

IDENTIFIKASI DATA DRIFTING PADA APLIKASI INTERNET OF THINGS (IoT)

Alessandro Luiz Kartika¹⁾, I Made Murwantara²⁾, Pujianto Yugopuspito^{3*)}

¹⁾Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pelita Harapan
E-mail: ak70029@student.uph.edu¹⁾

²⁾Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pelita Harapan
E-mail: made.murwantara@uph.edu²⁾

³⁾Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pelita Harapan
E-mail: yugopuspito@uph.edu³⁾

**Penulis Korespondensi*

Abstract – Internet of Things has become a high demand technology which also generated massive data that streaming into their collector in a real-time manner. However, the stability of streaming data become crucial as it will influence the analytical processing at the edge. In order to reduce the mismatch of data streaming, this work utilizing concept drift to identify when the stream has begin to inconsistent in certain criteria. To this we do the classification using Stochastics Gradient Descent (SGD). To identify the data drifting, we adopted Adaptive Windowing and Page-Hinkley method. Our result shows that Adaptive Windowing (ADWIN) is more sensitive and outperform the Page-Hinkley method using our dataset.

Keywords: *Internet of Things, Concept Drift, SGD Classifier, Adaptive Windowing (ADWIN), Page-Hinkley*

Abstrak – Internet of Things telah menjadi teknologi yang sangat dibutuhkan dimana juga menghasilkan data dalam skala sangat besar yang dialirkan secara waktu nyata. Stabilitas pengiriman data ke dalam pengumpul menjadi sangat penting karena berpengaruh pada proses analitik dilokasi akhir. Untuk mengurangi kesalahan ini maka penelitian ini mempergunakan metode Concept Drift untuk melakukan identifikasi kapan ketidakteraturan terjadi dengan beberapa kriteria tertentu. Untuk melakukan klasifikasi dipergunakan SGD. Kami mempergunakan metode ADWIN dan Page-Hinkley untuk melakukan identifikasi pergeseran data. Hasil menunjukkan ADWIN lebih sensitive dibandingkan Page-Hinkley dengan mempergunakan dataset yang kami miliki.

Kata Kunci: *Internet of Things, Concept Drift, SGD Classifier, Adaptive Windowing (ADWIN), Page-Hinkley*

PENDAHULUAN

Perkembangan teknologi yang sangat pesat adalah sesuatu hal yang tidak dapat

dihindari didalam kehidupan manusia, terutama di dalam bidang teknologi informasi. Terlihat pada banyaknya

orang yang membuat aplikasi *Internet of Things* (IoT) yang diciptakan dengan tujuan untuk membantu atau memudahkan hidup manusia. *Internet of Things* itu sendiri adalah sebuah sistem dimana benda-benda di dunia fisik dapat dihubungkan ke internet oleh sensor [1].

Dengan banyaknya penggunaan data *real-time stream* pada perangkat IoT yang digunakan sebagai pengambilan keputusan dapat mengalami permasalahan ketika melakukan pengambilan dan pengiriman data sehingga diperlukannya teknik untuk melakukan identifikasi terjadinya permasalahan tersebut.

Concept drift didalam *machine learning* dan *data mining* merupakan perubahan hubungan antara data input dan output seiring dengan perubahan waktu. Hal ini terjadi akibat dari perubahan kondisi sumber data, yang kemungkinan terganggu, dan proses pengiriman data yang menyebabkan pola data berubah secara waktu nyata, dimana kejadian ini harus dimonitor. Hal ini menjadi penting karena perubahan pola dan relasi data secara waktu nyata akan mempengaruhi analisa yang dilakukan sehingga hasil yang diperoleh akan menyimpang secara berkala. Perubahan data yang mengalir ini bisa terjadi dengan perlahan, terkadang stabil dan tidak, hingga terjadi perubahan secara mendadak. Oleh karena itu *Concept Drift* menjadi penting untuk mengatasi kesalahan pembacaan dari *streaming* data yang secara *real-time*.

Identifikasi penyimpangan yang terjadi kemungkinan memiliki pola dan pergerakan data yang berbeda. Untuk itu teknik yang sama belum tentu sesuai dengan kasus berbeda, demikian juga sebaliknya. Semua itu bergantung pada hasil identifikasi yang dilakukan secara sistematis. Untuk dapat mempergunakan metode yang sesuai dengan pola dan

relasi data maka dapat dipergunakan asumsi mengenai pola data. Hal lain yang perlu dilakukan identifikasi adalah tipe perubahan yang terjadi, apakah lambat atau tiba-tiba.

Tulisan ini mempergunakan model atau algoritma seperti *Hoeffding Tree*, *Hoeffding Adaptive Tree*, *SGD Classifier*, *Adaptive Windowing* (ADWIN), dan *Page-Hinkley* memang diperuntukan untuk mengevaluasi data stream dan mengidentifikasi penyimpangan data.

METODE PENELITIAN

Perancangan penelitian ini terbagi dalam tiga tahapan yakni tahapan pengambilan data, tahapan evaluasi data menggunakan metode *Hoeffding Tree*, *Hoeffding Adaptive Tree & SGD Classifier*, tahapan identifikasi penyimpangan data menggunakan metode *Adaptive Windowing* (ADWIN) & *Page-Hinkley*

Pengambilan Data

Pada tahapan awal yang akan dilakukan adalah mengambil data *streaming* dari sensor pada IoT. *Dataset* yang digunakan pada penelitian ini merupakan data yang berkaitan keadaan atau kondisi pada suatu tempat, seperti suhu, tekanan udara, intensitas cahaya, dll. *Dataset* berjumlah 70 ribu *tuple streaming data* yang diambil dalam kurun waktu 2 bulan. Data-data ini diperoleh atau diambil dari suatu sensor yang bernama ClimaStick 2.0. Sensor ini akan menangkap data-data setiap waktu, dan data-data tersebut sudah otomatis tersimpan pada *dashboard* atau aplikasi Thinger.io. Data yang digunakan pada penelitian ini merupakan data yang bersifat *streaming*, maksudnya adalah data-data tersebut diperoleh secara terus menerus.

Evaluasi Data

Setelah mendapatkan data dari tahapan sebelumnya, maka tahapan selanjutnya

adalah mengevaluasi data tersebut. Tahapan ini bertujuan untuk mengevaluasi dan menganalisa apakah data *streaming* yang telah diperoleh memiliki *error rate* yang tinggi atau tidak. Pengujian data ini pun menggunakan dua metode yang dikhususkan untuk data *streaming* yaitu: *Hoeffding tree* yang merupakan suatu *decision tree* yang mampu menguji atau menganalisa data dalam skala yang cukup besar, biasanya pemodelan ini dikhususkan untuk mengolah data yang bersifat *streaming* [2] *Hoeffding tree* ini dikembangkan oleh Domingos dan Hulten dengan tujuan mendesain sebuah *decision tree* yang dapat mengolah dataset yang sangat besar bahkan tak terbatas, tanpa menyimpan dataset tersebut (*online*). *Hoeffding tree* memanfaatkan sampel kecil yang cukup untuk menentukan atribut pemisahan yang optimal, secara matematis *Hoeffding tree* menggunakan *Hoeffding bound* [2].

Hoeffding adaptive tree ini merupakan salah satu metode yang dikembangkan dari *Hoeffding tree*. Metode ini merupakan *decision tree* untuk data *streaming* dengan klasifikasi *adaptive naive bayes* pada cabangnya. Data *streaming* tersebut diuji tanpa mengetahui ukuran/parameter yang tetap dari *sliding window* [3]. Tujuan dari *Hoeffding adaptive tree* memonitor tingkat *error* mayoritas dan disetiap cabangnya memiliki *naive bayes decision*, yang dimana *naive bayes decision* akan digunakan jika cabang-cabangnya lebih akurat pada kasus sebelumnya [2].

Sementara SGD [4] merupakan metode klasifikasi secara linier yang akan memperkirakan *loss* dari setiap sampel disatu waktu dan dengan seiring berjalan model akan selalu diperbarui.

Accuracy [5] merupakan metrik yang menunjukkan pada ketepatan tingkat prediksi benar dengan keseluruhan data.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Dimana TP (*True Positive*) = prediksi yang benar dan kelas positif, TN (*True Negative*) = prediksi yang benar dan kelas negatif, FP (*False Positive*) = prediksi yang salah dan kelas positif, sedangkan FN (*False Negative*) = prediksi yang salah dan kelas negatif.

Precision [6] merupakan metrik yang menunjukkan rasio tingkat prediksi benar positif dibandingkan dengan hasil yang diprediksi positif.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Recall [6] merupakan metrik yang menunjukkan rasio tingkat prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

Kappa [5] merupakan metrik menunjukkan kecocokan (*degree of agreement*) untuk data kualitatif atau menunjukkan kelas-kelas (kelompok data) yang berbeda. Metrik *kappa* biasanya digunakan untuk mengatasi masalah *multi-class* dan *imbalanced class*.

$$Kappa = \frac{p_o - p_e}{1 - p_e} \quad (4)$$

Dimana p_o (*observed agreement*) merupakan *accuracy* prediksi dari suatu metode/model sedangkan p_e (*expected agreement*) merupakan probabilitas kesempatan “*agreement*” dari kedua pengklasifikasian.

Identifikasi *Concept Drift*

Pada tahapan ini akan dilakukan pengidentifikasian penyimpangan (*Concept drift*) pada *dataset* tersebut. Setelah pada tahapan sebelumnya *dataset* tersebut diuji menggunakan kedua metode sebelumnya, dan *dataset* tersebut harus memiliki *error rate* yang rendah tidak lebih dari 0.5. Selanjutnya, pada *dataset*

tersebut akan dilakukan pengidentifikasi-an penyimpangan data (*Concept drift*) dengan menggunakan dua metode yaitu:

Adaptive Windowing (ADWIN) [7] melakukan identifikasi penyimpangan data (*Concept Drift*) menggunakan *adaptive sliding window*. Page-Hinkley (PH) [8] mempergunakan teknik *sequential analysis* yang dapat digunakan sebagai pendeteksi penyimpangan data (*Concept Drift*). Metode ini menghitung nilai *observed* dan rata-ratanya sampai suatu waktu yang ditentukan

HASIL DAN PEMBAHASAN

Bagian ini berisi penjelasan terkait hasil dan analisa dari perancangan penelitian yang telah dibuat pada bagian sebelumnya. Semua hasil dari rancangan penelitian yang telah didapat akan dibahas pada bab ini. Tak hanya itu saja pada bagian ini juga akan memaparkan analisis berdasarkan hasil dari rancangan penelitian

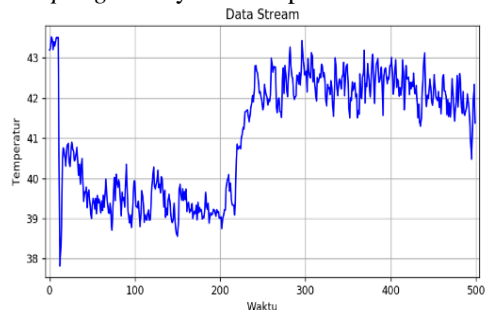
Hasil

Pada bagian ini akan dibagi menjadi beberapa bagian tahapan seperti pada perancangan penelitian pada bab sebelumnya. Mulai dari hasil pada tahapan pengambilan data stream pada perangkat IoT sampai tahapan identifikasi penyimpangan data (*concept drift*).

1). Pengambilan Data

Hasil pada bagian ini merupakan pengambilan data stream pada perangkat IoT sebanyak 70 ribu tuple yang diambil dalam kurun waktu kurang lebih 2 bulan. Data stream yang telah diambil merupakan data yang merekam perubahan suhu atau *temperature* pada suatu ruangan. Gambar 1. Memperlihatkan-

kan gambaran dari data *stream* dengan *sampling* sebanyak 500 tuple data.

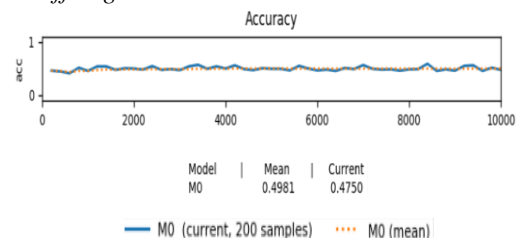


Gambar 1. Data *Stream* pada Perangkat IoT

2). Evaluasi Data

Hasil pada bagian ini memperlihatkan nilai *error rate* dari data *stream* yang telah diambil. Tak hanya itu hasil dari bagian ini memperlihatkan evaluasi performa kinerja dan keefetifkan kedua metode dalam hal *accuracy*, *precison*, *recall*, dan *kappa*. Lalu hasil dari perbandingan kedua metode tersebut akan ditunjukkan.

Ada beberapa *evaluator*/parameter yang digunakan untuk memperlihatkan hasil dari metode *hoeffding tree* seperti 10 ribu *sampling* data yang akan diuji, serta 100 sampel yang digunakan *train* model sebelum memulai mengevaluasi. Terlihat pada Gambar 2 hasil dari evaluasi data menggunakan metode *Hoeffding Tree*.

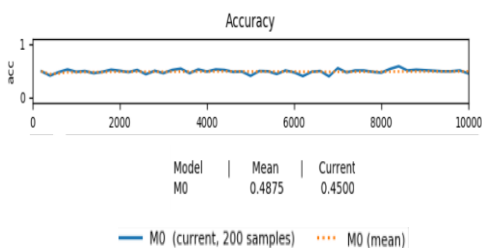


Gambar 2. Hasil *Accuracy* dari Evaluasi Data Metode *Hoeffding Tree*

Dapat terlihat bahwa pada Gambar 2 yang menunjukkan bahwa nilai rata-rata pada metrik *accuracy* dari 10 ribu sampel yang telah dievaluasi adalah sebesar 0,49 sedangkan nilai rata-rata dari 200 sampel terakhir (*current*) adalah sebesar 0.47.

b). *Hoeffding Adaptive Tree*

Bagian ini akan menunjukkan hasil evaluasi data menggunakan metode *hoeffding adaptive tree*. Beberapa *evaluator*/parameter yang digunakan untuk mendapatkan hasil dari metode *hoeffding adaptive tree*, seperti seperti 10 ribu *sampling* data yang akan dievaluasi, serta 100 sample yang digunakan train model sebelum memulai proses evaluasi. Ini menunjukkan hasil dari evaluasi data menggunakan metode *hoeffding adaptive tree*.



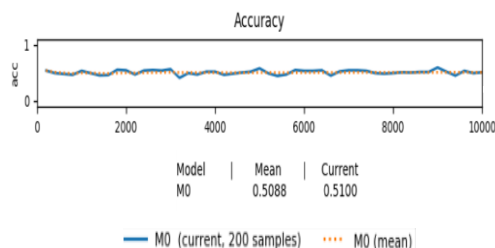
Gambar 3 *Accuracy* dari Evaluasi Data Metode *Hoeffding Adaptive Tree*

Pada Gambar 3, menunjukkan bahwa nilai rata-rata pada metrik *accuracy* dari 10 ribu sampel data *stream* yang telah dievaluasi adalah sebesar 0,48. Sedangkan nilai rata-rata dari 200 sampel terakhir (*current*) yang telah dievaluasi adalah sebesar 0,45.

c). *SGD Classifier*

Pada bagian ini akan menunjukkan hasil evaluasi data *stream* dengan menggunakan metode *sgd classifier*. Beberapa *evaluator*/parameter yang digunakan untuk mendapatkan hasil evaluasi dari metode *sgd classifier*, seperti maksimum sampel data *stream*

yang digunakan sebesar 10 ribu tuple data serta 100 sample yang digunakan untuk *train* model sebelum memulai proses evaluasi. Dibawah ini menunjukkan hasil evaluasi data dari metode *SGD classifier*.

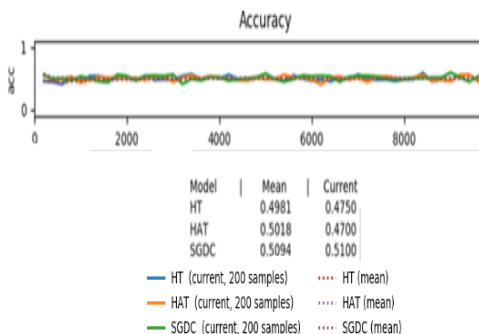


Gambar 4 *Accuracy* dari Evaluasi Data Metode *SGD Classifier*

Pada Gambar 4, menunjukkan bahwa nilai rata-rata pada metrik *accuracy* dari 10 ribu sampel data *stream* yang telah dievaluasi adalah sebesar 0,50. Sedangkan nilai rata-rata dari 200 sampel terakhir (*current*) yang telah dievaluasi adalah sebesar 0,51.

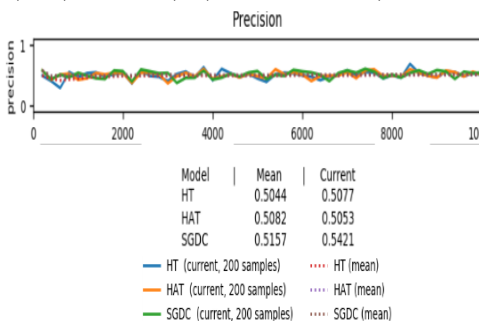
d). Perbandingan Ketiga Metode

Bagian ini akan memperlihatkan hasil evaluasi data *stream* perbandingan dari ketiga metode yaitu *hoeffding tree*, *hoeffding adaptive tree*, dan *SGD classifier*. Beberapa *evaluator*/parameter yang digunakan untuk mendapat hasil perbandingan evaluasi data dari ketiga metode, seperti maksimum sampel data yang digunakan sebesar 10 ribu tuple data serta 100 sample yang digunakan untuk *train* model sebelum memulai evaluasi.



Gambar 5 Hasil *Accuracy* dari Evaluasi Data Ketiga Metode

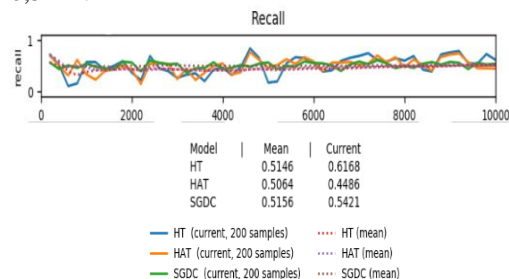
Pada Gambar 5, menunjukkan bahwa nilai rata-rata pada metrik *accuracy* dari ketiga metode dengan 10 ribu sampel data *stream* yang telah dievaluasi, yang masing-masing metode memiliki adalah sebesar HT = 0,4981, HAT = 0,5018, dan SGDC = 0,5094. Sedangkan, nilai rata-rata dari 200 sampel terakhir (*current*) yang telah dievaluasi dari ketiga metode masing masing adalah sebesar HT = 0,475, HAT = 0,47, dan SGDC = 0,51



Gambar 6 Hasil *Precision* dari Evaluasi Data Ketiga Metode

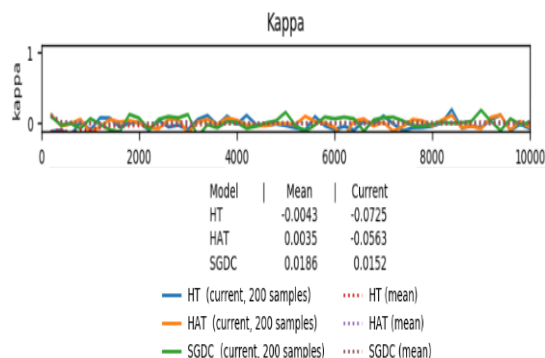
Pada Gambar 6, menunjukkan bahwa nilai rata-rata pada metrik *precision* dari ketiga metode dengan 10 ribu sampel data *stream* yang telah dievaluasi, yang masing-masing metode memiliki adalah sebesar HT = 0,5044, HAT = 0,5082, dan SGDC = 0,5157. Sedangkan, nilai rata-rata dari 200 sampel terakhir (*current*) yang telah dievaluasi dari ketiga metode masing-masing adalah sebesar HT =

0,5077, HAT = 0,5053, dan SGDC = 0,5421.



Gambar 7 Hasil *Recall* dari Evaluasi Data Ketiga Metode

Pada Gambar 7, menunjukkan bahwa nilai rata-rata pada metrik *recall* dari ketiga metode dengan 10 ribu sampel data *stream* yang telah dievaluasi, yang masing-masing metode memiliki adalah sebesar HT = 0,5146, HAT = 0,5064, dan SGDC = 0,5156. Sedangkan, nilai rata-rata dari 200 sampel terakhir (*current*) yang telah dievaluasi dari ketiga metode masing masing adalah sebesar HT = 0,6168, HAT = 0,4486, dan SGDC = 0,5421.



Gambar 8 Hasil *Kappa* dari Evaluasi Data Ketiga Metode

Pada Gambar 8, menunjukkan nilai rata-rata pada metrik *kappa* dari ketiga metode dengan 10 ribu sampel data *stream* yang telah dievaluasi, yang masing-masing metode memiliki adalah sebesar HT = -0,0043, HAT = 0,0035, dan SGDC = 0,186. Sedangkan, nilai

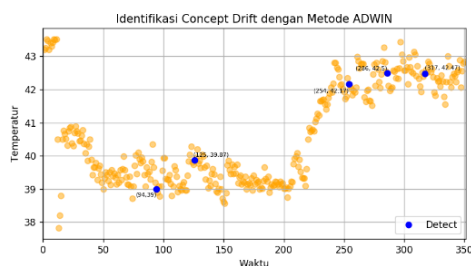
rata-rata dari 200 sampel terakhir (*current*) yang telah dievaluasi dari ketiga metode masing masing adalah sebesar $HT = -0,0725$, $HAT = -0,0563$, dan $SGDC = 0,0152$.

3). Identifikasi Data

Pada bagian ini akan menunjukkan hasil dari identifikasi penyimpangan data (*concept drift*) dari kedua metode yaitu *Adaptive Windowing* (ADWIN) dan *Page-Hinkley*.

a). ADWIN

Pada bagian ini akan menunjukkan hasil identifikasi penyimpangan data (*concept drift*) menggunakan metode (ADWIN). Sampel data yang digunakan sebanyak 350 tuple data, serta diasumsikan bahwa indeks data merupakan waktu (menit) ketika data *stream* tersebut terekam.



Gambar 9 Hasil Identifikasi *Concept Drift* dengan Metode ADWIN

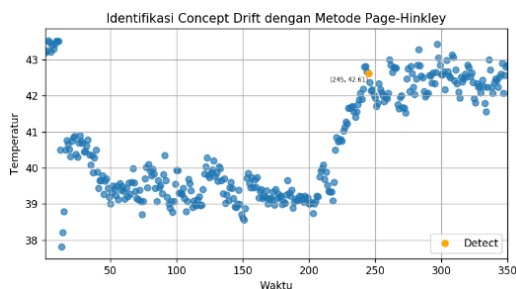
Mtd	Pre-train size	Accura	Precision	Recall
HT	100	0,498	0,504	0,514
	200	0,498	0,504	0,512
	300	0,499	0,505	0,512
HAT	100	0,496	0,502	0,518
	200	0,494	0,500	0,465
	300	0,496	0,503	0,531
SGD	100	0,509	0,515	0,515
	200	0,509	0,515	0,515
	300	0,509	0,516	0,516

Dapat terlihat pada Gambar 9, bahwa terjadi penyimpangan data (*concept drift*) yang terdeteksi pada indeks 94, 125, 254,

286, dan 317 ditandai dengan titik berwarna biru. Selain itu, sumbu x pada Gambar 9 tersebut merupakan *range* dari waktu, sedangkan sumbu y merupakan *range* dari *temperature*.

b). *Page-Hinkley*

Bagian ini akan menunjukkan hasil dari identifikasi penyimpangan data (*concept drift*) dengan metode *page-hinkley*. Sampel data yang digunakan sebanyak 350 tuple data, serta diasumsikan bahwa indeks data merupakan waktu (menit) ketika data *stream* tersebut terekam



Gambar 10 Hasil Identifikasi *Concept Drift* dengan metode *Page-Hinkley*

Dapat terlihat pada Gambar 10, bahwa penyimpangan data (*concept drift*) terdeteksi pada indeks 243 dan 325 ditandai dengan titik berwarna oranye. Selain itu, sumbu x pada Gambar 10 tersebut merupakan range dari waktu, sedangkan sumbu y merupakan *range* dari *temperature*.

Analisa

Tabel 1. Hasil Evaluasi HT, HAT dan SGD

Dapat terlihat pada Tabel 1, bahwa hasil evaluasi dari *hoeffding tree* dan *SGD Classifier* cenderung mengalami peningkatan seiring dengan penambahan parameter (*pre-train*) yang digunakan untuk melatih (*train*) model sebelum memulai evaluasi. Lalu total waktu (*train + test*) dan memori yang digunakan dari *hoeffding tree* dan *SGD Classifier*

cenderung stabil. Pada metode *hoeffding tree* hal tersebut terjadi karena

Indeks	Metode		Indeks	Metode		
	AW	PH		AW	PH	
0 - 2000	95	246	68000 -70000	68511	68356	
	127					
	255					
	287					
	318					
	575					
	607					
	639					
	735					
	799					
	831	775		68991	68442	
	991			69023		
	1119			957	69087	68680
	1151					
	1215					
	1279					
	1311					
	1343					
	1407					
	1439					
	1663					
	1695	69785			69919	
	1855					
	1887					
	1919					
	1951			69951		

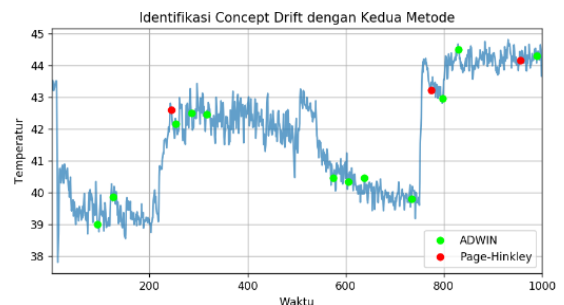
disebabkan *hoeffding tree* merupakan metode *incremental* yang dimana data atau sampel (*pre-train*) digunakan untuk memperluas pengetahuan model, oleh karena itu hasil metrik-metrik dari metode *hoeffding tree* juga cenderung meningkat. Begitu juga dengan *SGD Classifier* yang pada penelitian ini digunakan untuk mendemonstrasikan *compability* dengan metode *incremental*, oleh karena itu hasil metrik-metrik dari metode *SGD Classifier* juga cenderung meningkat. Dari hasil evaluasi dari metode *SGD Classifier* menunjukkan hasil yang paling tinggi disetiap metrik dibandingkan dengan kedua metode yang lain. Sedangkan pada metode *hoeffding*

adaptive tree, hasil evaluasi disetiap metrik menunjukkan ketidakstabilan seiring dengan peningkatan pada jumlah sampel (*pre-train*), serta total waktu (*train + test*) dan memori digunakan cenderung meningkat. Hal ini disebabkan karena metode *hoeffding adaptive tree* tidak memiliki ukuran yang tepat untuk menentukan *sliding window* dan ukuran tersebut sulit ditebak atau tergantung dari tingkat perubahan distribusi dari dataset tersebut.

Tabel 2, menunjukkan hasil identifikasi *concept drift* dari kedua metode *ADWIN* dan *page-hinkley* terhadap data *stream* sebanyak 70 ribu tuple data. Selain itu dapat terlihat bahwa metode *ADWIN* lebih banyakk mengidentifikasi *concept drift* dibandingkan dengan metode *Page-Hinkley*.

Tabel 2 Hasil Identifikasi *Concept Drift* dari Kedua Metode

Sementara pada gambar 11 menunjukkan hasil identifikasi *concept drift* dari kedua metode *ADWIN* dan *Page-Hinkley* pada 1000 tuple data *stream*.



Gambar 11 Identifikasi *Concept Drift* dari Kedua Metode

Terlihat pada Gambar 11, bahwa kedua metode memiliki perbedaan hasil dalam mengidentifikasi *concept drift*. Metode *ADWIN* dapat mengidentifikasi 12 penyimpangan data sedangkan metode *Page-Hinkley* dapat mengidentifikasi 3

penyimpangan data. Selain itu metode ADWIN cenderung mengidentifikasi penyimpangan data ketika terjadi kenaikan temperatur. Berbeda dengan metode ADWIN, metode *Page-Hinkley* cenderung mengidentifikasi penyimpangan data ketika terjadi penurunan temperatur.

KESIMPULAN

Penelitian ini dapat disimpulkan bahwa pada penelitian ini data *stream* yang digunakan mengalami perubahan data berdasarkan waktu secara bertahap sehingga membentuk pola *gradual*. Selain itu, berdasarkan hasil evaluasi data pada penelitian ini yang menggunakan ketiga metode yakni *Hoeffding Tree*, *Hoeffding Adaptive Tree*, dan *SGD Classifier* dan ketiga metode tersebut memiliki perbandingan antara nilai prediksi yang benar dengan nilai prediksi yang salah hampir sama. Metode *SGD Classifier* memiliki hasil nilai rata-rata yang paling tinggi pada setiap metrik yakni *accuracy*, *precision*, *recall*, dan *kappa*. Selain itu metode *SGD Classifier* paling sedikit dalam menggunakan memori. Pada hasil identifikasi penyimpangan data/*concept drift* dari kedua metode yakni ADWIN dan *Page-Hinkley*, metode ADWIN yang mengidentifikasi penyimpangan data/*concept drift* yang paling banyak dibandingkan dengan metode *Page-Hinkley*, tetapi metode ADWIN cenderung mengidentifikasi penyimpangan data/*concept drift* ketika terjadi peningkatan temperatur. Sedangkan, metode *Page-Hinkley* mengidentifikasi penyimpangan data ketika terjadi penurunan temperatur. Berdasarkan penelitian yang telah dilakukan masih terdapat kekurangan yang dapat dikembangkan kemudian hari seperti pada tahapan evaluasi data dapat menggunakan metode-metode terbaru yang dikhususkan untuk data stream sehingga hasil yang didapatkan lebih

baik. Selain itu pada tahapan pengidentifikasian penyimpangan data dapat menggunakan metode-metode terbaru atau metode yang lebih akurat dalam mengidentifikasi penyimpangan data.

DAFTAR PUSTAKA

- [1] K. Asthon, "That ' Internet of Things ' Thing," *RFID J.*, p. 4986, 2010.
- [2] P. K. Srimani and M. M. Patil, "Performance analysis of Hoeffding trees in data streams by using massive online analysis framework," *Int. J. Data Mining, Model. Manag.*, vol. 7, no. 4, pp. 293–313, 2015.
- [3] A. Bifet and R. Gavaldà, "Adaptive Parameter-free Learning from Evolving Data Streams," *Adv. Intell. Data Anal. VIII*, no. September, pp. 249–260, 2017.
- [4] H. Robbins and S. Monro, "A Stochastic Approximation Method," *Ann. Math. Stat.*, vol. 22, no. 3, pp. 400–407, 1951.
- [5] Y. Liu, Y. Zhou, S. Wen, and C. Tang, "A Strategy on Selecting Performance Metrics for Classifier Evaluation," *Int. J. Mob. Comput. Multimed. Commun.*, vol. 6, no. 4, pp. 20–35, 2014.
- [6] D. M. W. Powers, "Evaluation : From Precision , Recall and F-Factor to ROC , Informedness , Markedness & Correlation," no. December, 2007.
- [7] A. Bifet and R. Gavaldà, "Learning from time-changing data with adaptive windowing," *Proc. 7th SIAM Int. Conf. Data Min.*, no. April, pp. 443–448, 2007.
- [8] E. S. Page, "Continuous Inspection Schemes," *Biometrika*, vol. 41, no. 1/2, pp. 100–115, Oct. 1954.