

# Rancang Bangun Infrastruktur Big Data pada Institusi Pendidikan Tinggi Multi Kampus

<sup>1)</sup> Daniel Yeri Kristiyanto

STEKOM, Jl. Majapahit No.605, Pedurungan Kota Semarang, Jawa Tengah, Indonesia

E-mail: [daniel.jerry182@gmail.com](mailto:daniel.jerry182@gmail.com)

<sup>2)</sup> Bambang Suhartono

STEKOM, Jl. Majapahit No.605, Pedurungan Kota Semarang, Jawa Tengah, Indonesia

E-mail: [bambang@stekom.ac.id](mailto:bambang@stekom.ac.id)

## ABSTRACT

Nowadays is referred to as the era of big data where every individual or organization produces large amounts of data through various digital devices used. Higher education institutions continuously produce data and only store it in various data formats and physical files, thus creating massive raw data. Collection of data created from a long time produces big data, so it is difficult to be treated manually or processed using conventional data processing applications. Higher education is an educational institution that always produces data from time to time continuously through various affairs in each of its parts. Multi campus institutions have physical locations that are far apart from one to another with more than one physical infrastructure. Each campus branch produces its own data, this requires a thought about how to build a large data infrastructure, so that stakeholders can use various data from each branch of the campus for the analysis process of various needs of tertiary institutions. Multi campus big data infrastructure is the focus of this research, through the stages of big data sampling on multiple campuses using statistical and prototyping methods, so that data flow and data unity can be standardized in all branches of the campus by minimizing infrastructure constraints that are based on socioculture and spatial aspects.

**Keyword – Big Data, Sampling Techniques, Prototyping, Infrastructure, Multi Campus**

## PENDAHULUAN

Big data merupakan issue terkini dan telah menjadi fenomena yang telah masuk ke setiap aktifitas pribadi bahkan menjadi tren disetiap industri maupun institusi. Institusi pendidikan tinggi tidak terlepas dari trend big data pada saat ini. Kemajuan dalam bidang data yang mampu memberikan manfaat pada masa depan mampu mempengaruhi sistem pendidikan tinggi. Adopsi pemanfaatan big data pada level industri, serta dorongan pemerintah untuk memperkenalkan teknologi industri 4.0 secara langsung maupun tidak langsung memaksa penyelenggara perguruan tinggi untuk mengadopsi pemanfaatan teknologinya[1]. Penyelenggara perguruan tinggi kemungkinan telah sukses menyelenggarakan sistem pendidikannya menggunakan metode yang telah sejak lama digunakan untuk mengatur institusinya, yakni mengatur sistem pendidikan, mengumpulkan data, membersihkan data, dan mendistribusikan sejumlah besar data set aktivitas organisasi. Namun, ketika data tersebut dari waktu ke waktu menjadi data yang besar, tidak semua perguruan tinggi mampu menangani big data dengan alasan kekurangan sumber daya manusia dan infrastruktur teknologi big data. Adopsi teknologi big data pada institusi pendidikan tinggi menjadi salah satu keunggulan kompetitif dalam proses collecting dan analyzing sejumlah dataset

pada aktivitas multi disiplin yang bersifat scientific[2]. Perhatian utama dalam penyelenggaraan pendidikan tinggi adalah kemampuan untuk menghubungkan berbagai elemen yang berpotensi menghasilkan data besar dan dalam jangka waktu yang lama yang selaras dengan visi dan misi institusi pendidikan tinggi. Fokus utama penggunaan aplikasi big data pada institusi pendidikan tinggi adalah sebagai penghubung antara data yang besar dengan tujuan jangka panjang institusi dalam melayani stakeholdernya, sebagai contoh untuk mendukung tingkat kinerja pegawai, sebagai penampung pemikiran kritis atau ide-ide murni yang berasal dari civitas akademi, atau penelitian serta pengabdian yang telah dilakukan dan akan dilakukan untuk masyarakat. Penggunaan big data diharapkan oleh para alumnus untuk mencari pekerjaan setelah lulus dari perguruan tinggi, sedangkan perguruan tinggi dapat memanfaatkannya untuk sarana traceret studi. Gabungan dari kondisi ini dapat digunakan semaksimal mungkin untuk terus memanfaatkan data besar yang mampu dimanfaatkan dengan teknologi big data.

### A. Multi Kampus

Multi kampus memiliki pengertian yakni bahwa pelaksanaan proses belajar mahasiswa dan dosen dilaksanakan tidak hanya pada satu tempat atau

satu lokasi saja. Terdapat beberapa institusi pendidikan tinggi memiliki lokasi perkuliahan ada pada dua tempat atau lebih. Kebijakan pembangunan infrastruktur multi kampus berupa gedung perkuliahan diambil oleh pengelola perguruan tinggi dengan berbagai alasan yang disesuaikan dengan visi dan misinya. Kebijakan tersebut masuk ke dalam bagian strategi perguruan tinggi untuk dapat sedekat mungkin dengan konsumennya yakni calon mahasiswa. Pembangunan infrastruktur pada multi kampus terkait erat dengan aspek fisik-spasial, pengadaan sarana dan prasarana kuliah, penataan lanskap dan estetika, yang seluruhnya ditujukan untuk kenyamanan, keamanan yang mendukung proses perkuliahan secara maksimal. Keberadaan multi kampus memiliki keunggulan mengenai berbagai hal, misalkan saja adalah pengurangan beban penggunaan ruangan dan alat yang digunakan setiap harinya, suasana belajar sangat dinamis sesuai dengan tempat dan suasana yang tidak membosankan bagi mahasiswa, penyelenggara pendidikan dapat langsung mengetahui jumlah peserta didik pada masing-masing cabang, kampus cabang relatif lebih dekat dengan tempat tinggal peserta didik, yang secara otomatis mampu memangkas biaya harian yang dikeluarkan oleh mahasiswa. Staf pengajar atau dosen dapat dipekerjakan pada lokasi kampus cabang yang dekat dengan tempat tinggalnya sehingga hal ini dapat memangkas biaya variabel yang dikeluarkan oleh karyawan. Selain keunggulan model multi kampus tersebut tentu ada kelemahan keberadaan sistem multi kampus diantaranya: ketidak samaan jumlah sarana dan prasarana pada satu kampus, akan mengakibatkan kesulitan mahasiswa dan dosen memulai perkuliahan, sumber daya manusia yang terbatas yang dimiliki oleh penyelenggara pendidikan tinggi dapat mengakibatkan ketimpangan jumlah pengajar dan mata kuliah yang mampu di laksanakan pada sebuah kampus cabang, apabila sebuah kampus cabang tidak memiliki staf pengajar yang sesuai bidang maka biasanya penyelenggara pendidikan akan mengambil dari cabang lain yang justru mengakibatkan dosen atau staf pengajar bertambah biaya operasionalnya, urusan berbagai bidang mengakibatkan seorang pimpinan belum tentu dapat ditemui pada sebuah kampus cabang, hal ini menyulitkan apabila hendak mengurus administrasi yang melibatkan lebih dari satu orang tenaga pendidik maupun tenaga kependidikan, misalkan saja urusan tanda tangan pengesahan skripsi, atau tugas akhir.

## B. Adopsi Big Data Pada Multi Kampus

Adopsi teknologi big data beserta analisisnya pada institusi pendidikan tinggi multi kampus merupakan sesuatu yang relatif baru. Adopsi penggunaan big data akan sangat menarik pada multi kampus

dengan lebih dari 5 cabang, bahkan sampai 10 cabang kampus yang dimiliki oleh penyelenggara pendidikan. Uniknya berbagai kampus cabang tersebut tidak melanggar kaidah kelas jauh sebagaimana yang tertuang pada pasal 56 PP Nomor 30 Tahun 1990. Hal menarik lainnya adalah bagaimana halnya jika teknologi big data diaplikasikan untuk menangkap aktivitas data di seluruh kampus cabang tersebut yang jumlahnya melebihi 5. Infrastruktur dari seluruh kampus cabang perlu untuk di observasi sehingga adopsi big data ke dalam institusi pendidikan tinggi multi kampus dapat dilakukan. Hal menarik lainnya adalah bagaimana mengintegrasikan berbagai data penting pada setiap cabang menggunakan teknologi big data, sehingga aspek geografis dan spasial dapat diatasi melalui integrasi dari setiap kampus cabang[3]. Adopsi big data pada perguruan tinggi multi kampus dapat dimanfaatkan melalui data analytics untuk berbagai keperluan yang berkaitan dengan pengambilan keputusan. Produksi data pada berbagai departemen yang dimiliki sebuah perguruan tinggi dengan konsep multi kampus sedapat mungkin dapat dimaksimalkan untuk keunggulan kompetitif berdasarkan data-data yang telah dihasilkan.

Data analitik merupakan perkembangan ilmu mengenai data yang lebih disempurnakan dari teknologi data mining sebagai pendahulunya. Tren aplikasi berbasis online telah menciptakan data yang besar yang tak mampu lagi di selesaikan dengan metode data mining. Big data teknologi dapat digunakan untuk pengambilan keputusan pada institusi pendidikan tinggi multi kampus melalui data analitik, artinya bahwa untuk dapat mempelajari sebuah tren atau kecenderungan sebuah data yang besar dan bertipe time series maupun raw data maka langkah data analitik merupakan jalan yang ditempuh untuk menyelesaikan berbagai permasalahan terkait penggunaan data[4]. Selain untuk menyelesaikan permasalahan internal, data analitik dapat pula digunakan sebagai jurusan atau bahkan program studi yang baru, sehingga aplikasi antara aspek teoritikal dengan aspek aplikatif dapat terus berkesinambungan sehingga menciptakan sebuah keuntungan tersendiri bagi institusi pendidikan tinggi.

## C. Pemodelan Big Data Menggunakan Metode Prototype

Big data dapat didefinisikan sebagai data yang kompleks dari berbagai segi seperti volume, variasinya, kecepatannya, serta hubungannya dengan data lain. Aplikasi big data memiliki keunikan dibandingkan dengan aplikasi pengolah data lainnya, hal ini dapat dilihat dari segi kompleksitas dalam menangani data besar serta kemampuan memadukan berbagai instrumen yang terdapat pada aplikasi basis data

konvensional. Data besar yang telah memiliki standart tertentu dalam sistem basis data disebut sebagai data set. Data set berisi sejumlah data yang besar dan memiliki keterlibatan langsung dalam proses keamanan data, analisis data, dan interpretasi dan potensi inovasi data sebuah institusi[5]. Kemajuan inovasi penggunaan big data dapat dimunculkan melalui kerangka kerja dan sistem baru dalam manajemen organisasi perguruan tinggi multi kampus.

Pengambilan keputusan yang tepat dihasilkan oleh kemampuan analitik yang baik. Ketepatan dan kecepatan merupakan faktor yang sangat penting dalam pengolahan data set di dalam big data multi kampus, sehingga harapan kenyamanan dapat terwujud yang disesuaikan dengan pelayanan kampus. Big data melibatkan berbagai multidisiplin dan berbagai kombinasi inovasi baru penimbunan dan penanganan data yang lazim didalam big data analitik disebut sebagai volume dan speed. Big data membahas pula mengenai periode variasi sumber data dan tantangan mengelola kualitas sebuah informasi secara cepat dan tepat yang disebut dengan veracity[6].

## METODE

Penelitian ini menggunakan metode prototype dengan menyertakan model fisik kerja sistem dan fungsinya sebagai versi awal dari penerapan sistem. Pemodelan prototype dalam infrastruktur big data multi kampus menggambarkan beberapa elemen yang akan dikaitkan kedalam sistem big data melalui sarana atau media. Peneliti membagi kedalam dua bagian sistem prototyping, yakni prototype fisik dan non fisik.

Prototype fisik membahas mengenai aspek hardware yang dibutuhkan, sedangkan prototype non fisik membahas mengenai aspek software yang digunakan dalam big data. Tujuan penelitian ini adalah melaksanakan simulasi dari metode prototype sebelum diterapkan pada perguruan tinggi multi kampus. Metode prototype memiliki tujuan mengumpulkan sebanyak mungkin data dan informasi dari stakeholder atau pemangku kepentingan perguruan tinggi multi kampus, pemodelan prototype yang dihasilkan memberikan gambaran mengenai versi awal dari sebuah penerapan sistem informasi untuk menjamin keberlanjutan sebuah sistem informasi berbantuan komputer, sehingga pengembangannya tetap terus dapat dilaksanakan seiring berkembangnya kemajuan jaman serta perkembangan teknologi informasi[7]. Metode prototype secara aktif melibatkan seluruh bagian yang terlibat untuk menentukan dan memastikan fungsi, tujuan, dan kebutuhan sistem infrastruktur big data dapat diimplementasi dengan baik.

Pemodelan prototyping dilaksanakan melalui beberapa tahapan yang melibatkan beberapa

aspek yang berhubungan dengan implementasi teknologi big data pada multi kampus, adapun langkah-langkah yang ditempuh adalah sebagai berikut[8]:

### 1. Identifikasi Kebutuhan

Proses pengumpulan seluruh elemen yang terlibat ke dalam implementasi penerapan infrastruktur big data pada multi kampus. Elemen yang terlibat adalah seluruh stakeholder multi kampus. Identifikasi kebutuhan disesuaikan dengan tujuan implementasi aplikasi[9].

### 2. Proses Desain

Proses desain berfokus kepada hirarki sistem informasi yakni mencakup input, proses, output dan feedback[10]. Proses desain mempermudah user, administrator, serta super user memiliki hirarki yang jelas didalam sistem dengan tujuan integritas data.

### 3. Membuat Model Prototype

Proses pembuatan model prototype dimaksudkan untuk menyederhanakan sejumlah pengumpulan data yang kompleks melalui teknik sampling big data, sehingga mampu untuk dirancang, diimplementasi dengan sebaik mungkin, dengan sumber daya yang seminimal mungkin, sehingga kebutuhan pengguna dan pengembang dapat diakomodir[11].

### 4. Evaluasi Pengujian Sistem dan Perbaikan Kesalahan

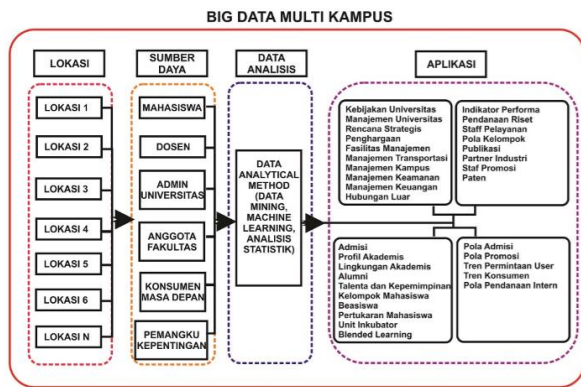
Tahap evaluasi pengujian sistem dan perbaikan kesalahan merupakan tahapan untuk mengukur sejauhmana sistem telah dikerjakan dan diterapkan, pada proses implementasi big data multi kampus evaluasi pengujian sistem dilakukan dengan memperhatikan komponen yang telah ditentukan sebelumnya apabila terdapat kesalahan dapat segera dilakukan perbaikan atau bahkan penambahan variabel dengan tetap memperhatikan tujuan dan aspek fungsionalitas sistem[12].

### 5. Implementasi Sistem

Implementasi sistem dilakukan apabila telah mencapai kata sepakat antar stakeholder dan telah mencapai tingkat fungsionalitas sistem yang selaras dengan tujuan sistem big data untuk diterapkan di multi kampus. Tahap implementasi juga memperhatikan aspek keberlanjutan jangka panjang[13].

## HASIL DAN PEMBAHASAN





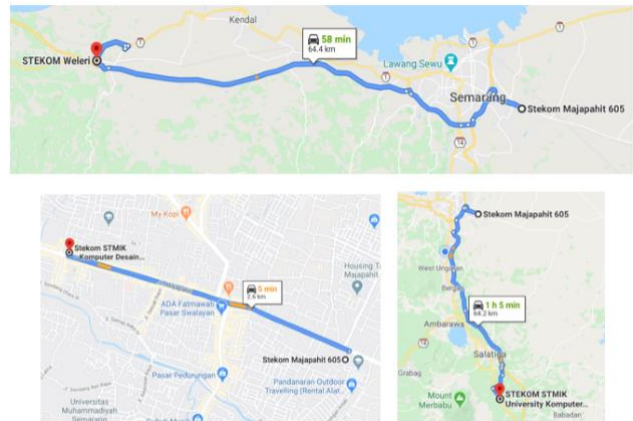
**Gambar 1. Desain Kerangka Kerja Big Data Multi Kampus**

### A. Identifikasi Kebutuhan, Proses Desain dan Model Prototype

Produksi data pada sistem multi kampus utamanya bergantung kepada data mahasiswa yang menjadi sumber data primer. Keputusan yang bersifat kritis maupun keputusan yang bersifat strategis utamanya ditujukan untuk memenuhi segala kebutuhan akademik mahasiswa. Data mulai dikumpulkan oleh perguruan tinggi sejak berdirinya institusi tersebut, meskipun pengelolaannya masih bersifat konvensional. Data yang dikumpulkan dan disimpan digunakan kembali, melalui berbagai proses yang terjadi selama proses penerimaan mahasiswa baru hingga mahasiswa lulus dari perguruan tinggi tersebut. Berbagai macam data tercipta dari proses mahasiswa belajar di kampus seperti halnya aplikasi data, perkuliahan, data registrasi, data kehadiran, data pembelajaran online, data kinerja, data kegiatan mahasiswa ekstrakurikuler, data magang mahasiswa, data kemampuan akademis, data kinerja penerimaan mahasiswa. Adopsi aplikasi Big Data dapat membantu institusi perguruan tinggi multi kampus meningkatkan pengalaman belajar di seluruh program studi yang ada, membantu mengurangi mahasiswa putus kuliah, serta meningkatkan angka kelulusan mahasiswa.

Agar sesuai dengan pengaplikasian big data pada multi kampus, maka digunakan kerangka kerja big data multi kampus seperti pada gambar 1. Kerangka kerja tersebut menunjukkan indikasi sejumlah proses transformasi data yang besar yang tercipta pada lingkungan multi kampus yang berasal dari berbagai sumber departemen maupun data yang berasal dari masing-masing lokasi multi kampus yang mampu untuk dianalisis menggunakan teknologi big data. Melalui implementasi big data pada multi kampus, analisis sejumlah data yang kompleks tersebut dapat dikategorikan pada aplikasi yang mencakup seluruh manajemen dan pelanggan dari institusi pendidikan tinggi multi kampus hanya dengan klik pada aplikasi. Manajemen dan analisis data yang

baik menghasilkan kepuasan pelanggan yang semakin meningkat.



**Gambar 2. Beberapa Gambaran Multi Kampus Berdasar Jarak**

Pengolahan data didalam teknologi big data merupakan proses pengumpulan nilai dan berbagai faktor yang saling berkaitan dan terkadang dapat menimbulkan pengertian yang berbeda. Namun, bagaimanapun juga ukuran basis data yang digunakan dalam teknologi ini dapat berkembang secara cepat dari waktu ke waktu. Implementasi big data diperlukan manakala data yang diolah pada sebuah institusi telah mencapai volume yang besar, memiliki variety yang beragam dan dibutuhkan velocity untuk dapat mengaksesnya dimana hal ini tak dapat dilakukan oleh instrumen-pengolah database konvensional. Identifikasi kebutuhan implementasi big data pada multi kampus terlihat pada aplikasi gambar 1. Data yang dikelola melalui aplikasi big data menjadi penting apabila data tersebut selalu dijaga keamanannya, dijaga transaksinya, mampu memberikan persepsi, dan memerlukan teknologi terkini dalam menanganinya. Jarak yang bervariasi antar cabang kampus memerlukan formulasi khusus dalam menangani integrasi data besar. Berbagai macam jenis data memicu meningkatnya varian data, sehingga diperlukan teknologi yang mampu menyeragamkan format data.

Varian data yang berasal dari transaksi multi kampus memiliki beberapa distribusi probabilitas yang bermuara kepada suatu populasi data yang disimpan dalam sistem big data. Berdasar kerangka kerja sebuah sampel data yang berasal dari berbagai cabang kampus memiliki sistem terdistribusi yang dapat digunakan untuk membentuk sebuah estimasi varian, sehingga pemodelan prototype dapat dibentuk. Sebuah formulasi sintetik diperlukan untuk mengatasi masalah big data yang kompleks dan besar. Ragam data di dalam adopsi big data multi kampus berpotensi memiliki sejumlah besar outlier. Pemodelan prototype yang digunakan adalah melalui pendekatan top down untuk desain metrik menggunakan map reduce framework.

Pendekatan top down dimaksudkan untuk mendekati desain matrik dari data ke teori. Pengaplikasian hadoop dapat digunakan dalam prototype adopsi big data pada multi kampus. Pemodelan prototype pada aplikasi big data multi kampus berdasar pada pemikiran bahwa data yang dikelola memiliki volume yang besar di berbagai departemen, dalam analisisnya sangat penting untuk memperhatikan bias seleksi data. Dalam tulisan ini penulis mengusulkan mengadopsi metode untuk mengurangi bias yang terkait sampel data pada multi kampus menggunakan invers sampling dengan memasukkan alat bantu informasi yang berasal dari data eksternal, yang kedua adalah mengintegrasikan data dengan teknik penggabungan sampel data besar dengan metode probabilitas independen. Tujuannya adalah menekan bias data dan dapat memiliki tingkat cakupan yang lebih baik. Identifikasi populasi finite  $\{y_i : i \in U\}$ , dimana  $y_i$  dimana  $i$  merupakan variabel yang diobservasi dan  $U$  yakni  $\{1, \dots, N\}$ , dimana  $N$  yang merupakan koresponding index. Big data sample dapat diinisiasi sebagai berikut  $\{y_i : i \in B\}$  dimana  $B \in U$ . khusus  $\delta_i = 1$  if  $i \in B$  dan  $\delta_i = 0$ , jika tidak maka penulis mengasumsikan  $y_i$  dapat diobservasi hanya jika  $\delta_i = 1$ . Estimasi dalam penelitian ini adalah mengenai mean populasi  $\bar{Y}_N = N^{-1} \sum_{i=1}^N Y_i$ . Berdasar kepada sample big data pada pengelolaan data multi kampus, dapat diestimasi sebagai  $\bar{Y}_N$  sama dengan  $\bar{Y}_B$  sehingga dapat ditulis  $N^{-1} \sum_{i=1}^N \delta_i Y_i$  dimana  $N_B = \sum_{i=1}^N \delta_i$  yang dikenal sebagai ukuran data B. Standar error  $\bar{Y}_B$  dapat ditulis melalui inisialisasi  $\{\delta_i : i \in U\}$  sehingga dapat dituliskan sebagai berikut  $\bar{Y}_B - \bar{Y}_N = \frac{1}{f_B} \text{Cov}(\delta, Y)$ , dimana  $f_B = N_B/N$  dan  $\text{Cov}(\delta, Y) = \sum_{i=1}^N (\delta_i - \delta_N)(y_i - \bar{Y}_N)$  dengan  $\delta_N = N^{-1} \sum_{i=1}^N \delta_i$  jadi secara keseluruhan persamaan dapat ditulis  $E\delta \{(\bar{Y}_B - \bar{Y}_N)^2\} = \frac{1}{f_B^2} \{Cov(\delta, Y)\}^2$  dimana  $E\delta$  mampu menunjukkan ekspektasi hubungan acak dari sejumlah big data.

Apabila mekanisme random untuk  $\delta_i$  berbasis kepada Bernoulli Sampling, dimana secara inklusi berbagai indikator merujuk kepada teknik Bernoulli dengan  $f_B$  independen, maka dapat dimaksimalkan dengan formula  $E\delta \{Cov(\delta, Y)\} = [E\delta \{Cov(\delta, Y)\}]^2 + \text{Var}\delta \{Cov(\delta, Y)\} = 0 + \frac{1}{N^2} \sum_{i=1}^N (y_i - \bar{Y}_N)^2 f_B (1-f_B) = \frac{1}{N} f_B (1-f_B) \sigma^2$  dengan ketentuan  $\sigma^2 = N^{-1} \sum_{i=1}^N (y_i - \bar{Y}_N)^2$ . jadi Bernoulli sampling dapat di sederhanakan  $E\delta \{(\bar{Y}_B - \bar{Y}_N)^2\} = \frac{1}{N_B} (1-f_B) \sigma^2$ , yang secara konsisten secara teori merujuk kepada Bernoulli sample dengan ukuran sampel  $n = N_B$ . Sehingga persamaan ke dua, pengambilan sample untuk multi kampus adalah  $E\delta \{(\bar{Y}_B - \bar{Y}_N)^2\} = \frac{1}{f_B} E\delta \{Cov(\delta, Y)^2 \text{Var}(\delta) \text{Var}(Y)\} = E\delta \{Cov(\delta, Y)^2\} \times (\frac{1}{f_B} - 1) \times \sigma^2$ , dimana

persamaan kedua diperoleh dari  $\text{Var}(\delta) = \frac{1}{N} \sum_{i=1}^N (\delta_i - \delta_N)^2 = f_B (1 - f_B)$ . Penentuan seleksi bias dari  $\bar{Y}_B$ , yakni  $E\delta \{Cov(\delta, Y)^2\}$ , yang merupakan titik kritis dimana disebut sebagai Data Defect Index (DDI), yang menentukan data awal dari simple random sampling, dibawah desain sample probabilitas yang sama sehingga  $E\delta (E_i) = f_B$ .  $E\delta \{Cov(\delta, Y)\} = 0$  dan DDI stabil pada  $O(1/N)$ , yang berimplikasi  $E\delta \{(\bar{Y}_B - \bar{Y}_N)^2\} = O(N_B^{-1})$ . Untuk desain sampling yang lain  $E\delta \{Cov(\delta, Y)\} \neq 0$ , DDI menghasilkan signifikansi  $O(1)$ , yang berimplikasi kepada  $E\delta \{(\bar{Y}_B - \bar{Y}_N)^2\} = O(N_B^{-1} N - 1)$ . Oleh sebab itu penulis tidak menggunakan sebuah non-probability sampling dengan desain  $E\delta \{Cov(\delta, Y)\} \neq 0$ , karena akan menyebabkan analisis subjek menjadi bias.

## B. Evaluasi Pengujian Sistem dan Perbaikan Kesalahan

Ketika distribusi variabel penelitian untuk sampel big data multi kampus diambil dengan cara berbeda dalam hal sistematiknya, maka dapat menghasilkan interpretasi yang berbeda dari populasi yang ditargetkan, sampel big data tidak selalu mewakili target populasi. Variasi data yang besar akan berpotensi memiliki bias yang tinggi, sehingga memperbesar potensi implementasi big data pada multi kampus, maka penulis mencoba menggunakan variabel bantu eksternal untuk sampel big data, sehingga bias data dapat di koreksi atau diminimalkan. Solusi yang penulis tawarkan adalah menggunakan pendekatan sampling inverse. Fase pertama adalah menggunakan sampel big data pada multi kampus yang memiliki potensi bias seleksi. Fase kedua adalah menciptakan sub sampel dari fase pertama yang berfungsi untuk melakukan iterasi pada sample fase pertama. Teknik inverse sampling pada permulaannya diusulkan sebagai sebuah metode untuk mendapatkan sampel acak sederhana dari sampel yang diperoleh dari desain pengambilan sampel yang lebih kompleks. Beberapa desain klasik, seperti pada teknik stratified sampling, algoritma inverse sampling dikemukakan oleh Hinkins et all dan Rao et all. Pada big data multi kampus penulis menggunakan inverse sampling untuk menyeleksi bias pada rancang bangun big data multi kampus.

Koreksi bias data menggunakan pendekatan inverse sampling yang diusulkan, penulis menggunakan eksternal informasi mengenai populasi target, baik dari cara sensus maupun dari sampel probabilitas untuk beberapa variabel bantu  $\chi$ . Penulis mendesain formulasi  $(\chi_i, y_i)$  dimana pada desain prototype big data multi kampus diinisiasi sebagai sampel (B) dan  $f(\chi)$  akan menjadi densitas untuk distribusi marginal  $\chi$  yang diperoleh dari sumber eksternal. Penulis mengasumsikan bahwa variabel bantu  $\chi$  memiliki momen kedua

yang terbatas, maka menarik apabila estimasi dibuat ke dalam  $\Theta = E(Y)$  yang berasal dari big data sample B. inklusi urutan pertama probabilitas untuk sampel data besar B tidak diketahui. Apabila ukuran sampel B terlalu besar, formulasi komputasi untuk menciptakan kalibrasi ekuasi menjadi sangat rumit dan sulit untuk dilakukan. Solusi one-step approximation dapat digunakan dalam kondisi ini. Berbasis kepada fase ke dua sampe B yang diinisialisasi menjadi (B2) dari variabel n dari big data sample B dimana  $\Theta B2 = n^{-1} \sum_{i \in B2} y_i$  yang merupakan formulasi estimasi desain tidak bias dari  $\Theta B1$ . Ide dasar memilih kondisional first-order inklusi probabiliti adalah  $\pi_2|1 = P(i \in B2 | i \in B)$  sama seperti  $\pi_2|1 = n W_i1, i \in B$ , dimana  $W_i1$  merupakan bobot yang penting untuk diperhatikan, maka untuk menjamin  $\pi_2|1 = \epsilon(0,1), i \in B$ , maka formulasi yang digunakan adalah  $n \leq 1 / \max_{i \in B} \{W_i1\}$ . Apabila  $\pi_2|1 = n W_i1, i \in B$  dapat digunakan dan formulasi  $\pi_2|1 = \epsilon(0,1), i \in B$  diketahui, maka fase ke dua untuk penentuan teknik sampling dapat digunakan. Fase ke dua untuk sampel B2 yang dapat diketahui, maka sample mean  $y_i$  pada B2 dapat diestimasi menggunakan  $\Theta$ .

**C. Integrasi Big Data Multi Kampus**

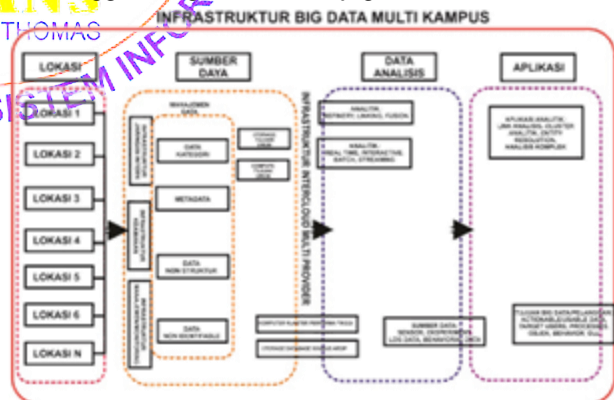
Integrasi big data multi kampus merupakan bidang kajian yang muncul, tujuannya adalah untuk menggabungkan informasi dari dua survei independen yang berasal dari populasi target yang sama. Cara yang diusulkan penulis adalah menggunakan asumsi instrumental variabel menggunakan error model untuk mengkombinasikan dua variabel survey. Survey integrasi data diusulkan dengan mengkombinasikan big data dengan data survey itu sendiri. Penulis mengasumsikan terdapat dua sumber data, yakni survey data dinotasikan sebagai A, dan big data dinotasikan sebagai B, dimana sekaligus sebagai subjek seleksi bias. Selain A dan B terdapat asumsi lain yakni  $\chi$  dan  $y$  yang diidentifikasi berasal dari data survey sementara maupun berasal dari big data multi kampus, dan  $n/NB = o(1)$ , dimana n adalah ukuran sampel A. penulis tertarik untuk dapat melakukan estimasi populasi rata-rata YN dengan cara mengkombinasi dua sumber yang telah diinisialisasi diatas. Oleh karena seleksi bias, maka sample YB yang berasal dari big data multi kampus dianggap bias, sehingga representasi dari struktur datanya adalah sebagai berikut:

**Tabel.1 Struktur Data Big Data Multi Kampus**

Data	Representasi Data	X	Y
A	Ya	✓	-
B	Tidak	✓	✓

Apabila kedua sampel data merupakan sampel probabilistik, maka sistesis data dapat digunakan untuk membuat imputasi nilai dari  $y_i$  pada sample

A. imputasi data bertujuan untuk mendapatkan pola dari suatu data dengan memperkirakan nilai missing pada data tersebut. Permasalahan missing data akan berakibat big data yang diimplementasi pada multi kampus berpotensi menghasilkan outlier yang terlalu besar yang berimplikasi kepada kesulitan mengintegrasikan data. Apabila variabel B merupakan sampel non-probability, maka pendekatan yang digunakan adalah imputasi massal menggunakan imputasi nearest neighbour untuk mengintegrasikan hasil survey. Proses imputasi ini dapat diartikan bahwa variabel x digunakan untuk menemukan nearest neighbour pada sampel big data variabel B, sehingga dapat menghasilkan value imputasi  $y_i$  untuk setiap element dari sampel A. apabila value imputasi  $y_i$  sudah tercipta untuk seluruh elemen sampel A, maka proses komputasi mengenai imputasi dapat dilakukan estimasi dengan  $\Theta = E(Y)$  untuk sampel A. formulasi dapat dibentuk melalui  $fB(y|x) = f(y|x)$  yang merupakan densitas bersyarat dari y ke variabel x untuk sampel big data B dan  $f(y|x)$  merupakan target populasi. Sebagai penerapan integrasi big data pada multi kampus dapat dicontohkan, apabila sampel big data B terdiri dari mahasiswa yang menggunakan cara pembayaran melalui bank tertentu, maka kita dapat memperoleh  $\delta_i$  dari A dengan menanyakan mahasiswa  $i$  yang menggunakan berbagai bank tersebut, sehingga berbagai informasi terkait sebuah item variabel mampu ditampung dan ditemukan korelasinya sehingga integrasi data terjaga.



**Gambar 3. Desain Prototype Infrastruktur Big Data Multi Kampus**

**KESIMPULAN**

Berdasarkan penelitian, implementasi dan pengujian, maka dapat diambil kesimpulan sebagai berikut :

1. Rancang bangun infrastruktur big data multi kampus dimulai dengan menentukan sampel dan formulasi big data, sehingga kebutuhan rancang bangun dapat dioptimalkan dengan baik.
2. Desain prototype sampel big data dapat diformulasikan menggunakan bantuan metode statistika yakni teknik sampling.



3. Prototype rancang bangun infrastruktur big data multi kampus dapat digunakan sebagai landasan membangun infrastruktur fisik pengaplikasian big data pada multi kampus.
4. Formulasi yang dibentuk melalui  $fB(y|x) = f(y|x)$  merupakan densitas bersyarat dari  $y$  ke variabel  $x$  untuk sampel big data  $B$  dan  $f(y|x)$  merupakan target populasi yang kesemuanya ditujukan untuk menjaga integritas data, apabila big data multi kampus diterapkan.

#### DAFTAR PUSTAKA

- [1] I. Kristanti and N. Umamah, "Jurnal Historica THE CHARACTER-BASED MODULES AND THEIR INFLUENCE ON HISTORICAL AWARENESS OF STUDENTS OF CLASS XI MIPA 4 SMAN PASIRIAN," *Hist. Lima.*, vol. 3, no. 2252, pp. 78–89, 2019.
- [2] A. Dukhanov, A. Boukhanovsky, T. Sidorova, and N. Spitsyna, "Big Data and Artificial Intelligence for Digital Humanities: An International Master Program via Trans-Eurasian Universities Network," *Procedia Comput. Sci.*, vol. 101, pp. 449–451, 2016.
- [3] D. Y. Kristiyanto, A. Iriani, S. Yulianto, and J. Prasetyo, "Visualisasi dan Interpretasi Database Engine Website Penilai Kinerja Karyawan Berbasis Online Transaction Processing (OLTP)," in *Prosiding SINTAK 2018*, 2018, no. Mvc, pp. 325–332.
- [4] D. Y. Kristiyanto, B. Suhartono, and A. Wibowo, "Digital Forensic InnoDB Database Engine for Employee Performance Appraisal Application," in *E3S Web of Conferences*, 2019, vol. 125, no. 2019, pp. 1–9.
- [5] S. T. Ng, F. J. Xu, Y. Yang, and M. Lu, "A Master Data Management Solution to Unlock the Value of Big Infrastructure Data for Smart, Sustainable and Resilient City Planning," in *Procedia Engineering*, 2017, vol. 196, no. June, pp. 939–947.
- [6] N. B. C. E. Jamil, I. Bin Ishak, F. Sidi, L. S. Affendey, and A. Mamat, "A Systematic Review on the Profiling of Digital News Portal for Big Data Veracity," *Procedia Comput. Sci.*, vol. 72, pp. 390–397, 2015.
- [7] J. Nelson, A. Berlin, J. Menold, and M. Parkinson, "The role of digital prototyping tools in learning factories," in *Procedia Manufacturing*, 2020, vol. 45, no. 2019, pp. 528–533.
- [8] H. Vestad and M. Steinert, "Creating your own tools: Prototyping environments for prototype testing," in *Procedia CIRP*, 2019, vol. 84, pp. 707–712.
- [9] S. Berhe, M. Maynard, and F. Khomh, "Software Release Patterns When is it a good time to update a software component?," *Procedia Comput. Sci.*, vol. 170, no. 2019, pp. 618–625, 2020.
- [10] L. Li *et al.*, "An integrated hardware/software design methodology for signal processing systems," *J. Syst. Archit.*, vol. 93, no. April 2018, pp. 1–19, 2019.
- [11] R. Bivand and K. Krivoruchko, "Big data sampling and spatial analysis: 'which of the two ladles, of fig-wood or gold, is appropriate to the soup and the pot?'," *Stat. Probab. Lett.*, vol. 136, no. xxxx, pp. 87–91, 2018.
- [12] D. Suleiman, M. Al-Zewairi, and G. Naymat, "An Empirical Evaluation of Intelligent Machine Learning Algorithms under Big Data Processing Systems," in *Procedia Computer Science*, 2017, vol. 113, pp. 539–544.
- [13] F. Fagerholm, A. Hellas, M. Luukkainen, K. Kyllönen, S. Yaman, and H. Mäenpää, "Designing and implementing an environment for software start-up education: Patterns and anti-patterns," *J. Syst. Softw.*, vol. 146, pp. 1–13, 2018.