

ABSTRACT

Alessandro Luiz Kartika (01082170029)

IDENTIFICATION DATA DRIFTING ON INTERNET OF THINGS (IoT) APPLICATION

(xii + 68 pages: 36 figures; 2 tables, 1 appendices)

Alongside with the development technology at once increase the need for actual & tangible information, IoT application has been developed and become popular in the community. The ability of IoT to provide actual and tangible information increases the level of confidence of IoT users in implementing things that require actual and tangible information, such as temperature, humidity, etc. However, several factors can cause a data drifting/concept drift in the process of collecting data, such as IoT sensor off, error, etc. Based on the aforementioned, there's a need for research on evaluating stream data and identify concept drift.

There are 3 processes in the research. Firstly, the process of data collecting from an IoT sensor-based Thingier.io that has been plugged in the area of the UPH campus. Secondly, the process of evaluating the data with the methods intended for streaming data, such as Hoeffding Tree, Hoeffding Adaptive Tree, and SGD Classifier. In the process of evaluating the data, several metrics have been the focus of this research, such as accuracy, precision, recall, and kappa. Thirdly, the process of identifying a data drifting/concept drift uses Adaptive Windowing (ADWIN) and Page-Hinkley.

This research produces two conclusions. Firstly, in the process of evaluating the data, SGD Classifier has been outperforming on the average value in every metric compared with other methods. The result of SGD Classifier are accuracy (0,5090), precision (0,5153), recall (0,5154), and kappa (0,0178). Lastly, in the process of identifying data drifting/concept drift, Adaptive Windowing (ADWIN) has been more identify data drifting/concept drift than Page-Hinkley. However, Adaptive Windowing (ADWIN) is more tend to identify data drifting/concept drift when the trend was in enhancement. Meanwhile, Page-Hinkley is more tend to identify data drifting/concept when the trend was in derivation.

Reference: 16 (1951 – 2017)

ABSTRAK

Alessandro Luiz Kartika (01082170029)

IDENTIFIKASI DATA DRIFTING PADA APLIKASI INTERNET OF THINGS (IoT)

(xii + 68 halaman: 36 gambar; 2 tabel; 1 lampiran)

Seiring dengan perkembangan teknologi sekaligus meningkatnya kebutuhan akan informasi yang aktual dan nyata, aplikasi IoT pun terus dikembangkan dan menjadi marak di tengah masyarakat. Kemampuan IoT dalam memberikan informasi secara aktual dan nyata yang membuat meningkatnya tingkat kepercayaan pengguna IoT dalam mengimplementasikan suatu hal yang membutuhkan informasi yang bersifat aktual dan nyata, seperti temperatur, kelembaban udara, dll. Akan tetapi ada beberapa faktor yang dapat menyebabkan terjadinya penyimpangan data/*concept drift* pada proses pengambilan data pada sensor IoT, seperti sensor IoT mati, dll. Berdasarkan hal tersebut, perlunya penelitian yang dapat mengevaluasi data *stream* yang telah ditangkap sensor dan mengidentifikasi penyimpangan data/*concept drift*.

Terdapat tiga tahapan dalam perancangan penelitian, tahapan pertama adalah pengambilan data yang berasal dari seperangkat alat IoT berbasis *Thingier.Io* yang terpasang di lingkungan kampus UPH. Tahapan kedua adalah mengevaluasi data dengan metode yang dikhususkan untuk mengolah data yang bersifat *streaming*, yakni *Hoeffding Tree*, *Hoeffding Adaptive Tree*, dan *SGD Classifier*. Dalam tahapan evaluasi data *stream* hanya berfokus pada metrik-metrik, seperti *accuracy*, *precision*, *recall*, dan *kappa*. Tahapan ketiga adalah mengidentifikasi penyimpangan data/*concept drift* dengan menggunakan metode *Adaptive Windowing (ADWIN)* dan *Page-Hinkley*.

Penelitian ini menghasilkan dua kesimpulan, yaitu hasil pertama adalah pada tahapan evaluasi data, metode *SGD Classifier* yang memiliki nilai rata-rata tertinggi dibandingkan kedua metode yang lain pada setiap metrik. Nilai rata-rata pada setiap metrik dengan metode *SGD Classifier* yakni *accuracy* (0,5090), *precision* (0,5153), *recall* (0,5154), dan *kappa* (0,0178). Hasil kedua adalah pada tahapan pengidentifikasian penyimpangan data/*concept drift*, metode *Adaptive Windowing (ADWIN)* lebih banyak mengidentifikasi penyimpangan data/*concept drift* dibandingkan dengan metode *Page-Hinkley*. Akan tetapi metode *Adaptive Windowing (ADWIN)* cenderung mengidentifikasi penyimpangan data/*concept drift* ketika terjadi peningkatan temperature, sedangkan metode *Page-Hinkley* mengidentifikasi penyimpangan data/*concept drift* ketika terjadi penurunan temperatur.

Referensi: 16 (1951 – 2017)