

## ABSTRACT

Farrell Nathaniel (01082170002)

### **IDENTIFIKASI TERJADINYA PENYIMPANGAN PADA LINGKUNGAN VIRTUAL DENGAN METODE ADAPTIVE WINDOWING**

(xiv + 52 pages; 14 figures, 18 tables, 1 appendices)

Concept Drift is a phenomena in data science which occurs when data changes in an unexpected way due to user or operational activity. This occurrence has to be observed at all times, especially on data stream, case example, a scalability process in cloud computing where scaling happens in real-time, data needs to be identified whether a drift happens when a scaling activity is done. Any relation between VM scaling process and drifting data will be observed. Hardware virtualization process could have caused the system to output inaccurate data.

The study will use CloudSim Plus, a fork of CloudSim 3. Using the provided simulation Vertical VM Scaling, the data will be outputted to a .csv file and the text will be cleansed. Once only the CPU Usage values remain, the data will be tested on drift detection algorithm ADWIN, provided by Python module *scikit-multiflow*.

After being tested on ADWIN, drift is detected on simulation data. Analysis found that after VM switches Cloudlet, system will start outputting 0% values. This is caused by system having to wait for the simulation to finish switching the Cloudlet, and the system end up having to wait until the next iteration to output the next data.

Reference: 14 (2007 – 2021)

## ABSTRAK

Farrell Nathaniel (01082170002)

### **IDENTIFIKASI TERJADINYA PENYIMPANGAN PADA LINGKUNGAN VIRTUAL DENGAN METODE ADAPTIVE WINDOWING**

(xiv + halaman; 14 gambar; 18 tabel; 1 lampiran)

*Concept Drift* adalah fenomena dalam *data science* yang terjadi ketika data berubah seiring waktu karena suatu aktivitas pengguna atau operasional. Hal ini perlu diamati pada data stream, pada contoh kasus proses skalabilitas dalam *cloud computing* di mana proses skalabilitas terjadi secara *real-time*, maka perlu diidentifikasi apakah terjadi penyimpangan pada data ketika aktivitas ini dilakukan. Juga akan diperhatikan jika adakah hubungan antara proses skalabilitas VM dan data yang menyimpang. Proses virtualisasi *hardware* dapat menyebabkan sistem tidak mengeluarkan data yang akurat.

Penelitian ini akan menggunakan CloudSim Plus, fork dari CloudSim. Menggunakan contoh simulasi scaling VM secara vertikal, data akan dioutput ke .csv lalu dicek output textnya. Setelah hanya tersisa nilai penggunaan CPU dari scaling vertikal, data diuji dengan algoritma deteksi drift ADWIN, menggunakan module Python *scikit-multiflow*.

Setelah diuji dengan ADWIN, *drift* terdeteksi dalam data simulasi. Dianalisa bahwa ketika terjadi pergantian Cloudlet dalam VM, sistem tidak mengeluarkan data yang benar, tetapi banyak mengeluarkan data bernilai 0% yang terjadi setelah sistem melakukan pergantian Cloudlet. Ini disebabkan oleh sistem yang terlambat mencatat nilai penggunaan CPU, karena simulasi harus menunggu pergantian Cloudletnya selesai sebelum memunculkan data selanjutnya.

Referensi: 14 (2007 – 2021)