-delapan dengan data radiasi sinar matahari dan meteorologi November 2016

Pada eksperimen ke-delapan dengan *batch* ke-lima data radiasi sinar matahari dan meteorologi pada gambar 4.67 dikatakan bahwa *trend* RH naik, WindSpeed konstan, tetapi yang lainnya turun. Beberapa *Statistical Event* kembali ditampilkan, dan terdapat *Significance Event Message* dimana parameter WindSeed mengalami fluktuatif yang signifikan.



Gambar 4.68 Eksperimen ke-sembilan dengan data radiasi sinar matahari dan meteorologi Desember 2016

Pada *batch* terakhir dikatakan bahwa *trend* parameter WindSpeed konstan, DHI menurun, sedangkan parameter lain naik, beberapa *Significance Event Message* yaitu kenaikan signifikan pada parameter DNI dan fluktuasi signifikan pada parameter GTI dan WindSpeed.

* + - 1. **Analisis Hasil Eksperimen**

Analisis eksperimen dengan menggunakan data radiasi sinar matahari dan meteorologi bulan Juli – Desember 2016 adalah sebagai berikut:

1. Analisis aspek *Readability*

Evaluasi dilakukan dengan menggunakan metode *Flesh Reading Ease* dengan menggunakan *Readability Analyzer* yang terdapat pada situs www.datayze.com dan aplikasi Grammarly. Hasil pengujian aspek *Readabilty* dapat dilihat pada tabel 4.48. Dimana rata-rata yang didapatkan adalah 21.935 yang kemudian berdasarkan tabel 4.41 didapat bahwa informasi dapat dipahami oleh kalangan sarjana keatas. Untuk hasil lebih lengkap terdapat didalam lampiran.

Tabel 4.48 Hasil evaluasi *Readability* dengan data radiasi sinar matahari dan meteorologi

| **No** | **Kode Dataset** | ***Flesch Reading Ease Score* (Grammarly)** | ***Flesch Reading Ease Score* (Datayze)** |
| --- | --- | --- | --- |
| **1** | **SD1** | 28 | 19.59 |
| **2** | **SD2** | 29 | 15.23 |
| **3** | **SD3** | 30 | 17.3 |
| **4** | **SD4** | 27 | 15.72 |
| **5** | **SD5** | 23 | 13.09 |
| **6** | **SD6** | 28 | 17.29 |
| **Rata-rata** | | 27.5 | 16.37 |
| **Rata-rata Keseluruhan** | | 21.935 | |

1. Analisis aspek *Computation Time*

Hasil perhitungan *Computation Time* pada data radiasi sinar matahari dan meteorologi terdapat pada tabel 4.49. Terlihat bahwa rata-rata *running time* sistem ini 2.043 detik. Untuk hasil lebih lengkap terdapat didalam lampiran.

Tabel 4.49 Hasil evaluasi *Computation Time* dengan data radiasi sinar matahari dan meteorologi

| **No** | **Kode Dataset** | ***Running Time* (*s*)** |
| --- | --- | --- |
| **1** | **SD1** | 1.84 |
| **2** | **SD2** | 2.08 |
| **3** | **SD3** | 2.08 |
| **4** | **SD4** | 1.98 |
| **5** | **SD5** | 2.03 |
| **6** | **SD6** | 2.25 |
| **Rata-rata** | | 2.043 |

1. Analisis kesamaan *Statistical Summary*

Hasil perbandingan *Statistical Summary* dengan data asli dalam csv terdapat pada tabel 4.50. Terlihat bahwa nilai minimun dan maksimum data sama dengan aslinya, hanya saja terjadi perbedaan pada rata-rata nilai keseluruhan, hal ini disebabkan pembagian data kedalam *batch* sehingga hasil rata-rata berbeda.

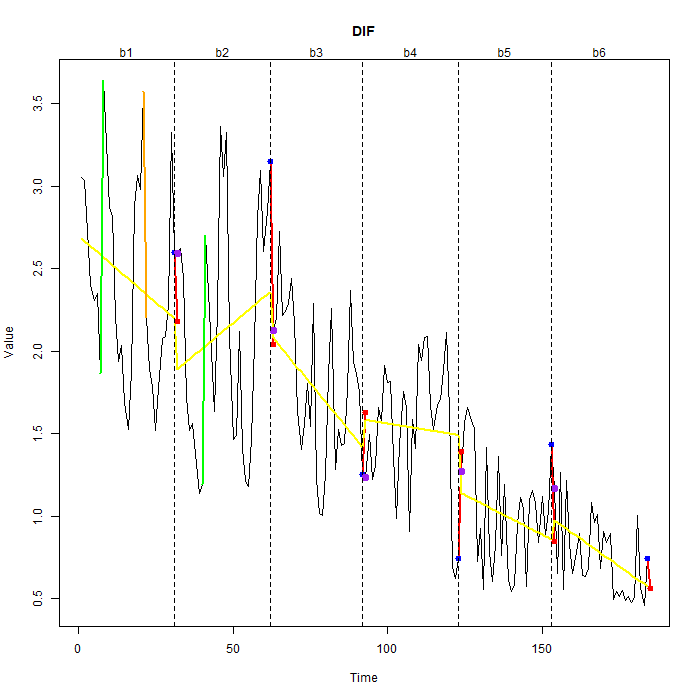
Tabel 4.50 Hasil perbandingan *Statistical Summary* data radiasi sinar matahari dan meteorologi

| ColName | MaxDate | MaxValue | Max Value Excel | MinDate | MinValue | Min Value Excel | Average | Average Excel | Selisih Average |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| GHI | 16/07/2016 00:00 | 8.506 | 8.506 | 04/12/2016 00:00 | 0.567 | 0.567 | 3.208701 | 4.920685 | -1.71198 |
| DNI | 08/08/2016 00:00 | 10.549 | 10.549 | 04/12/2016 00:00 | 0.034 | 0.034 | 4.737943 | 5.719549 | -0.98161 |
| DIF | 08/07/2016 00:00 | 3.643 | 3.643 | 30/12/2016 00:00 | 0.459 | 0.459 | 1.078649 | 1.614201 | -0.53555 |
| GTI | 24/08/2016 00:00 | 8.053 | 8.053 | 04/12/2016 00:00 | 0.537 | 0.537 | 4.747457 | 5.861951 | -1.11449 |
| Temperature | 05/09/2016 00:00 | 30.7 | 30.7 | 31/12/2016 00:00 | 7.2 | 7.2 | 13.55056 | 19.12391 | -5.57335 |
| AP | 24/12/2016 00:00 | 978.1 | 978.1 | 22/11/2016 00:00 | 947.4 | 947.4 | 964.0635 | 962.2495 | 1.814043 |
| RH | 16/12/2016 00:00 | 94.5 | 94.5 | 31/07/2016 00:00 | 21.3 | 21.3 | 72.98348 | 60.69837 | 12.28511 |
| WindSpeed | 18/12/2016 00:00 | 5.3 | 5.3 | 02/12/2016 00:00 | 0.9 | 0.9 | 2.166606 | 2.222826 | -0.05622 |
| PWAT | 27/08/2016 00:00 | 34.1 | 34.1 | 30/12/2016 00:00 | 5.4 | 5.4 | 15.73474 | 19.52609 | -3.79135 |

1. Analisis *Representative Text*

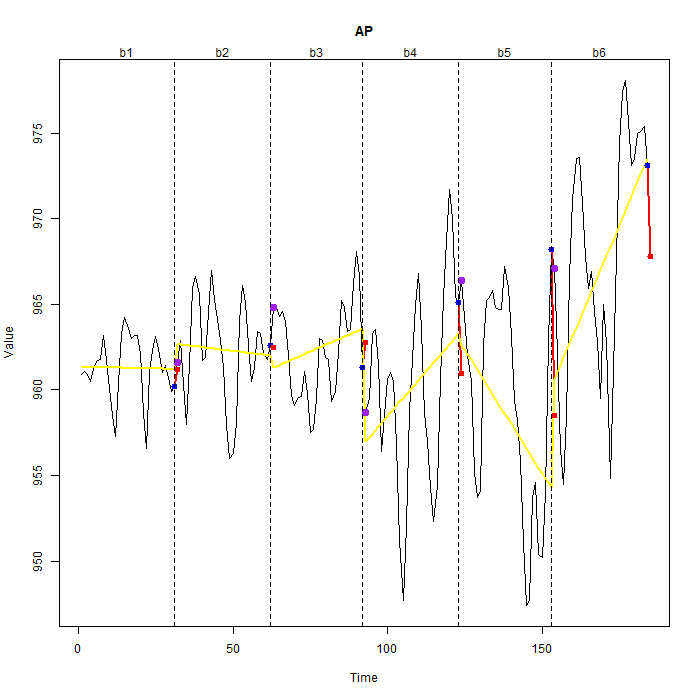
Pembahasan analisis *Representative Text* yang dilakukan sama seperti pada eksperimen sebelumnya, yaitu hanya membahas mengenai beberapa parameter dengan melihat pada *event* yang terjadi pada parameter tersebut. Parameter yang dibahas yaitu parameter DIF, AP, RH, PWAT, dan Temperature.

Sama dengan eksperimen sebelumnya, perbandingan *Representative Text* pada informasi dilakukan dengan cara membuat plot untuk data asli (berwarna hitam), plot *trend* dengan model *linear square* setiap *batch* (berwarna kuning), data terakhir/*current* diberi tanda biru, data prediksi *n+1* berwarna merah, data real n+1 dengan warna ungu, *Significance Event* *Message* untuk *extreme event* kenaikan berwarna hijau, penurunan berwarna jingga, *repeated event* dengangarisberwarna biru. Untuk keseluruhan plot terdapat pada lampiran.



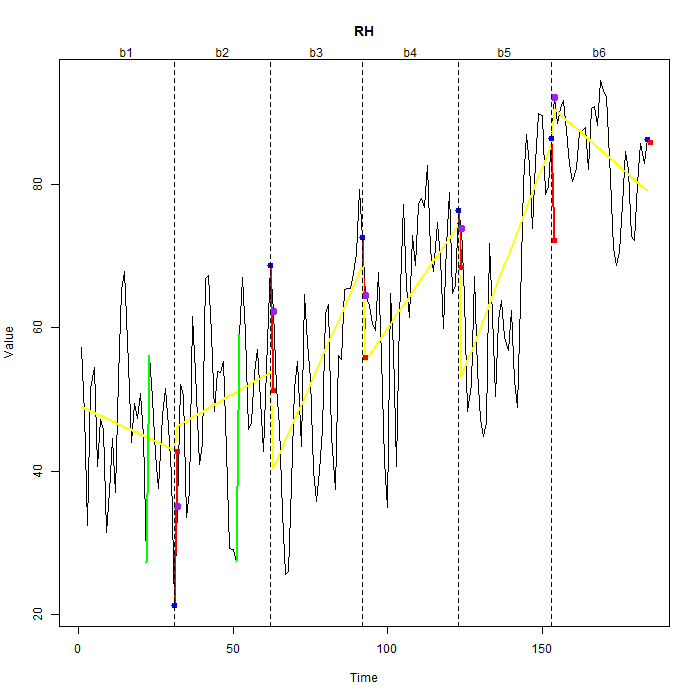
Gambar 4.69 Plot perbandingan *Representative Text* tentang Diffuse Horizontal Irradiance (DIF)

Pada parameter DIF, untuk *batch* pertama dikatakan bahwa *trend* menurun, sesuai dengan gambar 4.68, kemudian pada *batch* pertama juga dikatakan terjadi fluktuasi yang ekstrim, tergambar dengan garis warna jingga dan hijau pada gambar 4.68 *batch* pertama. Pada *batch* ke-dua dikatakan *trend* naik, namun dikatakan bahwa DIF dalam kondisi lebih rendah dari bulan kemarin, hal tersebut memang terjadi seperti pada gambar 4.68. Untuk *trend* pada *batch* selanjutnya dikatakan menurun sesuai dengan gambar 4.68.



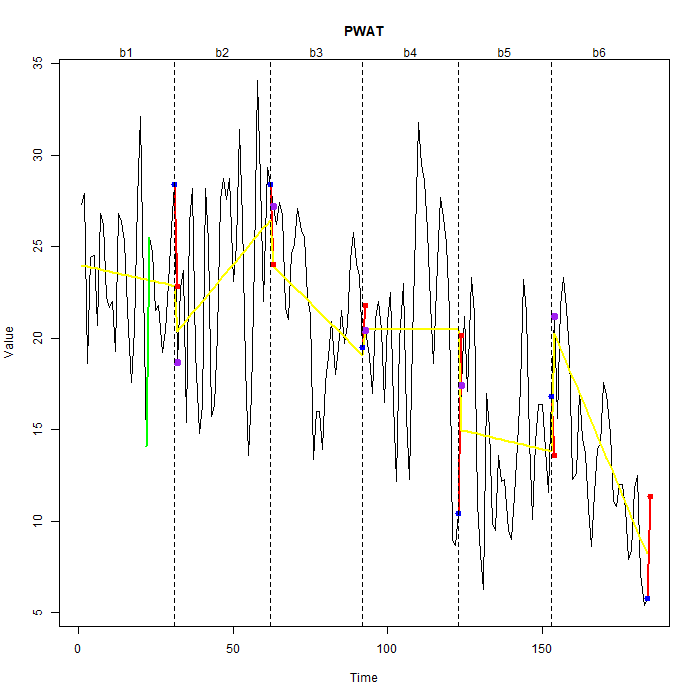
Gambar 4.70 Plot perbandingan *Representative Text* tentang Air Pressure (AP)

Pada *batch* pertama dan ke-dua dikatakan bahwa *trend* AP menurun, padahal terlihat pada gambar 4.70 bahwa *trend* AP konstan, hal ini terjadi karena data keseluruhan sudah muncul, namun bila kita hilangkan *batch* ke-tiga sampai ke-enam, maka kita akan melihat bahwa *trend* AP memang turun, karena toleransi dikatakan konstan bergantung pada nilai minimum dan maksimum data. Pada *batch* ke-dua dikatakan bahwa AP lebih fluktuatif dibandingkan dengan bulan kemarin, seperti yang terlihat pada gambar 4.70, pada *batch* ke-tiga dikatakan bahwa *trend* AP naik dan AP dalam kondisi yang lebih tinggi dibandingkan bulan kemarin, sesuai dengan apa yang ada pada gambar 4.70, pada *batch* ke-empat kembali dikatakan bahwa AP lebih fluktuatif dibandingkan bulan sebelumnya, hal tersebut terlihat dengan nilai minimum dan maksimum data baru diperbaharui pada *batch* ini. Untuk *batch* terakhir dikatakan bahwa *trend* AP naik dan dalam posisi lebih tinggi dari bulan bulan sebelumnya seperti yang terlihat pada gambar 4.70.



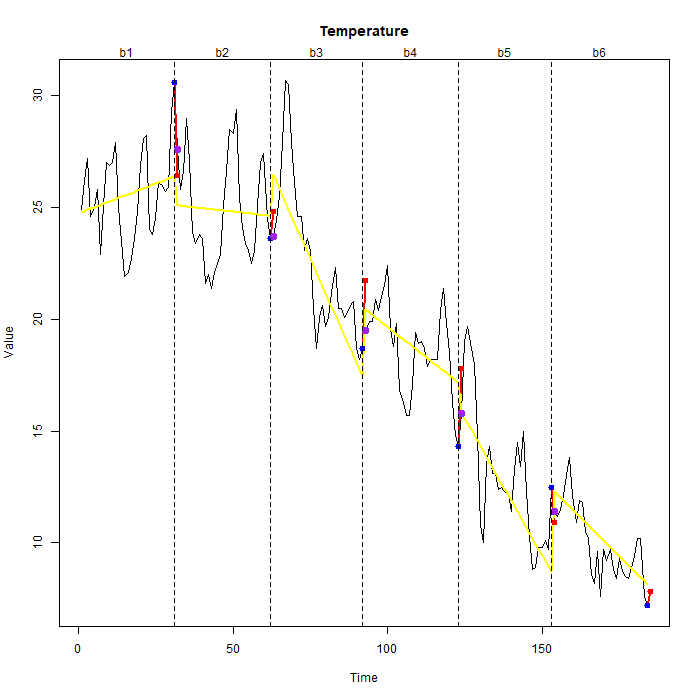
Gambar 4.71 Plot perbandingan *Representative Text* tentang Relative Humidity (RH)

Pada parameter RH *trend* pada *batch* pertama dan ke-enam dikatakan menurun, sedangkan untuk *batch* ke-dua sampai ke-lima dikatakan naik sesuai dengan gambar 4.71. Pada *batch* pertama dituliskan bahwa ada *Significance Event Message* pada tanggal 22 Juli, yaitu kenaikan yang cukup signifikan, sedangkan pada paragraf ke-dua dikatakan posisi RH dalam posisi terendah sejauh ini, terlihat pada gambar 4.71 dengan titik berwarna biru. Pada *batch* ke-dua juga terjadi *Significance Event Message* yaitu kenaikan yang signifikan pada tanggal 20 Agustus, serta pada paragraf kedua dikatakan RH dalam posisi tertinggi sejauh ini, terlihat pada gambar 4.71.



Gambar 4.72 Plot perbandingan *Representative Text* tentang Precipitable Water (PWAT)

Untuk parameter PWAT *trend* dikatakan konstan pada *batch* ke-empat, menurun pada *batch* pertama, ke-tiga, ke-lima, dan ke-enam, sedangkan *trend* dikatakan naik hanya terjadi pada *batch* ke-dua. *Significance Event Message* terjadi pada *batch* pertama dimana terdapat kenaikan yang signifikan. *Statistical Event* terjadi pada *batch* ke-dua, dengan *Representative Text* dikatakan parameter PWAT lebih fluktuatif dibandingkan dengan bulan lalu seperti yang terlihat pada gambar 4.72 dan pada *batch* ke-tiga sampai ke-enam dikatakan PWAT lebih rendah dibandingkan bulan-bulan sebelumnya, artinya ketika pada posisi *batch* ke-enam, parameter PWAT lebih rendah dibandingkan pada *batch* pertama sampai ke-lima.



Gambar 4.73 Plot perbandingan *Representative Text* tentang Temperature

Untuk parameter suhu *trend* dikatakan naik pada *batch* pertama saja, sedangkan *batch* lainnya dikatakan turun, sesuai dengan apa yang terlihat pada gambar 4.73. Tidak ada *Significance Event Message* yang terjadi pada paramter suhu. Pada *batch* pertama paragraf ke-dua dikatakan suhu pada data terakhir berada pada posisi tertinggi sejauh ini, terlihat pada gambar 4.73 dengan mengabaikan data yang belum teranalisis (*batch* ke-dua sampai ke-enam tidak dihitung). Pada eksperimen ke-tujuh dengan data *batch* ke-empat paragraf ke-dua dikatakan posisi suhu berada pada nilai terendah sejauh ini, sesuai dengan gambar 4.73 dengan hanya melihat *batch* pertama sampai ke-empat. Hal yang sama terjadi pada *batch* terakhir dimana pada paragraf ke-dua dikatakan bahwa suhu dalam kondisi paling rendah sejauh ini, terlihat pada gambar 4.73 *batch* ke-enam titik biru merupakan posisi terendah dibandingkan pada *batch* manapun.

Analisis keseluruhan mengenai kebenaran *Representative Text* untuk *trend* dan *event* terdapat pada tabel 4.51, untuk *Trend* ke-n dengan nilai “T” sebagai benar, dan “F” menunjukan salah dalam representasi data sedangkan untuk *Event* ke-n dengan nilai *x* sebagai jumlah event yang terjadi pada satu waktu (*extreme event*, *repeated event*, *statistical event¸* dan *highest/lowest point*), dengan nilai maksimum 4 dan minimum 0.

Tabel 4.51 Analisis *trend* dan *event* dengan data radiasi sinar matahari dan meteorologi

|  | GHI | DNI | DIF | GTI | Temperature | AP | RH | WindSpeed | PWAT |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| T1 | T | T | T | T | T | T | T | T | T |
| E1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 |
| T2 | T | T | T | T | T | T | T | T | T |
| E2 | 0 | 1 | 2 | 1 | 1 | 1 | 3 | 0 | 1 |
| T3 | T | T | T | T | T | T | T | T | T |
| E3 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 |
| T4 | T | T | T | T | T | T | T | T | T |
| E4 | 1 | 0 | 1 | 0 | 2 | 1 | 1 | 1 | 1 |
| T5 | T | T | T | T | T | T | T | T | T |
| E5 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 2 | 1 |
| T6 | T | T | T | T | T | T | T | T | T |
| E6 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 1 |

Sedangkan untuk hasil prediksi dilakukan hanya pada *batch* pertama sampai ke-lima, karena pada *batch* ke-enam tidak ada data real untuk dibandingkan. Analisis keseluruhan prediksi terdapat pada tabel 4.52, dengan adalah prediksi pada *batch* ke-n, dan adalah kondisi real padadata yang diprediksi, dengan tanda warna merah sebagai prediksi yang tidak tepat. Sehingga keseluruhan prediksi pada *batch* pertama dan ke-dua terdapat total 25 benar dan 20 salah.

Tabel 4.52 Analisis prediksi dengan data radiasi sinar matahari dan meteorologi

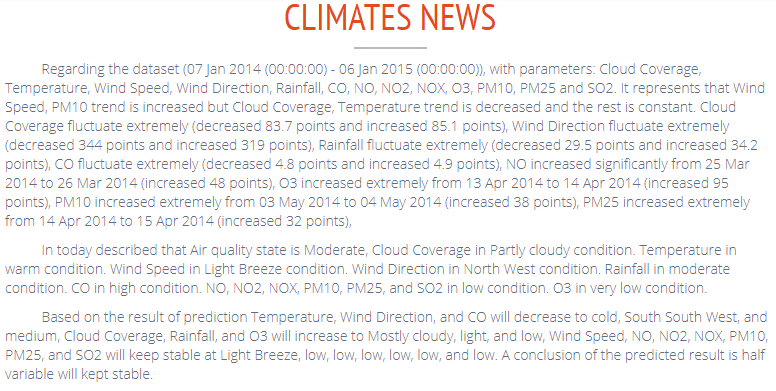
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | GHI | DNI | DIF | GTI | Temperature | AP | RH | WindSpeed | PWAT |
| P1 | high | high | low | very high | warm | medium | medium | Calm | medium |
| R1 | very high | high | medium | very high | hot | medium | low | Calm | low |
| P2 | high | high | low | very high | warm | medium | high | Calm | medium |
| R2 | high | high | medium | very high | warm | high | high | Calm | high |
| P3 | medium | medium | low | high | warm | medium | medium | Calm | medium |
| R3 | medium | high | low | very high | warm | medium | high | Calm | medium |
| P4 | medium | medium | low | high | warm | medium | high | Calm | medium |
| R4 | medium | medium | low | high | cold | high | high | Calm | medium |
| P5 | medium | medium | low | high | warm | medium | medium | Calm | medium |
| R5 | low | low | low | medium | cold | high | very high | Light Air | medium |

### Eksperimen dengan Data Klimatologi dan Kualitas Udara

Eksperimen menggunakan data kualitas udara dari MeteoGalicia yang digunakan pada penelitian DWP (Putra et al, 2017), dengan data tahunan 2014-2015, 2015-2016 dan 2016-2017 (3 *batch*).

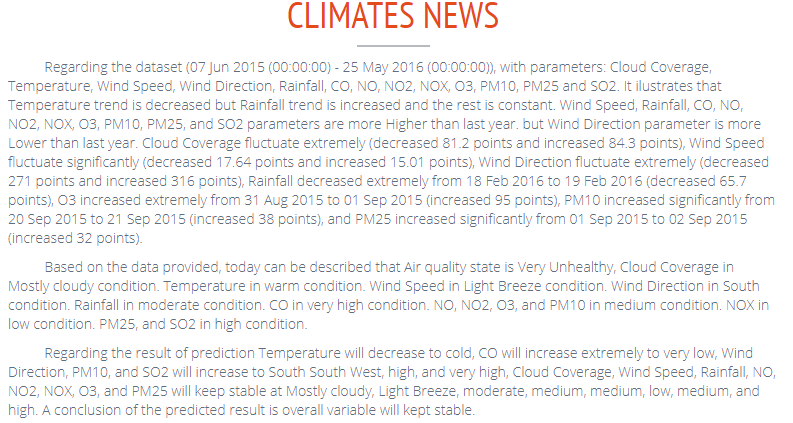
* + - 1. **Hasil Eksperimen**

Hasil *batch* pertama dengan data klimatologi dan kualitas udara pada tahun 2014-2015 (KK1) terdapat pada gambar 4.74.



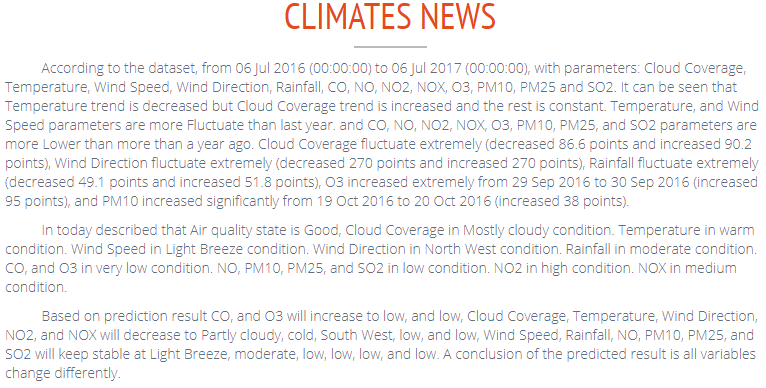
Gambar 4.74 Eksperimen ke-sepuluh dengan data klimatologi dan kualitas udara 2014-2015

Berdasarkan data pada tanggal 07 Januari 2014 hingga 06 Januari 2015, didapat bahwa *trend* WindSpeed dan PM10 naik, namun CloudCoverage dan Temparature turun, sedangkan parameter lainnya dalam keadaan konstan. *Significance Event Message* yang terjadi diantaranya fluktuasi ekstrim, kenaikan signifikan, dan kenaikan ekstrim pada beberapa parameter, dan juga kualitas udara pada hari itu dikatakan *moderate*, munculnya teks kualitas udara ini disebabkan terpenuhinya syarat untuk menghitung nilai PSI, teks keseluruhan dapat dilihat pada gambar 4.74.



Gambar 4.75 Eksperimen ke-sebelas dengan data klimatologi dan kualitas udara 2015-2016

Selanjutnya eksperimen *batch* ke-dua data klimatologi dan kualitas udara menghasilkan teks yang menjelaskan bahwa *trend* Temperature turun namun Rainfall naik, sedangkan parameter lainnya konstan. *Statistical Event* yang terjadi yaitu beberapa parameter lebih tinggi dibandingkan tahun lalu, serta WindSpeed lebih rendah dibandingkan tahun lalu. *Significance Event Message* yang terjadi diantaranya fluktuasi ekstrim, fluktuasi signifikan, penurunan yang ekstrim, kenaikan yang ekstrim, dan kenaikan yang signifikan pada beberapa parameter. Kondisi kualitas udara saat itu dikatakan *very unhelthy*.



Gambar 4.76 Eksperimen ke-duabelas dengan data klimatologi dan kualitas udara 2016-2017

Berdasarkan gambar 4.76 didapat bahwa *trend* Temperature menurun tetapi *trend* CloudCoverage naik, sedangkan parameter lainnya konstan, *Statistical Event Message* yang ditampilkan yaitu adanya fluktuasi data yang lebih besar dibandingkan tahun lalu pada parameter Temperature dan WindSpeed. *Significance Event Message* yang ditampilkan adanya fluktuasi ekstrim, kenaikan yang ekstrim dan kenaikan yang signifikan. Kualitas udara pada hari terkahir dalam data tahunan ini dikatakan *Good*.

* + - 1. **Analisis Hasil Eksperimen**

Analisis eksperimen dengan menggunakan data klimatologi dan kualitas udara pada tahun 2014-2015, 2015-2016, dan 2016-2017 adalah sebagai berikut:

1. Analisis aspek *Readability*

Evaluasi dilakukan dengan menggunakan metode *Flesh Reading Ease* dengan menggunakan *Readability Analyzer* yang terdapat pada situs www.datayze.com dan aplikasi Grammarly. Hasil pengujian aspek *Readabilty* dapat dilihat pada tabel 4.53. Dimana rata-rata yang didapatkan adalah 59.22 yang kemudian berdasarkan tabel 4.41 didapat bahwa informasi dapat dipahami oleh kalangan pelajar Sekolah Menengah Atas dan diatasnya. Untuk hasil lebih lengkap terdapat didalam lampiran.

Tabel 4.53 Hasil evaluasi *Readability* dengan data klimatologi dan kualitas udara

| **No** | **Kode Dataset** | ***Flesch Reading Ease Score* (Grammarly)** | ***Flesch Reading Ease Score* (Datayze)** |
| --- | --- | --- | --- |
| **1** | **KK1** | 60 | 57.77 |
| **2** | **KK2** | 59 | 55.17 |
| **3** | **KK3** | 63 | 60.39 |
| **Rata-rata** | | 60.67 | 57.77 |
| **Rata-rata Keseluruhan** | | 59.22 | |

1. Analisis aspek *Computation Time*

Hasil perhitungan *Computation Time* pada data klimatologi dan kualitas udara terdapat pada tabel 4.54. Terlihat bahwa rata-rata *running time* sistem ini 2.61 detik. Untuk hasil lebih lengkap terdapat didalam lampiran.

Tabel 4.54 Hasil evaluasi *Computation Time* dengan data klimatologi dan kualitas udara

| **No** | **Kode Dataset** | ***Running Time* (*s*)** |
| --- | --- | --- |
| **1** | **KK1** | 2.44 |
| **2** | **KK2** | 2.63 |
| **3** | **KK3** | 2.79 |
| **Rata-rata** | | 2.62 |

1. Analisis kesamaan *Statistical Summary*

Hasil perbandingan *Statistical Summary* dengan data asli dalam csv terdapat pada tabel 4.55. Terlihat bahwa nilai minimun dan maksimum data sama dengan aslinya, hanya saja terjadi perbedaan pada rata-rata nilai keseluruhan, hal ini disebabkan pembagian data kedalam *batch* sehingga hasil rata-rata berbeda.

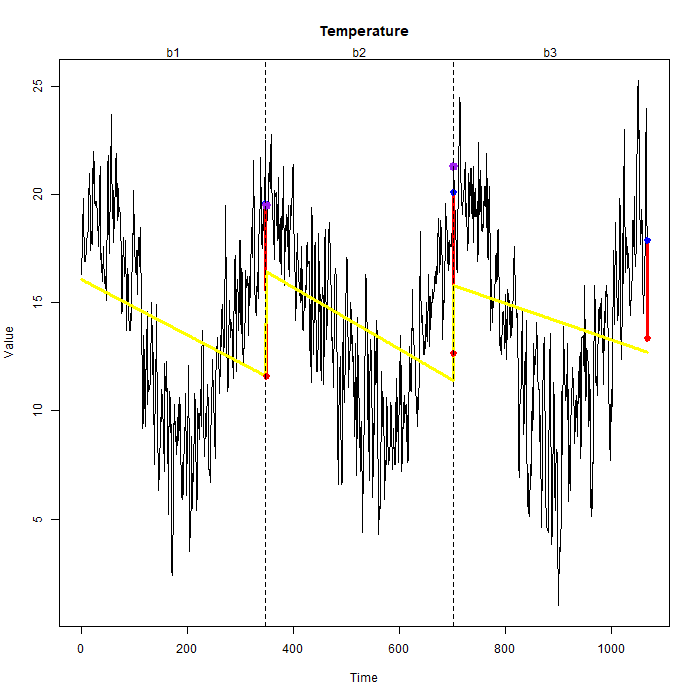
Tabel 4.55 Hasil perbandingan *Statistical Summary* data klimatologi dan kualitas udara

| ColName | MaxDate | MaxValue | Max Value Excel | MinDate | MinValue | Min Value Excel | Average | Average Excel | Selisih Average |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| CloudCoverage | 13/03/2014 00:00 | 100 | 100 | 23/11/2014 00:00 | 7.9 | 7.9 | 51.46901 | 53.15253 | -1.68352 |
| Temperature | 20/06/2017 00:00 | 25.3 | 25.3 | 19/01/2017 00:00 | 1 | 1 | 13.92797 | 13.90281 | 0.025159 |
| WindSpeed | 03/02/2017 00:00 | 29.84 | 29.84 | 28/10/2016 00:00 | 0.68 | 0.68 | 7.111337 | 7.285365 | -0.17403 |
| WindDirection | 18/09/2014 00:00 | 353 | 353 | 27/08/2015 00:00 | -1 | -1 | 220.7252 | 218.9504 | 1.774803 |
| Rainfall | 18/02/2016 00:00 | 67.7 | 67.7 | 09/01/2014 00:00 | 0 | 0 | 2.86441 | 2.960674 | -0.09626 |
| CO | 15/08/2015 00:00 | 22.46 | 22.46 | 28/09/2016 00:00 | 0 | 0 | 6.287033 | 8.260702 | -1.97367 |
| NO | 15/08/2015 00:00 | 88 | 88 | 28/09/2016 00:00 | 0 | 0 | 19.32335 | 21.28839 | -1.96504 |
| NO2 | 15/08/2015 00:00 | 96 | 96 | 28/09/2016 00:00 | 0 | 0 | 31.3927 | 33.31742 | -1.92471 |
| NOX | 15/08/2015 00:00 | 198 | 198 | 28/09/2016 00:00 | 0 | 0 | 50.79857 | 52.61517 | -1.8166 |
| O3 | 01/09/2015 00:00 | 117 | 117 | 28/09/2016 00:00 | 0 | 0 | 38.77772 | 40.91292 | -2.1352 |
| PM10 | 25/09/2015 00:00 | 62 | 62 | 17/09/2016 00:00 | 0 | 0 | 17.97652 | 20.01685 | -2.04034 |
| PM25 | 02/09/2015 00:00 | 54 | 54 | 28/09/2016 00:00 | 0 | 0 | 18.05228 | 20.05056 | -1.99828 |
| SO2 | 08/07/2015 00:00 | 35 | 35 | 28/09/2016 00:00 | 0 | 0 | 10.08824 | 12.11049 | -2.02225 |

1. Analisis *Representative Text*

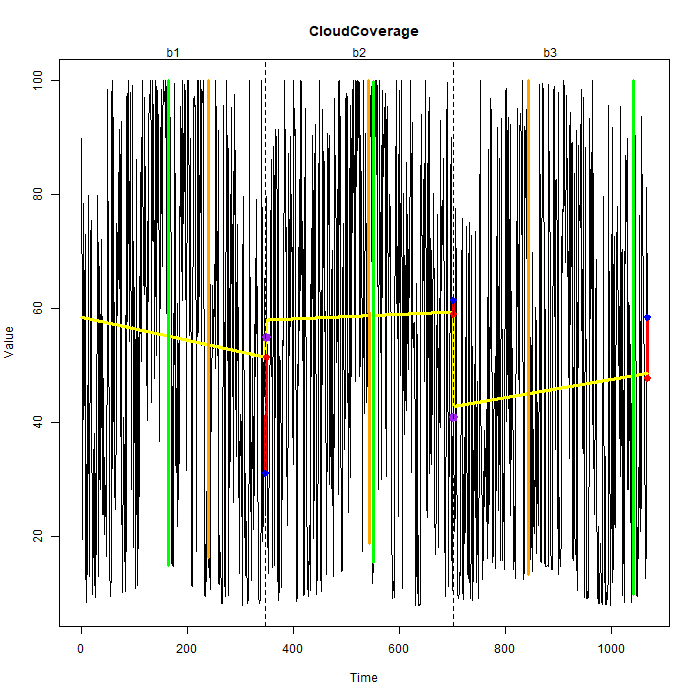
Pembahasan analisis *Representative Text* yang dilakukan sama seperti pada eksperimen sebelumnya, yaitu hanya membahas mengenai beberapa parameter dengan melihat pada *event* yang terjadi pada parameter tersebut. Parameter yang dibahas yaitu parameter Temperature, CloudCoverage, WindSpeed, dan Rainfall.

Sama dengan eksperimen sebelumnya, perbandingan *Representative Text* pada informasi dilakukan dengan cara membuat plot untuk data asli (berwarna hitam), plot *trend* dengan model *linear square* setiap *batch* (berwarna kuning), data terakhir/*current* diberi tanda biru, data prediksi *n+1* berwarna merah, data real n+1 dengan warna ungu, *Significance Event* *Message* untuk *extreme event* kenaikan berwarna hijau, penurunan berwarna jingga, *repeated event* dengangarisberwarna biru. Untuk keseluruhan plot terdapat pada lampiran.



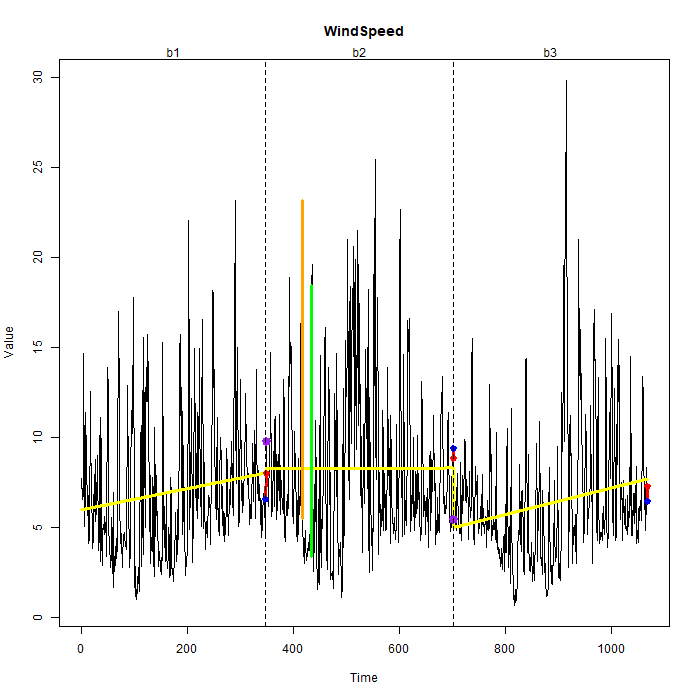
Gambar 4.77 Plot perbandingan *Representative Text* tentang Temperature

*Trend* Temperature dikatakan menurun pada setiap *batch* seperti pada gambar 4.77, pada *batch* ke-tiga dikatakan Temperatur lebih fluktuatif dibandingkan tahun-tahun sebelumnya, nampak pada gambar 4.77 nilai minimum dan maksimum tertinggi terjadi para *batch* ke-tiga.



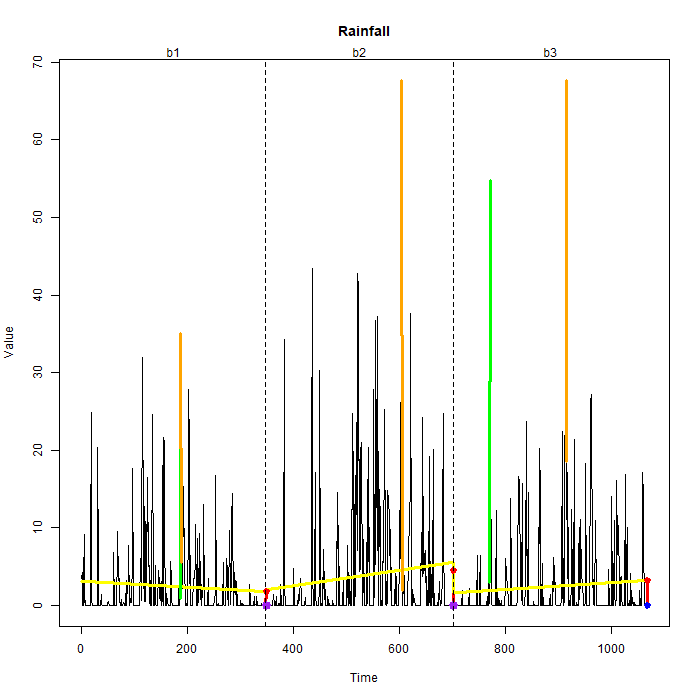
Gambar 4.78 Plot perbandingan *Representative Text* tentang Cloud Coverage (CloudCoverage)

Pada *batch* pertama, *trend* parameter CloudCoverage dikatakan menurun, sedangkan pada *batch* ke-dua dikatakan konstan, dan pada *batch* ke-tiga dikatakan naik, seperti yang tergambarkan dalam gambar 4.78. *Significance Event Message* yang sama muncul pada setiap *batch* yaitu fluktuatif yang ekstrim, sesuai yang tergambar pada gambar 4.78, kenaikan dan penurunan yang mendekati nilai minimum dan maksimum terjadi sangat berdekatan.



Gambar 4.79 Plot perbandingan *Representative Text* tentang Wind Speed (WindSpeed)

Pada *batch* pertama dan ke-tiga, *trend* parameter WindSpeed dikatakan naik, sedangkan pada *batch* ke-dua dikatakan konstan, seperti yang tergambarkan dalam gambar 4.79. *Statistical Event* terjadi pada *batch* ke-dua yaitu dikatakan bahwa parameter WindSpeed berada pada posisi yang lebih tinggi dari tahun kemarin, sedangkan pada *batch* ke-tiga dikatakan bahwa WindSpeed lebih fluktuatif dibandingkan tahun-tahun yang lalu, seperti yang terlihat pada gambar 4.79. *Significance Event Message* muncul pada *batch* ke-dua dimana terjadi fluktuasi yang signifikan seperti terlihat pada gambar 4.79.



Gambar 4.80 Plot perbandingan *Representative Text* tentang Rainfall

Pada parameter Rainfall *batch* pertama dan ke-tiga, *trend* dikatakan konstan, sedangkan pada *batch* ke-dua dikatakan naik, sesuai dengan yang terlihat pada gambar 4.80. Pada *batch* pertama dan ke-tiga ditampilkan *Significance Event Message* yaitu fluktuatif yang ekstrim, sedangkan pada *batch* ke-dua hanya terdapat penurunan yang ekstrim. Pada *batch* ke-dua *Statistical Event* terjadi dimana parameter Rainfall lebih tinggi dari tahun lalu.

Analisis keseluruhan mengenai kebenaran *Representative Text* untuk *trend* dan *event* terdapat pada tabel 4.56, untuk *Trend* ke-n dengan nilai “T” sebagai benar, dan “F” menunjukan salah dalam representasi data sedangkan untuk *Event* ke-n dengan nilai *x* sebagai jumlah event yang terjadi pada satu waktu (*extreme event*, *repeated event*, *statistical event¸* dan *highest/lowest point*), dengan nilai maksimum 4 dan minimum 0.

Tabel 4.56 Analisis *trend* dan *event* dengan data klimatologi dan kualitas udara

|  | CloudCoverage | Temperature | WindSpeed | WindDirection | Rainfall | CO | NO | NO2 | NOX | O3 | PM10 | PM25 | SO2 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| T1 | T | T | T | T | T | T | T | T | T | T | T | T | T |
| E1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 |
| T2 | T | T | T | T | T | T | T | T | T | T | T | T | T |
| E2 | 1 | 0 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 1 |
| T3 | T | T | T | T | T | T | T | T | T | T | T | T | T |
| E3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 |

Sedangkan untuk hasil prediksi dilakukan hanya pada *batch* pertama dan ke-dua, karena pada *batch* ke-tiga tidak ada data real untuk dibandingkan. Analisis keseluruhan prediksi terdapat pada tabel 4.57, dengan adalah prediksi pada *batch* ke-n, dan adalah kondisi real padadata yang diprediksi, dengan tanda warna merah sebagai prediksi yang tidak tepat. Sehingga keseluruhan prediksi pada *batch* pertama dan ke-dua terdapat total 9 benar dan 17 salah.

Tabel 4.57 Analisis prediksi dengan data klimatologi dan kualitas udara

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | CloudCoverage | Temperature | WindSpeed | WindDirection | Rainfall | CO | NO | NO2 | NOX | O3 | PM10 | PM25 | SO2 |
| P1 | Mostly cloudy | cold | Light Breeze | South South West | light | medium | low | low | low | low | low | low | low |
| R1 | Mostly cloudy | warm | Light Breeze | South West | moderate | very high | low | low | low | high | high | medium | high |
| P2 | Mostly cloudy | cold | Light Breeze | South South West | moderate | very low | medium | medium | low | medium | high | high | very high |
| R2 | Partly cloudy | warm | Light Air | North West | moderate | very low | very low | very low | very low | medium | low | low | very low |

### Analisis Perbandingan dengan Penelitian Sebelumnya

Penjelasan perbandingan penelitian ini dengan penelitian sebelumnya dari tabel 4.58 adalah sebagai berikut.

* + 1. Output

Keluaran dari sistem ini terdiri dari tiga paragraf, dimana paragraf pertama menceritakan rangkuman data dalam satu *batch*, paragraf ke-dua menceritakan deskripsi data terkini, dan paragraf ke-tiga memaparkan informasi prediksi. Sehingga dapat dikatakan bahwa dalam penelitian ini menjadi pelengkap dari semua konten penelitian sebelumnya dimana pada DWP (Putra et al, 2017) hanya terdapat 2 paragraf berisikan ringkasan data dan prediksi untuk besok. Sedangkan pada penelitan Ramos-soto (2015) hanya menghasilkan satu kalimat prediksi saja. Pada penelitian Gkatzia (2016), keluaran sistem berupa gambar dan 1 kalimat untuk pilihan kuis dalam game dan 1 kalimat untuk deskripsi sesuai gambar. Pada aplikasi FOG (Kittredge & Driedger, 1994) keluaran sistem berupa beberapa baris kalimat, tergantung pada konten yang ditampilkan. Keseluruhan keluaran sistem menggunakan bahasa Inggris.

* + 1. Realtime

Pada penelitian ini dikembangkan sistem D2T untuk data *streaming* yang menjadi hal baru dibandingkan dengan penelitan lainnya.

* + 1. Waktu data

Pada penelitian ini dilakukan pengembangan untuk membaca data secara dinamis, dalam periode waktu yang berbeda beda, dapat berupa data setiap detik, menit, jam, hari bahkan tahunan. Namun format waktu data harus mengikuti aturan pada sistem GNG.

* + 1. Evaluasi

Evaluasi pada penelitian DWP (Putra et al, 2017) dan Ramos-soto (2016) melakukan evaluasi *Relevance* dan *Truthfulness* dengan memberikan kuisioner kepada *expert* yang berisi pertanyaan mengenai kedua aspek tersebut. Pada penelitian ini dilakukan validasi data dengan membandingkan teks keluaran dengan visualisasi data dan menilai aspek *Readability* dan *Computation Time* seperti pada penelitian DWP (Putra et al, 2017). Sedangkan pada penelitian Gkatzia (2016) evaluasi dilakukan dengan mengevaluasi nilai dari hasil permainan.

Tabel 4.58 Perbandingan dengan penelitian sebelumnya

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Penelitian** | **Output** | **Realtime** | **Ringkas data** | ***Machine Learning*** | **Prediksi** | **Deskripsi data terkini** | **Waktu data** | **Evaluasi** | **Bahasa** | **Pengaplikasian Sistem** | **Readability** |
| **(Kittredge & Driedger, 1994)** | 1 – 11 baris | Tidak | Tidak | *Forecast Production  Assistant* (FPA) | Ya | Tidak | 48 jam | - | Inggris dan Prancis | Ya | 59.61 |
| **(Ramos-Soto et al., 2015)** | N kalimat (tergantung banyak data) | Tidak | Tidak | *Short-term Prediction* | Ya | Ya | General | Kuisioner *expert* (*Relevance & Truthfulness*) | Inggris, Spanyol dan Galician | Ya | 71.14 |
| **(Gkatzia et al., 2016)** | Gamnar dan 2 baris teks | Tidak | Tidak | - | Ya | Ya | Harian | Kuis dalam Game | Inggris | Ya | 47.43 |
| **DWP (Putra, 2017)** | 2 paragraf | Tidak | Ya | *Exponential Smoothing* dan *Gradient Descent* | Ya | Tidak | Tahunan | *Readability*, *Computation Time*, kuisioner expert (*Relevance & Truthfulness*) | Inggris | Tidak | 63.53 |
| **GNG** | 3 paragraf | Ya | Ya | *Piecewise Linear Approximation* dengan *Least Square Method* | Ya | Ya | General | *Readability*, *Computation Time*, dan Validasi Grafis | Inggris | Tidak | 35.425 |

# BAB V KESIMPULAN DAN SARAN

## Kesimpulan

Berdasarkan serangkaian proses penelitian yang telah dilakukan, maka dapat ditarik kesimpulan sebagai berikut:

1. Pengembangan sistem *Data-to-Text* (D2T) untuk data *streaming* dengan menggunakan *Mchine Learning* sangat bermanfaat, dimana sistem dapat bekerja tanpa masukan 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*.
2. Dalam implementasi menggunakan pemrograman bahasa R sebagai inti *Data-to-Text* merupakan pilihan yang tepat, karena sudah banyak *package* yang dapat digunakan dalam mendukung pengembangan sistem *Data-to-Text*.
3. Penelitian ini menjawab kekurangan penelitian sebelumnya yaitu DWP (Putra et al, 2017) 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.
4. 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 *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%.

## Saran

Berikut merupakan saran pada penelitian ini untuk pengembangan lebih lanjut:

1. Pengembangan *corpus* untuk kasus umum, atau menambahkan *corpus-corpus* untuk kasus khusus, seperti pada *Data Interpretation* sehingga dapat meningkatkan nilai *Readability*.
2. Menambahkan fitur untuk mendeteksi keterkaitan parameter seperti *Assosiation Rule*, untuk meningkatkan kualitas *output* sistem.
3. Menggunakan algoritma lain untuk memprediksi data *streaming* untuk meningkatkan akurasi prediksi dan dibandingkan dengan penelitian ini, mengingat akurasi yang dihasilkan masih terbilang cukup rendah.

# DAFTAR PUSTAKA

Athoillah, M., Irawan, M., I., & Imah, Elly, M. (2015). Study Comparison of SVM-, K-NN- and Backpropagation-Based Classifier for Image Retrieval. *Jurnal Ilmu Komputer Dan Informasi (Journal of Computer Science and Information)*, *8*(1), 11–18.

Babcock, B., Babu, S., Datar, M., Motwani, R., & Widom, J. (2002). Models and issues in data stream systems. *Proceedings of the Twenty-First ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems - PODS ’02*, 1. https://doi.org/10.1145/543613.543615

Banaee, H., Ahmed, M. U., & Loutfi, A. (2013). Towards NLG for Physiological Data Monitoring with Body Area Networks. *Proceedings of the 14th European Workshop on Natural Language Generation (ENLG’13)*, 193–197. Retrieved from http://www.aclweb.org/anthology/W13-2127

Bateman, J., & Zock, M. (2012). Natural Language Generation. *The Oxford Handbook of Computational Linguistics*, 1–21. https://doi.org/10.1093/oxfordhb/9780199276349.013.0015

Belz, A. (2007). Probabilistic Generation of Weather Forecast Texts. *Naacl-Hlt*, (April), 164–171. Retrieved from http://www.aclweb.org/anthology/N07-1021

Bowden, R., & Bullington, S. F. (1996). Development of manufacturing control strategies using unsupervised machine learning. *IIE Transactions (Institute of Industrial Engineers)*, *28*(4), 319–331. https://doi.org/10.1080/07408179608966279

Boyd, S. (1998). TREND: A System for Generating Intelligent Descriptions of Time-Series Data. *IEEE International Conference on Intelligent Processing Systems (ICIPS1998)*, 111. https://doi.org/10.1.1.57.3705

Budiharto, W. (2013). *Pengantar Praktis Pemrograman R untuk Ilmu Komputer*.

Castillo-Ortega, R., Marín, N., Martínez-Cruz, C., & Sánchez, D. (2014). A proposal for the hierarchical segmentation of time series. Application to trend-based linguistic description. *IEEE International Conference on Fuzzy Systems*, 489–496. https://doi.org/10.1109/FUZZ-IEEE.2014.6891840

Cliff, K. R., & Billy, K. M. (2017). Estimation of the Parameters of a Linear regression System Using the Simple Averaging Method. *Global Journal of Pure and Applied Mathematics*, *13*(11), 7749–7758.

Crowder, J. W., Moore, J. G., DeRose, L., & Franek, W. J. (1999). Air Pollution Field Enforcement. *STUDENT MANUAL*, 1–5.

Demir, S., Carberry, S., & McCoy, K. F. (2012). Summarizing Information Graphics Textually. *Computational Linguistics*, *38*(3), 527–574. https://doi.org/10.1162/COLI

Domingos, P., & Hulten, G. (2001). Catching up with the data: Research issues in mining data streams. *Workshop on Research Issues in Data Mining and Knowledge Discovery*, *470*. Retrieved from http://www.cs.cornell.edu/johannes/papers/dmkd2001-papers/p8\_domingos.pdf

Fallah-Ghalhary, G. A., Mousavi-Baygi, M., & Nokhandan, M. H. (2009). Annual Rainfall Forecasting by Using Mamdani Fuzzy Interface System. *Research Journal of Environmental Sciences*, *3*.

Gaber, M. M., Zaslavsky, A., & Krishnaswamy, S. (2005). Mining data streams: a review. *ACM Sigmod Record*, *34*(2), 18–26. https://doi.org/10.1145/1083784.1083789

Gama, J., & Gaber, M. M. (2007). *Learning from Data Streams: Processing Techniques in Sensor Networks*. https://doi.org/10.1007/3-540-73679-4\_7

Gatt, A., Portet, F., Reiter, E., Hunter, J., Mahamood, S., Moncur, W., & Sripada, S. (2009). From data to text in the neonatal intensive care Unit: Using NLG technology for decision support and information management. *AI Communications*, *22*(3), 153–186. https://doi.org/10.3233/AIC-2009-0453

Gkatzia, D., Lemon, O., & Rieser, V. (2016). Natural Language Generation enhances human decision-making with uncertain information. *The 54th Annual Meeting of the Association for Computational Linguistics*, 264. Retrieved from http://arxiv.org/abs/1606.03254

Gkatzia, D., Lemon, O., & Rieser, V. (2017). Data-to-Text Generation Improves Decision-Making Under Uncertainty. *IEEE Computational Intelligence Magazine*, *12*(3), 10–17. https://doi.org/10.1109/MCI.2017.2708998

Hallett, C., Power, R., & Scott, D. (2006). Summarisation and visualisation of e-Health data repos- itories Conference Item Repositories. *UK E-Science All-Hands Meeting*, 18–21.

Härdle, W., & Simar, L. (2007). *Applied Multivariate Statistical Analysis.* *Applied Statistics* (Vol. 22007). Berlin: Springer. https://doi.org/10.2307/2347962

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Ullah Khan, S. (2015). The rise of “big data” on cloud computing: Review and open research issues. *Information Systems*, *47*, 98–115. https://doi.org/10.1016/j.is.2014.07.006

Hospital, Z. (2003). Predicting hepatitis B virus – positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning. *Nature Medicine*, *9*(4), 416. https://doi.org/10.1038/nm843

Huby, J. (2010). Cloud Coverage. Retrieved May 20, 2018, from http://www.theweatherprediction.com/habyhints/189/

Hunter, J., Freer, Y., Gatt, A., Reiter, E., Sripada, S., Sykes, C., & Westwater, D. (2011). Bt-Nurse: Computer generation of natural language shift summaries from complex heterogeneous medical data. *Journal of the American Medical Informatics Association*, *18*(5), 621–624. https://doi.org/10.1136/amiajnl-2011-000193

Ihaka, R., & Gentleman, R. (2012). R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, *5*(3), 299–314. https://doi.org/10.1080/10618600.1996.10474713

Käll, L., Canterbury, J. D., Weston, J., Noble, W. S., & MacCoss, M. J. (2007). Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nature Methods*, *4*(11), 923–925. https://doi.org/10.1038/NMETH1113

Kittredge, R. I., & Driedger, N. (1994). Using Natural-Language Processing to Produce Weather Forecasts. *IEEE Expert-Intelligent Systems and Their Applications*, *9*(2), 45–53. https://doi.org/10.1109/64.294135

Kukich, K. (1983). Design of a knowledge-based report generator. *Proceedings of the 21st Annual Meeting on Association for Computational Linguistics -*, 145. https://doi.org/10.3115/981311.981340

Liddy, E. D. (2001). Natural Language Processing. *Natural Language Processing. In Encyclopedia of Library and Information Science, 2nd Ed. NY. Marcel Decker, Inc.*, 1–15.

Mishra, N., & Jain, E. A. (2014). Time Series Data Analysis for Forecasting – A Literature Review. *International Journal of Modern Engineering Research*, *4*(7), 1–5.

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2012). *Foundations of Machine Learning*. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 17). MIT Press. https://doi.org/10.1007/978-3-642-34106-9\_15

Muthukrishnan, S. (2005). Data Streams: Algorithms and Applications. *Foundations and Trends® in Theoretical Computer Science*, *1*(2), 117–236. https://doi.org/10.1561/0400000002

Ng, A. Y., Coates, A., Diel, M., Ganapathi, V., Schulte, J., Tse, B., … Liang, E. (2006). Autonomous inverted helicopter flight via reinforcement earning. *Springer Tracts in Advanced Robotics*, *21*, 363–372. https://doi.org/10.1007/11552246\_35

Paliouras, G., Papatheodorou, C., Karkaletsis, V., & Spyropoulos, C. D. (2002). Discovering user communities on the Internet using unsupervised machine learning techniques. *Interacting With Computers*, *14*(6), 761–791. https://doi.org/10.1016/S0953-5438(02)00015-2

Palpanas, T., Vlachos, M., Keogh, E., Gunopulos, D., & Truppel, W. (2004). Online amnesic approximation of streaming time series. *Data Engineering, 2004. Proceedings. 20th International Conference On*, 339–349. https://doi.org/10.1109/ICDE.2004.1320009

Portet, F., Reiter, E., Gatt, A., Hunter, J., Sripada, S., Freer, Y., & Sykes, C. (2009). Automatic generation of textual summaries from neonatal intensive care data. *Artificial Intelligence*, *173*(7–8), 789–816. https://doi.org/10.1016/j.artint.2008.12.002

Pressman, R. S. (2001). *Software Enginering A Practitioner’s Approach*. (B. Jones & E. Gray, Eds.) (FIFTH EDIT). Palgrave Macmillan.

Putra, B., Riza, L. S., & Wihardi, Y. (2017). Pengembangan Sistem Data-to-Text untuk Membangkitkan Berita Cuaca dengan Pendekatan Time-Series dalam R, 1–39.

Ramos-soto, A., Bugarin, A., & Barro, S. (2016). Fuzzy Sets Across the Natural Language Generation Pipeline. *Progress in Artificial Intelligence*, *5*(4), 261–276.

Ramos-Soto, A., Bugarín, A., & Barro, S. (2016). On the role of linguistic descriptions of data in the building of natural language generation systems. *Fuzzy Sets and Systems*, *285*, 31–51. https://doi.org/10.1016/j.fss.2015.06.019

Ramos-Soto, A., Bugarin, A., Barro, S., Gallego, N., Rodriguez, C., Fraga, I., & A.Saunders. (2015). Automatic Generation of Air Quality Index Textual Forecasts Using a Data-To-Text Approach. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9422*(May), 164–174. https://doi.org/10.1007/978-3-319-24598-0

Reddington, J., & Tintarev, N. (2011). Automatically Generating Stories from Sensor Data. *Proceedings of the 16th International Conference on Intelligent User Interfaces*, (November 2010), 407–410. https://doi.org/10.1145/1943403.1943477

Reiter, E. (1996). Building Natural-Language Generation Systems, 91–93. Retrieved from http://arxiv.org/abs/cmp-lg/9605002

Reiter, E. (2010). Natural Language Generation. *The Handbook of Computational Linguistics and Natural Language Processing*, 574–598. https://doi.org/10.1002/9781444324044.ch20

Reiter, E. (2011). An Architecture for Data-to-Text Systems. *Computational Intelligence*, *27*(1), 23–40. https://doi.org/10.1111/j.1467-8640.2010.00370.x

Riza, L. S. (2015). *Data Science and Big Data Processing in R: Representations and Software*.

Riza, L. S., Nasrulloh, I. F., Junaeti, E., Zain, R., Bayu, A., & Nandiyanto, D. (2016). gradDescentR : An R Package Implementing Gradient Descent and Its Variants for Regression Tasks. *Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, 125–129.

Rowlett, R. (2001). Beaufort Scales (Wind Speed). Retrieved May 20, 2018, from https://www.unc.edu/~rowlett/units/scales/beaufort.html

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, *3*(3), 210–229. https://doi.org/10.1147/rd.33.0210

Schneider, A. H., Mort, A., Mellish, C., Reiter, E., & Wilson, P. (2013). MIME - NLG in Pre-Hospital Care. *Fourteenth European Workshop on Natural Language Generation*, 152–156.

Shannon, C. E. (1950). A Chess-Playing Machine. *Scientific American*, *182*(2), 48–51.

Spector, P. (2004). An Introduction to R. *Statistical Computing Facility*, (x), 1–10.

Sripada, S. G., & Gao, F. (2007). Summarizing dive computer data: A case study in integrating textual and graphical presentations of numerical data. *Proceedings of the Workshop on Multimodal Output Generation (MOG-2007)*, 149–157. Retrieved from http://doc.utwente.nl/67074/1/mog07\_proceedings.pdf#page=157

Sripada, S. G., Reiter, E., Hunter, J., & Yu, J. (2001). A two-stage model for content determination. *Proceedings of the 8th European Workshop on Natural Language Generation-Volume 8*, *8*, 1–8. https://doi.org/10.3115/1117840.1117842

Sripada, S. G., Reiter, E., Hunter, J., & Yu, J. (2003). Generating English summaries of time series data using the Gricean maxims. *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD ’03*, 187. https://doi.org/10.1145/956755.956774

Stone, P., Sutton, R. S., & Kuhlmann, G. (2005). Reinforcement Learning for RoboCup-Soccer Keepaway. *Adaptive Behavior*, *13*(3), 165–188. https://doi.org/10.1177/105971230501300301

Sudrina, Y. (2016). Data Mining : tren Analisa data Berskala Besar Terkait Penelitian Ekologi. *Data Mining*, (Icdm), 1–6. Retrieved from https://www.researchgate.net/publication/307568588\_Data\_Mining\_tren\_Analisa\_data\_Berskala\_Besar\_Terkait\_Penelitian\_Ekologi

Tang, F., Brennan, S., Zhao, Q., & Tao, H. (2007). Co-Tracking Using Semi-Supervised Support Vector Machines. *Computer Vision*, 1–8.

Thomas, K. E., Sripada, S., & Noordzij, M. L. (2012). Atlas.txt: Exploring linguistic grounding techniques for communicating spatial information to blind users. *Universal Access in the Information Society*, *11*(1), 85–98. https://doi.org/10.1007/s10209-010-0217-5

Turian, J., Ratinov, L., & Bengio, Y. (2010). Word representations : A simple and general method for semi-supervised learning. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 384–394.

Turing, A. (1950). Introducci ón a la Inteligencia Artificial. *Intelligence*, *59*, 433–460.

Turner, R., Sripada, S., Reiter, E., & Davy, I. P. (2008). Using spatial reference frames to generate grounded textual summaries of georeferenced data. *Proceedings of the Fifth International Natural Language Generation Conference*, 16–24. https://doi.org/10.3115/1708322.1708328

Williamson, R., & Andrews, B. J. (2000). Gait Event Detection for FES Using Accelerometers and Supervised Machine Learning. *IEEE Transactions on Rehabilitation Engineering*, *8*(3), 312–319.

Ye, Q., Zhang, Z., & Law, R. (2009). Expert Systems with Applications Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Systems With Applications*, *36*(3), 6527–6535. https://doi.org/10.1016/j.eswa.2008.07.035

Yu, J., Reiter, E., Hunter, J., & Mellish, C. (2007). Choosing the content of textual summaries of large time-series data sets. *Natural Language Engineering*, *13*(1), 25–49. https://doi.org/10.1017/S1351324905004031

Zandlo, J., Spoden, G., Bouley, P., & Ruschy, D. (2001). Wind Direction and Degrees. Retrieved May 20, 2018, from http://snowfence.umn.edu/Components/winddirectionanddegreeswithouttable3.htm

Zanero, S., & Savaresi, S. M. (2004). Unsupervised learning techniques for an intrusion detection system. *Proceedings of the 2004 ACM Symposium on Applied Computing - SAC ’04*, 412. https://doi.org/10.1145/967900.967988